INSTRUMENT ENGINEERS' HANDBOOK

Fourth Edition

Process Control and Optimization

BÉLA G. LIPTÁK, Editor-in-Chief

Taylor & Francis
Taylor & Francis Group

ISA–The Instrumentation, Systems, and Automation Society

INSTRUMENT ENGINEERS' HANDBOOK

Fourth Edition

Process Control and Optimization

VOLUME II

INSTRUMENT ENGINEERS' HANDBOOK

Fourth Edition

Process Control and Optimization

VOLUME II

Béla G. Lipták

EDITOR-IN-CHIEF

ISA–The Instrumentation, Systems, and Automation Society

Taylor & Francis
Taylor & Francis Group
Boca Raton London New York

A CRC title, part of the Taylor & Francis imprint, a member of the Taylor & Francis Group, the academic division of T&F Informa plc.

This reference text is published in cooperation with ISA Press, the publishing division of ISA--Instrumentation, Systems, and Automation Society. ISA is an international, nonprofit, technical organization that fosters advancement in the theory, design, manufacture, and use of sensors, instruments, computers, and systems for measurement and control in a wide variety of applications. For more information, visit www.isa.org or call (919) 549-8411.

Published in 2006 by
CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-8493-1081-4 (Hardcover)
International Standard Book Number-13: 978-0-8493-1081-2 (Hardcover)
Library of Congress Card Number 2003048453

Library of Congress Cataloging-in-Publication Data

Instrument engineers' handbook / Béla G. Lipták, editor-in-chief.
p. cm.
Rev. ed. of: Instrument engineers' handbook. Process measurement and analysis. ©1995 and Instrument engineers' handbook. Process control. ©1995.
Includes bibliographical references and index.
Contents: v. 2 Process control and optimization.
ISBN 0-8493-1081-4 (v. 2)
1. Process control--Handbooks, manuals, etc. 2. Measuring instruments--Handbooks, manuals, etc. I. Lipták, Béla G. II. Instrument engineers' handbook. Process measurement and analysis.

TS156.8 .I56 2003
629.8--dc21 2003048453

Taylor & Francis Group is the Academic Division of T&F Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Dedicated to my colleagues, the instrument and process control engineers. It is hoped that by applying the knowledge found on these pages, we will make our industries more efficient, safer, and cleaner, and thereby will not only contribute to a happier future for all mankind, but will also advance the recognition and prestige of our profession.

CONTENTS

8 Control and Optimization of Unit Operations 1473

Appendix 2209

CONTRIBUTORS

The names of all the authors of all four editions of this handbook are listed at the beginning of each section of each chapter. Here, their academic degrees, titles, and positions that they held at the time of making their last contributions are also listed. Those authors who have participated in the preparation of this fourth edition are distinguished by an asterisk.

***JÁNOS ABONYI**	PhD, Assistant Professor, University of Veszprém, Hungary
ROBERT H. APPLEBY	BSMath, MSStat, Facilities Manager, Immaculate Conception Church
***JAMES B. ARANT**	BSChE, PE, Principal of J. B. Arant & Assoc. Consulting, U.S.A.
***TIBOR BAAN**	BME, CEO, Aalborg Instruments and Controls Inc.
MIGUEL J. BAGAJEWICZ	PhD, AIChe, Professor, University of Oklahoma, Norman, Oklahoma, U.S.A.
***STEVEN BAIN**	BscEE, PE, I&C Engineer, CH2M Hill
ROBERT J. BAKER	BS, Technical Director, Wallace & Tiernan Div. of Pennwalt Corp.
ROGER M. BAKKE	BS, MS, Senior Engineer, IBM Corp.
***ANDRÁS BÁLINT**	PhD, Assistant Professor, University of Veszprém, Hungary
***RUTH BARS**	PhD, Associate Professor, Budapest University of Technology and Economics, Hungary
CHET S. BARTON	PE, PSEE, Senior Process Automation Engineer, Jacobs Engineering, Baton Rouge, Louisiana, U.S.A.
***HANS D. BAUMANN**	PhDME, PE, President, H.B. Services Partners, LLC
CHESTER S. BEARD	BSEE, Author, Retired from Bechtel Corp.
***KEN BEATTY**	BSME, MEMME, Director of Advanced Product Development, Flowserve
***JONAS BERGE**	Engineer, SMAR, Singapore
JOHN W. BERNARD	BSChE, MSIE, PE, Manager, Systems Technology, The Foxboro Co.

***MICHAEL L. BERNS**	MSEE, PE, Electrical Engineer, Alcan Plastic Packaging
***PETER GRAHAM BERRIE**	PhD, Marketing Communications Manager, Endress + Hauser Process Solutions AG, Switzerland
***VIPUL BHSANAR**	BE, I&C Specialist, Consultant, Canada
PAUL B. BINDER	BSME, Program Manager, Leeds & Northrup
***TERRENCE BLEVINS**	MSEE, Principal Technologist, Emerson Process Management, U.S.A.
RICHARD W. BORUT	Senior Manager, M. W. Kellogg Co.
STUART A. BOYER	PE, BSc, EE, President, Iliad Engineering, Inc., Canada
***WALTER BOYES**	Editor in Chief, Control Magazine
AUGUST BRODGESELL	BSEE, President, CRB Systems Inc.
ROBERT D. BUCHANAN	BSCE, MS, Environmental Consultant
GEORGE C. BUCKBEE	PE, BSChE, MSChE, Control Engineer, Top Control, Clarks Summit, Pennsylvania, U.S.A.
ERIC J. BYRES	PE, Research Facility, British Columbia Institute of Technology, Canada
***FRED M. CAIN**	BSME, PE, Director of Engineering, Flowserve Corp.
ANTHONY M. CALABRESE	BSChE, BSEE, MSChE, PE, Senior Engineering Manager, M. W. Kellogg Co.
***OSCAR CAMACHO**	PhD, Professor, Universidad de Los Andes, Venezuela
BRYAN DELL CAMPBELL	BSEE, MEE, Electrical Engineer, Stanley Tools
DANIEL E. CAPANO	President, Diversified Technical Services, Inc., Stamford, Connecticut, U.S.A.
RICHARD H. CARO	BChE, MS, MBA, CMC Associates, Acton, Massachusetts, U.S.A.
***MICHAEL D. CHAPMAN**	AAS, Startup and Control System Eng. Supervisor, Bechtel Corp.
HARRY L. CHEDDIE	PE, BSc, Principal Engineer, Exida.com, Sarnia, Ontario, Canada.
***GEORGE S. CHENG**	PhD, Chairman, CTO, CyboSoft, General Cybernation Group, Inc.
***XU CHENG**	PhD, Emerson Process Management, Power and Water Solutions
WILLIAM N. CLARE	BSChE, MSChE, Engineering Manager, Automated Dynamics Corp.
SCOTT C. CLARK	BS, ChE, Project Engineer, Merck & Co., Inc., Elkton, Virginia, U.S.A.
JOHN R. COPELAND	BSEE, MSEE, PhDEE, Partner, Jackson Associates
ARMANDO B. CORRIPIO	BSChE, MSChE, PhDChE, PE, Professor of Chemical Engineering, Louisiana State University

NICHOLAS O. CROMWELL	BSChE, Principal Systems Architect, The Foxboro Co.
LOUIS D. DINAPOLI	BSEE, MSEE, Manager, Flowmeter Marketing and Technology, BIF Products of Leeds & Northrup Co.
ROBERT G. DITTMER	BSEE, Corporate Process Control Engineer, PPG Industries
ASEGEIR DRØIVOLDSMO	MS, Research Scientist, OECD Halden Reactor Project, Halden, Norway
SAMUEL G. DUKELOW	BSME, National Sales Manager, Bailey Meter Co.
***WATSON PARNELL DURDEN**	ASInstrTech, Systems Engineer, Westin Engineering
***RANDY A. DWIGGINS**	BCChE, MSSysE, PE, Performance Improvement Consultant, Invensys Systems
LAWRENCE S. DYSART	BSEE, Engineer, Robertshaw Controls Co.
SHIBI EMMANUEL	MTech, BTech, Head of I & C, Dar Al Riyablh Consultants, Al Khobar, Saudi Arabia
***HALIT EREN**	PhD, Curtin University of Technology, Australia
GEORG F. ERK	BSME, MSChE, PE, Consulting Engineer in Instrumentation and Control Systems, Retired from Sun Refining
EDWARD J. FARMER	BSEE, PE, President, Ed Farmer & Associates, Inc.
***BRIDGET ANN FITZPATRICK**	BSChE, MBA, Senior Consultant, Mustang Engineering
PAUL G. FRIEDMANN	BSChE, MSE, Control Consultant, CRB Systems
JEHUDA FRYDMAN	Manager, Electrical Engineering Section, Mobil Chemical Co., Plastics Div.
LUDGER FÜCHTLER	Dipl. Eng., Marketing Manager, Endress + Hauser Process Solutions AG, Reinach, Switzerland
***STEPHAN E. GAERTNER**	BSChE, PE, Sales Engineer, Clipper Controls
***WINSTON GARCIA-GABIN**	PhD, MSc, Professor, Universidad de Los Andes,Venezuela
***DAVID J. GARDELLIN**	BSME, PE, President and CEO of Onyx Valve Co.
CHARLES E. GAYLOR	BSChE, PE, Manager of Engineering Services, Hooker Chemical Corp.
***BRUCE J. GEDDES**	I&C Design Consultant, Calvert Cliffs Nuclear Power Plant
***JOHN PETER GERRY**	MSChE, PE, President, ExperTune Inc.
***ASISH GHOSH**	CE, MIEE, Vice President, ARC Advisory Group
***PEDRO M. B. SILVA GIRÃO**	Professor, Laboratótio de Medidas Eléctricas, Portugal
***ALAN CHARLES GIBSON**	BAS, ADE, Contract Instrument Engineer, Australia
***IAN H. GIBSON**	BSc, DipChE, Process, Controls & Safety Consultant, Australia

*RICHARD A. GILBERT	PhD, Professor of Chemical Engineering, University of South Florida
DAVID M. GRAY	BSChE, Specialist Engineer, Leeds & Northrup Instruments
JOSEPH A. GUMP	BSChE, Process Panel Manager, Control Products Corp.
BHISHAM P. GUPTA	BS, MSME, DSc, PE, Instrument Engineer, Aramco,Saudi Arabia
BERNHARD GUT	Dipl. Eng. Informatik, Head of Department for Communications, Endress + Hauser GmbH + Co., Germany
*GÖRGY GYÖRÖK	PhD, Vice Director, Budapest Polytechnic, Hungary
DIANE M. HANLON	BSEE, BSME, Engineer, E.I. DuPont de Nemours & Co.
HASHEM MEHRDAD HASHEMIAN	MSNE, ISA Fellow, President, Analysis and Measurement Services Corp., Knoxville, Tennessee, U.S.A.
HAROLD I. HERTANU	MSEE, PE, Senior Project Engineer, Malcolm Pirnie Inc.
*JENÖ HETHÉSSY	PhD, Associate Professor, Budapest University of Technology and Economics, Hungary
CONRAD H. HOEPPNER	BSEE, MSEE, PhD, PE, President, Industrial Electronics Corp.
HAROLD L. HOFFMAN	BSChE, PE, Refining Director, Hydrocarbon Processing
PER A. HOLST	MSEE, Director of Computing Technology, The Foxboro Co.
*KARLENE A. HOO	PhD, Associate Professor of Chemical Engineering,Texas Tech University
MICHAEL F. HORDESKI	BSEE, MSEE, PE, Control System Consultant, Siltran Digital
FRANKLIN B. HOROWITZ	BSChE, MSChE, PE, Process Engineer, Crawford & Russell Inc.
DAVID L. HOYLE	BSChE, Principal Field Application Engineer, The Foxboro Co.
DAVID W. HOYTE	MACantab, Consultant, IBM Corp.
JOHN C. HUBER	PhDEE, Laboratory Manager, 3M Co.
STUART P. JACKSON	BSEE, MSEE, PhD, PE, President, Engineering and Marketing Corp. of Virginia
RAJSHREE RANKA JAIN	BSChE, Applications Engineer, Capital Controls Co.
*JAMES EDWARD JAMISON	BScChE, PE, Lead Engineer, Instrumentation, Electrical & Control Systems, OPTI Canada Inc.
*BRUCE ALAN JENSEN	BS, MSChE, Technical Solutions Manager, Yokagawa Corp. of America
KENNETH J. JENTZEN	BSEE, MSEE, Manager, Project Engineering and Maintenance, Mobil Chemical Co.
DONALD R. JONES	BSME, PE, Consultant, Powell-Process Systems, Inc.

BABU JOSEPH	PhDChE, Professor of Chemical Engineering, Washington University
VAUGHN A. KAISER	BSME, MSE, PE, Member of Technical Staff, Profimatics Inc.
***BABAK KAMALI**	BSEE, Control and Automation Designer, PSJ Co., Iran
LES A. KANE	BSEE, Mechanical/Electrical Editor, Hydrocarbon Processing
GENE T. KAPLAN	BSME, Project Design Engineer, Automated Dynamics Corp.
DONALD C. KENDALL	BSChE, MSChE, Senior Systems Analyst, The Foxboro Co.
***PRABHAKAR Y. KESKAR**	PhDEE, PE, Principal Technologist, CH2M Hill
CHANG H. KIM	BSChE, Senior Group Leader, UniRoyal Inc.
PAUL M. KINTNER	BSEE, MSEE, PhDEE, Manager, Digital Products,Cutler-Hammer Inc.
KLAUS H. KORSTEN	Dipl. Eng., Marketing Manager, Endress + Hauser Process Solutions AG, Reinach, Switzerland
***BÉLA LAKATOS**	MSc, PhD, Professor, University of Veszprém, Hungary
JAMES W. LANE	BSChE, MSChE, PE, Executive Director, Tenneco Inc.
***CULLEN G. LANGFORD**	BSME, PE. President, Cullen G. Langford Inc.
AVERLES G. LASPE	BS, PE, Senior Control Application Consultant, Honeywell Inc.
DONALD W. LEPORE	BSME, Design Engineer, The Foxboro Co.
CHENG S. LIN	DrEng, PE, Consultant, Tenneco Inc.
***KLAUS-PETER LINDNER**	Dipl. Inform., Specialist for New Technologies, Endress + Hauser GmBH + Co., Switzerland
***BÉLA LIPTÁK**	ME, MME, PE, Process Control Consultant
DAVID H. F. LIU	PhDChE, Principal Scientist, J. T. Baker Inc.
MICHAEL H. LOUKA	BS, Section Head, IFE Halden Virtual Reality Centre, Halden, Norway
ORVAL P. LOVETT JR.	BSChE, Consulting Engineer, E.I. DuPont de Nemours & Co.
DALE E. LUPFER	BSME, PE, Process Automation Consultant, Engel Engineering Co.
***MATTHEW H. MACCONNELL**	BSChE, MBA, PE, Technology Manager, Air Products and Chemicals, Inc.
***JÁNOS MADÁR**	MSCSChE, Research Assistant, University of Veszprém, Hungary
VICTOR J. MAGGIOLI	BSEE, Principal Consultant, E.I. DuPont de Nemours & Co.
CHARLES L. MAMZIC	BSME, CA. Director, Marketing and Application Engineering, Moore Products Inc.

***M. SAM MANNAN**	PhD, PE, Associate Professor of Chemical Engineering, Texas A&M University
ALAN F. MARKS	BSChE, PE, Engineering Specialist, Control Systems, Bechtel Petroleum Inc.
***EDWARD M. MARSZAL**	BSChE, CFSE, PE, Principal Engineer, Exida
FRED D. MARTON	Dipl. Eng., Managing Editor, Instruments and Control Systems
***GREGORY K. MCMILLAN**	MSEE, Professor, Washington University of Saint Louis
***RICK MEEKER**	Process Control Solutions, Inc.
DANIEL T. MIKLOVIC	ASNT, BSEE, MSSM, CMfgE, MIMC, Senior Consultant, H.A. Simons Ltd.
***HERBERT L. MILLER**	MSME, PE, Vice President, Control Components
FERENC MOLNÁR	MSEE, Department Manager, Chinoin Pharmaceutical and Chemical Works Ltd., Hungary
CHARLES F. MOORE	BSChE, MSChE, PhDChE, Professor of Chemistry, University of Tennessee
JOHN A. MOORE	BSChE, PE, Senior Application Specialist, Leeds & Northrup Instruments
RALPH L. MOORE	BSME, MSME, MSInstE, PE, Instrument Engineer, E.I. DuPont de Nemours & Co.
***DOUG MORGAN**	BSChE, Project Engineer, Control Systems International, U.S.A.
PAUL W. MURRILL	BSChE, MSChE, PhDChE, PE, Vice-Chancellor, Louisiana State University
THOMAS J. MYRON, JR.	BSChE, Senior Systems Design Engineer, The Foxboro Co.
MIKLÓS (NICHOLAS) NAGY	ME, Chief Engineer, Flygt Submersible Pump Co.
***MARK NIXON**	BSEE, Chief Architect, Delta V Control System, Emerson Process Management
***DAVID S. NYCE**	BSEE, MBA, Owner, Revolution Sensor Company
***SHINYA OCHIAI**	BSME, MSME, PhD, PE, Adjunct Professor, Texas A&M University
RICHARD H. OSMAN	PE, Engineering Manager, Robicon Corp.
GLENN A. PETTIT	Manager, Plastic Instrumentation, Rosemount Engineering Co.
***MICHAEL J. PIOVOSO**	PhD, PE, Associate Professor, Pennsylvania State University
WALLACE A PRATT, JR.	BSEE, Chief Engineer, HART Communication Foundation, Austin, Texas, U.S.A.
***SUBRAMANIUM RENGANATHAN**	PhD, Vice Chancellor, Bharath Institute of Higher Education and Research, Chennai, India
***R. RUSSELL RHINEHART**	PhDChE, Professor and Head, School of Chemical Engineering, Oklahoma State University
JAMES B. RISHEL	BSME, President, Corporate Equipment

HOWARD C. ROBERTS	BAEE, PE, Consultant, Lecturer at University of Colorado, Denver
***ALBERTO ROHR**	EE, Dr.Eng., Consultant, Italy
***RUBÉN DARIO ROJAS**	PhD, Universidad de Los Andes, Venezuela
DERRICK KEITH ROLLINS, SR.	BS, MS, PhD, AIChE, Associate Professor, Iowa State University, Ames, Iowa, U.S.A.
***MICHEL RUEL**	PE, President, Top Control
DOUGLAS D. RUSSELL	BSEE, MSEE, Group Leader, The Foxboro Co.
DONALD R. SADLON	BSEE, General Manager, Automated Dynamics Corp.
CHAKRA J. SANTHANAM	BSChE, MSChE, PE, Management Staff, Arthur D. Little Inc.
PÉTER SZILÁRD SCHERMANN	Senior Consultant at Comprimo BV and Professor of Process Dynamics, University of Groningen, the Netherlands
WALTER F. SCHLEGEL	BE, Assistant Chief Process Engineer, Crawford & Russell Inc.
PHILLIP D. SCHNELLE, JR.	BSEE, MSEE, PE, Technical Service Engineer, E.I. DuPont de Nemours & Co.
MICHAEL R. SHAFFER	PhD, PE, Adjunct Professor, Florida Institute of Technology
***FRANCIS GREGWAY SHINSKEY**	BSChE, Process Control Consultant
JOSEPH P. SHUNTA	PhDChE, PE, Principal Consultant, E.I. DuPont de Nemours & Co.
BURBINDER B. SINGH	BE, MBA, MCSE, Consultant Control Systems Engineer, Chicago, Illinois, U.S.A.
***ROBERT J. SMITH II**	EET, Plant Engineer, Rock-Tenn Co.
DAVID A STROBHAR	PE, SBHFE, President, Beville Engineering, Inc., Dayton, Ohio, U.S.A.
***ANGELA ELAINE SUMMERS**	PhD, PE, President, Sis-Tech Solutions, LLC
***GUSZTÁV SZECSÖ**	BSEE, MSEE, PhD. Professor of Control Engineering University of Miskolc, Hungary
JAMES E. TALBOT	BSEP, Senior Account Executive, Lewis, Gilman & Kynett
***JAMES F. TATERA**	BS, MBA, Senior Process Analysis Consultant, Tatera and Associates Inc., U.S.A.
MAURO C. TOGNERI	BSEE, President, Powell-Process Systems
***G. KEVIN TOTHEROW**	BSEE, President, Sylution Consulting
THOMAS H. TSAI	BSE, PE, Director, Tenneco Inc.
***ANN TUCK**	BSME, Senior Control Systems Engineer, Bechtel Corp.
***ISTVÁN VAJK**	PhD, Professor, Budapest University of Technology and Economics, Hungary
JÁNOS (JOHN) VENCZEL	BSEE, MSEE, Senior RF Engineer, Automation Industries

IAN VERHAPPEN	PE, BSCE, Engineering Associate, Syncrude Canada Lts., Fort McMurray, Alberta, Canada
***STEFANO VITTURI**	Dr. Eng., Italian National Council of Research, Italy
***HAROLD L. WADE**	PhD, President, Wade Associates, Inc.
MICHAEL H. WALLER	BME, SM, ME, PE, Associate Professor, Miami University
CHARLES G. WALTERS, JR.	BSEE, Electrical/Instrument Engineer, Geraghty and Miller Inc.
MARVIN D. WEISS	BSE, MSChE, Assistant Professor of Mechanical Engineering, Valpariso University
PETER ERIC WELLSTEAD	BSc, MSc, DSc, FIEE, CE, Reader in Control Engineering, University of Manchester, England
***CURT W. WENDT**	BSET, PE, I&C Group Leader, CDM
***HARRY H. WEST**	PhD, Research Associate, Texas A&M University
ANDREW C. WIKTOROWICZ	BS, PE, President, Automated Dynamics Corp.
THEODORE J. WILLIAMS	BSChE, MSChE, MSEE, PhDChE, PE, Professor of Engineering, Director of Purdue Laboratory for Applied Industrial Control, Purdue University
ROBERT A. WILLIAMSON	BSME, BA, PE, Supervisor, The Foxboro Co.
***JOHN WILSON**	BSChE, Business Manager, Fisher Controls International, LLC
***WILLY WOJSZNIS**	PhD, Technologist, DeltaV Advanced Control, Emerson Process Management, U.S.A.

INTRODUCTION

This is the fourth edition of the *Instrument Engineers' Handbook* (*IEH*). This handbook serves the automation and control engineering (ACE) profession and has been published once every decade since 1969. The subject of the first volume is measurement; the second volume deals with control; and the third volume is devoted to the topics of software and digital networks.

In the introduction to each new edition, I give a summary of the key developments of the previous decade and point to the challenges we face in the coming decade. Before discussing the previous and the next decades, however, I will say a few words about the growing pains of the ACE profession. I will conclude this introduction with a brief summary of the history of this handbook.

AUTOMATION AND CONTROL ENGINEERING (ACE)

Ours is a very young profession. When the first edition of the *Instrument Engineers' Handbook* (*IEH*) was published, Marks' *Mechanical Engineers' Handbook* was in its fifth and Perry's *Chemical Engineers' Handbook* was in its sixth edition. It is partly for this reason that while people know what kind of engineer an ME or a ChE is, they have no idea what I do when I say that my field is process control or instrumentation. I just get a blank stare.

It is time for us to change that. The first step should be to use a name for our profession that people understand. It is time for our profession to develop a distinct identity.

When I was teaching at Yale, my course was offered under the Chemical Engineering Department. This was not because Yale had anything against our profession; it was simply because they did not know where to put my course. Even this handbook of mine proves the confusion about our identity, because Taylor & Francis publishes this handbook among its Electrical Engineering books. Once again, the reason is not that Taylor & Francis has something against our profession. No, the reason is that we have not yet developed our distinct identity.

"Automation" is a term that the wider public understands. Therefore, I would suggest that we consistently refer to our profession as Automation and Control Engineering (ACE). Together with that, the name of our professional society should also be changed to International Society of Automation (ISA). Another clarifying step could be to change the title of our society magazine from *InTech* to *AutomationTech* because "In" does not say much.

The potentials of the ACE profession are great. While as a profession we have been anonymous, we have already designed fully automated Mars explorers and fully optimized industrial plants. Now it is time to let the world know that we exist. It is time to add to the list of ME, EE, or ChE professional engineering licenses one for ACE engineers; it is time for universities to offer degrees in ACE engineering and for publishers to set up ACE departments.

We should not be shy about this. After all, no engineering profession can claim what we can. No engineering profession can offer to increase the global gross domestic product by trillions of dollars simply through optimization, without building a single new plant, while also *increasing* safety and *reducing* pollution. We can do that. We can increase productivity without using a single pound of additional raw material and without requiring a single additional BTU of energy. Yes, our profession does deserve a distinct identity.

DEVELOPMENTS OF THE LAST DECADE

These days, the computer is our main tool of control. The chapters of this volume describe how the computer is being used in optimizing our processes, providing self-diagnostics, and displaying status information in operator-friendly formats. Today, the World Wide Web provides access to great quantities of data; in the future it will also provide problem-solving capability, so that through the grid, every ACE engineer will have a supercomputer at his or her disposal.

During the last decade, the artificial separation between the roles of DCS, PLC, and PC packages has started to disappear because their separate roles of control (DCS), logic sequencing (PLC), or simulation, and business-related tasks (PC) are beginning to be integrated. I believe that in the near future DCS will simply mean digital control system. Once the digital bus protocols are integrated into a single global standard, the presently required interfacing cards (and the associated risk of mixup) will disappear, and therefore our control systems will become safer, simpler, and more effective.

In the paragraphs below I review some of the developments of the last decade.

Is the Age of the PID Over?

Designating a valve on a flow sheet as a temperature control valve (TCV) will not suspend the laws of nature, and this arbitrary designation will not, for example, prevent the valve from affecting the process pressure. Similarly, the number of available control valves in a process will not necessarily coincide with the number of process properties that need to be controlled. Multivariable herding or envelope control overcomes this limitation of uncoordinated single loop controllers and lets us control all variables that need to be controlled, while minimizing interactions.

The majority of our control loops are still keeping single process variables on set point, but the age of multivariable control has already arrived. In the majority of cases, we still tend to control levels, flows, and temperatures as if these single loops operated in a vacuum, but others are already recognizing that loops do interact and that the opening or closing of a control valve affects not only the one variable we are controlling. For these reasons, the decoupling of interactions based on relative gain calculations have become important tools in the tool boxes of ACE engineers.

Many of us have concluded that the single-loop mentality is wrong because our plants do not produce flows, levels, and temperatures; hence, the control of these variables should not be our ultimate goal. Our ultimate goal should be to optimize the productivity and safety of the whole plant. As a consequence, we are now thinking in terms of unit operation controllers. In these multivariable control systems, the total unit operation (be it a boiler, a distillation column, or a compressor) is being controlled.

The Set Point

Advances have also been made in rethinking the role of the set point. In one aspect, the single set point is often replaced by a set point gap, so that as long as the controlled variable is within that gap, the output is unaltered. This tends to stabilize sensitive loops, such as flow.

Another aspect in which the set point is treated differently today is its effect on the controller output. In many algorithms a change in set point does not change the proportional or derivative contributions to the output because the P and D modes act only on the measurement.

In other algorithms, while the set point change does affect the integral contribution to the output, the set point change is “feedforwarded” directly to the output to minimize reset windup. Reset windup is also minimized by external feedback, taken from the slave measurement in case of cascade loops, from the valve signal in case of selective loops, and from the inverse model in feedforward loops.

Dynamics and Dead Time

The dynamics of control are also better understood today. It is clear that for quarter amplitude damping the gain product of the loop should be about 0.5. This means that in order to keep this product constant, if the process gain doubles, the controller gain must be cut in half. This understanding is critical to the control of all nonlinear processes (heat transfer, chemical reaction, pH, etc.). Clearly understanding this goal also allows for gain adaptation based on either measurements or modeling.

Similarly, the role of dead time is also better understood today. Most ACE engineers know that the integral and derivative control modes must be tuned to match the dead time. Therefore, the control goal is to reduce the loop dead time to a minimum and keep it constant. If that is not possible because the process dead time must vary (transportation lag caused by displacing a fixed volume), it is necessary to match that variable dead time with adapted I and D settings. When the dead time is large, the regular PID algorithm is replaced with sample-and-hold or predictor algorithms.

Unit Operations Controllers

An already existing multipurpose reactor package (described in this volume) can be reconfigured through software modifications to become a stripper, distillation, or crystallizer unit controller. Other multivariable, envelope, and matrix control systems described in this volume are aimed at increasing the efficiency or the productivity of the process, while treating the individual variables — the temperatures, pressures, and levels — only as constraints.

There are hundreds of expert systems, all serving some form of optimization. From the perspective of their methods of operation, one can group them into model-based and model-free methods. They both control multivariable unit operations and because they both evaluate the total process, they also eliminate the interactions between the various controlled and manipulated variables.

Expert systems, which are used in unit operations controllers, are also useful in decoupling the interactions through relative gain and other techniques. Probably the greatest progress has occurred in the area of model-based control, which utilizes both steady-state and dynamic models and allows both for the prediction of process responses before they occur and for continual refinement of the model by empirical updating. In this regard neural networks, artificial intelligence, statistical process control, fuzzy logic, and empirical optimization strategies have all made some contribution.

Model-Based and Model-Free Control

Model-Based Control (MBC), Model Predictive Control (MPC), and Internal Model Control (IMC) are suited for the optimization of well-understood unit processes, such as heat transfer or distillation. Their performance is superior to that of model free systems because they are capable of anticipation and thereby can predict the process response to new situations. In this sense their performance is similar to that of feedforward control systems, while model-free systems behave in a feedback manner only.

The performance of a model-free expert system can be compared to the behavior of a tennis player. The tennis player does not necessarily understand Newton's laws of motion or the aerodynamic principles that determine the behavior of a tennis ball. The tennis player has simply memorized the results of a large number of past "process" responses. This is also the basis of most human learning. All the neural network–based software packages mimic this method of learning.

Neural networks, fuzzy logic, and statistical process control are all such model-free methods, which can be used without the need for knowing the mathematical model of the process. The major difference between fuzzy logic and neural networks is that the latter can only be trained by data, but not with reasoning. Fuzzy logic is superior from this perspective because it can be modified both in terms of the gain (importance) of its inputs and in terms of the functions of its inputs.

The main limitations of all model-free expert systems are their long learning period (which can be compared to the maturing of a child) and the fact that their knowledge is based solely on past events. Consequently, they are not prepared to handle new situations. Therefore, if the process changes, they require retraining because they can only anticipate repetitive events.

Artificial Neural Networks (ANN)

One of the tools used in building models is the Artificial Neural Network (ANN), which can usually be applied under human supervision or can be integrated with expert and/or fuzzy logic systems. Figure 1 shows a three-layer ANN network, which serves to predict the boiling point of a distillate and the Reid vapor pressure of the bottoms product of a

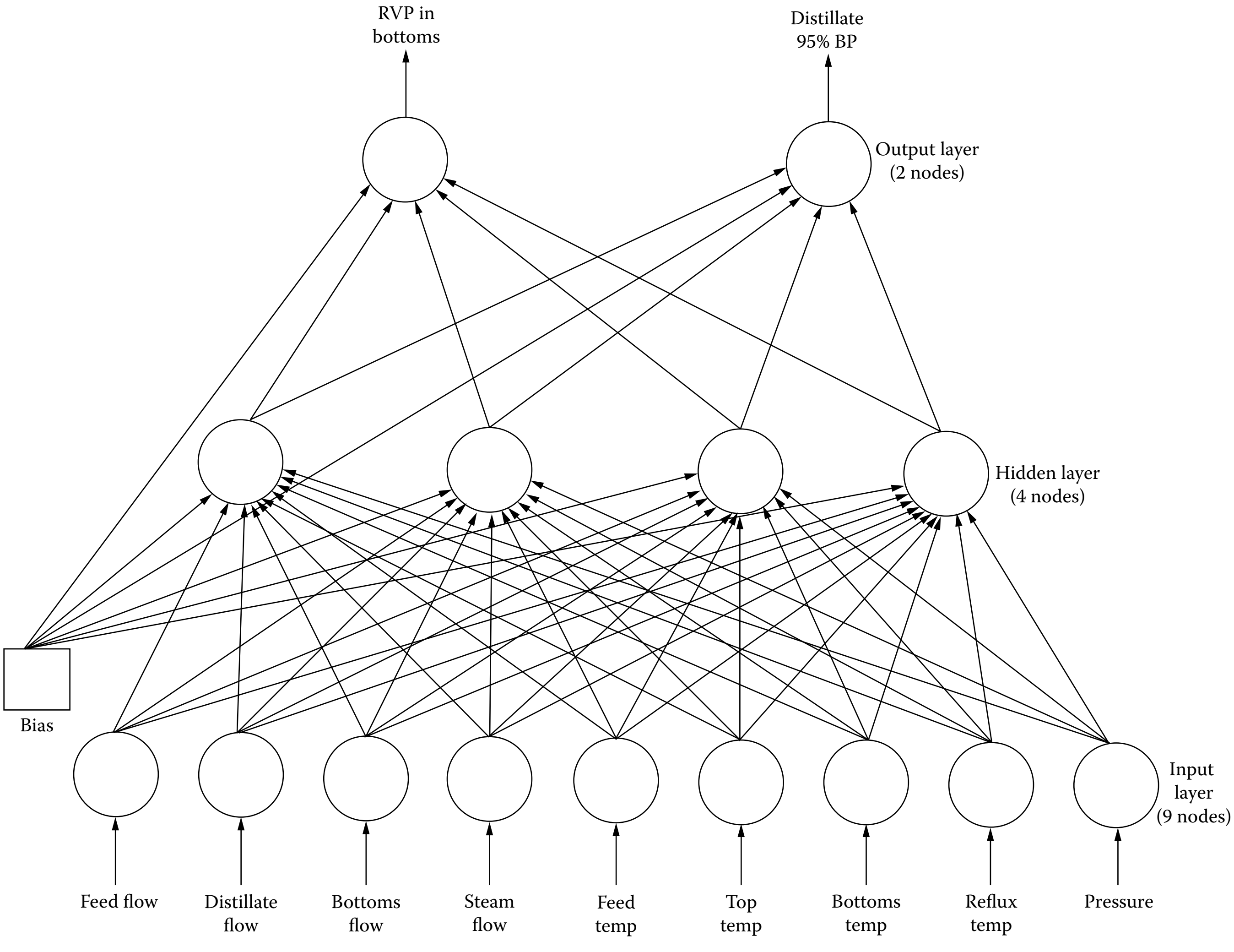


FIG. 1
A three-layer artificial neural network (ANN) can be used to predict the quality of overhead and bottoms products in a distillation column.

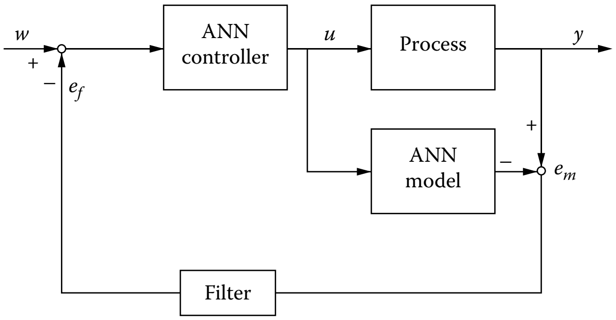


FIG. 2
The use of an artificial neural network in an IMC (Internal Model Control) application.

column. Such predictive ANN models can be valuable because they are not limited by either the unreliability or the dead time of analyzers.

The "personality" of the process is stored in the ANN network by the way the processing elements (nodes) are connected and by the importance assigned to each node (weight). The ANN is "trained" by example and therefore it contains the adaptive mechanism for learning from examples and to adjust its parameters based on the knowledge that is gained through this process of adaptation. During the "training" of these networks, the weights are adjusted until the output of the ANN matches that of the real process.

Naturally, these networks do need "maintenance" because process conditions change; when that occurs, the network requires retraining. The hidden layers help the network to generalize and even to memorize.

The ANN network is also capable of learning input/output and inverse relationships. Hence it is useful in building Internal Model Control (IMC) systems based on ANN-constructed plant models and their inverses. In a neural controller (Figure 2), the ANN can be used in calculating the control signal.

Herding Control

When a large number of variables is involved, model free herding control can be considered. This approach to control can be compared to the herding of sheep, where the shepherd's dog goes after one animal at a time and changes the direction or speed of the whole herd by always influencing the behavior of the least cooperative sheep.

I have successfully applied such herding algorithms in designing the controls of the IBM headquarters building in New York City. By herding the warm air to the perimeter (offices with windows) from the offices that are heat generators even in the winter (interior offices), the building became self-heating. This was done by changing the destination of the return air from one "hot" office at a time (the one that was the warmest in the building) and simultaneously opening the supply damper of the "coldest office" to that same header.

I have also applied the herding algorithm to optimize the process of computer chip manufacturing by eliminating the minute pressure differences that can cause dust-transporting drafts, which in turn can ruin the chips.

In general, herding control is effective if thousands of manipulated variables exist and they all serve some common goal.

Common-Sense Recommendations

While evaluating and executing such sophisticated concepts as optimized multivariable control, the ACE engineer's best friend is common sense and our most trusted teacher is still Murphy, who says that anything that can go wrong, will. In order to emphasize the importance of common sense, I will list here some practical recommendations:

- Before one can control a process, one must fully understand it.
- Being progressive is good, but being a guinea pig is not. Therefore, if the wrong control strategy is implemented, the performance of even the latest digital hardware will be unacceptable.
- An ACE engineer is doing a good job by telling plant management what they need to know and not what they like to hear.
- Increased safety is gained through backup. In case of measurement, reliability is increased by the use of multiple sensors, which are configured through median selectors or voting systems.
- If an instrument is worth installing, it should also be worth calibrating and maintaining. No device can outperform the reference against which it was calibrated.
- All man-made sensors detect relative values, and therefore the error contribution of references and compensators must always be included in the total error.
- Sensors with errors expressed as percent of the actual reading are preferred over those with percent of full-scale errors. If the actual percent of reading error increases as the reading moves down-scale, the loop performance will also deteriorate.
- It is easier to drive within the limits of a lane than to follow a single line. Similarly, control is easier and more stable if the single set point is replaced by a control gap.
- Process variables should be allowed to float within their safe limits as they follow the load. Constancy is the enemy of efficiency. Optimization requires efficient adaptation to changing conditions.
- Trust your common sense, not the sales literature. Independent performance evaluation based on the recommendations of international and national users' associations (SIREP-WIB) should be done *before* installation, not after it. The right time for "business lunches" is *after* start-up, not before the issue of the purchase order.
- Annunciators do not correct emergencies; they just throw those problems that the designers did not know how to handle into the laps of the operators. The smaller the annunciator, the better the design.

FUTURE NEEDS

I have already mentioned such needs as the establishment of our professional identity as automation and control engineers (ACE) or the need to integrate DCS, PLC, and PC hardware into a new digital control system (DCS+) that incorporates the features of all three. I have also briefly mentioned the need for bringing order into our "digital Babel" and to stop the trend toward software outsourcing (selling DCS systems without some software packages), which is like selling violins without the strings.

I would not be surprised if, by the end of the 21st century, we would be using self-teaching computers. These devices would mimic the processes taking place in our children's brains, the processes that allow babies to grow into Einsteins by learning about their environment. These devices would be watching the operation of a refinery or the landing and takeoff of airplanes and eventually would obtain as much knowledge as an experienced operator or pilot would have.

If this comes to pass, some might argue that this would be a step forward because machines do not forget; do not get tired, angry, or sleepy; and do not neglect their job to watch a baseball game or to argue with their spouse on the phone. This might be so, yet I would still prefer to land in a human-piloted airplane. I feel that way because I respect Murphy more than most scientists and I know that he is right in stating that "anything that can happen, will." For this reason, the knowledge of the past, which is the knowledge of the computer, might still not be enough.

In addition to new control tools, we will also have new processes to control. Probably the most important among these will be the fuel cell. The fuel cell is like a battery, except that it does not need recharging because its energy is the chemical energy of hydrogen, and hydrogen can come from an inexhaustible source, namely water. In order to gain the benefits of tripling the tank-to-wheel efficiency of our transportation systems while stopping global warming, we will have to learn to control this new process. The challenge involves not only the control of the electrolytic process that splits the water by the use of solar energy, but also the generation, storage, and transportation of liquid or slurry hydrogen.

As I was editing this reference set for the fourth time, I could not help but note the new needs of the process control industry, which are the consequences of the evolution of new hardware, new software, and new process technologies. Here, in the paragraphs that follow, I will describe in more detail what I hope our profession will accomplish in the next decade.

Bringing Order to the "Digital Babel"

In earlier decades, it took time and effort to reach agreement on the 3- to 15-PSIG (0.2- to 1.0-bar) pneumatic and later on the 4- to 20-mA DC standard current signal range. Yet when these signal ranges were finally agreed upon, the benefit was universal because the same percentage of full scale measurement was represented by the same reading on every device in the world.

Similarly, the time is ripe for a single worldwide standard for all digital I/O ranges and *digital communication protocols.* The time is ripe for such an internationally accepted digital protocol to link all the digital "black boxes" and to eliminate the need for all interfacing.

By so doing, we could save the time and effort that are being spent on figuring out ways to interconnect black boxes and could invest them in more valuable tasks, such as enhancing the productivity and safety of our processing industries.

Networks and Buses

Protocol is the language spoken by our digital systems. Unfortunately, there is no standard that allows all field devices to communicate in a common language. Therefore the creation and universal acceptance of a standard field bus is long overdue.

There is nothing wrong with, say, the Canadians having two official languages, but there is something wrong if a pilot does not speak the language of the control tower or if two black boxes in a refinery do not speak the same language. Yet the commercial goal of manufacturers to create captive markets resulted in some eight protocols. These control-oriented communication systems are supported by user groups, which maintain Internet sites as listed below:

AS-Interface	www.as-interface.com
DeviceNet	www.odva.org
HART	www.hartcomm.org
PROFIBUS	www.profibus.org
Found. Fieldbus	www.fieldbus.org
OPC Foundation	www.opcfoundation.com
WorldFIP	www.worldfip.org
ControlNet	www.controlnet.com
MODBUS	www.modbus.org
Ethernet TCP/IP*	www.industrialethernet.com

*TCP - Transmission control protocol
IP - Internet protocol

During the last decade, HART has become the standard for interfacing with analog systems, while Ethernet was handling most office solutions. SCADA served to combine field and control data to provide the operator with an overall view of the plant. While there was no common DCS fieldbus protocol, all protocols used Ethernet at the physical and TCP/IP at the Internet layer. MODBUS TCP was used to interface the different DCS protocols.

The layers of the communication pyramid were defined several ways. OSI defined it in seven (7) layers, #1 being the physical and #7 the application layer (with #8 being used for the "unintegrated," external components). The IEC-61512 standard also lists seven levels, but it bases its levels on physical size: (1) control module, (2) equipment, (3) unit, (4) process cell, (5) area, (6) site, and (7) enterprise.

As I noted earlier, in the everyday language of process control, the automation pyramid consists of four layers: #1 is the level of the field devices, the sensors and actuators; #2 is control; #3 is plant operations; and #4 is the level of business administration.

Naturally, it is hoped that in the next decade, uniform international standards will replace our digital Babel, so that we once again can concentrate on controlling our processes instead of protecting our plants from blowing up because somebody used the wrong interface card.

Software Outsourcing

Another problem in the last decade was the practice of some DCS vendors to sell their "violins without strings," to bid their packages without including all the software that is needed to operate them.

To treat software as an extra and not to include the preparation of the unique control algorithms, faceplates, and graphic displays in the basic bid can lead to serious problems. If the plant does not hire an engineering firm or system integrator to put these strings onto the DCS violin, the plant personnel usually cannot properly do it and the "control music" produced will reflect that. In such cases the cost of plugging the software holes can exceed the total hardware cost of the system.

The cause of another recurring problem is that the instructions are often written in "computerese."

In some bids, one might also read that the stated cost is for "hardware with software license." This to some will suggest that the operating software for the DCS package is included. Well, in many cases it is not; only its license is.

Similarly, when one reads that an analyzer or an optimization or simulation package needs "layering" or is in the "8th layer," one might think that the bid contains eight layers of fully integrated devices. Well, what this language often means is that the cost of integrating these packages into the overall control system is an extra.

So, on the one hand, this age of plantwide digital networks and their associated advanced controls has opened the door for the great opportunities provided by optimization. On the other hand much more is needed before the pieces of the digital puzzle will conveniently fit together, before these "stringless violinists" can be integrated into a good orchestra of automation.

Connectivity and Integration

Utilizing our digital buses, one can plug in a PC laptop or use a wireless hand tool and instantly access all the data, displays, and intelligence that reside anywhere in a DCS network. This capability, in combination with the ability for self-tuning, self-diagnosing, and optimizing, makes the startup, operation, and maintenance of our plants much more efficient.

The modern control systems of most newly built plants consist of four levels of automation. In the field are the intelligent and self-diagnosing local instruments (sensors, valves, motors, safety devices). This first level is connected by a number of data highways or network buses to the next level in this automation pyramid, the level of control. The third level is plant operations, and the fourth is the enterprise-wide business level.

The functions of the DCS workstations include control/operation, engineering/historian, and maintenance functions, while the enterprise-wide network serves the business functions. In addition, wireless hand tools are used by the roving operators, and external PCs are available to engineers for their process modeling and simulation purposes.

HARDWARE-RELATED IMPROVEMENTS NEEDED

I will discuss below some of the areas in which the next decade should bring improvements in the quality and intelligence of the components that make up our control systems. I will discuss the need for better testing and performance standards and the improvements needed in the sensors, analyzers, transmitters, and control valves. I will place particular emphasis on the potentials of "smart" devices.

Meaningful Performance Standards

The professional organizations of automation and control engineers (ACE) should do more to rein in the commercial interests of manufacturers and to impose uniform performance testing criteria throughout the industry. In the sales literature today, the meanings of performance-related terms such as inaccuracy, repeatability, or rangeability are rarely based on testing, and test references are rarely defined. Even such terms as "inaccuracy" are frequently misstated as "accuracy," or in other cases the error percentages are given without stating whether they are based on full scale or on actual readings. It is also time for professional societies and testing laboratories to widely distribute their findings so that these reliable third-party test results can be used by our profession to compare the performance of the various manufacturers' products.

We the users should also require that the manufacturers always state not only the inaccuracy of their products but also the rangeability over which that inaccuracy statement is valid. In other words, the rangeability of all sensors should be defined as the ratio between those maximum and minimum readings for which the inaccuracy statement is still valid.

It would also be desirable to base the inaccuracy statements on the performance of at least 95% of the sensors tested and to include in the inaccuracy statement not only linearity, hysteresis, and repeatability but also the effects of drift, ambient temperature, over-range, supply voltage variation, humidity, radio frequency interface (RFI), and vibration.

Better Valves

In the next decade, much improvement is expected in the area of final control elements, including smart control valves.

This is because the performance of the control loop is much affected not only by trim wear in control valves but also by stem sticking caused by packing friction, valve hysteresis, and air leakage in the actuator. The stability of the control loop also depends on the gain of the valve during the tuning of the loop.

In order for a control loop to be stable, the loop is tuned (the gain of the controller is adjusted) to make the gain product of the loop components to equal about 0.5. If the control valve is nonlinear (its gain varies with the valve's opening), the loop will become unstable when the valve moves away from the opening where it was when the loop was tuned. For this reason, the loop must be compensated for the gain characteristics of the valve; such compensation is possible only if the valve characteristics are accurately known.

For the above reasons, it is desirable that the users and the professional societies of ACE engineers put pressure on the manufacturers to accurately determine the characteristics of their valves. The other performance capabilities of the final control elements also need to be more uniformly defined. This is particularly true for the rangeability of control valves. For example, a valve should be called linear only if its gain (Gv) equals the maximum flow through the valve (Fmax) divided by the valve stroke in percentage (100%) throughout its stroke.

The valve manufacturers should also be required to publish the stroking range (minimum and maximum percentages of valve openings) within which the valve gain is what it is claimed to be (for a linear valve it is Fmax/100%). Similarly, valve rangeability should be defined as the ratio of the minimum and maximum valve Cvs, at which the valve characteristic is still what it is specified to be.

Smarter Valves

A traditional valve positioner serves only the purpose of maintaining a valve at its intended opening. Digital valve controllers, on the other hand, provide the ability to collect and analyze data about valve position, valve operating characteristics, and valve performance trending. They also provide two-way digital communication to enable diagnostics of the entire valve assembly and instrument. Section 6.12 in this handbook and the following Web pages provide more information: Metso Automation (http://www.metsoautomation.com/), (http://www.emersonprocess.com/fisher/products/fieldvue/dvc/index.html)

The potentials of smart valves are likely to be further increased and better appreciated in the next decade. The main features of a smart valve include its ability to measure its own:

- Upstream pressure
- Downstream pressure
- Temperature
- Valve opening position
- Actuator air pressure

Smart valves will also eliminate the errors introduced by digital-to-analog and analog-to-digital conversions and will guarantee the updating of their inputs about 10 times per second. In addition, they will be provided with the filters required to remove the errors caused by turbulence in these readings. As a consequence, smart valves will also be able to measure the flow, by solving equations, such as the one below for liquid flow:

$$Q = \frac{C_v}{\sqrt{\frac{SPGRAV}{\Delta P_A}}}$$

where

Q = Flow rate (GPM)
F_L = Recovery coefficient
C_v = Flow capacity factor
P_1 = Upstream pressure (PSIA)
P_v = Liquid vapor press (PSIA)
P_c = Critical pressure (PSIA)
ΔP_A = Valve pressure drop (PSI) or

If Choked:

$$\Delta P_A = F_L^2 \left[P_1 - \left(0.96 - 0.28 \sqrt{\frac{P_V}{P_C}} \right) \right] P_V$$

The smart valves of the coming decade will hopefully not only measure their own flows but will also measure them over a rangeability that exceeds most flowmeters (from 25:1 to over 100:1) because they in effect are variable-opening orifices.

If the sensors of the intelligent control valve are connected to a PID chip mounted on the valve, the smart valve becomes a local control loop. In that case, only the operator's displays need to be provided remotely, in a location that is convenient for the operator. In such a configuration, it will be possible to remotely reconfigure/recalibrate the valve as well as to provide it with any limits on its travel or to diagnose stem sticking, trim wear, or any other changes that might warrant maintenance.

"Smarter" Transmitters, Sensors, and Analyzers

In the case of transmitters, the overall performance is largely defined by the internal reference used in the sensor. In many cases there is a need for multiple-range and multiple-reference units. For example, pressure transmitters should have both atmospheric and vacuum references and should have sufficient intelligence to automatically switch from one to the other on the basis of the pressure being measured.

Similarly, d/p flow transmitters should have multiple spans and should be provided with the intelligence to automatically switch their spans to match the actual flow as their measurement changes.

The addition of "intelligence" could also increase the amount of information that can be gained from such simple detectors as pitot tubes. If, for example, in addition to detecting the difference between static and velocity pressures, the pitot tube was able to also measure the Reynolds number, it would be able to approximate the shape of the velocity profile. An "intelligent pitot tube" of such capability could increase the accuracy of volumetric flow measurements.

Improved On-Line Analyzers

In the area of continuous on-line analysis, further development is needed to extend the capabilities of probe-type analyzers. The needs include the changing of probe shapes to obtain self-cleaning or to improve the ease of cleaning by using "flat tips." The availability of automatic probe cleaners should also be increased and the probe that is automatically being cleaned should be made visible by the use of sight flow indicators, so that the operators can check the cleaner's effectiveness.

More and more analyzers should become self-calibrating, self-diagnosing, and modular in their design. In order to lower maintenance costs, analyzers should also be made more modular for ease of replacement and should be provided with the intelligence to identify their defective modules. The industry should also explore the use of multiple-probe fiber-optic analyzers with multiplexed shared electronics.

Improving Operators' Displays

The control rooms of the coming decades will be more operator-friendly and more enterprise-wide optimization oriented. The human-machine interfaces (HMIs) in the control rooms are only as good as the ability of the operators to use them.

The hand, the psychological characteristics, the hearing, and color discrimination capability of the human operator must all be part of the design. Even more importantly, the design should also consider the personality and education of the average operator. Therefore, a well-designed HMI is the operator's window on the process.

In past decades, the operator's window on the process was little more than the faceplate of an analog controller and an annunciator. Today, when a single operator is expected to oversee the operation of processes having hundreds if not thousands of variables, the operator must be provided with diagnostic, trend, and historical data in an easily understandable and familiar format.

For that purpose, it is advisable to provide in the control room a large display panel, on which (as one of the options) the operation of the whole plant can be displayed. Using that graphic flowsheet, the operator should have the ability to focus in on any unit operation of interest. As the operator focuses in on smaller and smaller sections of the plant, the information content of the display should increase. In the process of "focusing," the operator must be able to select subsystems, individual loops, or loop components, such as a single control valve. At each level of scale, the display should identify all abnormal conditions (by color and flashing), while providing all the trend and status data for all related variables.

Another essential feature of modern control systems is their suitability for smooth growth of the plant. This capability is very important in process control because plants are ever expanding and therefore their control systems must also grow with the expansions. A modular approach to operator stations makes the expansion of the plant easier.

Optimization and Advanced Controls

In some ways we have already passed the age of the single-loop PID control. Yet in the coming decade much more improvement is expected both in multivariable unit operations control and in model-based optimization.

We all know that it is time to stop controlling flows, pressures, and temperatures and to start controlling and optimizing pumping stations, boilers, and chemical reactors. In the next decade, hopefully we will see the development of the universal software package for the various unit operations that can be adapted to specific plants just by the entering of design data for the particular process.

Plant-Wide Modeling and Optimization

In addition to its role in providing better control, process modeling and simulation can also improve the training of operators. If the simulation model is accurate and if it integrates the dynamics of the process with that of its control system, it can be used to train operators for plant startup without risking the consequences of their inexperience. Needless to say, the building of good models is expensive, but once prepared, they are very valuable.

Business- or enterprise-wide optimization includes the model of not only the manufacturing process, but also the optimization of the raw material supply chain and of the packaging and product distribution chain. This is a higher and more important level of optimization because it requires the simultaneous consideration of all areas of optimization and the finding of enterprise-wide operation strategies, which will keep all areas within their optimum areas of operation.

Plant-wide optimization also involves more than the optimization of the unit processes because it must also consider documentation, maintenance, production scheduling, and quality management considerations. Plant-wide optimization requires the resolution of the conflicting objectives of the operations and strategies.

Naturally, it should be kept in mind that modeling and optimization can only be achieved when the control loops are correctly tuned, the measurements are sampled fast enough, and interactions between loops have been eliminated.

Efficiency and Productivity Controllers

This handbook already describes a large number of methods to increase the efficiency of unit operations. For example, in the section describing the methods of pumping station optimization, it is pointed out that the lifetime operating cost of a pumping station is about a *hundred times higher than its initial cost.* The returns on the optimization of other unit operations are similar, although not that high. It is for this reason that in the coming decade, optimization is expected to increase.

When using multivariable envelopes for unit operation optimization, the individual variables of levels, pressures, and temperatures become only constraints, while the overall goal is to maximize the efficiency or productivity of the controlled process. New software packages are needed to "educate" and give "personality" to today's multivariable controllers to transform these general-purpose units into chemical reactor, distillation tower, compressor, or any other type of unit operation controllers.

Unit Operation Controllers of the Future

The next decade could bring a building-block approach to control systems. In this approach all "empty boxes" could be very similar, so that a unit operations controller that was, say, to optimize a dryer, could be converted to control an evaporator or a pumping station just by loading into it a different software package and connecting a different set of I/Os. Once the particular software package was loaded, the unit controller would be customized by a menu-driven adapter package, organized in a question-and-answer format.

During the customization phase, the user would answer questions on piping configuration, equipment sizes, material or heat balances, and the like. Such customization software packages could not only automatically configure and tune the individual loops but could also make the required relative gain calculations to minimize the interaction among them.

HISTORY OF THE HANDBOOK

The birth of this handbook was connected to my own work: In 1962 — at the age of 26 — I became the Chief Instrument Engineer at Crawford & Russell, an engineering design firm specializing in the building of plastics plants. C&R was growing and with it the size of my department also increased. Yet, at the age of 26 I did not dare to hire experienced people because I did not feel secure enough to lead and supervise older engineers.

So I hired fresh graduates from the best engineering schools in the country. I picked the smartest graduates and I obtained permission from C&R's president, Sam Russell, to spend every Friday afternoon teaching them. In a few years C&R not only had some outstanding process control engineers but had them at relatively low salaries.

By the time I reached 30, I felt secure enough to stop disguising my youth. So I shaved off my beard and threw away my phony, thick-rimmed eyeglasses, but my Friday's notes remained. They still stood in a 2-foot-tall pile on the corner of my desk.

"Does Your Profession Have a Handbook?"

In the mid-1960s an old-fashioned Dutch gentleman named Nick Groonevelt visited my office and asked: "What is that pile of notes?" When I told him, he asked: "Does your profession have a handbook?" "If it did, would I bother to prepare all these notes?" I answered with my own question. (Actually, I was wrong in giving that answer, because Behar's *Handbook of Measurement and Control* was already available, but I did not know about it.) "So," Nick proposed, "let me publish your notes and then the instrument engineers will have a handbook!" In 1968 the first edition of the *Instrument Engineers' Handbook* (*IEH*) was published.

In 1968, the Soviet tanks — which I fought in 1956 in Budapest — were besieging Prague, so I decided to dedicate the three volumes of the *IEH* to the Hungarian and Czech freedom-fighters. A fellow Hungarian-American, Edward Teller, wrote the preface to the first edition; Frank Ryan — the editor of *ISA Journal* — wrote the introduction. Because of the publication of the first edition of the *IEH*, in 1973 I was elected the youngest ISA fellow ever.

Later Editions

By the end of the 1970s the world of process control had changed. Pneumatics were on the way out, and new solutions like DCS control and on-line analyzers proliferated. It was time to revise the handbook. The second edition was published in 1985. It was well received.

By the mid-1990s the handbook was ready for another updated edition. By that time the process control market was becoming globalized, "smart" instruments had evolved, and such hardware inventions as fiber-optic probes and throttling solenoid valves proliferated. So I stopped teaching at Yale, cut back on consulting, and prepared the third edition. In this edition I also added a third volume to the two-volume set to cover all the evolving digital software packages, communication networks, buses, and optimization packages.

Work on the fourth edition of the *IEH* started in the new millenium, and the first volume on measurement and analysis was published in 2003. The second volume is in your hands now.

During the nearly four decades of its existence, the *Instrument Engineers' Handbook* (*IEH*) has become the most widely used reference source for the automation and control (ACE) engineering profession. During this period, our experience and our knowledge of control principles have penetrated all the fields of modern science and technology. I hope that the three volumes of the *IEH* will continue to play a major role in spreading this knowledge and understanding.

The Contents of the *IEH* Volumes

In 1968, this handbook started out as a two-volume reference set and, in that respect, it has not changed. The first volume still deals with measurement, the second with control. What is new is that the third volume deals with digital networks and software systems.

This fourth edition updates the information content of the previously published sections, incorporates the new developments of the last decade, and broadens the horizons of the work from an American to a global perspective. In the first volume, one chapter was devoted to each major measured variable including the detection of flow, level, temperature, pressure, density, viscosity, weight, composition, and safety sensors. Each subchapter (section) was devoted to the discussion of a different method of making that measurement.

This second volume of the *IEH* deals with process control and covers both control hardware and control strategies. The hardware discussed includes transmitters, controllers, control valves, and displays, including the design of control rooms. The chapters on control systems provide in-depth coverage both of the theory of control and of the unit processes of pumping, distillation, chemical reaction, heat transfer, and many others. The individual sections (subchapters) begin with a flowsheet symbol and if the subject matter is a hardware item, start with a feature summary.

This summary provides quick access to specific information on the available sizes, suppliers, ranges, and inaccuracies of the devices covered in that section. The reader is advised to turn to the section of interest and, based on the information in the feature summaries, decide whether the costs, inaccuracies, and other characteristics meet the requirements of the particular application.

We know that there is no greater resource than the combined knowledge and professional dedication of a well-educated new generation. We live in an age when technology can make the difference in overcoming the social and environmental ills on this planet. We live in an age when an inexhaustible and nonpolluting energy technology must be developed. It is hoped that this handbook will make a contribution toward these goals and in addition will also improve the professional standing of automation and control engineers around the world.

Béla Lipták
Stamford, Connecticutt
(*liptakbelaieh4@aol.com*)

DEFINITIONS

Absolute (Dynamic) Viscosity (μ)	Constant of proportionality between applied stress and resulting shear velocity (Newton's hypothesis).
Absorbance (*A*)	Ratio of radiant energy absorbed by a body to the corresponding absorption of a blackbody at the same temperature. Absorbance equals emittance on bodies whose temperature is not changing. ($A = 1 - R - T$, where R is the reflectance and T is the transmittance.)
Absorption	The taking in of a fluid to fill the cavities in a solid.
Accumulation	In safety and relief valve terminology, accumulation is the pressure increase over the maximum allowable working pressure of a tank or vessel during discharge through the pressure relief valve. It is given either in percentage of the maximum allowable working pressure or in pressure units such as pounds per square inch or in bars.
Accuracy	See **Inaccuracy**, which is the term used in this handbook.
Adaptive Control	See **Control, Adaptive**
Admittance (*A*)	The reciprocal of the impedance of a circuit. The admittance of an AC circuit is analogous to the conductance of a DC circuit. (Units: siemens.)
Adsorption	The adhesion of a fluid in extremely thin layers to the surfaces of a solid.
Alarm	A signal designed to alert, inform, guide, or confirm deviations from acceptable system status.
Alpha Curve	In resistance bulb terminology, it is the relationship of the resistance change of an RTD vs. temperature. In European alpha curves, the alpha value is 0.00385 ohms per degree C, while in American curves it is 0.00392.
Ampacity	The current (amperes) a conducting system can support without exceeding the temperature rating assigned to its configuration and application.
Amperometric Titration	Titration in which the end point is determined by measuring the current (amperage) that passes through
Amperometry	The process of performing an amperometric titration. The current flow is monitored as a function of time between working and auxiliary electrodes while the voltage difference between them is held constant; in other designs, the current is monitored as a function of the amount of reagent added to bring about titration of an analyte to the stoichiometrically defined end point. Also called constant potential voltametry.
Amplifier	A device that enables an input signal to control a signal source independent of the signal and is thus capable of delivering an output that is related to and is greater than the input signal.
Apparent Viscosity	Viscosity of a non-Newtonian fluid under given conditions. Same as consistency.
Approach	The different between the wet-bulb temperature of the ambient air and the water temperature leaving a cooling tower. The approach is a function of cooling tower capacity; a large cooling tower will produce a closer approach (colder leaving water) for a given heat load, flow rate, and entering air condition. (Units: °F or °C.)
Artificial Neural Networks (ANNs)	An ANN can learn complex functional relations by generalizing from a limited amount of training data; hence, it can thus serve as black-box model of nonlinear, multivariable static and dynamic systems and can be trained by the input–output data of these systems. ANNs attempt to mimic the structures and processes of biological neural systems.

	They provide powerful analysis properties such as complex processing of large input/output information arrays, representing complicated nonlinear associations among data, and the ability to generalize or form concepts–theory.
Attenuation	Loss of communication signal strength.
Auto-Manual Station	A device that enables an operator to select either the output of a controller or a manually generated signal.
Backlash	In mechanical devices, it is the relative movement between interacting parts that results from the looseness of these parts when motion is reversed.
Backplane	Physical connection between individual components and the data and power distribution buses inside a chassis.
Backpressure	In relief and safety valve terminology, it is the pressure on the discharge side of a pressure relief valve. This pressure is the sum of the superimposed and the built-up backpressures. The superimposed backpressure is the pressure that exists in the discharge piping of the relief valve when the valve is closed.
Balanced Safety Relief Valve	A safety relief valve with the bonnet vented to atmosphere. The effect of backpressure on the performance characteristics of the valve (set pressure, blow-down, and capacity) is much less than on a conventional valve. The balanced safety relief valve is made in three designs: (1) with a balancing piston, (2) with a balancing bellows, and (3) with a balancing bellows and an auxiliary balancing piston.
Balling Degrees	Unit of specific gravity used in the brewing and sugar industries.
Balun (Balanced/ Unbalanced)	A device used for matching characteristics between a balanced and an unbalanced medium.
Band Pass Filter	An optical or detector filter that permits the passage of a narrow band of the total spectrum. It excludes or is opaque to all other wavelengths.
Bandwidth	Data carrying capacity, the range of frequencies available for signals. The term is also used to describe the rated throughput capacity of a given network medium or protocol.
Barkometer Degrees	Unit of specific gravity used in the tanning industry.
Baseband	A communication technique where only one carrier frequency is used to send one signal at a time. Ethernet is an example of a baseband network. Also called narrow-band. Contrast with **Broadband**.
Basic Control	Continuously executed algorithms that drive the process or equipment to a specified state and keep it there, such as indicators, regulatory and device controls, and interlocks.
Batch	A quantity of material produced by the single execution of a batch process. A batch also refers to the intermediate materials during the manufacturing process.
Baumé Degree	A unit of specific gravity used in the acid and syrup industry.
Bay	The area between two bents of lines of framing members; usually longitudinal.
Bent	A line of structural framework composed of columns or ties; a bent may incorporate diagonal bracing members; usually transverse.
Blackbody	The perfect absorber of all radiant energy that strikes it. A blackbody is also a perfect emitter. Therefore, both its absorbance and emissivity (E) are unity. A blackbody radiates energy in predictable spectral distributions and intensities, which are a function of the blackbody's absolute temperature.
Black Box Model	See **Empirical Model**.
Blowdown	In case of cooling towers, it is the water discharged to control the concentration of impurities in circulated water. (Units: percentage of circulation rate.)
Blowdown (Blowback)	In relief valves, it is the difference between the set pressure and the reseating (closing) pressure of a pressure relief valve, expressed in percent of the set pressure, in bars, or in pounds per square inch.
Bode Diagram	A plot of the logarithm of gain or magnitude ratio and a plot of the logarithm of phase angles against the logarithm of frequency for a transfer function.
Boiling Point Rise	This term expresses the difference (usually in °F) between the boiling point of a constant composition solution and the boiling point of pure water at the same pressure. For example, pure water boils at 212°F (100°C) at 1 atmosphere, and a 35% sodium hydroxide solution boils at about 250°F (121°C) at 1 atmosphere. The boiling point rise is therefore 38°F (21°C). In a Dühring plot, the boiling point of a given composition solution is plotted as a function of the boiling point of pure water.

Bolometer	Thermal detector that changes its electrical resistance as a function of the radiant energy striking it.
Bonding	The practice of creating safe, high capacity, reliable electrical connectivity between associated metallic parts, machines, and other conductive equipment.
Brightness Pyrometer	A device that uses the radiant energy on each side of a fixed wavelength of the spectrum. This band is quite narrow and usually centered at 0.65 microns in the orange-red area of the visible spectrum.
British Thermal Unit (BTU)	The amount of heat required to raise the temperature of one pound of water by one degree Fahrenheit at or near 60° Fahrenheit.
Brix Degree	A specific gravity unit used in the sugar industry.
Broadband	A communication technique that multiplexes multiple independent signals simultaneously, using several distinct carriers. A common term in the telecommunication industry to describe any channel having a bandwidth greater than that of a voice-grade channel (4 kHz). Also called wideband. Contrast with **Baseband**.
BTU "Dry"	The heating value of a gas expressed on a "dry basis." The common assumption is that pipeline gas contains 7 pounds (or less) of water vapor per million standard cubic feet.
BTU "Saturated"	The heating value of a gas expressed on the basis that the gas is saturated with water vapor. This state is defined as the condition when the gas contains the maximum amount of water vapor without condensation, when it is at base pressure and 60°F.
Bubble Point	The temperature at which a mixture of liquids first starts to boil.
Built-Up Backpressure	In connection with safety relief valves, the variable backpressure that develops as a result of flow through the pressure relief valve after it opens. This is an increase in pressure in the relief valve's outlet line caused by the pressure drop through the discharge headers.
Burning	In combustion-related terminology, burning is when the flame does not spread or diffuse but remains at an interface where fuel and oxidant are supplied in proper proportions.
Calibrate	To ascertain the outputs of a device corresponding to a series of input values that the device is to receive, measure, or transmit.
Calibration Cycle	The application of known values as inputs to a device and the registering of the corresponding output readings over the range of that device, in both ascending and descending directions.
Calibration Traceability	The relationship of the calibration of an instrument to that of one or more instruments calibrated and certified by a national standardizing laboratory.
Campaign	The total production run of one product, for example, for a single order or season, consisting of one or more lots.
Capacitance (*C*)	The amount of charge, in coulombs, stored in a system necessary to raise the potential difference across it 1 volt; represented by the SI unit farad.
Capacitor	Device consisting of two conductors electrically isolated by an insulator. The conductors are called plates, and the insulator is referred to as the dielectric. The larger the capacitor, the smaller its impedance, and the more AC current will flow through it.
Cascade Control	See **Control, Cascade**.
Characteristic Impedance	The impedance that, when connected to the output terminals of a transmission line appear to be infinitely long, for there are no standing waves on the line, and the ratio of voltage to current is the same for each point of the line [nominal impedance of wave-guide].
Chatter	Rapid, abnormal reciprocating variations in lift during which the disc contacts the seat.
Chronopotentiometry	When the potential difference between a metallic measuring electrode and a reference electrode is monitored as a function of time. At the measuring electrode an oxidation or reduction of a solution species is taking place.
Closing Pressure (Reseat Pressure)	In relief and safety valve terminology, the pressure, measured at the valve inlet, at which the valve closes, flow is substantially shut off, and there is no measurable lift.
Coax	Jargon meaning "coaxial cable," consisting of a center wire surrounded by low K insulation, surrounded by a second shield conductor. It has low capacitance and inductance for transmission of high-frequency current.
Co-Current Operation	The process feed and steam (or other utility fluid) follow parallel paths through such processes as an evaporator train.

Countercurrent Operation	The feed and steam enter the evaporator train at opposite ends.
Cold Differential Test Pressure (CDTP)	The pressure at which the PRV is adjusted to open during testing. The CDTP setting includes the corrections required to consider the expected service temperature and backpressure.
Cold Junction	See **Reference Junction**.
Combustion Air Requirement Index (CARI)	This dimensionless number indicates the amount of air required (stoichiometrically) to support the combustion of a fuel gas. Mathematically the Combustion Air Requirement Index is defined by the equation below: $\text{CARI} = \dfrac{\text{Air/Fuel Ratio}}{\sqrt{\text{SG}}}$
Common Mode Interference	See **Interference, Common Mode**.
Common Mode Rejection	The ability of a circuit to discriminate against a common mode voltage.
Common Mode Voltage	A voltage of the same polarity relative to ground on both sides of a differential input.
Common Resource	An equipment entity that services more than one unit, either simultaneously (shared-use resource) or one at a time (exclusive-use resource).
Compliance	The reciprocal of stiffness.
Conductance (*G*)	The reciprocal of resistance. (Units: Siemens [formerly "mhos"].)
Conductivity (*g*)	The reciprocal of resistivity. All solids and liquids have some degree of conductivity. For the purpose of this section, any material above 1 micro Siemen/cm will be considered to be conductive (this includes most metals and water with any ions).
Conformity	The degree or closeness to which one curve approximates another curve. Conformity can be expressed as independent, terminal based, or zero based. Independent conformity is obtained when the calibration curve is so positioned as to minimize the maximum deviation between it and the specified curve. Terminal-based conformity is obtained when the two curves are so positioned that their readings coincide at zero and full span. Zero-based conformity is when they coincide only at zero.
Consistency	In terms of viscosity, the resistance of a substance to deformation. It is the same as viscosity for a Newtonian fluid and the same as apparent viscosity for a non-Newtonian fluid.
Constant Backpressure	In connection with safety relief valves, the backpressure that does not change under any condition of operation, whether the pressure relief valve is closed or open.
Control Action	For a controller, the nature of the change in the output of the controller, which is affected by the controller's input.
Control Action, Derivative (Rate) (*D*)	A control action in which the controller output is proportional to the rate of change of the input.
Control Action, Integral (Reset) (*I*)	A control action in which the controller output is proportional to the time integral of the input. In this case the rate of change of the controller output is proportional to the input.
Control Action, Proportional (*P*)	A control action in which there is a continuous linear relationship between the controller's output and its input.
Control, Cascade	Control configuration in which the output of one controller (the "master") is the set point of another controller (the "slave").
Control, Differential Gap	Control in which the output of the controller remains at a minimum or maximum value until the controlled variable reaches the top limit of a band or gap, at which point the output reverses and stays in that state until the controlled variable crosses the bottom limit of the gap. At that point, the controller output returns to its original condition.
Control, Feedback	Control in which a measurement is compared to a set point to produce an error signal. This error is acted upon in such a way as to reduce the magnitude of the error.
Control, Feedforward	Control in which one or more conditions that are outside the feedback loop and have an influence on the controlled variable are converted into some corrective action in order to minimize the deviations from the set point of the controlled variable.
Control Horizon	The number of future manipulated variable moves that are taken into account in developing the optimal MPC solution.
Controlled Variable	The variable that is detected to originate the feedback signal for the controller.
Controller, Direct Acting	A controller that increases its output signal when its measured variable (input signal) increases.

Controller, On-Off	A two-position controller, of which one of the two discrete output signal values is zero.
Controller, Program	A controller that automatically adjusts its set point to follow a prescribed program.
Controller, Ratio	A controller that maintains a predetermined ratio between two variables.
Controller, Reverse Acting	A controller that decreases its output signal as the value of its measured variable (input signal) increases.
Controller, Sampling	A controller using intermittent readings of the controlled variable (error or set point) to affect control action.
Controller, Self-Operated (Regulator)	A controller that derives the required energy for its operation from the system that it controls.
Controller, Time Schedule	A controller in which the set point (or the reference-input signal) automatically follows some predetermined time schedule.
Control Module	A set of equipment and functions that can carry out basic control. For example, a control loop consisting of one or more sensors, actuators, and control functions is a control module. It is also the lowest-level equipment grouping that acts as a single entity from a control standpoint, but cannot execute procedural elements. May be an individual measurement (with suitable signal conditioning or state names) or a grouping of directly coupled actuators with their associated measurements, alarms, and control actions, including subordinate Control Modules as appropriate. Examples are an uncontrolled temperature, a flow control loop, an automatic block valve with limit switches, or a header containing interlocked (mutually exclusive) block valves.
Control, Optimizing	Control that does not hold the controlled variable constant but seeks and maintains a value of the controlled variable that will result in the most advantageous operation of the process.
Control Recipe	An equipment-specific recipe that defines the production of a single batch. It is usually derived from a master recipe.
Control, Supervisory	Control in which a supervisory controlled periodically applies some corrective action, such as set point changes, to a number of controllers.
Control System, Multi-Variable	A control system that uses two or more process variable measurement signals to affect control.
Control, Time Proportioning	Control in which the outputs are periodic pulses whose duration is varied in relation to the actuating error signal.
Control, Velocity Limiting	Control in which the rate of change of some variable is prevented from exceeding a predetermined limit.
Conventional Safety Relief Valve	A safety relief valve with the bonnet vented either to atmosphere or internally to the discharge side of the valve. The performance characteristics (set pressure, blow-down, and capacity) are directly affected by changes of the backpressure on the valve.
Coordination Control	Control functions existing at multiple levels to schedule batches and manage recipes, procedural control execution, equipment entity allocation, and batch data.
Corner Frequency	In the asymptotic form of the Bode diagram, that frequency that is indicated by a break point. It is the junction of two confluent straight lines asymptotic to the log gain curve.
Coulometry	A process of monitoring analyte concentration by detecting the total amount of electrical charge passed between two electrodes that are held at constant potential or when constant current flow passes between them.
Countercurrent Operation	The process feed and steam (or other utility fluid) enter at opposite ends as they flow through such processes as an evaporator train.
Counterflow Cooling Tower	A cooling tower in which the airflow is in the opposite direction from the fall of water through the water cooling tower.
CPVC (Chlorinated Polyvinyl Chloride)	A low cost, reasonably inert polymer used for some noninsertion sensors. It is easily solvent-welded. The max temperature range is about 225°F.
Cross-Flow Cooling Tower	A cooling tower in which airflow through the fill is perpendicular to the plane of the falling water.
Crystallography	How the atoms are arranged in the object; a direct relation exists between these arrangements and materials properties (conductivity, electrical properties, strength, etc.).
Curie (Ci)	A unit of radiation source size, corresponding to 37 billion disintegrations per second.
Damping	The progressive reduction or suppression of oscillation in a device or system.
Damping Factor	In the case of the free oscillation of a second-order linear system, it is the

	ratio of the greater by the lesser of a pair of consecutive swings of the output in opposite directions (without sign) about an ultimate steady-state value.
Data Servers	A standard interface to provide data exchange between field devices and data clients.
Deadband	The range through which an input can be varied without causing an observable change in the output. Deadband is the range through which the measurand may be varied by reversing direction without producing an observable response in the output signal.
Dead Time	See **Time, Dead**.
Dead Zone	See **Zone, Dead**.
Deflagration or Explosion	A process where the flame front advances through a gaseous mixture at subsonic speeds.
Deionized Water	Water of extremely high purity, with few ions to carry current. If exposed to air for any significant period, it will have a conductivity of about 5 micro Siemens/cm due to dissolved CO_2.
Demultiplexing	Separating of multiple input streams that were multiplexed into a common physical signal back into multiple output streams.
Derivative Gain	See **Gain, Derivative**.
Derivative Time	See **Time, Derivative**.
Design Pressure	This pressure is equal to or less than the maximum allowable working pressure. It is used to define the upper limit of the normal operating pressure range.
Detonation	A process where the advancement of the flame front occurs at supersonic speeds.
Device Description (DD)	In digital systems, a clear and unambiguous, structured text description that allows full utilization/operation of a field device by a host/master without any prior knowledge of the field device.
Device Description Language (DDL)	In the Foundation fieldbus technology concept, it is the definition and description of function blocks and their parameters.
Dew Point	Saturation temperature of a gas–water vapor mixture.
Dew Point Temperature (DPT)	The temperature at which condensation begins if air is cooled under constant pressure.
Dielectric	A material that is an electrical insulator or in which an electric field can be sustained with a minimum of dissipation of power. Dielectric materials include metal-oxides, plastics, and hydrocarbons.
Dielectric Compensation	A scheme by which changes in insulating liquid composition or temperature can be prevented from causing any output error. It requires a second sensor and homogeneous liquid.
Dielectric Constant	A unit expressing the relative charge storage capability of various insulators. Full vacuum is defined as 1.0, and all gases are indistinguishable from each other for practical purposes. TFE has a dielectric constant of 2.0, cold water about 80. It has no units because it is the ratio of the dielectric constant of a substance to that of vacuum. For the dielectric values of selected materials refer to Table 3.3p of the first volume of this handbook.
Diode	A two-terminal electronic (usually semiconductor) device that permits current flow predominantly in only one direction.
Direct Acting Controller	See **Controller, Direct Acting**.
Discontinuity	An abrupt change in the shape [or impedance] of a wave-guide [creating a reflection of energy].
Distance/Velocity Lag	A delay attributable to the transport of material or to the finite rate of propagation of a signal.
Dither	A useful oscillation of a small magnitude, which is introduced to overcome the effect of friction, hysteresis, or recorder pen clogging.
Drift	An undesired change in the output of a device, which occurs over time and is unrelated to the input, the environment, or the load. In the case of cooling towers it is the water loss due to liquid droplets entrained in exhaust air. Usually under 0.2% of circulated water flow rate.
Drift Eliminator	An assembly constructed of wood or honeycomb materials that serves to remove entrained moisture from the discharged air.
Droop	See **Offset**.
Dry-Bulb Temperature (TDB)	The temperature of air measured by a normal thermometer.
Dust-Ignition-Proof	Enclosed in a manner to exclude ignitable amounts of dust or amounts that might affect performance. Enclosed so that arcs, sparks, or heat otherwise generated or liberated inside of the enclosure will not cause ignition of exterior accumulations or atmospheric suspensions of dust.
Dynamic Gain	See **Gain, Dynamic**.

Dynamic Matrix	A matrix built from step responses to predict the changes in the process output that result from the moves of the manipulated variable over the control horizon.
Economy	In case of the evaporation process, this term is a measure of steam use and is expressed in pounds of vapor produced per pound of steam supplied to the evaporator train. For a well-designed evaporator system the economy will be about 10% less than the number of effects; thus, for a triple-effect evaporator the economy will be roughly 2.7.
Effective Coefficient of Discharge	This is a coefficient used to calculate the minimum required discharge area of the PRV.
Electrochemical Process	The changes in voltage or current flow that occur between two electrodes in a solution (electrolyte) over time. The oxidation or reduction of the analyte provides data that are related to concentration.
Electrolytic Probe	Probe that is similar to a galvanic probe, except that a potential is applied across the electrodes and the electrodes are not consumed. Dissolved oxygen detection is a primary application of this type of probe.
Electromagnetic Wave (Energy)	A disturbance that propagates outward from any electric charge that oscillates or is accelerated; far from the charge it consists of vibrating electric and magnetic fields that move at the speed of light and are at right angles to each other and to the direction of motion.
Electron Microscope	A scientific instrument that uses a beam of highly energetic electrons to examine objects on a very fine scale.
Element, Final Control	The controlling element that directly changes the value of the manipulated variable. It can be a control valve, a damper, an electric power modulator, or a variable-speed pump.
Element, Primary	The element that converts the measured variable into a form which can be measured.
Elevated Range, Elevation	See **Range, Suppressed Zero**.
Elevation, Zero	See **Zero Elevation**.
Emissivity or Emittance (*E*)	The emissivity of an object is the ratio of radiant energy emitted by that object divided by the radiant energy that a blackbody would emit at that same temperature. If the emittance is the same at all wavelengths, the object is called a gray body. Some industrial materials change their emissivity with temperature and sometimes with other variables also. Emissivity always equals absorption and it also equals 1 minus the sum of reflectance and transmittance: $(E = A = 1 - T - R)$
Empirical Model	This type of model can be used for processes for which no physical insight is available or used. This model structure belongs to families that are known to have good flexibility and have been "successful in the past." The parameters of the models are identified based on measurement data. A complete mechanistic model is constructed from *a priori* knowledge.
Equipment Control	The procedural, basic, and coordination control capability that enables an equipment entity to perform its function. It is not part of a recipe but may be directed by a recipe.
Equipment Entity	A set of process and control equipment and associated control capability that has been grouped together to provide some specified process or equipment functionality.
Equipment Module	A group of equipment that can carry out a finite number of specific minor processing activities. An equipment module may include one or more control modules. It is typically centered around a piece of process equipment, such as a weigh tank, heat exchanger, filter, or weighing scale. It is an equipment entity incorporating necessary devices, control modules, and application-specific states and modes to execute some basic process-oriented task or group of tasks. It may include equipment procedural elements (typically phase logic) and/or subordinate equipment modules but may not overlap other equipment modules. Examples are a common (exclusive-use) weigh tank, a recirculation/transfer pump within a unit, and a shared-use ingredient dosing system.
Equivalent Time Sampling (ETS)	The process that captures high-speed electromagnetic events in real time (nanoseconds) and reconstructs them into an equivalent time (milliseconds), which allows easier measurement with present electronic circuitry.
Ergonomics	An empirical approach to the study of human–machine interactions, with the

	objective of improving efficiency and reducing strain and discomfort of the operator.
Error	The difference between the measurement signal and its ideal value. A positive error denotes that the indication of the instrument is greater than the ideal value.
Error, Systematic	An error which, when a number of measurements are made under the same conditions, measuring the same value of a given quantity, either remains constant or varies according to some definite law when conditions change.
Error, Zero	The error of a device when it is operating at the lower range-value, which is commonly referred to as its zero.
Ethernet	A baseband local area network specification developed by Xerox Corporation, Intel, and Digital Equipment Corporation to interconnect computer equipment using coaxial cable and transceivers.
Evaporation Loss	Water evaporated from the circulating water into the atmosphere in the cooling process. (Unit: percentage of total GMP.)
Excitation, Maximum	The maximum excitation that can be applied to a device at rated operating conditions without causing damage or performance degradation beyond specified tolerances.
Exception Handling	Procedures and functions that deal with conditions outside the normal or desired behavior of a process.
Explosion Proof	All equipment that is contained within enclosures that are strong enough to withstand internal explosions without damage and tight enough to confine the resulting hot gases so that they will not ignite the external atmosphere. This is the traditional method of protection and is applicable to all sizes and types of equipment.
Fan Pitch	The angle a fan blade makes with the plane of rotation. (Unit: degrees from horizontal.)
Fan Stack (Cylinder)	Cylindrical or modified cylindrical structure in which the fan operates. Fan cylinders are used on both induced draft and forced draft axial-flow propeller-type fans.
Farad (*F*)	A unit of capacitance. Because this is a very large unit, a unit equal to one trillionth of a farad (called a pico Farad, symbol: “pF”) is commonly used in RF circuits.
Feedback Control	See **Control, Feedback**.
Feedforward Control	See **Control, Feedforward**.
FEP (Fluorinated Ethylene Propylene)	A fluorocarbon that is extremely chemically inert, melts at a reasonable temperature, and can be plastic-welded fairly easily. Hard to bond with adhesives. Max. temperature range limited to the 300°F (150°C) area.
Fieldbus	An all-digital, two-way, multi-drop communications system for instruments and other plant automation equipment.
Firewall	A router or access server, designated as a buffer between any public networks and a private network.
Flash Point	The lowest temperature at which a flammable liquid gives off enough vapors to form a flammable or ignitable mixture with air near the surface of the liquid or within the container used. Many hazardous liquids have flash points at or below room temperature. They are normally covered by a layer of flammable vapors that will ignite in the presence of a source of ignition.
Fluidity	Reciprocal of absolute viscosity; unit in the cgs system is the rhe, which equals 1/poise.
Floutter	Rapid, abnormal reciprocating variations in lift, during which the disc does not contact the seat.
Forced Draft Cooling Tower	A type of mechanical draft water cooling tower in which one or more fans are located at the air inlet to force air into the tower.
Forman	Vocal-tract resonance.
Formula	A part of recipe that include process inputs, process parameters, and process outputs. The list of process inputs, outputs, and data (operating set points, reported values, timing, etc.) required to execute the batch procedure.
Frequency, Damped	The frequency of a damped oscillatory response of a system resulting from a nonoscillatory stimulus.
Frequency Response Characteristics	The frequency-dependent relation, both in terms of gain and phase, between steady-state sinusoidal inputs and the resulting steady-state sinusoidal outputs.
Fuel Cells	Cells that convert the chemical energy of fuels such as hydrogen into electrical energy, while the electrode and the electrolyte remain unaltered. Fuel is converted at the anode into hydrogen ions, which travel through the electrolyte to

	the cathode, and electrons, which travel through an external circuit to the cathode. If oxygen is present at the cathode, it is reduced by these electrons, and the hydrogen and oxygen ions eventually react to form water.
Function Block	A logical processing unit of software that has one or more input and output parameters.
Fuzzy Logic Modeling	This type of model is used for processes that are not fully understood. It is a linguistically interpretable rule-based model, which is based on the available expert knowledge and measured data.
Gain, Closed Loop	The ratio of a change in output to a change in input of a closed loop system at a specified frequency.
Gain, Derivative (Rate Gain)	The ratio of the maximum gain that results from the proportional plus derivative actions in a controller to the gain due to the proportional mode alone.
Gain, Dynamic	For a sinusoidal signal, it is the magnitude ratio of the steady-state amplitude of the output signal from a device to the amplitude of the input signal to that device.
Gain, Loop	At a specified frequency, it is the ratio of the magnitude of the change in the feedback signal to the change in its corresponding error signal.
Gain (Magnitude Ratio)	For a linear system, it is the ratio of the amplitude of a steady-state sinusoidal output and the amplitude of the input that caused it.
Gain, Proportional	The ratio of the change in the output of a controller with proportional action to the change in the input of the controller.
Galvanic Probe	A probe in which no external voltage is applied across its electrodes and the current flows as the cell is depolarized when diffusion of the analyte occurs. Electrodes are consumed during this operation and require periodic replacement.
General Recipe	A type of recipe that expresses equipment- and site-independent processing requirements. It is the highest level of recipe.
Gray Body	An object with a constant emittance of less than unity. This emittance is constant at all wavelengths (over that part of the spectrum where the measurement takes place). This means that gray-body radiation curves are identical to those of a blackbody, except that they are dropped down on the radiated power density scale.
Gross Calorific Value	The heat value of energy per unit volume at standard conditions, expressed in terms of British thermal unit per standard cubic feet (BTU/SCF) or as kilocalorie per cubic Newton meters ($\text{Kcal/N} \cdot \text{m}^3$) or other equivalent units.
Ground	A conducting connection, whether intentional or accidental, between an electrical circuit or equipment and the earth, or to some conducting body that serves in place of earth. (See NFPA 70-100.)
Ground Fault Protector	Device used to open ungrounded conductors when high currents, especially those due to line-to-ground fault currents, are encountered.
Guard	The "electronic guard" (called a "shield" in some RF level literature) consists of a concentric metallic element with an applied voltage that is identical to the voltage on the conductor that it is "guarding." This negates the capacitance between the guarded conductor and the outside world.
Guided Wave Radar (GWR)	A contact radar technology where Time Domain Reflectometry (TDR) has been developed into an industrial-level measurement system where a probe immersed in the medium acts as the waveguide.
Hagen–Poiseuille Law	Defines the behavior of viscous liquid flow through a capillary.
HART	Is an open, smart field instrumentation protocol that is a de facto fieldbus. It imposes a 1200-bit-per-second digital signal on a twisted pair of wires carrying a 4- to 20-mA input.
Home Run Wiring	Wire between the cabinet where the Fieldbus host or centralized control system resides and the first field junction box or device.
Hub (Shared)	Multi-port repeater joining segments into a network.
Hunting	An undesirable oscillation of some appreciable magnitude, prolonged even after the external stimuli that caused it disappear.
Hygrometer	An apparatus that measures humidity.
Hygroscopic Material	A material with great affinity for moisture.
Hysteresis	That property of an element or system that shows the dependence of its output for a given excursion of the input on the history of prior excursions and on the direction of the current traverse. Hysteresis in a damper or valve is caused mostly by friction. Hysteresis in

	a process transmitter is the maximum difference between readings taken at the same measurand with upscale and downscale approaches to the readings. It is due to inelastic qualities of the sensing element, friction, and backlash.
Impedance	Maximum voltage divided by maximum current in an alternating current circuit. Impedance is composed of resistive, inductive, and capacitive components. Like direct current circuits, the quantity of voltage divided by current is expressed in ohms.
Inaccuracy	The degree of conformity of a measured or indicated value to a recognized standard value. It is usually expressed in terms of measured variable, percent of span, percent of upper-range value, or percent of actual reading.
Inaccuracy Rating	The quantity that defines the limit that errors will not exceed while the device is operating under specified conditions. Can be expressed in terms of the measured variable (for example ±0.5°F), in terms of percent span, in percent upper-range value, in percent of scale length, or in percent of actual output reading.
Infrared	That portion of the spectrum whose wavelength is longer than that of red light. Only the portion between 0.7 and 20 microns gives usable energy for radiation detectors.
Insulation Resistance	The resistance measured across the insulation at reference operating conditions when a specified direct current voltage is applied. The goal of the measurement is to determine whether the expected leakage current will or will not be excessive.
Integral Action Rate (Reset Rate)	For PI or PID controllers, it is the ratio of the initial rate of output change caused by the integral action to the change in steady-state output caused by the proportional action. Reset rate is often expressed as the number of repeats per minute because it equals the number of times the proportional correction is repeated by the integral action every minute.
Integral Controller	See **Controller, Integral**.
Interface	(1) Shared boundary. For example, the physical connection between two systems or two devices. (2) Generally, the point of interconnection of two components and the means by which they must exchange signals according to some hardware or software protocol.
Interoperability	A marketing term with a blurred meaning. One possible definition is the ability for like devices from different manufacturers to work together in a system and be substituted one for another without loss of functionality at the host level (HART).
Intrinsic Safety	When under all conditions, the available energy is limited to levels that are too low to ignite the hazardous atmosphere. This method is useful only for low-power equipment such as instrumentation, communication, and remote-control circuits.
Kinematic Viscosity (υ)	Dynamic viscosity/density = $\upsilon = \mu/\rho$.
Lambda	The desired closed loop time constant, often set to equal the loop lag time.
Latency	Latency measures the worst-case maximum time between the start of a transaction and the completion of that transaction.
Lift	The rise of the disc in a pressure relief valve.
Line Driver	Inexpensive amplifier and signal converter that conditions digital signals to ensure reliable transmissions over extended distances without the use of modems.
Linearity, Independent	The maximum deviation of a calibration curve from a straight line that is so positioned as to minimize that maximum deviation.
Linearity, Terminal-Based	The maximum deviation of a calibration curve from a straight line, if the line is so positioned that they coincide at the upper and lower range values.
Linearity, Zero-Based	The maximum deviation of a calibration curve from a straight line, if the line is so positioned that they coincide at the lower range value and to minimize the maximum deviation.
Linear Programming	A mathematical technique for solving a set of linear equations and inequalities in order to maximize or minimize an additional function called an objective function.
Loop Gain Characteristics	For a closed loop, it is the characteristic curve of the ratio of the change in the return signal to the change in the corresponding error signal.
Loop Transfer Function	This value of a closed control loop is obtained by taking the ratio of the Laplace transform of the return signal

	to the Laplace transform of the corresponding error signal.
Lot	Products produced by a set of similar batches, usually using the same master recipe. A collection of batches prepared using the same recipe. Typically, all batches of a lot are prepared from the same homogeneous source of raw material.
Louvers	Assemblies installed on the air inlet faces of a tower to eliminate water splash-out.
Lower Explosive Limit (LEL)	The lowest concentration of gas or vapor in air where, once ignition occurs, the gas or vapor will continue to burn after the source of ignition has been removed.
Lower Range-Limit	See **Range-Limit, Lower**.
Low-Pass Filters	Filters that are used to remove high-frequency interference or noise from low-frequency signals.
Makeup	In cooling towers it is water added to replace loss by evaporation, drift, blowdown, and leakage. (Unit: percentage of circulation rate.)
Manchaster	A digital signaling technique that contains a signal transition at the center of every bit cell.
Manipulated Variable	The variable that is manipulated in order to reduce the controller's error and thereby to bring the controlled variable closer to set point.
Manufacturing Range	A range around the specified burst pressure within which the marked or rated burst pressure must fall. Manufacturing range is not used in ISO standards.
Master Recipe	A recipe for producing a batch of products using a particular set of process equipment.
Maximum Allowable Operating Pressure (MAOP)	The maximum pressure expected during normal operation.
Maximum Allowable Working Pressure (MAWP)	This is the maximum pressure allowed for continuous operation. As defined in the construction codes (ASME B31.3) for unfired pressure vessels, it equals the design pressure for the same design temperature. The maximum allowable working pressure depends on the type of material, its thickness, and the service conditions set as the basis for design. The vessel may not be operated above this pressure or its equivalent at any metal temperature other than that used in its design; consequently for that metal temperature, it is the highest pressure at which the primary pressure relief valve can be set to open.
Measurand	The physical parameter to be measured.
Mechanical Draft Cooling Tower	A tower through which air movement is effected by one or more fans. There are two general types of such towers: those that use forced draft with fans located at the air inlet and those that use induced draft with fans located at the air exhaust.
Mechanical Emissivity Enhancement	Mechanically increasing the emissivity of a surface to near-blackbody conditions (using multiple reflection).
Micron	Equivalent to 0.001 millimeters or 10,000 Ångstrom units. A unit used to measure wavelengths of radiant energy.
Model-Based Control (MBC)	In model-based control (MBC), a process model is used to make control decisions. The controller uses this model of the process to calculate a value for the manipulated variable, which should make the controlled variable behave in the desired way. The "inverse" nomenclature arises from how the model is used. In a normal modeling approach, one specifies the process input, and the model predicts the process output response. In contrast, MBC determines the process input (manipulated variable) that will cause a desired process output response (controlled variable value) to occur. This is the model inverse.
Model Predictive Control (MPC)	A model-based control technique that uses process output prediction and calculates consecutive controller moves in order to satisfy control objectives.
Modem	Modulator–demodulator. Device that converts digital and analog signals. At the source, a modem converts digital signals to a form suitable for transmission over analog communication facilities. At the destination, the analog signals are returned to their digital form. Modems allow data to be transmitted over voice-grade telephone lines.
Modulation	The result of a process whereby a characteristic of one wave is varied in accordance with some characteristics of another wave.
Morphology	The shape and size of the particles making up the object; a direct relation exists between these structures and materials properties (ductility, strength, reactivity, etc.)

Multiple-Effect Evaporation	Multiple-effect evaporations use the vapor generated in one effect as the energy source to an adjacent effect. Double- and triple-effect evaporators are the most common; however, six-effect evaporation can be found in the paper industry, where kraft liquor is concentrated, and as many as 20 effects can be found in desalinization plants.
Multiplexing	A method that allows multiple logical signals to be transmitted simultaneously across a single physical channel. Compare with **Demultiplexing**.
Narrow-Band Pyrometer	A radiation pyrometer that is sensitive to only a narrow segment of wavelengths within the total radiation spectrum. Optical pyrometers are one of the devices in this category.
Natural Draft Cooling Tower	A cooling tower in which air movement is essentially dependent upon the difference in density between the entering air and internal air. As the heat of the water is transferred to the air passing through the tower, the warmed air tends to rise and draw in fresh air at the base of the tower.
Net Calorific Value	The measurement of the actual available energy per unit volume at standard conditions, which is always less than the gross calorific value by an amount equal to the latent heat of vaporization of the water formed during combustion.
Network	All of the media, connectors, and associated communication elements by which a communication system operates.
Neurons	A brain cell that passes information by receiving and transmitting electrical impulses. Nodes in neural networks serve similar functions.
Neutral Zone	See **Zone, Neutral**.
Newton	The internationally accepted unit of force, defined as the force required to accelerate one kilogram by one m/sec^2. It equals 0.2248 pound-force or about 4 ounces.
Nodes	Processing elements in neural networks.
Nominal Tonnage (Cooling)	In cooling towers, one nominal ton corresponds to the transfer of 15,000 BTU/hr (4.4 kW) when water is cooled from 95 to 85°F (35 to 29.4°C) by ambient air having a wet-bulb temperature of 78°F (25.6°C) and when the water circulation rate is 3 GPM (11.3 lpm) per ton.
Nonfragmenting Disc	A rupture disc design that, when burst, does not eject fragments that could interfere with the operation of downstream equipment (i.e., relief valves).
Nonincendive Equipment	Equipment that in normal operations does not constitute a source of ignition. Therefore, its surface temperature shall not exceed ignition temperature of the specified gas to which it may be exposed, and there are no sliding or make-and-break contacts operating at energy levels capable of causing ignition. Used for all types of equipment in Division 2 locations. Relies on the improbability of an ignition-capable fault condition occurring simultaneously with an escape of hazardous gas.
Nonlinearity	Nonlinearity is expressed as a percent of full range output (%FRO) and is determined from the maximum difference between a datum on the output vs. measurand plot and a best straight line (BSL) drawn through the data.
Normal Mode Rejection	The ability of a circuit to discriminate against normal mode voltage. It can be expressed as a dimensionless ratio, as a scalar ratio, or in decibels (20 times the $\log_{10}$ of that ratio).
Normal Mode Voltage	A voltage across the input terminals of a device.
Offset	When in a controller the set point is fixed and in the steady state there is a deviation, it is called offset. The change in offset that results from a no-load to a full-load condition is often called "droop."
Oil Immersion	Equipment is submerged in oil to a depth sufficient to quench any sparks that may be produced. This technique is commonly used for switch-gears, but it is not utilized in connection with instruments.
On-Off Controller	See **Controller, On-Off**.
Open Loop Gain	The steady-state gain of a control loop when the other control loop(s) is(are) in manual. (Their control valve opening is constant.)
Operating Pressure	The operating pressure of a vessel is the pressure, in pounds per square inch gauge, to which the vessel is usually subjected in service. A processing vessel is usually designed for a maximum allowable working pressure, in pounds per square inch gauge, that will provide a suitable margin above the operating pressure in order to prevent any undesirable operation of the relief device. It is suggested that this margin be approximately 10%, or 25 PSI (173 kPa), whichever is greater. Such a margin will

	be adequate to prevent the undesirable opening and operation of the pressure relief valve caused by minor fluctuations in the operating pressure.
Operating Pressure Margin	The margin between the maximum operating pressure and the set pressure of the PRV.
Operating Pressure Ratio	The ratio of the maximum operating pressure to the set pressure of the PRV.
Operating Ratio of a Rupture Disc	The ratio of the maximum operating pressure to the marked burst pressure expressed as a percentage (common U.S. definition). The ratio of the maximum operating pressure to the minimum of the performance tolerance expressed as a percentage (common ISO definition).
Operation	A procedure that controls the execution of a number of phases in a batch. A major programmed processing action or set of related actions, normally consisting of one or more phases.
Optical Pyrometer	Also called brightness pyrometer, it uses a narrow band of radiation within the visible range (0.4 to 0.7 microns) to measure temperature by color matching and other techniques.
Optimizing Control	See **Control, Optimizing**.
Overpressure	The pressure increase over the set pressure of the primary relief device. When the set pressure is the same as the maximum allowable operating pressure (MAOP), the accumulation is the same as the overpressure. Pressure increase over the set pressure of the primary relieving device is overpressure. *Note*: From this definition it will be observed that when the set pressure of the first (primary) safety or relief valve is less than the maximum allowable working pressure of the vessel, the overpressure may be greater than 10% of set pressure.
Overrange Limit	The maximum input that can be applied to a device without causing damage or permanent change in its performance.
Partial Pressure	In a mixture of gases, the partial pressure of one component is the pressure of that component if it alone occupied the entire volume at the temperature of the mixture.
Pascal-Seconds (Pas)	Internationally accepted unit of absolute (dynamic) viscosity. Pas = Newton-sec/m^2 = 10 poise = 1000 centipoise.
PDVF, Polyvinylidene Fluoride	This fluorocarbon has substantially lower temperature limits than the others (250°F or 120°C) and is less inert chemically. It is dissolved by the ketones (acetone, MEK, MIBK) and attacked by benzene and high concentrations of sulfuric acid. The most insidious enemy is caustic, which causes brittleness and cracking. It has much better toughness and abrasion resistance, than the other fluorocarbons, as well as unique electrical properties ($K = 8$).
PE (Polyethylene)	A low temperature insulation that is compatible with a wide range of corrosives but is attacked by most petroleum products. Generally limited to situations where fluoro- and chlorocarbons are not allowed, such as the tobacco and nuclear power industries. Max. allowable temperature is in the 180°F (80°C) area.
PEEK (Polyether Etherketone)	A high-temperature, injection-molded polymer that is chemically quite inert. This material has wide chemical application. Temperature capability is high at 450 to 500°F (225 to 260°C). Avoid any liquids with "phenol" in their name. Adhesive bonding to the molded parts would be difficult.
Perceptron	A transfer function used in some neural networks.
PFA (Per-Fluoro-Alkoxy)	A fluorocarbon that is quite inert chemically, melts at a fairly high temperature, and is easily plastic welded. It can be used up to 550°F (290°C), but as a probe insulation it is generally limited to 350°F (175°C) due to bonding limitations with the metal rod.
Phase	A set of logic steps that completes a major processing function, such as charge, mix, heat, and reaction. A batch is usually in a stable state at the end of a phase. It is the lowest level of procedural control to accomplish a process-oriented task. Phases may be further subdivided into equipment-oriented steps and transitions for executing its defined task, as described in European standard IEC 60848 (1988) for specification of sequential function charts. Normally, the phase boundaries represent points of process transition, hold, or activity. The boundaries define major milestones and possible points of safe intervention. Phases may exist either as part of a recipe procedure (recipe phase) or independently for equipment control (equipment phase); however, any constituent steps are always part of the equipment phase.

Phase Difference Sensor (PDS)	A contact radar technology; unlike TDR-based systems, which measure using sub-nanosecond time intervals, PDS derives level information from the changes in phase angle.
Phase Shift	Of a signal, it is a change of phase angle with transmission. Of a transfer function, it is a change of phase angle with test frequency.
Phoneme	The sound of a human voice.
Photodetector	Measures thermal radiation by producing an output through release of electrical changes within its body. Photodetectors are small flakes of crystalline materials, such as CdS or InSb, which respond to different portions of the spectrum, consequently showing great selectivity in the wavelengths at which they operate.
Pixel (“Picture Element”)	Square dot, used in machine vision and camera technology.
Plenum	Air distribution ducting, chamber, or compartment.
Poise (μ)	Unit of dynamic or absolute viscosity (dyne-sec/cm^2).
Poiseuille (pi)	Suggested name for the new international standard unit of viscosity, the Pascal-second.
Polarography	Process for monitoring the diffusion current flow between working and auxiliary electrodes as a function of applied voltage as it is systematically varied. Concentration of analyte allows for flow of the diffusion current, which is linearly dependent on the analyte concentration. Polarography can be applied using direct current, pulsed direct current, or alternating current voltage excitation wave forms. Dissolved oxygen determination is an example of an application for which polarography is used.
Potentiometry	When no current is passing between electrodes. Examples: ORP, pH, selective-ion electrodes. The potential difference (at zero current) is monitored between the measuring and reference electrodes.
Potting	Potting compound completely surrounding all live parts and thereby excluding the hazardous atmosphere has been proposed as a method of protection. There is no known usage except in combination with other means.
Power Factor	The ratio of total true power (watts) to the apparent power total rms (root-mean-square) volt-amperes.
PP (Polypropylene)	Similar to PE. Used for low cost, and where fluoro- and Chlorocarbons are excluded. Max. temperature is in the area of 200°F.
Pressure, Design	The pressure that is used in the design of a device for the purpose of determining the required minimum wall thickness and other characteristics for a given maximum working pressure (MWP) at a given temperature.
Pressure, Maximum Working (MWP)	The very maximum pressure that is permissible in a device under any circumstances during operation, at a specified temperature.
Pressure Relieving Device	The broadest category in the area of pressure relief devices, includes rupture discs and pressure relief valves of both the simple spring-loaded type and certain pilot-operated types.
Pressure Relief Valve (PRV)	A generic term that might refer to relief valves, safety valves, and pilot-operated valves. The purpose of a PRV is to automatically open and to relieve the excess system pressure by sending the process gases or fluids to a safe location when its pressure setting is reached.
Pressure, Surge	It is the sum of the operating pressure plus a pressure increment that develops for a very short time while pumps are starting or valves are closing.
Primary Standard	A measuring instrument calibrated at a national standard laboratory such as NIST and used to calibrate other sensors.
Procedural Control	Control that sequentially directs subordinate procedural elements or basic controls to execute the steps required by its defined process-oriented task.
Procedure	A user-defined set of instructions that define the strategy for making a single batch of a particular type or grade of product.
Process Cell	A set of equipment required for production of one or more batches. It usually consists of one or more units. It is a grouping of equipment that comprises one or more complete trains and defines the immediate local domain for production scheduling (analogous to a work cell in discrete manufacturing).
Process Inputs	Identity and quantity of raw materials and other resources required to make a batch. Other resources include energy and manpower requirements.
Process Outputs	Identity and quantity of products and/or energy produced at the end of a batch.
Process Parameters	Variables, such as temperature, pressure, and time, that are set points and

	comparison values needed for the production of a batch.
Proof	A specific gravity unit used in the alcohol industry.
Proportional Band	In a proportional-only controller, it is the change required in the input signal to result in a full range change in the controller's output.
Proportional Controller	See **Controller, Proportional**.
Protocol	Formal description of a set of rules and conventions that govern how devices on a network exchange information. In communications it is a set of conventions or rules that must be adhered to by both communicating parties to ensure that information being exchanged between two parties is received and interpreted correctly.
Prosodic Characteristics	The pitch of voice; the duration and intensity of speech.
Prosody	Accent and voice modulation.
Psychrometer	An instrument used primarily to measure the wet-bulb temperature.
(P)TFE (Tetra-Fluoro-Ethylene)	In the abbreviation, the "P" stands for "polymerized." It is the oldest, highest temperature, and most inert fluorocarbon probe insulation. Extremely difficult to adhesive bond, it is usable up to 550°F (290°C), but on probes, its temperature limit is determined by the type of bonding to the probe rod (300, 450, or 550°F). This is the most common probe insulation in the industry. Since it never melts (it disintegrates producing HF at 600 +°F), it is difficult to fabricate, is impossible to plastic-weld, and exhibits a high degree of microporosity. Can be destroyed by butadiene and styrene monomer.
Purging, Pressurization, Ventilation	These processes depend upon the maintenance of a slight positive pressure of air or inert gas within an enclosure so that the hazardous atmosphere cannot enter. Relatively recent in general application, it is applicable to any size or type of equipment.
Quevenne Degrees	A specific gravity unit used in the expressing the fat content of milk.
Raceway	A general term for enclosed channels, conduit, and tubing designed for holding wires and cables.
Radar (Radio Detection and Ranging)	A system using beamed and reflected radio-frequency energy for detecting and locating objects, measuring distance or altitude, navigating, homing, bombing, and other purposes; in detecting and ranging, the time interval between transmission of the energy and reception of the reflected energy establishes the range of an object in the beam's path.
Radio Frequency	A frequency that is higher than sonic but less than infrared. The low end of the RF range is 20 kHz, and its high end is around 100,000 MHz.
Radio Frequency Interference (RFI)	A phenomenon where electromagnetic waves from one source interfere with the performance of another electrical device.
Range	The region in which a quantity is measured, received, or transmitted. The limits of this region are the lower and upper range-values.
Range, Cooling	For a cooling tower, it is the difference between the temperatures of water inlet and outlet. For a system operating in a steady state, the range is the same as the water temperature rise through the load heat exchanger. Accordingly, the range is determined by the heat load and water flow rate, not by the size or capability of the cooling tower.
Range, Elevated Zero	A range in which the zero value of the measured variable is greater than the lower range-value. Sometimes the term "suppressed range" is also used, but "elevated zero" is preferred.
Range, Suppressed-Zero	A range in which the zero value of the measured variable is less than the lower range value. Sometimes the term "elevated range" is also used, but "suppressed zero range" is preferred.
Range-Value, Lower	The lowest value of the measured variable less that the particular device is adjusted to measure.
Range-Value, Upper	The highest value of the measured variable that the device is adjusted to measure.
Rate Control Action	See **Control Action, Derivative**.
Rated Relieving Capacity	It is the maximum relieving capacity of the PRV. These data are normally provided on the nameplate of the PRV. The rated relieving capacity of the PRV exceeds the required relieving capacity and is the basis for sizing the vent header system.
Ratio Controller	See **Controller, Ratio**.
Ratio Pyrometer	See **Two-Color Pyrometer**.
Reactance (*X*)	That part of the impedance of a circuit that is caused by either capacitance or inductance or both. (Units: ohms.)

Rear Mount	A technique for making long inactive sections, by mounting the probe on the end of a pipe, with its coax cable running through the pipe to the top of the tank. The coax must survive the process temperature, so it is often of high-temperature construction.
Recipe	A set of procedures and formula variables that specify the production of a batch. There are four types of recipes: general, site, master, and control. A recipe is the complete set of data and operations that define the control requirements for a particular type or grade of final or intermediate product. Specifically, each recipe comprises a header, formula, procedure, and equipment requirements.
Reference Junction	The thermocouple junction that is at a known or reference temperature. It is that point at which the thermocouple extension wires are connected to the lead wires or to an instrument.
Reference Junction Compensation	The means by which the effect of temperature variations at the reference junction is corrected for.
Reflectance or Reflectivity (*R*)	The percentage of the total radiation falling on a body that is directly reflected without entry. Reflectance is zero for a blackbody, and nearly 100% for a highly polished surface. ($R = 1 - A - T$, where A is the absorbance and T is the transmissivity.)
Regulator	See **Controller, Self-Operated**.
Relative Gain (RG)	The ratio of the steady-state gain of the loop with other loops in manual, divided by the steady-state gain of the loop when the other loops are in automatic.
Relative Gain Array	A matrix of dimensionless gain ratios giving one RG value for each pairing of manipulated and controlled variables.
Relative Humidity	The ratio of the mole fraction of moisture in a gas mixture to the mole fraction of moisture in a saturated mixture at the same temperature and pressure. Or, the ratio of the amount of moisture in a gas mixture to the amount of moisture in a saturated mixture at equal volume, temperature, and pressure.
Relative Viscosity	Ratio of absolute viscosity of a fluid at any temperature to that of water at 20°C (68°F). Since water at this temperature has a μ of 1.002 cp, the relative viscosity of a fluid equals approximately its absolute viscosity in cp. Since density of water is 1, kinematic viscosity of water equals 1.002 cSt at 20°C.
Relief Valve	An automatic pressure relieving device actuated by the static pressure upstream of the valve, which opens in proportion to the increase in pressure over the operating pressure. It is used primarily for liquid service.
Relieving Pressure	The sum of opening pressure plus overpressure. It is the pressure, measured at the valve's inlet, at which the relieving capacity is determined.
Reopening Pressure	The opening pressure when the pressure is raised as soon as practicable after the valve has reseated or closed from a previous discharge.
Repeatability	For full range traverses of a number of consecutive measurements, approaching from the same direction, it is the closeness of agreement among the outputs of the sensor when measuring the same value under the same operating conditions. It is usually expressed in percent of span. A more accurate term for it would be nonrepeatability.
Reproducibility	For a number of consecutive measurements, approaching from both directions, it is the closeness of agreement among the outputs of the sensor when measuring the same value under the same operating conditions over a period of time. It is usually expressed in percent of span. A more accurate term for it would be nonreproducibility.
Reset Control Action	See **Control Action, Integral**.
Resistive Component	AC current can be separated into two components; the portion that is in phase with the excitation voltage is the resistive component.
Resistivity (*ρ*)	It is the property of a conductive material that determines how much resistance a unit cube will produce. (Units: ohm-centimeters.)
Resolution	The least interval between two adjacent discrete details that can be distinguished one from the other.
Resonance	A condition evidenced by large oscillatory amplitude, which results from a small amplitude periodic input, when the frequency of that input is approaching one of the natural frequencies of the system.
Response Time	The time it takes for the output of a device, resulting from the application of a specified input under specified

	operating conditions to move from its initial value to within some specified percentage of its final steady-state value.
Reverse Acting Controller	See **Controller, Reverse Acting**.
Richter Degrees	A specific gravity unit used in the alcohol industry.
Riser	In case of a cooling tower, it is the piping that connects the circulating water supply line from the level of the base of the tower or the supply header to the tower inlet connection.
RMS Value	See **Value, RMS**.
Roentgen	A unit for expressing the strength of a radiation field. In a 1-Roentgen radiation field, 2.08 billion pairs of ions are produced in a cubic centimeter of air.
Roentgen Equivalent Man (rem)	A unit of allowable radiation dosage, corresponding to the amount of radiation received when exposed to 1 roentgen over any period of time.
Root Valve	The first valve off the process.
Rupture Tolerance	For a rupture disc it is the tolerance range on either side of the marked or rated burst pressure within which the rupture disc is expected to burst. Rupture tolerance may also be represented as a minimum–maximum pressure range. Also referred to as performance tolerance in ISO standards.
Safety Relief Valve	An automatic pressure-actuated relieving device suitable for use as either a safety or relief valve.
Safety Valve	An automatic pressure-relieving device actuated by the static pressure upstream of the valve and characterized by rapid and full opening or pop action. It is used for steam, gas, or vapor service.
Sand Filling	All potential sources of ignition are buried in a granular solid, such as sand. The sand acts, in part, to keep the hazardous atmosphere away from the sources of ignition and, in part, as an arc quencher and flame arrester. It is used in Europe for heavy equipment. It is not used in instruments.
Saturated Solution	A solution that has reached the limit of solubility.
Saturation	A condition where RF current from a probe-to-ground is determined solely by the impedance of the probe insulation. Increased conductivity in the saturating medium, even to infinity, will not cause a noticeable change in that current or in the transmitter output.
Saturation Pressure	The pressure of a fluid when condensation (or vaporization) takes place at a given temperature. (The temperature is the saturation temperature.)
Saybolt Furol Seconds (SFS)	Time units referring to the Saybolt viscometer with a Furol capillary, which is larger than a universal capillary.
Saybolt Universal Seconds (SUS)	Time units referring to the Saybolt viscometer.
Scale Factor	The value of the scale divisions on an instrument. To compute the value of the measured variable, the number of scale divisions is multiplied by this scale factor.
Scaling	The conversion from engineering units to fractions or percentages.
Sealing	The atmosphere is excluded from potential sources of ignition by sealing them in airtight containers. This method is used for components such as relays, not for complete instruments.
Seal-Off Pressure	The pressure, measured at the valve inlet after closing, at which no further liquid, steam, or gas is detected at the downstream side of the seat.
Segment	The section of a network that is terminated in its characteristic impedance. Segments are linked by repeaters to form a complete network.
Self-Regulation	The property of a system, which permits attainment of equilibrium after a disturbance without the intervention of a controller.
Sensitivity	The ratio of the change in output to the change of the input that causes it after a steady state been reached.
Sensor	An input device that provides a usable output in response to the input measurand. (A sensor is also commonly called a sensing element, primary sensor, or primary detector. The measurand is the physical parameter to be measured.)
Service	Term used by NFPA-70 (NEC) to demarcate the point at which utility electrical codes published by IEEE (NESC) take over. Includes conductors and equipment that deliver electricity from utilities.
Servomechanism	An automatic feedback control device in which the controlled variable is a mechanical position or some derivative of that position.
Set Point	An input variable of a controller that sets the desired value of the variable that is being controlled.

Set Pressure (Opening Pressure)	The pressure at which the relief valve is set to open. It is the pressure measured at the valve inlet of the PRV, at which there is a measurable lift, or at which discharge becomes continuous as determined by seeing, feeling, or hearing. In the pop-type safety valve, it is the pressure at which the valve moves more in the opening direction compared to corresponding movements at higher or lower pressures. A safety valve or a safety relief valve is not considered to be open when it is simmering at a pressure just below the popping point, even though the simmering may be audible.
Shear Viscometer	Viscometer that measures viscosity of a non-Newtonian fluid at several different shear rates. Viscosity is extrapolated to zero shear rate by connecting the measured points and extending the curve to zero shear rate.
Signal, Analog	A signal representing a continuously observed variable.
Signal, Digital	Information represented by a set of discrete values in accordance with a prescribed law.
Signal-to-Noise Ratio	The ratio of the amplitude of a signal to the amplitude of the noise. The amplitude can be a peak or an rms value, as specified.
Signum	A transfer function used in some back-propagation neural networks.
Sikes Degree	A specific gravity unit used in the alcohol industry.
Simmer (Warn)	The condition just prior to opening at which a spring-loaded relief valve is at the point of having zero or negative forces holding the valve closed. Under these conditions, as soon as the valve disc attempts to rise, the spring constant develops enough force to close the valve again.
Single-Effect Evaporation	Single-effect evaporation occurs when a dilute solution is contacted only once with a heat source to produce a concentrated solution and an essentially pure water vapor discharge.
Site Recipe	A recipe that includes site-specific information, such as local language and locally available raw materials.
Slab	A set of nodes.
Smart Field Device	A microprocessor-based process transmitter or actuator that supports two-way communications with a host; digitizes the transducer signals; and digitally corrects its process variable values to improve system performance. The value of a smart field device lies in the quality of data it provides.
Span	The difference between the upper and lower range-values.
Specific Humidity	The ratio of the mass of water vapor to the mass of dry gas in a given volume.
Specific Viscosity	Ratio of absolute viscosity of a fluid to that of a standard fluid, usually water, both at the same temperature.
Spectral Emissivity	The ratio of emittance at a specific wavelength or very narrow band to that of a blackbody at the same temperature.
Split Ranging	A configuration in which, from a single input signal, two or more signals are generated or two or more final control elements are actuated, each responding consecutively, with or without overlap, to the magnitude of the input signal.
Standard Air	Dry air having a density of 0.075 lb/cu. ft. at 70°F and 29.92 in. Hg.
Start-to-Leak Pressure	For a safety relief valve it is the pressure at the valve inlet at which the relieved fluid is first detected on the downstream side of the seat before normal relieving action takes place.
Steady State	A variable is at steady-state condition when it is exhibiting negligible change over a long period of time.
Stiction (Static Friction)	The resistance to the starting of motion. When stroking a control valve, it is the combination of sticking and slipping.
Stiffness	In the case of a spring element, it is the ratio of the change in force or torque to the resulting change in deflection.
Stoke	Unit of kinematic viscosity υ (cm^2/sec).
Stress	Force/Area (F/A).
Subchannel	In broadband terminology, a frequency-based subdivision creating a separate communications channel.
Subsidence Ratio	The ratio of the peak amplitudes of two successive oscillations of the same sign, measured in reference to an ultimate steady-state value.
Superimposed Backpressure	Backpressure that is present in the discharge header before the pressure relief valve starts to open. It can be constant or variable, depending on the status of the other PRVs in the system.
Suppression	See **Range, Elevated-Zero**.
Suppression, Zero	See **Zero Suppression**.
Surge Pressure	See **Pressure, Surge**.
Switched Hub	Multiport bridge joining networks into a larger network.
Systematic Error	See **Error, Systematic**.

System, Linear	A system is linear if its time response to several simultaneous inputs is the sum of their individual (independent) time responses.
Tapping	See **Dither**.
Teflon, TFE, FEP, and PFA	Most people interchange the name Teflon with TFE. This is completely incorrect, but understandable. TFE was the first fluorocarbon polymer to carry the trade name “Teflon” at E.I. DuPont. DuPont chose to use the “Teflon” trade name for a whole family of fluorocarbon resins, so FEP and PFA made by DuPont are also Teflon. To complicate the matter, other companies now manufacture TFE, FEP, and PFA, which legally cannot be called Teflon, since that name applies only to DuPont-made polymers.
Thermopile	Measures thermal radiation by absorption to become hotter than its surroundings. It is a number of small thermocouples arranged like the spokes of a wheel, with the hot junction at the hub. The thermocouples are connected in series and the output is based on the difference between the hot and cold junctions.
Throughput	The maximum number of transactions per second that can be communicated by the system.
Time Constant (*T*)	If a first-order system is responding to a step or an impulse, T is the time required to complete 63.2% of the total rise or decay. In higher-order systems, there is a time constant for each of the first-order components of the process.
Time, Dead	The time interval between the initiation of an output change or stimulus and the start of the resulting observable response.
Time Domain Reflectometry (TDR)	A TDR instrument measures the electrical characteristics of wideband transmission systems, subassemblies, components, and lines by feeding in a voltage step and displaying the superimposed reflected signals on an oscilloscope equipped with a suitable time-base sweep.
Timeout	Event that occurs when one network device expects to hear from another network device within a specified period of time, but does not. The resulting timeout usually results in a retransmission of information or the dissolving of the session between the two devices.
Time, Ramp Response	The time interval by which an output lags an input, when both are varying at a constant rate.
Time, Settling	After a stimulus to a system, the time required for the output of the system to enter and remain within a specified narrow band centered on its steady-state value. If the stimulus is a step or impulse, the band is often specified as ±2%.
Topology	Physical arrangement of network nodes and media within an enterprise networking structure or the surface features of an object or “how it looks,” its texture; there is a direct relation between these features and materials properties (hardness, reflectivity, etc.).
Total Emissivity	The ratio of the integrated value of all spectral emittances to that of a blackbody.
Train	A grouping within one process cell of units and associated lower-level equipment that is capable of making a batch of material. A train may define a single equipment path for a batch or multiple possibilities, of which one will be selected based on availability during execution of the control recipe. Multiple batches can be processed simultaneously in the same train (but different units).
Tralles Degrees	A specific gravity unit used in the alcohol industry.
Transducer	A device that receives information in the form of one quantity and converts it to information in the form of the same or another quantity. This general definition also applies to primary elements or transmitters. An input transducer produces an electrical output, which is representative of the input measurand. Its output is conditioned and ready for use by the receiving electronics. (The terms “input transducer” and “transducer” can be used interchangeably.)
Transfer Function	A statement of influence of an element or system, in a mathematical, tabular, or graphical form. This influence can be that of an element or system on a signal or action, which is compared at input and output terminals.
Transient	The behavior of a variable during transition between two steady states.
Transient Overshoot	It is the maximum overshoot beyond the final steady-state value of an output, which results from a change in an input.
Transistor	Three-terminal, solid state electronic device made of silicone, gallium-arsenide or germanium and used for

	amplification and switching in integrated circuits.
Transmittance or Transmissivity (*T*)	The percentage of the total radiant energy falling on a body that passes directly through it without being absorbed. Transmittance is zero for a blackbody and nearly 100 percent for a material like glass in the visible spectrum region. ($T = 1 - A - R$, where A is the absorbance and R is the reflectance.)
Transmitter	A transducer that responds to a measured variable generated by a sensor and converts it to a standardized transmission signal. This signal is a function of only the measured variable. The term "transmitter," as commonly used with industrial process control instrumentation, has a narrower definition than those of a sensor or transducer: A transmitter is a transducer that responds to a measured variable by means of a sensing element and converts it to a standardized transmission signal that is a function only of the measured variable.
Twaddell Degree	A specific gravity unit used in the sugar, tanning, and acid industries.
Two-Color Pyrometer	Measures temperature as a function of the radiation ratio emitted around two narrow wavelength bands. Also called ratio pyrometer.
Unit	A major piece of process equipment with its associated equipment modules. Mixers, storage tanks, and reactors are examples of units. The associated equipment modules include pumps, valves, heat exchangers, and agitators that are closely associated with the major process equipment. Units operate relatively independently of one another. They are equipment that contains and performs some major processing activity or activities (e.g., react, crystallize, make solution) on one batch at a time. A unit normally comprises a major piece of equipment and directly associated control modules and/or equipment modules that are not shared with other units.
Unit Procedure	A major programmed processing action or set of related actions, normally consisting of one or more operations. Unit procedures are naturally related to a distinct regime of production: for example, all processing carried out in one batch unit of a multiunit production line.
Unit Recipe	A part of a recipe that defines a part of batch production requirements within a unit. It usually includes a number of operations and phases.
Upper Explosive Limit (UEL)	The highest concentration of gas or vapor in air in which a flame will continue to burn after the source of ignition has been removed.
Upper Range-Limit	The upper limit of the value of the measured variable that a particular instrument is capable of measuring.
Varactor	Voltage-sensitive capacitor.
Variable Backpressure	Backpressure that varies due to changes in operation of one or more pressure relief valves connected into a common discharge header.
Variable, Controlled	See **Controlled Variable**.
Variable, Manipulated	See **Manipulated Variable**.
Velocity Gradient (Shear)	Rate of change of liquid velocity across the stream—V/L for linear velocity profile, dV/dL for nonlinear velocity profile. Units are $V-L$ = ft/sec/ft = sec^{-1}.
Velocity Head	The velocity head is calculated as $v^2/2g$, where v is the flowing velocity and g is the gravitational acceleration (9.819 m/s^2 or 32.215 ft/s^2 at 60 degrees latitude).
Velocity Limit	The limit that the rate of change of a particular variable may not exceed.
Virtual Field Device (VFD)	It is used to remotely view local device data described in an object dictionary. A typical device will have at least two VFDs.
Voltage, Common Mode	See **Common Mode Voltage**.
Voltage, Normal Mode	See **Normal Mode Voltage**.
Water Loading	In case of cooling towers it is the water flow divided by effective horizontal wetted area of the tower. (Unit: GPM/ft^2 or m^3/hr m^2.)
Wave-Guide	A device that constrains or guides the propagation of electromagnetic waves along a path defined by the physical construction of the wave-guide; includes ducts, a pair of parallel wires, and a coaxial cable.
Wet-Bulb Temperature (WBT)	If a thermometer bulb is covered by a wet, water-absorbing substance and is exposed to air, evaporation will cool the bulb to the wet-bulb temperature of the surrounding air. This is the temperature read by a psychrometer. If the air is saturated with water, the wet-bulb, dry-bulb, and dew-point temperatures will all be

	the same. Otherwise, the wet-bulb temperature is higher than the dew-point temperature but lower than the dry-bulb temperature.
White Box Modeling	This type of modeling is feasible if a good understanding of the process exists. In such cases, the dynamic models are derived based on mass, energy, and momentum balances of the process.
Wide Band (Total) Pyrometer	A radiation thermometer that measures the total power density emitted by the material of interest over a wide range of wavelengths.
Wobbe Index	AGA 4A defines the Wobbe Index as a numerical value, which is calculated by dividing the square root of the relative density (a key flow orifice parameter) into the heat content (or BTU per std. cubic foot) of the gas. Mathematically, the Wobbe Index is defined by the equation below: $\text{Wobbe Index} = \dfrac{\text{calorific value}}{\sqrt{\text{specific gravity}}}$
(E)Xtensible Markup Language (XML)	A computer authoring language for publishing documents through the World Wide Web on the Internet. For use in automation, it is better and more flexible than hypertext markup language (HTML).
Zero Elevation	When the zero of a range is elevated, the amount of its elevation is the quantity by which the zero of the measured variable exceeds the lower range-value.
Zero Suppression	When the zero of a range is suppressed, the amount of the suppression is the quantity by which the zero of the measured variable is below the lower range-value.
Zone, Dead	A range of input through which the output remains unchanged. This holds true if the input signal is rising or dropping.
Zone, Neutral	For two-position controllers (switches), the neutral intermediate zone is the range of input values in which the previously existing output value is maintained. If the output is *A* at low inputs and *B* at high ones, on a rising input, *A* is maintained until the input reaches the value corresponding to the set point of switch *A* plus this zone, while, when the input is dropping, the switch will change its output from *B* to *A* when the input has passed through the dead zone and reached the set point of switch *A*.

SOCIETIES AND ORGANIZATIONS

AATCC	American Association of Textile Chemists and Colorists
ACC	American Chemistry Council
ACGIH	American Conference of Governmental Industrial Hygenists
ACS	American Chemical Society
AGA	American Gas Association
AIA	Automatic Imaging Association
AIChE	American Institute of Chemical Engineers
AMTEX	American Textile Partnership
ANSI	American National Standards Institute
AOCS	American Oil Chemists’ Society
APHA	American Public Health Association
API	American Petroleum Institute
ARI	Air Conditioning and Refrigeration Institute
ASA	American Standards Association
ASCE	American Society of Civil Engineers
ASM	Abnormal Situation Management Consortium
ASME	American Society of Mechanical Engineers
ASRE	American Society of Refrigeration Engineers
ASTM	American Society for Testing and Materials or ASTM International
Awwa	American Water Works Association
BSI	British Standards Institution
CARB	California Air Resources Board
CCITT	Consultative Committee for International Telegraphy and Telephony
CCSDS	Consultative Committee for Space Data Systems
CDC	Centers for Disease Control (United States)
CENELEC	European Committee for Electrotechnical Standardization
CIE	Commission International del’Eclairage
CII	Construction Industry Institute
CIL	Canadian Industries Limited
CNI	ControlNet International
CPAC	Center for Process Analytical Chemistry
CSA	Canadian Standards Association
DARPA	Defense Advanced Research Projects Agency
DIERS	Design Institute for Emergency Relief Systems
DIN	Deutsche Institut für Normung (German Standards Institute)
DOD	Department of Defense (United States)
DOE	Department of Energy
DOT	Department of Transportation
EBF	European Batch Forum
ECMA	European Computer Manufacturers Association
EEMUA	Engineering Equipment and Materials Users Association
EIA	Electronic Industries Association
EIA/TIA	Electrical Industries Alliance/Telecommunications Industries Alliance
EPA	Environmental Protection Agency
EPRI	Electric Power Research Institute
EXERA	The Instrument Users’ Association in France
FCC	Federal Communications Commission
FCI	Fluid Control Institute
FDA	Food and Drug Administration
FF	Fieldbus Foundation
FIA	Fire Insurance Association
FM	Factory Mutual
FMRC	Factory Mutual Research Corporation
FPA	Fire Protection Association
FSEC	Florida Solar Energy Center
GERG	Groupe Européen de Rechérches Gaziéres (European Gas Research Group)
GRI	Gas Research Institute
HCF	HART Communication Foundation

IAEI	International Association of Electrical Inspectors
IBEW	International Brotherhood of Electrical Workers
ICE	Institute of Civil Engineers
ICEA	Insulated Cable Engineers Association
ICTS	International Consortium of Telemetry Spectrum
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
IETF	Internet Engineering Task Force
IGT	Institute of Gas Technology
INPO	Institute for Nuclear Power Operation
IPTS	International Practical Temperature Scale
IrDA or IRDA	Infrared Data Association
ISA	Instrumentation, Systems, and Automation Society
ISO	International Standards Organization
ISSEP	International Soros Science Education Program
ISTM	International Society for Testing Materials
ITA	Instrumentation Testing Association
JBF	Japanese Batch Forum
JPL	Jet Propulsion Laboratory
KEPRI	Korean Electric Power Research Institute
LCIE	Laboratoire Central des Industries Electriques
LPGA	National LP-Gas Association
MCA	Manufacturing Chemists Association
NAMUR	German standardization association for process control (Normen-Arbeitsgemeinschaft für Meß- und Regeltechnik in der Chemischen Industrie)
NASA	National Aeronautics and Space Administration
NBFU	National Board of Fire Underwriters
NBS	National Bureau of Standards
NEMA	National Electrical (Equipment) Manufacturers Association
NEPSI	National Supervision and Inspection Center for Explosion Protection and Safety Instrumentation
NFPA	National Fire Protection Association
NIOSH	National Institute of Occupational Safety and Health
NIST	National Institute of Standards and Technology
NSC	National Safety Council
NSPA	National Spa and Pool Association
NSPE	National Society of Professional Engineers
NRC	Nuclear Regulatory Commission
ODVA	Open DeviceNet Vendor Association
OSHA	Occupational Safety and Health Administration
OTS	Office of Technical Services
PCT	Patent Cooperation Treaty
PNO	Profibus User Organization
SAE	Society of Automotive Engineers
SAMA	Scientific Apparatus Manufacturers Association
SIREP	The Instrument Users' Association in the United Kingdom
TAPPI	Technical Association of the Pulp and Paper Industry
TIA	Telecommunications Industries Alliance
TUV	Technischer überwachungs Verein (Technical Inspection Association)
UA	United Association of Journeyman and Apprentices of the Plumbing and Pipe Fitting Industry of the United States and Canada
UL	Underwriters Laboratories, Inc.
USASI	USA Standard Institute
USNRC	U.S. Nuclear Regulatory Commission
VDMA	Verband Deutscher Maschinen und Anlagenbau e.V.
WBF	World Batch Forum
WEF	Water Environment Federation
WIB	The International Instrument Users' Association
WIDO	World Intellectual Property Office

ABBREVIATIONS, NOMENCLATURE, ACRONYMS, AND SYMBOLS

NOTES

1. Whenever the abbreviated form of a unit might lead to confusion, it should not be used and the name should be written out in full.
2. The values of SI equivalents were rounded to three decimal places.
3. The words meter and liter are used in their accepted English spelling form instead of those in the standards, namely, metre and litre, respectively.

1oo1	one out of one
1oo2	one out of two
1oo2D	one out of two with diagnostics
2oo2	two out of two
2oo3	two out of three
2oo3d	two out of three with diagnostics
2D	two-dimensional
3D	three-dimensional

A

a	acceleration
A	1) area; 2) ampere, symbol for basic SI unit of electric current; also admittance
Å	Ångstrom (= 10^{-10} m)
AA	atomic absorption
AAS	atomic absorption spectrometer
abs	absolute (e.g., value)
ABS	acrylonitrile-butadiene-styrene
AC, ac, a-c	alternating current
A/C	air to close
ACFM	volumetric flow at actual conditions in cubic feet per minute (= 28.32 alpm)
ACL	asynchronous connection-less
ACM	automatic control mode
ACMH	actual cubic meter per hour
ACMM	actual cubic meter per minute
ACS	analyzer control system
ACSL	advanced continuous simulation language
A/D	analog-to-digital, also analog-to-digital converter
AD	actuation depth
ADC	analog-to-digital converter
ADIS	approved for draft international standard circulation
ADPCM	adaptive differential pulse-code modulation
AE	analyzer element
A&E	alarm and event
AES	atomic emission spectrometer
AF or a-f	audio frequency
AFC	alkaline fuel cell
AFD	adjustable frequency drive
AGA3	American Gas Association Report No. 3
AGC	automatic generation control or automatic gap control
ai	adobe illustrator
AI	analog input or artificial intelligence
AI-AT	analog input–air temperature
AI-RT	analog input–return temperature
a(k)	white noise
ALARA	as low as reasonably achievable
ALARP	as low as reasonably practicable
AliS	alternate lighting of surfaces
ALP	low pressure air
Alpm	actual liters per minute
ALSW	admissible load supply well
alt	altitude
AM	amplitude modulation or actual measurement or alarm management
AMC	annual maintenance contract or adaptive model controller or auto-manual cascade
AMLCD	active matrix LCD
amp	ampere; also A, *q.v.*
AMPS	advanced mobile phone system
AMS	asset management solutions or analyzer maintenance solutions
a/n	alpha numeric
ANN	artificial neural network
ANS	adaptive neural swarming

AO	analog output
A/O	air to open
AOTF	acousto-optical tunable filters
AP	access point
APC	automatic process control or advanced process control
APDU	application (layer) protocol data unit
API	application programming interface or absolute performance index
°API	API degrees of liquid density
APM	application pulse modulation
APSL	air pressure switch, low
AR	auto regressive
ARA	alarm response analysis
ARIMA	auto regressive integrated moving average
ARP	address resolution protocol
ARX	auto regressive with external inputs (model)
ARW	antireset windup
AS	adjustable speed
ASCI	autosequentially commutated current-fed inverter
ASCII	American Standard Code for Information Interchange
AS-I	actuator sensor interface
ASIC	application-specific integrated chips
ASK	amplitude shift keying
ASU	air separation unit
asym	asymmetrical; not symmetrical
AT	air temperature or analyzer transmitter
ATG	automatic tank gauging
atm	atmosphere (= 14.7 psi)
ATP	adenosine triphosphate
ATR	attenuated total reflectance
AUI	attachment unit interface
AUTRAN	automatic utility translator
aux	auxiliary
AV	auxiliary (constraint) variable
AVC	air velocity controller
AVR	automatic voltage regulator
AWG	American wire gauge

B

B	bottom product flow rate
B2B	business-to-business
°Ba	Balling degrees of liquid density
BAC	biologically activated carbon
bar	1) barometer; 2) unit of atmospheric pressure measurement (= 100 kPa)
bara	bar absolute
barg	bar gauge
bbl	barrels (= 0.1589 m^3)
BCD	binary coded decimal
BCM	backup communication module
BCS	batch control system
BD	blow down
°Bé	Baumé degrees of liquid density
BEP	best efficiency point
BFO	beat frequency oscillator
BFW	boiler feed water
Bhp, BHP	brak horsepower (= 746 W)
BIBO	bounded input, bounded output
°Bk	Barkometer degrees of liquid density
blk	black (wiring code color for AC "hot" conductor)
BMS	burner management system or boiler management system
BO	basic operation
BOD	biochemical oxygen demand
bp or b.p.	boiling point
BPCS	basic process control system
BPS or bps	bits per second
BPSK	binary phase shift keying
Bq	becquerel, symbol for derived SI unit of radioactivity, joules per kilogram, J/kg
°Br	Brix degrees of liquid density
BSL	best straight line
BSTR	batch stirred tank reactor
BTR	block transfer read
BTU	British thermal unit (= 1054 J)
BWD	backwash drain
BWG	Birmingham wire gauge
BWR	backwash return
BWS	backwash supply

C

c	1) velocity of light in vacuum (3×10^8 m/s); 2) centi, prefix meaning 0.01
C	coulombs or symbol for discharge coefficient, also capacitance
°C	Celsius degrees of temperature
ca.	*circa*: about, approximately
CAC	channel access code
CAD	computer-aided design
Cal	calorie (gram, = 4.184 J); also g-cal
CAMAC	control advance moving average
CAN	control area network or control and automation network
CAPEX	CAPital EXpenditure
CARI	combustion air requirement index
CAS	cascade
CATV	community antenna television (cable)
CBM	condition-based maintenance
CBT	computer-based timing
cc	cubic centimeter (= 10^{-6} m^3)
CC	cooling coil
CCD	charge-coupled device
CCF	common cause failure or combination capacity factor
ccm	cubic centimeter per minute
CCR	central control room
Ccs	constant current source

CCS	computer control system or constant current source
CCTV	closed circuit television
CCW	counterclockwise
cd	candela, symbol for basic SI unit of luminous intensity
CD	compact disk, collision detector, compel data or dangerous coverage factor, cold deck
CDDP	cellular digital packet data
CDF	cumulative distribution function
CDMA	code division multiple access
CDPD	cellular digital packet data
CDT	color detection tube
CDTP	cold differential test pressure
CE	Conformité Europeéne (European Conformity) applicable to electrical safety
CEHED	chlorination–caustic extraction–hypochlorite bleaching–caustic extraction–chlorine dioxide
CEMS	continuous emissions monitoring system
CENP	combustion engineering nuclear power
CF	cleanliness factor or cubic foot
CFA	continuous flow analyzer
CFE	cartridge filter effluent
CEM	cause and effect matrix
CFM or cfm	cubic foot per minute (28.32 lpm)
CFR	Code of Federal Regulations
CF/Yr	cubic foot per year
CHP	combined heat and power
CHS	chemical sludge
CHWP	chilled water pump
CHWR	chilled water return
CHWS	chilled water supply
Ci	curie (= 3.7 × 10^{10} Bq)
CI	cast iron or corrosion inhibitor
CIM	computer-integrated manufacturing
CIO	channel input–output
CIP	1) computer-aided production; 2) control and information protocol (an application layer protocol supported by DeviceNet, ControlNet, and Ethernet/IP); or 3) clean in place
CJ	cold junction
CL	clamp on
CL1	electrically hazardous, Class 1, Division 1, Groups C or D
CLD	chemiluminescence detector
CLOS	common Lisp object system
CLP	closed-loop potential factor
cm	centimeter (= 0.01 m) or cubic meter
CM	condition monitoring or communication (interface) module or control module
CMF	Coriolis mass flowmeter
CMMS	computerized maintenance management system
CMOS	complementary metal oxide semiconductor
CMPC	constrained multivariable predictive control
cmph	cubic meter per hour
CMR	common mode rejection
CMRR	common mode rejection ratio
CMS	carbon molecular sieve
CMV	common-mode voltage
CNC	computerized numerical control
CNI	ControlNet International
Co	cobalt
CO	controller output or carbon monoxide or contact output
CO_2	carbon dioxide
CO_2D	carbon dioxide demand
COD	chemical oxygen demand
COF	coefficient of haze
COM	component (or compiled) object model
COND	conductivity
COP	coefficient of performance
cos	cosine, trigonometric function
COT	coil outlet temperature
COTS	commercial off-the-shelf
COV	coil outlet velocity
cp or c.p.	1) candle power; 2) circular pitch; 3) center of pressure (cp and ctp may also be used for centipoises)
cpm	cycles per minute; counts per minute
cps	1) cycles per second (= Hz); 2) counts per second; 3) centipoises (= 0.001 Pa.S)
CPS	computerized procedure system
CPU	central processing unit
CPVC	chlorinated polyvinyl chloride
CR	corrosion rate
CRC	cyclical redundancy check or cyclic redundancy code (an error detection coding technique based upon modulo-2 division; sometimes misused to refer to a block check sequence type of error detection coding)
CRDS	cavity ring-down spectroscopy
CRH	cold reheat
CRLF	carriage return-line feed
CRT	cathode ray tube
Cs	cesium
CS	1) carbon steel; 2) constant speed; 3) chlorine solution
CSD	crystal size distribution
CSH	constant speed held
CSL	car seal lock
CSMA/CD	carrier sense, multiple (medium) access with collision detection
CSO	car seal open
CSS	central supervisory station
CSSD	compatible single side band
cSt	centi stoke
CSTR	continuous-stirred tank reactor
CT	cooling tower or the product of *C* for disinfectant concentration and *T* for time of contact in minutes

CTDMA	concurrent time domain multiple access
CTMP	chemi-thermo-mechanical pulp
CTWP	cooling tower water pump
CV	controlled variable or control valve
CVAAS	cold vapor atomic absorption spectroscopy
CVF	circular variable filters
cvs	comma-separated variables
CW	clockwise
CWA	Clean Water Act
CWR	cooling water return
CWS	cooling water supply
CWT	centralized waste treatment

D

d	1) derivative; 2) differential, as in dx/dt; 3) deci, prefix meaning 0.1; 4) depth; 5) day
D	diameter; also dia and ϕ or derivative time of a controller or distillate flow rate
DA	data access or direct action or difference in aeration
D/A	digital-to-analog
DAC	device access code or digital-to-analog converter
DACU	data acquisition and control unit
DAE	differential algebraic equation
DAMPS	digital advanced mobile phone system
DAS	data acquisition system
DB or dB	decibels
dBa	decibels with "*A*" weighing to approximate the human ear
DBB	double-block and bleed
DBPSK	differential binary phase shift keying
DC or dc	direct current
DC	diagnostic coverage
DCAC	direct contact after cooler
DCE	data communications equipment
DCOM	distributed COM
DCS	distributed control system
DD	data definition or device description
D/DBP	disinfectants/disinfection byproducts
DDC	direct digital control
DDE	dynamic data exchange
DDL	device description language (an object-oriented data modeling language currently supported by PROFIBUS, FF, and HART)
DEDED	chlorine dioxide treatment–caustic extraction–chlorine dioxide treatment–caustic extraction–chlorine dioxide treatment
deg	degree; also ° ($\pi/180$ rad)
DEMUX	demultiplexer
Deoxo	deoxidation unit
DES	data encryption standard
DF	direction of flow
DFIR	diffused infrared
DFR	digital fiber-optic refractometer
DFT	digital (or discrete) Fourier transforms
DG	directed graph
DH	data highway
DH+	data highway plus (high-speed peer-to-peer link)
DHCP	dynamic host configuration protocol
DI	discrete (digital) input
dia	diameter; also D and ϕ
DIAC	dedicated inquiry access code
DIR	diffused infrared
DIS	draft international standard
DIX	Digital-Intel-Xerox (DIX is the original specification that created the de facto Ethernet standard; IEEE 802.3 came later, after Ethernet was established)
d(k)	unmeasured disturbance
D(k)	measured disturbance
DLE	data link escape
DLL	dynamic link library
DLPD	digital light processor display
Dm or dm	decimeter
DM	delta modulation
DMA	dynamic mechanical analyzer or direct memory access
DMC	dynamic matrix control(ler)
DMFC	direct methanol conversion fuel cell
DMM	digital multi-meter
DN	diameter nominal, the internal diameter of a pipe in rounded millimeters
DO	dissolved oxygen or discrete (digital) output
DOAS	differential optical absorption spectroscopy
d/p cell	differential pressure transmitter (a Foxboro trademark)
DP	decentralized peripheral
DPC	damper position controller
DPCM	differential pulse code modulation
DPD	*N*,*N*-Diethyl-*p*-phenylenediamine
DPDT	double pole double throw (switch)
dpi	dots per inch
DPS	differential pressure switch
DQPSK	differential quadrature phase shift keying
DR	decay ratio
DSB	double side band
DSL	digital subscriber line
DSN	distributed sensor network
DSP	digital signal processing
DSR	direct screen reference
DSSS	direct sequence spread spectrum
DST	dirty service trim
DSTM	dual-scan twisted nematic
DT or dt	dead time (second or minutes) or delay time
DTB	draft tube baffle
DTC	digital temperature compensation or dead time compensator
DTD	document type definition
DTE	data terminal equipment
DTGS	deuterated tryglycine sulfate

DTL	diode-transistor logic
DTM	device type manager (an active-X component for configuring an industrial network component; a DTM “plugs into” an FDT)
DU	dangerous component failure occurred in leg but is undetected
DV	disturbance variable
DVC	digital valve control or discrete valve coupler
DVM	digital voltmeter
DWS	dewatered sludge

E

e	1) error; 2) base of natural (Naperian) logarithm; 3) exponential function; also exp $(-x)$ as in e^{-x}
E	1) electric potential in volts; 2) scientific notation as in $1.5E-03 = 1.5 \times 10^{-3}$; 3) tray efficiency in distillation
E{.}	expected value operator
E&I	electrical and instrumentation
EA	evolutionary algorithm or exhaust air
EAD	exhaust air damper
EAF	exhaust air fan
EAI	enterprise application integration
EAM	enterprise asset management
EAPROM	erasable alterable programmable read-only memory
EBCDIC	extended binary code for information interchange
EBR	electronic batch records
ECD	electron capture detector
ECG	electrocardiogram
ECKO	eddy-current killed oscillator
ECLiPS	English control language programming software
ECN	effective carbon number
ECTFE	ethylene chloro-tetra-fluoro-ethylene (Halar)
ED	explosive decompression
EDD	electronic device description
EDS	electronic data sheet (DeviceNet)
EDTA	ethylenediaminetetraacetic acid
EDXRF	energy dispersive x-ray fluorescence
E/E/PE	electrical/electronic/programmable electronic
E/E/PES	electrical/electronic/programmable electronic system
EEPROM	electrically erasable programmable read only memory
EFB	external feedback
EFD	engineering flow diagram
EFRVD	expected flow rate value through dryer
e.g.	*exempli gratia*: for example
EHC	electro-hydraulic control or extended horizon adaptive controller
EHM	equipment health management
e(k)	feedback error
E.L.	elastic limit or enthalpy logic
ELD	electroluminescent display
Em	minimum error
EM	equipment module
Emf or EMF	1) electromotive force (volts); 2) electromotive potential (volts)
EMI	electro-magnetic interference
EMI/RFI	electromagnetic and radio frequency interference
em(k)	process/model error
EN	European standard
ENB	Ethernet card
EP	evolutionary programming or equipment phase
EPA	enhanced performance architecture
EPC	engineering-procurement-construction (firm or industry)
EPCM	engineering, procurement, and construction management (companies)
EPDM	ethylene propylene diene terpolymer
EPROM	erasable programmable read only memory
EPS	electronic pressure scanner, encapsulated postscript file, emergency power supply, or expanded polystyrene
EPSAC	extended prediction self-adaptive controller
EQ or eq	equation
ER	external reset
ERM	enterprise resource manufacturing
ERP	enterprise resource planning or effective radiated power
ERW	electric-resistance-welded
ES	evolutionary strategy
ESD	emergency shutdown (system)
ESN	electronic serial number
ESP	environmental simulation program
ETFE	ethylene-tetrafluoroethylene copolymer (Tefzel)
ETM	elapsed time meter
ETMI	expected total mass input
ETMP	expected total mass produced
ETS	equivalent time sampling
EU	engineering unit
EVOP	evolutionary operation or evolutionary optimum
EWMA	exponentially weighed moving average
Exp	exponential function as in exp $(-at) = e^{-at}$; also *e*

F

F	frequency; also freq or filter, also farad, symbol for derived SI unit of capacitance, ampere · second per volt, A · s/V, also feed flow rate
°F	Fahrenheit degrees [$t°C = (t°F - 32)/1.8$]
FAH	flow alarm, high

FAL	flow alarm, low
FAN	fuzzy, analytical intelligence, neural network
FARM	frame acceptance and reporting mechanism
FAT	factory acceptance testing
FB	friction bound
FBAP	function block application process (FF)
FBD	function block diagram
FBG	fiber bragg grating
FBT	fire box temperature
FC	1) flow controller; 2) fail closed; 3) forced circulation; 4) fire command panel
FCC	fluid catalytic cracking unit
FCCU	fluidized catalytic cracking unit
FCOR	filtering and correlation (method)
FCS	frame check sequence or field control system
FCV	flow control valve
FD	forced draft
FDE	fault disconnection electronics
FDL	fieldbus data link
FDM	frequency division multiplexing
FDMA	frequency division multiple access
FDS	flame detection system or fines density sensor
FDT	field device tool (a MS Windows-based framework for engineering and for configuration tools)
FE	1) final element; 2) flow element; 3) filter effluent
FED	field emission display
FEED	front end engineering and design
FEGT	furnace exit gas temperature
FEP	fluorinated ethylene propylene
FES	fixed end system
FF	1) Foundation Fieldbus; 2) feedforward
FF-HSE	Foundation Fieldbus, high-speed Ethernet
FFIC	flow ratio controller
FFT	fast Fourier transform
FGB	fiber bragg grating
FGD	flue gas desulfurization
FGR	flue gas recirculation
FH	frequency hopping
Fhp	fractional horsepower (e.g., $^{1}/_{4}$ HP motor)
FHSS	frequency hopped spread spectrum
FI	1) flow indicator; 2) fail intermediate; 3) filter influent
FIA	flow injection analyzer
FIC	flow indicator controller
FID	flame ionization detector
FIE	flame ionization element
FIFO	first-in, first-out
Fig. or FIG.	figure
FIR	finite impulse response (model)
FISCO	Fieldbus Intrinsic Safety COncept
fl.	fluid
FL	fail last (position)
fl. oz.	fluid ounces (= 29.57 cc)

FM	frequency modulated or Factory Mutual
FMAX	maximum feed availability
FMCW	frequency-modulated carrier wave
FMEA	failure mode and effects analysis
FMEDA	failure modes, effects, and diagnostic analysis
FMS	1) fieldbus message specification; 2) fieldbus messaging services/system
FNC	function byte
FO	1) fiber optic; 2) fail open
FOP	fiber optic probe
FOPDT	first order plus dead time
FOSS	first order system series
FOV	field of view
fp or f.p.	freezing point
FPC	fine particle content
FPD	flame photometric detector
FPGA	field programmable gate arrays
FPM or fpm	feet per minute (= 0.3048 m/m)
fps or ft/s	feet per second (= 0.3048 m/s)
FRC	flow recording controller
FRM	frequency response method
FRO	full range output
FS or fs	full scale or flow switch or full stroke
FSA	fault state active
FSC	fail safe controller
FSH	flow switch, high
FSHL	flow switch, high-low
FSK	frequency shift keying
FSL	flow switch, low
FSNL	full speed no load
FSPF	first stage permeate flow
FST	full stroke testing
FT	Fourier transfer or flow transmitter
FTA	fault tree analysis
FTC	fail to close
FTIR	Fourier transform infrared
FTL	fuzzy technology language
FTNIR	Fourier near infrared
FTO	fail to open
FTP	file transfer protocol
FTS	fault tolerant system
FTU	formazin turbidity unit
FW	finished water

G

g	acceleration due to gravity (= 9.806 m/s^2) or conductivity
G	giga, prefix meaning 10^9 or process gain or conductance
GA	genetic algorithm
gal.	gallon (= 3.785 liters)
GB	gigabyte, 1,000,000,000 bytes
GbE	gigabit Ethernet
Gbps or GBPS	gigabits per second
G_c	feedback controller transfer function
GC	gas chromatograph

g-cal	gramcalorie, *q.v.*; also cal
GD	group of dryers
G_d	unmeasured disturbance transfer function
G_D	measured disturbance transfer function
G_D	approximate feedforward transfer function model
GEOS	geosynchronous Earth orbit satellites
GFC	gas filter correlation
G_ff	feedforward controller transfer function
GHz	giga-Hertz
GI	galvanized iron
GIAC	general inquiry access code
GLR	gas-to-liquid ratio
G-M	Geiger–Mueller tube, for radiation monitoring
G_m	model transfer function
GMC	generic model control
GMR	giant magneto resistive
GMV	generalized minimum variance
GOSUB	go to subroutine
GOX	gaseous oxygen
GP	genetic programming
G_p	process transfer function or process gain
GPa	giga-Pascal (10^9 Pa)
GPC	generalized predictive control
GPCP	general process control programming
GPH or gph	gallons per hour (= 3.785 lph)
GPC	generalized predictive controller
GPM or gpm	gallons per minute (= 3.785 lpm)
GPS	global positioning satellite
gr	gram
grn	green (wiring code color for grounded conductor)
GSC	gas-solid chromatography
GSD	Profibus version of an electronic data sheet
GT	gas turbine
GTG	gas turbine generator
GTO	gate-turn-off thyristor
GUI	graphical user interface
GWR	guided wave radar
Gy	gray, symbol for derived SI unit of absorbed dose, joules per kilogram, J/kg

H

h	1) height; 2) hour; 3) suffix indicating heavy key component; 4) hour
H	1) humidity expressed as pounds of moisture per pound of dry air; 2) henry, symbol of derived SI unit of inductance, volt · second per ampere, V · s/A; 3) high; 4) humidifier
H1	field-level fieldbus, also refers to the 31.25 Kbps instrinsically safe SP-50, IEC61158-2 physical layer
HAD	holographic autostereoscopic display or historical data access
HART	highway addressable remote transducer
HAZOP	HAZard and OPerability studies
HC	horizontal cross-connect or heating coil
HCN	hydrogen cyanide
HD	hot deck
HEC	header error check
HF	hydrogen fluoride or hydrofluoric acid
HFE	human factors engineering
HFT	hardware fault tolerance
HGBP	hot gas bypass
HGC	hydraulic gap control
hh	suffix indicating heavier key component
HH	high-high
hhv	higher heating value
HIC	hand indicating controller
HIPPS	high-integrity pressure protection system
HIPS	high-integrity protection systems
HIS	human interface station
HIST	host interface system test
HK	heavy key
HLL	high-level logic
HLLS	high-low limit switch
HMI	human–machine interface
HMP	hexametaphosphate
HMSD	hexamethyldisiloxane
hor.	horizontal
HP or hp	horsepower (U.S. equivalent is 746 W) or high pressure
HPBF	high performance butterfly valves
HPD	hybrid passive display
HPLC	high pressure (or precision) liquid chromatography
hr	hour
H&RA	hazard & risk analysis
HRH	hot reheat
HRL	hysteresis, repeatability and linearity
HRSG	heat recovery steam generator
HS	hand switch or hot standby
HSE	high-speed Ethernet (host-level fieldbus)
HSF	hydrofluosilic acid
his	human system interface
HTG	hydrostatic tank gauging
HTML	hyper text markup language
HTTP	hyper text transfer protocol
HV	high voltage
HVAC	heating, ventilation, and air conditioning
H/W	hardware
HWD	height, width, depth
HWP	hot water pump
HWS	hot water supply
Hz	Hertz, symbol for derived SI unit of frequency, one per second (l/s)

I

I	integral time of a controller in units of time/repeat
IA	instrument air or impedance amplifier

IAC	inquiry access code
IAE	integral of absolute error
IAQ	indoor air quality
ibid	in the same place
IC	integrated circuitry, intermediate cross-connect, initial condition, or inorganic carbon
I&C	instrumentation and control or information and control
ICA	independent computing architecture
ICCMS	inadequate core cooling monitoring system
ICMP	Internet control message protocol
ICP	inductively coupled plazma
ID	inside diameter or induced draft
IDCOM	identification command
i.e.	*id est*: that is
IE	integral error
I&E	instrument and electrical
IEC	interactive evolutionary computation
IEH	Instrument Engineers' Handbook
IETF	Internet engineering task force
IF	intermediate frequency
IFS	initiate fault state
IGBT	insulated gate bipolar transistor
IGV	inlet guide vane
IIS	Internet information server
IL	instruction list
ILD	instrument loop diagrams
IMC	internal model control or injection cycle
iMEMS	integrated micro-electro-mechanical system
IN	insertion
in.	inch (= 25.4 mm)
IrGaAs	iridium gallium arsenide
in-lb	inch-pound (= 0.113 N × m)
I/O	input/output
IP	Internet protocol or ionization potential or intermediate pressure
I-P	current-to-pressure conversion
IPA	isopropyl alcohol
IPL	independent protection layer
IPS	in-plane switching
IPTS	international practical temperature scale
IR	infrared
IRQ	interrupt request queue
IS	intermediate system or intrinsic safety
ISAB	ionic strength adjustment buffer
ISE	integral of squared error or ion selective electrode
ISFET	ion-selective field effect transistor
ISM	industrial, scientific, medical
ISP	Internet service provider or interoperable system provider
IT	information technology (as in IT manager or IT department) or current transmitter
ITAE	integral of absolute error multiplied by time
ITD	indium tin oxide
ITSE	integral of squared error multiplied by time
ITT	intelligent temperature transmitters
IWF	initial weighing factor
IXC	intereXchange carrier

J

J	joule, symbol for derived SI unit of energy, heat or work, newton-meter, N · m
JB	junction box
JIT	just-in-time manufacturing
JSR	jump to subroutine
JT	Joule Thomson
JTU	Jackson turbidity unit

K

k	kilo, prefix meaning 1000
K	coefficient, also dielectric constant or process gain
°K or K	degrees Kelvin, symbol for SI unit of temperature or process gain (dimensionless)
Kbs, Kbps	kilo bits per second
KBs	kilo bytes per second
k-cal	kilogram-calories (= 4184 J)
kg	kilogram, symbol for basic SI unit of mass
kg-m	kilogram-meter (torque, = 7.233 foot-pounds)
KHP	potassium acid phthalate
kip	thousand pounds (= 453.6 kg)
kJ	kiloJoule
km	kilometer
KOH	potassium hydroxide
K_p	proportional gain of a PID controller
kPa	kilo-Pascals
K_u	ultimate controller gain
kVA	kilovolt-amperes
KVSP	Kalrez valve stem packing
kW	kilowatts
KWD	kilowatt demand
kWh	kilowatt-hours (= 3.6 × 10^6 J)
KWIC	Kilowatt indicating controller

L

l	suffix indicating light key component
L	1) liter (= 0.001 m^3 = 0.2642 gallon), 2) length; 3) inductance, expressed in henrys; 4) low; 5) reflux flow rate
L2F	laser two-focus anemometer
Lab	CIE functions for lightness, red/green, blue/yellow
LACM	local automatic control mode
LAG	liquid argon
LAN	local area network
LAS	link active scheduler (FF)
Lat	latitude

Lb or lb	pound (= 0.4535 kg)
lbf	pound force
lbm	pound mass
LC	level controller or liquid chromatography
LCD	liquid crystal display
Lch	CIE functions for lightness, chroma, hue
LCI	load commutated inverter
LCM	life cycle management
LCSR	loop current step response
LCV	level control valve
LD	ladder diaphragm or load
LDA	laser Doppler anemometer
LDAD	load distribution algorithm (dryers)
LDAW	load distribution algorithm (wells)
LDP	large display panel
LEC	local exchange carrier or lower explosive limit
LED	light emitting diode
LEL	lower explosive limit
LEO	low Earth orbit satellites
LEPD	light emitting polymer display
LF	linear feet
L/F	reflux-to-feed ratio
LFL	lower flammable limit
LGR	liquid-to-gas ratio
LI	level indicator
LIC	level indicator controller
LIDAR	laser induced doppler absorption radar or light detection and ranging
LIFO	last-in first-out
lim. or lim	limit
lin.	linear
LIN	liquid nitrogen
liq.	liquid
LK	light key
ll	suffix indicating lighter key component
L/L	lead-lag
LL	leased lines or low-low
LLC	logical link control
LLK	lighter than light key
LLOI	low-level operator interface
lm	lumen, symbol for derived SI unit of luminous flux, candela.steradian, cd.sr
ln	Naperian (natural) logarithm to base e
LNG	liquified natural gas
LO	lock open or local override
LOC	limiting oxygen concentration
log or $\log_{10}$	logarithm to base 10; common logarithm
LOI	local operation interface
long.	longitude
LOPA	layers of protection analysis
LOR	local-off-remote
LOS	line of sight
LOX	liquid oxygen
LP	1) liquified petroleum; 2) linear programming; 3) low pressure
LPC	1) large particle content; 2) line predictive coding
LPG	1) liquified petroleum gas; 2) liquified propane gas
lph	liters per hour (0.2642 gph)
lpm	liters per minute (0.2642 gpm)
LPR	linear polarization resistance
LPV	linear parameter varying
LQG	linear quadratic Gaussian
LRC	1) longitudinal redundancy check; 2) level recording controller
LRL	lower range limit
LRV	lower range value
LS	1) level switch; 2) lime slurry
LSB	least significant bit
LSI	large-scale integrator
LSL	lower specification limit
LTI	linear time-invariant
LV	low voltage
LVDT	linear variable differential transformer
LVN	limiting viscosity number
lx	lux, symbol for derived SI unit of illumination, lumen per square meter, lm/m^2

M

m	1) meter, symbol for basic SI unit of length; 2) milli, prefix meaning 10^{-3}; 3) minute (temporal); also min
M	1) thousand (in commerce only); Mach number; 2) molecular weight; 3) mole; 4) mega, prefix meaning 10^6
mA or ma	milliamperes (= 0.001 A)
MAC	medium access control or model algorithm control or main air compressor
MACID	medium access control identifier
MAE	minimum absolute error
MAOP	maximum allowable operating pressure
MAP	manufacturing automation (access) protocol
MAU	media access unit
MAWP	maximum allowable working pressure
Max	maximum
MB	megabyte, 1,000,000 bytes or megabits
MBC	model-based control
MBPC	model-based predictive control
MBPS, mbps, Mbps, Mbs, MBs	megabits per second
MC	main cross-connect
MCFC	molten carbonate fuel cell
mCi or mC	millicuries (= 0.001 Ci)
m.c.p.	mean candle power
MCP	main control panel or manual control mode
MCR	master control relay
MCT	mercury cadmium telluride or minimum cycle time

MDBS	mobile data base station
MDIS	mobile data intermediate system
MDPH	maximum pressure difference (header towell)
m/e	mass-to-energy ratio
med.	medium or median
MEDS	medium Earth orbit satellites
MEMS	micro electro mechanical structures
m.e.p.	mean effective pressure
MES	manufacturing execution system, management execution system, or mobile end station
MeV	mega-electron volt
MF	micro filtration
MFAC	model free adaptive control
MFC	model free control or microbiological fuel cell
MFD	mechanical flow diagram
MFE	magnetic flux exclusion
mfg	manufacturer or manufacturing
MFI	melt flow index or melt factor index
MFR	manual flow rate
mg	milligrams (= 0.001 gr)
MGD	million gallons per day
mg/l	milligrams per liter
mho	outdated unit of conductance, replaced by Siemens (*S*), *q.v.*
MHz	megahertz
mi	miles (= 1.609 km)
MI	melt index
MIB	management information base
micro	prefix = 10^{-9}; also μ (mu) or µm and sometimes u, as in ug or µg, both meaning microgram (= 10^{-9} kg)
micron	micrometer (= 10^{-6} m)
MIE	minimum ignition energy
MIMO	multiple-input multiple-output
MIMOSA	machinery information management open system alliance
min	1) minutes (temporal); also m; 2) minimum, 3) mobile identification number
MIR	multiple internal reflection
MIS	management information system
ml	milliliters (= 0.001 l = 1 cc)
MLR	multiple linear regression
mm	millimeters (= 0.001 m) or millimicron (= 10^{-9} m)
MMAC	multiple model adaptive control
mmf	magnetomotive force in amperes
MMI	man–machine interface
mmpy	millimeters per year
MMS	machine monitoring system or manufacturing message specification
MMSCFD	million standard cubic feet per day
MMV	manually manipulated variable
MOC	management of change
MODBUS	a control network
MODEM	modulator/demodulator
MOEA	multiobjective evolutionary algorithm
MOGA	multiobjective genetic algorithm
MOGP	multiobjective genetic programming
mol	mole, symbol for basic SI unit for amount of substance
mol.	molecules
MOON	M out of N voting system
MON	motor octane number
MOS	metal oxide semiconductor
MOSFET	metallic oxide semiconductor field-effect transistor
MOV	metal oxide varistor or motor-operated valve or most open valve
MOVC	most open valve control
mp or m.p.	melting point
MP	medium pressure
MPa	mega Pascal (10^6 Pa)
MPC	model predictive control
MPEC	mathematical problem with equilibrium constraints
MPFM	multiphase flowmeter
mph or MPH	miles per hour (1.609 km/h)
MPHC	model predictive heuristic controller
MPM or mpm	meters per minute
mps or m/s	meters per second
MPS	manufacturing periodic/aperiodic services
Mpy	mills per year
mR or mr	milliroentgens (= 0.001 R)
MRAC	model reference adaptive control
mrd	millirads (= 0.001 rd)
mrem	milliroentgen-equivalent-man
MRP	material requirement planning or manufacturing resource planning or material/product planning
ms	milliseconds (= 0.001 s)
MS	mass spectrometer or Microsoft
MSA	metropolitan statistical areas
MSB	most significant bit
MSC	monitoring and sequential control
MSD	most significant digit
MSDS	material safety data sheet
MT	measurement test
MTBE	methyl tertiary butyl ether
MTBF	mean time between failures
MTSO	mobile telephone switching offices
MTTF	mean time to fail
MTTFD	mean time to fail dangerously
MTTFS	mean time to spurious failure
MTTR	mean time to repair
MTU	master terminal unit
MUX	multiplexer
MV	minimum variance or manipulated variable
MVA	multiple-domain vertical alignment
MVC	minimum variance controller or multivariable control

MW	megawatts (= 10^6 W)
MWC	municipal waste combustors
MWD	molecular weight distribution

N

n	1) nano, prefix meaning 10^6; 2) refractive index; 3) number of trays
N	newton, symbol for derived SI unit of force, kilogram-meter per second squared, $kg \cdot m/s^2$
N_0	Avogadro's number (= 6.023×10^{23} mol^{-1})
N-16	nitrogen-16
NA	numeric aperture
NAAQS	National Ambient Air Quality Standards
NAP	network access port/point
NARMAX	nonlinear autoregressive w/exogenous moving average input nodes
NARX	nonlinear autoregressive w/exogenous input nodes
NAT	network address translation
NB	nominal bore, internal diameter of a pipe in inches
NBJ	nonlinear Box-Jenkins
NC or N/C	normally closed (switch contact)
NC	numeric controller
NCAP	networking capable application processors
NDIR	nondispersive infrared
NDM	normal disconnect mode
NDT	nondestructive testing
NEC	National Electrical Code
NESC	National Electrical Safety Code
NEXT	near end cross-talk
nF	nano filtration
NFI	near field imaging
NFIR	nonlinear finite impulse response
NIC	network interface card
NIP	normal incident pyrheliometer
NIR	near infrared
Nm or nm	nanometer (10^{-9} meter)
NMR	nuclear magnetic resonance
NMV	normal mode voltage
NO or N/O	normally open (switch contact)
NOE	nonlinear output error
NPN	transistor with base of p-type and emitter and collector of n-type material
NPS	nominal pipe size, the internal diameter of a pipe in inches
NPSH	net positive discharge head
NPSHA	net positive discharge head available
NPSHR	net positive discharge head required
NPT	network time protocol
NR	narrow range
NRM	normal response mode
NRZ	nonreturn to zero (NZR refers to a digital signaling technique)
NS	nominal pipe size, the internal diameter of a pipe in inches
NTC	negative temperature coefficient
NTP	network time protocol or normal temperature and pressure, corresponding to 1 atm. absolute (14.7 psia) and 0°C (32°F)
NTSC	national television standards code
NTU	nephalometric turbidity unit
NUT	network update time

O

OA	operational amplifier or outside air
OAC	operator automatic control
OAD	outside air damper
OCD	orifice-capillary detector
OD	outside diameter or oxygen demand
ODBC	open database connectivity or communication
OEM	original equipment manufacturer
OES	optical emission spectrometer
oft or OFT	optical fiber thermometry
ohm	unit of electrical resistance; also Ω (omega)
OI	operator interface
OI-F	operator interface for filtering
OI-P	plant operator interface
OIU	operator interface unit
OJT	on-the-job training
OL	overload
OLE	object linking and embedding
OLED	organic LED
OLE_DB	object linking and embedding data base
OMAC	open modular architecture controls
OMMS	optical micrometer for micro-machine
ON	octane number
OP	output or operating point
OPAM	online plant asset management
OPC	object link embedding (OLE) for process control
OPEX	OPerational EXpenditure
OP-FRIR	open path Fourier-transform infrared
OP-HC	open path hydrocarbon
OP-TDLAS	open path tunable diode-laser absorption spectroscopy
OP-UV	open path ultraviolet
Or	orange (typical wiring code color)
OREDA	Offshore Reliability Data Handbook
ORP	oxidation-reduction potential
OS	operator station or operating system
OSFP	open shortest path first
OSI	open system interconnection (model)
OSI/RM	open system interconnect/reference model
OT	operator terminal or open tubular
OTDR	optical time domain reflectometer
OTSG	once-through steam generator
oz	ounce (= 0.0283 kg)
OZ	ozone

OZA	ozonated air
OZW	ozonated water

P

p	1) pressure; 2) pico, prefix meaning 10^{-12}, also resistivity
P&ID	piping and instrumentation diagram
Pa	pascal, symbol for derived SI unit of stress and pressure, newtons per square meter, N/m^2
PA	plant air, phase angle, or pole assignment
PAC	path average concentration or process automation controllers
PAFC	phosphoric acid fuel cell
PAH	pressure alarm, high
PAL	phase alternating line or pressure alarm, low
PAM	pulse amplitude modulation
PAN	personal area network
Pas	Pascal-second, a viscosity unit
PAS	process automation system (successor to DCS) or project application specification
PB	proportional band of a controller in % (100%/controller gain) or push button
PC	personal computer (MS-Windows based) or pressure controller
PCA	principal component analysis
PCB	printed circuit board
PCC	pressure correction control
PCCH	pressure correction control mode
PCCS	personal computer control system
PCDD	polychlorinated dibenzo-*p*-dioxin
PCDF	polychlorinated dibenzo furans
PCL	peer communication link
PCM	pulse code modulation
PCR	principal component regression
PCS	process control system or personal communication services
pct	percent; also %
PCT	Patent Cooperation Treaty
PCTFE	polychlorotrifluoroethylene
PCV	pressure control valve
PD	positive displacement or proportional and derivative or percentage detected or position detector
PDA	personal digital assistant or photodiode array
PDD	pulsed discharge detector
PDF	probability density function, probability of failure or portable document file
PDIC	pressure differential indicating controller
PDLCP	partial differential linear complementary problem
PdM	predictive maintenance
PDM	pulse duration modulation
PDP	plasma display panel
PDS	phase difference sensor
PDU	protocol data unit
PDVF	polyvinylidene fluoride
PE	polyethylene or penalty on error or pressure element
PED	pressure equipment directive
PEEK	poly ether ether ketone
PEL	permissible exposure level
PEMFC	proton exchange membrane fuel cell or polymer electrolyte membrane fuel cell
PES	programmable electronic system
Pf	picofarad (= 10^{-12} F)
PF or p.f.	power factor
PFA	per-fluoro-alkoxy copolymer (a form of Teflon)
PFC	procedure functional chart or procedure function chart
PFD	process flow diagram or probability of failure on demand
PFDavg	average probability of failure on demand
PFPD	pulsed flame photometric detector
PGC	process gas chromatograph
PGNAA	prompt gamma neutron activation analysis
pH	acidity or alkalinity index (logarithm of hydrogen ion concentration)
PHA	process hazard analysis
pi or pl	Poiseuille, a viscosity unit
PI	proportional and integral or pressure indicator
P/I	pneumatic-to-current (conversion)
P&I	piping and instrument (diagram)
PIC	pressure indicating controller or path integrated concentration
PID	proportional, integral, and derivative (control modes in a classic controller) or photoionization detector
P&ID	piping (process) and instrumentation diagram (drawing)
PI-MDC	path integrated minimum detectable concentration
PIMS	process information management system
PIO	program input–output
PIP	process industry practices
PIR	precision infrared radiometer
PL	preload
PLC	programmable logic controller
PLD	programmable logic devices
PLED	polymeric LED
PLL	phase locked loop
PLS	physical layer signaling or partial least squares
PM	photo multiplier or penalty on moves or phase modulation
PMA	physical medium attachment
PMBC	process model based control
PMD	photomultiplier detector
PMF	probability mass function

PMLCD	passive matrix LCD
PMMC	permanent magnet moving coil
PMS	plant monitoring system
PMT	photo-multiplier tube or photometer tube
PN	pressure, nominal
PNP	transistor with base of n-type and emitter and collector of p-type material
PO	polymer
POF	positive opening feature
POL	problem oriented language
POPRV	pilot operated pressure relief valve
POTW	publicly owned treatment works
PP	polypropylene or pole placement
ppb or PPB	parts per billion
PPC	process procedure chart
PPD	pounds per day
ppm or PPM	parts per million or pulse position modulation
ppmV	volumetric parts per million
PPP	point-to-point protocol
ppt	parts per trillion
PRBS	pseudo random binary sequence
PRC	pressure recording controller or production cycle
PRD	pressure relief device
Precip	precipitate or precipitated
PRV	pressure relief valve or pressure reducing valve
PS	power supply (module) or pressure switch or partial stroke
PSA	pressure swing adsorption
PSAT	pre-startup acceptance test
PSD	power spectral density or photosensitive device
PSE	pressure scale effect
PSG	phosphosilicate glass
PSH	pressure switch, high
PSI	pre-startup inspection
psi or PSI	pounds per square inch (= 6.894 kPa)
PSIA or psia	absolute pressure in pounds per square inch
PSID or psid	differential pressure in pounds per square inch
PSIG or psig	above atmospheric (gauge) pressure in pounds per square inch
PSK	phase shift keying
PSL	pressure switch, low
PSM	process safety management
PSS	programmable safety system
PSSR	pre-startup safety review
PST	partial stroke testing
PSTN	public switched telephone network
PSU	post-startup
PSV	pressure safety valve
pt	point or part or pint (= 0.4732 liter)
PT	pressure transmitter
PTB	Physikalisch-Technische Bundesanstalt
PTC	positive temperature coefficient or programmable telemetry controller
PTFE	polytetrafluoroethylene (conventional Teflon)
PU	per unit
PUVF	pulsed ultraviolet fluorescence
PV	process variable (measurement) or the HART primary variable
PVC	polyvinyl chloride
PVDF	polyvinylidene fluoride
PVHI	process variable high (reading or measurement)
PVLO	process variable low (reading or measurement)
PW	private wire
PWM	pulse width modulation
PWR	pressurized water reactor
PZT	lead-zirconate-titanate ceramic

Q

q	1) rate of flow; 2) electric charge in coulombs, C
q^{-1}	backward shift operator
Q	quantity of heat in joules, J or electric charge
°Q	Quevenne degrees of liquid density
QA	quality assurance
QAM	quadrature amplitude modulation
QBET	quench boiler exit temperature
QCM	quartz crystal microbalance
QEV	quick exhaust valve
QMS	quality management system
QoS	quality of service
QPSK	quadrature phase shift keying
qt	quart (0.9463 liter)
q.v.	*quod vide*: which see
QV	quaternary variable

R

r	radius; also rad
R	1) resistance, electrical, ohms; 2) resistance, thermal, meter-kelvin per watt, $m \cdot K/W$; 3) gas constant ($= 8.317 \times 10^7$ erg $\cdot$ mol^{-1},$°C^{-1}$); 4) roentgen, symbol for accepted unit of exposure to *x* and gamma radiation, ($= 2.58 \times 10^{-4}$ C/kg)
r^2	multiple regression coefficient
Ra	radium
RA	return air or reaction (failure) alarm or reverse action
RACM	remote automatic control mode
Rad	1) radius; also r; 2) radian, symbol for SI unit of plane angle measurement or symbol for accepted SI unit of absorbed radiation dose, (= 0.01 Gy)
RAD	return air damper
RADAR	radio detection and ranging

RAID	redundant array of inexpensive disks
RAM	random access memory
RASCI	responsible for, approves, supports, consults, informed
RAT	return air temperature
RBC	remote base communication module
RCF	rotational centrifugal force
RCU	remote control unit
R&D	research and development
RD	rupture disk
RDP	remote desktop protocol
Re	Reynolds number
REC	roll eccentricity compensation
ReD	Reynolds number corresponding to a particular pipe diameter
redox	oxidation–reduction
rem	measure of absorbed radiation dose by living tissue (*r*oentgen *e*quivalent *m*an)
rev	revolution, cycle
RF or rf	radio frequency or return fan
RFC	request for comment (an Internet protocol specification) or rotational centrifugal force or rising film concentrator
RFF	remote fiber fluorimetry
RFI	radio frequency interference
RFQ	request for quote
RG	relative gain
RGA	residual gas analyzer or relative gain array or relative gain analysis
RGB	red, green, and blue
RGM	reactive gaseous mercury
RH	relative humidity or reheated or reheater
RHA	relative humidity alarm
RHC	receding horizon adaptive controller or reheat coil
RHCV	relative humidity control valve
RHE	relative humidity element
RHPZ	right half plane zero
RI	refractive index
RIO	remote operator interface
RIP	routing information protocol
RJT	ring joint type
r(k)	set point
RLL	relay ladder logic
RMS or rms	root mean square (square root of the mean of the square) or rotary mirror sleeves
RNG	ring number
RO	reverse osmosis
ROI	return on investment
ROM	read-only memory
RON	research octane number
RPC	remote procedure call (RFC1831)
RPG	remote password generator
RPM or rpm	revolutions per minute
rps	revolutions per second

R_r	reaction rate, slope
RRF	risk reduction factor
RRT	relative response time (the time required to remove the disturbance)
RS	recommended standard
RSA	rural service areas
RSS	root sum squared
RSSI	received signal strength indicator
RTD	resistance temperature detector
RTL	resistance-transistor logic
RTO	real time optimization or operation
RTOS	real time operating system
RTR	remote transmission request
RTS	ready (or request) to send
RTS/CTS	request to send/clear to send
RTU	remote terminal unit
RUDS	reflectance units of dirt shade
RV	relief valve
RVDT	rotary variable differential transformer
RVP	Reid vapor pressure
RW	raw water
RWF	raw water flow
RWS	remote work station

S

S or s	1) second, symbol for basic SI unit of time; 2) Laplace variable; 3) Siemens (Siemens/cm), symbol for unit of conductance, amperes per volt, A/V; 4) separation
SA	supply air
SAN	surface acoustic wave
SAP	service access point
SASFC	single automatic stage flow control
sat.	saturated
SAT	site acceptance test or supervisory audio tone or supply air temperature
SAW	surface acoustic wave
S.B.	set-point bandwidth
SC	system codes or speed controller
SCADA	supervisory (system) control and data acquisition
SCCM	standard cubic centimeter per minute
SCD	streaming current detector or sulfur chemilumenesce detector
SCE	saturated calomel electrode
SCFH	standard cubic feet per hour
SCFM	standard cubic feet per minute (air flow at 1.0 atm and 70°F)
SCM	station class mark
SCMM	standard cubic meter per minute
SCO	synchronous connection oriented
SCOT	support coated open tubular (column)
SCR	silicone-controlled rectifier or selective catalytic reduction
SCS	sample control system

SD	component in leg has failed safe and failure has been detected
S/D	shut down (valve)
SDIU	Scanivalve digital interface unit
SDN	send data with no acknowledgement
SDS	smart distributed system
SEA	spokesman election algorithm or slurry density sensor
Sec or sec	seconds; also s
SEI	system efficiency index
SELV	source: extra low voltage
SER	sequence of event recorder
SF	supply fan
S/F	smoke and fire (detector)
SFC	sequential function chart or system function configuration or static frequency converter
SFD	system flow diagram or start of frame delimiter
SFF	safe failure fraction
SFI	sight flow indicator
SFR	spurious failure rate
SG or SpG	specific gravity; also sp. gr.
S/H or S&H	sample and hold
SH	superheated or superheater
SHE	standard hydrogen electrode
SHS	sample handling system
SI	system international
SIC	speed indicating controller
SID	system identification digit (number)
SIF	safety instrumented function
SIG	special interest group
SIL	safety integrity level
sin	sine, trigonometric function
SIS	safety instrumented system
SISO	single-input single output
SKU	stock keeping units
SLAMS	state and local air monitoring stations
SLC	safety life cycle or single loop controller
slph	standard liters per hour
slpm	standard liters per minute
SMC	sliding mode control
SMCr	sliding mode controller
SMR	specialized mobile radio
SMTP	simple mail transfer (management) protocol
S/N	signal-noise (ratio)
SNCR	selective non-catalytic reduction
SNG	synthetic natural gas
SNMP	simple network management protocol
SNR	signal-to-noise ratio
SNTP	simple network time protocol
SOA	safe operation area
SOAP	simple object access protocol (an Internet protocol that provides a reliable stream-oriented connection for data transfer)
SOE	sequence of events
SOER	sequence of events recorder
SOFA	secondary over fire air
SOFC	solid oxide fuel cell
SONAR	sound navigation ranging
SOP	standard operating procedure
SP	set point or Smith predictor
SPC	statistical process control
SPDT	single pole double pole throw (switch)
Sp. G or sp. gr.	specific gravity; also SG
Sph	starts per hour
SPI	serial peripheral interface
SPL	1) sound pressure level or 2) sound power level
SPR	set-point rate
SPRT	standard platinum resistance thermometer
SPST	single pole single throw (switch)
Sq	square
SQC	statistical quality control
SQL	structured (or standard, or sequential) query language
Sr	steradian, symbol for SI unit of solid angle measurement
SRD	send and request data with reply
SRS	safety requirements specification
SRV	safety relief valve
SS	stainless steel or selector switch
SSB	single side band
SSC	solid state contactor
SSE	size scale effect
SSF	Saybolt seconds furol
SSL	secure socket layers
SSR	solid state relay
SSU	Saybolt seconds universal
ST	structured text or steam turbine
STC	self-tuning controller
std.	standard
STEL	short term exposure limit
STEP	standard for the exchange of product model data
STG	steam turbine generator
STIG	steam injection gas turbine
Stm.	steam
STP	shielded twisted pair or standard temperature and pressure, corresponding to 70°F (21.1°C) and 14.7 psia (1 atm. Abs.)
STR	spurious trip rates or self-tuning regulator
SU	security unit or component in leg has failed safe and failure has not been detected
SUS	Seybold universal seconds or stochastic uniform selection
SV	secondary variable or safety valve or solenoid valve
S/W or SW	software
SWIR	short wave infrared
s^2_y	sample variance of output *y*

T

t	1) ton (metric, = 1000 kg); 2) time; 3) thickness
T	1) temperature; 2) tera, prefix meaning 10^{12}; 3) period (= 1/Hz, in seconds); 4) tesla, symbol for derived SI unit of magnetic flux density, webers per square meter, Wb/m2
$T^{1}/_{2}$	half life
TAH	temperature alarm, high
TAL	temperature alarm, low
Tan	tangent, trigonometric function
Tanh	hyperbolic tangent
TAS	thallium-arsenic-selenide
Tau	process time constant (seconds)
TBM	tertiary butyl mercaptan
TBP	true boiling point
t/c	thermal coefficient of linear expansion
TC	thermocouple, temperature controller or total carbon
TCAM	timer/counter access modules
TCD	thermal conductivity detector
TCI	track command in
TCO	track command out
TCP	transmission control protocol
TCP/IP	transmission control protocol/Internet protocol
TCV	temperature control valve
t_d	process dead time (seconds)
T_d	derivative time (in seconds) of a PID controller
TD	time delay
TDLAS	tunable diode laser absorption spectroscopy
TDM	time division multiplexing
TDMA	time division multiple access
TDP	dew point temperature
TDR	time domain reflectometry
TE	temperature element
T/E	thermoelectric
TEM	transmission electron microscope
TFELD	thin film electroluminescent display
TFT	thin film transistor
TG	thermogravimetry
TGQ	total gas supply
TH	upper limit of comfort zone
THC	total hydrocarbon
THR	total heat release
Ti	integral time (in seconds) of a PID controller
TI	test interval (time interval between tests) or temperature indicator
TIC	temperature indicating controller or total inorganic carbon
TIFF	tagged image file format
TISAB	total ionic strength adjustment buffer
TL	lower limit of comfort zone
TLV	threshold limit value
TMP	1) thermo-mechanical pulp; 2) transmembrane pressure
TMR	triple modular redundancy
TMT	tube metal temperature
TN	total nitrogen or twisted nematic
t_o (t_d)	process dead time (seconds)
TOC	total organic carbon
TOD	total oxygen demand
TOF	time of flight
TOP	technical and office protocol
TP	turbine protection
TPD	tons per day
TQM	total quality management
TR	temperature recorder or time delay relay
T/R	transmit/receive
TRC	temperature recording controller
TRI	track reference in
TRO	track reference out
T.S.	tensile strength
TSA	temperature swing adsorption
TSH	temperature switch, high
TSL	temperature switch, low
TSM	thermal stress monitoring
TSR	terminate and stay resident
TT	temperature transmitter or transit time
TTC	tungstram titanium carbide or time to close
TT&C	telemetry, tracking and command
TTFM	transit time flow measurement
TTL	transistor-transistor logic or time to live
TTO	time to open
TTP	through the probe
t_u	ultimate period
TV	tertiary variable
°Tw	Twadell degrees of liquid density
TWA	time weighed average
TWB	wet bulb temperature
TWM	technical working method
TY	temperature relay
τ (Tau)	process time constant (seconds)
τF	PV filter time constant F

U

u	prefix = 10^{-6} when the Greek letter μ is not available
UART	universal asynchronous receiver transmitter
UBET	unbiased estimation
UCMM	unconnected message manager
UDP	user/universal data/datagram protocol
UEL	upper explosive limit
UF	ultra filtration
$u_{fb}(k)$	feedback controller output
UFD	utility flow diagram
$u_{ff}(k)$	feedforward controller output
UFL	upper flammable limit
UGS	underground gas storage
UHF	ultra high frequency
UHSDS	ultra high speed deluge system
u(k)	controller output

UML	universal modeling language
UPS	uninterruptable power supply
UPV	unfired pressure vessel
URL	upper range limit
URV	upper range value
USART	universal synchronous/asynchronous receiver transmitter
USB	universal serial bus
USL	upper specification limit
USV	unloading solenoid valve
UTP	unshielded twisted pair
UTS	ultimate tensile stress
UUP	unshielded untwisted pair
UV	ultraviolet
UVS	uninterruptible voltage supply
UV-VIS-NIR	ultraviolet-visible-near infrared

V

v	velocity
v or V	volt, symbol for derived SI unit of voltage, electric potential difference and electromotive force, watts per ampere, W/A
VA	vertical alignment
Vac	a.c. voltage
VAV	variable air volume
VBA	visual basis for applications
VCO	voltage controlled oscillator
VCR	virtual communication relationship
VDC	volts DC
VDF	vacuum fluorescent display
VDT	video display tube
VDU	video display unit or visual display unit
vert.	vertical
VF	vacuum fluorescent
VFD	variable frequency drive or vacuum fluorescent display or virtual field device
VFIR	very fast infrared
VHF	very high frequency
VIS	visible
V-L	vapor–liquid (ratio)
V/M	voltmeter
VME	Virsa Module Europe (IEEE 1014-1987)
VMS	vibration monitoring system
VOC	volatile organic compounds or volatile organic carbon
VP	valve position
VPA	valve position alarm
VPC	valve position controller
VPN	virtual private network
VPS	valve position switch
VR	virtual reality
VRL	very low frequencies
VRML	virtual reality modeling language
vs.	versus
VSA	vacuum swing absorption
VSD	variable speed drive
V&V	verification & validation
VVVF	variable voltage variable frequency drive

W

w	1) width; 2) mass flow rate
W	1) watt, symbol for derived SI unit of power, joules per second, J/s; 2) weight; also wt water; 3) water
WAN	wide area network
Wb	weber, symbol for derived SI unit of magnetic flux, volt · seconds, V · s
WCOT	wall coated open tubular (column)
WDXRF	wavelength dispersion x-ray fluorescence
WF	weighing factor
WFS	water flow switch
WG	standard (British) wire gauge
Wh	white (wiring code color for AC neutral conductor)
WI	Wobble Index
WLAN	wireless local area network
WPAN	wireless personal area network
WS	work station
wt	weight; also W

X

x	molar fraction of light component in bottom product
X	reactance in ohms
XLS	limit switch
XML	eXtensible markup language
XP	blade pitch position
x-ray	electromagnetic radiation
XRF	x-ray fluorescence
XSCV	superheat control valve
XSET	start-up setting
XYZ	tri-stimulus functions

Y

Y	expansion factor, or molar fraction of light component in distillate
y(k)	process output
yd	yard (= 0.914 m)
yr	year

Z

z	molar fraction of light component in feed
Z	1) atomic number (proton number); 2) electrical impedance (complex) expressed in ohms
ZAFC	zinc air fuel cells
ZC	position controller
ZEB	zero energy band
ZIC	position indicating controller
ZOH	zero order hold
ZSC	limit switch – closed

ZSCO	limit switch – closed/open
ZSO	limit switch – open
ZSRG	zero signal reference grid
ZT	position transmitter or zone temperature
∂	partial derivative

MISCELLANEOUS LETTER SYMBOLS

α (alpha)	1) geometric angle; 2) radiation particle (helium atom); 3) linear expansion coefficient; 4) average relative volatility between the key components across the column
β (beta)	radiation particle (electron)
γ (gamma)	1) electromagnetic radiation; 2) surface tension; also σ (sigma)
Δ (delta)	difference, change or deviation from a steady-state condition
ε (epsilon)	1) emmissivity; 2) linear strain, relative elongation $\varepsilon \times \Delta 1/1^\circ$
η (eta)	1) efficiency; 2) viscosity (absolute); also μ
θ (theta)	thermal resistance
λ (lambda)	1) thermal conductivity; 2) wavelength; 3) relative gain
Λ (lambda)	relative gain array
μ (mu)	1) viscosity (absolute); also η; 2) linear attenuation coefficient; 3) prefix, micro = 10^{-6}; also u; 4) mμ: millimicron (10^{-9} m)
μm	micron (10^{-6} m)
ν (nu)	viscosity, kinematic
π (pi)	1) surface pressure; 2) constant = 3.1416…
ρ (rho)	1) density; 2) resistivity
σ (sigma)	1) surface tension; also γ; 2) conductivity; 3) normal stress; 4) nuclear capture cross section
Σ (sigma)	summation
τ (tau)	1) time delay; 2) shear stress; 3) time constant
ω (omega)	angular velocity expressed in radians per second
Ω (omega)	ohm
ϕ	diameter, also dia and D
~	alternating current
′	minute, angular or temporal
″	second, angular
$\perp$	perpendicular to, normal to
$\|$	parallel
%	percent; also pct

GREEK ALPHABET

Α, α	alpha
Β, β	beta
Γ, γ	gamma
Δ, δ	delta
Ε, ε	epsilon
Ζ, ζ	zeta
Η, η	eta
Θ, θ	theta
Ι, ι	iota
Κ, κ	kappa
Λ, λ	lambda
Μ, μ	mu
Ν, ν	nu
Ξ, ξ	xi
Ο, o	omicron
Π, π	pi
Ρ, ρ	rho
Σ, σ	sigma
Τ, τ	tau
Υ, υ	upsilon
Φ, ϕ	phi
Χ, χ	chi
Ψ, ψ	psi
Ω, ω	omega

General

1

1.1
ANALOG VS. DIGITAL INSTRUMENTS 3

1.2
COMPUTER CONFIGURATIONS OF SUPERVISORY UNITS 15

1.3
COMPUTERS — SIMULATION BY ANALOG AND HYBRID SYSTEMS 26

1.4
ELECTRONIC VS. PNEUMATIC INSTRUMENTS 41

1.5
HUMAN ENGINEERING 46

1.6
PROCESS ALARM MANAGEMENT 59

1.7
SPEECH SYNTHESIS AND VOICE RECOGNITION 64

1.8
WIRING PRACTICES AND SIGNAL CONDITIONING 71

1.1 Analog vs. Digital Instruments

W. P. DURDEN (1995) **H. EREN** (2005)

Types: Analog and digital instruments

Costs: A few dollars to thousands of dollars depending on functionality and complexity

Sensors: Common sensors for both types; additionally, IC (integrated circuitry) and intelligent sensors for digital instruments

Components: Purely analog components for analog instruments; analog and digital components and microprocessors for digital instruments

Displays: Needle indicator for analog instruments, numeric light emitting diode (LED) or liquid crystal display (LCD) for digital instruments

Communications: Wired (e.g., 20 mA) for analog instruments; wired, RF, microwave, optical, sonic, etc. for digital instruments

Networking: Limited network capabilities for analog instruments; unlimited capabilities for digital instruments

Vendors (partial list): Analytical Measurements, Inc. (www.analyticalmeasurements.com)
Athena Controls, Inc. (www.athenacontrols.com)
Brighton Electronics, Inc. (www.brighton-electronics.com)
Crompton Instruments (www.crompton-instruments.com)
Devar, Inc. (www.devarinc.com)
Dresser Instruments (www.dresserinstruments.com)
Dunham Instruments (www.dunhaminstruments.com.au)
EDL, Inc. (www.edl-inc.com)
Encore Electronics, Inc. (www.encore-elec.com)
Hioki Corp. (www.hioki.co.jp)
Industrial Instruments & Supplies, Inc. (www.iisusa.com)
Kahn Instruments, Inc. (www.kahn.com)
Morehouse Instrument Co. (www.morehouseinst.com)
Ono Sokki Technology, Inc. (www.onosokki.net)
Precision Devices, Inc. (www.predev.com)
PTC Instruments (www.ptc1.com)
R & R Instrumentation, Inc. (www.rrinst.com)
Taylor Precision Products LP (www.taylorusa.com)
Weschler Instruments (www.weschler.com)
Weston Aerospace (www.westonaero.com)

Digital Instruments

Alldos Eichler GmbH. (www.alldos.com)
AMETEK Instruments (www.ametek.com)
A&S Company, Inc. (www.detection.com)
D&D Security Products, Inc. (www.ddsp.com)
Davis Inotek Instruments (www.davisontheweb.com)
Draeger Safety (www.draeger.com)
Dranetz Technologies, Inc. (www.dranetz.com)

Dwyer Instruments, Inc. (www.dwyer-inst.com)
Dynalco Controls (www.dynalco.com)
Entran Devices, Inc. (www.entran.com)
Fluke Electronics (www.flukecanada.ca)
Garmin, Ltd. (www.garmin.com)
Hanna Instruments, Inc. (www.hannainst.com)
Hoyt Electrical Instrument Works, Inc. (www.hoytmeter.com)
Intercomp (/www.intercomp.com)
Keithley Instruments, Inc. (www.keithley.com)
Koehler Instrument Co., Inc. (www.koehlerinstrument.com)
Leader Instruments Corp. (www.leaderusa.com)
MC Miller Co. (www.mcmiller.com)
NEC San-ei Instruments, Ltd. (www.necsan-ei.co.jp)
NeTech Corporation (www.gonetech.com)
Nihon Kohden (www.nihonkohden.com)
Polar Instruments, Inc. (www.polarinstruments.com)
Protek T&M (www.hcprotek.com)
SE International, Inc. (www.seintl.com)
Sensotec, Inc. (www.sensotec.com)
Sierra Instruments, Inc. (www.sierrainstruments.com)
SunTech Medical Instruments (www.suntechmed.com)
Texas Instruments (www.ti.com)
Turner BioSystems (www.turnerbiosystems.com)
Valhalla Scientific, Inc. (www.valhallascientific.com)
Wagner Instruments (www.wagnerforce.com)
Warren-Knight Instrument Co. (www.warrenind.com)
Winters Instruments, Inc. (www.winters.ca)
Yokogawa Corp. (www.yokogawa.com)

INTRODUCTION

Instruments are manmade devices that measure the parameters of physical variables. They may be analog, digital, or a combination of the two. Nowadays, most instruments are digital because of their advantages over analog counterparts. However, the front ends of many instruments are still analog; that is, most signals from sensors and transducers and the first stage of signal processing are still analog. Nevertheless, it is important to mention that in recent years digital instruments operating purely on digital principles have been developing fast. Today's smart sensors based on digital principles contain the complete signal condition circuits and the sensors in a single chip. The outputs of smart sensors can directly be interfaced with other digital devices.

Sensors and transducers are the most basic and primary elements of all instruments. They respond to physical variations to produce continuous outputs that convey useful information. The outputs may be in the forms of amplitudes, frequencies, or phase differences of currents, voltages, power, or energy. As in the case of all signal-bearing systems, in both analog and digital instruments, there are useful signals that respond to the physical phenomena and unwanted signals that are imposed as various forms of noise.

In the last 10 to 15 years or so, due to rapid progress in integrated circuit (IC) technology and the availability of low-cost analog and digital components and microprocessors, considerable progress in digital instruments has taken place. Therefore, this section will concentrate primarily on digital instruments. However, signals and signal processing will be introduced first since they are common to both types of instruments. In the following sections, the underlying principles of analog and digital instruments will be discussed separately, some examples will be given, and a comparison between the two will be provided at the end of the chapter.

AN OVERVIEW OF SIGNALS AND SIGNAL PROCESSING

All instruments operate on signals detected by sensors from physical phenomena. One of the main differentiators between analog and digital instruments is the signal processing, which requires different theoretical approaches and hardware. The signals generated by sensors can be processed in three different ways:

1. By analog techniques that directly deal with analog signals
2. By converting analog signals into digital forms and implementing the systems as digital instruments
3. By dealing with the signals purely in digital forms as digital-to-digital inputs and outputs

A *signal* is defined as "any physical quantity that varies with time, space, or any other independent variable." Most signals occurring in real world are analog in nature, i.e., they provide a continuous stream of information about the physical quantity. Common electrical signals include variations in voltages, currents, frequencies, electric and magnetic properties, phase relations, etc. On the other hand, the physical phenomena may be temperature, position, sound, vibration, chemical reactions, biological activity, light intensity, and so on.

Analog signals are processed by *analog signal processing* techniques, which can be described as a body of theoretical and practical methods that can be implemented. These techniques include but not restricted to amplifying, filtering, transmitting, estimating, detecting, modulating, demodulating, analyzing, displaying, and reproduction of signals. On the other hand, due to easy and cost-effective availability of advanced microprocessors and the supporting components, the majority of modern instruments are digital and require *digital signal processing*. Digital systems provide powerful processing and data handling capabilities simply by software and/or firmware implementations. Also, once the signals are converted to digital formats, they can be managed by any computer or digital system. This provides a wide range of possibilities for data processing, communications, storage, and visual displays.

CONTINUOUS AND DIGITAL SIGNALS

Continuous signals, also known as analog signals, are defined for every value of time from $-\infty$ to $+\infty$. Continuous signals can be periodic or nonperiodic. In periodic signals, the signal repeats itself in an oscillatory manner, e.g., the sinusoidal waveform (as in Figure 1.1m), which can be expressed as:

$$x(t) = X_m \sin(\omega t) \quad -\infty < t < \infty \qquad \textbf{1.1(1)}$$

where $x(t)$ is a time-dependent signal, $\omega = 2\pi ft$ is the angular frequency, and X_m is the maximum value.

The signals can be periodic but not necessarily sinusoidal; for example, they may be triangular, sawtooth, or rectangular waveforms. The periodic signals can be expressed as a combination of pure sinusoidal waveforms known as the *Fourier series.* That is:

$$\begin{aligned} x(t) = X_0 + X_1 \sin(\omega_1 t + \phi_1) + X_2 \sin(\omega_2 t + \phi_2) + \cdots \\ + X_n \sin(\omega_n t + \phi_n) \end{aligned} \qquad \textbf{1.1(2)}$$

where $\omega_1, \omega_2, \ldots, \omega_n$ are the frequencies (rad/s); $X_0, X_1, \ldots, X_n$ are the maximum amplitudes of respective frequencies; and $\phi_1, \phi_2, \ldots, \phi_n$ are the phase angles.

In Equation 1.1(2), the number of terms may be infinite, and the higher the number of elements the better the approximation. Nevertheless, high-frequency contents can have significant effect on digital signal processing, as will be explained next.

Digital Signals

Digital signals are arrays of numbers that are used in computational analysis of systems and signals. Digital signals can either be generated or directly be derived from analog signals using A/D (analog to digital) converters.

Analog-to-Digital In the A/D conversion process, analog signals are first sampled at discrete time intervals to obtain a sequence of samples that are spaced uniformly in time. This is achieved by multiplying a continuous signal by a periodic train of Dirac delta functions spaced T seconds apart. T represents the sampling period; its reciprocal is the sampling rate ($f_s = 1/T$). After the analog signals have been sampled, *quantization* is necessary to put the discrete-time signal into discrete numerical values. This is followed by *coding* to convert the quantized values to binary sequences.

For example, a discrete-time sinusoidal signal may be expressed in terms of sequences as:

$$x(nT) = A\sin(\omega nT + \theta), \qquad \textbf{1.1(3)}$$

where n is the integer variable termed the *sample number*, and T is the time interval between samples known as the *sampling period.*

The Fourier series representation of a periodic discrete-time signal $x(n)$ sequence is:

$$x(n) = \sum_{k=0}^{N-1} c_k e^{j2\pi kn/N} \qquad \textbf{1.1(4)}$$

where N is the fundamental period.

The coefficient c_k is given by:

$$c_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\exp\left[\frac{-j2\pi kn}{N}\right] \qquad \textbf{1.1(5)}$$

The frequency-domain of discrete-time signals that are not necessarily periodic can be expressed in a discrete Fourier transform (DFT) as:

$$X(\omega T) = \sum_{n=-\infty}^{\infty} x(n)\exp[-j\omega nT] \qquad \textbf{1.1(6)}$$

and

$$x(n) = \frac{1}{2\pi}\int_0^{2\pi} X(\omega T)\exp[-j\omega Tn]d(\omega T) \qquad \textbf{1.1(7)}$$

These equations are essential in signal processing for digital instruments where analog-to-digital signal conversion takes place.

In all digital instruments, the Nyquist sampling theorem must be observed; that is, "the number of samples per second must be at least twice the highest frequency present in the continuous signal," Equation 1.1(2). As a rule of thumb, depending on the significance of the high frequencies, the

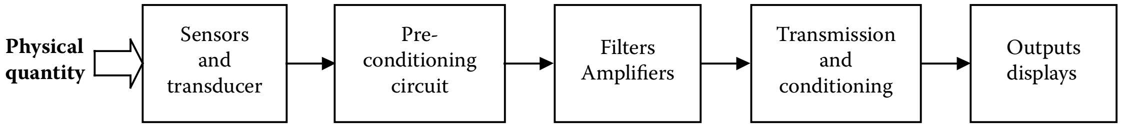


FIG. 1.1a
Functional blocks of an analog instrument.

sampling must be about 5 to 10 times the highest frequency component in the signal.

ANALOG INSTRUMENTS

Analog instruments are characterized by their continuous signals. A purely analog instrument measures, transmits, displays, and stores data in analog form. The signal processing is realized by analog components that are integrating together as functional blocks, as illustrated in Figure 1.1a. Some examples of functional blocks are bridges, amplifiers, filters, oscillators, modulators, offset circuits, level converters, and buffers.

Three basic components are common in all types of analog signal processing: resistors, capacitors, and inductors. The main function of a resistor is to limit the current. The current will flow through a capacitor only if the voltage changes across it. In the case of inductors, voltage is established only as a result of a change in the current across it. Other analog components, including semiconductor devices such as diodes, transistors, operational amplifiers, and rectifiers, are based on these three basic elements.

Two basic types of semiconductor devices exist: bipolar and metal oxide semiconductors (MOS). Many modern analog components and circuits are manufactured in integrated circuit (ICs). ICs contain all basic circuit components such as resistors, transistors, and diodes densely packed on silicon chips. The operational amplifier (op amp) is an example of an IC.

Operational amplifiers are fundamental building blocks of analog signal processing circuits. They are made either as integrated (monolithic) circuits or hybrid circuits (combination of monolithic and discrete parts). An op amp is made from hundreds of transistors, resistors, and capacitors in a single chip. It has two inputs (inverting and noninverting) and a single output. A good operational amplifier has the following properties:

- High input resistance, hundreds of megaohms or few gigaohms
- Low output resistance, less than few ohms or fraction of one ohm
- Low input offset voltage, a few millivolts or micro-ohms
- Low input bias current, a few picoamps
- Very high open loop gain, 10^4 to 10^6
- Low intrinsic noise
- High common mode rejection ratio (CMRR)
- Low sensitivity to changes in power voltage
- A broad operating frequency range
- High environmental stability

Operational amplifiers can be configured as inverting or noninverting amplifiers. In addition, they can perform many other functions, for example as multipliers, adders, limiters, and filters.

Another version of op amps is the instrumentation amplifiers, which are essentially high-performance differential amplifiers that consist of several closed loops within the chip. Instrumentation amplifiers are used extensively in applications where sensor signals are extremely weak and noisy. An instrumentation amplifier has improved CMRR (up to 160 dB), high input impedances (e.g., 500 MΩ), low output impedance, low offset currents and voltages, and better temperature characteristics than common op amps do.

ANALOG SIGNAL PROCESSING

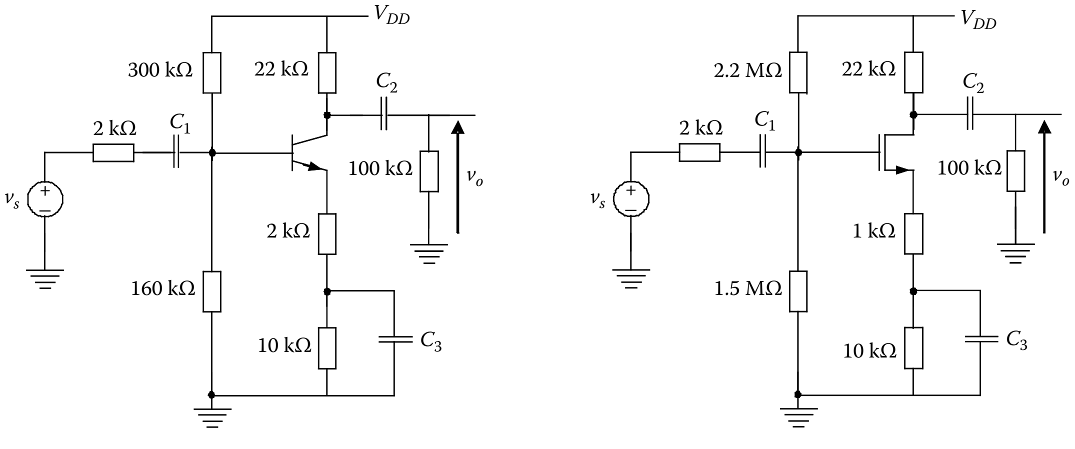
A small number of the most essential functional blocks necessary for analog signal processing, such as amplifiers, adders, integrators, filters, modulators, and demodulators, will be explained next.

Amplifiers An amplifier is a device that provides amplification of the input signal in terms of voltage, current, or power. Many different types of amplifiers exist, including voltage, power, current, transconductance, and transresistance amplifiers, and, in terms of frequency range, audio (15 Hz to 20 kHz), radio frequency (RF = 20 kHz to 0.1 GHz), and video (10 Hz to 6 MHz) amplifiers. As an example, Figure 1.1b illustrates two single-transistor amplifiers, bipolar and MOSFET (metal oxide semiconductor field effect transistor). See Figure 1.1b.

Adders An adder (summer) is a circuit, based on op amps, whose output is the weighted algebraic sum of all its inputs. In the case of the voltage adder, in Figure 1.1c, inputs V_1, V_2, and V_3 are added to give an inverted output, that is:

$$V_0 = -\frac{R_0}{R_1}V_1 - \frac{R_0}{R_1}V_2 - \frac{R_0}{R_1}V_3 \qquad \mathbf{1.1(8)}$$

More inputs can be connected without affecting the response of the existing inputs since the inverting input is held at a virtual ground by the feedback mechanism.



(a) (b)

FIG. 1.1b
Single-transistor amplifiers: (a) bipolar common-emitter, (b) MOSFET common source.

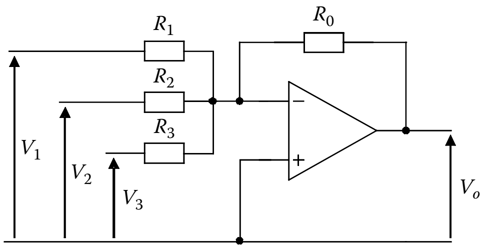


FIG. 1.1c
An adder with three inputs.

Integrators An integrator can be defined as a circuit whose output rate of change is proportional to the applied input. The output of an integrator for an input V_i is:

$$\frac{dV_0}{dt} = kV_i \quad \text{or} \quad V_0 = k\int V_i dt \qquad \mathbf{1.1(9)}$$

This equation shows that the output V_o is obtained by integration of the input V_i. Figure 1.1d illustrates an integrator based on an op amp. The integrator is an important building block in the design and implementation of active filters.

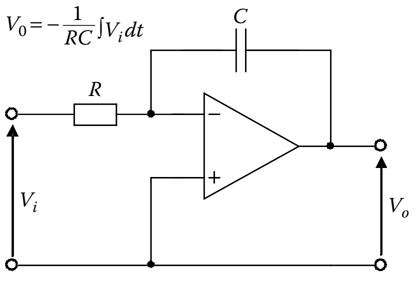


FIG. 1.1d
An integrator.

Differentiators can be obtained by changing the positions of R and C.

Filters Analog filters are essential parts of both analog and digital instruments. They are used to extract the useful information from the sensors, as most sensors generate broadband signals containing various forms of noise. Filters are useful for the elimination of external and internal noise that interferes with operation of an instrument.

There are two basic types of analog filters, passive and active. Passive filters are made from purely passive components, such as resistors, capacitors, and inductors. Active filters contain active elements such as op amps in addition to the passive components. The filters can be configured to be low-pass, high-pass, band-pass and band-stop, or notch filters. The subject of filters is an important but vast topic; therefore, some limited detail is provided below.

Low-Pass Active Filters A typical example of a low-pass filter is given in Figure 1.1e. In the process of an active low-pass filter design, the cutoff frequency of the filter must be determined first. The cutoff frequency is the frequency at which the gain of the filter is about –3 dB.

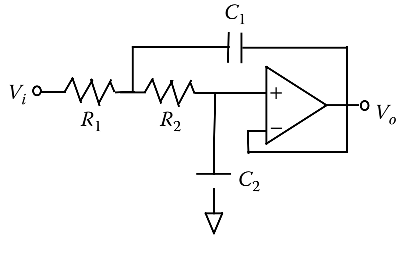


FIG. 1.1e
A low-pass filter.

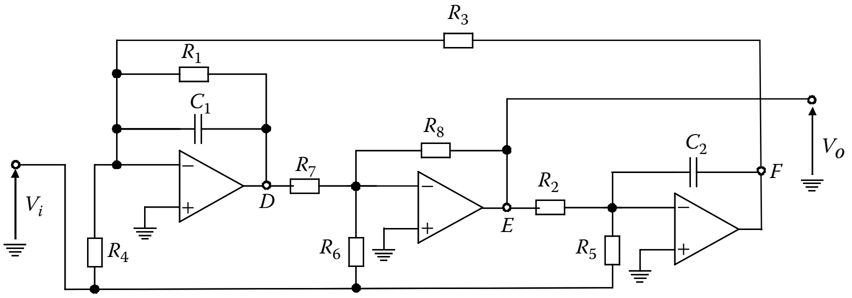


FIG. 1.1f
A three-amplifier biquad filter.

The choice of resistors in filters is critical. As a rule of thumb, resistors R_1 and R_2 are selected as 25 kΩ for a 10 kHz or 250 kΩ for 100 Hz filter. The values of capacitors can be calculated for the selected frequency by substituting the selected resistance values. The calculations are repeated to obtain the nearest available capacitors by changing the resistor values.

High-Pass Active Filters High-pass active filters are analogous to low-pass filters and pass only high frequencies.

Band-Pass or Band-Stop Active Filters A straightforward way of building a band-pass or a band-stop filter is by combining low-pass and high-pass filters in cascaded form.

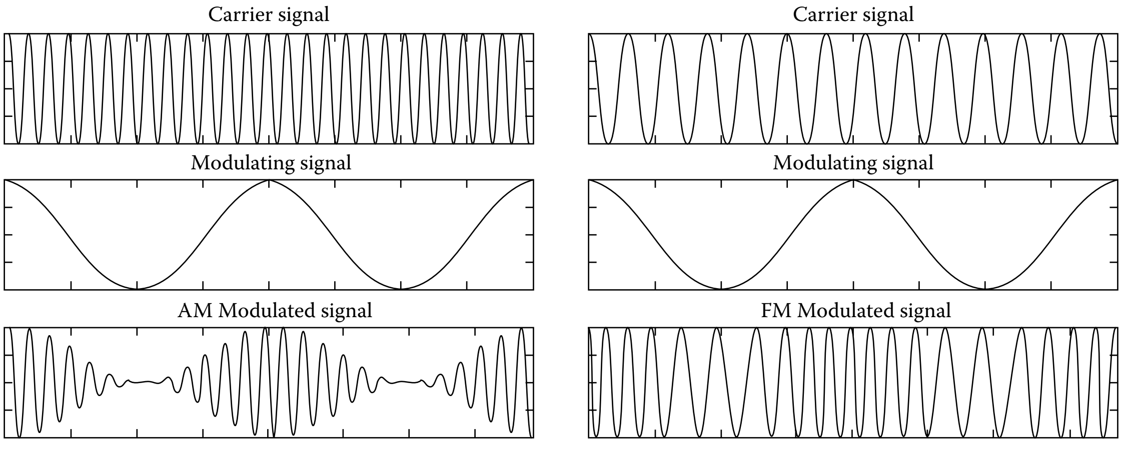
Notch Filters These filters are usually used to suppress single sharp frequencies such as 50 Hz. The best performance is obtained by adjusting component values while monitoring the circuit performance on an oscilloscope.

Various design techniques are used to configure analog filters. Some of these design techniques include Butterworth, Chebyshev, and Bessel–Thomson filters. The complexity of filters increases as the order of the filter (number of poles) increases. A typical complex active filter is illustrated in Figure1.1f.

Modulators and Demodulators In many electronic systems, modulations and demodulations are essential for allowing convenient and efficient transmission and reception of information. Basically, *modulation* is the process of encoding information on a carrier signal such that its amplitude, frequency, or phase varies with the message signal, known as the modulating signal. There are three types of modulation: amplitude modulation (AM), frequency modulation (FM), and phase modulation. Examples of AM and FM are illustrated in Figure 1.1g. *Demodulation* is the process by which information is extracted from the carrier signal.

APPLICATION AND EXAMPLES OF ANALOG INSTRUMENTS

Analog instruments are used in many applications where digital readouts are required. Examples of such applications are control panels, testing systems, manufacturing and production equipment, dashboards in all types of transportation



(a)

(b)

FIG. 1.1g
Two types of modulation, (a) AM and (b) FM.

FIG. 1.1h
An analog multimeter. (Courtesy of Hioki Corporation, www.hiokiusa.com.)

equipment, laboratories, explosion-proof areas, consumer products, and machinery. A typical example of analog instruments is the multimeter, as illustrated in Figure 1.1h.

DIGITAL INSTRUMENTS

In most modern instruments, the original analog information acquired from a physical variable is converted to digital form to obtain a digital instrument. Analog-to-digital (A/D) converters are used for signal conversion together with appropriate multiplexers and sample-and-hold devices. The multiplexers enable the connection of many sensors and transducers to the same signal-processing media. The typical building blocks of a digital instrument are illustrated in Figure 1.1i.

Digital systems are particularly useful in performing mathematical operations, numeric displays, and storing and transmitting information. Digital instruments consist of three essential subsystems: converters, general-purpose processors together with mass storage and communication peripherals, and application-specific software. In this section, some essential features of digital instruments will be discussed, including signal conversion, intelligent and IC sensors, basic hardware, inputs and outputs, communications and networks, virtual instruments, and software.

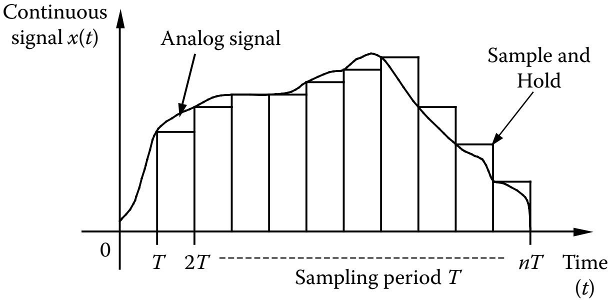


FIG. 1.1j
Analog-to-digital conversion process.

Signal Conversion

Analog-to-Digital Conversion Analog-to-digital (A/D) and digital-to-analog (D/A) converters are used to interface the analog environment (the real world) and the digital environment. Analog signals obtained from a myriad of real-world sources, including temperature, motion, sound, and electrical signals, are converted to digital forms by A/D converters. After the conversion, the signals are processed using many different algorithms such as digital filters and statistical techniques. When necessary, the processed signals then can be converted back to analog forms to drive devices requiring analog inputs such as mechanical loads and sound speakers.

A/D conversion involves three stages: sampling, quantization, and encoding. Once the analog signals are sampled, quantization takes place to determine the resolution of the sampled signals. In coding, the quantized values are converted to binary numbers. Figure 1.1j illustrates a typical A/D conversion process of an analog signal.

All modern A/D converters produce digital output values in the integer format. These can be in binary codes or the twos-complementary form. The length of the output word defines theoretical resolution of the A/D converter and the ranges that an A/D converter can produce. The smallest change in the input voltage (V_{LSB}) that A/D can convert is:

$$V_{LSB} = \frac{V_{\max} - V_{\min}}{2^n} \qquad \textbf{1.1(10)}$$

where n is the number of bits produced by the A/D converter and $V_{\max}$ and $V_{\min}$ are the maximum and minimum input voltages that an A/D converter can handle correctly.

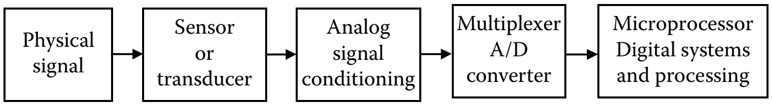


FIG. 1.1i
Basic building blocks of a digital instrument.

Due to a diverse range of application requirements, many different types of A/D converters exist. The most popular are counter ramp, successive approximation (serial), flash (parallel), and delta–sigma (Δ/Σ) converters.

Quantization A number of quantization methods exist, one of which is *amplitude quantization*, defined as the process by which a signal with a continuous range of amplitudes is transformed into a signal with discrete amplitudes. In this method, the process is usually memoryless; that is, the quantizer does not depend on previous sample values to predict the next value. Although this is not an ideal solution, it is commonly used in practice.

The precision of the quantizer is limited by the number of discrete amplitude levels available to approximate the continuous signal amplitudes. These levels are referred to as representation or reconstruction levels, and the spacing between two representation levels is called the *step size.* Quantization, like sampling, introduces some degree of uncertainty in the value of the analog signal. For A/D converters with large number of bits (8 bits or more) this effect can modeled as an additive noise. The amount of quantization noise relative to the signal, or Signal-to-Noise Ratio (SNR), depends on the nature and magnitude of the input signal. Quantization noise can be both nonlinear and signal dependent. However, provided that the step size is sufficiently small, it is reasonable to assume that the quantization error is a uniformly distributed random variable and its effect is similar to that of thermal noise.

In addition to quantization noise, A/D converters have other sources of noise, such as internal voltage reference noise and thermal amplifier noise. In practice, most converters rarely achieve their theoretical noise floor. Therefore, when selecting an A/D converter, one uses the number of bits of converter as a rough estimate of its resolution. Nevertheless, in practical applications, a close examination of the converter specifications should reveal its true performance.

Coding Coding is a process of representing the finite output states of the quantizer by a sequence of *n* bits. The codes formed using bits are normally known as *binary codes* and can be of two types: unipolar and bipolar.

Unipolar codes are used to represent unipolar quantities, that is, quantities with a predefined sign (positive or negative). The most common unipolar codes are *natural binary, BCD* (Binary-Coded Decimal), and the *Gray Code.* In order to illustrate the coding process, the natural binary code is described below.

In the natural binary code, each bit of a number represents a successive power of two according to the position that the bit occupies. Given a number of *n*-digits (represented by D_{n-1}, D_{n-2}, … D_1 and D_0), the value of the number is determined by the following expression:

$$D_{n-1}D_{n-2}\dots D_1D_0 = D_{n-1}\times 2^{n-1} + D_{n-2}\times 2^{n-2} + \cdots + D_1\times 2^1 + D_0\times 2^0 \qquad \mathbf{1.1(11)}$$

D_{n-1} is the Most Significant Bit (MSB), and D_0 is the Least Significant Bit (LSB).

However, the output code of an A/D converter is normally interpreted as a fraction of its Full Scale (FS) value, which is considered as the unit. So the binary codes from an A/D conversion process must be interpreted in fractional numbers, which is equivalent to dividing the value of each bit by 2^n.

Digital-to-Analog Converters Analog signal generation from the digital world may be part of the measurement process or may be required for control or sensor stimulation purposes. Digital signals are converted to analog waveforms by D/A converters and must be filtered and amplified before being applied to the appropriate output transducers. However, some new types of instrumentation now deal exclusively with digital signals. These include logic and computer network analyzers, logic pattern generators, and multimedia and digital audio devices. Such instruments usually include high-speed digital interfaces to move data in and out of the instrument at high speeds.

Digital information requires digital-to-analog converters (D/A) for the conversion of the binary words into appropriately proportional analog voltages. A typical D/A converter consists of three major elements:

1. Resistive or capacitive networks
2. Switching circuitry, operational amplifiers, and latches
3. A reference voltage

Also, an important requirement is the switching of the reference voltage from positive to negative for bipolar signals.

Intelligent and IC Sensors

In recent years, progress in electronic instruments has been making a major turnaround due to the availability of IC sensors in the form of micro- and nanosensors. Most of these sensors are smart sensors that have intelligence due to integration of complex digital processors into the same chip. A general scheme of a smart sensor is illustrated in Figure 1.1k. In this particular example, the sensor is under a microprocessor control. Excitation is produced and modified depending on the required operational range. The processor may contain

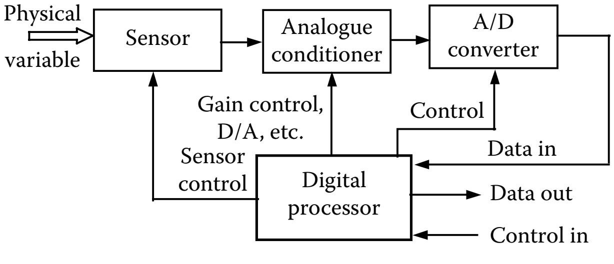


FIG. 1.1k
General form of smart sensors.

full details of the transducer characteristics in ROM (read only memory) enabling the correct excitation, gain, and so on.

Basic Hardware

Microprocessors and micro-controllers constitute the heart of almost all types of modern digital instruments and instrumentation systems. They play an important role in data acquisitions, data processing, and control. The applications of microprocessors and micro-controllers in instruments and instrumentation systems can be categorized according to the following roles:

1. *Data handling* functions that include data acquisition, data processing, information extraction, data compression, interpretation, recording, storage, and communication
2. *Instrumentation control,* which includes sensors, actuators, system resources, and process controls
3. *Human-machine interface,* one of the significant roles of digital systems, to provide a meaningful and flexible user interface to the operators for ergonomic information and control display
4. *Experimentation and procedural development,* consisting of commissioning, testing, and general prototyping of the targeted system under investigation

Micro-controllers and digital signal processors are special microprocessors that have built-in memory and interface circuits within the same IC. Due to smaller sizes, simplicity and low cost, and power efficiency, micro-controllers are commonly used in electronic instruments and the associated instrumentation systems. Therefore, in this section, detailed information on micro-controllers will be provided.

Micro-Controllers Many micro-controllers are single-chip devices that have memory for storing information and are able to control read/write functions and manipulating data. The performance of a micro-controller is classified by its size, that is, the number of bits that the CPU (central processing unit) can handle at a time. Many types or families of micro-controllers are offered by a diverse range of manufacturers. Therefore, in order to select appropriate micro-controllers for a specific task, it is vital to understand the differences in the characteristics of different families.

The software for micro-controllers to perform specific tasks can be written in either assembly codes or high-level languages such as Basic or C. The program written in other digital devices can be downloaded into the system memory through the serial communication port of the micro-controller. All high-level programs are compiled into machine language for execution. Significant numbers of compilers are available on the market for the different families of micro-controllers.

Digital Signal Processors Digital signal processors (DSPs) are specialized microprocessors that involve partial computer architecture and fulfill fast operational needs. A simplified block diagram of a typical DSP processor is given in Figure 1.1l.

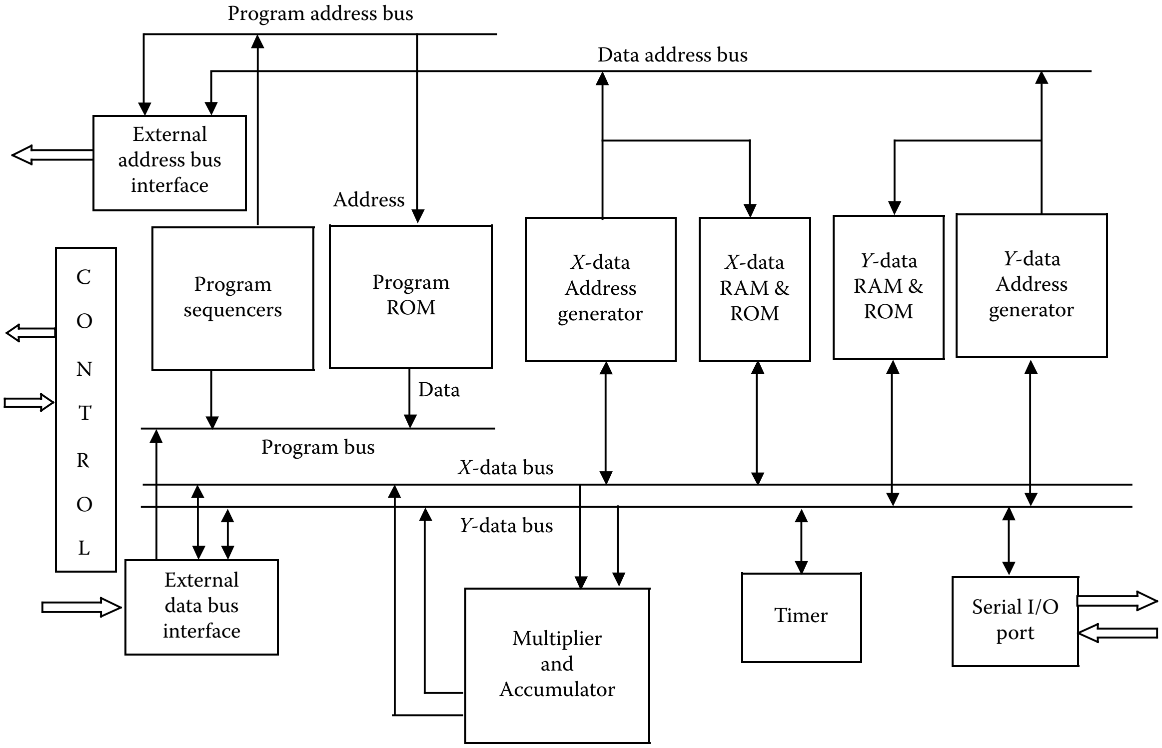


FIG. 1.1l
Block diagram of a typical DSP.

DSPs incorporate special hardware features that are capable of speeding up calculation of digital filters, fast Fourier transforms (FFTs), and other frequency-based algorithms. Modern general-purpose microcomputers can also address the needs of digital signal processing if some of the necessary hardware and special instructions are added. As a result, the distinction between the DSP processor and microprocessors, in general, is in the degree of specialization. This particular DSP has two data memories, denoted as X and Y (Figure 1.11). For the implementation of signal processing, say in a finite impulse response (FIR) filter, X-memory can be used to store the samples of input signals and Y-memory to store the impulse response.

One of the main characteristics of the DSP processors is the capability of handling integers for which they are designed. Most of the existing DSP processors are either 16- or 32-bit devices. For example, the Motorola DSP56300 processors are 24-bit integer processors that offer a compromise between inexpensive 16-bit and powerful 32-bit devices.

Inputs and Outputs

The data communication between micro-controller and outside world is realized through input and output (I/O) pins. The format of data transmission of a micro-controller can be universal asynchronous receiver transmitter (UART) — a port adapter for asynchronous serial communications. A diverse range of I/O hardware and software is available. Some of these I/Os suitable for instruments and instrumentation systems are explained below:

1. *Universal Synchronous/Asynchronous Receiver Transmitter (USART)*—a serial port adapter for either asynchronous or synchronous serial communications. Communications using a USART are typically much faster than those using UARTs.
2. *Synchronous serial port*—Synchronous serial ports are used to communicate with high-speed devices such as memory servers, display drivers, additional A/D ports, etc. In addition, they are suitable to implement simple micro-controller networks.
3. *Serial Peripheral Interface (SPI)* (Motorola) — A synchronous serial port functioning as an enhanced UART.
4. *I^2C bus—Inter-Integrated Circuit bus* (Philips)—is a simple two-wire serial interface that was originally developed for 8-bit applications. The I^2C bus is a two-line, multi-master, multi-slave network interface with collision detection. Up to 128 devices can be connected on the network, and they can be spread out over 10 m. The two lines of the network consist of the serial data line and the serial clock line. Each node on the network has a unique address, which accompanies any message passed between nodes. Since only two wires are needed, it is easy to interconnect a number of devices.
5. *Controller Area Network (CAN)* is a multiplexed wiring scheme that was developed jointly by Bosch and Intel for wiring in automobiles. The CAN specification is being used in industrial control in both North America and Europe.

Data-Acquisition Boards A data-acquisition board is a plug-in I/O board that allows the treatment of a set of analog signals by a computer. Thus, these boards are the key elements to connect a computer to a process in order to measure or control it. Data-acquisition boards normally operate on conditioned signals, that is, signals which have already been filtered and amplified by analog circuits. However, some data-acquisition boards can deal with the signals directly generated by sensors. These boards include all the necessary signal-conditioning circuits.

The number of analog inputs normally managed by a data-acquisition board is eight or 16. This system is made up of an analog multiplexer, an amplifier with programmable gain, a sample-and-hold circuit, and an A/D. All the elements of this system are shared by all the analog input channels. Some data-acquisition boards also offer analog outputs to carry out control operations. When available, the normal number of analog outputs is two.

In addition to analog inputs and outputs, data-acquisition boards also supply digital I/O lines and counters. Digital lines are used for process control and communication with peripheral devices. Counters are used for applications such as counting the number of times an event occurs or generating a time base.

Communications and Networks

In many applications, instruments are networked by computers to measure many variables of a physical process. The resulting arrangement for performing the overall measurement function is called the *measurement system.* In measurement systems, instruments operate autonomously but in a coordinated manner. The information generated by each device is communicated between the instruments themselves or between the instrument and other devices such as recorders, display units, base stations, or a host computer. Currently, the coordination of instruments is largely realized by digital techniques.

In digital instrumentation systems, the transmission of data between devices is realized relatively easily by using serial or parallel transmission techniques. However, as the measurement system becomes large by the inclusion of many instruments, communication can become complex. To avoid this complexity, message interchange standards are used, such as RS-232, USB, EIA-485, and IEEE-488.

For long-range data transmission, modem, microwave, or radio frequency (RF) media are selected. In these cases, various modulation techniques are used to convert digital signals to suitable formats. For example, most modems use frequency-shift keyed (FSK) modulation. The digital interface with modems uses various protocols such as MIL-STD-188C to transmit signals in simplex, half-duplex, or full duplex forms depending on the directions of the data flow. The simplex interface transmits data in one direction, whereas full duplex transmits it in two directions simultaneously.

In industrial applications, several standards for digital data transmission are available. These are commonly known in engineering literature as *fieldbuses*. Among many others, some of these standards, such as the WordFIP, Profibus, and Foundation Fieldbus, are widely accepted and used. The fieldbuses are supported by hardware and software (for example, National Instruments chips and boards) that allow increases in the data rates suitable with high-speed protocols.

Virtual Instruments and Software

Traditional instrumentation systems are made up of multiple stand-alone instruments that are interconnected to carry out a determined measurement or control an operation. Functionality of all these stand-alone instruments can be implemented in a digital environment by using computers, plug-in data-acquisition boards, and supporting software implementing the functions of the system. The plug-in data-acquisition boards enable the interface of analog signals to computers, and the software allows programming of the computer to look and function as an instrument. The systems implemented in this way are referred to as *virtual instrumentation systems.*

The major advantage of virtual instrumentation is its flexibility; changing a function simply requires modification of supporting software. However, the same change in a traditional system may require to adding or substituting a stand-alone instrument, which is more difficult and expensive. For example, a new analysis function, such as a Fourier analysis, can be easily added to a virtual instrumentation system by adding the corresponding software to the analysis program. However, to do this with a traditional instrumentation system, a new stand-alone instrument (for example, spectrum analyzer) would have to be added to the system.

Virtual instruments also offer advantages in displaying and storing information. Computer displays can show more colors and allow users to quickly change the format of displaying the data that is received by the instrument. Virtual displays can be programmed to resemble familiar instrument panel components, including buttons and dials. In Figure 1.1m, an example of a virtual display corresponding to a two-channel oscilloscope is shown. Computers also have more mass storage than stand-alone instruments do; consequently, virtual instruments offer more flexibility in data handling.

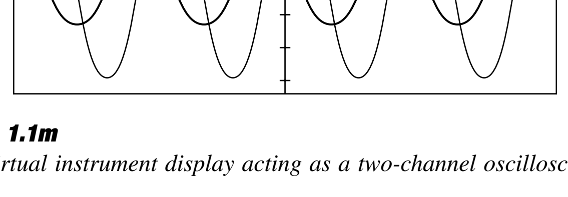


FIG. 1.1m
A virtual instrument display acting as a two-channel oscilloscope.

Software for Virtual Instruments To develop the software of a virtual instrumentation system, a programming language or special software can be used. However, the option of using a traditional programming language (C, for example) can generate several problems, including difficulty in programming the graphical elements and difficulty in learning the language.

Currently, a more utilized option is the Microsoft Visual Basic programming language, which runs under the Windows operating system. Visual Basic has become quite popular in the development of virtual instrumentation systems because it combines the simplicity of the Basic programming language with the graphical capabilities of Windows. Another important reason for its popularity is that it is an open language, meaning that it is easy for third-party vendors to develop products that engineers can use with their Visual Basic programs. Several companies now sell Dynamic Link Libraries (DLLs), which add custom controls, displays, and analysis functions to Visual Basic. The controls and displays mimic similar functions found on stand-alone instruments, such as toggle switches, slide controls, meters, and LEDs. By combining various controls and displays, any virtual instrument can easily be programmed.

A third option for developing virtual instrumentation systems is to use a software package specifically designed for the development of these systems, such as LabView or PC600. The crucial advantage of these packages is that it is not necessary to be a Windows programmer to use them. Several packages of this type exist, but only one, LabView by National Instruments, has reached great diffusion.

LabView is an entirely graphical language based on two concepts: the *front panel* and the *block diagram.* It is extendible, so new functional modules can be added to a program. The modules can also be written using general-purpose languages, such as C or C++. These languages provide great flexibility to program functions that perform complex numerical operations on the data.

The applications of virtual instruments are gaining momentum in instrumentation systems. Nowadays, palm and laptop computers are widely available and they can easily be equipped with commercial data-acquisition boards such as the PCMCIA interface cards. Together with a suitable interface, software, and sensors, these computers function just as any digital instrument would.

Application and Examples of Digital Instruments

Digital instruments have all the functionality of their analog counterparts and are used in all types of instrumentation systems. Modern digital instruments are rapidly replacing analog instruments due to advantages in communications, data handling, and networking capailities, as will be explained in the next section.

An example of digital instrumentation is a blood pressure recorder "Comparison of Analog and Digital Instruments,"

FIG. 1.1n
A digital ambulatory blood pressure recorder. (Courtesy of SunTech Medical, www.suntechmed.com.)

illustrated in Figure 1.1n. This wireless outdoor system operates on solar power. It measures temperature, humidity, pressure, wind speed and direction, and rainfall. A wireless link allows the transmission of data up to 48 km. Several instruments can be networked together.

COMPARISON OF ANALOG AND DIGITAL INSTRUMENTS

Digital hardware allows programmable operation; therefore, it is possible to modify the functions of an instrument through the software, which increases flexibility and reliability. However, it is worth mentioning that some signals with extremely wide bandwidths require intensive real-time processing in the digital environment. In these cases, analog instruments may be the only solution.

Many modern digital instruments have built-in intelligence; hence, they can make self calibrations, self diagnosis, and corrections in acquired data. A new trend is that calibrations of instruments can be conducted over the Internet by using appropriate sites of manufacturers and standards authorities.

Digital instruments are suitable in applications where multiple measurements are necessary, as in the case of process industry, for two main reasons:

1. Ease in communication and networking by means of remote communication methods such as RF, microwave, the Internet, and optical techniques
2. Availability of almost infinite memory for data storage

In digital instruments, however, converting an analog signal allows certain errors (for example, quantization error) to be introduced. The extent of error depends on the sampling rates and number of bits of the A/D converter. By the same token, digital instruments minimize the reading error that can occur in analog displays. Errors such as parallax and shadowing are also eliminated. Digital instruments also reduce the margin for error due to inaccurate interpolation. Analog meters can be very sensitive to movement or rough handling, while digital meters are immune to such dynamic effects.

One of the distinct advantages of analog instruments is the easy interpretation of displays. The operator can get an intuitive feel about the variables being measured at a glance. There is no such spatial reference with digital displays, which require mental interpretation. This requires an additional step in the thought process and also some familiarity with the equipment. This additional step of interpretation may not be tolerable in some applications, such as driving cars, where instantaneous decisions may need to be made. In addition, analog instruments are relatively cheaper for the same functionality due to their simplicity.

Bibliography

Brignell, J., and White, N., *Intelligent Sensor Systems,* Rev. ed., Philadelphia: Institute of Physics Pub., 1996.

Cavicchi, T. J., *Digital Signal Processing,* New York: John Wiley & Sons, 2000.

Dyer, S. A. (Ed.), *Survey of Instrumentation and Measurement,* New York: John Wiley & Sons, 2001.

Eren, H., "Analog and Discrete Input/Output, Costs, and Signal Processing," *Instrumentation Engineers Handbook,* 4th ed., B. Liptak, Ed., Boca Raton, FL: CRC Press, 2002, pp. 123–141.

Fowler, K.R., *Electronic Instrument Design—Architecting for the Life Cycle,* New York: Oxford University Press, 1996.

Garrett, P. H., *Advanced Instrumentation and Computer I/O Design: Real-Time System Computer Interface Engineering,* New York: IEEE Press, 1994.

Grover, D., and Deller, R. J., *Digital Signal Processing and the Microcontroller,* Upper Saddle River, NJ: Prentice Hall PTR, 1999.

Haykin, S., and Kosko, B. (Eds.), *Intelligent Signal Processing,* New York: IEEE Press, 2001.

IEEE, *Spectrum,* Piscataway, NJ: IEEE, 1990–2002.

IEEE, *Transactions on Instrumentation and Measurement,* Piscataway, NJ: IEEE, 1990–2002.

Liptak, B., *Instrumentation Engineers Handbook,* 4th ed., Boca Raton, FL: CRC Press, 2002.

Madisetti, V. K., and Williams, D. B., *The Digital Signal Processing Handbook*, Boca Raton, FL: CRC Press, 1998.

Pallás-Areny, R., and Webster, J. G., *Sensors and Signal Conditioning,* 2nd ed., New York: John Wiley & Sons, 2001.

Pallás-Areny, R., and Webster, J. G., *Analogue Signal Processing,* New York: John Wiley & Sons, 1999.

Paton, B. E., *Sensors, Transducers, and LabView,* Upper Saddle River, NJ: Prentice Hall PTR, 1999.

Potter, D., "Smart Plug and Play Sensors," *IEEE Instrumentation and Measurement Magazine,* Vol. 5(1), pp. 28–30, 2002.

Predko, M., *Handbook of Microcontrollers,* New York: McGraw-Hill, 1999.

Randy. F., *Understanding Smart Sensors,* 2nd ed., Boston: Artech House, 2000.

Soloman, S., *Sensors Handbook,* New York: McGraw-Hill, 1999.

Webster, J. G. (Ed.), *Wiley Encyclopedia of Electrical and Electronic Engineering,* New York: John Wiley & Sons, 1999.

Webster, J. G. (Ed.), *The Measurement, Instrumentation, and Sensors Handbook,* Boca Raton, FL: CRC Press, 1999.

1.2 Computer Configurations of Supervisory Units

J. W. LANE, C. S. LIN, T. H. TSAI (1995) **S. RENGANATHAN** (2005)

Types of Software Packages:	A. Control system design B. Process monitoring and control C. Process optimization D. Statistical process control
Partial List of Software Suppliers:	Allen-Bradley (B) (www.ab.com) DMC (C) (www.dmc.com) Gensym Corp. (D) (www.gensym.com) Icom (A) (www.icom.net.au) Iconics (B) (www.iconics.com) Setpoint (C) (www.setpointusa.com) T/A Engineering (B) (www.ta-eng.com) U.S. Data (B) (www.usdata.com) Wonderware (B) (www.wonderware.com)

INTRODUCTION

The first use of digital computers was for acquiring data from the plant and logging that data without influencing plant operations. These systems were called data acquisition systems (DAS).

Supervisory control evolved, as the information collected by DAS was processed in order to generate optimal set points for the individual control loops. These set points were first sent to existing analog controllers and later to the digital control algorithms within DCS systems. This method of operation is called supervisory control, which these days often also incorporates optimization functions.

The discussion in this section will first cover the design when supervisory computers are operating with controllers, followed by a description of supervisory computers in combination with DCS systems.

HISTORY OF SUPERVISORY CONTROL

In 1957, the Ramo–Woodridge Corporation introduced the RW 300, the first solid-state digital computer designed for process control, at the National Meeting of the American Institute of Chemical Engineers in Los Angeles. The computer was designed based upon a magnetic drum containing 8064 words of 20-bit length. The external input/output supported 540 one-bit digital inputs, 1024 analog inputs, 540 digital outputs, and 36 analog outputs.

The RW 300 was designed to interface directly with the existing analog control system of a process. The role of this system was an enhancement to the existing process control instrumentation, and it functioned as a supervisory control device.

The software program consisted of models of the process in varying degrees of sophistication, which were updated based upon real-time data and then perturbed using optimization or maximization routines. These were usually simple linear programming algorithm perturbation routines. The models provided new sets of operating conditions (set points), which would improve the process performance according to some performance criteria. These new conditions were then outputted to the process. Essentially, the process was moved from one set of steady-state conditions to another in response to changes in such uncontrolled disturbances as feed characteristics, ambient conditions, heat transfer coefficients, and catalyst activity.

In spite of the seemingly small memory capacity of the RW 300 as compared to today's standards, the RW 300 was successfully installed in a number of sophisticated chemical and refinery processes. The very first process control system to be successfully installed was on the Texaco phosphoric acid polymerization unit at Port Arthur, Texas.

Within the next few years, several companies introduced their versions of the process control computer. These companies included IBM, General Electric, Foxboro, Control Data Corporation, and many others. The first process control computer to perform dynamic regulatory control of a process was installed at the Riverside Cement Company's Victorville plant in the United States and controlled two cement kilns and raw feed blending.

The first improvement in the field of the process control computer was an increase in magnetic drum capacity. Memory size first doubled to 16K, then 32K, and then kept rising. The next improvement in the process control computer design was the introduction of high-speed core memory, which increased the computational speed from milliseconds to microseconds. A giant leap in technology came with the replacement of solid-state circuitry (diodes, resistors, and transistors) with silicon integrated circuits. The new technology offered greatly increased speed, capacity, and most of all, reliability.

Another chapter in the evolution of the process control computer came with the development of direct digital control (DDC), wherein the function of the analog instrumentation was incorporated in the computer, and the analog controllers were thereby eliminated. The motivation was to reduce the control system cost and increase design flexibility for new processes. This new design concept initially met with mixed reviews, primarily because the computers—especially the software, which had become complicated—were not sufficiently reliable, and in most cases, when the computer went down, so did the process.

However, as minicomputers began to replace the old process control computers, it became economically feasible to provide redundancy by installing dual computers. This was frequently done. Then a major advancement came in 1976 when Honeywell announced the first distributed digital control system (DCS), named the TDC 2000. The hallmark of the system was reliability based upon redundancy in microprocessor-based controllers, redundancy in communications, and redundancy in operator interface.

Generic Features of DCS

Currently, most large-scale process plants such as oil refineries, petrochemical complexes, and various other processing plants are controlled by microcomputer-based DCS (distributed control systems). As is discussed in detail in Chapter 4, these systems generally include the following features:

1. Cathode ray tube (CRT)-based operator consoles and keyboards, which are used by plant operators or engineers to monitor and control the process
2. Controllers, multifunction control modules, and programmable logic controllers (PLCs), which provide the basic control computation or operation
3. A communication network, which is used to transfer the information between control modules and operator consoles across the node on the network
4. I/O (Input/Output) modules, which are used to convert the field instrumentation signals from analog to digital and digital to analog form for controller modules and console displays
5. Fieldbus communication links, which are used for communication between remote I/O devices and control modules
6. Historical module, which is used for data storage for pertinent process data or control data and for online data retrieval or archiving
7. Computer interface, which is used for communication between the nodes on the DCS network and the supervisory computer
8. Software packages for real-time monitoring, control, reporting, graphics, and trending

The generic arrangement for these components is shown in Figure 1.2a. This arrangement is a natural outcome for

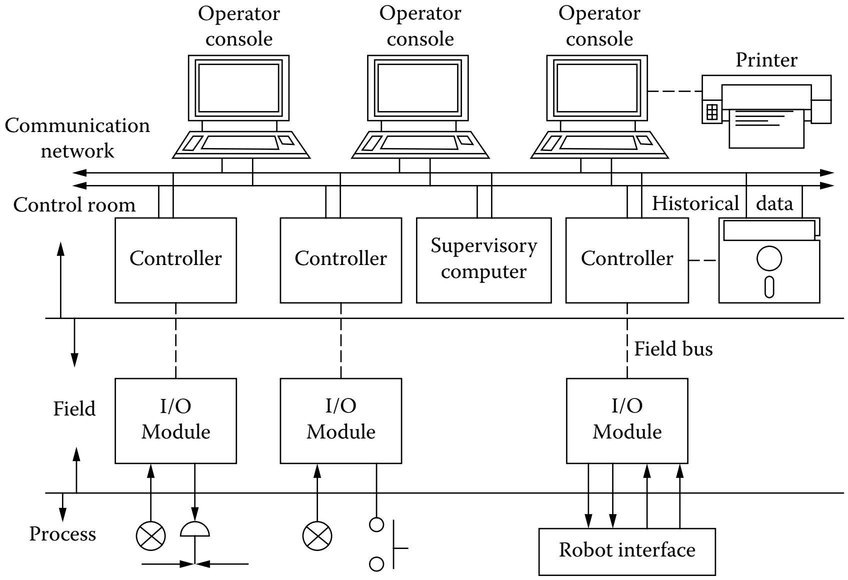


FIG. 1.2a
Generic features of DCS.

large-scale process control because the parallel operations of the control tasks are required for both continuous and batch process control. Therefore, the distributed control is a natural way of satisfying the parallel nature of the process from a geographical point of view.

DCS replaces conventional control systems to perform basic control. The supervisory control functions that in the past were performed by the computer systems described above, plus additional functions such as online information evaluation, are usually still performed by a supervisory computer, which is linked to a DCS.

COMPUTER-SUPERVISED ANALOG CONTROLLERS

The computer's capability to perform complex mathematical calculations and to make logical decisions provide a unique opportunity for improving the performance of any process through the application of supervisory control. All processes are affected by disturbances; some can be measured or calculated, others are not. Examples of disturbances are feed composition and concentration variations, ambient conditions, drifts in catalyst activities, heater and heat exchanger fouling, and economics.

Functions

Supervisory control has three fundamental functions:

1. Calculation of the present state of the process in terms of key process parameters such as yield, severity, and efficiency.
2. Calculation of critical process constraints, such as compressor capacity, column flooding limits, explosion limits, vacuum condenser capacity, and raw material.
3. Calculation of new process conditions, in terms of set points, which will meet the requirements of the objective function. The objective function can be one of a number of goals such as maximizing throughput, yield, and profit or minimizing deviation.

Tasks

The supervisory control computer requires:

1. A process model
2. Cost functions
3. Optimization algorithms
4. Constraints and limits

The supervisory computer typically performs the following types of tasks:

- Determines the process operating constraints, such as a column flooding condition in distillation or a surge condition of a compressor. Basic material balance, energy balance, or heat transfer calculations are utilized in creating the required process models. Because of the complex nature of the process itself, the computation of the model is a difficult task and might require iterative computation to satisfy convergent criteria.
- Determines the present operating state of the process based on the online, real-time information based on temperatures, pressures, and feed characteristics to obtain reactor yields or determines the desired state according to the constraints and optimization criteria.
- Determines the optimum control strategy based on the online, real-time information to achieve the control command by adjusting the manipulated variables or set points at the DCS level. Once such an algorithm is used in the supervisory computer, the response of the process to control strategy commands will gradually move this process from an initial state to a final desired state, avoiding the process constraints and following the optimum path such that one of the following objective functions is obtained: minimum cost, minimum energy, maximum yield, or maximum profit.
- Predicts impending alarms based on rigorous mathematical models using present and past history, process data, and control commands. Anticipating alarm conditions in advance of the process reaching these conditions is a vital function of supervisory control. Proper grouping of the pertinent process variables in a preclassified manner allows the severity of the alarms to be identified and allows actions to be pinpointed that can then be initiated quickly by operators.

Supervisory Computer for Analog Loops

The functions of a supervisory computer control system are shown in Figure 1.2b. It controls four process loops, but can be extended to as many loops as needed. Typically, 50 loops are assigned to one computer.

In this configuration, if analog controllers are being supervised, the proportional, integral and derivative (PID) settings of the analog controllers are kept intact. They are not changed. The process variables are measured, and their transmitted measurements are sent not only to the analog controller but also to the analog to digital (A/D) converter of the multiplexer.

These signals are processed by the computer, which takes into consideration the optimal control algorithm, the external constraints, and the management decisions, while generating the set points for the corresponding analog controllers. These set points are sent back to their corresponding controllers through the D/A converter of the demultiplexer. This operation is carried out for all the loops in a sequential, cyclic fashion on time-sharing basis.

Process Models

The model of a process can be a simple first-order linear one or a more complicated multivariable nonlinear model. It is essential to obtain a fairly accurate process model, and for

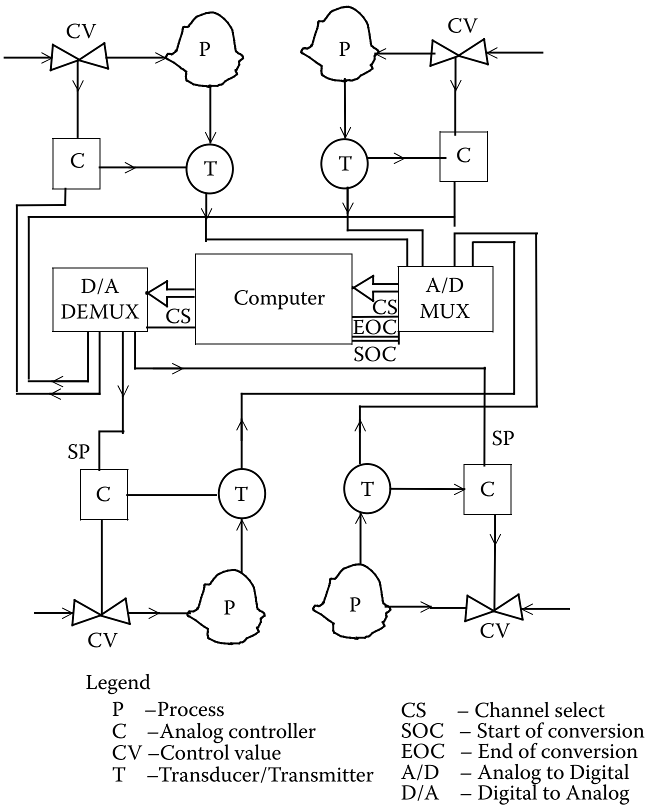


FIG. 1.2b
Block schematic of supervisory computer control for four loops.

many chemical processes such as distillation or boiler operation, obtaining an accurate process model is a difficult task. In some cases, particularly in old, existing plants, obtaining an accurate and complete model is not possible. When conventional methods of obtaining a model fail, heuristic models are advised.

Some of the heuristic models include the fuzzy logic–based model, the neural network–based model (see Section 2.18) and the genetic algorithm–based neural models. These models can be obtained by mapping a set of inputs and their corresponding outputs from the process. There are also neural or fuzzy controllers, which can generate their own optimal control algorithms. The behavior of these nonconventional models and controllers is similar to that of humans; these controllers are well suited for controlling complicated and nonlinear processes.

Conventional Models Process models can be obtained by conducting a process reaction test. Many higher-order processes have over-damped responses. Such a response is similar to the response of a first-order with dead time process. In other words, the higher-order process can often be approximated by models that are first order plus dead.

If there are oscillations in the system, then it can be approximated by an under-damped second-order + dead time model. The parameters of the dead time and the first- or second-order approximation can be experimentally determined.[1] The different methods available are impulse response methods, step response methods, and frequency response methods.[1]

For multivariable systems, if one can identify all the inputs and outputs and if they are measurable, then the least square method can be used to obtain a state model of the process. This will be an offline model or post-flight model. A number of experiments can be conducted for a sufficiently long time to obtain the input and output data, and those values are used in the least squares (LS) algorithm.[2,3] In systems where the parameters are likely to change with time, it is preferred to have online estimation of the system parameters. The recursive least square algorithm, discussed in References 2 and 3, can be used.

Optimization The main purpose of supervisory control is to obtain optimal performance of the overall process. When the supervisory computer works out and generates the set points for individual local controllers, coordinating the operation of the individual loops is an important task. The overall process performance will only be optimal if all the equipment constraints are taken into consideration and raw material availability, market needs, and management requirements are all satisfied.

When the supervisory computer generates the set points for individual analog controllers, it is important that these controllers be properly tuned. The tuning of PID and other controllers for various criteria are discussed in detail in Chapter 2 and are summarized in Table 1.2c.

In 1/4 decay ratio response, the oscillations of the controller output are such that for a small disturbance the second positive peak is one-fourth of the amplitude of the first positive peak. The PID controller settings are tuned to provide this output response. In the integral tuning criteria listed in Table 1.2c, the integral values are minimized.

COMPUTER-SUPERVISED DCS SYSTEMS

Most of the DCS software has evolved from the classical analog control. Therefore, many sophisticated supervisory control functions, such as calculating operating constraints for process units or determining heat transfer coefficient trends for maintaining heat exchangers, cannot be easily performed at the DCS level due to hardware or software restrictions. A review of supervisory control techniques and strategies is provided later in this section.

Production Monitoring and Control

Production monitoring and control for a given operation may include any or all of the following functions:

- Order entry/assignment
- Scheduling
- Production reporting

TABLE 1.2c
Criteria for Controller Tuning

1. Specified Decay Ratio, Usually $^1/_4$	$\text{Decay ratio} = \dfrac{\text{second peak overshoot}}{\text{first peak overshoot}}$
2. Minimum Integral of Square Error (ISE)	$\text{ISE} = \int_0^\infty [e(t)]^2 dt$, where $e(t)$ = (set point process output)
3. Minimum Integral of Absolute Error (IAE)	$\text{IAE} = \int_0^\infty \lvert e(t)\rvert\, dt$
4. Minimum Integral of Time and Absolute Error (ITAE)	$\text{ITAE} = \int_0^\infty \lvert e(t)\rvert\, t\, dt$

- Quality measurement and control by statistical process control or statistical quality control (SPC/SQC)
- Warehouse and shipping management
- Inventory management
- Customer service (order status, etc.)

Traditionally, these functions were either performed by manual operation or by the application of management information systems (without the availability of online data). As computers became more powerful and less costly, it became technically feasible to include many or all of the above functions in an integrated online information and control system that includes several computers linked by one or a number of sophisticated communication networks.

Because of the increasing pressure from worldwide competitions, these integrated systems, sometimes called CIM (Computer Integrated Manufacturing), are rapidly becoming a necessary tool for many major operations. One example is for a mature company to allocate each order to one of its three refineries in an optimum manner. In order to do this task correctly, the most current information must be available from all three refineries, including all processing unit loads and efficiencies, maintenance schedules, and inventories.

CIM systems are complicated and will not be discussed in detail in this section. However, it is important to note that a supervisory control computer may be only a small part of a network system, and "system integration" can be an important added task, demanding both technical and managerial attention.

Online Information System

An online information system may include any or all of the following functions:

- Data collection, checking, and verification
- Data reconciliation
- Data storage and retrieval

Many DCS systems include only limited online information functions, perhaps providing, for example, a history module as part of a standard DCS function. In most cases, such systems do not have sufficient capacity to meet users' needs, and a computer-based data archiving system is necessary.

Similarly, some DCS packages do not have sufficient storage capacity nor processing speed for processing a large quantity of data, and they normally do not support the desired database formats required by many users. In addition, many computer systems are supported by huge quantities of standard software packages, making the implementation of a real-time information system affordable. Many of the standard packages are not available in the DCS environments.

SUPERVISORY CONTROL TECHNIQUES

One of the most important jobs for an engineer responsible for implementing supervisory control is to make sure that the computer functions and DCS functions support each other.

Supervisory Control Algorithms

As was discussed earlier, before the DCS became popular in process industries, supervisory control was used to command single-loop analog controllers. These controllers served to achieve certain limited goals, such as to obtain a uniform temperature profile for a given multi-pass furnace or to determine the optimum blending for gasoline products.

The functions that the analog controllers could not accomplish were more-or-less delegated to the digital computer to perform. These included the necessary logic, sequence, or analytic computations to improve the process control operation.

From this experience, the practical value of writing the control algorithm in velocity form becomes clearer. Coordinated supervisory control with integrated feedback or feedforward combined in a velocity control algorithm can be activated or deactivated without creating perturbations to the process. The use of velocity algorithms in a computer-supervised DCS system is illustrated in Figure 1.2d and is explained below.

Advantage of Velocity Algorithm The velocity algorithm implemented in the supervisory computer can be converted to the position algorithm, which is usually applied at the DCS level. The control command, u, in the supervisory computer at the current sampling time, n, can be expressed as follows:

$$u(n) = \text{Sat}[u(n-1) + \Delta] \qquad \textbf{1.2(1)}$$

where Δ = incremental output from computer command; $u(n)$ = current computer position command at sampling instant n; $u(n-1)$ = past sampled position command; Sat(x) = upper limit if $x \geq$ upper limit, = lower limit if $x \leq$ lower limit, = x if lower limit $< x <$ upper limit.

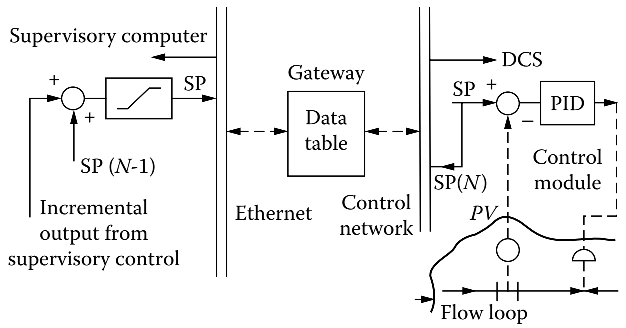


FIG. 1.2d
Velocity algorithm for supervisory control.

The incremental value can be the combination of feedforward or feedback as follows:

$$\Delta = \sum_i k_i \Delta fb_i = \sum_j k_j \Delta ff_j \qquad \mathbf{1.2(2)}$$

where Δfb_i = incremental value from the i^{th} feedback loop; Δff_j = incremental value from j^{th} feedforward loop; k_i = assignable constant for i^{th} loop; k_j = assignable constant for j^{th} loop.

The advantage of this algorithm for initialization of the control output command from the supervisory computer is that initialization can be obtained simply by setting Δ equal to 0 and making $u(n-1)$ equal the value obtained from DCS (for example, a set point value for a slave loop).

Antiwindup Control Strategy For supervisory control, the command from the supervisory computer to the set point for the control loop at the DCS level is essentially a form of cascade control. Caution is needed when control loops are in cascade. The upper or cascade master control command (supervisory computer command) requires information from the lower or secondary loop at the DCS level in order to command correctly. This information includes valve saturation status and lower loop current set point value at the DCS.

This information, which resides at the DCS level, requires constant monitoring by the supervisory computer of the real-time information communicated through interface modules on the DCS network. It is very important to prevent the computer command from causing set point windup at the DCS level. The block diagram for antiwindup is shown in Figure 1.2e.

This algorithm protects the upper loop command from windup. This protection is accomplished by freezing the supervisory control command output to the DCS set point at its last value if the DCS output to the valve is saturated and if the direction of the upper loop output increment would cause further valve saturation at the DCS level. Otherwise the upper loop within the supervisory computer will not be frozen at its last value.

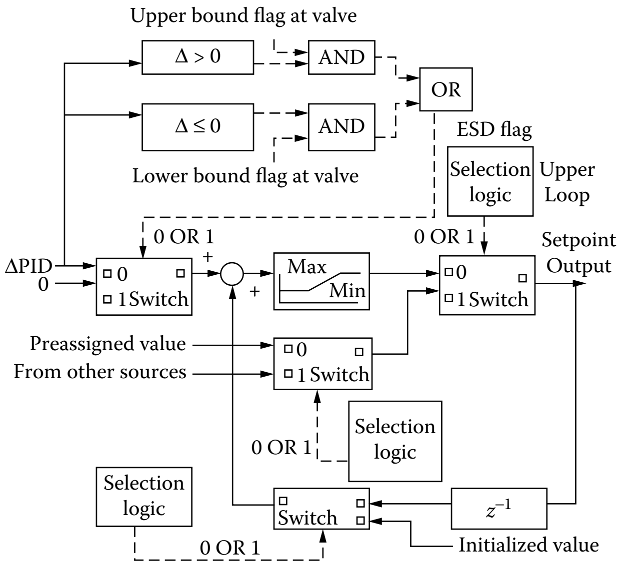


FIG. 1.2e
Supervisory control with antiwindup.

This antiwindup algorithm can be written by modifying Equation 1.2(1) as follows:

$$u(n) = \mathrm{Sat}[u(n-1) + (1 - sw) * \Delta] \qquad \mathbf{1.2(3)}$$

where $sw = 1$ if antiwindup logic is true and $sw = 0$ if antiwindup logic is false.

The switching logic is defined as:

$$I_+ = N_+ (\Delta > 0)$$
$$I_- = N_- (\Delta \le 0)$$
$$sw = (I_+ \text{ .AND. } I_u) \text{ .OR. } (I_- \text{ .AND. } I_1)$$

where

$$\begin{aligned} N_+ (\Delta > 0) &= 1 \quad \text{if} \quad > 0 \\ &= 0 \quad \text{if} \quad \le 0 \\ N_- (\Delta \le 0) &= 1 \quad \text{if} \quad \le 0 \\ &= 0 \quad \text{if} \quad > 0 \end{aligned}$$

I_u = 1 if valve8 position reaches upper bound
= 0 if valve position at upper bound is not reached
I_1 = 1 if valve position reaches lower bound
= 0 if valve position at lower bound is not reached

Combined Discrete and Continuous Algorithm For industrial process control, the control algorithm, if arranged in a structured manner, can be implemented by using a user-friendly configuration arrangement instead of writing program code for each special algorithm. Such an arrangement can be extremely useful to facilitate the implementation of supervisory control by control engineers.

Through the experience gained using several DCS systems implemented at the supervisory computer level, the generalized

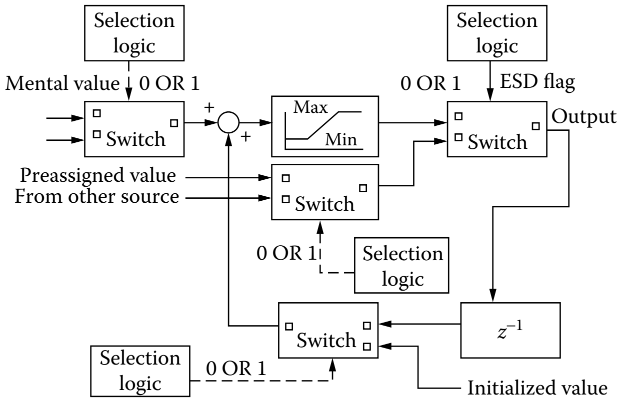


FIG. 1.2f
Generalized output control algorithm.

output control algorithm shown in Figure 1.2f was found to be useful and practical. It is designed to satisfy the application of control output for batch or continuous process control and is combined with emergency shutdown logic.

Algorithm Initialization For supervisory control, the upper-level (supervisory) control algorithm must be initialized when the output from the supervisory computer commanding the lower loops in the DCS is initiated or when an emergency process condition requires that the supervisory control command be prevented from acting in an unrealistic manner. Because the control algorithm is structured in the velocity algorithm form, initialization is simple and universal. That is, $\Delta = 0$, and $u(n-1)$ = initial value from destination loop at the DCS level.

System Failure When the supervisory computer fails, the control commands generated at the computer level cannot be transferred to the DCS level. The designer of the integrated control system should utilize a watchdog timer or communication link message to flag the lower DCS loops as a precautionary step, to prevent unnecessary disturbance to the process due to supervisory computer failure.

The intent of the supervisory control command should be designed so that fault-tolerant control can be achieved in the event that the computer fails. One approach, which can be utilized to achieve fault tolerance, is to implement the generalized output control algorithm at the DCS level, as shown in Figure 1.2f. Proper manipulation of the switch blocks and proper use of the "computer failed" flag at the DCS level ensures that the process control actions at the DCS level will not be perturbed by a computer failure.

DCS and Supervisory Computer Displays In order to judge whether or not the supervisory computer or DCS is performing properly, the displays on the CRT screen become the only man-machine interface that can be utilized for performance evaluation.

There are typically hundreds of display page screens provided in both DCS and the supervisory computer. These operational display pages can be easily accessed through the keyboard, touch screen, mouse, or trackball to arrive at the destination screen for display. The display access method and the display architecture should be similar for both the supervisory computer and the DCS screen displays. The supervisory control monitor and the operator console at the DCS level should have consistent operating procedures and should be easy for the operators to use.

Display Access Method The access methods should be designed such that a minimum number of keystrokes are required to arrive at the destination display, where the controller of interest resides. One convenient access method is to use a tree-structured, multilevel display architecture. This is shown in Figure 1.2g.

For example, if a two-level structure is used for 16 elements (pages) in each level, then the total number of pages of screen will be 16 × 16 = 256 pages. In such case, each screen page at the second level can cover 16 elements, and each element can be any digital or analog input/output combination. Therefore, a total of 4096 combined I/O points can be displayed at the second level. If the correct search path is selected, only two keystrokes are required to arrive at the destination controller for operation.

Display Feature for Supervisory Control Traditional analog controllers on the panel board are arranged in cluster of rows or columns. In the DCS, controllers are usually arranged in groups of 8, 12, 16, or 32 for screen displays in a one-dimensional horizontal arrangement or a two-dimensional matrix arrangement (Figures 1.2i and 1.2j). However, if the same type of the group display for DCS is used for a supervisory control application, the information that resulted from the supervisory computer calculations cannot be properly displayed to inform the operators.

A typical example is the multi-input/multi-output control commands from the supervisory computer to the lower loops in DCS. This type of control algorithm does not exactly fit

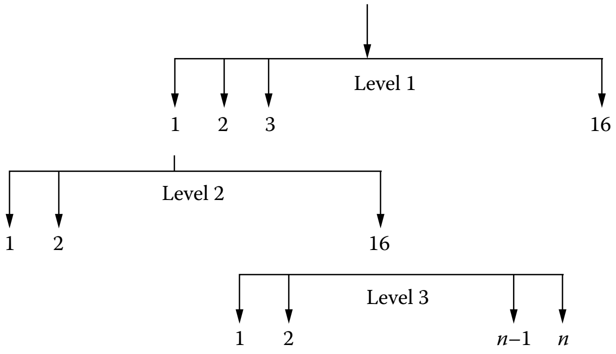


FIG. 1.2g
Multilevel display architecture.

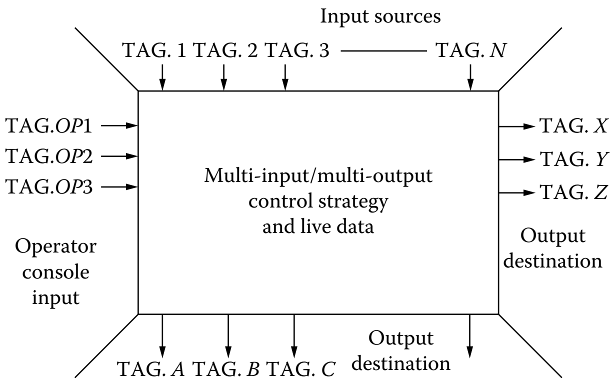


FIG. 1.2h
Multi-input/multi-output strategy display.

the conventional single-input, single-output structure. For this reason a new type of display for the multi-input/multi-output control display structure should be created. This display would indicate the source and destination of information for different locations, magnitude, type of process signals, etc.

One of the simplest ways to represent the information for supervisory strategies is to create a multi-input/multi-output control strategy display. This display is shown in Figure 1.2h.

Alarm Access Architecture Several types of alarms exist for each process variable from the DCS, including high, low, high-high, low-low, and high rate of change. These alarms will appear on the alarm screen at the DCS operator's console in chronological order with tag names, time of occurrence, magnitude of alarmed variable, alarm limit, and so on. The alarm buffer can be created to accommodate the burst of alarms occurring in a short period of time.

For supervisory computers, the alarm management program should be created differently in order to speed the alarm access and to assist the alarm diagnostic operation (to help the operators pinpoint alarm cause and location). To do so, the priority classification or grouping of the pertinent process alarm information is required.

Tag names belonging to an important process unit or equipment should be grouped to give a group alarm flag so that any individual alarm or alarms within that group can be identified logically. This will help speed the decision making of the operator when time is critical. The alarm access architecture can also be grouped according to geographical regions of the process plant or unit so that field operators can be informed (by the control room console operator) to take precautionary measures when necessary.

Voice Input Man-Machine Interface In process control, voice input and output can be an important method for conveying pertinent information about real-time process conditions to the operator. The cause or location of alarms can be announced without viewing the screen display, which clearly can be an advantage when the operator is away from the console. When process variables are approaching an alarm limit, voice output can be generated to alert the operator of the impending alarm condition.

Because of the parallel nature of process control activities, the input methods for man-machine interface should be improved to take advantage of both hands and feet to speed up the operational maneuver.

Advanced Control Strategies

Advanced control, as it is defined here, refers to the multi-input/multi-output control algorithm. The computations usually involve a combination of continuous and discrete algorithms. These algorithms can be implemented either at the DCS or at the supervisory level or at a combination of both of these levels.

Simple PID control algorithms are usually implemented at the DCS level. Those algorithms are structured with single-input/single-output characteristics. Robust multi-input/multi-output algorithms are often implemented at the supervisory computer level and should be carefully structured to satisfy multi-periodic sampled data control requirements.

Advanced Control at the DCS Level Advanced control strategies may fit into the software architecture of DCS systems with sophisticated controllers. Such control modules usually include multi-input and multi-output capabilities inside the control module. Therefore, if the advanced control algorithm is programmed with multi-input/multi-output that have assignable sequence computation steps, and the periodic computation is synchronous with the periodic execution of the control module (such as less than a 30-s range), then the advanced control should be implemented at the DCS level. Examples include ratio controls, feedforward controls, logical sequencing, and start-up and shutdown ramping operations during the transition period.

A typical application of the multi-input/multi-output algorithm is heater pass temperature balancing control combined with the total feed-flow control. The multi-input/multi-output algorithm in this case is structured to fit the single-input/single-output type, suitable for implementation at the DCS level with single-input/multi-output algorithm structure combined to form the multi-input/multi-output algorithm.

The execution of the multi-input/multi-output control period in this case is identical with that of the DCS controller. The control period is usually a multiple basic scan rate of the given controller module in the DCS. The size of the program is usually small. Consequently, the CPU time limit can be satisfied easily.

Advanced Control in the Supervisory Computer For control strategies requiring substantial computation time to complete the program execution, the regular DCS controller is not suitable. This function may involve a process model or iterative computations to obtain the convergent solution. In this case a dedicated computer is required.

To accomplish this computation task, the periodic execution of the program is typically greater than a 30-s period. This period is a function of such process response dynamic parameters as the time constant, pure transportation delay, program execution elapse time, interaction with the lower loop control execution frequencies, DCS network speed, update frequency of the gateway data table, and external periodic interrupt.

Because of the physical distances between the input sources at the DCS and the supervisory computer and because of the multiple scan frequencies involved, the precise timing for the arrival for the commands at the DCS loop set points cannot easily be predicted. It is up to the designer of the advanced control strategy to minimize the timing problem.

An example of advanced control being implemented in supervisory computers and using the multi-input/multi-output concept is gasoline blending with optimal product cost control. The criteria for optimal cost can be minimum product quality, maximum profit subject to the availability of the blend component tank inventories, or delivery time constraints.

An online optimization algorithm, such as a linear programming method or gradient method, can be utilized to obtain the solution for commands that can be manipulated. The commands to the set points of the flow ratio parameters can be so directed by the supervisory computer that optimum volume ratios are maintained for the components.

Mixed Implementation For certain situations, the DCS control module cannot facilitate the advanced control algorithm because of the infrequent execution or the excessive size of the program. In such a situation it is necessary to partition the multi-input/multi-output structure to fit the actual hardware constraint. An example of such a partitioned structure is batch control, where the sequence control can be handled by a supervisory computer, while regulatory and ramping controls can be handled by the DCS control.

COMPUTER INTERFACE WITH DCS

Hardware

Supervisory computers are interfaced to the DCS through a gateway, which is constructed to receive the data or messages from an originating source (see Chapter 4). The gateway resides on nodes of the local control network or data communication network of the DCS. Usually, the communication network topology is either a ring or bus type.

The nodes on the communication network consist of the operator console, engineering console, process control unit, controller subsystem, and gateway. The gateways are usually configured through the DCS from the engineering console to establish a list of variables associated with the tag names of interest in the DCS node on the gateway data table. The gateway device collects the listed data from the DCS communication network or data highway and transfers them to the supervisory computer through a different communication link or network. At the same time, the gateway also receives the data from the supervisory computer or nodes on the computer local area network (LAN) and transfers these to the nodes of the DCS where the information was requested.

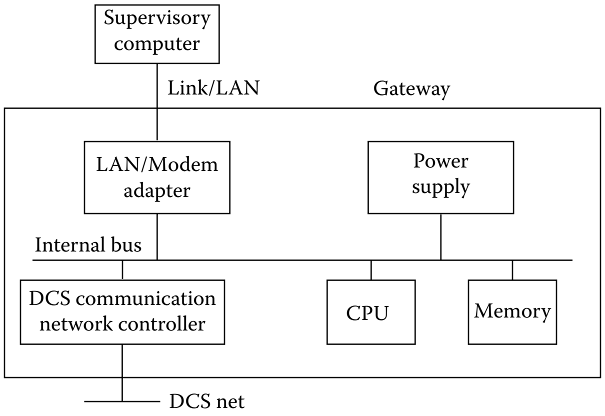


FIG. 1.2i
Generic gateway.

The generic hardware arrangement of the gateway with respect to the supervisory computer and the DCS network is shown in Figure 1.2i. Typical data transmission speeds for the DCS network are from 1 to 10 million bits per second (MBS).

Two types of computer connections are explained here using Honeywell's configuration of the gateway to illustrate the functionality of each type. Type one, called a computer gateway, communicates with a computer via a dedicated link, as shown in Figure 1.2j. Type two is the plant network module, which communicates with computers through LAN. The computer LAN usually differs from the proprietary DCS network in communication speed and protocol. This is shown in Figure 1.2k.

The physical connection of the LAN is typically a bus-type topology; however, the logical connection of the

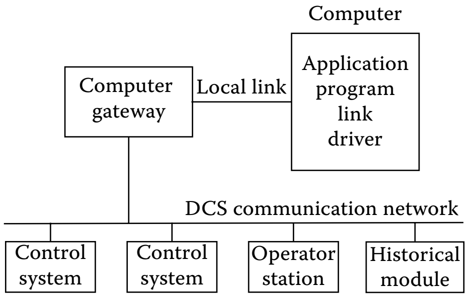


FIG. 1.2j
Gateway for supervisory computer and DCS.

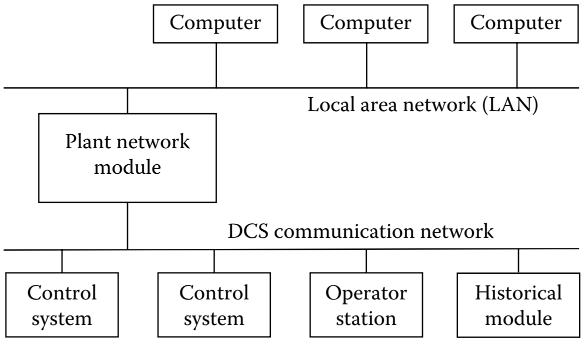


FIG. 1.2k
Gateway for computer LAN and DCS.

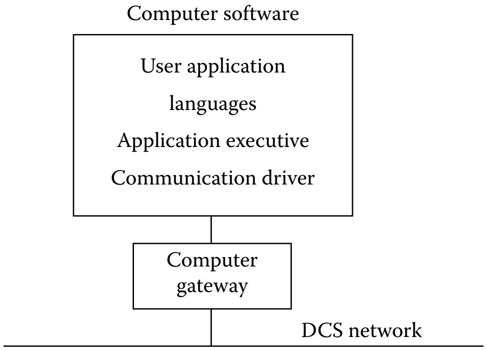


FIG. 1.2m
Computer software structure.

computer-to-computer gateway or to plant network modules can be of the star-type topology.

One approach for computer-integrated manufacturing is to design the hardware and software for the gateway such that integration of manufacturing or process data and plant management information can be achieved easily. Such an interface gateway or network module requires:

1. Highly efficient information communication between the DCS network data or messages and the data or messages in the computers on the LAN nodes
2. High security and reliability for both the DCS network access and the computer LAN access

The actual architecture to this type of hardware connection is shown in Figure 1.2l.

Software

The computer interface module or gateway functions such that the commands from the computer can be transmitted to accomplish read/write processing of data or messages in the data table at the interface module or at the nodes on the DCS. Usually, the data classified for reading are variables such as set points (SP), process variables (PV), control loop outputs (OP), digital inputs, digital outputs, control parameters, device status, controller status, or alarm status. However, the data to the DCS, which is classified for writing, usually are much more restrictive for security reasons.

For a computer, the structure of the software architecture is are shown schematically in Figure 1.2m. The data representation in the interface module is converted into the computer format for transmission.

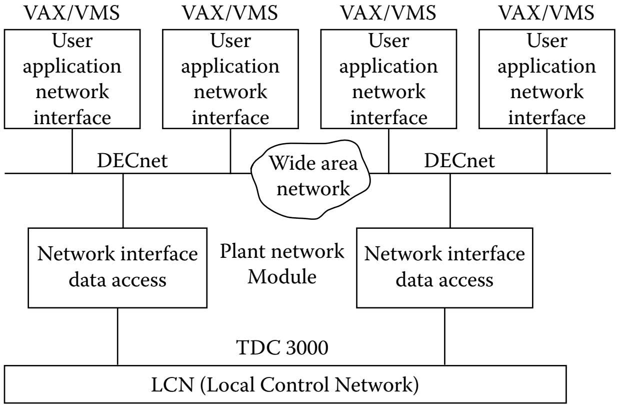


FIG. 1.2l
Honeywell TDC 3000 and VAX computers.

CONCLUSIONS

At the present time, most DCS vendors are able to interface their proprietary DCS systems with computers. The supervisory computer can be used for data analysis and control commands. Most of the supervisory computers are handling the time-critical tasks. However, some personal computers are able to access directly from historical files for engineering analysis of the process using a database software package.

The supervisory computer can assist the operator with the following tasks: performing the supervisory control; alarm classification and prediction; process performance evaluation; process diagnostics; communication routing to other host computers; and online optimization of production.

When the supervisory computer is used with conventional analog controllers, it can be used for data analysis and generation of set point control commands.

References

1. Renganathan, S., *Transducer Engineering,* Chennai, India: Allied Publishers Ltd., 2003.
2. Mendel, J. M., *Discrete Techniques of Parameter Estimation,* New York: Marcel Dekker, 1973.

3. Renganathan, S., "Adaptive Fixed Memory Indentification of Discrete Data Systems," I.I.T., Madras, India.
4. Johnson, C., *Process Control Instrumentation Technology,* 4th ed., New Dehli: Prentice Hall of India Pvt. Ltd., 1996.

Bibliography

Adelman, M. L., "Programmable Controller/Personal Computer Automation Control System," *InTech,* Vol. 34, July 1987, pp. 37–42.

Berry, A. P., "DCS Enhancements for Advanced Simulation," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Dartt, S. R., "Assess Today's Option Is Distributed Process Control," *Chemical Engineering Progress,* October 1990, pp. 61–70.

Gallun, S. E., Matthews, C. W., Senyard, C. P., and Later, B., "Windup Protection and Initialization for Advanced Digital Control," *Hydrocarbon Processing,* June 1985, pp. 63–68.

Shaw, G., "What to Look for in an Industrial PC," *InTech,* March 1993.

Shaw, J. A., "Design Your DCS to Reduce Operator Error," *Chemical Engineering Progress,* February 1991, pp. 61–65.

Shook, P., "PCs Getting Plant Floor Respect," *Control,* April 1993.

Stafford, U., Serianni, A. S., and Varma, A., "Microcomputer Automated Reactor for Synthesis of Carbon 13 Labeled Monosaccharides," *AIChE Journal,* Vol. 36, December 1990, pp. 1822–1828.

Tsai, T. H., Lane, J. W., and Lin, C. S., *Modern Control Techniques for the Processing Industries,* New York: Marcel Dekker, 1986.

1.3 Computers — Simulation by Analog and Hybrid Systems

P. A. HOLST (1970, 1985) **B. G. LIPTÁK** (1995) **S. RENGANATHAN** (2005)

Types:	A. Analog B. Hybrid
Size:	Small — under 50 amplifiers Medium — from 50 to 150 amplifiers Large — over 150 amplifiers
Components:	Integrators, adders, multipliers, function generators, resolvers, diodes, switches, relays, digital attenuators, logic gates, flip-flops, counters, differentiators, pulse generators, and coefficient potentiometers
Interface Components:	Analog-to-digital and digital-to-analog converters, multiplexers, track-store amplifiers, and logic signal registers
Inaccuracy:	For static, linear analog units, ±0.005% to ±0.1% For static, nonlinear analog units. ±0.01& to ±2%
Frequency:	Clock cycle frequency, 100 kHz to 2 MHz. Conversion rate. 10 KHz to 1 MHz. Full power signal bandwidth, from DC to 25 to 500 kHz
Costs:	Computers with 20 to 100 amplifiers cost from $7500 to $30,000. Units with more than 100 amplifiers cost from $20,000 to $150,000. For 10 to 50 channels of hybrid interface, the cost ranges from $15,000 to $75,000.
Partial List of Suppliers:	ABB Kent-Taylor (B) (www.abb.com/us/instrumentation) Acromag Inc. (A) (www. acromag.com) Adtech Instrument (A,B) (www.adtech.info) AGM Electronics Inc. (A,B) (http://agmelectronics.com/AGMContact.asp) Ametek, Rochester (B) (www. ametekapt.com) ASC Computer Systems (A) (www.asclubbock.com) Celesco Transducer Products (B) (www.celesco.com) Compudas Corp. (B) (www.compudas. com) Computrol Inc. (B) (www.computrol.com) Devar Inc. (A) (www.devarinc.com) Electronic Associates Inc. (B) (http://dcoward.best.vwh.net/analog/eai.htm) Emerson Process, Daniel Div. (A) (www.easydeltav.com) Gould Inc. (B) (www. pumpsmar.net) ILC Data Device Corp. (B) (www.electrobase.com) Invensys, Foxboro (A, B) (www.foxboro.com) MTS Systems Corp. (B) (www.mtssensors.com) Siemens, Moore Products (A, B) (www.sea.siemens.com) Voice Computer (B) (www. webdesk. com/voice-computer-control/) Xycom Automation Inc. (B) (www. xycom. com).

This section first describes the components and the operation of both analog and hybrid computers and then covers some of the simulation techniques for first order, second order, PID, and other applications.

The computer is a versatile tool for many applications ranging from simple calculation to sophisticated control of large-scale process plants. Computers are classified as analog and digital. Analog computers accept continuous signals and perform many operations on them, such as addition, subtraction, multiplication, integration, and simulation of systems.

THE ANALOG COMPUTER

Analog computers work on continuous signals and consist of operational amplifiers, capacitors, resistors, potentiometers, diodes, switches, squarer cards, and patch cards. All the fundamental functions of computation, such as addition, subtraction, multiplication, integration, differentiation, and generation of different functions, can be carried out with an analog computer.

The heart of an analog computer is the operational amplifier. The operational amplifier is a single-ended, high-gain DC coupled wide bandwidth unit, which has a very high open-loop gain, on the order of 10^5 to 10^8. All the computer signals are referenced to a common ground. Offsets, drifts due to temperature variation, aging, and electronic noise are the main problems in the operational amplifier circuits. Hence special precautions are taken to reduce or eliminate them. Special care should be taken to minimize the electromagnetic and static coupling between different computing units and signal sources.

The basic analog computing units are the inverter, summer, integrator, and multiplier.

Operational Amplifier

The main component of an analog computer is the operational amplifier, popularly called the op-amp, which is represented by a symbol shown in Figure 1.3a.

Its static gain μ is of the order of 10^5 to 10^8 and is a single-ended input and output representation. Its other special features are:

- High input impedance
- High DC gain
- Low power requirement
- Low noise and drift
- Stability and ruggedness

With the advent of integrated circuits (IC), operational amplifiers are now available as chips, such as the units IC741, IC741C, and IC LM 308. The ICs are quite compact, and many problems such as offset, drift, and noise associated with discrete component models have been minimized or eliminated.

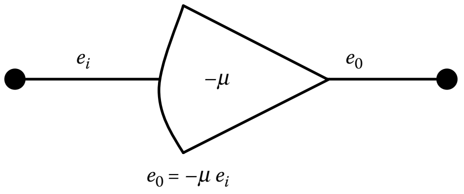


FIG. 1.3a
Symbol for the operational amplifier (OP-AMP).

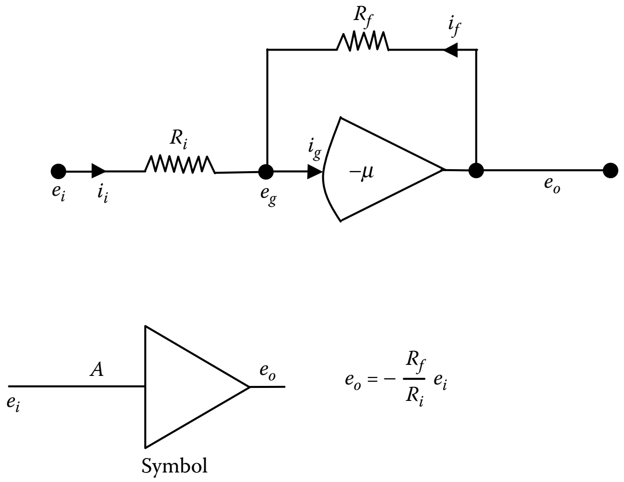


FIG. 1.3b
The circuit of and the symbol for an inverter.

In analog computers, operational amplifiers are used for three basic purposes:

- To generate the necessary computing functions
- To amplify the input signal level
- To provide isolation and unloading between the different input and output signals within the computing units

The Inverter

This operational amplifier can be used in conjunction with resistors to invert and multiply a signal. The circuit for and the symbol of the inverter are shown in Figure 1.3b. It may be noted that e_g is very near earth potential because of the large value of μ and a finite e_0 value. The current flowing into the op-amp grid is very small, on the order of micro amps.

According to Kirchoff's current law:

$$i_i + i_f = i_g \tag{1.3(1)}$$

when $i_g = 0$, $i_1 = -i_f$, $\frac{e_i}{R_i}$ and $i_f = \frac{e_0}{R_f}$.

Therefore

$$e_0 = -\frac{R_f}{R_i} e_i$$

If $R_f > R_i$, the input is inverted and amplified.

The Summer

The operational amplifier can be used as a summer by using the circuit shown in Figure 1.3c. Assuming μ to be large, i_g can

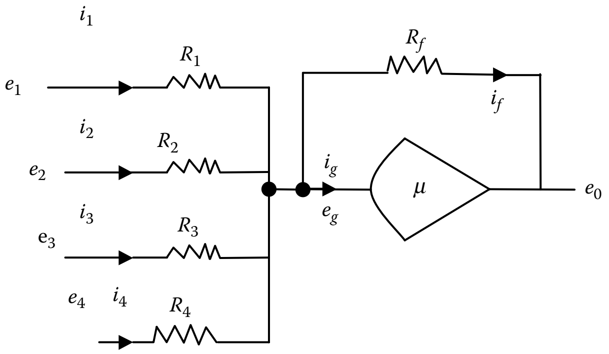


FIG. 1.3c
The circuit of a summer.

be neglected. Applying Kirchoff's equation, Equation 1.3(2) is obtained:

$$i_f = (i_1 + i_2 + i_3 + i_4) \qquad \textbf{1.3(2)}$$

As e_g is near ground potential

$$i_f = e_0/R_f,\ i_1 = e_1/R_1,\ i_2 = e_2/R_2,\ i_3 = e_3/R_3 \quad \text{and} \quad i_4 = e_4/R_4 \qquad \textbf{1.3(3)}$$

Then

$$e_0/R_f = -[e_1/R_1 + e_2/R_2 + e_3/R_3 + e_4/R_4] \qquad \textbf{1.3(4)}$$

$$e_0 = -[R_f/R_1e_1 + R_f/R_2e_2 + R_f/R_3e_3 + R_f/R_4e_4] \qquad \textbf{1.3(5)}$$

$$e_0 = -[g_1e_1 + g_2e_2 + g_3e_3 + g_4e_4] \qquad \textbf{1.3(6)}$$

The symbol for a summer is given in Figure 1.3d.

If the open loop gain μ is low, the summing junction potential e_g must be considered in Equation 1.3(2):

$$(e_0 - e_g)/R_f = -[(e_1 - e_g)/R_1 + (e_2 - e_g)/R_2 + (e_3 - e_g)/R_3 + (e_4 - e_g)/R_4] \qquad \textbf{1.3(7)}$$

Since

$$e_0 = -\mu e_g \qquad \textbf{1.3(8)}$$

Substituting 1.3(8) in 1.3(7), one obtains:

$$e_0\,(1/R_f + 1/\mu R_f) = -[e_1/R_1 + e_0/\mu R_1 + e_2/R_2 + e_0/\mu R_2 + e_3/R_3 + e_0/\mu R_3 + e_4/R_4 + e_0/\mu R_4] \qquad \textbf{1.3(9)}$$

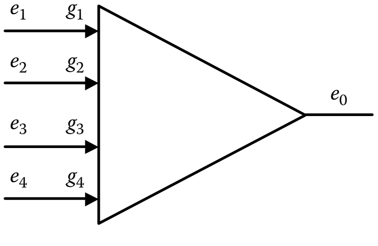


FIG. 1.3d
The symbol for a summer.

i.e.,

$$e_0\,[1/R_f + 1/\mu R_f + 1/\mu R_1 + 1/\mu R_2 + 1/\mu R_3 + 1/\mu R_4] = -[e_1/R_1 + e_2/R_2 + e_3/R_3 + e_4/R_4] \qquad \textbf{1.3(10)}$$

$$e_0 = -f_u\,[g_1e_1 + g_2e_2 + g_3e_3 + g_4e_4] \qquad \textbf{1.3(11)}$$

where

$$\begin{aligned} g_1 &= R_f/R_1,\ g_2 = R_f/R_2,\ g_3 = R_f/R_3,\ g_4 = R_f/R_4 \\ 1/f_u &= 1/\mu(1 + \mu + g_1 + g_2 + g_3 + g_4) \end{aligned} \qquad \textbf{1.3(12)}$$

f_u is the correction factor.

The analog computer circuits and their symbols, which are used for basic computation, are summarized in Figure 1.3e.

Frequency Response of the OP-AMP

It may be noted that the open loop amplification (open loop gain) of the OP-AMP decreases as the signal frequency increases. Hence the analog computation must be carried out within the corresponding frequency limits. For higher frequencies, the gain falls off. When the signal frequency is 100 kHz, the open loop gain becomes 10^2 as shown in Figure 1.3f. The correction factor for this gain is 101/100. That means the error in computation is 0.01%

Analog Circuits for Differential Equations

Consider the equation below, in which a is a positive constant and $f(t)$ is an arbitrary function of time:

$$dx/dt = ax + f(t) \qquad \textbf{1.3(13)}$$

Let us assume that $-dx/dt$ is available as voltage from some amplifier. This voltage is connected to an integrator as shown in Figure 1.3g. The output of this integrator is x. Then this output is connected to the summer as also shown in Figure 1.3g. The output of this summer is $-[ax + f(t)]$, which is equal to $-dx/dt$. By connecting the output of the summer to the input of the integrator, the equality of the equation 1.3(13) has been electrically established.

Figure 1.3g can be redrawn as in Figure 1.3h. Once this circuit is made and the circuit is switched on, $x(t)$ is obtained as a function of time, which is the solution of the differential equation. It may be noted that the initial condition is assumed to be zero.

The detailed circuit for a first-order differential equation is described in Figure 1.3i.

There is no fixed procedure for constructing computer circuits to solve differential equations. There are many satisfactory ways to arrive at a computer circuit to solve a given differential equation.

The analog computer circuit shown in Figure 1.3h is for solving a first-order differential equation with a zero initial condition. But normally, differential equations will have boundary conditions or initial conditions. A first-order differential equation will have one initial condition, a second-order will have two initial conditions, and so on.

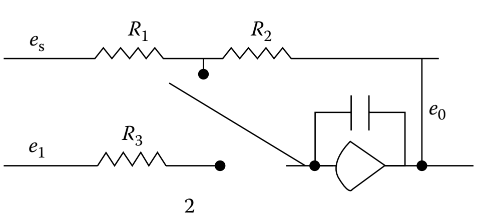
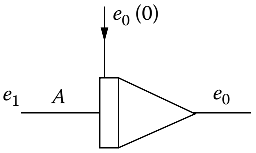
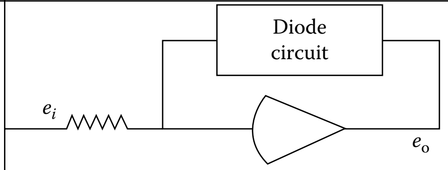
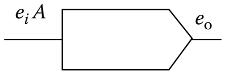
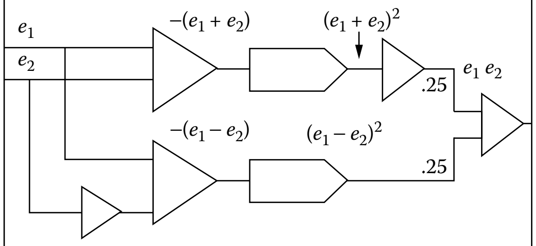
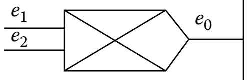
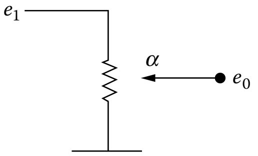
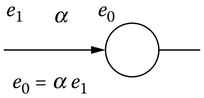
Integrator with initial condition	 $e_0 = -\left[\frac{1}{R_3C}\int_o^t e_i dt + \frac{R_2}{R_1}e_s\right]$	 $e_o = -A\int_o^t e_i\,dt + e_o$ $A = \frac{1}{R_3c}$; $e_o(o) = -\frac{R_2}{R_1}e_s$	To have an initial condition for integration the capacitor is charged initially to a required value by choosing R_2 & R_1. The switch is initially at point 1 for a sufficient time so that the capacitor C is charged to $R_2/R_1\,e_s$. When we want to integrate the signal e_s the switch is moved from position 1 to position 2 and then the integration starts.
Squarer	 $e_0 = -Ae_i^2$ Square function generation	 $e_0 = -Ae_i^2$	This device approximates a function $y = f(x)$ by a number of straight line segments. The electrical components used in a square function generator are diodes, resistors, and potentiometers.
Multiplier		 $e_0 = -\,e_1\,e_2$	This circuit is known as a quarter square multiplier $e_0 = \frac{1}{4}\left[(e_1+e_2)^2 - (e_1-e_2)^2\right]$ $= e_1e_2$ Now the circuit is available as a device which produces a current proportional to the product of two voltages.
Signal divider		 $e_0 = \alpha\,e_1$ $0 \le \alpha \le 1$	

FIG. 1.3e
Summary of analog computer circuits and their symbols used for basic computation.

For the first-order system described by Equation 1.3(13), assume the initial condition to be $x(0) = x_0$. This requires that the voltage from the integrator be held at x_0 until the integration starts. There are two ways of obtaining the initial condition. One way is to connect a battery of the required initial condition value between the grid and the output as shown in Figure 1.3j.

Initially, the switch is closed for a sufficiently long time for the capacitor to become charged to the initial condition value x_0. Two relay-operated switches are placed in the circuit as shown in Figure 1.3j. When the initial condition is to be established, switch 1 is closed and switch 2 is simultaneously opened. When the equation is to be solved, a suitable relay system opens switch 1 and closes switch 2 at the same time.

Another way of obtaining the initial condition voltage is shown in Figure 1.3k. In this case, when the computer is put in the initial condition (IC) mode, the switches S_1 and S_2 are put in position 1. This is done simultaneously by a master switch. A voltage V_s of desired magnitude and polarity is connected to the R_1 and R_2 circuit. Normally, R_1 and R_2 values will be 0.1 MΩ. When the switch S_1 is in position 1, $e_0 = -(R_2/R_1)V_s$, and when $R_1 = R_2$, e_0 becomes the negative of V_s.

Then, according to the time constant of the RC circuit, the capacitor is charged to $-e_0$. In this case the time constant $RC = 0.1$ s, and hence one can expect that in five times the time constant or say in 0.5 s the capacitor will be charged to V_s voltage, which becomes the initial condition for the integrator.

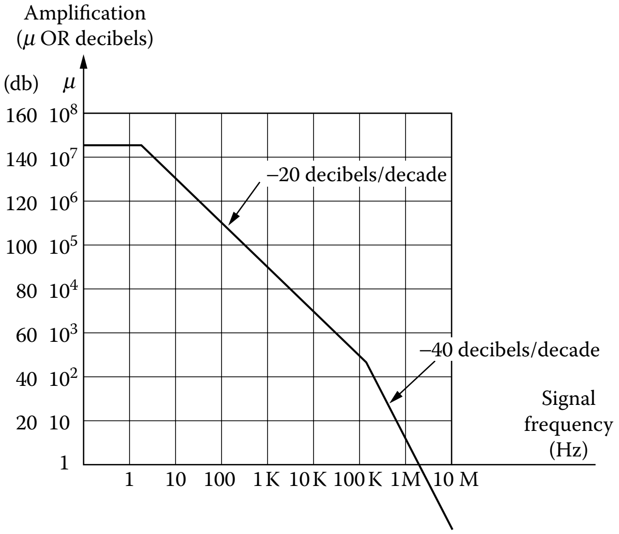


FIG. 1.3f
The open-loop gain characteristic is such that as the signal frequency increases, the open-loop gain (amplification) of the OP-AMP decreases.

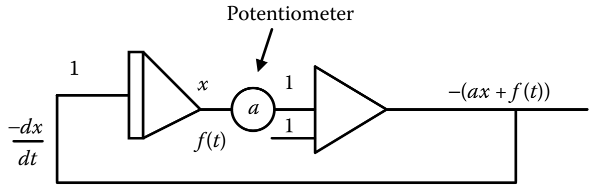


FIG. 1.3g
The circuit for solving Equation 1.3(13).

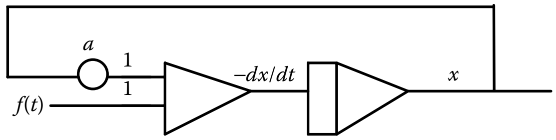


FIG. 1.3h
The computer circuit for solving first-order differential equations.

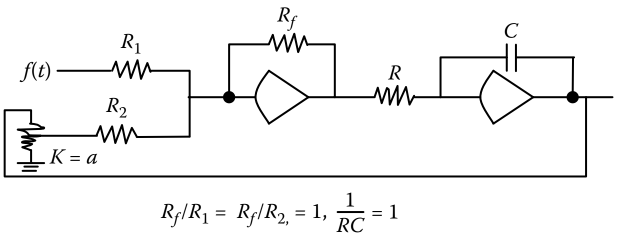


FIG. 1.3i
The detailed circuit for solving first-order differential equations.

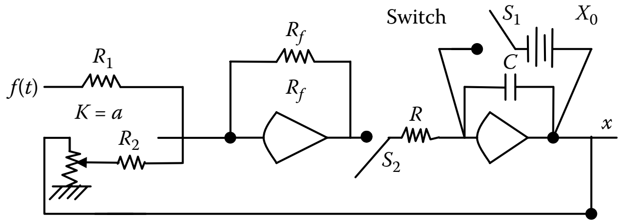


FIG. 1.3j
Analog computer circuit for solving the equation dx/dt = ax + f(t) with initial condition.

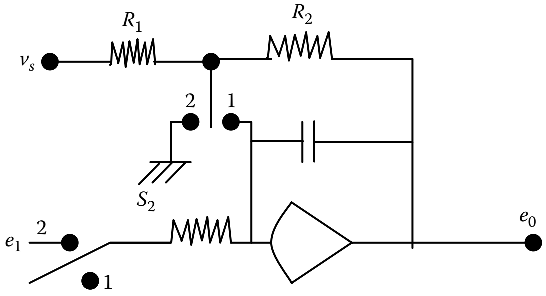


FIG. 1.3k
The circuit for an integrator with IC value.

When the computer is placed in the operate mode, switches S_1 and S_2 are switched to position 2 and the voltage is integrated. In this state, resistor R_1 provides a load for the initial condition voltage because one side of the resistor is grounded through switch S_1.

Figure 1.3l shows the symbol for an integrator with an IC value. It may be noted that the reference voltage is opposite in sign to the initial condition voltage because of the sign reversal at the amplifier.

Magnitude and Time Scaling

To get satisfactory results utilizing the analog computer, magnitude scaling and time scaling are needed.

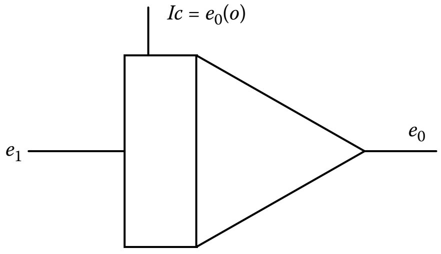


FIG. 1.3l
The symbol for an integrator with IC value.

Magnitude Scaling During the solution of a problem in an analog computer, the output voltage from any amplifier should not exceed the operating range of that amplifier. For many semiconductor analog computers the nominal voltage range is +15 to –15 V. If, at any point of the solution, the output voltage of any amplifier exceeds this value, the amplifier response may not be linear. Often, it is possible to estimate the maximum value of x, say x_m, which occurs during the solution. This maximum value should be reduced to the maximum output voltage of the amplifier, which is say 15 V, by selecting the suitable scaling factor of $k_x = 15/x_m$, so that the amplifier does not saturate.

For example, if $x_m = 5$ units, then $k_x = 15/5 = 3$.

By using magnitude scaling, full use is made of the voltage range available in an analog computer and also the overloading and saturation problems are taken care of. Scale factor for the derivates of the dependent variable

$$dx/dt, \qquad d^2x/dt^2$$

are also to be chosen by estimating their maximum values. To fully understand the concept of magnitude scaling, consider the second-order differential equation:

$$d^2x/dt^2 + a\,dx/dt + bx = f(t) \tag{1.3(14)}$$

where x is the dependent variable and t is the independent variable. The coefficients a and b are positive constants. The dependent variable and its derivates are available as voltages at the output of amplifiers on the computer. Magnitude scale factors are used to correlate the voltages from amplifiers to the dependent variable and its derivates.

In case of a vibration problem, x may represent displacement, dx/dt can stand for velocity, and d^2x/dt^2 for acceleration.

for acceleration.

$$\text{Let } x1 = dx/dt \quad \text{and} \quad x2 = \frac{d^2x}{dt^2}$$

if the maximum anticipated values for x_m, $x1_m$, and $x2_m$. In that case the scale factors are:

$$k_1 = 15/x_m,\ k_2 = 15/x1_m, \quad \text{and} \quad k_3 = 15/x2_m \tag{1.3(15)}$$

Equation 1.3(14) can be rewritten as:

$$\frac{d^2x}{dt^2} = -a\frac{dx}{dt} - bx + f(t) \tag{1.3(16)}$$

Using the scale factors, the variables at the output of the amplifiers are $[k_1x]$, $[k_2x1]$, and $[k_3x2]$. Rewriting the equation using $[k_1x]$, $[k_2x1]$, and $[k_3x2]$, we get

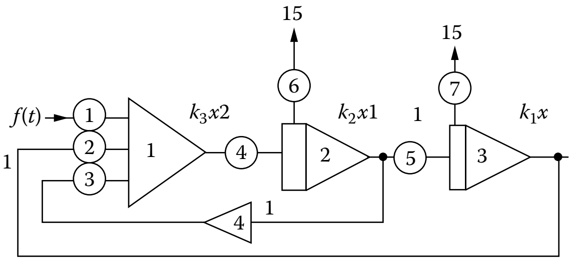
$$[k_3x2]/k_3 = -a[k_2x1]/k_2 - b[k_1x]k_1 + f(t) \tag{1.3(17)}$$

or

$$[k_3x2] = -a(k_3/k_2)[k_2x1] - b(k_3/k_1)[k_1x] + k_3f(t) \tag{1.3(18)}$$

The analog computer circuit for Equation 1.3(18) is shown in Figure 1.3m, which is a generalized circuit for solving a second-order differential equation with magnitude scaling.

Time Scaling It may be noted that the analog computer integrates with respect to time measured in seconds. But the independent variable in a differential equation representing a real time problem may be time in minutes or length in feet. The time needed for the solution of a particular problem may be 30 minutes, in which case 30 minutes of operators' time and the computer time would be needed, if time scaling was not used to speed up the solution.



Potentiometer number	1	2	3	4	5	6	7	
Values of its setting	k_3	$b\,k_3/k_1$	$a\,k_3/k_2$	k_2/k_3	k_1/k_2	$k_2x/1(0)/15$	$k_1x/(0)/15$	← Magnitude scaling
Values of its setting	k_3	$b\,k_3/k_1$	$a\,k_3/k_2$	$k_2/b\,k_3$	$k_1/b\,k_2$	$k_2x1(0)/b15$	$k_1x/(0)/15$	← Magnitude and time scaling

FIG. 1.3m
A generalized analog computer circuit for solving a second-order differential equation with magnitude scaling (Equation 1.3[18]).

For some fast-acting processes, such as transients in circuit breakers, slowing down the problem is needed to capture and record the variation. Hence time scaling is needed whenever it is necessary to speed up or slow down the process when it is studied in an analog computer.

If one designates the computer time as t and the real time as τ, then the time scale factor can be defined as the ratio of $t/\tau = \beta$. The steps in the procedure that was adopted for time scaling are listed below:

Step 1. Assuming $\beta = 1$, magnitude scale the problem as described earlier.
Step 2. Choose β, $\beta > 1$ when speeding up the solution or $\beta < 1$ to slow down the solution. The actual values can be selected from knowing the real process.
Step 3. Multiply the gain of each input to an integrator by I/β
Step 4. Modify the time in the forcing function from $f(t)$ to $f(\tau/\beta)$.

The above procedure can be applied to Equation 1.3(18) to implement the time scaling:

$$\left[\beta^2 k_3 d^2 x/d^2\tau\right] = -a(k_3/k_2)\left[\beta k_2 \frac{dx}{d\tau}\right] - b(k_3/k_1)[k_1 x] + k_3 f(\tau/\beta) \qquad \textbf{1.3(19)}$$

The initial conditions for potentiometer settings are given in the table at the lower part of Figure 1.3m.

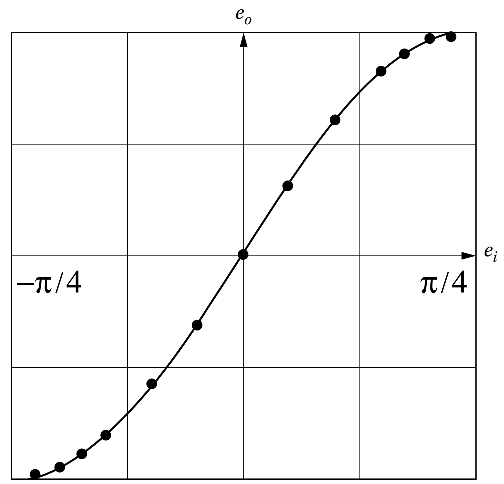
Nonlinear Components

For solving nonlinear equations using analog computers, one needs nonlinear components[1] such as function generators, multipliers, dividers, and switches. These nonlinear elements significantly extend the usefulness of a computer by allowing nonlinear equations to be solved with the same ease as linear ones.

Function Generator The most common function generator used is a diode function generator.[2] This circuit approximates a function $y = f(x)$ by a number of straight-line segments of different scopes. To generate a functional relationship $e_o = \sin(2e_i)$, a number of line segments with different slopes are needed, as shown in Figure 1.3n.

Multiplier To obtain the instantaneous product of two time-varying voltages, a multiplier is needed. A convenient symbol for the multiplication block is shown in Figure 1.3e. Three types of multipliers are used in computers.[2] They are:

1. The quarter-square multiplier
2. The servo mechanical multiplier
3. The time division multiplier



Example of function $e_0 = \sin(2\,e_i)$

FIG. 1.3n
The response of a sine function generator.

Quarter-Square Multiplier The identity of the quarter-square multiplier can be described as:

$$xy = 1/4[(x+y)^2 - (x-y)^2] \qquad \textbf{1.3(20)}$$

The analog computer circuit that corresponds to the above equation is shown in Figure 1.3o. Today, the quarter-square card is available from many manufacturers; this card gives a current output, which is proportional to the product of two voltages divided by 10.

The Analog Computer Unit

Commercially available analog computer equipment is box-type equipment with an orderly collection of amplifiers, power supplies, potentiometers, reference power supplies, diodes, switches, and some function generators, quarter-square cards, etc. The connections from the various components are brought to a panel called a patch board. The patch board is removable, and many spare patch boards are available. This facility is for wiring a computer circuit when the computer is to operate on another problem. The wired computer circuits can also be stored and used later.

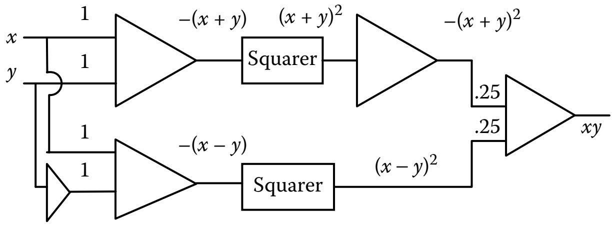


FIG. 1.3o
Analog computer circuit for a quarter-square multiplier.

In many analog computers, the resistors and capacitors used to construct summers and integrators are not accessible to the user. They are located inside the computer, and only their terminals are brought out to the patch board.

A reference power supply is provided for establishing initial conditions on the integrators. X-Y plotters, oscilloscopes, and strip chart recorders are used to record or observe the response or solution.

HYBRID COMPUTERS

A hybrid computer is a combination in hardware and software of one or more analog and digital computers. It aims at providing faster, more efficient, and more economical computational power than is available with computers of either type alone. The results depend to a large extent on the exchange of information between the analog and the digital computers and on the compatibility in operations and mutual interactions between the two parts.

A hybrid computer provides for the rapid exchange of information between the parallel and simultaneous computations and simulations within the analog computer and the serial and sequential operations of the digital computer. This information exchange links the two computational domains and offers the combined advantages of the fast and flexible analog computer with the precise and logic-controllable digital computer.

The extent of the information exchange between the two parts and the sophistication of the control structures and instruction repertoires determine the capability and the capacity of the hybrid computer. Best results are obtained when both computers are designed and developed with hybrid applications as the major purpose. If a hybrid computer is made up of general-purpose analog and digital computers, with an interface tailored to these, the resulting hybrid computer often poses severe limitations in equipment complement and operational features.

Hardware

The distinguishing feature of a hybrid computer is the interface that connects the analog and digital computers. The interface consists of data communication channels in which information is passed between the two computational parts. The interface does not carry out computations, but it may contain equipment that incorporates computational units, such as multiplying digital-to-analog converters.

The interface contains a number of conversion channels in which information is converted between an analog signal (voltage) and an encoded numerical (discrete) digital computer representation, according to programmed instructions and hardware executions. The number of conversion channels states the total parallel capacity of the interface, for conversions in both directions. In practice, the number of A/D (analog-to-digital) channels may differ from the number of D/A channels, depending on applications and implementations.

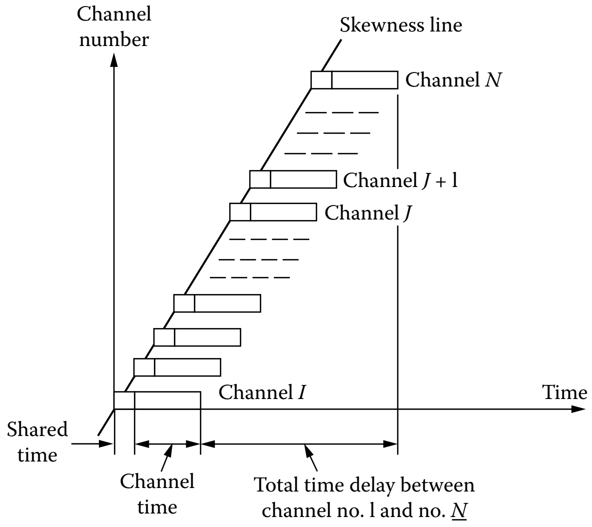


FIG. 1.3p
The skewness of converted data.

Since the conversion channels link parallel and concurrent analog computer variables with serial and sequential program steps in the digital computer, the interface must provide storage facilities, by which information (variables) can be stored, while all channels are being prepared (loaded). This way, all conversions can take place simultaneously, in terms of analog variables, and sequentially, in terms of digital variables.

If the converted information is not buffered, it reflects computational variables or conditions that are not concurrent or coreferenced in time. Such skewness is indicated in Figure 1.3p, showing the effect of a sequential conversion capability (among several time-shared units). To a degree, and at the cost of computer time, the skewness effects may be reduced by a common hold mode for all analog computing units, during which hold the information conversion may take place. Fast, accurate mode control facilities that allow rapid computer interruption and resumption of the analog program are necessary for this.

Figure 1.3q shows an example of an A/D conversion channel in which the analog signal is tracked and stored in analog form. The conversion channel utilizes a multiplexer and converter that are shared among a number of channels, typically 24 to 36, and it reads the converted information into a programmed (controlled) memory location in the digital computer. From this memory location the digital program can then obtain the converted information.

An example of a D/A conversion channel is shown in Figure 1.3r. It consists of a buffer register for the digital information and converter unit. The buffer register holds the digital information until the moment of conversion, when it is loaded into the converter with other D/A channels. For single-channel or continuous D/A conversion, the buffer register is sometimes bypassed, and the digital information read directly (jammed) into the converter.

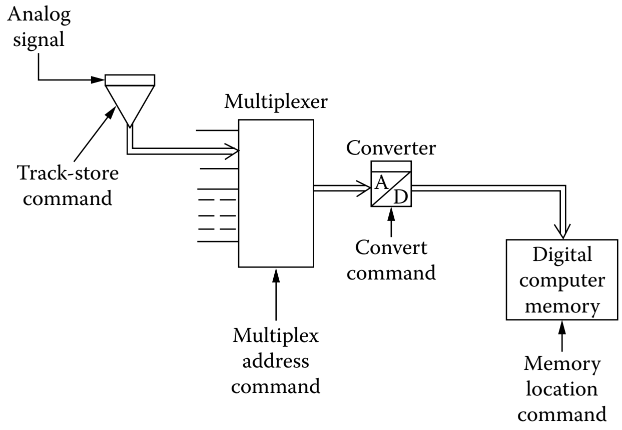


FIG. 1.3q
The components of an A/D channel.

The analog conversion output is a direct signal output or, as in a multiplying D/A, the product of an analog signal input and the converted D/A information. The conversion moment is only triggered by digital computer instructions, or indirectly, by interrupts form the analog computer through the digital computer program.

A number of status channels in the interface provide binary information exchange between the analog and digital computers. This information relates to the status of particular computational variables or computing units and to the condition and progress of the programmed functions. Status channels are used to ensure proper relationships in terms of operation sequences, timing of program events, steps, and coordination of the two hybrid computer parts.

There are three main types of status channels, depending on the importance or immediacy of the reported conditions:

1. Status indicating, which at all times corresponds in status to the condition of its input variable
2. Status retaining, which will retain the predetermined status of the input variable from the first time it is set or activated until the status channel is interrogated by the receiving computer and the channel is reset or deactivated
3. Status responding, which will respond immediately to the condition of the input variable when this is set or activated and will cause the receiving computer to interrupt its task or change its modes or functions according to the programmed actions

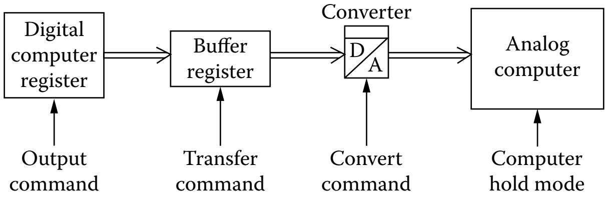


FIG. 1.3r
The components of a D/A channel.

Status channels are generally furnished in the interface as one-way signal lines, which must be assigned to the conditions or computing units that are to be reported during the programming.

The interaction and control of operations between the two computers are handled by binary command channels. The command channels represent direct, fixed interactive links between the control systems or command sources of the two computers. They control the executions of programs, such as start and stop computations and iterations, and the initialization or reset of programmed functions and routines.

Operation

The operational efficiency of a hybrid computer depends on the command and control (instruction) structures of the two computers. Analog and digital computers are sufficiently different in organization and operation to present profound problems in terms of command characteristics and functional orientation. In a hybrid computer, the analog and digital computers are linked together on the fundamental control level, providing facilities for mutual interruption and interaction of tasks.

In terms of analog computer control, the hybrid computer permits the digital computer program to carry out the control system functions previously outlined. These functions span the setup and checkout of analog and hybrid programs, the initialization and presetting of conditions for the computations and simulations, and the measurement, recording, and monitoring of the analog computer variables and functions during productive runs. Of prime importance in a hybrid computer is the ability of the computer programs to govern the progress of the computations and to take the appropriate control actions, depending on the obtained results and responses.

The digital computer instruction repertoire normally includes many bit-handling instructions designed to facilitate the exchange of information through the status channels, command channels, and direct mode control and command channels. These instructions permit multiple-level priority assignments and convenient handling of the exchanged information, which can be either in a converted data format or in single-bit format.

The fast access to information in the digital computer memory through direct access or cycle stealing is important for high-speed conversion channel utilization and efficient hybrid computations.

Software

Good and complete software is required to attain the optimum operational efficiency and utilization of the hybrid computer. This is especially important in hybrid computer applications since the complexity and extent of many hybrid programs and the sophistication of the instruction repertoire and control routines in the computers otherwise limit the usefulness and understanding of the computer capabilities.

The software is designed with practical problem-solving objectives in mind, such as handling the system's functions and presenting conditions and variables in concise and efficient formats. For best results, the software is written for a particular hybrid computer and is defined and specified according to the characteristics of the hardware configuration. The software consists of three types of system programs:

1. Batch-oriented hybrid computer operation programs, organized and utilized for large, complex hybrid problems, with extensive program setup and checkout demands
2. Conversational computer operation programs, designed for experimental and development-oriented type hybrid computations and simulations
3. Utility routines designed for efficient and convenient setup, checkout, and documentation of intermediate or limited hybrid computer programs, such as test programs, experimental circuit evaluations, and hybrid program developments

The software enables the operator to carry out a hybrid computation or simulation of the specific problem with a minimum of effort. In a hybrid application, this includes determining the signal connections and tie-lines, calculating the scale factors and coefficients of analog variables, adjusting coefficient units to the appropriate values prior to or during computations, and selecting the states or modes of the analog computing units.

An important aspect of the software is its capabilities for readout and documentation. It must be able to obtain the values within the hybrid program and to decode or interpret these in the language of the problem (such as in engineering units or mathematical terms). Finally, it must be able to make this information available to the operator, either on CRT, graphic display, trend curve, or in other forms.

PROCESS SIMULATION

The simulation of most processes, reactions, and plants requires the developing of a representation or building a model of common elements. For this model, each element is adapted to the particular functions it must simulate. When all the parts have been defined and assembled, the overall effects and constraints may be imposed, such as material and energy balances, operating modes, limitations, instrumentation, and control system characteristics.

Laplace Transforms

For linear control systems, Laplace transforms can be used to obtain their transfer functions. Laplace impedances can be used when working with Laplace transfer functions for system simulation. Information on Laplace impedance is available in most books on circuit theory or network analysis,[3] while these Laplace impedances are referred to as complex impedances in control system books.[4]

For the two terminal elements like resistance, capacitance, and inductance the impedance are given by R, $1/Cs$ and Ls, respectively. If these complex impedances are connected in series, the total impedance is the sum of the individual complex impedances, and if these are connected in parallel, then the parallel resistance rule is applied. It may be noted that complex impedances are valid only for such transfer function approaches where the initial conditions are assumed to be zero. For example, a series combination of R and C has a total impedance of

$$Z = R + 1/Cs = (RCs + 1)/Cs \qquad \textbf{1.3(21)}$$

The impedance of R and C connected in parallel is obtained by equating the reciprocal of the parallel impedance, i.e.,

$$1/Z = 1/R + Cs = (1 + Rcs)/R$$

Then

$$Z = R/(1 + RCs) \qquad \textbf{1.3(22)}$$

First-Order System Simulation The first-order transfer function $A/(1 + sT)$ can be simulated by the circuit shown in Figure 1.3s:

$$Y/X = Z_f/Z_i = R/R_1(1 + Rcs) = A/(1 + Ts) \qquad \textbf{1.3(23)}$$

where $A = R/R_1$ and $T = RC$.

Another way of obtaining a simulation circuit is as follows: First convert the Laplace function into a time function:

$$(1 + Ts)Y = AX \qquad \textbf{1.3(24)}$$

Then, in order to get the time domain equation, replace s and d/dt in Equation 1.3(24):

$$\begin{aligned} y + Td/dt\ y &= Ax \\ d/dt\ y &= -y/T + A/Tx \end{aligned} \qquad \textbf{1.3(25)}$$

The computer circuit for this given in Figure 1.3t.

Second-Order System Simulation In order to simulate a second-order system, which is in the form of $K/(1 + sT_1)(1 + sT_2)$,

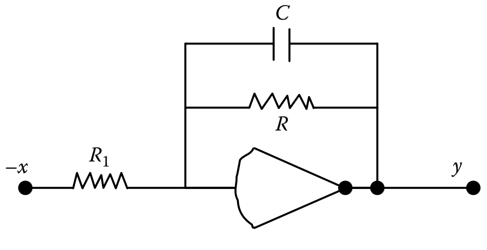


FIG. 1.3s
This circuit simulates a first-order transfer function A/(1 + sT).

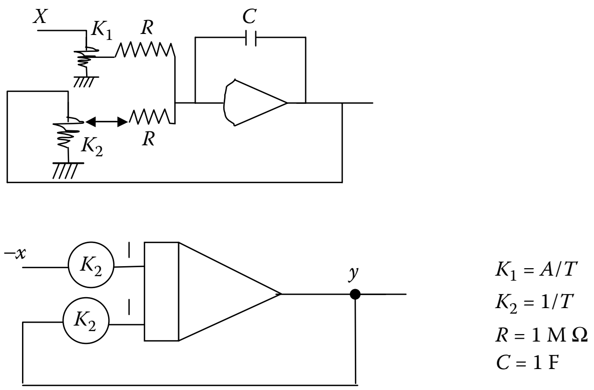


FIG. 1.3t
This alternate circuit also simulates a first-order transfer function A/(1 + sT).

one can connect two first-order system circuits (which are shown in Figure 1.3s) in cascade, but such a circuit cannot be used to simulate underdamped systems.

In order to simulate a second-order system of the form

$$G(s) = Y/X = k\left(s^2 + 2\xi w_n s + w_n^2\right) \qquad \mathbf{1.3(26)}$$

where $\xi < 1$, then the circuit shown in Figure 1.3u can be used. To obtain this circuit, one can first rewrite the transfer function into a differential equation form, using the following steps.

Cross-multiplying Equation 1.3(26)

$$s^2Y + 2\xi w_n sY + w_n^2 Y = kX \qquad \mathbf{1.3(27)}$$

and replacing s by d/dt, one obtains:

$$\frac{d^2}{dt^2}y + 2\xi w_n \frac{dy}{dt} + w_n^2 y = kx \qquad \mathbf{1.3(28)}$$

Keeping d^2/dt^2

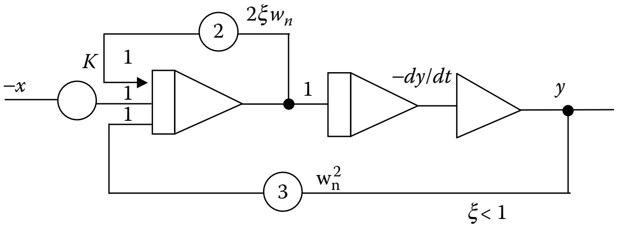


FIG. 1.3u
This circuit simulates an underdamped second-order system.

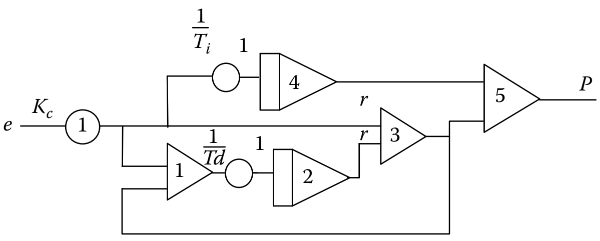


FIG. 1.3v
Analog simulation of a PID controller.

on the left-hand side and shifting all other terms to the right and multiplying the whole equation by −1, we obtain:

$$-\frac{d^2}{dt^2}y = 2\xi w_n \frac{dy}{dt} + w_n^2 y - kx \qquad \mathbf{1.3(29)}$$

For Equation 1.3(29), the circuit is shown in Figure 1.3u.

Simulation of PID Controller The transfer function that represents a proportional, integral, and derivative (PID) controller is:

$$\frac{P(s)}{E(s)} = K_c\left[\frac{1 + T_D s}{1 + T_D s/r} + \frac{1}{T_1 s}\right] \qquad \mathbf{1.3(30)}$$

When r is made very large, the transfer function reduces to

$$\frac{P(s)}{E(s)} - = k_c\left[1 + T_D s + \frac{1}{T_1 s}\right] \qquad \mathbf{1.3(31)}$$

This is the transfer function of an ideal PID controller.[4] The analog computer circuit that simulates it is shown in Figure 1.3v.

Lag Functions

A common element in the process and plant simulations is the lag, defined by the transfer function (in Laplace notation) as:

$$H(s) = \frac{Y(s)}{X(s)} = \frac{K}{1 + \tau s} \qquad \mathbf{1.3(32)}$$

where K is the steady-state gain and τ is the time constant. In general, the gain and time constant will vary with other simulation variables, such as flows, temperatures, and pressures. A generalized circuit for the lag function is shown at the top of Figure 1.3w. Other versions of the lag are also shown in the figure with different aspects and advantages, such as independent adjustment of gain and time constant, bipolar output, and simplicity and independent adjustment (but poorly scaled in many cases).

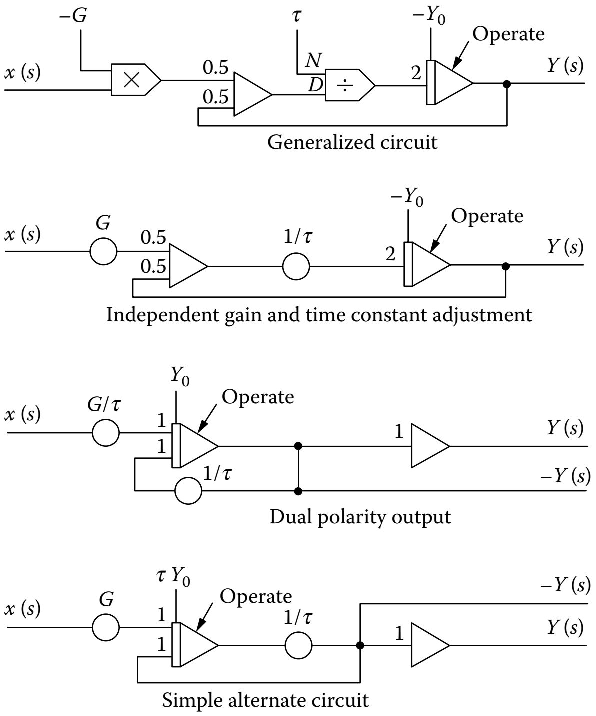


FIG. 1.3w
The lag function.

Lead-Lag Element Another commonly used element is the lead-lag, for which the transfer function is:

$$H(s) = K\frac{s}{1+\tau s} \tag{1.3(33)}$$

or

$$H(s) = K\frac{1+\tau_1 s}{1+\tau_2 s} \tag{1.3(34)}$$

Figure 1.3x shows a generalized circuit for the transfer function described in Equation 1.3(33), with variable gain K and time constant τ.

The class of lead-lag transfer functions described by Equation 1.3(33) may be simulated with circuits of the type shown in Figure 1.3y. The generalized arrangement used for individual time constants and gain variations by signal inputs often results in poor integrator scaling, especially for large τ_2 values. The common circuit permits the ratio between τ_1 and τ_2 to be adjusted over a wide range.

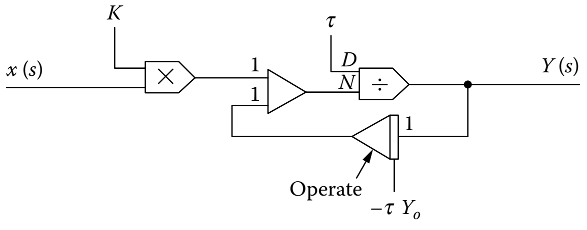


FIG. 1.3x
The lead-lag function for Equation 1.3(33).

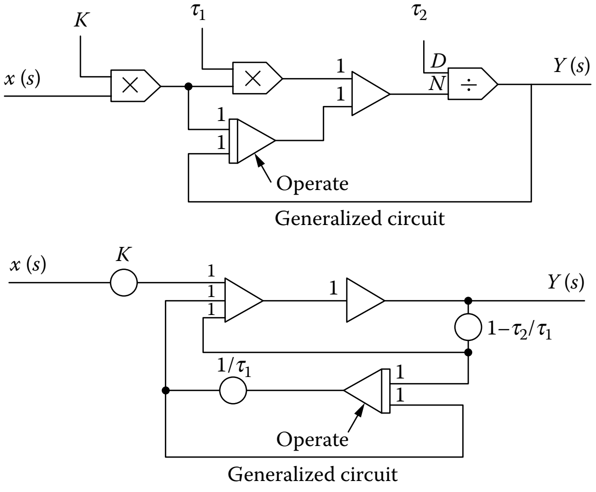


FIG. 1.3y
The lead-lag function for Equation 1.3(34).

Delays, Transport Lags In many analog computer simulations, a signal delay effect, or simulation of a transport lag, such as the flow of fluids through pipes or the movement of material on a conveyor, is desired. Delays or dead times such as this often represent important dynamic factors and influence the stability and operation of the plant or process. In general, there is no true analog computer "model" for a transportation delay, and for most applications some form of approximation of the delay function must suffice.

The most commonly known approximations are the Padé functions. These are often unsatisfactory in process and plant simulations, and other empirically determined functions are used in their place. Figure 1.3z shows a delay approximation, for which the time domain responses are better than those of the Padé functions and give more stable, damped, and consistent performance. The delay approximation is the low-pass type, with the transfer function:

$$H_2(s) = \frac{3.45 - 0.345(\tau s)}{3.45 + 3.18(\tau s) + (\tau s)^2} \tag{1.3(35)}$$

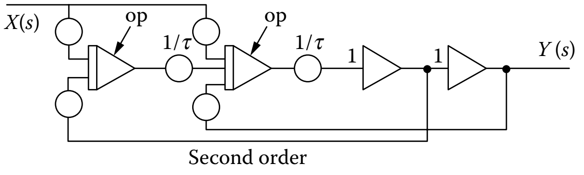


FIG. 1.3z
Approximation of the transport lag.

In Figure 1.3z, the transportation delay time (τ) can be varied by adjusting the coefficient potentiometers. If the delay time is a function of a simulation variable, such as the flow rate in a pipe, multipliers must replace the potentiometers. For the circuit described, the range of variation is generally limited to less than 100 : 1, without rescaling or rearranging the gain distributions. The shortest delay is determined by the highest input gains available (or by gain–integrator speed combinations).

Control System Simulation

The simulation of instrumentation control systems includes measuring or detecting devices, the controllers, and the actuating or manipulating output elements. Most measuring devices may be simulated by one or more simple lags for dynamic representation, with added nonlinearities for device characteristics, sensitivity, or operating point modeling.

For special functions, such as logarithmic input–output relationships (such as pH measurements), the general-purpose function-generating units may be used or the functions generated by implicit techniques. When the measuring device puts out a discontinuous (pulse-type) signal, signal-generating circuits may be used, as discussed later in this section.

Interacting PID Controller A direct, three-mode (PID) controller has the transfer function:

$$M(s)=\frac{100}{\text{PB}}\left[1+\frac{1}{T_i s}\right]\left[1+\frac{T_d s}{1+T_0 s}\right][R(s)-C(s)] \qquad \textbf{1.3(36)}$$

where $M(s)$ = controller output in Laplace representation; $C(s)$ = measured (controlled) input variable; $R(s)$ = reference (set point); PB = proportional band (in percent); T_i = reset (integration) time; T_d = derivative (rate) time; and T_0 = stabilizing (filtering) time constant.

The filter time constant is usually made as small as possible and is often a fraction of the derivative time T_d (such as $T_d/16$).

Figure 1.3aa shows simulation circuits for direct and reverse acting three-mode controllers, with the initial value M_0 corresponding to the controller output when it was in "manual."

The controller transfer function in Equation 1.3(36) is of the interacting type, which is the common case for most industrial applications.

Noninteracting PID Controller A mathematically noninteracting controller is expressed by:

$$M(s)=\left[\frac{100}{\text{PB}}+\frac{1}{T_i s}+\frac{T_d s}{1+T_0 s}\right][R(s)-C(s)] \qquad \textbf{1.3(37)}$$

Here all the three modes can be independently adjusted.

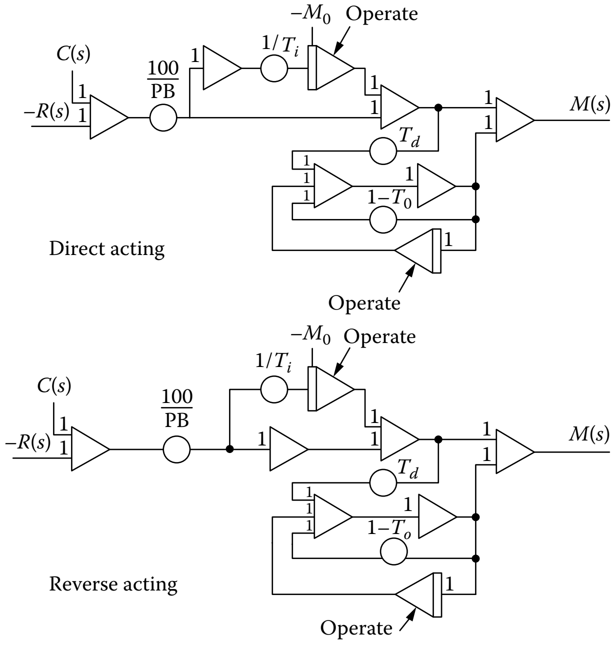


FIG. 1.3aa
Interacting PID controller described by Equation 1.3(36).

In general, the proportional band should have an *overall* adjusting effect, which leads to the practical noninteracting controller with the transfer function:

$$M(s)=\frac{100}{\text{PB}}\left[1+\frac{1}{T_i s}+\frac{T_d s}{1+T_0 s}\right][R(s)-C(s)] \qquad \textbf{1.3(38)}$$

This circuit arrangement is shown in Figure 1.3bb.

More modern controllers often limit the derivative action so that it will respond only to changes in the measured variable and not to rapid set point changes, which could upset the process. This characteristic is described in the modified transfer function:

$$M(s)=\frac{100}{\text{PB}}\left[1+\frac{1}{T_i s}\right]\left[R(s)-\left(1+\frac{T_d s}{1+T_0 s}\right)C(s)\right] \qquad \textbf{1.3(39)}$$

which is of the interacting kind.

Hybrid Simulation

A hybrid controller can simulate direct digital control of a process. In that case the individual PID control algorithms can be simulated by the digital computer, while the process is simulated using an analog computer, and the two are interconnected for system analysis and design.

Simulation of Direct Digital Control Direct digital control may be simulated with a controller circuit as shown in Figure 1.3cc, where the measured variable $C(s)$ and the set point $R(s)$ are sampled at fixed intervals (Δt) and held in

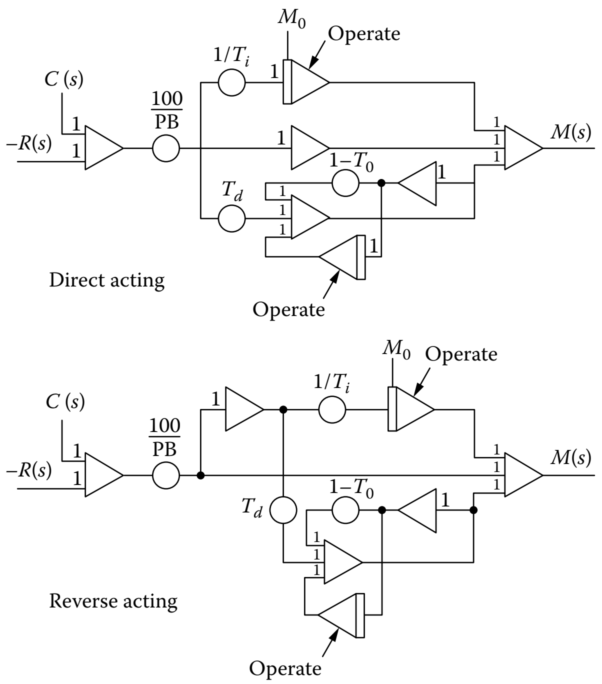


FIG. 1.3bb
Practical noninteracting controller described by Equation 1.3(38).

track-store units. The controller algorithm for this case is given by Equation 1.3(40):

$$\Delta m = \frac{100}{\text{PB}}\left[\left(1+\frac{\Delta t}{T_i}\right)(r_n - c_n) - r_{n-1} + c_{n-1} - \frac{T_d}{\Delta t}(c_n - 2c_{n-1} + c_{n-1})\right] \qquad \textbf{1.3(40)}$$

with Δm being the change in the output, which is computed in the time interval. This control corresponds to the one given in Equation 1.3(39), with derivative action responding to the measured variable only. The track-store operation is controlled by the pulse P occurring once during each time interval Δt to produce an output pulse Δm of amplitude (height) corresponding to the desired change in the control variable.

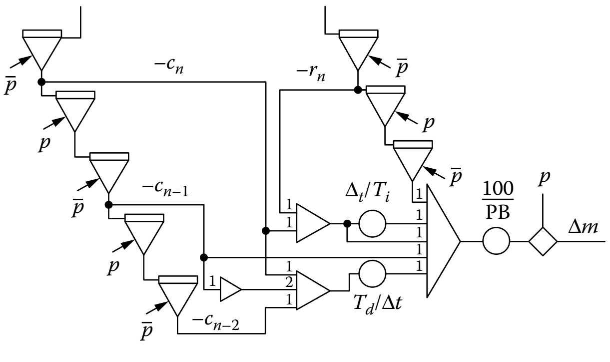


FIG. 1.3cc
Analog simulation of direct digital control based on Equation 1.3(40).

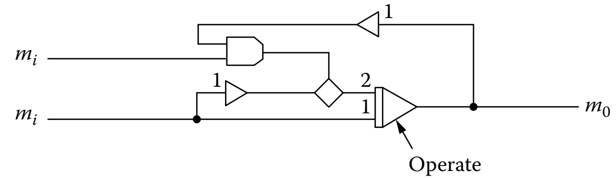
Simulation of Control Valves The most important manipulating output element is the control valve, which can have several distinct functional aspects in an analog computer simulation. The flow characteristics can be linear, equal percentage, quick opening, or butterfly type.

Except for the linear valve, the flow characteristics must be generated by a special analog computer circuit (by implicit techniques), or programmed in a function-generating unit, for a true representation (see Chapter 3 for details). Additional effects such as velocity limiting or backlash in valve stem movements must be included in the model.

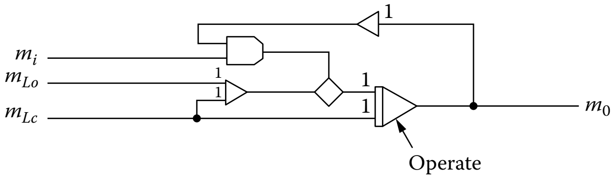
The dynamic performance of the control valve may be represented by a time lag of first or second order, with a limited velocity in stem movement. The time lags are simulated by circuits that were described under transfer functions, lags, and lead-lags. Velocity limiting may be expressed by the equation

$$\frac{dm_0}{dt} = \text{LOW}\left(m_L; \frac{dm_i}{dt}\right) \qquad \textbf{1.3(41)}$$

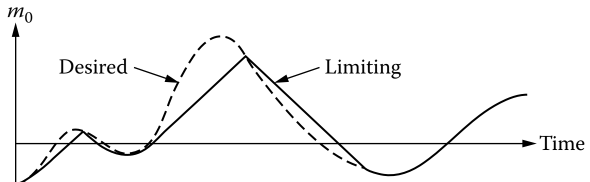
where m_0 is the stem position, m_L the velocity limit, and m_i the input stem position of an ideal, unconstrained valve. The simulation circuit for velocity limiting is shown in Figure 1.3dd.



Common limiting



Direction-dependent limiting



Response of velocity limited system

FIG. 1.3dd
The analog simulation circuit and response of a velocity-limited system.

References

1. Coughanour, D. R., and Koppel, L.B., *Process Systems Analysis and Control,* New York: McGraw-Hill, 1965.
2. Jackson, A. S., *Analog Computation,* New York: McGraw-Hill, 1960.
3. Van Valkanburg, M. E., *Network Analysis,* Eaglewood Cliffs, NJ: Prentice Hall, 2000.
4. Ogata, K., *Modern Control Engineering,* New Dehli: Prentice Hall of India, 2002.

Bibliography

Adams, J., et al., "Mathematical Modeling of Once-Through Boiler Dynamics," IEEE Winter Power Meeting, Paper 31 TP65-177, 1965.

Ashley, J. R., *Introduction to Analog Computation,* New York: John Wiley & Sons, 1963.

Box, G., and Draper, N., *Empirical Model-Building and Response Surface,* New York: Wiley, 1987.

Brignell, J., and White, N., *Intelligent Sensor Systems,* Rev. ed., Bristol, UK: Institute of Physics Pub., 1996.

Cavicchi, T. J., *Digital Signal Processing,* New York: Wiley, 2000.

Dyer, S. A. (Ed.), *Survey of Instrumentation and Measurement,* New York: Wiley, 2001.

Ebert, K., et al., *Computer Applications in Chemistry,* Weinheim, Germany: VCH, 1989.

Eren. H., "Instruments," *Survey of Instrumentation and Measurement,* S. A. Dyer, 2001.

Fowler, K.R., *Electronic Instrument Design — Architecting for the Life Cycle,* New York: Oxford University Press, 1996.

Garrett, P. H., *Advanced Instrumentation and Computer I/O Design: Real-Time System Computer Interface Engineering,* New York: IEEE Press, 1994.

Grover, D., and Deller, R. J., *Digital Signal Processing and the Microcontroller,* Upper Saddle River, NJ: Prentice Hall PTR, 1999.

Hamell, E., "Modeling and Simulation," *InTech,* November 1982.

Haykin, S., and Kosko, B., (Eds.), *Intelligent Signal Processing,* New York: IEEE Press. 2001.

Howe, R. M., *Design Fundamentals of Analog Computer Components,* New York: Van Nostrand, 1961.

James, M. L., Smith, G. M., and Wolford, J. C., *Analog Computer Simulation of Engineering System,* Scranton, PA: International Textbook Company, 1966.

Jenness, R. R., *Analog Computation and Simulation: Laboratory Approach,* Boston: Allyn and Bacon, 1965.

Kasparek, S., *Computer Graphics and Chemical Structures,* New York: Wiley, 1990.

Korn, G., *Interactive Dynamic Systems Simulation,* New York: McGraw-Hill, 1989.

Lupfer, D. E., "Analog Computer Cuts Distillation Costs," *ISA Journal,* August 1966.

MacKay, D. M., and Fisher, M. E., *Analog Computing at Ultra-High Speed,* New York: John Wiley & Sons, 1962.

Madisetti, V. K., and Williams, D. B., *The Digital Signal Processing Handbook,* Boca Raton, FL: CRC Press, 1998.

McDonald, J. P., and Kwatny, H. G., "Design and Analysis of Boiled Turbine-Generator Controls Using Optimal Linear Regulatory Theory," *JACC*, 1972.

McDonald, J. P., and Kwatny, H. G., and Space, J. H., "A Non-Linear Model For Reheat Boiler–Turbine–Generator Systems, Parts I and II," *JACC,* 1971.

Mostafa, E. S., "An Operator Training Simulator," *Instrumentation Technology,* January 1982.

Pallás-Areny, R., and Webster, J. G., *Analogue Signal Processing,* New York: John Wiley & Sons, 1999.

Pallás-Areny, R., and Webster, J. G., *Sensors and Signal Conditioning,* 2nd ed., New York: John Wiley & Sons, 2001.

Paton, B. E., *Sensors, Transducers, and LabView,* Upper Saddle River, NJ: Prentice Hall PTR, 1999.

Peterson, G. R., *Basic Analog Computation,* New York: Macmillan, 1967.

Randy, F., *Understanding Smart Sensors,* 2nd ed., Boston: Artech House, 2000.

Rogers, A. E., and Connolly, T. W., *Analog Computation in Engineering Design,* New York: McGraw-Hill Book Co., 1960.

Scrivers, D. B., "Special Purpose Analog Computers," *ISA Journal,* July 1966.

Simpkin, L. J., and Wooden, L. F., "Use of Analog Computations of Power Plant Automatic Control Systems," IS, The Ninth National Power Instrumentation Symposium, Denver, CO, May 1964.

Tomovic, R. J., and Karplus, W. J., *High Speed Analog,* New York: John Wiley & Sons, 1962.

Vassos, B., and Ewing, G., *Analog and Digital Electronics,* New York: Wiley, 1985.

Webster, J. G. (Ed.), *The Measurement, Instrumentation, and Sensors Handbook,* Boca Raton, FL: CRC Press, 1999.

1.4 Electronic vs. Pneumatic Instruments

H. L. FELDMAN (1969) **R. K. KAMINSKI** (1985) **B. G. LIPTÁK** (1995, 2005)

The debate concerning the relative merits of electronic versus pneumatic instrumentation took place in the middle of the 20th century. As can be seen from the references for this section, most of the arguments were published during that period, and little additional knowledge has been obtained in more recent decades. Although the debate has been concluded, and the process control industry has decisively moved into the electronic and digital age, knowing the arguments — the advantages and disadvantages of pneumatics versus electronics — can still be educational. For this reason, and because pneumatics are still used in the heating, ventilation, and air conditioning industry and in the control valve portion of control loops, this section has not been deleted from this handbook.

FACTORS TO BE CONSIDERED

The main advantages of pneumatic systems over electronic instruments lies in their safety in hazardous locations and in the availability of qualified maintenance personnel. However, electronic systems are favored when computer compatibility, long transmission distances, or very cold ambient conditions are expected, or when long-term maintenance is to be minimized (Figure 1.4a).

Perhaps "Pneumatics versus High Technology" would have been a better title for this section. If a process plant uses digital electronic control, it is very difficult to justify the use of pneumatics. Naturally, when evaluating the advisability of pneumatics, one must evaluate the options of analog electronics, digital controllers with or without computers, distributed control systems, and programmable logic controllers. The relative merits of analog vs. digital controls are covered in Section 1.1 of this volume. In the discussion here, references are provided for further reading on the subject.

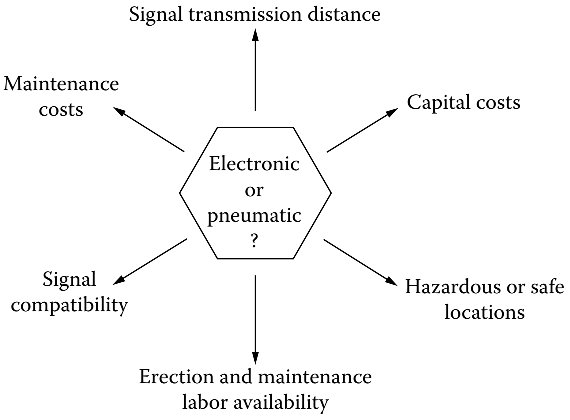


FIG. 1.4a
Electronic versus pneumatic systems.

THE OLD DEBATE

In the debate some decades ago, some authors promoted electronics[1–4] and others pneumatics.[5–9] Table 1.4b summarizes the views on both pneumatic and electronic instruments that are expressed in the first nine referenced articles.

In the area of maintenance, the referenced literature[1–18] does not present convincing evidence that electronics guarantees reduced maintenance costs, but it does suggest that electronic instrumentation is easier to maintain and that the life of electronic instruments is longer. This is because of the wear on the pivot points, the change in spring elasticity and diaphragm stretch in pneumatic designs. Maintenance publications, such as those of ISA, have contributed to making maintenance easier by promoting modular component designs and self diagnostics.

The argument that pneumatic instruments need dry, clean air but electronic systems do not is not correct because many electronic control system include I/P (current-to-air) converters and electro-pneumatic valve positioners, which also require dry and clean air. Deuschle,[14] in his detailed cost comparison, points out that the capacity required for the air compressor serving an electronic control system is about 60% of that required for a pneumatic system. For a specific application, the rating approach proposed by Tompkins[17] can be considered for comparing the choice of pneumatics versus electronics.

HVAC DAMPER ACTUATORS

The relative performance of electronic and pneumatic damper actuators will be discussed below.

Pneumatic Actuators

Pneumatic dampers are usually throttled over a 5 PSI spring range such as from 8 to 13 PSIG. Over this range, the damper will move from the closed to the open position if there is no

TABLE 1.4b
Viewpoints Expressed on Electronics versus Pneumatics†

Features	*Electronics Superior*	*Pneumatics Superior*	*About Equal*	*Remarks*
Lower initial hardware cost		11,12,13,14,15,16,17*		
Lower installation cost	10, 12, 13, 14		11	Ref. 13: Cost for either varies all over the map
Lower total installed cost			12, 14, 15, 16*	
Simpler system design		13, 15, 16*		
Shorter check-out and start-up		11*		
Shorter training period		11*		
Higher dependability (reliability)			11, 15, 16*	
Less affected by corrosive atmospheres		13, 14, 17*		Ref. 14: Air acts as a purge
Lower maintenance	12, 14, 17	11	*	Ref. 12 cites user experience
More compatible with control valves		13, 14, 18*		
Greater accuracy	11, 12, 16*			
Superior dynamic response	11, 12, 13, 14, 15, 16, 17, 18*			Ref. 13: Normally, the fastest response is not necessary
Better suited for long transmission distances	11, 12, 13, 14, 15, 17, 18*			
Superior computer capability	11, 12, 13, 14, 17, 18*			Ref. 13: This is the primary reason for selecting electronic instruments

† Numbers refer to the references listed at the end of this section.
* An asterisk marks the preference of this author.

friction and no process load, which resist the movement of the damper. Usually both are present.

Hysteresis

Resistance forces (F_R) are present, not only because there is friction between the stem and the hole through which it passes but because there also is linkage stiffness and axle bind at the rotary joints. Therefore, because the air is both the control signal and the power source, its pressure must rise beyond that required to break free and initiate movement, which results in overshoot. On the return stroke, the air signal pressure must drop below the spring force (F_S) before the spring will overcome friction and start returning the damper. On an actuator with a 4” stroke, this can result in a hysteresis of 1”.

Spring Range Shift

When a process load (F_L) exists on the damper, it will shift the throttling range of the damper. If the actuator area is 15 in.^{2}, each pound of air signal pressure will generate 15 lb of added force on the diaphragm (F_A). Therefore, if the process load (F_L) is, say, 45 lb, that load will shift the throttling range of the damper by 3 PSIG. Depending on the failure position of the damper, either the lower or the upper limit of the range will be shifted. Consequently, if the process load (F_L) is 45 lb, the throttling of a fail closed (FC) damper will not occur between 8 and 13 PSIG, but the damper will throttle between 5 and 13 PSIG. Similarly a fail open (FO) damper will throttle between 8 and 16 PSI.

Direct Coupled Electronic Actuator

When a 2- to 10-VDC signal, generated by a 4- to 20-mADC signal passing through a 500-ohm resistor, directly operates the damper, performance is improved. This is because the direct coupled electronic actuator has a resolution of 160:1, allowing a positioning accuracy that can be as high as 120:1,[28] which is much superior to that of a pneumatic actuator. This actuator uses an internal potentiometer to verify its position with respect to the control signal. Consequently, it is unaffected by friction and process load forces and obeys the relationship between control signal and internal potentiometer within +/–0.05 VDC.

Pneumatic vs. Electronic

If the pneumatic actuator has no positioner, the problems of hysteresis and spring range shift will make its performance inferior to that of a direct coupled electronic design. If the pneumatic actuator is furnished with a positioner, its performance is improved. The amount of improvement depends on the quality of the positioner, which will determine how close the loop performance will approach that of the electronic actuator. In any case, it is not likely that the positioning accuracy of a pneumatic actuator will approach or reach that of an electronic one.

The relative costs of the installed pneumatic and electronic actuators are a function of the quality of the pneumatic positioner used. If the positioner is of reasonably high quality, the installed costs will be similar.

CONTROL SIGNALS

Putting together a pneumatic system can be relatively simple. The 3- to 15-PSIG (0.2- to 1.0-bar) air pressure signal range permits the teeing in of different manufacturers' equipment with very few problems. Other pressure ranges are seldom used, except when there is reason to amplify the range, such as to 6 to 30 PSIG.

Much has been written about pneumatic transmission lags.[19–23] In this regard the electronic system is superior; however, pneumatics systems have been successfully operated with 250- to 500-ft (75- to 150-m) lines. When boosters are used, distances of 1000 ft (300 m) and more can be considered.[21, 22] The longer distances can be attained by specifying such requirements as higher air capacities and the use of volume boosters and 3/8 in. (9.38 mm) OD tubing. In order to reduce transmission distances, field-mounted controllers can also be used. According to Buckley,[22] serious limitations on performance are caused by:

1. Control valves without positioners
2. The use of 1/4 in. (6.25 mm) rather than 3/8 in. (9.38 mm) OD tubing
3. Restrictions in manual/automatic switch blocks and plug-in manifolds
4. Inadequate valve positioner air capacities
5. Single-acting (instead of double-acting) positioners on cylinder operators
6. Inadequate field air supplies
7. Multiple process lags

The most common analog electronic control or transmission signal is 4 to 20 mA DC. Considerable engineering must go into all-electronic systems because of impedance restrictions and polarity, power supply, shielding, and grounding requirements. In digital control systems, the buses and networks require substantial engineering, particularly in the area of the interfacing of different suppliers' products.

The transmission range of the electronic system can be a mile or more with no time lag. However, this feature is important only in a minority of control loops.

With regard to electrical noise, pneumatic instruments are, of course, immune. Electronic systems do experience problems if shielding and grounding are inadequate. In recent years, many manufacturers have designed instruments protected against RFI (radio frequency interference).

CONVERTERS

A pneumatic system does not need additional hardware to operate an air-operated control valve, whereas an I/P converter must be installed if an electronic loop is to operate a pneumatic control valve. Because many transmitter outputs are inherently electronic, pneumatic loops also require I/P converters for applications involving such transmitters on temperature, flow, or analytical measurements.

Electronic systems are superior for data loggers, computers, distributed control, and/or programmable controller applications. The pneumatic system would require I/Ps, P/Is, and pneumatic scanning devices for these applications.

ELECTRICAL SAFETY

In many process industries, some locations are hazardous because of the presence of flammable gases, combustible dusts, and ignitable fibers. Therefore, equipment in hazardous areas must be designed so that it will not cause a fire or explosion (see Section 7.1 in Volume 1 of this handbook for details). Manufacturers and users agree that pneumatic instruments are inherently safe to use in installations in hazardous areas. On the other hand, analog and digital electronic instruments require special protection features if they are to operate in a hazardous environment.

If the area is Class I explosion-proof, where explosive or ignitable mixtures (in air) of flammable gases or vapors constitute the hazard, the electronics must meet different requirements. Divisions 1 and 2 are defined below:

Division 1: Location is likely to have flammable mixtures present under normal conditions or the operating process equipment is likely to require frequent maintenance
Division 2: Location is likely to have flammable mixtures present only under abnormal conditions

The initial and installed costs for electronic systems will be a function of the electrical classifications and of the selected approaches, which are used to meet the requirements of the

National Electrical Code. The choices include explosion-proof housings, purging, and intrinsically safe designs. For Division 2, nonincendiary, nonarcing, and hermetically sealed equipment can also be considered. (Refer to Chapter 7 in the *Measurement* volume and to References 24 to 26 in this section for more information.)

THE MODERN DEBATE

In selecting a control system one must examine the advantages and drawbacks of both analog and digital alternatives,[27] which was the topic of Section 1.1 in this chapter. Digital controls are compatible with computers, distributed control systems, programmable controllers, and digital controllers.

References

1. Mathewson, C. E., "Advantage of Electronic Control," *Instruments and Automation,* February 1955, pp. 258–269.
2. "Sun Oil Opts for Electronic Instrumentation," *Canadian Controls and Instrumentation,* February 1968, pp. 29–31.
3. "CPI Swing to Electronic Controls," *Chemical Week,* April 13, 1968, p. 53.
4. Hileman, J. R., "Practical Aspects of Electronic Instruments," *Instrumentation Technology,* May 1968, pp. 43–47.
5. "A Plug for Pneumatics," *Chemical Week,* December 14, 1968, pp. 77 and 79.
6. Maas, M. A. "Pneumatics: It's Not Dead Yet," *Instruments and Control Systems,* November 1974, pp. 47–52.
7. Martin, R. L., "Simple Solutions to Control Problems," *Chemical Engineering,* May 22, 1978, pp. 103–111.
8. Bailey, S., "Pneumatic Process Controls: An Eternal Case of Borrowed Time," *Control Engineering,* October 1978, pp. 39–42.
9. Farmer, E., "Pneumatics in a Digital World," *Instruments and Control Systems,* March 1979, pp. 31–35.
10. Teed, J. T., "Comparative Installation Costs for Miniature Pneumatic and Electronic Instruments," Paper 1-LA61, Fall Instrument-Automation Conference, Instrument Society of America, 1961.
11. Palmer, O. J., "Instrumentation — Pneumatic Versus Electronic, Which Way To Buy?" Paper 5.4-4-64, 19th Annual Conference, Instrument Society of America, 1964.
12. Savage, H. L., "Electronic vs. Pneumatic Control Systems," unpublished paper, 1964.
13. Lynch, B., "Comparison of Electronic and Pneumatic Control Systems," Paper presented at the 6th Annual Maintenance Management Symposium, Instrument Society of America, 1971.
14. Deuschle, R., "Pneumatic? Electronic? Which to Buy?" *Instruments and Control Systems,* February 1975, pp. 51–54.
15. Lommatsch, E. A., "Pneumatic versus Electronic Instrumentation," *Chemical Engineering,* June 21, 1976, pp. 159–162.
16. Hordeski, M., "When Should You Use Pneumatics, When Electronics?" *Instruments and Control Systems,* November 1976, pp. 51–55.
17. Tompkins, J., "Pneumatic or Electronic Instrumentation?" *Instruments and Control Systems,* January 1979, pp. 55–57.
18. Marcovitch, M. J., "Whither Pneumatic and Electronic Instrumentation?" *Chemical Engineering,* Deskbook Issue, October 13, 1979, pp. 45–49.
19. Harriott, P., *Process Control,* New York: McGraw-Hill, 1964, pp. 204–210.
20. Warnock, J. D., "How Pneumatic Tubing Size Influences Controllability," *Instrumentation Technology,* February 1967, pp. 37–43.
21. Buckley, P. S., "Dynamic Design on Pneumatic Control Loops, Part I: Analysis by Segments," *Instrumentation Technology,* April 1975, pp. 33–40.
22. Buckley, P. S., "Dynamic Design of Pneumatic Control Loops, Part II: Application Principles and Practices," *Instrumentation Technology,* June 1975, pp. 39–42.
23. Shinskey, F. G., *Process Control Systems,* 2nd ed., New York: McGraw-Hill, 1979, pp. 38–40.
24. Magison, E. C., *Electrical Instruments in Hazardous Locations,* 3rd ed., Research Triangle Park, North Carolina: Instrument Society of America, 1978.
25. Oudar, J., "Intrinsic Safety," *Measurements and Control,* October 1981, pp. 153–160.
26. Garside, R., *Intrinsically Safe Instrumentation: A Guide,* Research Triangle Park, North Carolina: Instrument Society of America, 1982.
27. Kaminski, R. K., "The Basic Control Choice," *Instrumentation Technology,* July 1973, p. 41.
28. Belimo Aircontrols Inc., *Electronic vs. Pneumatic Actuation,* Danbury, CT, 2003.

Bibliography

Bailey, S. J., "Process Control 1979: A Year of Drastic Change, and More To Come," *Control Engineering,* October 1979.

Bailey, S. J., "The Programmable Controller Is Finding Its In-Line Niche," *Control Engineering,* February 1980.

Bowron, J., "DCS vs. PLC — Enough Already!" *InTech,* September 1992.

Cocheo, S., "How to Evaluate Distributed Computer Control Systems," *Hydrocarbon Processing,* June 1981.

"The Configurations of Process Control: 1979," *Control Engineering,* March 1979.

D'Abreo, F. A., "Replacing Installed Process Control Computers," *Instrumentation Technology,* July 1982.

Dartt, S. R., "Distributed Digital Control — Its Pros and Cons," *Chemical Engineering,* September 6, 1982.

Giles, R. G., and Bullock, J. E., "Justifying the Cost of Computer Control," *Control Engineering,* May 1976.

Hall, J., and Kuhfeld, R., "Industry Leaders Look at 1982," *Instruments and Control Systems,* January 1982.

Hickey, J., "The Interconnect Dilemma," *Instrument and Control Systems,* December 1981.

Kompass, E. J., "The Long-Term Trends in Control," *Control Engineering,* September 1979.

Ledgerwood, B. K., "Trends in Control," *Control Engineering,* May 1982.

McAdams, T. H., "Digital Control at Half the Cost of Analog," *Instruments and Control Systems,* April 1977.

McNeil, E. B., "A User's View of Instrumentation," *Chemical Processing,* October 1982.

Merritt, R., "Large Control Systems: Big Price Tags but Even Bigger Returns," *Instrument and Control Systems,* December 1981.

Miller, T. J., "New PCs Exemplify Trends to Smaller Size and Distributed Network Compatibility," *Control Engineering,* January 1983.

Morris, H. M., "Operator Convenience Is Key as Process Controllers Evolve," *Control Engineering,* March 1981.

Neal, D., "The Datalogger Moves into Control," *Instruments and Control Systems,* November 1982.

Pluhar, K., "International Manufacturers Offer Many Choices in Process Control Systems," *Control Engineering,* January 1980.

Rapley, D. E., "How to Select Distributed Control and Data Systems," *Instrumentation Technology,* September 1980.

Sheridan, T. B., "Interface Needs for Coming Industrial Controls," *Instrumentation Technology,* February 1982.

Smith, D. R., "Is Your Control Computer Obsolete?" *Instruments and Control Systems,* February 1982.

"Special Report on Programmable Controllers," *Plant Services,* June 1982.

Stanton, B.D., “Designing Distributed Computer Control Systems,” *Hydrocarbon Processing,* June 1981.

Stanton, B.D., “Reduce Problems in New Control System Design,” *Hydrocarbon Processing,* August 1982.

Thurston, C. W., “Experience with a Large Distributed Control System,” *Control Engineering,* June 1979.

Warenskjold, G. A., “Microprocessor vs. Analog Control for the Nicholasville Water Treatment Facility,” Annual Conference, Instrument Society of America, 1978.

Williams, T. J., “Two Decades of Change — A Review of the 20-Year History of Computer Control,” *Control Engineering,* September 1977.

1.5 Human Engineering

M. C. TOGNERI (1985, 1995) **B. G. LIPTÁK** (2005)

THE CONTROL ROOM OF THE 21ST CENTURY

Human–machine interfaces (HMI) can allow a single operator to monitor thousands of control loops from a single operator's station. The HMI is the operator's window on the process that is being controlled. The information content of the early HMIs was that of the faceplates of analog controllers, while today's computer displays can provide not only data on the process parameters and alarms but can also furnish trend charts and historical data, in addition to generating reports and analyzing or diagnosing complex systems.

Modern control rooms and their equipment will be discussed in more detail in Chapter 4. They serve to simultaneously enhance efficiency and safety by providing advanced alarm systems, compact workstations, and large display panels (Figure 1.5a). Another essential feature of modern control systems is their ability to allow smooth growth, which is referred to as their scalability. This capability is very important in process control, because processing plants are ever expanding and therefore their control systems must also grow with them. A modular approach to operator stations can conveniently serve this purpose (Figure 1.5b), as will also be discussed in more detail in Chapter 4.

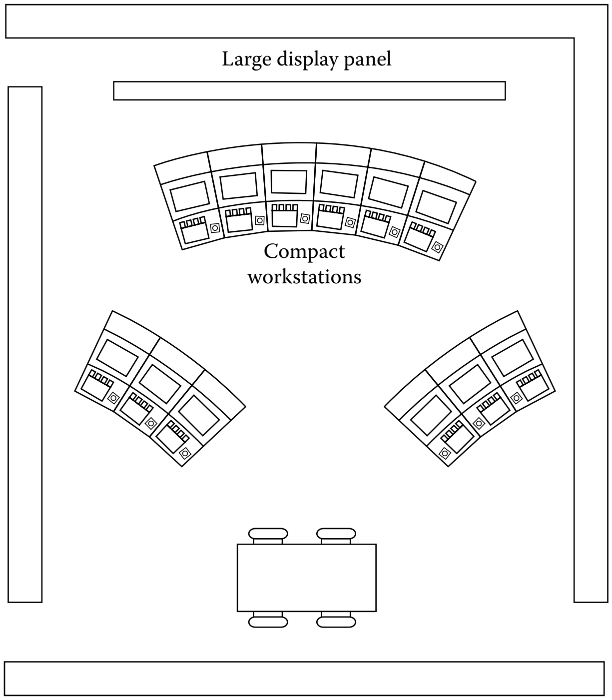


FIG.1.5a
The layout of a modern control room.

The value of these modern tools depends on the human operator's ability to use them. The hand, psychological characteristics, hearing, and color discrimination capability of the human operator must all be part of human engineering, which is the topic of this section.

INTRODUCTION

Human engineering, also known as engineering psychology, human factors engineering, and ergonomics, is probably the most basic form of engineering because the fashioning of any tool to suit the hand is the goal of human engineering. The discipline began as a feedback system in the sense that modification was left to future modification as needed instead of being made part of the original design. This approach sufficed when technological progress was sufficiently slow to allow the sequential development of several generations of improved

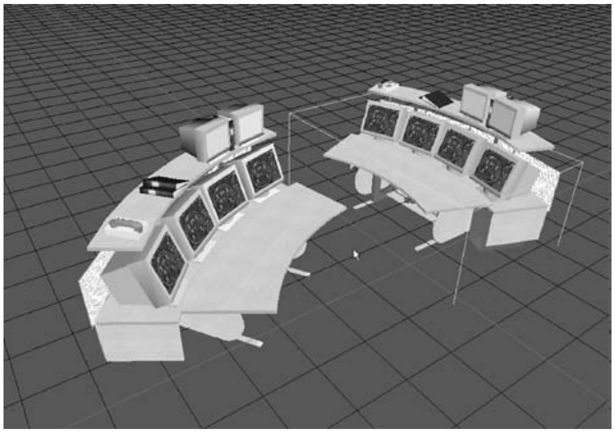


FIG. 1.5b
The growth of control rooms is served by the availability of modular operator's workstations.

equipment. The efficiency and cost of tools were such that improper original design was tolerable.

Current technology has supplied highly efficient, complex, and therefore expensive tools, the ultimate success of which rests with our ability to use them. Today's tools are required not only to fit the human hand but also to reinforce many physiological and psychological characteristics, such as hearing, color discrimination, and signal acceptance rates. The complexity and cost of tools make the trial-and-error feedback approach to improvement unacceptable.

World War II introduced the first attempt at technical innovation and the first serious effort to treat human engineering as a discipline. Postwar economic pressures have been almost as grudging of critical mistakes as war has been; thus, the role of human engineering has increased with technological progress. A few highlights of human engineering described in this section will aid the instrument engineer in evaluating the impact of this discipline in process plants.

Applicability extends from instruments as sophisticated as those used in the Apollo moon flight to those as prosaic as the kitchen stove.

Man–Machine System

The man–machine system is a combination of people and equipment interacting to produce a desired result. The majority of industrial control systems are indeed sophisticated man–machine systems.

Because the purposes of an instrument and control system are limited, a satisfactory definition is difficult to make. Subdividing the purpose into its constituent goals is an important first step in clarifying the goals. Many operators cannot supply much more information than to criticize that which is currently used or available. Learning from mistakes, although necessary, is not a desirable engineering approach.

The backbone of a man–machine system is a flow chart of man-related activities. This approach concentrates the design efforts on improving the capability of the operator to handle information and control — not just on improving instruments. Proper design criteria are obtained by asking what should be done, not what has been done. The machine should be adapted to the man. Two constraints exist:

1. The abilities of the average operator
2. The amount of information and control required by the process

Sight, hearing, and touch allow the operator to control information storage and processing, decision making, and process control. Information is processed by man and by machine in a similar manner (Figure 1.5c).

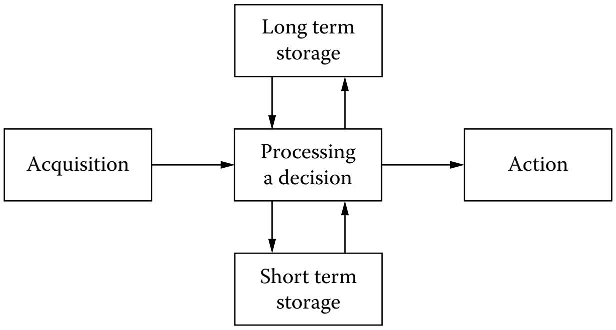


FIG. 1.5c
Information processing: man or machine.

Although people and machines process information similarly, their abilities to execute diverse functions vary greatly. Humans are better than machines at qualitatively processing information, an aptitude that develops on the basis of experience or judgment. The ability of humans to store enormous quantities of data (10^{15} bits) and to use these facts subconsciously in decision making allows them to reach decisions with a high probability of being right, even if all necessary information is not available. Machines are suited for quantitative processing and storage of exact data and provide rapid access. Machine decisions must be based on complete and exact data.

In a process control system, the differences are best applied by training the operator in emergency procedures and letting the machine store data on limits that, when exceeded, trigger the emergency. Table 1.5d summarizes the characteristics of men and machines. Since information is the primary quantity processed by a man–machine system, a brief account of informational theory is necessary.

Information Theory

The discipline of information theory defines the quantitative unit of evaluation as a binary digit or "bit." For messages containing a number of equal possibilities (n), the amount of information carried (H) is defined as:

$$H = \log_2 n, \qquad \textbf{1.5(1)}$$

and the contact of a pressure switch (PS) carries one bit of information. Therefore:

$$H_{\text{PS}} = \log_2^2 \text{(alternatives)} = 1 \qquad \textbf{1.5(2)}$$

The nature of most of the binary information in process control is such that probabilities of alternatives are equal. If this is not so, as with a pair of weighted dice, the total information carried is reduced. In an extreme case (if the switch were shorted), the probability of a closed contact is 100%, the probability of an open contact is 0%, and the total information content is zero.

CHARACTERISTICS OF MAN

Much has been learned about man as a system component since organized research in human factors was initiated during World War II. Anthropometric (body measurement)

TABLE 1.5d
Comparing the Characteristics of Man to Machine

Man Excels in	*Machines Excel in*
Detection of certain forms of very low energy levels	Monitoring (of both men and machines)
Sensitivity to an extremely wide variety of stimuli	Performing routine, repetitive, or very precise operations
Perceiving patterns and making generalizations about them	Responding very quickly to control signals
Detecting signals in the presence of high noise levels	Exerting great force, smoothly and with precision
Ability to store large amounts of information for long periods — and recalling relevant facts at appropriate moments	Storing and recalling large amounts of information in short periods
Ability to exercise judgment when events cannot be completely defined	Performing complex and rapid computation with high accuracy
Improvising and adopting flexible procedures	Sensitivity to stimuli beyond the range of human sensitivity (infrared, radio waves, and so forth)
Ability to react to unexpected low-probability events	Doing many different things at one time
Applying originality in solving problems, e.g., alternative solutions	Deductive processes
Ability to profit from experience and alter course of action	Insensitivity to extraneous factors
Ability to perform fine manipulation, especially when misalignment appears unexpectedly	Ability to repeat operations very rapidly, continuously, and precisely the same way over a long period
Ability to continue to perform even when overloaded	Operating in environments that are hostile to man or beyond human tolerance
Ability to reason inductively	

and psychometric (psychological) data are now available to the designer of control systems, as are many references to information concerning, for instance, ideal distances for perception of color, sound, touch, and shape. The engineer should keep in mind that generally the source of data of this sort is a selected group of subjects (students, soldiers, physicians, and so on), who may or may not be representative of the operator population available for a given project.

Body Dimensions

The amount of work space and the location and size of controls or indicators are significant parameters that affect operator comfort and output and require knowledge of body size, structure, and motions (anthropometry).

Static anthropometry deals with the dimensions of the body when it is motionless. Dynamic anthropometry deals with dimensions of reach and movement of the body in motion.

Of importance to the instrument engineer is the ability of the operator to see and reach panel locations while he or she is standing or seated. Figure 1.5e and Table 1.5f illustrate dimensions normally associated with instrumentation. The dimensions given are for the American male population. When operator populations other than American males are involved, the dimensions must be adjusted accordingly.

Information Capability

Human information processing involves sensory media (hearing, sight, speech, and touch) and memory. The sensory channels are the means by which man recognizes events (stimuli). The range of brightness carried by a visual stimulus, for example, may be considered in degrees and expressed in bits, and discrimination between stimulus levels is either absolute or relative. In relative discrimination, the individual compares two or more stimuli; in absolute discrimination, a single stimulus is evaluated.

Man performs much better in relative than in absolute discrimination. Typically, a subject can separate 100,000 colors when comparing them with each other; the number is reduced to 15 colors when each must be considered individually. Data in Table 1.5g deal with absolute discrimination encountered in levels of temperature, sound, and brightness.

In addition to the levels of discrimination, sensory channels are limited in the acceptance rates of stimuli. The limits differ for the various channels and are affected by physiological and psychological factors. A study by Pierce and Karling suggests a maximum level of 43 bits per second.

Several methods of improving information transmission to man include coding (language), use of multiple channels (visual and auditory), and organization of information (location of lamps; sequence of events). Simultaneous stimuli, similar signals with different meanings, noise, excessive intensity, and the resulting fatigue of the sensors — stimuli that approach discrimination thresholds — are all detrimental to the human information process.

The limit of the operator's memory affects information processing and is considered here in terms of long and short span. Work by H.G. Shulmann indicates that short-term memory

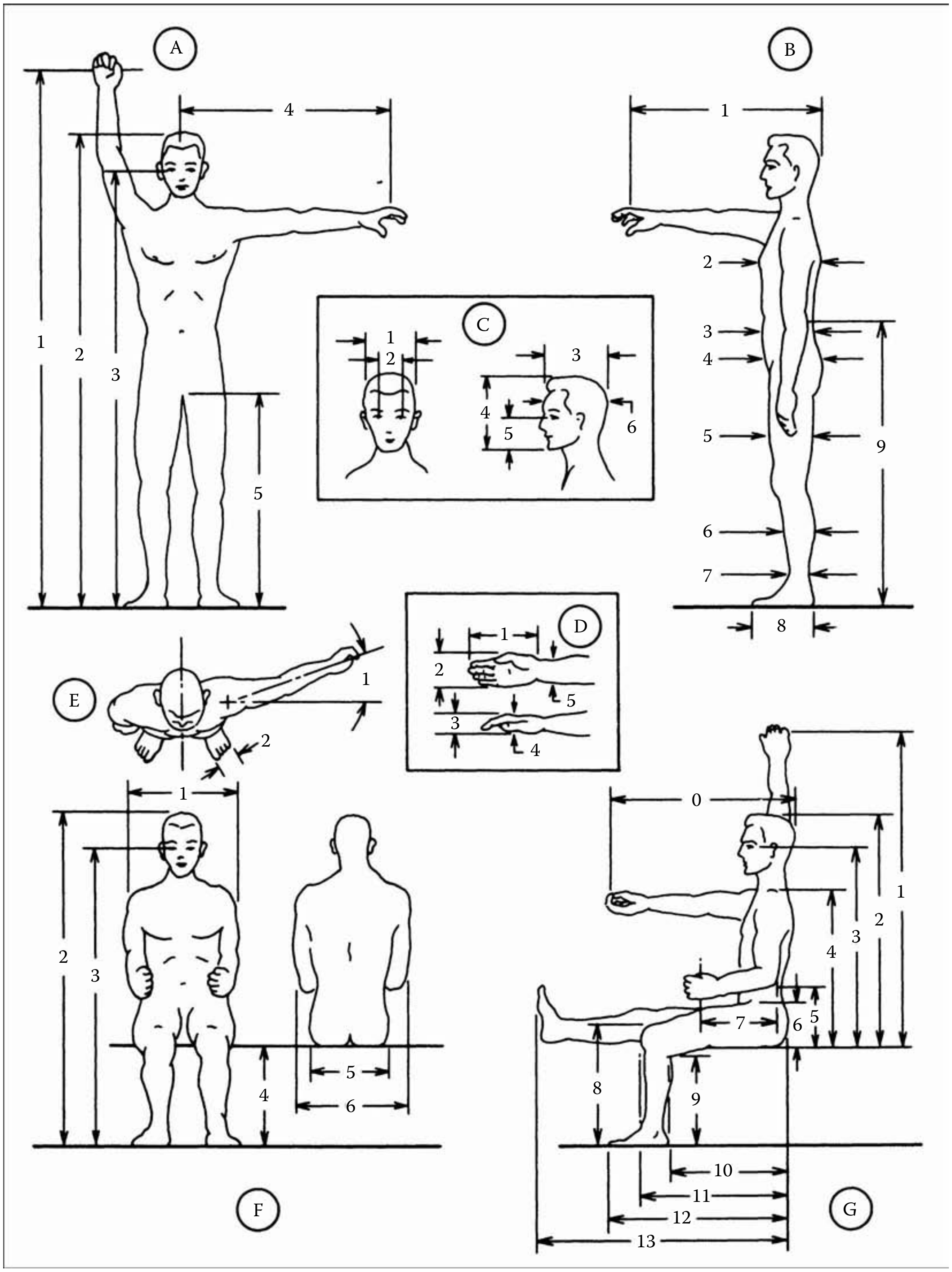


FIG. 1.5e
Body dimension chart. Dimensions are summarized in Table 1.5f.

uses a phonemic (word sounds) and long-term memory uses a semantic (linguistic) similarity. These approaches tend to make codes that are phonemically dissimilar less confusing for short-term recall and make codes that are semantically different less confusing for long-term recall. Training, for example, uses long-term memory.

Performance (Efficiency and Fatigue)

Motivation, annoyance, physical condition, and habituation influence the performance of primarily nonphysical tasks. These influences make the subject of efficiency and of fatigue controversial. Both the number and frequency of occurrence

TABLE 1.5f

Male Human Body Dimensions

Selected dimensions of the human body (ages 18 to 45). Locations of dimensions correspond to those in Figure 1.5e.

			Dimension in Inches (Meters) Except Where Noted	
		Dimensional Element	*5th Percentile*	*95th Percentile*
		Weight in pounds (kg)	132 (59.4)	201 (90.5)
A	1	Vertical reach	77.0 (1.9)	89.0 (2.2)
	2	Stature	65.0 (1.6)	73.0 (1.8)
	3	Eye to floor	61.0 (1.5)	69.0 (1.7)
	4	Side arm reach from center line of body	29.0 (0.7)	34.0 (0.8)
	5	Crotch to floor	30.0 (0.75)	36.0 (0.9)
B	1	Forward arm reach	28.0 (0.7)	33.0 (0.8)
	2	Chest circumference	35.0 (0.87)	43.0 (1.1)
	3	Waist circumference	28.0 (0.7)	38.0 (0.95)
	4	Hip circumference	34.0 (0.8)	42.0 (1.1)
	5	Thigh circumference	20.0 (0.5)	25.0 (0.6)
	6	Calf circumference	13.0 (0.3)	16.0 (0.4)
	7	Ankle circumference	8.0 (200 mm)	10.0 (250 mm)
	8	Foot length	9.8 (245 mm)	11.3 (283 mm)
	9	Elbow to floor	41.0 (1.0)	46.0 (1.2)
C	1	Head width	5.7 (143 mm)	6.4 (160 mm)
	2	Interpupillary distance	2.27 (56.75 mm)	2.74 (68.5 mm)
	3	Head length	7.3 (183 mm)	8.2 (205 mm)
	4	Head height	—	10.2 (255 mm)
	5	Chin to eye	—	5.0 (125 mm)
	6	Head circumference	21.5 (0.54)	23.5 (0.59)
D	1	Hand length	6.9 (173 mm)	8.0 (200 mm)
	2	Hand width	3.7 (92.5 mm)	4.4 (110 mm)
	3	Hand thickness	1.05 (26.25 mm)	1.28 (32 mm)
	4	Fist circumference	10.7 (267.5 mm)	12.4 (310 mm)
	5	Wrist circumference	6.3 (39.7 mm)	7.5 (186 mm)
E	1	Arm swing, aft	40 degrees	40 degrees
	2	Foot width	3.5 (87.5 mm)	4.0 (100 mm)
F	1	Shoulder width	17.0 (0.4)	19.0 (0.48)
	2	Sitting height to floor (std chair)	52.0 (1.3)	56.0 (1.4)
	3	Eye to floor (std chair)	47.4 (1.2)	51.5 (1.3)
	4	Standard chair	18.0 (0.45)	18.0 (0.45)
	5	Hip breadth	13.0 (0.3)	15.0 (0.38)
	6	Width between elbows	15.0 (0.38)	20.0 (0.5)
G	0	Arm reach (finger grasp)	30.0 (0.75)	35.0 (0.88)
	1	Vertical reach	45.0 (1.1)	53.0 (1.3)
	2	Head to seat	33.8 (0.84)	38.0 (0.95)
	3	Eye to seat	29.4 (0.7)	33.5 (0.83)
	4	Shoulder to seat	21.0 (0.52)	25.0 (0.6)

TABLE 1.5f Continued
Male Human Body Dimensions

Selected dimensions of the human body (ages 18 to 45). Locations of dimensions correspond to those in Figure 1.5e.

		Dimension in Inches (Meters) Except Where Noted	
	Dimensional Element	*5th Percentile*	*95th Percentile*
5	Elbow rest	7.0 (175 mm)	11.0 (275 mm)
6	Thigh clearance	4.8 (120 mm)	6.5 (162 mm)
7	Forearm length	13.6 (340 mm)	16.2 (405 mm)
8	Knee clearance to floor	20.0 (0.5)	23.0 (0.58)
9	Lower leg height	15.7 (393 mm)	18.2 (455 mm)
10	Seat length	14.8 (370 mm)	21.5 (0.54)
11	Buttock to knee length	21.9 (0.55)	36.7 (0.92)
12	Buttock to toe clearance	32.0 (0.8)	37.0 (0.93)
13	Buttock to foot length	39.0 (0.98)	46.0 (1.2)

Note: All except critical dimensions have been rounded off to the nearest inch (mm).

TABLE 1.5g
Absolute Discrimination Capability

(Maximum Rates of Information Transfer)

Modality	*Dimension*	*Maximum Rate (Bits/Stimulus)*
Visual	Linear extent	3.25
	Area	2.7
	Direction of line	3.3
	Curvature of line	2.2
	Hue	3.1
	Brightness	3.3
Auditory	Loudness	2.3
	Pitch	2.5
Taste	Saltiness	1.9
Tactile	Intensity	2.0
	Duration	2.3
	Location on the chest	2.8
Smell	Intensity	1.53
	Multidimensional Measurements	
Visual	Dot in a square	4.4
	Size, brightness, and hue (all correlated)	4.1
Auditory	Pitch and loudness	3.1
	Pitch, loudness, rate of interruption, on-time fraction, duration, spatial location	7.2
Taste	Saltiness and sweetness	2.3

of stimuli to which the operator responds affect efficiency and fatigue.

Performance increases when the number of tasks increases from a level of low involvement to one of high involvement. When rates of information processing are raised beyond the limits of the various senses, performance rapidly deteriorates.

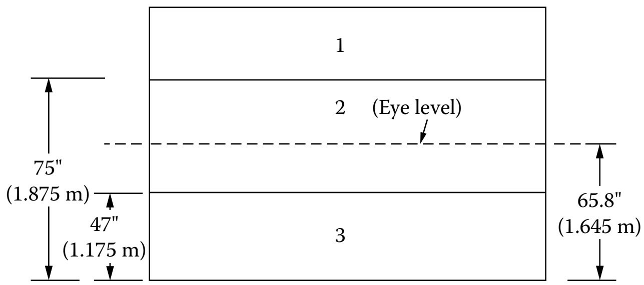
APPLICATION OF HUMAN ENGINEERING

The seemingly commonplace task of defining the man–machine system is more difficult to formalize than expected because:

1. Operators themselves are too closely involved for objectivity.
2. Engineers tend to make assumptions based on their technical background.
3. Equipment representatives focus their attention mainly on the hardware.

The definition can be more highly organized around the purpose, components, and functions of the system. The purpose of the system is its primary objective; it significantly affects design. In a system for continuous control of a process, operator fatigue and attention span are of first importance. Attention arousal and presentation of relevant data are important when the objective is the safety of personnel and equipment, and the operator must reach the right decision as quickly as possible.

Components of the system are divided into those that directly confront the operator and those that do not. The distinction is used to locate equipment. For example, a computer process console should be located for operator convenience,



Variables are grouped by function

- Indication and controls in area 2.
- Seldom used controls or push buttons in area 3.
- Indicators (analog or digital) only in area 1.
- Alarm and on/off indication in area 1.

FIG. 1.5h
Panel for standing operator. Variables are grouped by function, with indication and controls in area 2, seldom-used controls or pushbuttons in area 3, indicators (analog or digital) only in area 1, and alarm and on-off indicators in area 1.

but maintenance consoles should be out of the way so as to reduce interference with operation.

System functions are primarily the operator's responsibility, and operator function should be maximized for tasks that involve judgment and experience and minimized for those that are repetitive and rapid. Table 1.5d is an efficient guideline for task assignment.

Statistics of Operator Population

As was mentioned in the introduction to this section, the statistical nature of human engineering data influences individual applications. A specific problem demands full evaluation of the similarity among operators. For example, if color blindness is not a criterion for rejecting potential operators, color coding must be backed by redundant means (shape) to ensure recognition because one of four subjects has some degree of color blindness.

Variations and exceptions are to be expected and recognized since people are significantly different in physical size, native aptitude, and technical acculturation. Control systems and instruments should be designed for the user. A frequent mistake made by the instrument engineer is to assume that the operator will have a background similar to his or hers.

Some operators may have never ridden a bicycle, much less driven a car or taken a course in classical physics. In Arctic regions, massive parkas and bulky gloves change all the figures given in body dimension tables. It is only in Western culture that left-to-right motion of controls and indicators is synonymous with increasing values.

Setting Priorities

The diverse functions of the system confronting the operator vary in importance. The most accessible and visible areas should be assigned to important and frequently used items. Vertical placement is critical and dependent on the normal operating mode (standing, sitting, or mixed). Determination of optimum viewing zones is centered on the average eye level for standing (approximately 66 in. or 1.65 m) and sitting (approximately 48 in. or 1.2 m). Above these areas, visibility is still good, but accessibility falls off (Figures 1.5h and 1.5i). At the lower segment, even grocers know that objects are virtually hidden. Canting the panel helps to make subordinate areas more usable. (For more on control panel design, refer to Chapter 4).

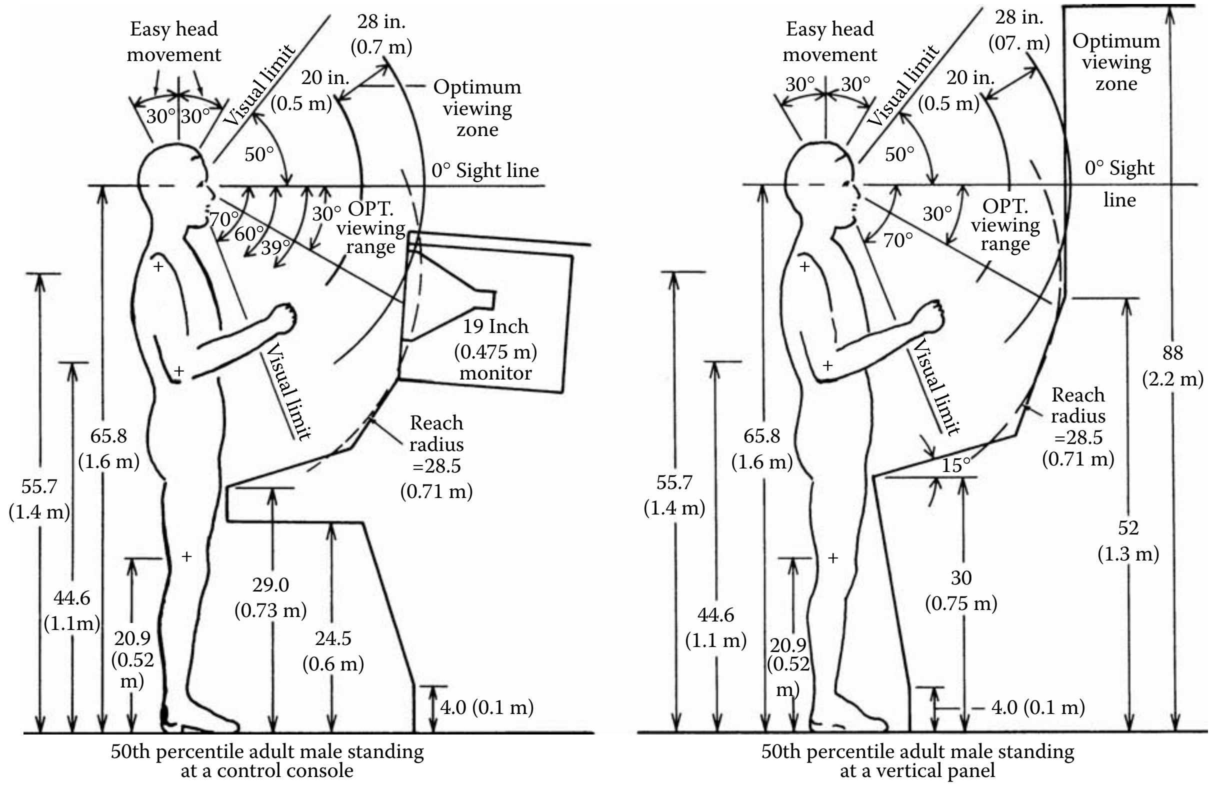
Instruments should be arranged from left to right (normal eye scan direction) in (1) spatial order, to represent material movement (storage and distribution); (2) functional order, to represent processing of material (distillation, reaction, and so on); and (3) sequential order, to modify the functional order approach to aid the operator in following a predetermined sequence even under stress.

Wrap-around consoles (now in common use with the advent of distributed control systems) with multifunction displays (such as cathode ray tubes) do not require operator movement; however, the primary (most used or critical) displays and controls should be centered in front of the sitting operator with auxiliary or redundant instruments on either side. Reach considerations are much more important when the operator is normally sitting (Figure 1.5j).

Locating similar functions in relatively the same place gives additional spatial cues to the operator by associating locating with function. On this account it is advisable to divide alarms into groups related to meaning, including (1) alarms requiring immediate operator action (for example, high bearing temperatures), which must be read when they register; (2) alarms for abnormal conditions that require no immediate action (standby pump started on failure of primary), which must be read when they register; and (3) status annunciator for conditions that may or may not be abnormal. These are read on an as-needed basis.

Alarms can be separated from other annunciators, have a different physical appearance, and give different audible cues. For example, group 1 alarms can have flashing lamp windows measuring 2 × 3 in. (50 × 75 mm) with a pulsating tone. Group 2 alarms can have flashing lamp windows measuring 1 × 2 in. (25 × 50 mm) with a continuous tone, and group 3 can have lamp windows measuring 1 × 1 in. (25 × 25 mm) without visual or auditory cues.

Alarms are meaningless if they occur during shutdown or as a natural result of a primary failure. Whether usable or not, they capture a portion of the operator's attention. Several alarms providing no useful information can be replaced by a single shutdown alarm. Later, the cause of the shutdown will be of keen interest to maintenance personnel; to provide detailed information about the cause, local or back-of-panel annunciators can be used. CRT displays are generally multifunctional, and rapid operator association (especially in emergencies) with a particular situation or function can be visually facilitated by unique combinations of shapes, colors, or sounds.

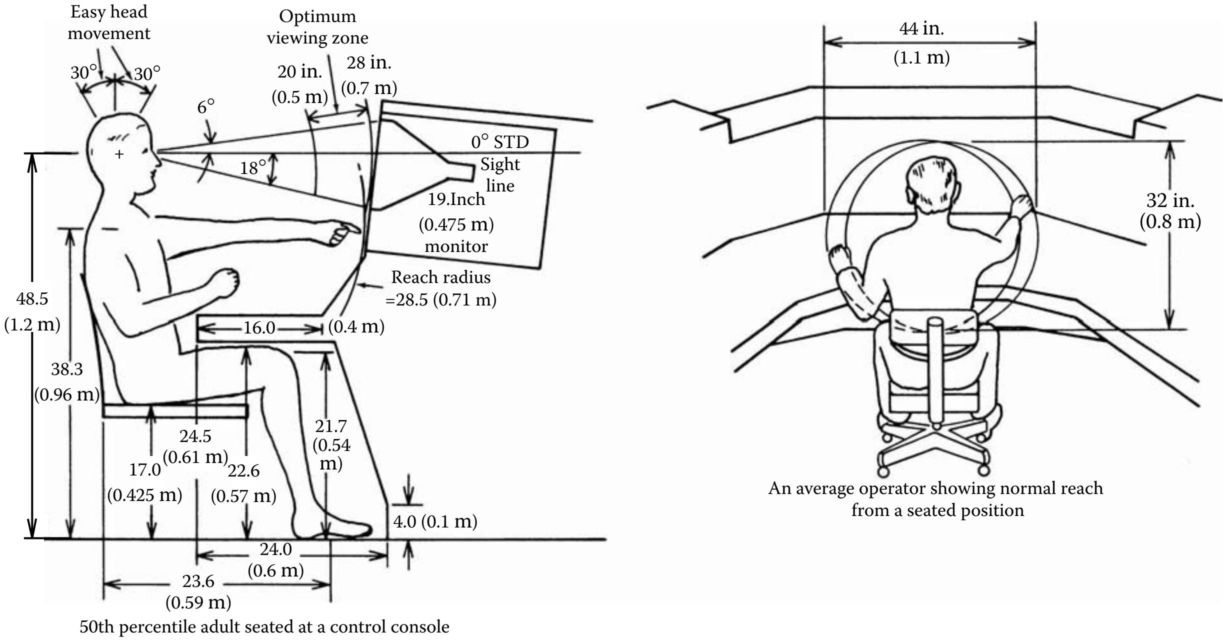


50th percentile adult male standing at a control console

50th percentile adult male standing at a vertical panel

FIG. 1.5i

Left, 50th percentile adult male standing at a control console. Right, 50th percentile adult male standing at a vertical panel.



50th percentile adult seated at a control console

An average operator showing normal reach from a seated position

FIG. 1.5j

Left, 50th percentile adult male seated at a control console. Right, normal reach from a seated position.

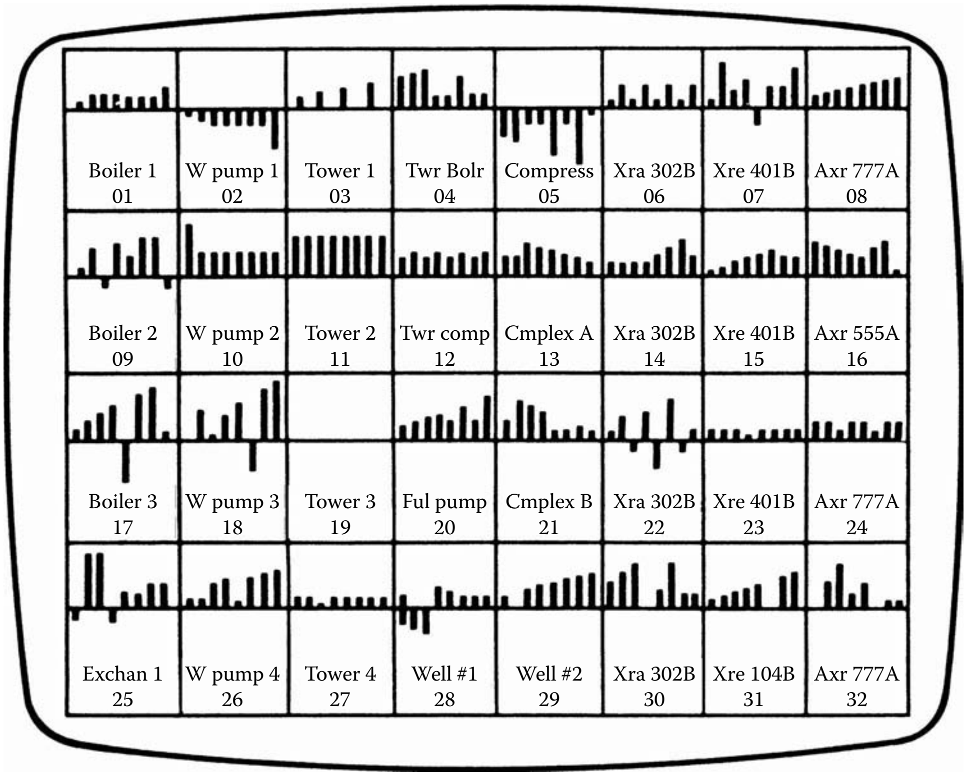


FIG. 1.5k
CRT overview display.

Hardware Characteristics

Digital indicators are best when rapid reading or obtaining accurate values is of utmost concern. However, they give only a useless blur at a high rate of change. Analog indications allow rapid evaluation of changing values. The reading can be related to zero and full-scale for interpretation of rate of change (estimated time to reach minimum or maximum levels). Correspondingly, analog information is difficult to read when exact values are needed.

Display patterns should be natural to the eye movement—in most cases, left-to-right. An operator scanning only for abnormal conditions will be aided by this arrangement. The horizontal scan line approach proved successful in several major lines of control instruments and is now widely used by “overview” displays in distributed control systems (Figure 1.5k).

An alternative system is shown in Figure 1.5l, in which two rows are combined to give the effect of one line containing twice the number of readings. In our culture, horizontal scanning is eight times faster than vertical scanning.

Viewing distances with cathode ray tubes depend on resolution (the number of scan lines) and size of characters used in the display. Normal broadcast uses 525 scan lines, which translates to approximately 1.8 lines per millimeter. Color CRTs used as operator displays vary widely, from 38 to 254 lines per inch (1.52 to 10 lines per millimeter). As a general rule, 10 lines per character provide optimal character recognition. Words can be identified with a smaller number of lines per character. Typical color CRTs associated with process control have eight lines per character.

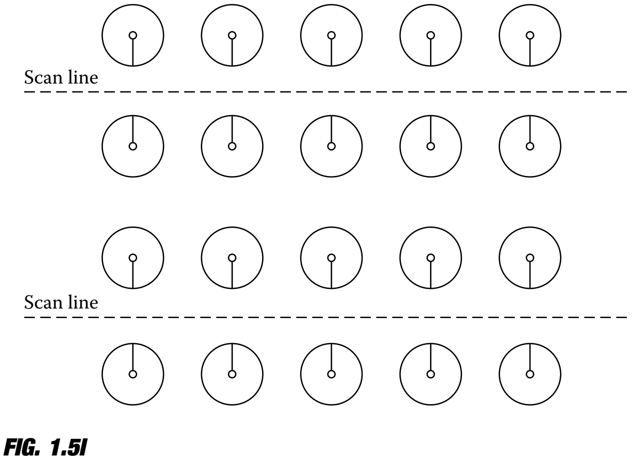


FIG. 1.5l
Dial orientation enabling two rows of instruments to be scanned at one time.

Viewing distance as it relates to character height can be determined using the following formula:

$$\text{optimum viewing distance (inches)} = \frac{\text{character height (inches)}}{0.00291} \tag{1.5(3)}$$

The interactive effect of color, contrast, and lines per character adversely influences the viewing distance.

Black-and-white screens are visible at nearly any light level, provided that reflected light is eliminated by a hood or a circularly polarized filter. Flicker resulting from slow refresh rate is a potential fatigue factor. Rates above 25 flickers per second are reasonably stable. Color CRTs require a lighting level above 0.01 lumen to ensure that color vision (retinal rods) is activated and color discrimination is possible.

Display of information through cathode ray tubes can be graphic (pictorial or schematic representation of the physical process) or tabular (logical arrangement of related variables by function). The graphic format is useful when the controlled process changes in operating sequence or structure (batch reactors, conveyor systems, and so forth) or if the operator is not familiar with the process. The tabular display format conveys large quantities of information rapidly to an experienced operator in continuous processes (distillation columns, boilers, and so forth).

Operator response in modern control systems occurs predominantly through specialized keyboards. Key activation can be sensed by the operator through tactile, audible, or visual feedback. Tactile feedback is usually found in discrete or conventional pushbutton keys and is well suited for rapid multiple operations, such as entry of large quantity of data. Membrane-, laminate-, or capacitance-activated keyboards do not provide adequate tactile feedback; therefore, audible feedback, such as a “beep” or “bell” sound, is usually provided. Keyboards with audible feedback are well suited for low-frequency usage, such as may be encountered in the operation of a process control system. Audible feedback may be confusing if adjacent keyboards with identical audible feedback are used simultaneously.

Visual feedback is useful when several operations use multiple select keys and a common activate key. Blinking or color change of the selected operation reduces the probability of an operator error.

Information Coding

Several factors affect the ability of people to use coded information. The best method to predict difficulty is by the channel center load, which uses the basic information formula of Equation 1.5(1). By this method the informational load is determined by adding the load of each character.

In a code using a letter and a number, the character load is:

$$\text{Letter (alphabet) } L_h = \text{Log}_2\ 26 = 4.70$$
$$\text{Number (0 to 9) } N_h = \text{Log}_2\ 10 = 3.32$$
$$\text{Code information } (C_h) = 4.70 + 3.32 = 8.02$$

This method applies when the code has no secondary meaning, such as when A 3 is used for the first planet from the sun, Mercury. When a code is meaningful in terms of the subject, such as O 1 for Mercury (the first planet), the information load is less than that defined by the formula.

Typical laboratory data indicate that when a code carries an information load of more than 20, a person has difficulty in handling it. Tips for code design include:

1. Using no special symbols that have no meaning in relation to the data
2. Grouping similar characters (HW5, not H5W)
3. Using symbols that have a link to the data (M and F, not 1 and 2, for male and female)
4. Using dissimilar characters (avoiding zero and the letter O and I and the numeral one)

The operator environment can be improved by attending to details such as scale factor, which is the multiplier relating the meter reading to engineering units. The offset from midscale for the normal value of the value of the variable and the desired accuracy of the reading for the useful range are constraints. Within these constraints, the selected scale factor should be a one-digit number, if possible. If two significant digits are necessary, they should be divisible by (in order of preference) 5, 4, or 2. Thus, rating the numbers between 15 and 20 as to their desirability as scale factors would give 20, 15, 16, 18 (17, 19).

Operator Effectiveness

The operator is the most important system component. His or her functions are all ultimately related to those decisions that machines cannot make. The operator brings into the system many intangible parameters (common sense, intuition, judgment, and experience) — all of which relate to the ability to extrapolate from stored data, which sometimes is not even consciously available. This ability allows the operator to reach a decision that has a high probability of being correct even without all the necessary information for a quantitative decision.

The human engineer tries to create the best possible environment for decision making by selecting the best methods of information display and reducing the information through elimination of noise (information type), redundancy (even though some redundancy is necessary), and environmental control.

Operator Load

The operator’s load is at best difficult to evaluate because it is a subjective factor and because plant operation varies. During the past 30 years, automation and miniaturization have exposed the operator to a higher density of information, apparently increasing operator efficiency in the process. Since increases in load increase attention and efficiency, peak loads will eventually be reached that are beyond the operator’s ability to handle.

Techniques for reducing operator load include the simultaneous use of two or more sensory channels.

Vision is the most commonly used sensory channel in instrumentation. Sound is next in usage. By attracting the operator’s attention only as required, it frees the operator from continuously having to face the display and its directional

TABLE 1.5m
General Illumination Levels

Task Condition	*Level (foot candles [lux])*	*Type of Illumination*
Small detail, low contrast, prolonged periods, high speed, extreme accuracy	100 (1076)	Supplementary type of lighting; special fixture such as desk lamp
Small detail, fair contrast, close work, speed not essential	50 to 100 (538 to 1076)	Supplementary type of lighting
Normal desk and office-type work	20 to 50 (215 to 538)	Local lighting; ceiling fixture directly overhead
Recreational tasks that are not prolonged	10 to 20 (108 to 215)	General lighting; random room light, either natural or artificial
Seeing not confined, contrast good, object fairly large	5 to 10 (54 to 108)	General lighting
Visibility for moving about, handling large objects	2 to 5 (22 to 54)	General or supplementary lighting

message. The best directionality is given by impure tones in the frequency range of 500 to 700 Hz.

Blinking lights, pulsating sounds, or combination of both effectively attract an operator's attention. The relative importance of these stimuli can be greatly decreased if their use is indiscriminate. (Blinking, as an example, should not be used in normal operator displays on a CRT.)

Increased frequency of a blinking light or a pulsating sound is normally associated with increased urgency; however, certain frequencies may have undesirable effects on the operator. Lights changing in intensity or flickering (commonly at frequencies of 10 to 15 cycles) can induce epileptic seizures, especially in stressful situations.

Information quantity can be decreased at peak load. In a utility failure, for example, more than half the alarms could be triggered when many of the control instruments require attention. It would be difficult for the operator to follow emergency procedures during such a shutdown. A solution gaining acceptance in the industry involves selectively disabling nuisance alarms during various emergencies. Some modern high-density control rooms have more than 600 alarms without distinction as to type or importance.

Environment

The responsibilities of human engineering are broad and include spatial relations — distances that the operator has to span in order to reach the equipment and fellow operators. Controlling temperature and humidity to maintain operator and equipment efficiency is also an important responsibility of the human engineer.

Light must be provided at the level necessary without glare or superfluous eye strain. Daylight interference caused by the sun at a low angle can be an unforeseen problem. Last-minute changes in cabinet layout can create undesirable shadows and glare from glass and highly reflective surfaces. Flat, brushed, or textured finishes reduce glare. Poor illumination can be reduced by light colors, low light fixtures, and special fixtures for special situations. Fluorescent lighting, because of its 60-Hz flicker rate, should be supplemented by incandescent lamps to reduce eye fatigue, particularly important in the presence of rotating equipment, in order to eliminate the strobing effect. Table 1.5m and Table 1.5n give detailed information on typical lighting applications.

TABLE 1.5n
Specific Recommendations, Illumination Levels

Location	*Level (foot candles [lux])*
Home	
Reading	40 (430)
Writing	40 (430)
Sewing	75 to 100 (807 to 1076)
Kitchen	50 (538)
Mirror (shaving)	50 (538)
Laundry	40 (430)
Games	40 (430)
Workbench	50 (538)
General	10 (108) or more
Office	
Bookkeeping	50 (538)
Typing	50 (538)
Transcribing	40 (430)
General correspondence	30 (323)
Filing	30 (323)
Reception	20 (215)
School	
Blackboards	50 (538)
Desks	30 (323)
Drawing (art)	50 (538)
Gyms	20 (215)
Auditorium	10 (108)
Theatre	
Lobby	20 (215)
During intermission	5 (54)
During movie	0.1 (1.08)
Passenger Train	
Reading, writing	20 to 40 (215 to 430)
Dining	15 (161)
Steps, vestibules	10 (105)
Doctor's Office	
Examination room	100 (1076)
Dental–surgical	200 (2152)
Operating table	1800 (19,368)

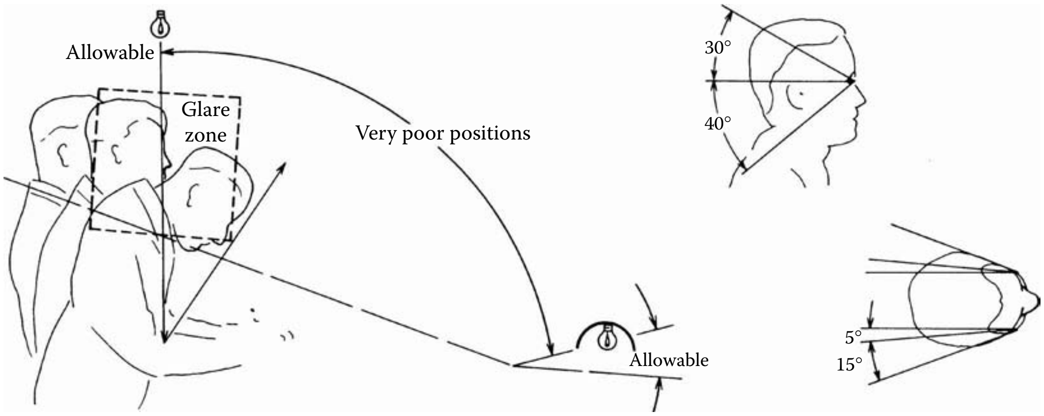


FIG. 1.5o
Typical lighting chart.

Glare is the most harmful effect of illumination (Figure 1.5o). There is a direct glare zone that can be eliminated, or at least mitigated, by proper placement of luminaires and shielding, or, if luminaires are fixed, by rearrangement of desks, tables, and chairs. Overhead illumination should be shielded to approximately 45 degrees to prevent direct glare. Reflected glare from the work surface interferes with most efficient vision at a desk or table and requires special placement of luminaires.

Eyeglasses cause disturbing reflections unless the light source is 30 degrees or more above the line of sight, 40 degrees or more below, or outside the two 15-degree zones, as shown in Figure 1.5o.

Noise and vibration affect performance by producing annoyance, interfering with communication, and causing permanent physical damage. Prolonged exposure to high sound levels causes hearing loss, both temporary and permanent. The noise exposure values in Table 1.5p are those stated in the Walsh Healy Public Contracts Act as being permissible. They are derived from the curves in Figure 1.5q and reflect the variation in sensitivity with frequency.

Annoyance and irritability levels are not easily determined because they are subjective factors, and habituation significantly affects susceptibility. A quantitative tolerance limit has not yet been established. One aspect of background noise is deterioration of speech communication.

Noise reduction may take the form of reducing noise emission at the source, adding absorbent material to the noise path, or treating the noise receiver (having operators wear protective equipment, such as ear plugs and ear muffs). The last precaution reduces both speech and noise, but the ear is afforded a more nearly normal range of sound intensity and thus can better recognize speech. Figure 1.5r illustrates typical speech interference levels.

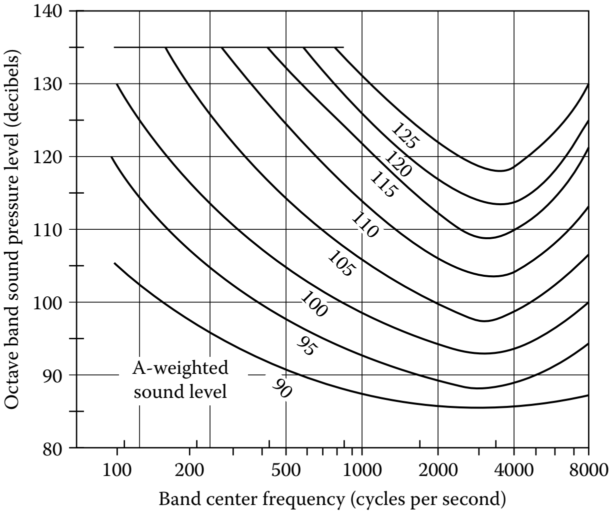


FIG. 1.5q
Equivalent sound level contours.

TABLE 1.5p
Permissible Noise Exposures[*]

Duration per Day (hours)	*Sound Level (dBA)*	*Duration per Day (hours)*	*Sound Level (dBA)*
8	90	1½	102
6	92	1	105
4	95	½	110
3	97	¼ or less	115
2	100		

[*] When the daily noise exposure is composed of two or more periods of noise exposure of different levels, their combined effect should be considered, rather than the individual effect of each. If the sum of the following fractions: $C1/T1 + C2/T2 \ldots Cn/Tn$ exceeds unity, then the mixed exposure should be considered to exceed the limit value. Cn indicates the total time of exposure at a specified noise level, and Tn indicates the total time of exposure permitted at that level.

Exposure to impulsive or impact noise should not exceed 140 dbA peak sound pressure level.

The speech interference level should be less than that given in order to have reliable conversation at the distance and voice levels shown.

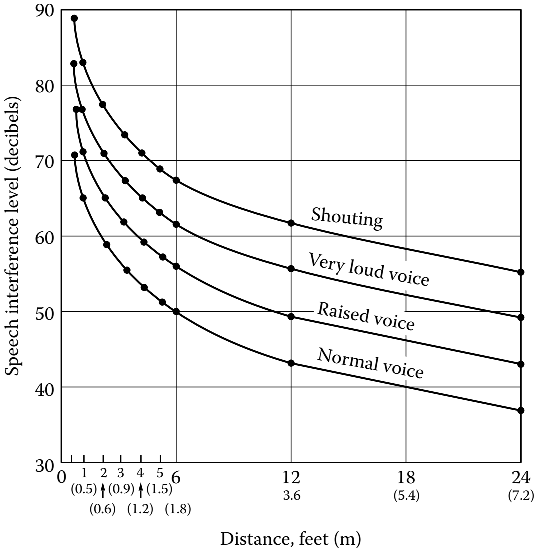


FIG. 1.5r
Speech interference levels.

SUMMARY

This section has described not only the tools needed, but the solutions that will fulfill the engineer's mandate from the industrial sector which must come out of his or her own creativity. The engineer must develop an original approach to each new assignment and introduce creative innovations consistent with sound engineering practices. Decisions must not be based solely on data from current projects reflecting only what has not worked, and human engineering must not be left to instrument manufacturers who, however well they fabricate, still do not know how the pieces must fit together in a specific application. Complications of data on mental and physical characteristics must be approached cautiously because they may reflect a group of subjects not necessarily like the operators in a specific plant.

It is the engineer's task to ensure that the limitations of men and machines do not become liabilities and to seek professional assistance when means of eliminating the liabilities are not evident.

Bibliography

"Application of Ergonomics to Occupational Safety," Report No. 712F, WSA Inc., San Diego, CA, April 1978, Retrieval #USGPB82-154253.

Chapanis, A., *Man–Machine Engineering,* Belmont, CA: Wadsworth, 1965.

Collier, S. G., and Green, M., "Verification and Validation of Human Factors Issues in Control Room Design and Upgrades (HWR-598)," Halden, Norway: OECD Halden Reactor Project, 1999.

Degani, A., et al., "Modes in Human–Machine Systems: Review, Classification, and Application," Human Factors Research and Technology Web page, NASA, http://human-factors.arc.nasa.gov, 2001.

Farmer, E., "Design Tips for Modern Control Rooms," *Instruments and Control Systems,* March 1980.

Geer, C. W., "Human Engineering Procedure Guide," Report No. AD-A 108643 D180-25471-1, Boeing Aerospace Co., Seattle, WA, September 1981.

Green, M., et al., "A Method for Reviewing Human Factors in Control Centre Design," Oslo: Norvegian Petroleum Directorate, 2000.

Harmon, D., and Shin, Y. C., "Joint Design Development of the KNGR Man–Machine Interface," Taejon, Korea: KEPRI & Combustion Engineering, 1998.

Human Factors Evaluation of Control Room Design and Operator Performance at Three Mile Island-2, U.S. Dept. of Commerce NUREG/CR-2107, Volumes 1, 2, and 3, January 1980.

Human Factors Review of Nuclear Power Plant Control Room Design, EPRI NP-309, Palo Alto, California, March 1977.

IEEE 1289, "IEEE Guide for the Application of Human Factors Engineering," Piscataway, NJ: IEEE Press, 1998.

ISO, "Ergonomic Design of Control Centres — Part 1, 2, and 3," Brussels: CEN, Management Centre ISO, 2000.

Ivergard, T., *Handbook of Control Room Design and Ergonomics,* London: Taylor and Francis, 1989.

Louka, M., et al., "Human Factors Engineering and Control Room Design" IFE/HR/E-98/031, Halden, Norway: OECD Halden Reactor Project, 1998.

McCormick, E. J., *Human Factors Engineering,* 2nd ed., New York: McGraw-Hill, 1964.

McCready, A. K., "Man–Machine Interfacing for the Process Industries," *InTech,* March 1982.

Military Standard — Human Engineering Design Criteria for Military Systems, Equipment and Facilities, MIL-STD 14728 and Later Addenda.

Nixon, C. W., "Hearing Protection Standards," Report No. AFAMRL-TR-80-60, Air Force Aerospace Medical Research Lab (Wright-Patterson Air Force Base Ohio), 1982; Retrieval #USGAD-A110 388/6.

NUREG/CR-6633, "Advanced Information System Design: Technical Basis and Human Factors Review Guidance," USNRC, March 2000.

Rehm, S., et al., "Third International Congress on Noise Effects as a Public Health Problem, Biological Disturbances, Effects on Behavior," Report No. BMFT-FB-HA-81-013, Mainz University (Germany), October 1981, Retrieval #USGN82-15597/9.

Rumph, P., "Improved Information Interfacing Turns Operators into Process Managers," *Pulp & Paper,* August 1998.

Shulmann, H. C., *Psychological Bulletin,* Vol. 75, 6, June 1971.

Stephanidis, C., and Sfyrakis, M., "Current Trends in Man–Machine Interfaces," in P. R. V. Roe (Ed.), *Telecommunications for All,* Brussels: ECSC-EC-EAEC, 1995.

Stubler, W. F., et al., "Human Systems Interface and Plant Modernization Process," NUREG/CR-6637, Washington, D.C.: U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 2000.

Technical Report 1001066, "Human Factors Guidance for Digital I&C Systems and Hybrid Control Rooms: Scoping and Planning Study," Palo Alto, CA: Electric Power Research Institute (EPRI), November 2000.

Thompson, B. J., "Preparing for Computer Control," *Instruments and Control Systems,* February 1980.

Treurniet, W. C., "Review of Health and Safety Aspects of Video Display Terminals," Report No. CRC-TN-712-E, Communications Research Centre (Ottawa, Ontario, Canada), February 1982.

Woodson, V. J., and Conover, H. N., *Human Engineering Guide for Equipment Designers,* 2nd ed., Berkeley: University of California Press, 1964.

Worthington, S., "The Personalization of Manufacturing: Redefining the Human Interface," Paper #91-0610, 1991 ISA Conference, Anaheim, CA, October 1991.

1.6 Process Alarm Management

M. S. MANNAN, H. H. WEST (2005)

INTRODUCTION

Recently, major incidents have been caused or worsened by the excessive and undisciplined growth in the quantity of alarms and for that reason, effective alarm management (A.M.) has moved to the top of many plant project lists.

Many currently installed computer control systems provide only limited information regarding abnormal operation. The operator information is often not enhanced by the computer control system, and the operator is forced to search out trends and use intuition and experience to evaluate abnormal plant status.

Alarm growth is a natural outcome of the increased information load provided by modern control systems. However, if alarms are not managed in a disciplined manner, uncontrolled alarm growth can result, leading to ineffective and potentially dangerous alarm situations.

A structured approach to alarm management has emerged to increase alarm effectiveness and thereby overall plant safety.

History of Alarm Systems

In the not-too-distant past, alarm systems consisted of a few selected process measurements, which were hard-wire connected to panel board mounted annunciators or indicator lights, which activated when the measurements exceeded some predefined limits. These panels provided alarm annunciation to the plant operator. The panels were large but limited in capacity and thereby tended to limit the number of configured alarms.

Modern distributed control systems (DCS) and programmable electronic systems (PES) are capable of defining limits for each field measurement. Furthermore, calculated parameters, such as rate of change or a combination of field measurements could also have defined limits. Therefore, computer-based control systems have the capability to vastly increase the number of configured alarms.

Alarms not only increase the amount of information provided to the operator but can often also be a source of operator overload and confusion. There have been a number of major incidents that might have been prevented if the plant operator had recognized critical alarms among the flood of alarms that were activated.[1] One notable example of an alarm problem was the Three Mile Island accident in 1979, where important alarms were missed because of the flood of alarms received. Another example of alarm overload was the Texaco refinery explosion at Milford Haven in 1994, where, in the last 10 minutes prior to the explosion, the operators had to respond to alarms that were annunciating at a rate of about three per second.

Definitions

The most common definition of an alarm is “a signal designed to alert, inform, guide or confirm deviations from acceptable system status.” Similarly, an alarm system is defined as a system for generating and processing alarms to be presented to users.

Another common definition is that an alarm system is designed to direct the operator’s attention toward significant aspects of the current plant status. In other words, an alarm system should serve to help the operator to manage his time.

Because the DCS system processes information on a tag point-by-tag point basis, alarms and non-safety related events, such as condition monitoring alerts, equipment fault status, or journal logging events for engineering or maintenance attention, have often been described as alarms.

Hence, the A.M. system must address events that need operator attention and various other non-safety related events. Figure 1.6a provides a graphic description of the relationship between alerts and safety-related alarms. The distinction that is emerging between the two can be stated as follows: An *Alarm* requires operator action, while an *Alert* is informational in its nature and may not require a specific action by the operator.

Alarm Basics

The purpose of alarm systems is to alert operators to plant conditions that deviate from normal operating limits and thereby require timely action or assessment. Hence, the purpose of an alarm is action.

Matters that are not worthy of operator attention should not be alarmed but rather placed into a logging journal database for the potential use by other plant staff.

An alarm is useless if the operator does not know what action is required in response to it. Similarly, if the operator knows the correct action but cannot implement the action, the alarm is also useless. Useless alarms should not exist.

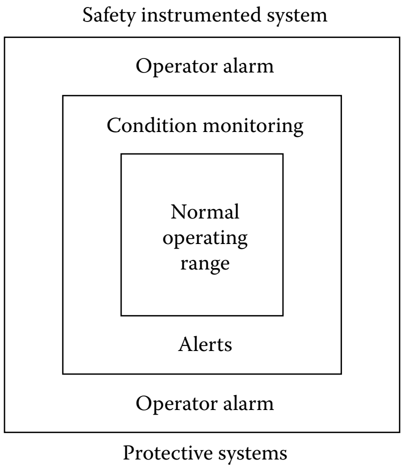


FIG. 1.6a
Graphic description of the relationship between alerts and safety-related alarms.

Logging may be a suitable alternative to be used to record non-safety related discrepancy events to prevent the unnecessary sounding of alarms. A system for assessing the significance of logged events to ensure timely intervention by maintenance personnel is absolutely required.

A.M. COSTS AND DESIGNS

The Honeywell Abnormal Situation Management Consortium[2] members estimated that a medium-sized petrochemical company consisting of about six facilities (refineries, chemical plants, etc.) might accumulate \$50 million to \$100 million in yearly losses due to plant upsets and other production-related incidents. They also noted that losses caused by plant upsets cost about 3 to 5% of the total throughput of the plant. The US National Institute of Standards and Technology estimated that plant upsets cost the national economy about \$20 billion annually. These estimates may actually underestimate the real situation. Not included in these cost estimates are the rare tragic incidents such as the Phillips Pasadena disaster or the Piper Alpha tragedy.

A.M. and Safety Instrumentation

Although alarm systems may not always have safety implications, they do have a role in enabling operators to reduce the need for automatic shutdowns initiated by safety systems. However, in cases where a risk reduction factor of 10^{-1} failures on demand is claimed, the safety system includes both the alarm system and the operator. Hence, the total safety system requires a suitable safety integrity level.[3,4]

Many Layers of Protection Analysis (LOPA) safety integrity level determination studies have incorporated alarm management as an existing layer of protection, thereby assigning the alarm system components some designated level of reliability.

Alarms that are not designated as safety should be carefully designed to ensure that they fulfill their role in reducing demands on the safety-related systems.

For all alarms, regardless of their safety designation, attention is required to ensure that under abnormal conditions or under severe emergency situations, the alarm system remains effective and the limitations of the speed of human response are recognized in its design.

Alarm Set Points

The type of alarm and its set point must be established to enable the operator to make the necessary assessment and take the required timely action.

Engineering analysis of the potential timing of events that generate alarms requires good control system engineering. Then alarm set points must be documented and controlled in accordance with the alarm system management control. Timing of alarms and potential operator reaction time are both critical. Figure 1.6b illustrates how alerts can progress into alarms and eventually into safety shutdowns.

The recommended best practice is that the engineering analysis of alarm set points be coordinated with the plant hazard and operability (HAZOP) studies.

Alarm Presentation

The alarm interface provided for the operator must be effective.[5] Alarms may be presented on annunciation panels, individual indicators, CRT screens, or programmable display devices. Alarms lists should be carefully designed to ensure that high-priority alarms are readily noted and recognized, while low-priority alarms are not overlooked, and that the list remains readable even during times of high alarm activity or when alarms occur repeatedly.

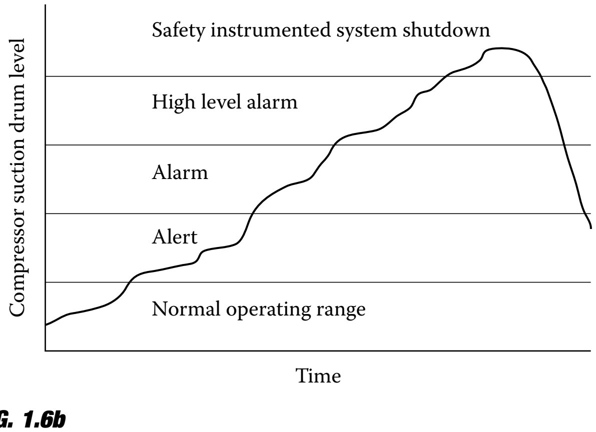


FIG. 1.6b
Alerts can progress into alarms and eventually into safety shutdowns.

Alarms should be prioritized in terms of which alarms require the most urgent operator attention. Alarms should be presented within the operators' field of view and use consistent presentation style (color, flash rate, inscription convention). The alarms should be presented in an explicit manner, should be distinguishable from other alarms, have the highest priority, and when activated, remain on view all the time.

The presentation of alarms should not exceed that which the operator is capable of acting upon, or alternatively the alarms should be prioritized and presented in such a way that the operator may deal with the most important alarms first and without distraction from the others.

Each alarm should provide sufficient information to the operator so that the operator will know the condition that caused the alarm, the plant affected, the corrective action that is required, priority of the alarm involved, the time of alarm initiation, and the alarm status.

The visual display device may be augmented by audible warnings. Where there are multiple audible warnings, they should be designed so that they are readily distinguished from each other and from emergency alarm systems. They should be designed to avoid distraction of the operator in high workload situations.

Alarm Processing

The alarms should be processed in such a manner as to avoid operator overload at all times (avoiding alarm floods). The alarm processing should ensure that fleeting or repeating alarms do not result in operator overload even under the most severe conditions. Alarm processing techniques such as filtering, the use of dead-band, de-bounce timers, and shelving have been successfully exploited by industry. For shelving alarms, a rule of thumb that has been used by some plants is that an alarm is shelved if it is repeated 10 times in 10 minutes.

Applicable alarm processing techniques include grouping and first-out alarms, suppression of lower-priority alarms (e.g., suppressing the high alarm when the high-high activates) suppression of out-of-service plant alarms, suppression of selected alarms during certain operating modes, automatic alarm load shedding, and shelving.

Care should be taken in the use of shelving or suppression to ensure that controls exist to ensure that alarms are returned to an active state when they are relevant to plant operation.

IMPROVING THE PLANT'S A.M.

In order to improve alarm management (A.M.) in a plant, it is first necessary to identify the symptoms of the problems that exist. Next it is desirable to note available elements of A.M. as the potential tools of correction. Once this been done, an A.M. improvement and rationalization project can be initiated. An example of such a project will be described later in this section.

Symptoms of A.M. Problems

The list below identifies some of the symptoms that might signal problems with the design of the A.M. system:

- Minor operating upsets produce a significant number of alarm activations.
- Routine operations produce a large number of alarm activations that serve no useful purpose.
- Alarm activations occur without need for operator action.
- Major operating upsets produce an unmanageable number of alarm activations.
- Operators do not know what to do when a particular alarm occurs.
- Occasional high alarm activation rates occur.
- A large number of configured high-priority alarms occur.
- Alarm conditions are not corrected for long periods of time.
- Alarms chatter or are frequently repeated transient.
- No plant-wide philosophy for the alarm system management has been established.
- There are no guidelines for when to add or delete an alarm.
- Operating procedures are not tied to alarm activations.
- There are active alarms when the plant is within normal operating parameters.
- There is a lack of documentation on alarm set points.
- The system does not distinguish safety, operational, environmental, and informational events.
- The schedule of alarm testing is erratic.
- A large number of defeated alarms occurs.
- Any operator or engineer can add an alarm or change the set points on his or her own authority.

The Tools and Elements of A.M.

Some of the essential elements of alarm management are:

- An alarm management philosophy document
- Written operating procedures for alarm response
- Ownership of the alarm system
- Alarm prioritization
- Alarm configuration records
- Alarm presentation
- Alarm logical processing, such as filtering and suppression
- Informational alert management
- Operator training
- Alarm testing
- Management of changes made to the alarm system

Management systems should be in place to ensure that the alarm system is operated, maintained, and modified in a controlled manner. Alarm response procedures should be available, and alarm parameters should be documented.

The performance of the alarm system should be assessed and monitored to ensure that it is effective during normal and abnormal plant conditions. The monitoring should include evaluation of the alarm presentation rate, operator acceptance and response times, operator workload, standing alarm count and duration, repeat or nuisance alarms, and operator views of operability of the system. Monitoring may be achieved by regular and systematic auditing.

A.M. Improvement and Rationalization Project

Alarm rationalization is a systematic procedure for the analysis and improvement of the DCS capability used to alert operators of alarm conditions. This process normally results in a reduction in the total number of configured alarms. Rationalization may occasionally identify the need for new alarms or changes in the alarm priority. Alarm rationalization is a structured procedure that involves a project team, with representatives from operations, maintenance, engineering, and safety.

The rationalization procedure includes a dozen essential steps, each of which is listed and described under separate titles in the paragraphs that follow:

Assessing the Current Alarm System Performance Typical metrics needed to assess the current and then later the improved alarm system include:

- Rate of alarm activation
- Pattern of alarm activations
- Priority alarm activations
- Number of standing alarms
- Time before acknowledge
- Spread between trip points
- Time before clearing
- Chattering alarms
- Correlated alarms
- Number of disabled alarms
- Total number of alarms
- Alarms per operator
- Alarm frequency by shift
- Alarm rate during emergencies
- Fraction of unacknowledged alarms
- Average time to return to normal
- Nuisance alarms
- Disabled alarms
- Bypassed alarms
- Shelved alarms

Operational metrics must also be used, such as production rate, off-quality production, number of upsets, and any other factors that the plant considers important or relevant. Safety and environmental metrics that must be used are the number of plant shutdowns, number of incidents/near misses, releases to the atmosphere, and pressure relief activations or releases to flare.

The amount of information needed to develop these metrics is daunting, thereby requiring special software to sort the alarm journals in an efficient manner. Many DCS and Supervisory Control and Data Acquisition (SCADA) systems and vendors provide alarm management features or products. However, some of these are either primitive or do not provide the necessary purpose by themselves, though they are generally improving.

There also are some add-on alarm software products on the market that enhance a control system's basic alarm capabilities by providing online A.M., advanced logical processing of alarms, alarm pattern recognition, and dynamic reconfiguration of the alarm system for varying operating conditions.

Developing an A.M. Philosophy Document A consistent, comprehensive alarm management procedure/philosophy is necessary before beginning an alarm rationalization project.

This procedure typically covers alarm type (quality, safety, environmental, maintenance information), method of prioritization of alarms, alarm logical processing methods, and testing requirements.

Reviewing the Basis for Alarm Set Points Evaluation of the alarm configuration file and the accompanying engineering reports is required to verify the basis for the alarm set points. If not available, then an engineering study must be made to recreate the basis for the alarm set points.

Identifying the purpose of the alarm and its correlation to other alarms is particularly important. Analysis of the alarm purpose includes defining the consequences of inaction to alarm notification.

Analyzing Alarm Histories Dynamic analysis of the alarm journals of several previous months is required to get a statistically valid sample. Disadvantages include correlation with operational, process, or equipment events which may not be available in the electronic database.

Prioritizing the Alarms The significance of the alarm must be determined through a ranking scheme or identified during the HAZOP study. The process hazard/risk analysis must determine the level of importance that is associated with the operator detecting the alarm and performing the expected action. Prioritization helps ensure that the operator understands the importance of the alarm itself as well as the importance of the alarm in relationship to other alarms.

The prioritization scheme is generally limited to the control systems capabilities along with third-party alarm management software on the system. The number of prioritization levels should be kept to a minimum to minimize operator confusion. Typical alarm priority categories can be critical, high, medium, and low.

Incorporating Operator Actions in Procedures In addition to ensuring that each alarm has a defined operator response, operator reliability must be enhanced by lowering the workload, reducing the number of false alarms, and making the alarm displays obvious and the operator responses simple. Management should be made responsible to ensure that the operators are well trained and that their performance is tested.

Considering Advanced Logical Processing Techniques In addition to grouping alarms, various suppression techniques and artificial intelligence techniques may be piloted and then incorporated into the alarm system.

Updating the Alarm Presentation Techniques The latest techniques in DCS graphic design and control system design should be considered to enhance operator effectiveness.

Implementing the Rationalization Project Since the rationalization may add or remove alarms and change presentation techniques or configuration parameters, it is necessary to have an implementation plan that involves the participation of the appropriate personnel, such as the operators and operating staff.

Benchmarking the New Alarm System Once the alarm rationalization is implemented, the final system should be evaluated to determine the degree of success of the rationalization effort.

For example, the alarm rate should be less than one per 10 minutes (some plants suggest that one per 5 minutes is manageable). Similarly the distribution of priority alarms should be less than five high alarms per shift and less than two such alarms per hour. Journal alerts should not exceed 10 per hour.

Auditing The goal of a good audit process is to keep the alarm system manageable and in control. Online dynamic alarm monitoring systems can assist in this. If the plant is not following a comprehensive alarm management procedure, the alarm system may go out of control again in the future.

Managing the Change Including the alarm system within the scope of the plant Management of Change program will assist in retarding uncontrolled alarm growth.

References

1. "Alarm Systems, a Guide to Design, Management and Procurement," Engineering Equipment and Materials Users Assn. Publication No. 191.
2. Nimmo, I., "Abnormal Situation Management," *Chemical Engineering Progress,* September 1995.
3. "Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems," IEC 61508.
4. "Functional Safety: Safety Instrumented Systems for the Process Industry Sector," IEC 61511.
5. "Better Alarm Handling," *The British Health & Safety Executive,* 1999.
6. ISA TR91.00.02, "Criticality Classification Guideline — Enunciator Sequences and Specifications," ANSI/ISA S18.1 1992.

1.7 Speech Synthesis and Voice Recognition

D. H. F. LIU (1995) **B. G. LIPTÁK** (2005)

THE NATURE OF SOUND

Sound normally reaches the ear through pressure waves in the air. Vibratory motion at frequencies between 16 Hz and 20 kHz is recognized as audible sound. Sound pressure levels and velocities are quite small, as listed in Table 1.7a. Sound pressures are measured in root-mean-square (RMS) values. The RMS values are arrived at by taking the square root of the arithmetic mean of the squared instantaneous values.

The RMS pressure of spherical sound waves in a free field of air can be described as:

$$P_r = P_0/r \text{ dynes/cm}^2$$

where P_r is the RMS sound pressure at a distance r from the sound source and P_0 is the RMS pressure at a unit distance from the source.

INTRODUCTION

Several companies and universities have developed voice recognition systems. These systems and others built under grants from the Defense Advanced Research Projects Agency (DARPA) couple the ability to convert speech into electronic text with the artificial intelligence to understand the text. In these systems the computer acts as an agent that knows what the users want and how it is to be accomplished, including the voice output. Altogether, this type of speech synthesis allows untrained manufacturing workers the uninterrupted use of their hands, which could translate into increased productivity and cost savings.

SPEECH SYNTHESIS

Figure 1.7b shows the components of a computer voice response system. A synthesis program must access from a stored vocabulary a description of the required sequence of words. It must put them together, with the proper duration, intensity, and inflection for the prescribed text. This description of the connected utterances is given to a synthesizer device, which generates a signal for transmission over a voice circuit. Three different techniques can be used for generating the speech: (1) adaptive differential pulse-code modulation (ADPCM), (2) formant synthesis, and (3) test synthesis. Their typical bit rates for vocabulary storage are shown in Table 1.9c.

TABLE 1.7a
Mechanical Characteristics of Sound Waves

	RMS Sound Pressure ($dynes/cm^2$)	*RMS Sound Particle Velocity (cm/sec)*	*RMS Sound Particle Motion at (1,000 Hz cm)*	*Sound Pressure Level (dB 0.0002 bar)*
Threshold of hearing	0.0002	0.0000048	0.76×10^{-9}	0
	0.002	0.000048	7.6×10^{-9}	20
Quiet room	0.02	0.00048	76.0×10^{-9}	40
	0.2	0.0048	760×10^{-9}	60
Normal speech at 3'	2.0	0.048	7.6×10^{-6}	80
Possible hearing impairment	20.0	0.48	076.0×10^{-6}	100
	200	4.80	760×10^{-6}	120
Threshold of pain	2000	48.0	7.6×10^{-3}	140
Incipient mechanical damage	20×10^3	480	76.0×10^{-3}	160
	200×10^3	4800	760×10^{-3}	180
Atmospheric pressure	2000×10^3	48000	7.6	200

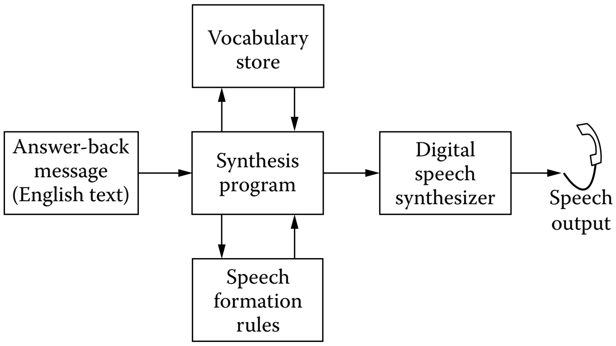


FIG. 1.7b

Block diagram of a computer voice response system. (Courtesy of Bell Laboratories.)

Adaptive Differential Pulse-Code Modulation (ADPCM)

Figure 1.7d shows the schematic of a simple ADPCM decoder. The synthesis program pulls the digitally coded words from the disk, where they are stored in the required sequence, and supplies them to an ADPCM decoder, which produces the analog signal. The only control of prosody (accent and voice modulation) that can be effected in the assembling of the message is the adjustment of time intervals between successive words, and possibly of their intensities. No control of voice pitch or merging of vocal resonance at word boundaries is possible.

This technique is suitable for messages requiring modest vocabulary size and very lax semantic constraints. Typical messages are voice readouts of numerals, such as telephone numbers, and instructions for sequential operations, such as equipment assembly or wiring.

Formant Synthesis

Formant synthesis uses a synthesis model in which a word library is analyzed and parametically stored as the time variations of vocal-tract resonance, or formants. Speech formants are represented as dark bars on sound spectrogram in the bottom panel of Figure 1.7e. These frequency variations and voice pitch periods can be analyzed by computer, as shown in the upper two panels of the figure.

TABLE 1.7c

Features of the Different Coding Techniques Used in Speech Generation

Coding Technique	*Data Rate (bits/s)*	*Duration of Speech in 10 Bits of Storage (min)*
ADPCM	20k	1
Formant synthesis	500	30
Text synthesis	700	240

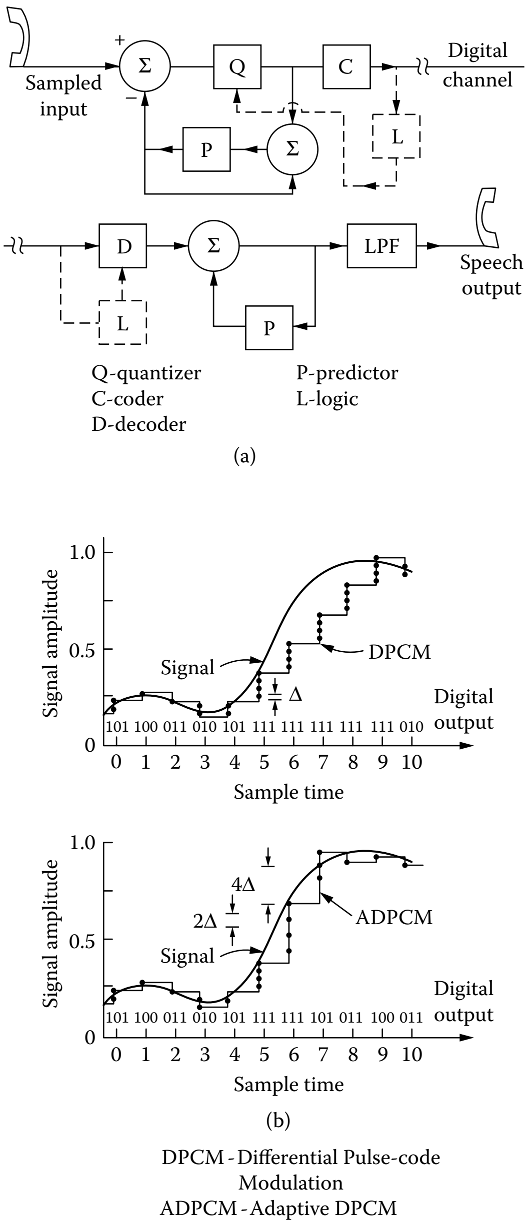


FIG. 1.7d

Adaptive differential coding of speech. (a) Block diagram of coder. (Adapted from U.S. Patent 3,772,682.) (b) Comparison of waveforms coded by 3-bit DPCM and ADPCM. (Courtesy of Bell Laboratories.)

In formant synthesis, the synthesizer device is a digital filter, as shown in Figure 1.7f. The filter's transmission resonance or antiresonance are controlled by the computer, and its input excitation is derived by programmed rules for voice pitch and for amplitudes of voiced sounds and noise-like unvoiced sounds. The parametric description of the speech information provides the control and modification of the prosodic characteristics (pitch, duration, intensity) of the synthetic speech and produces smooth formant transitions at successive word boundaries. This technique has also been

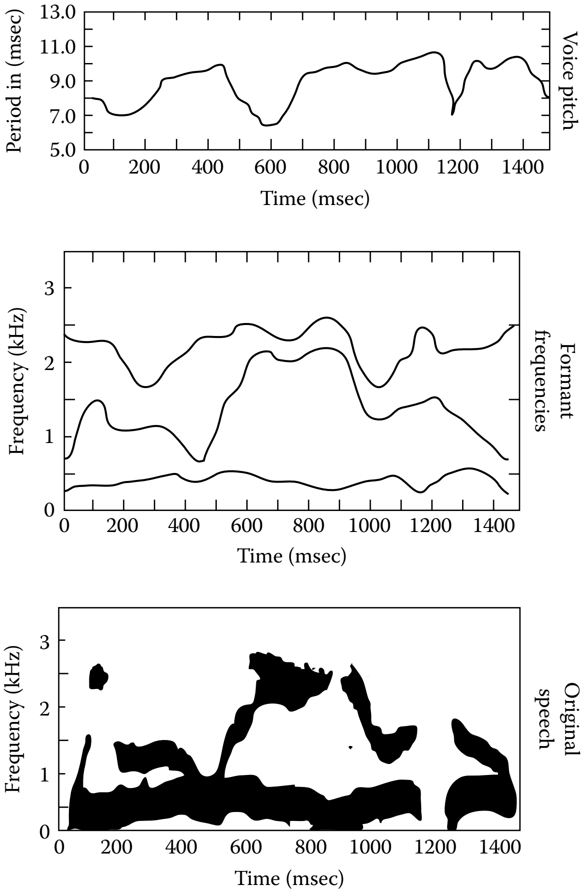


FIG. 1.7e

Analysis of the sentence "We were away a year ago." The sound spectrogram in the bottom panel shows the time variation of vocal-tract resonances or formants (dark areas). Computer-derived estimates of the formant frequencies and the voice pitch period are shown in the top two panels. (Courtesy of Bell Laboratories.)

used experimentally for generating automatic-intercept messages and for producing wiring instructions by voice.

Text Synthesis

Text synthesis is the most forward-looking voice response technique. It provides for storing and voice-accessing voluminous amounts of information or for converting any printed message into spoken form. It generates speech from an information input corresponding to a typewriter rate. In one design, the vocabulary is literally a stored pronouncing dictionary. Each entry has the word, a phonetic translation of the word, word stress marks, and some rudimentary syntax information.

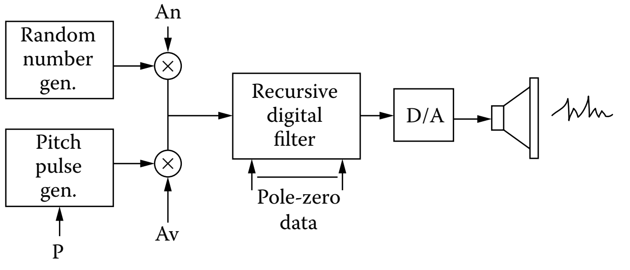


FIG. 1.7f

Block diagram of a digital formant synthesizer. (Courtesy of Bell Laboratories.)

The synthesis program includes a syntax analyzer that examines the message to be generated and determines the role of each word in the sentence. The stored rules for prosody then will calculate the sound intensity, sound duration, and voice pitch for each phoneme (the sound of human voice) in the particular context.

Figure 1.7g shows a dynamic articulatory model of the human vocal tract. The calculated controls cause the vocal-tract model to execute a sequence of "motions." These motions are described as changes in the coefficients of a wave equation for sound propagation in a nonuniform tube. From the time-varying wave equation, the formants of the deforming tract are computed iteratively. These resonant frequencies and the calculated pitch and intensity information are sent to the same digital formant synthesizer shown in Figure 1.7f. The system generates its synthetic speech completely from stored information. Language text can be typed into the system and the message can be synthesized online.

VOICE RECOGNITION

Figure 1.7h illustrates one general structure for a machine to acquire knowledge of expected pronunciations and to compare input data with those expectations. A precompiling stage involves the user's speaking sample pronunciations for each allowable word, phrase, or sentence, while identifying each with a vocabulary item number. Later, when an unknown utterance is spoken, it is compared with all the lexicon of expected pronunciations to find the training sample that it most closely resembles.

Word Boundary Detection

Silent pauses permit easy detection of word onsets at transitions from no signal to high energy and word offsets as energy dips below the threshold. Word boundary confusions can occur, when short silences during stop consonants (/p,t,k/) may resemble intended pauses. A word such as "transportation" could appear to divide into three words: "trans," "port," and "ation." Recognizers must measure the duration of the pause to distinguish short consonantal silences from longer deliberate pauses.

Feature Extraction

Acoustic data provide features to detect contrasts of linguistic significance and to detect segments such as vowels and consonants. Figure 1.7i shows a few typical acoustic parameters used in recognizers. One can extract local peak values, such as 1 in the top panel, as an amplitude measure. The sum of the squares

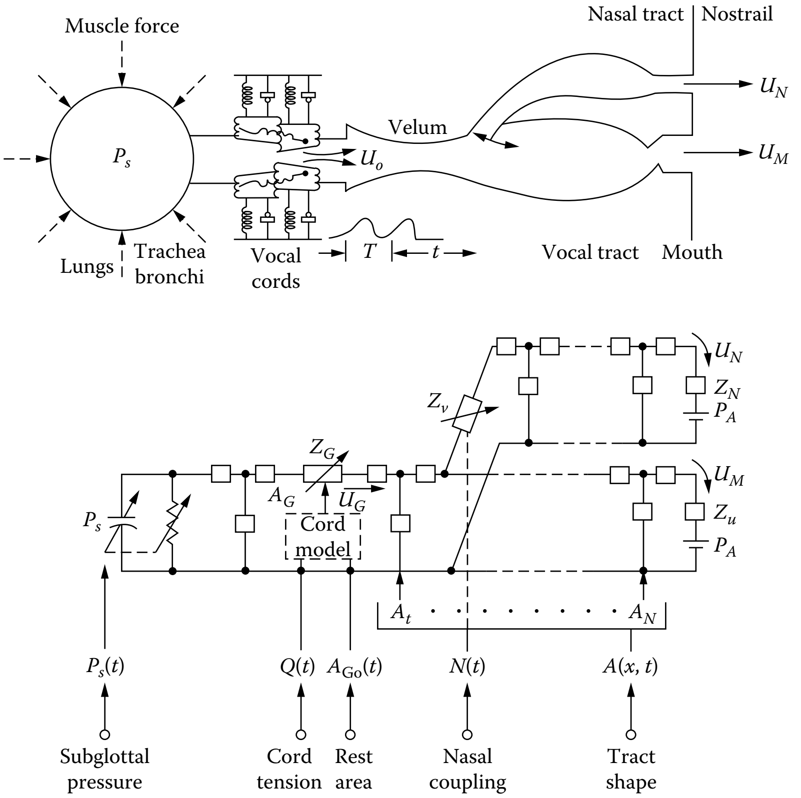


FIG. 1.7g

Speech synthesizer based upon computer models for vocal cord vibration, sound propagation in a yielding-wall tube, and turbulent flow generation at places of constriction. The control inputs are analogous to human physiology and represent subglottal air pressure in the lungs (P_s); vocal-cord tension (Q) and area of opening at rest (A_{Go}); the cross-sectional shape of the vocal tract $A(x)$; and the area of coupling to the nasal tract (N). (Courtesy of Bell Laboratories.)

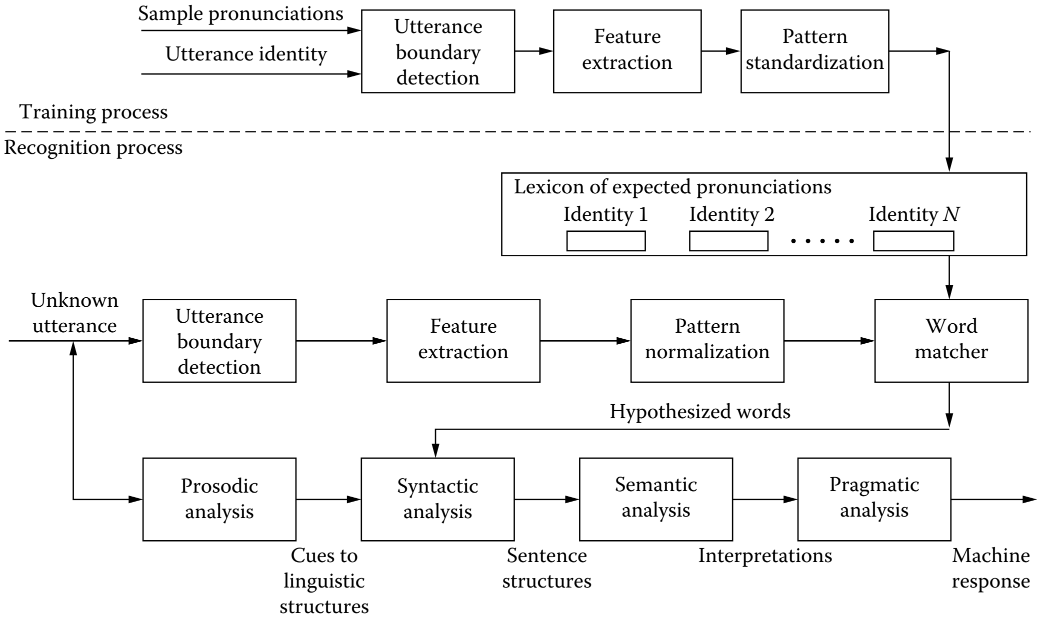


FIG. 1.7h

Typical structure for a trainable speech recognizer that could recognize words, phrases, or sentences.

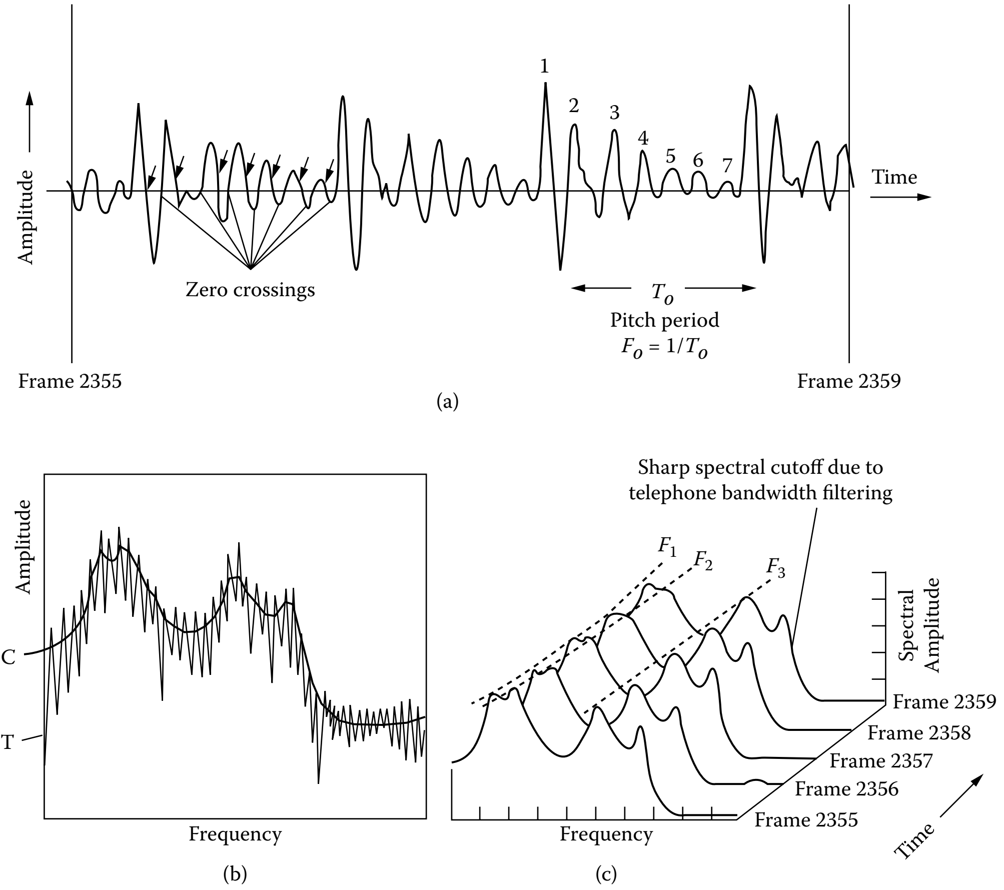


FIG. 1.7i

Typical acoustic parameters used in speech recognizers. (a) Time waveform showing parameters that can be extracted. (b) Frequency spectrum of the waveform of a, with the detailed frequency spectrum derived from a fast Fourier transform (FFT) smoothed by linear predictive coding (LPC) to yield a smooth spectrum from which formants can be found as the spectral peaks. (c) Smoothed LPC spectra for five successive short time segments, with formants F_1, F_2, and F_3 tracked as the spectral peaks.

of all waveform values over a time window provides a measure of energy. Vowel-like smooth waveforms give few crossings per unit time; noiselike segments give many crossings per unit time.

The time between the prominent peaks at the onsets of pitch cycles determines the pitch period T_0, or its inverse, the rate of vibration of the vocal cords, called fundamental frequency or F_0. Resonance frequencies are indicated by the number of peaks per pitch period. The third pitch period shows seven local peaks, indicating a ringing resonance of seven times the fundamental frequency. This resonance is the first formant or vocal tract resonance of the /a/-like vowel. It is the best cue to the vowel identity.

The frequency content of the short speech is shown in panel (a). The superimposed results of two methods for analyzing the frequency of speech sample are given in panel (b). The jagged Fourier-frequency spectrum, with its peak at harmonics of the fundamental frequency, is determined using a fast Fourier transform (FFT), and the exact positions of major spectral peaks at the formants of the speech are determined using an advanced method called linear predictive coding (LPC). The peaks indicate the basic resonance of the speaker's vocal tract and can be tracked as a function of time in panel (c). This indicates the nature of the vowel being articulated.

Pattern Standardization and Normalization

A word spoken on two successive occasions has different amplitudes. A recognizer must realign the data so that proper portions of utterance are aligned with corresponding portions of the template. This requires nonuniform time normalization, as suggested by the different phonemic durations. Dynamic programming is a method for trying all reasonable alignments and yields the closest match to each template. Another pattern normalization involves speaker normalization, such as moving one speaker's formants up or down the frequency axis to match those of a "standard" speaker that the machine has been trained for.

Word Matching

The above processes yield an array of feature values versus time. During training, this array is stored as a template of expected pronunciation. During recognition, a new array can be compared with all stored arrays to determine which word is closest to it. Ambiguities in possible wording result from error and uncertainties in detecting the expected sound structure of an utterance. To prevent errors in word identifications, recognizers will often give a complete list of possible words in decreasing order of agreement with the data.

High-Level Linguistic Components

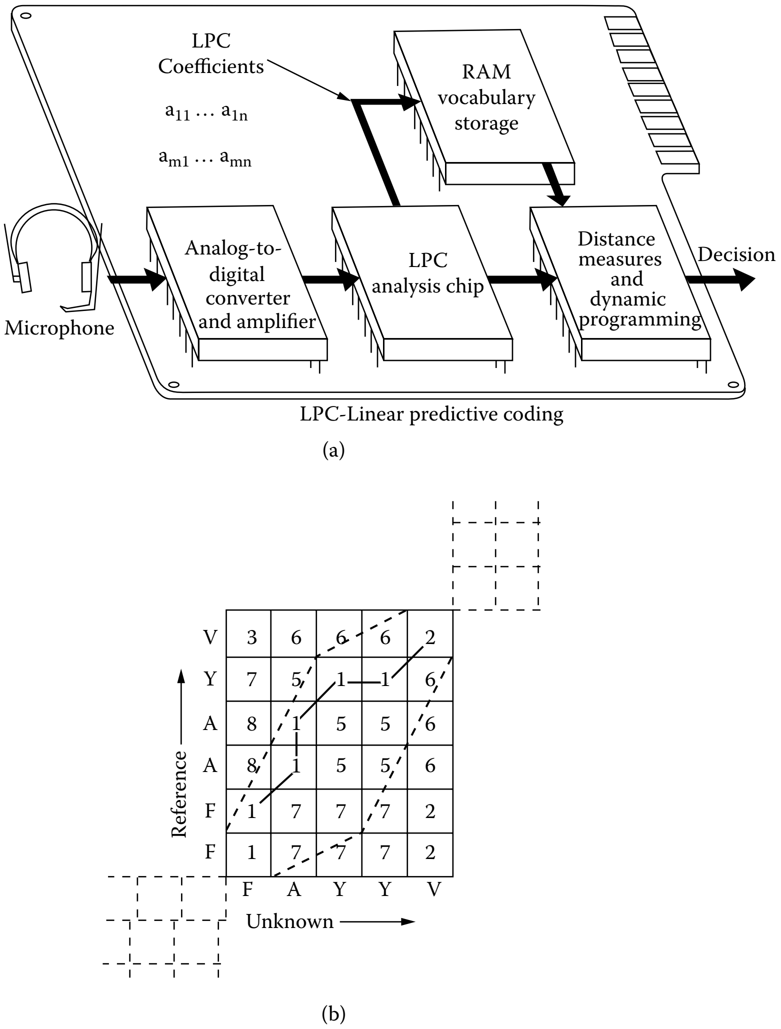
Speech recognizers need to limit the consideration of alternative word sequences. That is the primary purpose of high-level linguistic components such as prosodics, syntax, semantics, and pragmatics. Prosodic information such as intonation can help distinguish questions from commands and can divide utterances into phrases and rule out word sequences with incorrect stress patterns.

A syntactic rule can disallow ungrammatical sequences, like "plus divide" or "multiply clear plus." A semantic rule might disallow meaningless but grammatical sequences such as "zero divide by zero." A pragmatic constraint might eliminate unlikely sequences, such as the useless initial zero in "zero one nine."

It is possible to restructure the recognizer components to avoid propagation of errors through successive stages of a recognizer by having all the components intercommunicate directly through a central control component that might allow syntax or semantics to affect feature-extraction or word matching processes, or vice versa.

PRACTICAL IMPLEMENTATION

Figure 1.7j illustrates a typical circuit-board implementation of a method of word recognition using LPC coding for spectral data representation and dynamic programming for time alignment and word matching. An analog-to-digital converter



	F	A	Y	Y	V
V	3	6	6	6	2
Y	7	5	1	1	6
A	8	1	5	5	6
A	8	1	5	5	6
F	1	7	7	7	2
F	1	7	7	7	2

FIG. 1.7j

Practical speech synthesizer based on linear predictive coding (LPC) and dynamic programming. (a) Major components of a circuit-board speech recognizer using LPC analysis. (b) Matrix of speech sound differences between reference words and word portions of unknown inputs. Alternative alignments are allowed within the parallelogram, and the path with least accumulated distance (heavy line) is chosen for best alignment of reference and input words.

chip amplifies and digitizes the microphone handset signal. An LPC chip performs the acoustic analysis and determines the necessary coefficients for specifying an inverse filter that separates such smooth resonance structure from the harmonically rich impulses produced by the vocal cords.

A matrix of these LPC coefficients versus time represents the articulatory structure of a word. Word matching can be based on comparing such coefficients with those throughout each word of the vocabulary for which the recognizer is trained.

As shown in Figure 1.7j, the analysis frames of the input are on the horizontal axis; those of a candidate reference word from the training data are on the vertical axis. The distances between the respective frames of the input and reference words are entered into the corresponding intersection cells in the matrix. The distances along the lowest-accumulated-distance alignment of reference and input data are accumulated. If that distance is less than the distance of any other reference word inserted in the illustrated reference word, then that reference word is accepted as the identity of the input.

Dynamic programming is a method of picking that path through all successive distance increments that produces the lowest accumulated distance. As shown in the heavy line in Figure 1.7j, a couple of frames of "f"-like sounds of the reference word may be identified with a single "f"-like frame of the input, or two inputs framed of "y"-like sounds may be associated with one such frame in the reference.

Dynamic programming is also applicable to word sequences, looking for beginning and end points of each word, and matching with reference words by best paths between such end points. The same procedure can be applied to larger units in successive steps, to yield "multiple-level" dynamic programming, which is used in some commercial recognition devices.

Bibliography

Bristow, G., *Electronic Speech Recognition,* New York: McGraw-Hill, 1986.

Bristow, G., *Electronic Speech Synthesis,* New York: McGraw-Hill, 1984.

Burnett, D. C., et al., "Speech Synthesis Markup Language," http://www.w3.org/TR/speech-synthesis, copyright 2002.

Dutoit, T., *An Introduction to Text-To-Speech Synthesis,* Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.

Flanagan, J. L., "Synthesis and Recognition of Speech: Teaching Computers to Listen," *Bell Laboratory Record,* May–June 1981, pp. 146–151.

http://www.3ibm.com/software/speech/

Levinson, S. E., and Liberman, M. Y., "Speech Recognition by Computers," *Scientific American,* April 1981, pp. 64–76.

Yarowsky, D., "Homograph Disambiguation in Speech Synthesis," *Proceedings, 2nd ESCA/IEEE Workshop on Speech Synthesis,* New Paltz, NY, 1994.

1.8 Wiring Practices and Signal Conditioning

C. H. KIM (1970) **R. G. DITTMER** (1985) **B. G. LIPTÁK** (1995, 2005)

INTRODUCTION

Although a variety of signal transmission methods are discussed in Chapter 3 of this handbook, the majority of measurement signals are still transmitted over wires. These wires can be dedicated or shared as in the case of buses in digital networks, which are described in Volume 3 of this handbook.

Consequently, proper wiring practices play an important role in the total performance of process control systems, and this section is dedicated to that topic. The goal of signal conditioning is to protect the information content of the signal from various forms of noise and interference. The types of noise and some methods of noise protection have already been discussed in Volume 1, in Section 4.13, which covers the subject of thermocouple noise protection.

In this section, a systematic discussion of the many aspects of electric signal protection and conditioning will be provided. It will start with the definitions of the types of noise encountered and will continue with the description of many applications of noise reduction. These discussions will cover the protection of grounds and wiring, the guarding of amplifiers and power transformers, common-mode rejection in amplifiers, and the protection of some specific components, such as thermocouples, multiplexers, and A/D (analog-to-digital) converters. The discussion will follow a system approach, which will start from the transducer through the signal transmission line to the signal conditioner, multiplexer, and/or A/D converters.

ELECTRIC NOISE

The ability of a process control system to perform is directly dependent upon the quality of the measured variables. This quality is dependent upon the elimination or attenuation of noise that can deteriorate the actual transducer signal. The amount of noise present is expressed as the signal-to-noise ratio, which is the ratio of signal amplitude to noise amplitude, where these amplitudes are expressed as either peak or RMS values. In the worst case, the noise can actually be of higher amplitude than the transducer signal. This type of signal, of course, would be of no value, because its useful control information content is undetectable. The wiring practices and signal conditioning techniques described here are meant to remove the noise signal from the transducer signal, ensuring an accurate measured variable signal for process control.

Although a multitude of signal noise sources exist, the four main sources are:

1. Uncontrollable process disturbances that are too rapid to be reduced by control action
2. Measurement noise resulting from such factors as turbulence around flow sensors and instrument noise
3. Stray electrical pickup, such as from AC power lines, pulses from power switching, or RF (radio frequency) interference
4. A/D conversion error

Reduction of Noise

Noise adversely affects both analog and digital electronic equipment, but in digital systems the noise effects are compounded because of the scanning or snapshot nature of digital systems.

The typical frequency ranges of some noise components are shown in Figure 1.8a. Stray electrical pickup can be minimized by good engineering and installation practices such as proper shielding, screening, grounding, and routing of wires. Even so, noise will still exist at the input, but its effect on the controlled and manipulated variables can be significantly reduced through judicious selection of filtering.

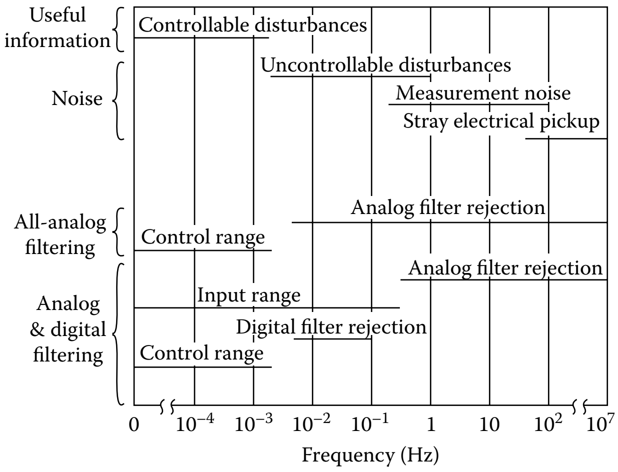


FIG. 1.8a
Frequency ranges of normal input signals, noise, and different filter types.

In most cases, relatively high-frequency noise (i.e., above approximately 5 Hz) is removed by conventional analog filters while the remaining noise occupying the range of 0.002 to 5 Hz may require either special analog filtering and/or digital filtering.

It is good practice to follow these three steps to reduce noise:

1. Estimate the dynamic characteristics (amplitude and bandwidth) of the noise.
2. Determine its effect on the controlled and manipulated variables.
3. Select the best filtering option (analog and/or digital filtering).

This section is devoted to a close examination of noise suppression techniques. Noise suppression is a system problem, and its discussion therefore should not be limited to signal transmission only but must be applied consistently from the transducer through the signal transmission line to the signal conditioner, multiplexer, and/or analog-to-digital converters.

There are several ways to reduce the effects of noise on control systems, including line filtering, integrating, digital filtering, and improving the signal-to-noise ratio, but none of these will completely solve the problem without proper cable selection and installation.

Careful examination of the various sources of noise shows that some can be eliminated by proper wiring practice. This implies good temperature compensation, perfect contacts and insulation, carefully soldered joints, use of good-quality solder flux, and use of the same wire material. However, additional precautions are needed to eliminate pickup noise entirely.

Noise Types

As shown in Figure 1.8b, signal leads can pick up two types of external noises: normal-mode and common-mode interference.

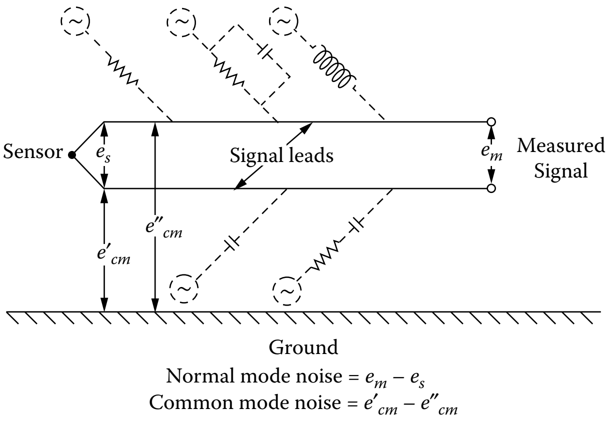


FIG. 1.8b
Electrical noise interference on signal leads.

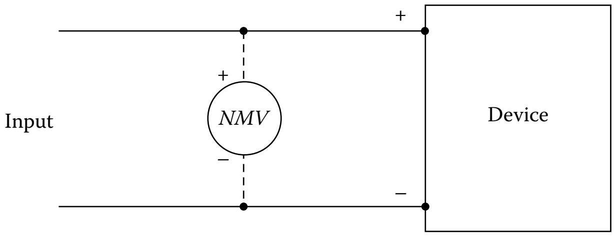


FIG. 1.8c
Definition of normal-mode voltage (NMV).

Interference that enters the signal path as a differential voltage across the two wires is referred to as normal-mode interference (Figure 1.8c). The normal-mode noise cannot be distinguished from the signal, which is coming from the transducer.

Common-mode voltage is a voltage of the same polarity on both sides of a differential input relative to ground (Figure 1.8d). Interference that appears between ground and both signal leads, as an identical voltage, is referred to as common-mode interference. Common-mode rejection is the ability of a circuit to discriminate against a common-mode voltage. This can be expressed as a dimensionless ratio, a scalar ratio, or in decibels as 20 times the $\log_{10}$ of that ratio.

Noise Sources

The most common sources of noise interference are described below:

Inductive Pickup This category includes 60-Hz noise from power lines and 120-Hz noise from fluorescent lighting, as well as high-frequency noise generated by electric arcs and pulse transmitting devices. The worst man-made electrical noise results from the opening or closing of electrical circuits that contain inductance. The amplitude of the transients generated in an inductive circuit is given in Equation 1.8(1) below:

$$E_i = L(d_i/d_t) \qquad \textbf{1.8(1)}$$

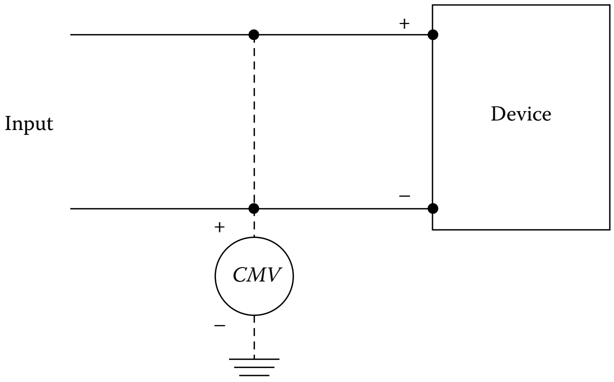


FIG. 1.8d
Definition of common-mode voltage (CMV).

TABLE 1.8e
Average Noise Conditions in Various Industries

Description		Industry Chemical	 Steel	 Aerospace
Normal signal levels (mV)		10 to 100	10 to 100	10 to 1000
Possible or expected noise level	Normal mode (mV)	1 to 10	2 to 7	1 to 10
	Common mode (V)	4 to 5	4 to 5	4 to 5
Desired noise rejection	Normal mode	10^3:1	10^3:1	10^3:1
	Common mode	10^6:1	10^6:1	10^6:1

This amplitude is calculated as the inductance (L) times the rates of change in the switched current. Closing an inductive circuit can cause a transient of twice the input voltage, and opening such a circuit can cause a transient as high as 10 times the supply voltage. When switching inductive loads from an AC supply, the noise transients are proportional to the instantaneous value of the AC supply.

Common-Mode Noise Inadequate common-mode rejection due to line unbalance will convert the common-mode noise interference into a normal-mode noise signal.

Table 1.8e shows the average noise conditions that exist in various industries. Some chemical plants have experienced as high as 60 volts common-mode noise interference. Electrical noise interference can be a severe problem in industries such as steel, power, and petroleum, where power consumption is high and complex electrical networks exist.

Other Noise Sources Other noise sources include:

Failure to distinguish between a ground and a common line. The ground currents that exist in a "daisy-chain" equipment arrangement are primarily noise signals induced into the equipment, or perhaps generated by some equipment, and they find their way back to earth. The resulting voltage drops in the ground wire place each piece of equipment at a different potential with respect to earth. This potential difference is easily coupled into the signal lines between equipment.

Common impedance coupling "ground loops." Placing more than one ground on a signal circuit produces a ground loop that is a very good antenna, and noise signals induced in the loop are easily coupled into the signal lines. This can generate noise that will completely obscure the useful signal.

Electrostatic coupling to AC signals. The distributed capacity between signal conductors and from signal conductors to ground provides a low-impedance path for cross-talk and for signal contamination from external sources.

Ineffective temperature compensation. Ineffective temperature compensation in transducers, lead wire systems, amplifiers, and measuring instruments can create changes in system sensitivity and drift in their zero. This is especially important in strain gauge transducers, resistance-type temperature sensors, and other balanced-bridge devices.

Loading the signal source. When a transducer or other signal source is connected in parallel with an amplifier or measuring device with low input impedance, the signal voltage is attenuated because of the shunting effect. This can considerably decrease the sensitivity of the system.

Variable contact resistance. All resistance-type transducers, as well as bridge-type circuits, are susceptible to changes in contact resistance. The measuring instrument is unable to distinguish between a resistance change in the sensor and a resistance change in the external wiring.

Conducted AC line transients. Large voltage fluctuations or other forms of severe electrical transients in AC power lines (such as those caused by lightning) are frequently conducted into electronic systems by AC power supply cords.

Conduction pickup. Conduction pickup refers to interfering signals that appear across the input terminals of the receiver, because of leakage paths caused by moisture, poor insulation, etc.

Thermoelectric drift. The junction of two dissimilar metal wires will create thermoelectric drift, which varies with changes in temperature. This is more critical in DC circuits, which operate at the micro volt level.

Electrochemically generated corrosion potentials. These potentials can often occur, especially at carelessly soldered junctions.

Use of the same common line. Using the same common line for both power and signal circuits or between two different signal lines can cause the appearance of transients on the signal and can also cause cross-talk between the two signals.

GROUNDING, WIRING, FILTERING

Ground Systems

Good grounding is essential for normal operation of any measurement system. The term "grounding" is generally defined as a low-impedance metallic connection to a properly designed ground grid, located in the earth. In large equipment, it is very difficult to identify where the ground is.

On standard all-steel racks less than 6 ft (1.8 m) in length, differences in potential of up to 15 volts peak to peak have been measured. Stable, low-impedance grounding is necessary to attain effective shielding of low-level circuits, to provide a stable reference for making voltage measurements, and to establish a solid base for the rejection of unwanted common-mode signals.

In a relatively small installation, two basic grounding systems should be provided. First, all low-level measurements and recording systems should be provided with a stable system ground. Its primary function is to assure that electronic enclosures and chassis are maintained at zero potential. A satisfactory system ground can usually be established by running one or more heavy copper conductors to properly designed ground grids or directly to earth grounding rods.

Signals can be measured with respect to the system reference ground only if the input signals are fully floating with respect to ground. In this case, the stable system ground fulfills the task of providing a base for common-mode noise rejection.

Signal Ground The other important ground is the signal ground. This system is necessary to ensure a low-noise signal reference to ground. This ground should be a low-impedance circuit providing a solid reference to all low-level signal sources and thus minimizing the introduction of interference voltages into the signal circuit.

The signal ground should be insulated from other grounding systems, and it is generally undesirable to connect it to the system ground at any point (Figure 1.8f). In a single-point grounding system, no current flows in the ground reference, and if the signal cable is properly selected, noise due to large and hard-to-handle low-frequency magnetic fields will not exist. It should be emphasized that a signal circuit should be grounded at one point and at one point only, preferably at the signal source (Figure 1.8g).

By connecting more than one ground to a single signal circuit, as shown in Figure 1.8f, a ground loop is created. Two separate grounds are seldom, if ever, at the same potential. This differential generates a current flow that is in series with the signal leads. Thus the noise signal is combined with the useful signal. These ground loops are capable of generating noise signals that can be 100 times larger than the typical low-level signal.

In off-ground measurements and recording, the cable shield is not grounded, but it is stabilized with respect to the useful signal though a connection to either the center tap or the low side of the signal source. Appropriate insulation is needed between the shield and the outside of the cable because the shield is driven by a voltage, which is off ground.

It is important that electric racks and cabinets be connected to a proper system ground and not be allowed to contact any other grounded element in the building. Guidelines on grounding can be summarized as follows:

1. Intentional or accidental ground loops in either the signal circuit or the signal cable shield will produce excessive electrical noise in all low-level circuits and will destroy the useful signal.
2. Every low-level data system should have a stable system ground and a good signal ground.
3. The signal circuit should be grounded at only one point.
4. The signal cable shield should not be attached to more than one grounding system.
5. A floating signal circuit and its signal cable shield should always be grounded at the signal source only.

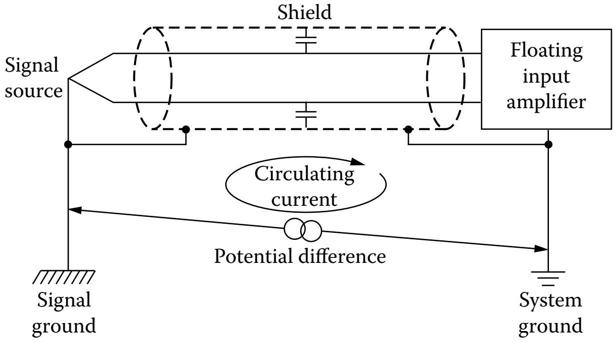


FIG. 1.8f
An example of the incorrect grounding of floating signal circuit. The ground loop is created by multiple grounds in a circuit and by grounding the shield at both ends.

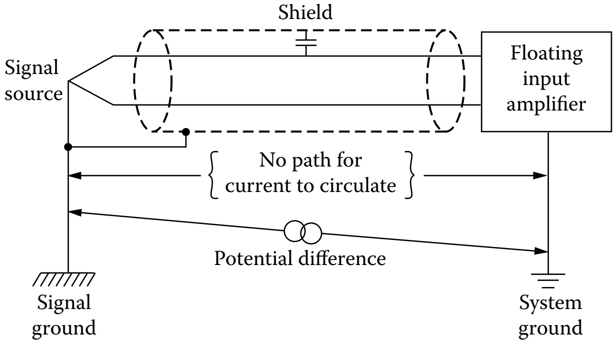


FIG. 1.8g
Illustration for showing the correct grounding of a floating signal circuit. By grounding the shield at the signal end only, the ground loop through the signal lead is eliminated.

Grounding and Safety A low-resistance, noncurrent-carrying metallic connection to ground should be established and maintained from every exposed metallic surface that can possibly become connected to an electrical circuit. Electrical connection could occur because of a fault, such as a loose wire making electrical contact, or as a result of leakage of current through insulation.

Grounding is usually accomplished by bonding all elements together in a system terminated at the ground connection where power enters the premises. It may be a bare or green insulated wire. More often it is the conduit enclosing the wires. It must be securely joined, electrically and mechanically, to each piece of equipment. It is connected at the service entrance to the grounded circuit conductor (white wire) and to ground. Instead of connection to a ground at the entrance connection, other suitable earth ground connections are acceptable. Equipment mounted directly on the structural steel of a building is considered effectively grounded. Water pipes have been considered effective grounds. However, because of the increasing use of plastic pipe for water connection, this is no longer an unquestionable ground.

Grounding serves two distinct safety-related purposes. First, since the ordinary power circuit has one side grounded, a fault that results in electrical contact to the grounded enclosure will pass enough current to blow a fuse. Second, the possibility of a shock hazard is minimized since the low-resistance path of a properly bonded and grounded system will maintain all exposed surfaces at substantially ground potential.

Grounding is effective against hazards from leakage currents. All electrical insulation is subject to some electrical leakage. This may rise to a significant level as insulation deteriorates with age or as layers of conductive dust accumulate in the presence of high humidity. A proper grounding system with low electrical resistance will conduct leakage currents to ground without developing significant potential on exposed surfaces.

Grounding of exposed metal surfaces is distinct from the grounded conductor of ordinary power wiring. The latter is a current-carrying ground, capable of developing significant potential, particularly on long lines and with surge or even short-circuit currents. Grounding systems are substantially noncurrent carrying, except for possible leakage currents.

Potential can build up only during the time required for a fuse to blow as a result of a specific fault that results in direct contact between power and grounding systems. Grounding is customarily not required for signal circuits where either maximum voltage is 30 V, or maximum current under any circumstance cannot exceed 5 mA.

Wiring

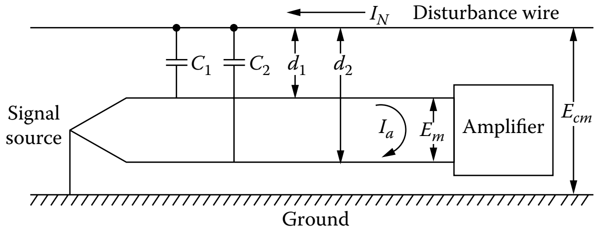
Another important aspect of reducing noise pickup involves the wiring system used to transmit the signal from its source to the measuring device or computer.

In less demanding low-frequency systems, where the signal bandwidth is virtually steady state and system accuracy requirements are not very high, two-wire signal leads will normally suffice. Otherwise, a third wire, or shield, becomes necessary.

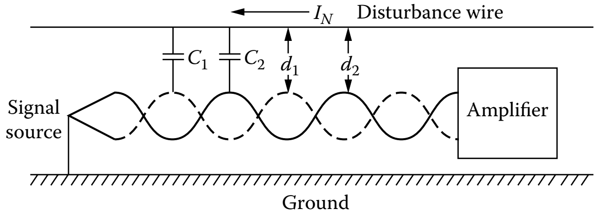
Shielding Where top performance is required, the shield is run all the way from the signal source to the receiving device. As already mentioned, the shield should be grounded at the signal source and not at the receiver because this arrangement provides maximum rejection of the common-mode noise.

The cable shield reduces electrostatic noise pickup in the signal cable, improves system accuracy, and is indispensable in low-level signal applications where high source impedance, good accuracy, or high-frequency response is involved. As the signal frequency approaches that of the noise, which is usually at 60 Hz, filtering can no longer be used to separate noise from the useful signal. Therefore, the only practical solution is to protect the signal lines and prevent their noise pickup in the first place.

Elimination of noise interference due to magnetic fields can be accomplished by wire-twisting (transpositions). If a signal line consisting of two parallel leads is run along with a third wire carrying an alternating voltage and an alternating current, the magnetic field surrounding the disturbing line will be intercepted by both wires of the signal circuit. Since these two wires are at different distances from the disturbing line, a differential voltage will be developed across them. If the signal wires are twisted (Figure 1.8h), the induced disturbing voltage will have the same magnitude and cancel out.



Where $d_1 \neq d_2$. Thus, $C_1 \neq C_2$, $\therefore$ difference in distance from disturbance wire creates a differential voltage, and a noise signal will be induced into the signal leads.



Where $d_1 = d_2$. Thus $C_1 = C_2$, $\therefore$ the induced disturbing voltages will have the same magnitude and cancel out when twisted wires are used.

FIG. 1.8h
The use of twisted wires eliminates the noise interference that might otherwise be caused by the magnetic fields of disturbing wires.

In order to prevent noise pickup from electrostatic fields, an effective shield must surround the low-level signal conductors. One type of shield consists of a woven metal braid around the signal pair, which is placed under an outside layer of the insulation. This type of shield gives only 85% coverage of the signal line and is adequate for some applications if the signal conductors have at least ten twists per foot (33 twists per meter). Its leakage capacity is about 0.1 picofarad per foot (0.3 picofarad per meter).

At the microvolt signal levels, this kind of shielding is not satisfactory. Another type of signal cable is shielded with a lapped foil (usually aluminum-Mylar tape) type shield, plus a low-resistance drain wire. This type of shielding provides 100% coverage of the signal line, reducing the leakage capacity to 0.01 picofarad per foot (0.03 picofarads per meter). To be most effective, the wires should have six twists per foot (20 twists per meter).

The lapped foil shield is covered with an insulating jacket to prevent accidental grounding of the shield. Care should be exercised to prevent the foil from opening at bends.

This type of low-level signal cable provides the following benefits:

1. There is an almost perfect shield between the signal leads and ground.
2. Magnetic pickup is very low because the signal leads are twisted.
3. Shield resistance is very low because of the low-resistance drain wire.

Feedback Currents and Shield Grounding In most measurements a voltage difference is detected relative to a reference potential. This is usually achieved by sending the signal through an amplifier, which provides the required gain. Such single-ended amplifiers will change the measurement signal in relation to the same reference potential, which is maintained both at their inputs and outputs.

With charge- and voltage-detecting single-ended amplifiers, the nonsignal related and unbalanced current flow, which is normal-mode noise pickup and is flowing in the zero reference conductor, must be limited. When such an amplifier and its power supply are suspended within a shield that they do not contact, capacitive couplings will occur. As it is shown in Figure 1.8i, because these couplings are at different voltages, they will cause a current flow back to the inputs (feedback) of the amplifier.

If the signal zero reference is grounded (as in Figure 1.8j) and allowing the shield to float does not solve the problem. This is because a current flow will be produced in the input side of conductor 3 by the potential difference of the earth between conductor 3 and the shield capacitance at the opposite end.

Figure 1.8k shows the solution to eliminating such feedback currents. Directly referencing the signal common to the shield will eliminate the capacitive feedback to the amplifier.

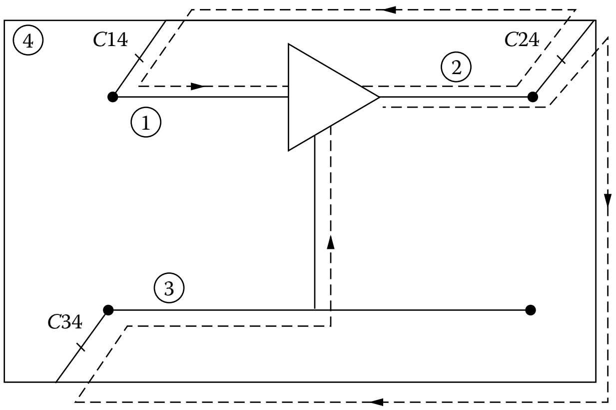


FIG. 1.8i
When the shield is ungrounded, the single-ended amplifier will receive feedback currents.

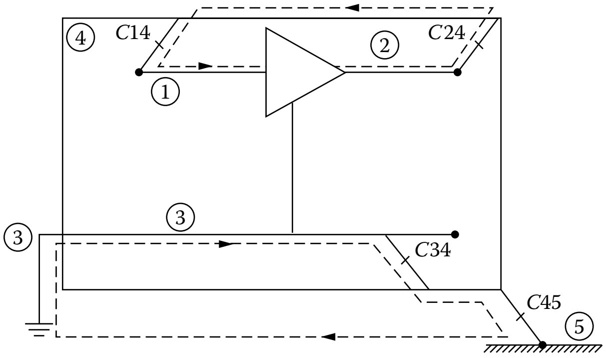


FIG. 1.8j
The feedback current problem is not solved by grounding the common conductor if the shield is not grounded.

Digital Communication Shielding Some 20 digital control system buses are in use worldwide. Table 1.81 lists their requirements for shielding. Most use unshielded twisted pair wiring (UTP). While only a minority requires the use of shielded twisted pairs (STP), shielding will improve the performance of all of them.

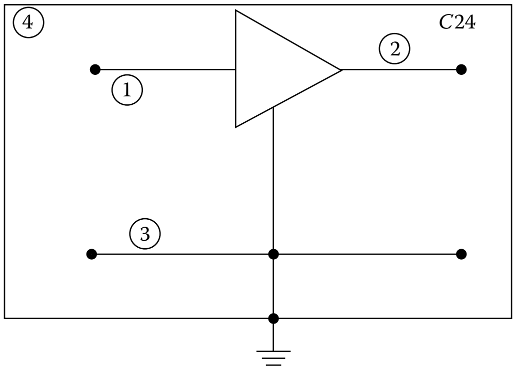


FIG. 1.8k
The solution is to properly ground both the shield and the common conductor.

TABLE 1.8l
Digital Network Cable Specifications

Network Name	*Associated Standards*	*Cable*
PROFIBUS DP/PA	EN 50170/DIN 19245 part 3 (DP), part 4(PA), IEC 1158-2(PA)	UTP
INTERBUS-S	DIN 19258, EN50.254	UTP
ControlNet		COAX
DeviceNet	ISO 11898 and 11519	UTP, COAX
Remote I/O		UTP
ARCNET	ANSI/ATA 878.1	UTP
AS-I		UTP
Foundation Fieldbus—H1	ISA SP50/IED 61158	UTP
Foundation Fieldbus—HSE	IEEE 802.3u, TCP and UDP	UTP
IEC/ISA SP50 Fieldbus	IEC 1158/ANSI 850	UTP
Seriplex		4 wire shielded
WorldFTP	IEC 1158-2	UTP
SDS	ISO 11989	UTP
CANopen		UTP
Ethernet	IEEE 802.3	UTP
Modbus Plus		UTP
Modbus	EN 1434-3, IEC 870-5	UTP
Genius Bus		STP

Often, whole cabinets or complete rooms are also shielded. Raised computer floors with electrically bonded stringers are regularly used to help with this task. Zero signal reference grids (ZSRG) are also provided by such building structural elements as Q-decking and rebar. The proper bonding of non-continuous grids, the use of special gaskets, and wave-guide ventilation covers are usually also desirable.

Shielded rooms should usually be grounded at a single point with respect to the outside world. This requires the installation of surge protectors so that bridging is provided for the isolation gaps during lightning strikes.

Wire Costs and Installation Typical costs of the three types of twisted signal wires are shown in Table 1.8m. The use of ordinary conduit is not likely to reduce noise pickup. Its serious limitation is in obtaining or selecting a single ground point. If the conduit could be insulated from its supports, it would provide a far better electrostatic shield. The following general rules should be observed in installing low-level signal circuits:

TABLE 1.8m
Costs of Three Twisted Signal Wire Designs

Recommended for the Following Type Noise Source	*Single Pair Cable Type*	*Cost ($) per Foot (per Meter)*
Magnetic fields	Twisted	.20 (.60)
Magnetic and electrostatic	Twisted with metal braid; 85% shield coverage	.40 (1.20)
Magnetic and electrostatic	Twisted with aluminum/Mylar; 100% shield coverage	.25 (.75)

1. Never use the signal cable shield as a signal conductor, and never splice a low-level circuit.
2. The signal cable shield must be maintained at a fixed potential with respect to the circuit being protected.
3. The minimum signal interconnection must be a pair of uniform, twisted wires, and all return current paths must be confined to the same signal cable.
4. Low-level signal cables should be terminated with short, untwisted lengths of wire, which expose a minimum area to inductive pickup.
5. Reduce exposed circuit area by connecting all signal pairs to adjacent pins in the connector.
6. Cable shields must be carried through the connector on pins adjacent to the signal pairs.
7. Use extra pins in the connector as a shield around signal pairs by shorting pins together at both ends and by connecting to the signal cable shield.
8. Separate low-level circuits from noisy circuits and power cables by a maximum physical distance of up to 3 ft (0.9 m) and definitely not less than 1 ft (0.3 m).
9. Cross low-level circuits and noisy circuits at right angles and at the maximum practical distance.

10. Use individual twisted shielded pairs for each transducer. Thermocouple transducers may be used with a common shield when the physical layout allows multiple pair extension leads.
11. Unused shielded conductors in a low-level signal cable should be single-end grounded with the shield grounded at the opposite end.
12. High standards of workmanship must be rigidly enforced.

Filtering

Filtering is required to stabilize the input signal and to remove AC noise components (particularly 60-Hz noise) resulting from direct connection of normal-mode AC signals, from normal-mode noise pickup, or from conversion of common-mode noise to normal-mode noise due to line unbalance.

A reliable, low-cost, and effective filter for analog inputs is the balanced resistance–capacitance filter, shown in Figure 1.8n. Its ability to eliminate AC components increases exponentially with the frequency of the noise signal components. Common-mode noise rejection of about 40 decibels is possible with this type of filter, with decibel defined as

$$\text{decibel} = 20\log\frac{\text{inlet noise amplitude}}{\text{outlet noise amplitude}} \qquad \mathbf{1.8(2)}$$

Filtering action causes a time delay (time constant, $T = RC$) between a signal change at the transducer and the time of recognition of this change by the measuring device. As the time delay may be on the order of 1 s or more, there are situations in which this has to be reduced (systems with high-frequency response). In order to decrease only the time delay but not the filtering efficiency, inductance–capacitance filters may be used (Figure 1.8o). This increases the filter cost.

Since a filter limits the bandwidth of the transmitted signal, it might be desirable to use more complicated and expensive filters when higher-frequency AC transducer signals are involved. Fortunately, this type of situation seldom occurs.

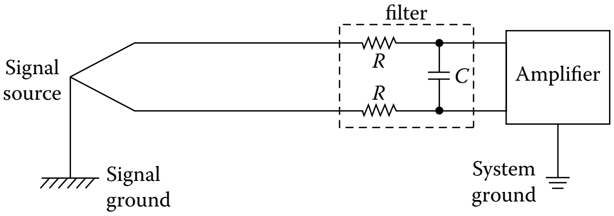


FIG. 1.8n
For analog signals, a low-cost and effective filter design is the balanced resistance–capacitance one.

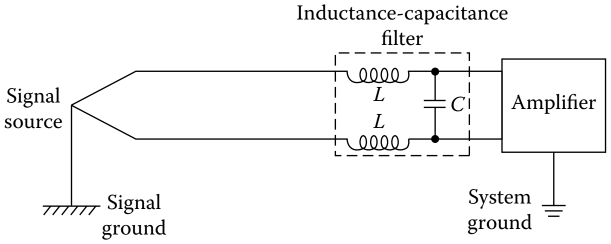


FIG. 1.8o
The use of an inductance–capacitance filter reduces the time delay while maintaining filtering efficiency.

APPLICATIONS

The following paragraphs discuss the application of noise protection techniques to transducers, thermocouples, multiplexers, and A/D converters.

Transducer Signals

The common-mode rejection of a good low-level signal amplifier is on the order of 10^6 to 1. This value is always decreased when the signal source and input signal leads are connected to the amplifier. Signal source impedance and signal lead impedance have a shunting effect on the signal input to the amplifier. From the point of view of maximum power transfer, equal impedance on both sides of a circuit is an ideal termination.

In the transmission of voltage signals, however, current flow has to be reduced to as close to zero as possible. This is achieved by selecting amplifiers with high input impedance and using transducers with low output impedance. Good instrumentation practice dictates that the input impedance of the amplifier be at least ten times the output impedance of the signal source.

However, low signal source impedance has an adverse effect on the pickup signal, because the low impedance of the transducer will shunt it. An additional step is to ground the transducer whenever possible and to run a balanced signal line to the amplifier input.

There are three basic techniques to condition transducers with high output impedance: high-impedance voltage amplifiers, charge amplifiers, and transducer-integrated unloading amplifiers.

Voltage Amplifiers Voltage amplifiers must have an amplifier input impedance that is very high compared with the source impedance. In this way, the amplifier's effect on the phase and amplitude characteristics of the system is minimized. A typical transducer with output amplifier is shown in Figure 1.8p. High impedance is practical for an amplifier, but an equally significant portion of the load is in the interconnecting cable itself, and it does not take much cable to substantially lower the available voltage.

In applying a voltage amplifier, low-frequency system response must be considered. The voltage amplifier input

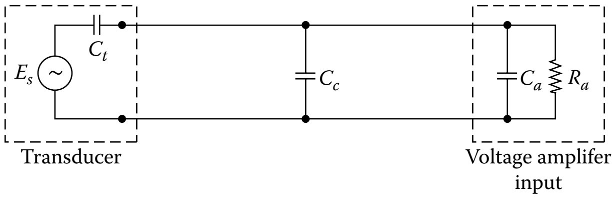


FIG. 1.8p
A typical transducer with voltage output amplifier. E_s = transducer voltage source; C_t = transducer capacitance; C_c = signal cable shunt capacitance; C_a = amplifier input capacitance; R_a = amplifier input resistance.

resistance (R_a) in combination with the total shunt capacitance forms a high-pass first-order filter with a time constant (T) defined by

$$T = R_a(C_t + C_c + C_a) \qquad \textbf{1.8(3)}$$

Cut-off frequency can become a problem when it approaches information frequency at low source capacitance (short cable or transducers with very low capacitance) or at lower amplifier-input resistance.

Charge Amplifiers Charge amplifiers have been widely used in recent years. This approach avoids the cable capacitance effects on system gain and frequency response. The typical charge amplifier shown in Figure 1.8q is essentially an operational amplifier with integrating feedback. A charge amplifier is a device with a complex input impedance that includes a dynamic capacitive component so large that the effect of varying input shunt capacitance is swamped, and the output is the integral of the input current.

Filtering of the resultant signal on both the low and high ends of the information band is desirable at times so that it is possible to get a rather high order of rejection without affecting information by using band pass filters. The addition of a resistor in parallel with the feedback capacitor in a charge amplifier will decrease the closed-loop gain at low frequencies, resulting in the desired high-pass filter characteristics.

Unloading Amplifiers Unloading amplifiers integrally mounted in a transducer housing (Figure 1.8r) have become available from transducer manufacturers because neither voltage amplifiers nor charge amplifiers offer a very satisfactory solution to the conditioning problem for systems with very high input capacitance (usually a result of very long lines).

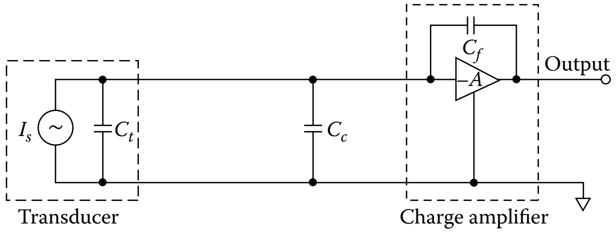


FIG. 1.8q
Transducer with charge output amplifier. I_s = transducer current source; C_f = feedback capacitance; C_t = transducer capacitance; C_c = signal cable shunt capacitance.

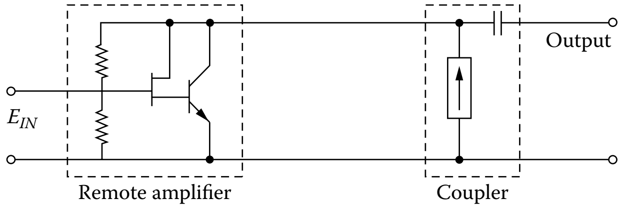


FIG. 1.8r
In comparison to voltage or to charge amplifiers, the unloading amplifier, which is located at the transducer, reduces the input capacitance.

With the voltage amplifier, signal-to-noise ratio suffers, because capacitance loading decreases the available signal. In a charge amplifier, the signal-to-noise ratio suffers because of the increased noise level (input noise is a direct function of input capacitance). Thus, remote signal conditioning appears to offer a satisfactory solution to the accommodation of long data lines. If closely located to the transducer, voltage-responding or charge-responding amplifiers are equally effective.

However, these techniques decrease the dynamic range capability (changing input amplifier gain to accomplish a range change is not possible) and restrict the high-frequency response signal amplitudes due to the limited current capability of the remote amplifier. When these two limitations are overcome, almost all of the signal-conditioning equipment will be at the remote location.

In summary, the following steps are recommended for low-level signals:

1. Select a signal source with low output impedance.
2. Select an amplifier (or measuring device) with high output impedance.
3. Use a balanced line from signal source to amplifier input (maximum allowable unbalance is 100 ohm/1000 ft or 0.33 ohm/m).
4. Keep signal cables as short as possible.
5. Use remotely located amplifiers when long signal cables are required (except for thermocouple and RTD signals).
6. Select a signal source that can be grounded (thermocouples, center-tapped sensors, etc.).

It is evident that common-mode rejection must be maintained at a high level in order to attain noise-free results from low-level signal sources.

Differential Amplifier An example where two zero references exist is a detector in a self-powered transmitter, which is interfaced to a digital I/O system. Other systems requiring

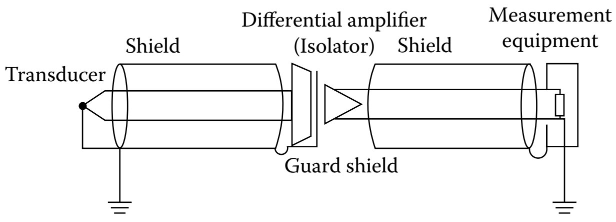


FIG. 1.8s
The application of a differential amplifier, which is also called an isolator.

differential amplifiers are grounded thermocouples, strain gauges, and data link applications.

It often happens that a common signal is the I/O reference to earth, while a self-powered transmitter generates an output, which is referenced to the ground of its own power supply. The difference between the zero references of the I/O and the transmitter's power supply will appear as common mode noise, which can cause normal-mode pickup.

In such cases, one should use a differential and not a single-ended I/O. If a differential I/O cannot be used, an external differential amplifier, also called an isolator, can be installed in the signal lines. In that case, each side of this isolator should be independently shielded and each shield should be grounded to its respective signal ground. Figure 1.8s illustrates the application of a differential amplifier.

Amplifier Guard In a normal measuring device, a signal line is connected to a differential DC amplifier with floating inputs. Generally, these amplifiers are provided with an internal floating shield that surrounds the entire input section, as shown in Figure 1.8t. This floating internal shield is called a "guard shield" or simply a "guard."

The guard principle requires that the amplifier guard be driven at the common-mode voltage appearing at the amplifier inputs. The most effective way to do this is as follows:

1. Connect the amplifier guard to the signal cable shield and make sure that the cable shield is insulated from the chassis ground or from any extension of the system ground.
2. Connect the ground of the signal source to the signal cable shield. In this way, the amplifier guard and signal cable shield are stabilized with respect to the signal from the source.
3. Connect the signal cable shield and its tap to the signal source and also to the signal ground, which should be as close as possible to the signal source. This will limit the maximum common-mode voltage. The signal pair must not be connected to ground at any other point.
4. Connect the amplifier chassis, the equipment enclosure, the low side of amplifier output, and the output cable shield to the system ground.

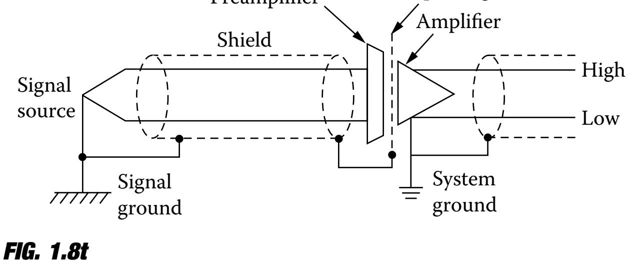


FIG. 1.8t
Correct grounding of an isolating amplifier.

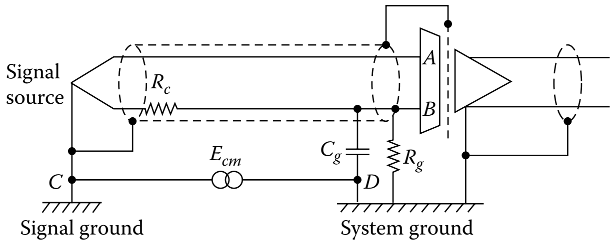


FIG. 1.8u
The design of a guard-shielded measuring system. R_c = cable unbalance resistance; C_g = measuring circuits/system ground capacitance; R_g = measuring circuits/system ground resistance; E_{cm} = common-mode voltage.

Power Transformer Guard It is important to avoid strong magnetic fields emanating from the power supply transformer. To avoid capacitive coupling, the transformer should be provided with at least two or three shielding systems. The third or final shield should be connected to the power supply output common. The inner shield should be connected to the signal ground.

Common-Mode Rejection Measurement

A typical configuration of a guard-shielded measuring system is shown in Figure 1.8u. The common-mode rejection of the system is measured by taking the ratio of the common-mode voltage (E_{cm} between points C and D) divided by the portion of E_{cm} that appears at the amplifier inputs A and B. Almost all of the remainder of the voltage E_{cm} will appear across C_g and R_g, in which case the common-mode rejection (CMR) is approximately:

$$CMR_{\mathrm{AC}} = 1/2FC_gR_c \qquad \mathbf{1.8(4)}$$

$$CMR_{\mathrm{DC}} = R_g/R_c \qquad \mathbf{1.8(5)}$$

It can be seen from these relations that the common-mode rejection is dependent on the value of R_c, which is the unbalance of the transducer and signal lines. In an ideal case, when R_c is zero, the common-mode rejection will be infinite.

Measuring by direct methods that portion of E_{cm} that appears across R_c is not possible since it is unlikely that any instrument could be connected to the source to measure the voltage across R_c without also changing the current through it. Thus, when testing the guarding of a system, the component

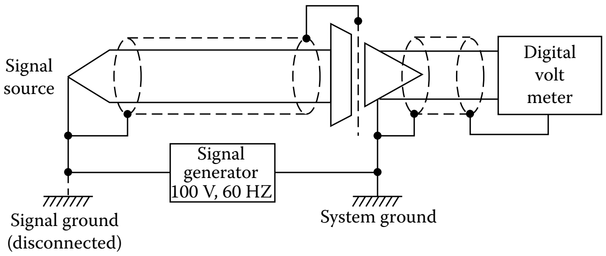


FIG. 1.8v
A method for measuring common-mode rejection.

due to E_{cm} at the amplifier output is measured. This value is divided by the amplifier gain, and it is assumed that the quotient is a good measure of the voltage across R_c.

A practical setup for measuring rejection in a common-mode system is shown in Figure 1.8v. With the signal source disconnected from the signal ground, a 100-V, 60-Hz signal from a signal generator is applied between the system ground and signal source. The change in the digital voltmeter (DVM) reading will show the effect of this common-mode signal on the system. To obtain the effective voltage across the amplifier inputs, this change is divided by the amplifier gain. The common-mode rejection is then obtained by dividing the signal generator voltage with the assumed voltage at the amplifier input. For example,

Amplifier gain (G) = 1000
Signal generator voltage (E'_{cm}) = 100 V
DVM reading before applying E_{cm} (e'_{cm}) = 2.0 V
DVM reading after applying E_{cm} (e''_{cm}) = 2.1 V

The increase in voltage at the amplifier output (e_{cm}) due to E_{cm} is:

$$e_{cm} = e''_{cm} - e'_{cm} = 2.1 - 2.0 = 0.1 \text{ volts} \tag{1.8(6)}$$

The voltage at the amplifier input (e'_{cm}) due to E_{cm} is:

$$e'_{cm} = \frac{e_{cm}}{G} = \frac{0.1}{1000} = 0.0001 \text{ volts} \tag{1.8(7)}$$

Therefore, the common-mode rejection of the system is:

$$CMR = \frac{E_{cm}}{e'_{cm}} = \frac{100}{0.0001} = \frac{1{,}000{,}000}{1} = 10^6:1 \tag{1.8(8)}$$

When dealing with AC common-mode signals applied to DC measuring instruments, it is important to consider the effect of inherent noise rejection in the measuring instrument (or amplifier). Many DC instruments have an input filter that allows the undisturbed measurement of DC signals in the presence of AC noise. Such instruments are said to have normal-mode interference rejection.

Thus, if common-mode rejection is measured by the indirect method just described, the apparent common-mode rejection will be the product of the actual common-mode rejection and the filter rejection. In the earlier example, if a filter with 10:1 normal-mode rejection were inserted in front of the amplifier input, the apparent common-mode rejection would increase to 10^7:1 (Ref. 2).

Thermocouple Signals

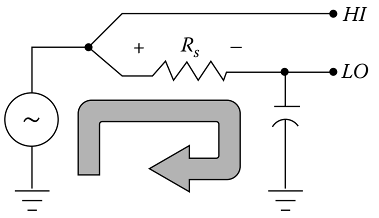
The thermocouple signal is very weak — a 1-degree change in temperature results in only a few millionths of a volt change in output. Because of this, precautions must be taken against errors due to stray currents resulting from the proximity of electrical wiring (common-mode noise) or from capacitive secondary grounds (normal-mode interference).

Common-Mode Noise Common-mode noise (Figure 1.8w) appears on both thermocouple signal wires and therefore can be filtered out as 60-Hz (or 50-Hz) harmonic noise.

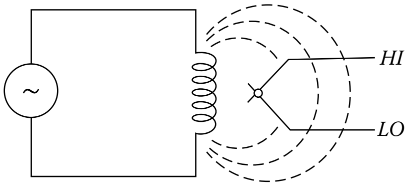
The filter does reduce the interference dramatically, but it also causes the voltmeter to be sluggish when responding to a step change. It is also possible to eliminate the common-mode interference by using twisted wire leads, because each time the wire is twisted, the flux-induced current is inverted.

Another recommended form of protection against any type of common-mode noise is guarding and shielding. If the shield surrounding the lead wires is connected to the guard surrounding the voltmeter, the interfering current caused by AC interference does not flow through the thermocouple lead resistance but instead is shunted.

Naturally, when thermocouples are scanned, the scanner guard must be switched to the shield of the thermocouple being read to eliminate ground loops. Harmonics can also be removed by integrating the incoming signal over the power line cycle in an integrating analog-to-digital converter or



Normal mode



Common mode

FIG. 1.8w
Normal-mode noise that enters only one of the lead wires is more difficult to remove than common-mode noise, which acts on both leads.

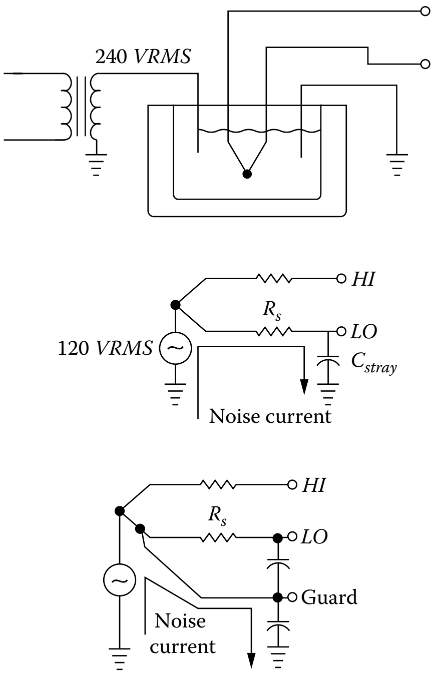


FIG. 1.8x
The normal-mode noise is reduced by the addition of a guard wire.[3]

voltmeter. In short, common-mode noise is relatively easy to remove.

Normal-Mode Noise The same cannot be said about normal-mode noise. Normal-mode noise interference can occur in the measurement of the temperature in a molten metal bath, which is heated by electric current. In this case, the thermocouple junction is in direct contact with a common-mode noise source. In addition, the capacitive ground (C-stray) from the LO terminal of the thermocouple to the chassis causes a current flow in the low lead and an associated normal-mode noise voltage across the resistance R_s (Figure 1.8x).

If a guard lead wire is connected directly to the thermocouple, the current flowing in the LO lead through the resistance R_s is drastically reduced. Therefore, the worst form of interference is DC offset caused by a DC leakage current; whatever normal-mode noise remains in the system cannot be distinguished from the measurement and, in case of weak signals, even a small amount of noise can represent a large amount of interference.

Complex TC Conditioning One of the important practices in eliminating noise in low-level signal systems is the arrangement of external circuits so that noise pickup will appear equally on both sides of the signal pair, remain equal, and appear simultaneously at both input terminals of the differential amplifier (or measuring device). In this case it can be rejected by common-mode protection without affecting the frequency response or the accuracy of the system. If a noise signal is allowed to combine with the useful signal, it can only be separated from the useful signal by selective filtering.

Thermocouple signals are susceptible to noise problems because they are often in contact with the process or structure being measured. Additional thermoelectric potentials developed in the bimetallic leads should be accounted for when measuring a signal from a thermocouple. The best thermocouple materials are not the best electrical conductors, yet lead wires must be made of thermocouple materials all the way back to the reference junction.

To eliminate the need to account for each bimetallic wire junction, one can compare the output of the thermocouple with the output of an identical thermocouple in a controlled-temperature environment. The balance of the circuit in that case can be made of copper wires creating an identical couple in each lead and producing a zero net effect when both couples are at the same temperature.

The use of a constant-temperature reference junction is the most accurate method of compensation, but it is not the most convenient. A simulated constant-temperature reference can be obtained with devices containing compensating junctions, millivolt sources, and temperature-sensitive resistors. These devices can also permit the use of copper lead wires instead of thermocouple leads.

A more complex conditioning circuitry is shown in Figure 1.8y, which performs the functions of thermocouple signal conditioning, noise discrimination, balancing, ranging, and standardizing. This circuit must be tied to a calibration scheme for the system output. The balance circuit shown accomplishes both variable offset and calibration.

The signal-conditioning circuitry makes it possible to convert the transducer output to make use of the optimum range of the measuring system for the best linearity, accuracy, and resolution.

Because thermocouples are often in electrical contact with the device or structure of which the temperature is being measured and with an intentional ground at the sensor, differential input amplifiers with very good common-mode rejection are required to effectively discriminate against noise in the structure or grounding system.[1]

Low-Level Signal Multiplexing

Conventional noise rejection techniques, such as twisting leads, shielding, and single-point grounding, have been discussed. It was pointed out that input guarding involves isolating the transducer input signal from the common-mode voltage between signal leads and ground. This change from differential to single-ended signal can occur at any place in the system after the multiplexer. Early versions of multiplexing systems employed crossbar switches, relays, and similar electromechanical devices for low-level input signal multiplexing.

Electromechanical Multiplexing Mercury-wetted relays, which are actuated by rotating magnets, can be used as electromechanical multiplexers. They have the advantage of high

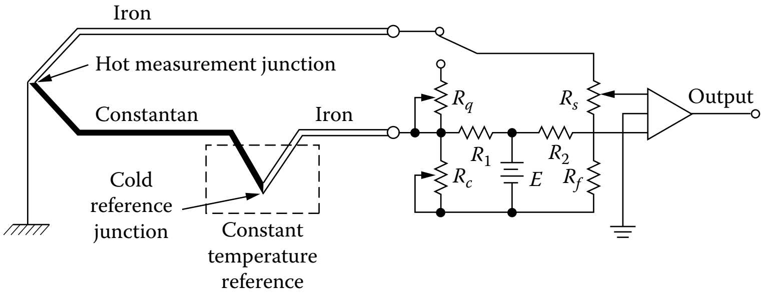


FIG. 1.8y
A complex thermocouple conditioner performs the functions of TC signal conditioning, noise discrimination, balancing, ranging, and standardizing. E = a precision isolated power supply; R_1 and R_2 = precision bridge resistors, low values (5 or 10 ohms); R_c = balance control variable resistor, large resistance value; R_f = reference resistor, large resistance value; R_q = equivalent line resistor for use during stabilizing and calibrating; R_s = span adjustment potentiometer.

cross-talk isolation, because of their high (10^{14}) ratio of open-circuit to closed-circuit resistance. They are also accurate because of their very low contact resistance.

Their disadvantages include their low speed (a few hundred points per second) and their vibration sensitivity, if it exceeds 0.75 g. DCS multiplexers, operating on analog process signals, in the past often used such electromechanical multiplexers.

Passive Filter Multiplexing A simple passive filter (RC) circuit, as shown in Figure 1.8z, can be designed to reject common-mode noise from about 40 decibels, at the selected frequency. More sophisticated passive networks (such as parallel T or notch filters) improve noise rejection, but it is hard to obtain 60-decibel noise rejection with passive circuits.

Because of deficiencies in the noise rejection capabilities of the earlier approaches, the limitations in scan rates, and the ever-increasing use of data acquisition systems, many devices have been developed with extended capability to cover the spectrum of present-day requirements. Each general category contains subsets of devices using the same basic switching element but offering application-dependent variations. The most important variations are the programmable range (i.e., gas chromatograph signal) switching and input grounding.

Three multiplexing techniques are commonly used: (1) flying capacitor multiplexer, (2) three-wire multiplexer, and (3) solid-state multiplexer.

Flying Capacitor Multiplexing The capacitive-transfer (flying capacitor) switching arrangement, shown in Figure 1.8aa, is simple, economical, and capable of great noise rejection (including random noise spikes). This system is limited to applications where signal bandwidth requirements are narrow (on the order of 0.5 to 1 Hz), due to a necessarily large transfer capacitance to minimize the effect on system resolution of charge losses and of delays during charging, settling, and digitizing.

In the capacitive transfer circuit, between scans, a set of normally closed contacts connects the low-leakage transfer capacitor across the input signal. Common practice is to short the amplifier input during this between-scan period to avoid stray pickup (due to a high-impedance open circuit).

When the multiplexer selects the input for scan, the amplifier input short is removed, and the contacts switch to connect the transfer capacitor to the amplifier input. This transfer circuit introduces input attenuation and phase lag, which must be considered in circuit design. An additional

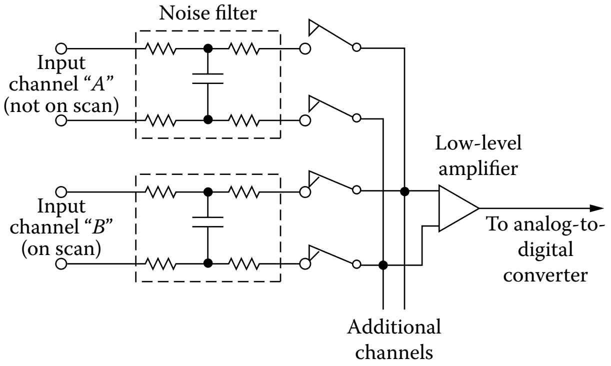


FIG. 1.8z
Circuit showing the passive filter multiplexer.

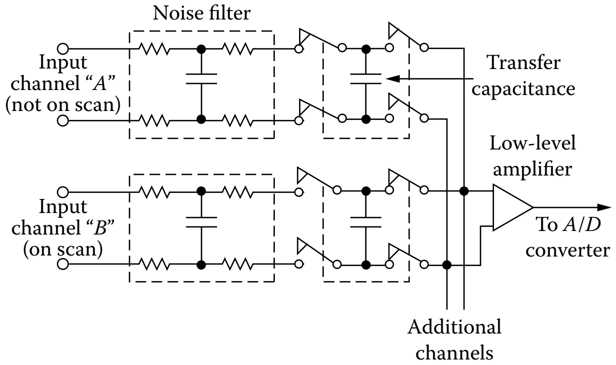


FIG. 1.8aa
Circuit describing the flying capacitor multiplexer.

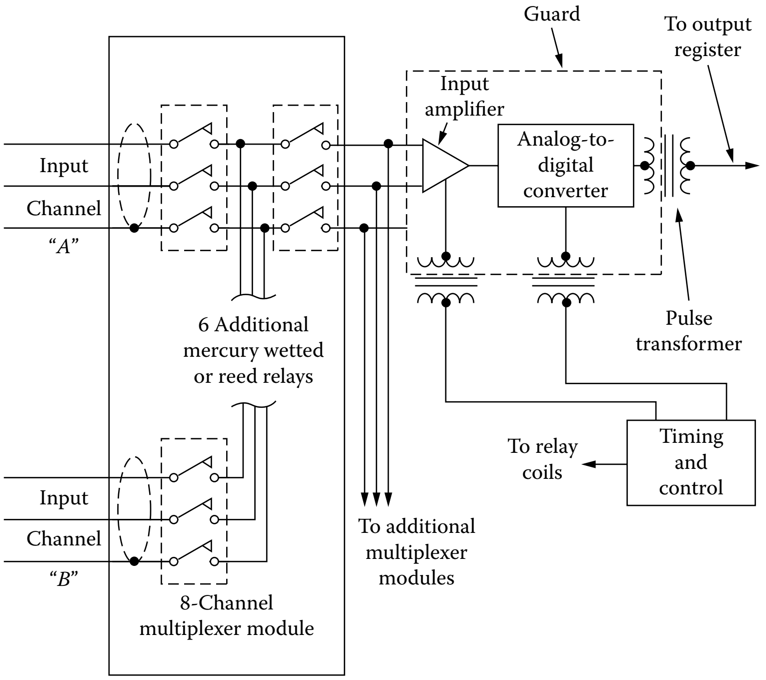


FIG. 1.8bb
Circuit describing the three-wire multiplexer, which provides high common-mode voltage rejection.

RC filter is usually necessary to achieve acceptable common-mode rejection of 100 to 120 decibels.

Three-Wire Multiplexing The three-wire multiplexing system requires the transducer lead wires to be shielded with the shield terminated at the signal ground. This guard shield must be carried through the multiplexer, at each point, up where the differential signal is transferred to a ground-referenced signal.

In Figure 1.8bb the input amplifier and analog-to-digital converter are enclosed within the guard. The serial digital data are transmitted through a shielded pulse transformer to an output register for presentation in parallel form. Relay coils are matrixed and controlled by address logic.

This system is used when input signal bandwidths of several hertz are required, since filtering is not essential to obtain good common-mode rejection. Common-mode rejection from DC to 60 Hz is about 100 decibels at reduced input bandwidth.

Solid-State Multiplexing Compared to electromechanical multiplexers, solid-state units do not have speed- or vibration-related limitations, but their early models did suffer from cross-talk due to their low open-circuit resistance. The switches in most newer multiplexers are MOSFET (metallic oxide semiconductor field-effect transistor) and are satisfactory for multiplexing digital signals. The costs of solid-state multiplexers using large-scale IC (integrated circuitry) circuits is lower than that of the electromechanical ones. Similarly, their power consumption and size are also smaller. Another advantage is that because their switching is electronic, the address of the next channel can be programmed in any order desired.

A solid-state multiplexing system is shown in Figure 1.8cc. In a typical high-performance solid-state multiplexer, each input has a matched pair of transistor switches terminating at the primary of shielded input transformers, which are driven through an isolation transformer. One cycle of a square wave of peak amplitude equal to the input signal level is transferred across the transformer by alternately pulsing the switches.

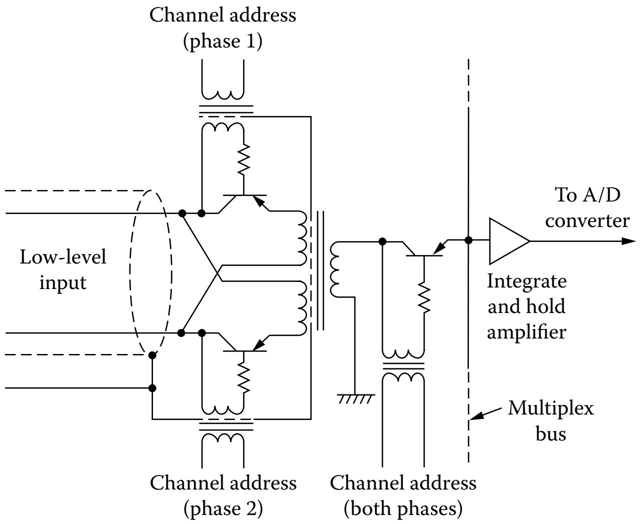


FIG. 1.8cc
Solid-state multiplexing system for high common-mode rejection provided with integrate and hold amplifier.

TABLE 1.8dd
Feature Summary of Commonly Used Multiplexers

Features	*Capacitive-Transfer*	*Three-Wire*	*Solid-State*
Cost	Low	Low	High
Input signal ranges	±50 to ±500 mV	±5 mV to ±5 V	From 5 mV full scale to 100 mV full scale
Scan rates (points per second)	Up to 200	Up to 200	Up to 20,000[a]
Operable common-mode environment (volts)	Up to 200 to 300	Up to 200 to 300	Up to 500
Common-mode rejection from DC to 60 Hz (decibels)	100 to 120[b]	100 to 120	120
Accuracy	±0.1% full scale at 1 to 10 samples per second scan rate; ±0.25% full scale at 10 to 50 samples per second	±0.1% full scale for all scan rates	±0.1% full scale[c]

[a] 10-mV resolution in high common-mode environments.
[b] With a two-section filter.
[c] Overall accuracy.

This signal is synchronously rectified to preserve original input polarity integrated over the cycle period, and it is amplified and held for digitizing by an integrator and hold amplifier. The cost of the system is relatively high, making the application economically impractical unless high common-mode tolerance is required. Common-mode rejection from DC to 60 Hz of 12 decibels is easily obtained.

Selection of any multiplexing system should be based on the performance, reliability, and cost of a particular application. Table 1.8dd summarizes the features of the discussed systems.

Digital Multiplexing When there are many channels, the multiplexer must be fast even if the incoming signals are slow. Figure 1.8ee illustrates such a fast multiplexer configuration, where one analog-to-digital converter (ADC) is used for each channel. The outputs of the converters are connected to a common bus through interface units.

In such a configuration, the digital processor, which is used to process the digital outputs of the converters, uses the same bus. This way the digital processor can alternately access the outputs of each ADC and thereby can multiplex their output signals.

Data Acquisition Boards The data acquisition board allows a computer to access a set of analog signals that are connected to that plug-in board. Most of these boards operate on signals that have already been filtered and amplified, although some boards can directly accept the output signals of sensors.

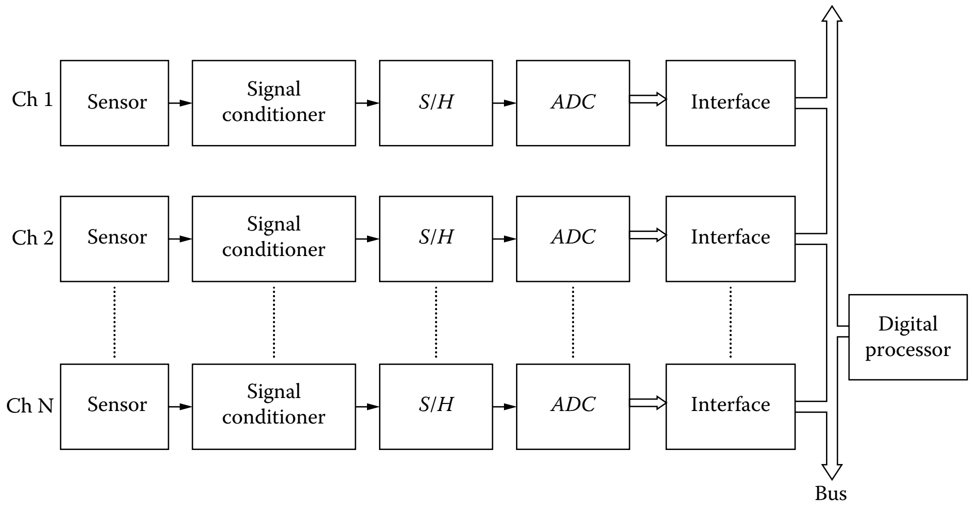


FIG. 1.8ee
Multi-channel data acquisition using digital multiplexing.

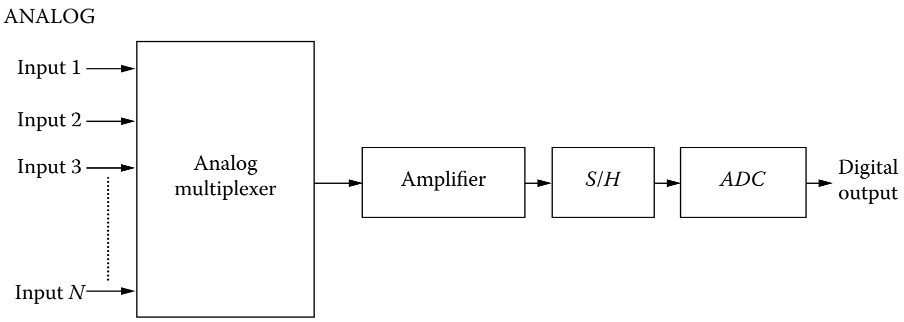


FIG. 1.8ff
Data acquisition boards commonly allow access digital computers to reach 8 or 16 analog inputs.

Such boards are provided with the required signal conditioning circuits.

Usually 8 or 16 analog inputs can be connected to a data acquisition board. Figure 1.8ff illustrates such a system, which consists of an analog multiplexer, an amplifier, a sample and hold circuit, and an analog-to-digital converter. Some data acquisition boards also provide (usually two) analog output channels for control purposes.

A/D Converters

Just as microprocessors have evolved in sophistication, so have A/D converters. Eight-bit resolution devices common in the 1960s provided a resolution of about +/– 0.4%. In the year 2000 the first 21-bit resolution A/D was used in a temperature transmitter, providing a resolution of +/– 0.00005%. D/A converters have also evolved with resolutions increasing from 8-bit up to the 18-bit versions used in the better transmitters beginning in the year 2000.

The result of combining these technologies is a universal transmitter that accepts inputs from any TC, RTD, mV, resistance, or potentiometer signal, checks its own calibration on every measurement cycle, has minimal drift over a wide ambient temperature range, incorporates self-diagnostics, and is configured using pushbuttons or simple PC software. The reconfiguration process is quick and convenient, and it tends to allow for lower inventories by making the transmitters interchangeable.

Noise Rejection in A/D Converters The dominant noise in A/D converters is line frequency noise. One approach toward reducing this noise is to integrate the input signal. The integrating technique relies on A/D converter hardware. The operation of the A/D converter is such that it converts the continuously monitored measurement into a pulse train and totals the number of pulses it receives. If the measurement time is equal to the period of the line frequency, integration yields the true value of signal level, and the line frequency noise effect becomes zero, as shown in Figure 1.8gg.

This method is usually applied to slow multiplexers (e.g., 40 points per second scan rate) and is suitable for applications in process monitoring and control where high sampling frequencies are not required. In appropriate situations, it provides noise rejection of about 1000 to 1 (or 60 decibels) at 60 Hz and offers good rejection of other frequency noises.

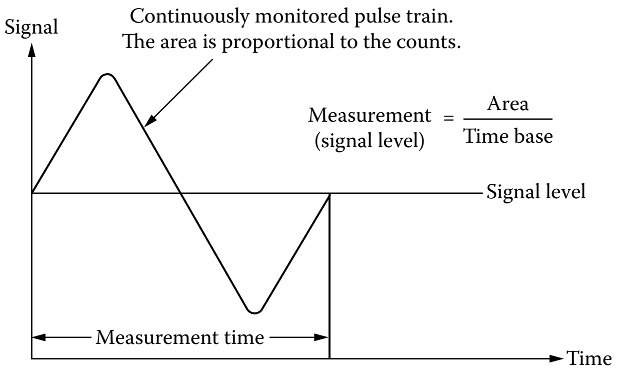


FIG. 1.8gg
Line frequency noise reduction by integrating the input signal of an A/D converter.

CONCLUSIONS

Electrical noise problems are not unique to new installations. Even systems that were in satisfactory operation can develop noise as a result of burned or worn electrical contacts, defective suppressors, loose connections, and the like. Any equipment used to identify the noise source must have the frequency response and rise time capability to display noise signals. Minimum requirements would be 10 MHz bandwidth and 35 nanosecond rise time.

Once the noise signal or a shielding problem is identified, its elimination involves applying the basic rules discussed earlier. The most important ones are listed below:

1. Select signal sources with low output impedance and with grounding capability. Never use the shield as a signal conductor and never splice a low-level circuit.

2. Use only top-quality signal cable, in which the signal pair is twisted and protected with a lapped foil shield, plus a low-resistance drain wire. Low-level signal cables must be terminated with short, untwisted wire, which expose a minimum area to inductive pickup.
3. Provide a good low-resistance system ground and a stable signal ground located near the signal source. Use individual twisted shielded pairs for each transducer. Unused shielded connector in a low-level signal cable should be single-end grounded with the shield grounded at the opposite end.
4. Ground the signal cable shield and signal cable circuit at the source only, and keep them from contacting ground at any other point. Signal cable shield must be maintained at a fixed potential with respect to the circuit being protected.
5. Thermocouple transducers may be used with common shield when the physical layout allows multiple pair extension leads.
6. Connect all signal pairs to adjacent pins in a connector and carry the cable shields through it on adjacent pins. Provide shield around signal pairs in the connector by shorting pins together at both ends and by connecting to signal cable shields.
7. Separate low-level circuits from power cables by 3 ft if possible (1 ft minimum), and cross noisy circuits at right angles and at a maximum practical distance.
8. Select appropriate multiplexing and amplifier circuits having input guarding systems of at least 10^6:1 common-mode rejection at 60 Hz.
9. Provide triple-shielded power supplies for measuring devices.
10. Measure the common-mode rejection of the system by an indirect method whenever the rejection capability is in doubt.
11. Pay particular attention to the quality of the workmanship and system wiring practices.
12. Perform the analog-to-digital conversion as close to the signal source as practical.

References

1. Lovuola, V. J., "Preventing Noise in Grounded Thermocouple Measurements," *Instruments and Control Systems,* January 1980.
2. Deavenport, J. E., "EMI Susceptibility Testing of Computer Systems," *Computer Design,* March 1980.
3. Hewlett-Packard Application Note 290, "Practical Temperature Measurement," Palo Alto, CA, Latest revision.

Bibliography

Adler, C. B., "Reliability Aspects of Temperature Measurement," ISA Conference, Chicago, 2001.

Demler, M. J., *High Speed Analog-to-Digital Conversion,* New York: Academic Press, 1991.

Denny, H., "Grounding for the Control of EMI," Virginia: DWCI, 1983.

Hoeschelle, D. F., *Analog-to-Digital and Digital-to-Analog Conversion Techniques*, New York: John Wiley & Sons, 1994.

Johnson, R., "Measuring the Hot, Difficult and Inaccessible," *Control Engineering,* June 2001.

Kaufman, A. B., "Third Party Junctions in TCs," *Instruments and Control Systems,* December 1981.

Klipec, B., "Reducing Electrical Noise in Instrument Circuits," *IEEE Transactions on Industry and General Applications,* 1967.

Lovoula, V. J., "Preventing Noise in Grounded Thermocouple Measurements," *Instruments and Control Systems,* January 1980.

Michalski, L., et al., *Temperature Measurement,* 2nd ed., London: John Wiley & Sons, 2001.

Morrison, R., and Lewis, W. H., *Grounding and Shielding in Facilities,* New York: John Wiley & Sons, 1990.

Nalle, D. H., "Signal Conditioning," *Instruments and Control Systems,* February 1969.

National Electrical Code, National Fire Protection Association, Art. 500—503, Quincy, MA, 2002.

Schmid, H., *Electronic Analog/Digital Conversions,* New York: Van Nostrand Reinhold, 1970.

Simmons, J. P., *Soares Book on Grounding,* 7th ed., Dallas, TX: IAEI, 1999.

Stoffel, J., "DCS Melds Analog, Digital I/O," *Control Engineering,* October 1991.

Van de Plassche, R., *Integrated Analog-to-Digital and Digital-to-Analog Converters,* Dordrecht, The Netherlands: Kluwer Academic Publishers, 1994.

Volbrecht, A., and Gordon, W., "Temperature Measurement: Making Sense of it All," *Sensors,* June 1998.

Control Theory

2

2.1
CONTROL BASICS 96

2.2
CONTROL MODES—PID CONTROLLERS 114

2.3
CONTROL MODES—PID VARIATIONS 124

2.4
CONTROL MODES—DIGITAL PID CONTROLLERS 130

2.12
INTERACTION AND DECOUPLING 205

2.13
MODEL-BASED CONTROL 209

2.14
MODEL-BASED PREDICTIVE CONTROL PATENTS 214

2.15
MODEL-FREE ADAPTIVE (MFA) CONTROL 224

2.16
MODELING AND SIMULATION OF PROCESSES 234

2.17
MODEL PREDICTIVE CONTROL AND OPTIMIZATION 242

2.18
NEURAL NETWORKS FOR PROCESS MODELING 253

2.19
NONLINEAR AND ADAPTIVE CONTROL 265

2.20
OPTIMIZING CONTROL 274

2.21
PID ALGORITHMS AND PROGRAMMING 284

2.22
PROCESS GAINS, TIME LAGS, REACTION CURVES 296

2.23
RATIO CONTROL 305

2.24
REAL-TIME PERFORMANCE ASSESSMENT 311

2.25
RELATIVE GAIN CALCULATIONS 318

2.26
ROBUSTNESS: A GUIDE FOR SENSITIVITY AND STABILITY 323

2.27
SAMPLED DATA CONTROL SYSTEMS 326

2.28
SELECTIVE, OVERRIDE, AND LIMIT CONTROLS 336

2.29
SELF-TUNING CONTROLLERS 345

2.35
TUNING PID CONTROLLERS 414

2.36
TUNING LEVEL CONTROL LOOPS 432

2.37
TUNING INTERACTING LOOPS, SYNCHRONIZING LOOPS 442

2.38
TUNING BY COMPUTER 446

2.1 Control Basics

C. F. MOORE (1970, 1985) **B. G. LIPTÁK** (1995)
K. A. HOO, B. G. LIPTÁK, M. J. PIOVOSO (2005)

INTRODUCTION

This section begins with some control-related definitions of some basic terms and concepts. This is followed by some introductory discussion of process dynamics and closed loop feedback control. For an in-depth discussion of analog and digital PID (proportional-integral-derivative) variations, closed loop responses, and the various model-based expert and optimizing control systems, the reader should turn to the later sections of this chapter.

During the first industrial revolution, the work done by human muscles was gradually replaced by the power of machines. Process control opened the door to the second industrial revolution, where the routine functions of the human mind and the need for the continuous presence of human observers were also taken care of by machines. Today, during the third industrial revolution, we are designing computers that can optimize and protect the safety of our processes not only by self-tuning, but also by applying neural networks capable of self-learning the characteristics of the controlled process, similar to the way a baby learns about the outside world.

This way, the human operator is relieved of the bulk of tedious and repetitive physical and mental tasks and is able to concentrate on the more creative aspects of the operating industry. In the third industrial revolution the traditional goal of maximizing the quantity of production will be gradually replaced by the goal of maximizing the quality and durability of the produced goods, while minimizing the consumption of energy and raw materials and maximizing recycling and reuse. Optimized process control will be the motor of this third industrial revolution.

HISTORY OF PROCESS CONTROL

The fly-ball governor was the first known automatic control system. It was installed on Watts' steam engine over 200 years ago in 1775. As shown in Figure 2.1a, the fly-balls detected the speed of shaft rotation and automatically opened up the steam supply when a drop was registered in that speed. The users of the fly-ball used it without understanding why it works. Another century went by before James Clark Maxwell in 1868 prepared the first mathematical analysis of the fly-ball governor.

The spreading use of steam boilers contributed to the introduction of other automatic control systems, including various steam pressure regulators and the first multiple-element boiler feed water systems. Here again the application of process control was ahead of its theory, as this first feedforward control system was introduced at a time when even the term "feedforward" had not yet been invented, let alone understood. Actually, the first general theory of automatic control, written by Nyquist, did not appear until 1932.

In 1958 Donald P. Eckman's classic *Automatic Process Control* was published, and even after that, for several decades, most universities, and academia as a whole, treated the teaching of process control as if it were just another course in mathematics. Some still do and teach process control as if we lived in the frequency and not in the time domain.

It was not until the late 1960s that the first edition of this handbook and F. G. Shinskey's classic *Process Control Systems* started to give recognition to process control analysis in the time domain. This is no surprise as progress was always made by the practical users and not by the mathematically oriented theoreticians. One should also understand that before one can control a process one must fully understand the process, and for this reason this section will start with the description of the "personalities" of different processes.

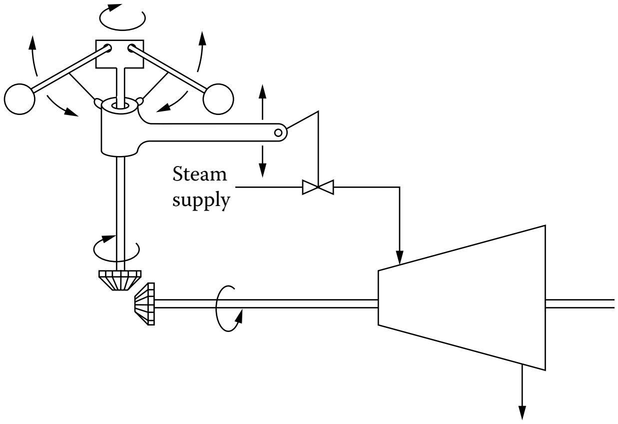


FIG. 2.1a
The fly-ball governor keeps shaft rotation constant by opening the seam supply valve when rotation slows and closing it when the rotation rises.

PROCESS DYNAMICS[1]

It is self-evident that the pilots of an ocean liner and a supersonic airplane need to have different personalities because the natures and the characteristics of their vehicles are different. This is also true in industrial process control. The pilot (the controller) must be correctly matched to the process that it controls. In order to do that, the designer of a process must understand the "personality" of the process.

Most processes contain resistance, capacitance, and deadtime elements, which determine their dynamic and steady-state responses to upsets. Before the control of processes is discussed, these three elements of the "personalities" of processes will be described.

To describe the personalities of processes, block diagrams are used (Figure 2.1b). The two main symbols used in any block diagram are a circle and a rectangular box. The circle represents algebraic functions such as addition or subtraction and is always entered by two lines but exited by only one.

The rectangular box always represents a dynamic function, such as a controller, where the output is the controlled variable (c) and is a function of both time and of the input or manipulated variable (m). Only one line may enter and only one line may leave a rectangular block. The "system function," the "personality" of the process component, is placed inside the block, and the output is determined by product of the system function and the input.

Resistance-Type Processes

Pressure drop through pipes and equipment is the most obvious illustration of a resistance-type process. Figure 2.1c illustrates the operation of a capillary flow system, where flow is linearly proportional to pressure drop. This process is described by a steady-state gain, which equals the resistance (R). Therefore as the input (m = flow) changes from zero to m, the output (c = head) will go in an instantaneous step from zero to $c = Rm$.

Laminar resistance to flow is analogous to electrical resistance to the flow of current. The unit of resistance is sec/m^2 in the metric and sec/ft^2 in the English system and can be calculated using Equation 2.1(1):

$$R = \frac{dh}{dq} = \frac{128\nu L}{g\pi D^4} \quad \text{sec/ft}^2 \qquad \textbf{2.1(1)}$$

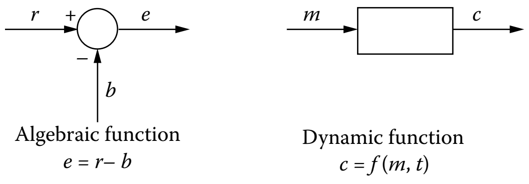
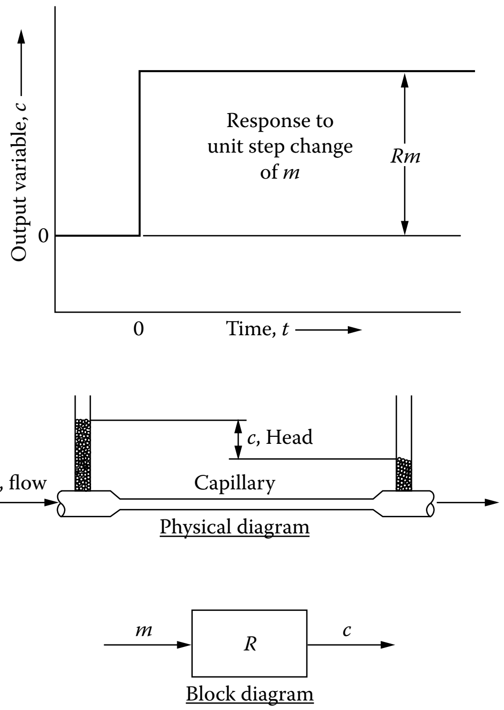


FIG. 2.1b
Block diagram descriptions of algebraic and dynamic functions.



$c = Rm$

Where c = Output variable (head)
R = Resistance
m = Input variable (flow)

FIG. 2.1c
Physical example of a resistance element and its block diagram.

where h = head, ft; ν = kinematic viscosity, ft^2/sec; L = length of tube or pipe, ft; D = inside diameter of pipe, ft; q = liquid flow rate, ft^3/sec; μ = absolute viscosity, lb-sec/ft^2 = $\gamma\nu/g$; γ = fluid density, lb/ft^3.

When the flow is turbulent, the resistance is a function of pressure drop, not of the square root of pressure drop, as was the case with laminar (capillary) flow. A liquid flow process usually consists of a flow-measuring device and a control valve in series, with the flow (c) passing through both. Figure 2.1d illustrates such a process. As can be seen from the block diagram, the flow process is an algebraic and proportional (resistance) only process. The manipulated variable (m) is the opening of the control valve, while the controlled variable (c) is the flow through the system. A change in m results in an *immediate and proportional* change in c. The amount of change is a function of the process gain, also called process sensitivity (K_a).

The load variables of this process are the upstream and downstream pressures (u_0 and u_2), which are independent, uncontrolled variables. A change in either load variable will also result in an *immediate and proportional* change in the controlled variable (c = flow). The amount of change is a function of their process sensitivity or gain (K_b).

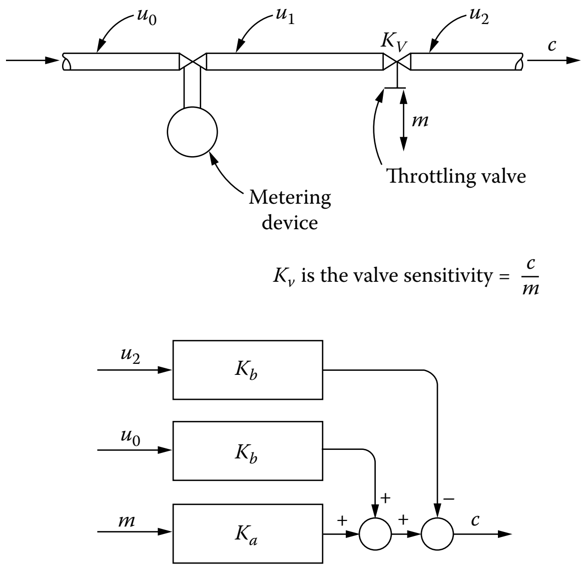


FIG. 2.1d
The relationship between the manipulated variable (m) and the controlled variable (c) in a resistance (flow) process, shown both in its physical and in its block diagram representations.

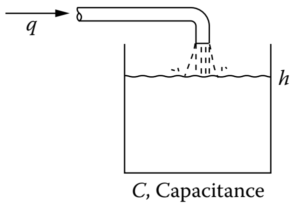
Capacitance-Type Processes

Most processes include some form of capacitance or storage capability. These capacitance elements can provide storage for materials (gas, liquid, or solids) or storage for energy (thermal, chemical, etc.). Thermal capacitance is directly analogous to electric capacitance and can be calculated by multiplying the mass of the object (W) with the specific heat of the material it is made of (Cp). The units of thermal capacitance are BTU/°F in the English or Cal/°C in the metric system.

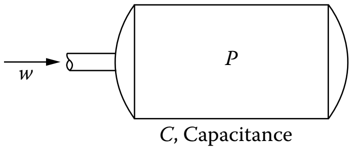
The capacitance of a liquid or a gas storage tank can both be expressed in area units (ft^2 or m^2). Figure 2.1e illustrates these processes and gives the corresponding equations for calculating their capacitances. The gas capacitance of a tank is constant and is analogous to electric capacitance. The liquid capacitance equals the cross-sectional area of the tank at the liquid surface, and if the cross-sectional area is constant, the capacitance of the process is also constant at any head.

A tank having only an inflow connection (Figure 2.1f) is a purely capacitive process element. In such a process the level (c) will rise at a rate that is inversely proportional to the capacitance (cross-sectional area of the tank) and after some time will flood the tank. The level (c) in an initially empty tank with a constant inflow can be determined by multiplying the inflow rate (m) with the time period of charging (t) and dividing that product with the capacitance of the tank ($c = mt/C$).

Figure 2.1f illustrates such a system and describes both the system equations and the block diagram for the capacitive element. In arriving at the system function the operational notation of the differential equation is used, using the substitution of $s = d/dt$. The system function therefore is $1/Cs$ and the output (c = head) is obtained by multiplying the system function with the input (m).



Liquid capacitance is defined by $C = \frac{dv}{dh} ft^2$



Gas capacitance is defined by

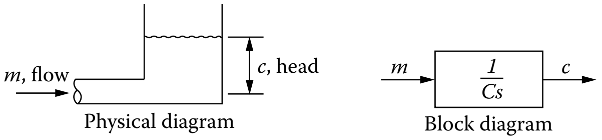
$$C = \frac{dv}{dp} = \frac{V}{nRT} ft^2$$

Where v = Weight of gas in vessel, LB.
V = Volume of vessel, ft^3
R = Gas constant for a specific gas, ft/deg
T = Temperature of gas, deg
p = Pressure, lb/ft^2
n = Polytropic exponent is between 1.0 & 1.2 for uninsulated tanks

FIG. 2.1e
Definitions of liquid and gas capacitance.

Resistance and Capacitance

Resistance and capacitance are perhaps the most important effects in industrial processes involving heat transfer, mass



$$C\frac{dc}{dt} = m = (Cs)c = m \therefore c = \left(\frac{1}{Cs}\right) m$$

Where C = Capacitance
c = Output variable (head)
t = Time
m = Input variable (flow)
$s = \frac{d}{dt}$ = differential operator

FIG. 2.1f
Example of a pure capacitance element.

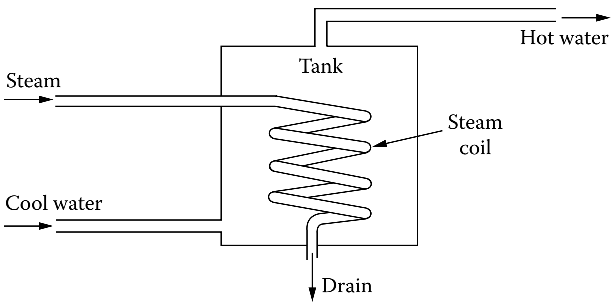


FIG. 2.1g
A water heater has both resistance and capacitance effects.

transfer, and fluid flow operations. Those parts of the process that have the ability to store energy or mass are termed "capacities," and those parts that resist transfer of energy or mass are termed "resistances." The combined effect of supplying a capacity through a resistance is a time retardation, which is very basic to most dynamic systems found in industrial processes.

Consider, for example, the water heater system shown in Figure 2.1g, where the capacitance and resistance terms can be readily identified. The capacitance is the ability of the tank and of the water in the tank to store heat energy. A second capacitance is the ability of the steam coil and its contents to store heat energy. A resistance can be identified with the transfer of energy from the steam coil to the water due to the insulating effect of a stagnant layer of water surrounding the coil.

If an instantaneous change is made in the steam flow rate, the temperature of the hot water will also change, but the change will not be instantaneous. It will be sluggish, requiring a finite period of time to reach a new equilibrium. The behavior of the system during this transition period will depend on the amount of material that must be heated in the coil and in the tank (determined by the capacitance) and on the rate at which heat can be transferred to the water (determined by the resistance).

Process Gain

The system function is the symbolic representation of how a process component changes its output in response to a change in its input. For example, if the system function is only a gain, only a constant (K_c, G, or some other symbol) would appear inside the block. The process gain is the ratio between the change in the output (*dc*) and the change in the input that caused it (*dm*). If the input (*m*) to the block is a sinusoidal, the output (*c*) will also be a sinusoidal.

The process gain can be the product of a steady-state gain (*Kp*) and a dynamic gain (*gp*) component. If the gain varies with the period of the input (exciting) sinusoidal, it is called dynamic gain (*gp*), while if it is unaffected by this period it is called steady-state gain (*Kp*). Therefore, if the process gain can be separated into steady-state and dynamic components, the system function can be given inside the block as (*Kp*)(*gp*).

The dynamic gain (*gp*) appears as a vector having a scalar component *Gp* and a phase angle. It will be shown shortly that capacitance-type process elements do introduce phase shifts in such a way that the peak amplitude of an input sinusoidal does not cause a simultaneous peak amplitude in the output sinusoidal.

Process Time Constant

Combining a capacitance-type process element (tank) with a resistance-type process component (valve) results in a single time-constant process. If the tank was initially empty and then an inflow was started at a constant rate of *m*, the level in the tank would rise as shown in Figure 2.1h and

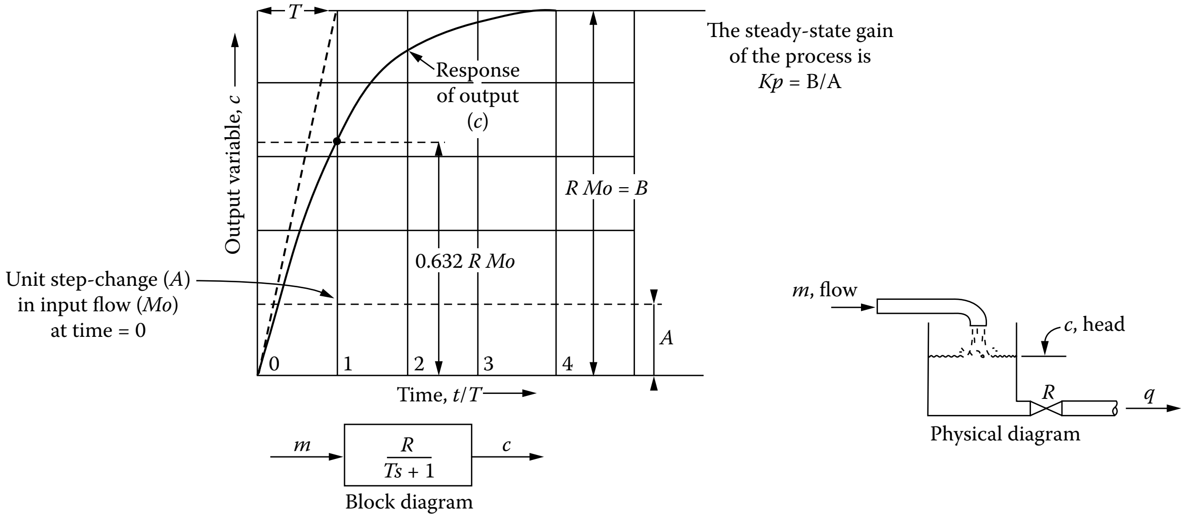


FIG. 2.1h
A single time-constant process consists of a capacitance and a resistance element. The time it takes for the controlled variable (c) to reach 63.2% of its new steady-state value is the value of the time constant.

would eventually rise to the steady-state height of $c = Rm$ in the tank.

In order to develop the overall system function, one has to combine the capacitance element of Figure 2.1f with a resistance element. If the input variable is the inflow (m) and the output variable is the level (c), the tank capacitance equals the difference between inflow (m) and outflow (q):

$$C(dc/dt) = m - q \qquad \textbf{2.1(2)}$$

In the fluid resistance portion of the system, the output (the head c) equals the product of the outflow (q) and the resistance (R). Because $c = qR$, therefore $q = c/R$. Substituting c/R for q in Equation 2.1(2) and multiplying both sides by R gives

$$RC(dc/dt) + c = Rm \qquad \textbf{2.1(3)}$$

The unit of R is time divided by area and the unit of C is area; therefore, the product RC has the unit of time. *This time (T) is called the time constant of the process.* It has been found experimentally that after one time constant the value of the output variable (c) of a single time-constant process will reach 63.2% of its final value.

Process elements of this description are common and are generally referred to as first-order lags. The response of a first-order system is characterized by two constants: a time constant T and a gain K. The gain is related to the amplification associated with the process and has no effect on the time characteristics of the response. The time characteristics are related entirely to the time constant. The time constant is a measure of the time necessary for the component or system to adjust to an input, and it may be characterized in terms of the capacitance and resistance (or conductance) of the process.

In their responses (Figure 2.1h), two characteristics distinguish first-order systems:

1. The maximum rate of change of the output occurs immediately following the step input. (Note also that if the initial rate were unchanged the system would reach the final value in a period of time equal to the time constant of the system.)
2. The actual response obtained, when the time lapse is equal to the time constant of the system, is 63.2% of the total response.

These two characteristics are common to all first-order processes.

Characteristic Equations of First-Order Systems A first-order system is one in which the system response to a constant change in the input is an exponential approach to a new constant value. Such a system can be described by a first-order differential equation:

$$\frac{dy(t)}{dt} + ay(t) = bu(t) \qquad \textbf{2.1(4)}$$

where $u(t)$ and $y(t)$ are the deviations of the inputs and outputs, respectively, and a and b are constants. The response of such a system with $a = 1$ and $b = 2$ to a unit step change (value 1 for $t \geq 0$ and 0 for $t < 0$) in the input is shown in the top panel of Figure 2.1i.

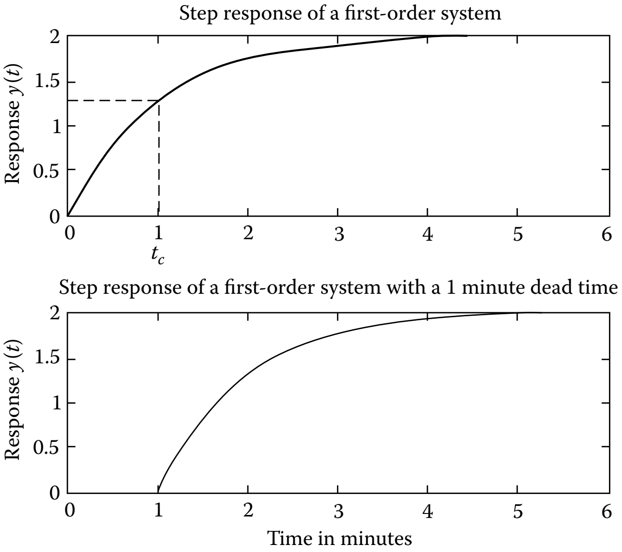


FIG. 2.1i
Step response of a first-order system with and without dead time.

This result can be demonstrated from the solution of the differential equation in Equation 2.1(4). The solution of this differential equation is important because it provides insight into the operation of the system and nature of the control problem.

The solution of any linear differential equation can be broken into the sum of two parts: the homogeneous and the particular solution. The homogeneous solution is that expression that when substituted for $y(t)$ in the left-hand side of Equation 2.1(4) produces the value zero. The particular or steady-state solution is that part of the total solution that is strictly due to the forcing input, $u(t)$.

To establish the homogeneous solution, assume the solution to be of the form

$$y_h(t) = Ke^{pt} \qquad \textbf{2.1(5)}$$

with $u(t) = 0$ in Equation 2.1(4). Substituting Equation 2.1(5) into the left-hand side of Equation 2.1(4) imposes the following condition on p:

$$p + a = 0 \qquad \textbf{2.1(6)}$$

which is satisfied with $p = -a$. It is clear from Equation 2.1(5) that for $a < 0$, the homogeneous solution becomes unbounded. The total solution, which is the sum of the homogeneous and particular solutions, becomes dominated by the homogeneous solution and not the particular solution.

Systems in which the homogeneous solutions grow without bound are said to be *unstable*. The homogeneous solution does not depend on the form of the forcing input and hence it is a characteristic of the system. In fact, a polynomial equation such as Equation 2.1(6) that defines $y_h(t)$ is called the *characteristic equation.*

The particular solution, $y_p(t)$, is that part of the total solution that is due entirely to the forcing function, $u(t)$. To establish $y_p(t)$, assume that it has the same functional form as $u(t)$, provided that the homogeneous solution does not have the forcing function as part of it. For example, if $u(t)$ is a ramp, assume that $y_p(t)$ is also a ramp with unknown intercept and slope. If $y_h(t)$ contains a term that is a ramp, then assume that the particular solution is quadratic in time with unknown coefficients. Substitute the assumed form into Equation 2.1(4) and solve for the unknowns.

Example

As an example, suppose that $u(t)$ in Equation 2.1(4) is a unit step. Assuming that the system is stable, the output will eventually reach a constant. Therefore, the particular solution is assumed to be a constant, α. Substituting α for $y(t)$ and 1 for $u(t)$ in Equation 2.1(4), yields that $\alpha = b/a$. Note that in Figure 2.1i, the particular solution is 2, obtained with $b = 2$ and $a = 1$.

The speed of response of a first-order system is determined by the *time constant, t_c*. The time constant is the time at which the exponent in the homogenous solution is −1. From Equation 2.1(5) and Equation 2.1(6), this corresponds to $1/a$. Figure 2.1i illustrates the time constant and output for the first-order system given in Equation 2.1(4).

If t_c is small, then the particular solution decays rapidly and the steady-state solution is quickly reached. Such a system is said to be "fast." On the other hand, if t_c is large, then the system reaches the steady state very slowly and is said to be "slow" or "sluggish."

First-Order System with Dead Time A first-order system coupled with dead time, or process delay, makes an excellent model for many process systems. The bottom panel of Figure 2.1i illustrates a first-order system with a 1-min dead time. Furthermore, this model is not difficult to determine. A simple step test in which the input to the system is changed from one value to another while the system is at some steady-state output gives sufficient information to determine the model.

Suppose the input is changed by an amount Δ. Let the output reach its new steady-state value, and divide the difference between the new and old steady-state outputs by Δ. This computed response is called the *step response.* The dead time is the time at which the output first starts to respond to the change in the input. If the exponent in the solution is evaluated at the time constant, the exponent of the particular part would be −1. This corresponds to the decay of 63.2% or $1-e^{-1}$ of the initial contribution of the particular solution. Thus, the time constant can be determined as the difference between the time at which the response is at 63.2% of the final value and the dead time.

In Figure 2.1i, 63.2% of the final response, denoted by the horizontal dashed line, has a value of 1.264 and yields a time constant of one. Experience shows that first-order plus dead time systems can be controlled by proportional plus integral (PI) controllers.

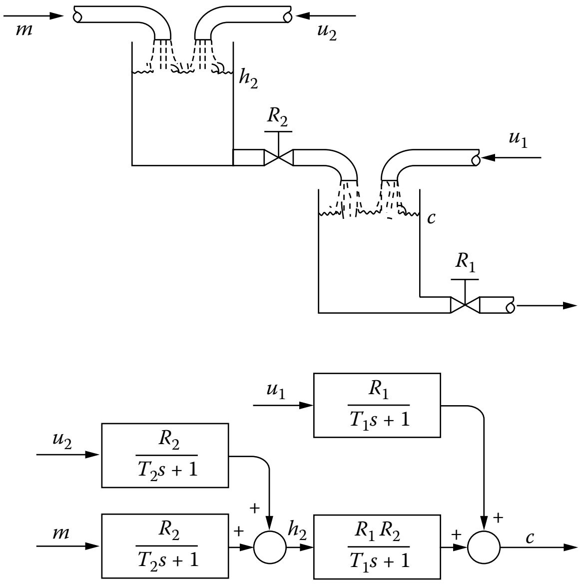


FIG. 2.1j
The physical equipment and the block diagram of a process with two time constants in series.

Multiple Time-Constant Processes Figure 2.1j illustrates a process where two tanks are connected in series and therefore the system has two time constants operating in series. The response curve of such a process ($r = 2$ in Figure 2.1k) is

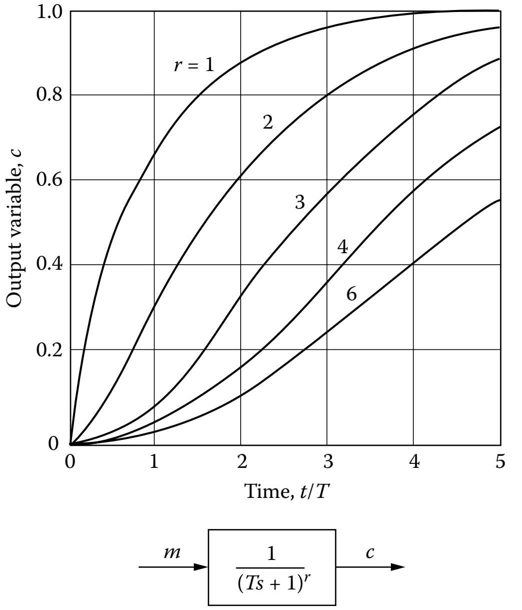


FIG. 2.1k
The responses of processes slow as the number of their (equal size) time constants in series increases.

slower than that of the single time-constant process ($r = 1$ in Figure 2.1k) because the initial response is retarded by the second time-constant. Figure 2.1k illustrates the responses of processes having up to six time constants in series. As the number of time constants increases, the response curves become slower (the process gain is reduced), and the overall response gradually changes into an S-shaped reaction curve.

Characteristic Equations of Second-Order Systems Second-order systems are characterized by second-order differential equations. These systems are more complex than first-order ones. A second-order system might be viewed as the solution of the following differential equation:

$$\frac{d^2y(t)}{dt^2}+2\xi\omega_n\frac{dy(t)}{dt}+\omega_n^2y(t)=\omega_n^2u(t) \qquad \textbf{2.1(7)}$$

Again, $y(t)$ is the process output (controlled variable) and $u(t)$ is the input, which is usually the reference input function. The parameters ξ and ω_n in this equation dictate the nature of the solution. The quantity ξ is called the *damping ratio* and ω_n is the *natural frequency.* If ξ is zero, then the solution is a sinusoid that oscillates at ω_n radians per second (Figure 2.1l). Thus, the system is an oscillator producing a sinusoidal response at the natural frequency of the oscillator.

The damping ratio is the ratio of the *damping constant* to the natural frequency. The damping constant determines the rate at which the oscillations decay away. As the damping constant increases, the speed of response increases and the oscillations decay more rapidly.

Figure 2.1m illustrates the effect of the damping ratio on the response of a second-order system. The top left panel of Figure 2.1m corresponds to the case of the damping ratio being zero. Note that the oscillations will continue forever and the steady-state solution is embedded in the response. This is an *undamped* response.

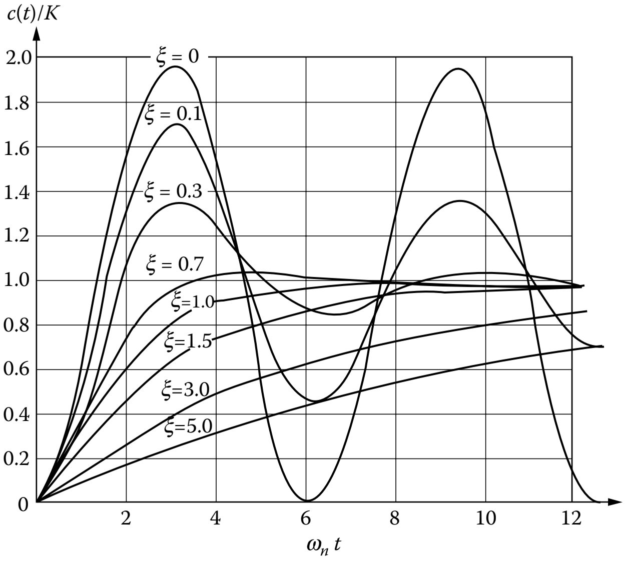


FIG. 2.1l
Illustration of the effect of damping ratio variations from 0 to 5 on a second-order system.

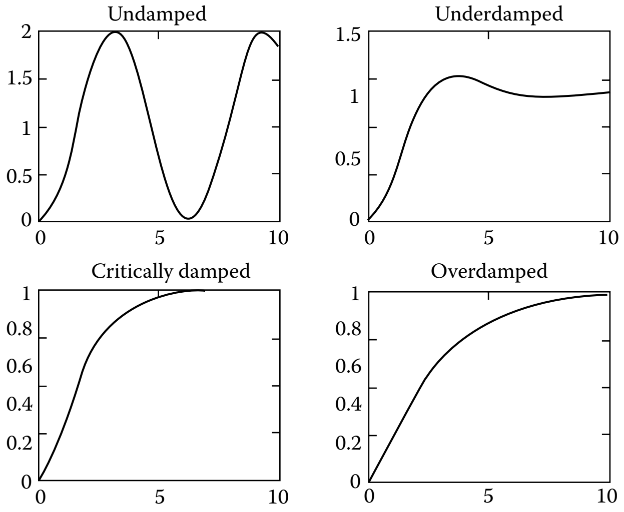


FIG. 2.1m
The effect of the damping ratio on second-order systems.

The top right panel corresponds to $0 < \xi < 1$ and produces an *underdamped output.* The speed of the response has deteriorated slightly in that the output $y(t)$ is a little slower in getting to the desired output 1. Note that the response oscillates, but the oscillation decays leaving only the steady-state solution.

The bottom left panel corresponds to being *critically* damped ($\xi = 1$). The response does not exceed the steady-state value of 1, but the speed of the system is noticeably slower than either the undamped or underdamped cases. The bottom right panel is the *overdamped* case ($\xi > 1$). Again, the response system is much slower.

From a control perspective, if the response is allowed to exceed the target value the system response is faster due to the underdamped nature of the system behavior. In fact, many heuristic tuning methods seek parameters that produce an underdamped response provided that the overshoot, or the amount that the output exceeds its steady-state response, is not too large.

If the second-order system also has dead time associated with it, then its response is not unlike that of the first-order system. That is, the response is delayed by the dead time, but the overall approach to steady state for different values of the damping ratio remains the same. A first-order system controlled with a PI controller may produce underdamped second-order system behavior.

Dead-Time Processes A contributing factor to the dynamics of many processes involving the movement of mass from one point to another is the transportation lag or dead time.

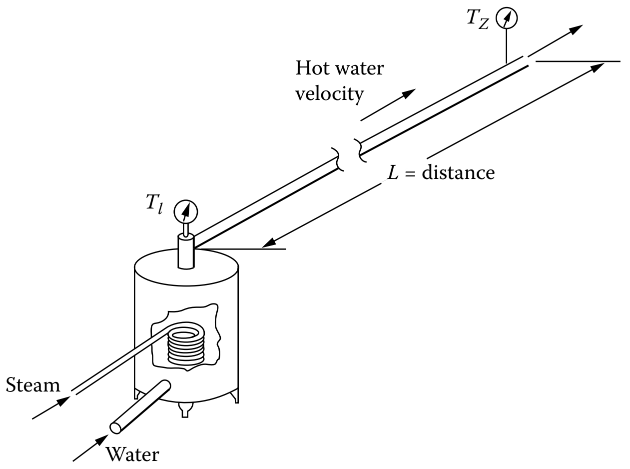


FIG. 2.1n
Transportation lag introduces a dead time in the water heater process.

Consider the effect of piping on the hot water to reach a location some distance away from the heater (Figure 2.1n).

The effect of a change in steam rate on the water temperature at the end of the pipe will not depend only on the resistance and capacitance effects in the tank. It will also be influenced by the length of time necessary for the water to be transported through the pipe. All lags associated with the heater system will be seen at the end of the pipe, but they will be delayed. The length of this delay is called the transportation lag (L) or dead time. The magnitude is determined as the distance over which the material is transported (l) divided by the velocity at which the material travels (v). In the heater example

$$L = v/l \qquad \mathbf{2.1(8)}$$

Dead time is the worst enemy of good control, and the process control engineer should therefore make a concentrated effort to minimize it. The effect of dead time can be compared to driving a car (the process) with closed eyes or with the steering wheel disconnected during that period. The goal of good control system design should be both to minimize the amount of dead time and to minimize the ratio of dead time to time constant (L/T).

The higher this ratio, the less likely it is that the control system will work properly, and once the L/T ratio reaches 1.0 ($L = T$), control by traditional PID (proportional–integral–derivative) strategies is unlikely to work at all. The various means of reducing dead time are usually related to reducing transportation lags. This can be achieved by increasing the rates of pumping or agitation, reducing the distance between the measuring instrument and the process, eliminating sampling systems, and the like.

When the nature of the process is such that the L/T ratio must exceed unity, or if the controlled process is inherently a dead-time process (a belt feeder for example), the traditional PID control must be replaced by control based on periodic adjustments, called sample-and-hold type control (see the next sections of this chapter).

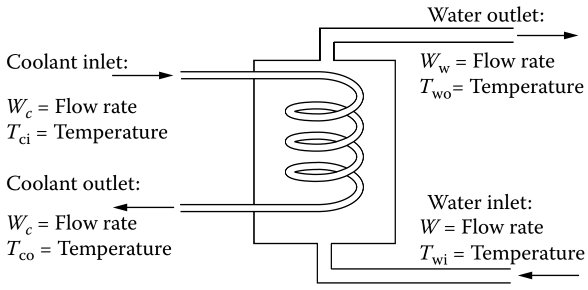


FIG. 2.1o
Process variables in a simple water cooler.

Process Variables Many external and internal conditions affect the performance of a processing unit. These conditions may be detected in terms of process variables such as temperature, pressure, flow, concentration, weight, level, etc. Most processes are controlled by measuring one of the variables that represent the state of the process and then by automatically adjusting another variable(s) that determine that state. Typically, the variable chosen to represent the state of the system is termed the "controlled variable" and the variable chosen to control the system's state is termed the "manipulated variable."

The manipulated variable can be any process variable that causes a reasonably fast response and is fairly easy to manipulate. The controlled variable should be the variable that best represents the desired state of the system. Consider the water cooler shown in Figure 2.1o. The purpose of the cooler is to maintain a supply of water at a constant temperature. The variable that best represents this objective is the temperature of the exit water, T_{wo}, and it should be selected as the controlled variable.

In other cases, direct control of the variable that best represents the desired condition is not possible. Consider the chemical reactor shown in Figure 2.1p. The variable that is

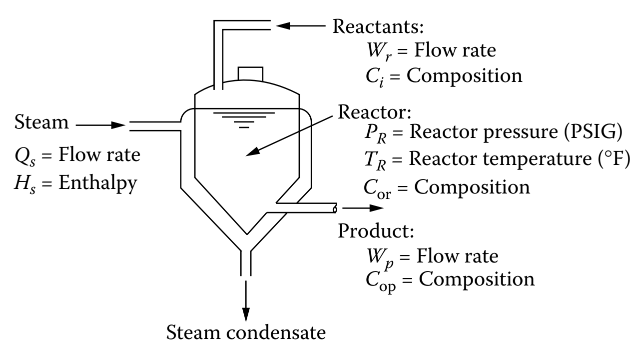


FIG. 2.1p
Process variables in a simple chemical reactor.

directly related to the desired state of the product is the composition of the product inside the reactor. However, in this case a direct measurement of product composition is not always possible.

If product composition cannot be measured, some other process variable is used that is related to composition. A logical choice for this chemical reactor might be to hold the pressure constant and use reactor temperature as an indication of composition. Such a scheme is often used in the indirect control of composition.

Degrees of Freedom To fully understand the "personality" of a process, one must also know the number of variables that can be independently controlled in a process. The maximum number of independently acting automatic controllers that can be placed on a process is the degrees of freedom (*df*) of that process. Mathematically, the number of degrees of freedom is defined as

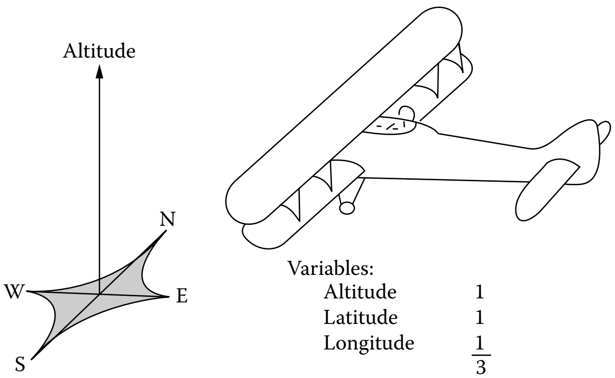
$$df = v - e \qquad \textbf{2.1(9)}$$

where df = number of degrees of freedom of a system; v = number of variables that describe the system; and e = number of independent relationships that exist among the various variables.

It is easy to see intuitively that a train has only one degree of freedom because only its speed can be varied, while boats have two and airplanes have three (see Figure 2.1q).

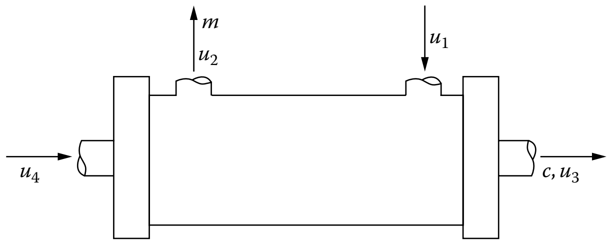
The degrees of freedom of industrial processes are more complex and cannot always be determined intuitively. In the case of a liquid-to-liquid heat exchanger, for example (Figure 2.1r), the total number of variables is six, while the number of defining equations is only one—the first law of thermodynamics, which states the conservation of energy.

Therefore, the degrees of freedom of this process are 6 – 1 = 5. This means that if five variables are held constant, this will result in a constant state for the sixth variable, the outlet temperature (c). Therefore, the maximum number of automatic controllers that can be placed on this process is five. Usually one would not exercise this option of using five controllers, and in the case of a heat exchanger of this type, one might use only one control loop.



Variables:	
Altitude	1
Latitude	1
Longitude	1
	3
Equations:	
None	–0
Degree of freedom:	3

FIG. 2.1q
The degrees of freedom of an airplane.



6 Variables:

c = heated fluid outlet temperature
m = heating fluid flow rate
u_1 = heating fluid inlet temperature
u_2 = heating fluid outlet temperature
u_3 = heated fluid flow rate
u_4 = heated fluid inlet temperature

FIG. 2.1r
The degrees of freedom of a liquid-to-liquid heat exchanger.

One would select the controlled variable (c) to be the process property that is the most important, because it has the most impact on plant productivity, safety, or product quality. One would select the manipulated variable (m) to be that process input variable that has the most direct influence on the controlled variable (c), which in this case is the flow rate of the heating fluid. The other *load variables* (*u_1 to u_4*) are uncontrolled independent variables, which, when they change, will upset the control system, and their effects can only be corrected in a "feedback" manner.

This means that a change in load variables is not responded to until it has upset the controlled variable (c). In order to fully describe this process, one would also consider such *system parameters* as the mass or the specific heat of the liquids, but such parameters are not considered to be variables.

When the process involves a phase change, the calculation of the degrees of freedom follows Gibbs's phase rule, stated in Equation 2.1(10):

$$n = n_c - n_p + 2 \qquad \textbf{2.1(10)}$$

where n = number of chemical degrees of freedom; n_c = number of components; and n_p = number of phases.

For example, if the process is a boiler producing *saturated* steam (Figure 2.1s), the number of components is one (H_2O), the number of phases is two (water and steam), and, therefore, the number of degrees of freedom is n = 1 – 2 + 2 = 1. Consequently, only one variable can be controlled: temperature or pressure, but not both. If a boiler produces

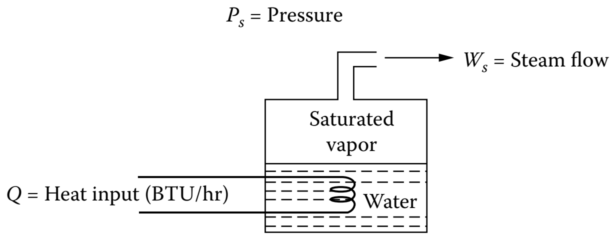
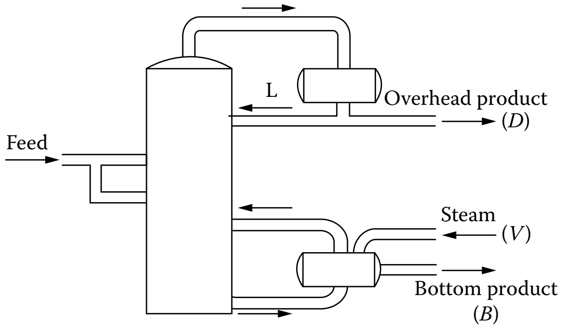


FIG. 2.1s
A saturated steam boiler has only one degree of freedom.

superheated steam, the number of degrees of freedom is two, and therefore both temperature and pressure can be independently controlled.

When the process is more complex, such as is the case of binary distillation, the calculation of the degrees of freedom also becomes more involved. Figure 2.1t lists 14 variables of this process, not all are independent. Since there are two components and two phases at the bottom, feed and overhead, Gibbs's law states that only two of the three variables (pressure, temperature, and composition) are independent. Therefore, the number of independent variables is only 11. The number of defining equations is two (the conservation of mass and energy), and therefore, the number of degrees of freedom for this process is 11 − 2 = 9. Consequently, not more than nine automatic controllers can be placed on this process.



Apparent variables:	Independent variables:
C_1 = Overhead temperature	2
C_2 = Overhead pressure	
C_3 = Overhead composition	
C_4 = Overhead flow rate	1
U_1 = Bottom temperature	2
U_2 = Bottom pressure	
U_3 = Bottom composition	
U_4 = Bottom flow rate	1
U_5 = Feed temperature	2
U_6 = Feed pressure	
U_7 = Feed composition	
U_8 = Feed percent vapor	1
U_9 = Feed flow rate	1
m = Steam flow rate (heat input)	1
	11

FIG. 2.1t
In a binary distillation process the number of independent variables is 11 and the number of degrees of freedom is nine.

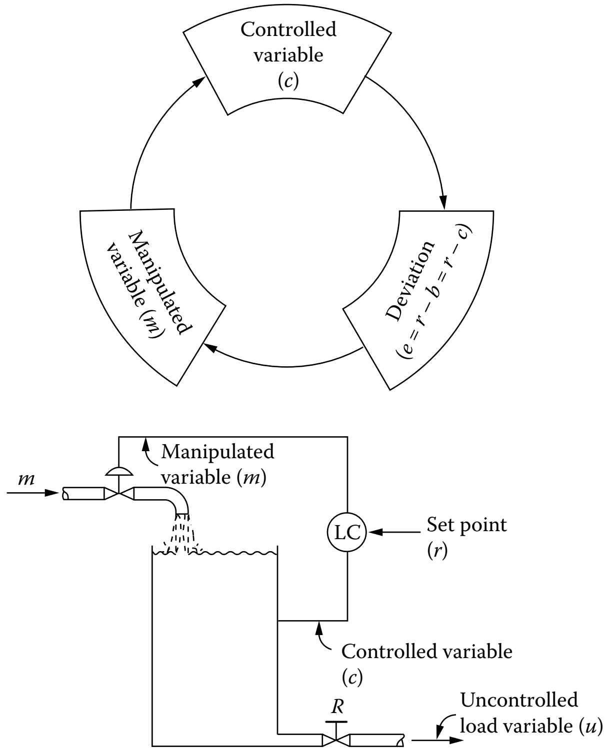


FIG. 2.1u
The addition of a controller closes the automatic control loop.

CLOSING THE LOOP[2]

After gaining a good understanding of the "personality" of the process and after having identified the controlled (*c*), manipulated (*m*), and uncontrolled load (*u*) variables, one can proceed to "close the loop" by installing an automatic controller onto the process.

The task of a controller is to measure the controlled variable *c* (say the level, in the case of Figure 2.1u), compare that measurement to some desired target value (*r* = set point), and, if there is an error (*c* does not equal *r*), modify its output signal (*m*) and thereby change the manipulated variable *m* (input to the process) in a way that will reduce the deviation from the set point.

Oscillation and Phase Shift

When a closed loop is in sustained oscillation, the period and amplitude of oscillation are constant, and therefore, the total phase shift in the loop must be 360 degrees. This state is called *sustained oscillation* or *marginal stability.*

In order to keep a ball bouncing at the same period and amplitude, it has to be struck at phase intervals of 360 degrees, when it is at its highest point. If it is struck sooner, the period of oscillation will be shortened. If the ball is the process and the hand is the controller, the goal is to push the ball back toward set-point whenever it is moving away from it.

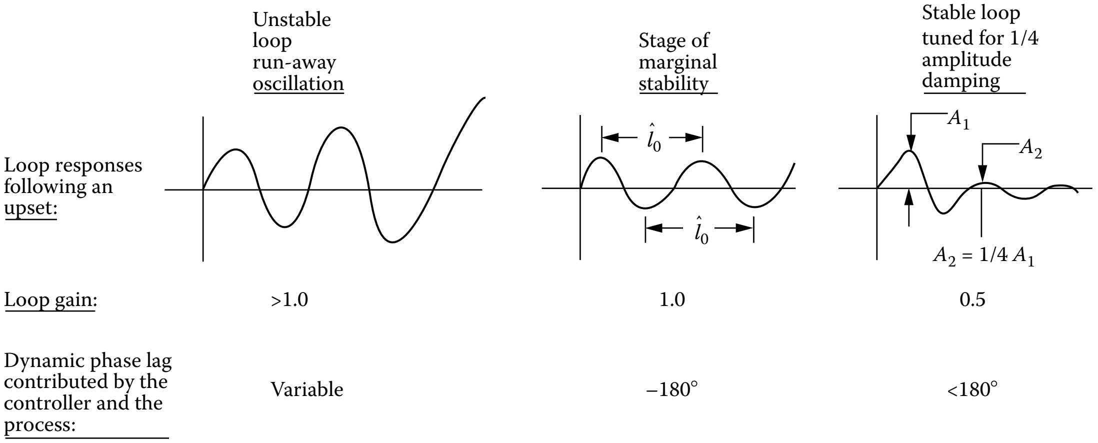


FIG. 2.1v
Control loops respond to upsets differently as a function of the loop gain and of the phase lag contributions of both the process and the controller.

Controlling a process is different from bouncing a ball only in that the force applied by the controller is applied continuously and not in periodic pulses. On the other hand, the two processes are similar from the perspective that in order to sustain oscillation the application of these forces must be displaced by 360 degrees (Figure 2.1v). The sources of this phase lag can be the process, the controller dynamics (modes), and the negative feedback of the loop.

The response of a process to a sinusoidal input is also sinusoidal if the process can be described by linear differential equations (a linear process). The phase shift in such processes is usually due to the presence of capacitance elements. For example, in the case of the process shown in Figure 2.1w, the process causes the output (c = outflow from the tank) to lag behind and oscillate with a smaller amplitude than the input of the process (m = inflow to the tank).

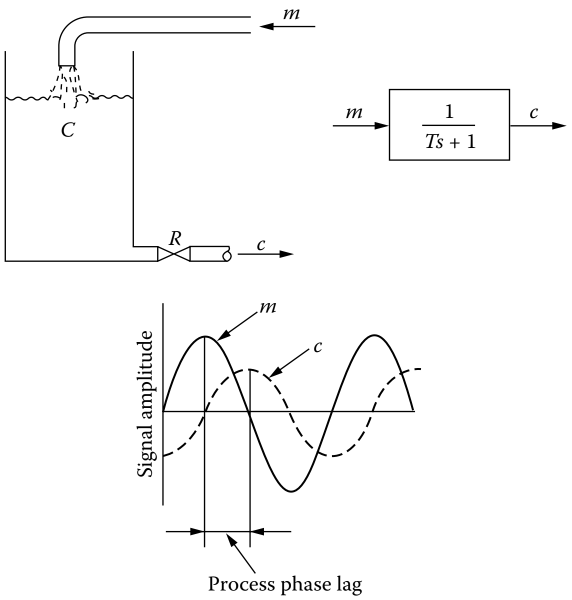


FIG. 2.1w
The outflow lags behind the inflow in this process because fluid must accumulate in the tank before the level will rise. This accumulation in turn will increase the outflow due to the resulting higher hydrostatic head.

Loop Gain and Tuning

It was shown in Figure 2.1v that sustained oscillation results if the loop gain is 1.0, and quarter amplitude damping results if the loop gain is 0.5. The goal of the tuning of most process control loops is to obtain quarter amplitude damping. This will result if the product of all the gains in the loop comes to 0.5. This end result is achieved through tuning, which is the process of finding the controller gain, which will make the overall gain product 0.5.

The controller gain (and most other gains also) consists of a steady-state component (K_c = the proportional setting), which is unaffected by the period of oscillation and the dynamic gain (gc), which varies with the period of the input (error) sinusoidal. The $(K_c)(gc)$ product is the total gain of the controller (Gc). Therefore, slow processes can be controlled with high-gain controllers, while controllers on fast processes must have low gains.

An example of a slow (low-gain) process is space heating, where it takes a long time (accumulation of the heat input provided by manipulated variable) to cause a small change in the controlled variable (the room temperature). If the combined gain of that process is 0.02, the controller gain required to provide quarter amplitude damping is 25 ($0.02 \times 25 = 0.5$).

If, on the other hand, the controlled process is a flow process, or a pH process near neutrality, the process gain is high. In such cases the controller gain must be low so that the total loop gain is 0.5.

The higher the controller gain, the more corrective action the controller can apply to the process in response to a small deviation from set point and the better the quality of the resulting control will be. Unfortunately, if the gain product of controller and process reaches unity, the process becomes unstable and undamped oscillations (cycling) will occur. Therefore, it is not possible to tightly control fast (high-gain) processes without cycling. It is easier to obtain tight control on slow, low-gain processes, because the use of high-gain controllers does not destroy stability.

The loop gain is the product of all the gains in the loop, including sensor, controller, control valve, and process. In a properly tuned loop, the product of all these gains is 0.5. What makes tuning difficult is that the process gain often varies with process load. For example, in heat transfer processes, when the heat load is low and the heat transfer surface available to transfer the heat is large, the transfer of heat is performed efficiently; therefore, under those conditions, this process is a high-gain process.

As the load rises, the same heat transfer process becomes a low-gain process because the fixed heat transfer area becomes less and less sufficient to transfer the heat. Therefore, as shown in Figure 2.1x, the gain of a heat transfer process (*Gp*) drops as the load rises.

Tuning such a system can be a problem because in order to arrive at an overall loop gain of 0.5, the controller should apply a high gain when the load is high and a low gain when the load is low. Standard controllers cannot do that because they have been tuned to provide a single gain. Therefore, if the loop was tuned (controller gain was selected) at high loads, the loop will cycle when the load drops, and if the loop was tuned at low loads, the loop will not be able to hold the process on set point (will be sluggish) when the load rises.

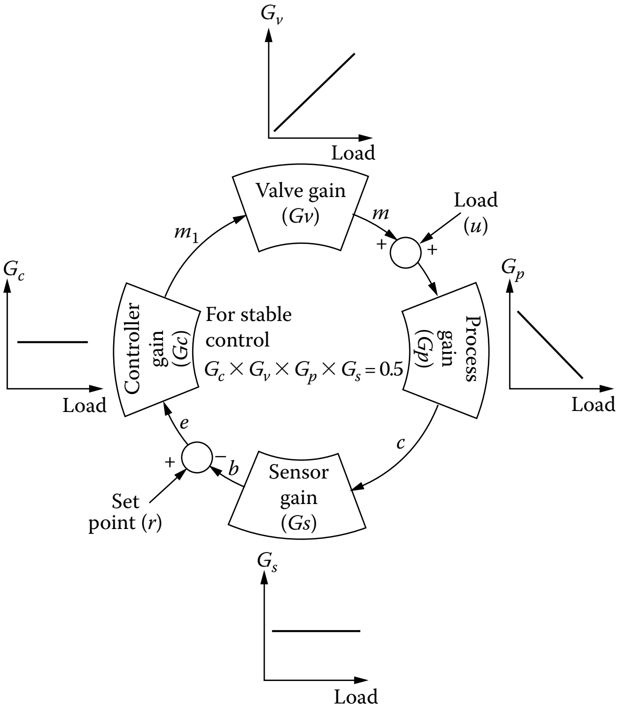


FIG. 2.1x

If the process gain varies with load, such as is the case of a heat transfer process, the gain product of the loop can be held constant by using a valve whose gain variation with load compensates for the process gain variation.

One way to compensate for this effect is to install a control valve in the loop (an equal percentage control valve), which increases its gain as the load rises. When the gain of the process drops, the gain of the valve increases and the total loop gain remains relatively unaffected.

Feedback Control

Two concepts provide the basis for most automatic control strategies: feedback (closed-loop) control and feedforward (open-loop) control. Feedback control is the more commonly used technique of the two and is the underlying concept on which most automatic control theory used to be based. Feedback control maintains a desired process condition by measuring that condition, comparing the measurement with the desired condition, and initiating corrective action based on the difference between the desired and the actual conditions.

The feedback strategy is very similar to the actions of a human operator attempting to control a process manually. Consider the control of a direct contact hot water heater. The operator would read the temperature indicator in the hot water line and compare its value with the temperature desired (Figure 2.1y). If the temperature was too high, he would reduce the steam flow, and if the temperature was too low, he would increase it. Using this strategy, he would manipulate the steam valve until the error is eliminated.

An automatic feedback control system would operate in much the same manner. The temperature of the hot water is measured and a signal is fed back to a device that compares the measured temperature with the desired temperature. If an error exists, a signal is generated to change the valve position in such a manner that the error is eliminated.

The only real distinction between the manual and automatic means of controlling the heater is that the automatic controller is more accurate and consistent and is not as likely to become tired or distracted. Otherwise, both systems contain the essential elements of a feedback control loop.

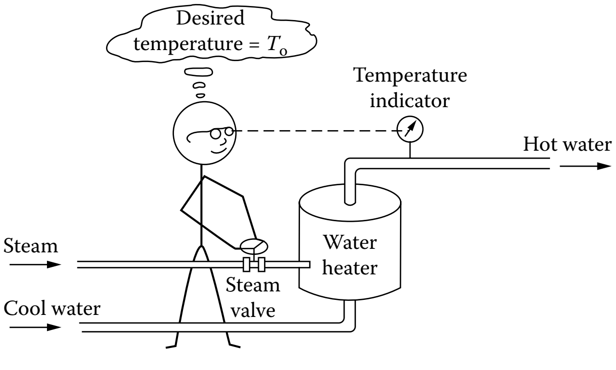


FIG. 2.1y

Manually accomplished feedback control.

Feedback control has definite advantages over other techniques in relative simplicity and potentially successful operation in the face of unknown contingencies. In general, it works well as a regulator to maintain a desired operating point by compensating for various disturbances that affect the system, and it works equally well as a servo system to initiate and follow changes demanded in the operating point.

Feedback Control Objectives

First, it is desirable that the output follow the desired behavior. Note that in all the subplots in Figure 2.1m, the response to a unit step change approaches the steady-state value of 1, which corresponds to the magnitude of the input step response. Because the output response eventually reaches the steady-state value, the *steady-state error,* or the difference between the desired final output and the actual one, is zero.

Second, almost always, the steady-state error should be zero to a step input, or constant targets, as inputs. In some cases, such as the case of a ramp input, it also is desirable for the steady-state error to be zero or nearly so. There may be an upper limit on the magnitude that is tolerable when no disturbances are present. However, in the presence of disturbances the steady-state error can become larger.

Third, the speed of response is important. From the discussion in connection with equation 2.1(7), viz. the solution of the differential equation, the steady-state is attained as the homogeneous portion of the solution of the differential equation approaches zero. A control system can affect the rate at which this happens. If the response of the system is sluggish, then the output (control action) of the controller is not changing enough in magnitude in response to the difference between the desired and actual output. By changing the parameters of the controller, the magnitude of the control action and the speed of response can be increased in response to control errors.

Fourth, the physical limitations of the plant constrain the ability of the controller to respond to input command changes. Another measure of the controller's speed is the *settling time.* The settling time is defined as the time after which the control system will remain within a given percentage of the desired final value when there are no outside disturbances. Figure 2.1z illustrates a "2% settling time," meaning the time it takes for a step response to approach the final steady-state value within 2%.

Lastly, note that (Figure 2.1l) the step change responses of a second-order system all have an overshoot, when the damping ratio of the system is less than one. Overshoot is defined as the percentage by which the peak response value exceeds the steady-state value (peak value of step response – steady-state value)/(steady-state value). A small overshoot can be acceptable, but large overshoots are not.

The PID Controller

In Sections 2.2, 2.3, and 2.4 detailed descriptions are provided of the various analog and digital PID algorithms and therefore only some of the basic aspects are discussed here.

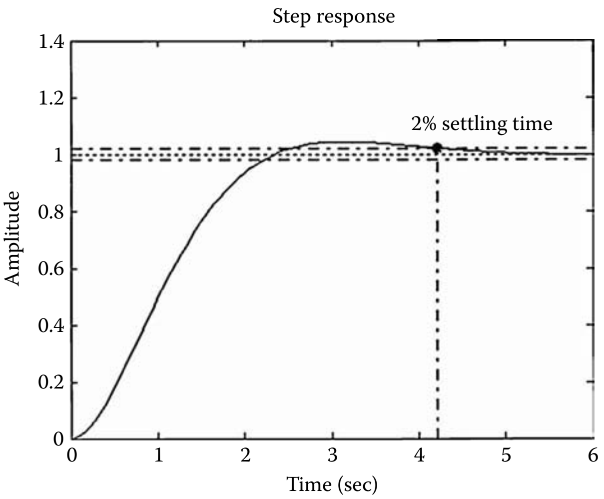


FIG. 2.1z
The step response shown has a 4.2-second "2% settling time."

As was shown earlier, the dynamic behavior of many processes can be modeled adequately as first- or second-order systems. Thus, the PID is an ideal tool with which to control such systems. Furthermore, the PID is easily understood and tuned by trained operators.

Consider the feedback control system with a plain proportional-only controller shown in Figure 2.1aa. Assume that the process to be controlled is a static system with a gain K_p. For a proportional-only controller, the controller output is the product of the error signal ($e = r - c$) and of the proportional gain K_c. That is,

$$u = K_c e \qquad \textbf{2.1(11)}$$

The closed-loop response is the relationship between the output (controlled variable), c, and the reference or set point input, r. This relationship is

$$c = \frac{K_c K_p}{1 + K_c K_p} r \qquad \textbf{2.1(12)}$$

Note that if r is a constant, say one, the controlled output is less than one. Thus, there is a nonzero steady-state error to constant inputs. This is not surprising because if $r = c$, then $e = r - c = 0$ and the output of the controller would also be

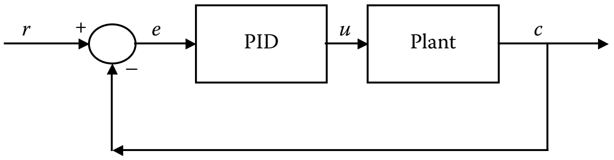


FIG. 2.1aa
Block diagram of a PID feedback control loop.

zero. This produces a contradiction in that c would be forced to zero, which, in general, is not the value of r.

Therefore, the plain proportional controller reduces but does not eliminate the error. Note from Equation 2.1(12) that as the controller gain increases, the controlled output, c, approaches the referenced input r more and more closely. Thus, the steady-state error, e, becomes smaller as K_c is made larger. But since K_c can never be infinite, the error is never zero.

The Derivative Mode To better understand the effect of derivative action, consider the situation in which the controller shown in Figure 2.1aa is proportional plus derivative (PD). In this case, the controller output is given by

$$u(t) = e(t) + T_d \frac{de(t)}{dt} \qquad \textbf{2.1(13)}$$

if the proportional gain K_c is 1 and the derivative gain is T_d, which is called the derivative time. Figure 2.1bb illustrates the controlled variable's step response, the error and derivative of the error signal, when the set point (reference) is $r = 1$.

The vertical lines in the top two plots of Figure 2.1bb are in locations where the controlled variable signal and the error signal have local maximums and minimums or points where the derivative is zero. Note that the controlled variable response has exceeded the set point (target value) of one. The excess overshoot is due to the presence of a certain momentum in the response of the system; the controller did not turn the input around in time to stop the system from exceeding the desired value.

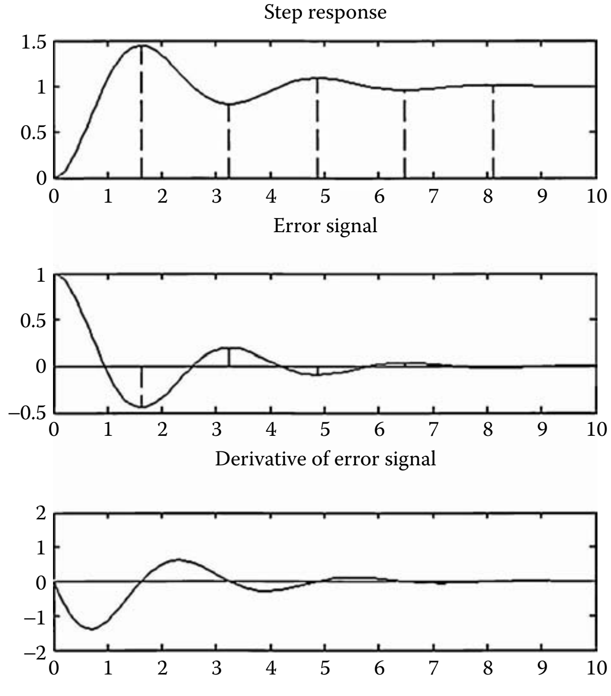


FIG. 2.1bb

The step response of a PD controller showing the responses of the controlled variable (top), the error (center), and the derivative of the error (bottom).

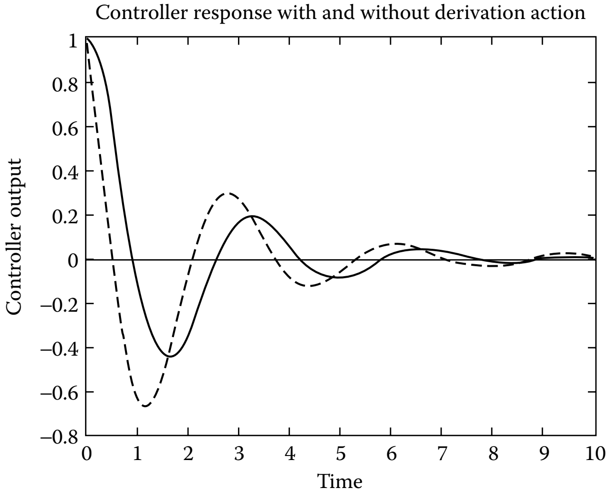


FIG. 2.1cc

The "anticipation" of the PD controller can be seen by noting the controlled variable response of a plain proportional controller (solid line) and that of a PD controller (dashed line) to the same step upset.

This is seen in the error signal that has not changed sign until the controlled variable output has exceeded the set point. Note also that the derivative is zero at the peak values and of opposite sign to the value of the error signal. When the error is added to a constant times the derivative, the result is a signal that changes sign earlier in time, that is, before the output has actually reached the steady-state value. Figure 2.1cc illustrates this.

The PD controller is used in applications where overshoot cannot be tolerated, such as pH neutralization. The reduction or elimination of overshoot can often be accomplished without significantly affecting the settling time of the closed-loop system. The primary disadvantage of derivative mode is its sensitivity to noise. Noise is generally of high frequency, and differentiating just amplifies it. The controller output can become cyclic or unstable, which can have a detrimental effect on the longevity of actuators such as valves or motors.

Integral Mode Nearly all controllers have some form of integral action. Integral action is important because it corrects based on the accumulated error, which is the area under the error curve. If the error goes to zero, the output of the integrator is the constant area that had accumulated up to that point.

Consider the feedback system illustrated in Figure 2.1aa. The task of the integral term in the PID algorithm is to find the manipulated variable (the input to the plant) needed to drive the steady-state error to zero when the set point (reference input) is constant.

When the error is zero, both the proportional term and the derivative term contribute nothing to the controller output.

Only the integral term provides any input to the controller output; only the integrator drives the manipulated variable to compensate for the area under the past error curve.

In summary, the PID controller produces an output defined as

$$u(t) = K_p\left(e(t) + \frac{1}{T_i}\int e(\tau)d\tau + T_d\frac{de(t)}{dt}\right) \qquad \textbf{2.1(14)}$$

where K_p is the proportional gain; T_d is the derivative time; and T_i is the integral time.

The integral time can be viewed as the amount of the time it takes for the integral component to make the same contribution as the proportional term. If the integral time is short, the integral contribution to the PID output is large and too much integral gain (T_i too small) can cause the system to become unstable.

Feedforward Control

Feedforward control is another basic technique used to compensate for uncontrolled disturbances entering the controlled process. Both feedback and feedforward control are discussed in detail in Section 2.8 and therefore only an introduction is given here. In this technique the control action is based on the disturbance input into the process without considering the condition of the process. In concept, feedforward control yields much faster correction than feedback control does, and in the ideal case compensation is applied in such a manner that the effect of the disturbance is never seen in the controlled variable, the process output.

A skillful operator of a direct contact water heater could use a simple feedforward strategy to compensate for changes in inlet water temperature by detecting a change in inlet water temperature and in response to that, increasing or decreasing the steam rate to counteract the change (Figure 2.1dd). This same compensation could be made automatically with an inlet temperature detector designed to initiate the appropriate corrective adjustment in the steam valve opening.

The concept of feedforward control is very powerful, but unfortunately it is difficult to implement in most process control applications. In many cases disturbances cannot be accurately measured, and therefore pure feedforward cannot be used. The main limitation of feedforward is due to our inability to prepare perfect process models or to make perfectly accurate measurements.

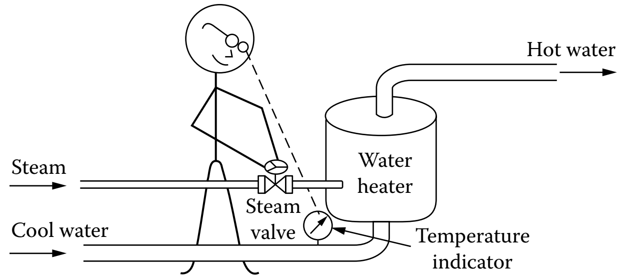


FIG. 2.1dd
The concept of feedforward control implemented by a human operator.

Because of these limitations, pure feedforward would accumulate the errors in its model and would eventually "self-destruct." The main limitations of feedback control are that feedback cannot anticipate upsets but can only respond to them after the upsets have occurred, and that it makes its correction in an oscillating, cycling manner.

It has been found that combining feedback and feedforward is desirable in that the imperfect feedforward model corrects for about 90% of the upset as it occurs, while feedback corrects for the remaining 10%. With this approach, the feedforward component is not pushed beyond its abilities, while the load on the feedback loop is reduced by an order of magnitude, allowing for much tighter control.

Feedforward Response

Ideally the feedforward correction would be so effective that a disturbance would have no measurable effect on the controlled variable, the process output. As an example, consider a first-order system in which there is a measurable disturbance. Suppose that a process disturbance occurs at time $t = 5$ seconds, as shown in the top segment of Figure 2.1ee, and causes the PID controller to generate a corrective action as shown in lower part of Figure 2.1ee. Note that while the controller will eliminate the disturbance, it will do that only after it has occurred.

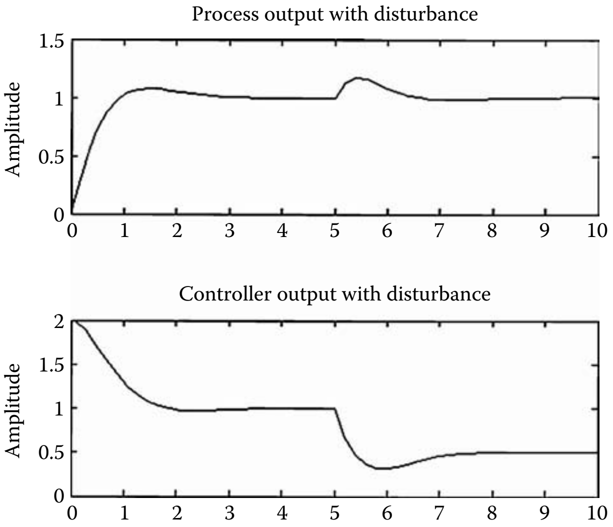


FIG. 2.1ee
If a step disturbance occurs at t = 5, the controlled variable of a first-order process responds to that upset as shown in the top portion of the figure. The bottom part shows the response of a feedback PID controller to such an upset, which generates the manipulated variable.

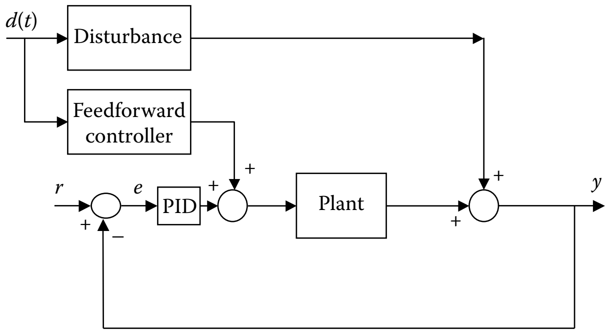


FIG. 2.1ff
Block diagram representation of a combination of a feedback PID control loop with a feedforward compensator.

Now consider a control system that includes a 100% effective feedforward controller as shown in Figure 2.1ff. The bottom part of Figure 2.1gg shows the response of a perfect feedforward compensator to an upset that has occurred at $t = 5$. The middle section of the figure shows the feedback PID output and the top portion shows the controlled variable, which remains undisturbed.

Note that the compensator does not respond until the disturbance occurs at time $t = 5$ seconds. With the feedforward compensation that is added to the PID controller output (Figure 2.1ff), the system exactly compensates for the disturbance. In practice, exact compensation is not possible. However, much of the effect of a disturbance can be mitigated by a judiciously designed feedforward compensator.

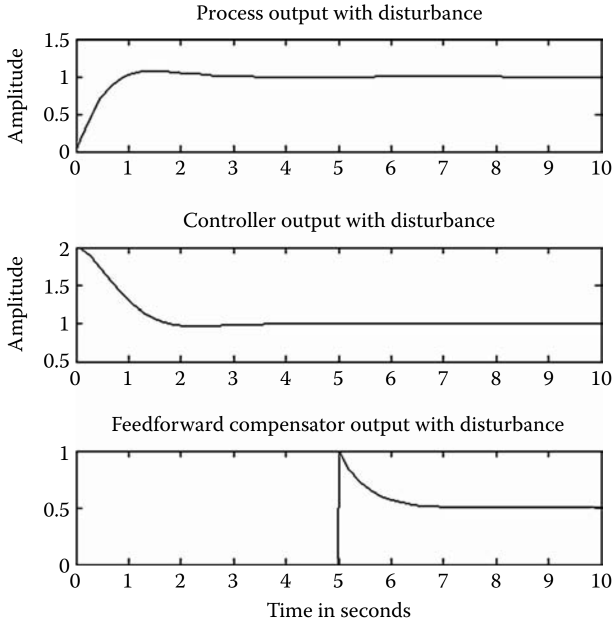


FIG. 2.1gg
The response of the control loop shown in Figure 2.1ff to a process upset that occurs at t = 5. In the bottom part, the output of the feedforward compensator is shown. When this is added to the output of the feedback PID controller, the upset is fully corrected and as shown in the top curve, the controlled variable is not upset at all.

Cascade Control

Section 2.6 discusses cascade control in detail; therefore, only an introduction is given here. In cascade control two controllers are used to control a single process variable. For example, consider the composition control of the product from a continuously stirred reactor in which an exothermic chemical reaction takes place. If the compensation is directly related to the temperature within the reactor and if the composition itself is difficult to measure online, then the reactor temperature is often chosen as the controlled variable. To control that temperature, the cooling water flow to the jacket is usually manipulated by the master (outer) temperature controller, which is adjusting the set point of the slave (inner) flow controller. Together these two controllers form a cascaded control loop, as shown in Figure 2.1hh.

The control loop for the cooling water flow is called the *inner loop.* The *outer loop* is the controller for the reactor temperature. The inner loop must have a response time that is faster than that of the outer loop. The outer loop controller provides the set point for the inner loop. The inner loop must be faster than the outer loop, so that the set point will not change too fast for the inner loop dynamics to follow; if it did, stability problems would result.

Filtering

Filters are needed to eliminate noise from such signals as the controlled variable. If the noise in the measurement signal is not reduced, it can pass through the controller and cause cycling and eventual damage to the control valve or other final control element.

Noise tends to be of a high frequency and changes more rapidly in time than does the controlled process variable. The task of a filter is to block the rapidly changing component of the signal but pass the slowly changing component, so that the filter output will have less noise than does the raw signal.

Filtering may have little or no effect on control performance if the closed-loop system response time is slow compared to the response time of the filter. In other words, if the

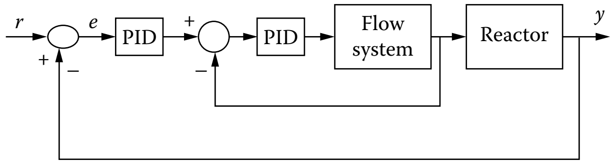


FIG. 2.1hh
Block diagram of a cascade control loop.

filter reaction time is assumed to be instantaneous compared to the speed of response of the closed-loop system, then there is no dynamic effect on the closed-loop system. On the other hand, if the closed-loop is fast relative to the speed of response of the filter, then the dynamic effects of the filter will have an effect.

Therefore, filtering can complicate the controller and can limit the response of a PID controller. One way to compensate for this is to tune the controller with the filter inserted into the feedback loop. Although the PID controller may give better performance without the filter, the effectiveness of the filter depends on the nature and amount of the noise and the desired closed-loop dynamics.

ADVANCED CONTROLS

The scope of the discussion in this first section was limited to basics of process control, and therefore such subjects as multivariable loops, interaction and decoupling, optimization, artificial intelligence, statistical process control, relative gain calculations, neural networks, model-based and model-free controls, adaptive control, fuzzy logic and statistical process control have not been covered. These topics and many others are discussed in the sections that follow.

References

1. Eckman, D. P., *Automatic Process Control,* New York: John Wiley & Sons, 1958.
2. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Bibliography

Arkun, Y., and Downs, J., “A General Method to Calculate Input–Output Gains and the Relative Gain Array for Integrating Processes,” *Computers & Chemical Engineering,* 14, Vol. 10, 1990, pp. 1101–1110.

Balakrishman, R., “Dead Time Controller,” ISA/94 Technical Conference, Anaheim, 1994.

Balchen, J. G., and Mummè, K. I., *Process Control Structures and Applications,* New York: Van Nostrand Reinhold, 1988.

Butterworths, W. R., *Advanced Process Control,* New York: McGraw-Hill, 1981.

Chang, J.-W., and Yu, C.-C., “The Relative Gain for Non-Square Multivariable Systems,” *Chemical Engineering Science* 45, Vol. 5, 1990, pp. 1309–1323.

Cutler, C. R., and Ramaker, B. L., “Dynamic Matrix Control—A Computer Control Algorithm,” *ACC Conference Proceedings,* San Francisco, California, 1980.

D’Azzo, J. J., and Houpis, C. H., *Linear Control System Design, Conventional and Modern,* New York: McGraw-Hill, 1988.

Doherty, J., “Detecting Problems with SPC,” *Control,* November 1990, pp. 70–73.

Dorsey, J., *Continuous and Discrete Control Systems: Modeling, Identification, Design, and Implementation,* New York: McGraw-Hill, 2002.

Franklin, G. F., Powell, J. D., and Emami-Naeini, A., *Feedback Control of Dynamic Systems,* 2nd ed., Reading, MA: Addison-Wesley, 1991.

Frerichs, D., “Integrating Diagnostic Expert System with Statistical Process Control in a Modern Distributed Control System,” paper presented at ISA-90, Instrument Society of America, 1990.

Grant, E., and Leavenworth, R., *Statistical Quality Control,* New York: McGraw-Hill, 1988.

Holmes, D. S., “Time Series Analysis Overcomes SPC Problems,” *Control,* February 1991, pp. 36–38.

Hoo, K. A., Piovoso, M. J., Schnelle, P. D., and Rowan, D. A., “Process and Controller Performance Monitoring: Overview with Industrial Applications,” *IEEE Journal of Adaptive Control & Signal Processing,* Vol. 17, No. 7–9, 2003, pp. 635–662.

Hougen, J. O., *Measurements and Control Applications for Practicing Engineers,* Boston: Cahners Books, 1972.

Johnson, C. D., *Process Control Instrumentation Technology,* New York: Wiley & Sons, 1988.

Jones, B. E., *Instrumentation, Measurement and Feedback,* New York: McGraw-Hill, 1977.

Joseph, B., et al., “Knowledge Acquisition and Knowledge Representation in an Expert System for Control of a Batch Process,” paper presented at AIChE National Meeting, Houston, 1991.

Kecman, V., *State-Space Models of Lumped and Distributed Systems,* Heidelberg: Springer-Verlag, 1988.

Kelly, K. P., “Should You Be Training Your Operators with Simulation Software,” *Control,* June 1992, pp. 40–43.

Koenig, D. M., *Control and Analysis of Noisy Processes,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Korn, G., *Interactive System Simulation,* New York: McGraw-Hill, 1989.

Kuo, B. C., *Automatic Control,* Englewood Cliffs, NJ.: Prentice Hall, 1982.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

MacFarlane, A. G. J., and Hung, Y. S., “Quasi-Classical Approach to Multivariable Feedback Systems Design,” 2nd IFAC Symposium, Purdue University, September 1982, pp. 39–48.

McAvoy, T. J., *Interaction Analysis,* Research Triangle Park, NC: Instrument Society of America, 1983.

McColl, M., and Collins, S., “Deadtime Compensation Algorithm,” ISA/93 Technical Conference, Chicago, 1993.

McMillan, G., *Tuning and Control Loop Performance,* ISA, 2nd ed., 1990, pp. 45–55.

Morely, R., “Faster Than Real Time,” *Manufacturing Systems,* May 1992, p. 58.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1986.

Nise, N. S., *Control System Engineering,* New York: John Wiley & Sons, 2000.

Nyquist, H., “Regeneration Theory,” *Bell System Technical Journal,* Vol. 13, 1952, p. 1.

Oakland, J., *Statistical Process Control,* New York: Wiley, 1986.

Ogata, K., *Modern Control Engineering,* Englewood Cliffs, NJ.: Prentice Hall, 1970.

Perron, S., “The New Generation Language for Controls Design,” *Control Engineering,* February 1992.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1991, p. 136.

Rich, S., et al., “Development of a Diagnostic Expert System for a Whipped Toppings Process,” *Journal of Loss Prevention in Process Industry,* Vol. 2, July 1989.

Rijnsdorp, J. E., MacGregor, J. F., Tyreus, B. D., and Takamatsu, T., *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes,* Oxford: Pergamon Press, 1990.

Roffel, B., and Chin, P., *Computer Control in the Process Industries,* Chelsea, MI: Lewis Publishers, 1989.

Rowan, D., “On-Line Expert Systems in Process Industries,” *AI/Expert,* August 1989.

Ryskamp, C. J., “New Control Strategy Improves Dual Composition Control,” *Hydrocarbon Processing,* May 1981.

Seborg, D., et al., *Process Dynamics and Control,* New York: Wiley, 1989.

Shinskey, F. G., *Chemical Process Control II: Proceedings,* T. J. McAvoy et al., Eds., Amsterdam: Elsevier, 1986.

Shinskey, F. G., *Distillation Control,* 2nd ed., New York: McGraw-Hill, 1984.

Shinskey, F. G., *Process Control Systems,* New York: McGraw-Hill, 1996.

Skogestad, S., Lundstrom, P., and Jacobsen, E. W., "Selecting the Best Distillation Control Configuration," *AIChE Journal* 36, May 1990, pp. 753–764.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984, p. 310.

Vasbinder, E. M., and Hoo, K. A., "A Decision-Based Approach to Plantwide Control Structure Synthesis," *Industrial & Engineering Chemistry Research,* Vol. 42, No. 20, 2003, pp. 4586–4598.

Vervalin, C. H., "Training by Simulation," *Hydrocarbon Processing,* December 1984, pp. 42–49.

2.2 Control Modes—PID Controllers

C. F. MOORE (1970, 1985) **B. G. LIPTÁK** (1995) **R. BARS, J. HETTHÉSSY** (2005)

INTRODUCTION

A controller compares its measurement to its set point (r) and, based on the difference between them (e = error), generates a correction signal to the final control element (e.g., control valve) in order to eliminate the error. The way a controller responds to an error determines its character, which significantly influences the performance of the closed-loop control system.

The different ways in which a controller may behave are the different control modes. They include on/off, floating, proportional (P), integral (I), differential (D), and many others. These control modes will be described in this section.

A controller is called direct acting if its output increases when its measurement rises and is called reverse acting if its output decreases when its measurement rises.

ON/OFF CONTROL

The oldest control strategy is to use a switch for control. For example, if the controlled process is the room temperature, the switch would turn on the heat source when the temperature is low and turn it off when the desired comfort level is reached.

A perfect on/off controller is "on" when the measurement is below the set point. Under such conditions the manipulated variable is at its maximum value. When the measured variable is above the set point, the controller is "off" and the manipulated variable is at its minimum value.

$$\text{if} \begin{cases} e > 0 & \text{then } m = \max \\ e < 0 & \text{then } m = \min \end{cases} \qquad \mathbf{2.2(1)}$$

In most practical applications, due to mechanical friction or to the arcing of electrical contacts, the error must exceed a narrow band (around zero error) before a change will occur. This band is known as the differential gap, and its presence is usually desirable to minimize the cycling of the process. Figure 2.2a shows the response of an on/off controller to a sinusoidal input.

The block diagram shown in Figure 2.2b can be used to analyze the behavior of an on/off controller. A dead time and a time constant can represent *the* control response. The on/off control of heating processes can be represented with such dynamics. The on/off controller can be implemented by a relay having a hysteresis of h. If the error signal is greater than h, the relay is switched on, providing output value u_o and if the error goes below $-h$, it is switched off.

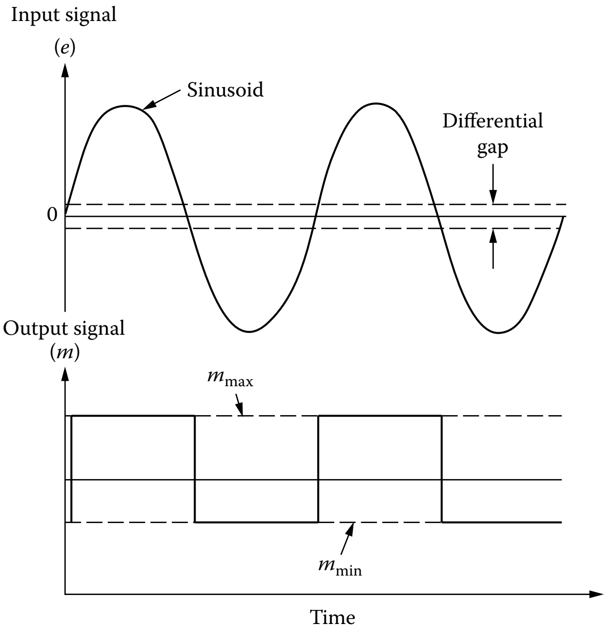


FIG. 2.2a
The response of a two-position controller.

Figure 2.2c shows the output (manipulated variable $-u$) and control (controlled variable $-y$) signals of the on/off control system. The relay is first switched on until the controlled variable (y) reaches the value $r + h$. At that point it

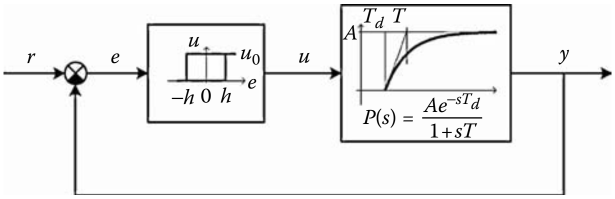


FIG. 2.2b
Block diagram of an on/off control system.

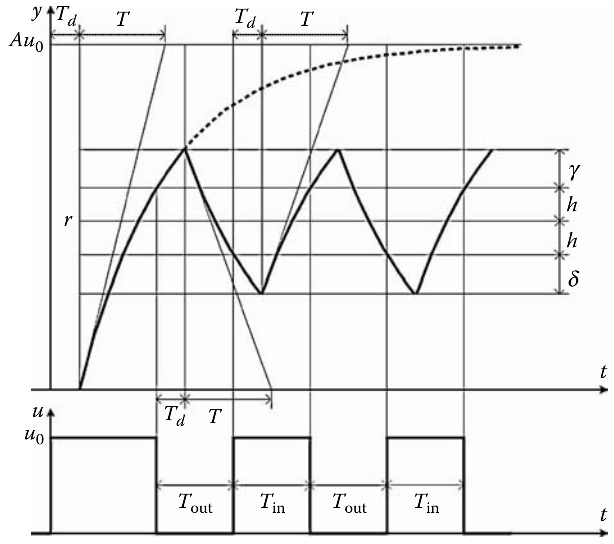


FIG. 2.2c
*The on/off output signal (*u*) of the controller results in a cycling response of the controlled variable (*y*).*

switches off, but the controlled variable keeps increasing through the duration of the switch's dead time by the amount γ. Then it starts decreasing exponentially. When the output (y) has dropped to $r - h$, the relay is switched on again, but the controlled variable (y) will continue to decrease for the period of the dead time by a value of δ.

This cycle is repeated as the controlled variable (y) is oscillating around a mean value. The amplitude of the oscillations is smaller if both the hysteresis band and the dead time are small. With a small hysteresis band the frequency of the oscillations becomes higher. It can be observed that the mean value of the controlled variable (y) is shifted a bit from the reference value (set point – r). This shift depends on the value of the reference signal.

The performance of the on/off control system can be improved by applying a negative feedback around the relay with a proportional lag element as shown in Figure 2.2d. This feedback causes the oscillations to have a higher switch-over frequency and a smaller amplitude range. The addition of an integrating (I) or a proportional-plus-integral (PI) mode to the on/off controller could further improve its performance.

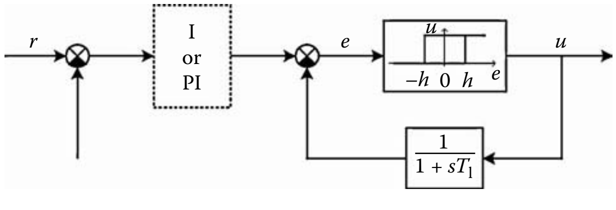


FIG. 2.2d
The performance of an on/off controller can be improved by the addition of an inner feedback.

SINGLE-SPEED FLOATING CONTROL

For self-regulating processes with little or no capacitance, the single-speed floating controller can be used. The output of this controller can either be increasing or decreasing at a certain rate. Such control is commonly used when the final control element is a single-speed reversible motor.

The controller is usually provided with a neutral zone (dead zone), and when the measurement is within that zone, the output of the controller is zero. This dead zone is desirable because otherwise the manipulated variable would be changing continually in one direction or the other.

With this controller, the output to the reversible motor is either forward, reverse, or off. Mathematically the single-speed floating control is expressed as

$$\text{if}\begin{cases} e > +\varepsilon/2 \\ e < -\varepsilon/2 \\ -\varepsilon/2 \le e \le \varepsilon/2 \end{cases} \quad \text{then}\begin{cases} m = +T_i t + M_{01} \\ m = -T_i t + M_{02} \\ m = M_{03} \end{cases} \qquad \mathbf{2.2(2)}$$

where t = time; m = controller output (manipulated variable); e = controller input (error signal); ε = constant defining neutral zone; T_i = controller speed constant; and M_{o1}, M_{o2}, M_{o3} = constants of integration.

The response of a single-speed floating controller to a sinusoidal input is shown in Figure 2.2e.

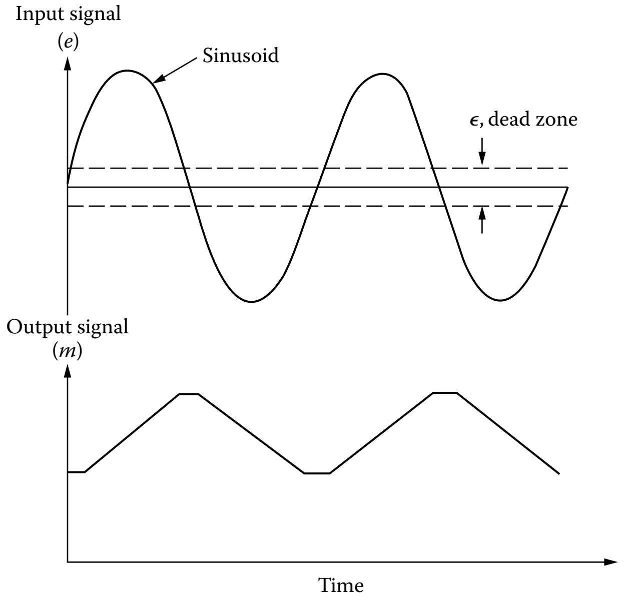


FIG. 2.2e
The response of a single-speed floating controller to a sinusoidal measurement signal.

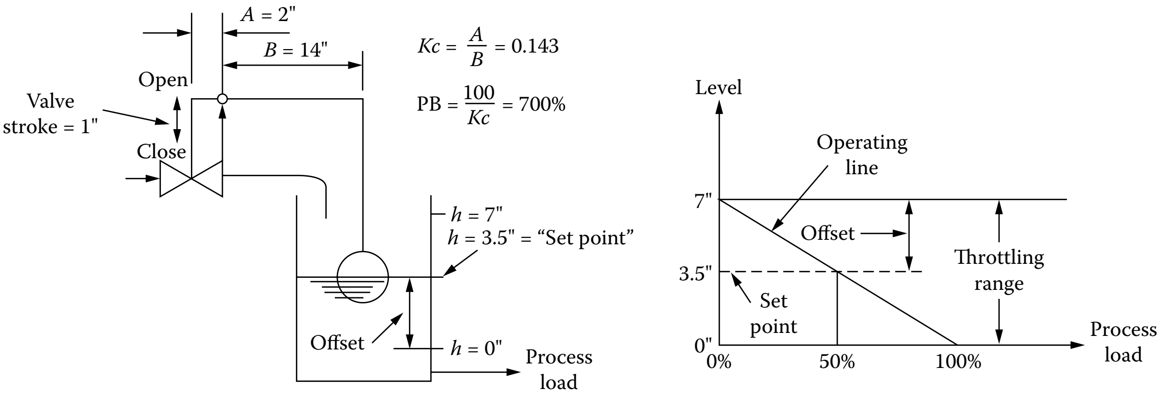


FIG. 2.2f

A plain proportional controller cannot maintain its set point because in order to operate at a load other than 50%, it first has to move the manipulated variable (inflow) away from 50%, which corresponds to its set point.

THE PROPORTIONAL CONTROL MODE

The proportional (P) mode alone is the simplest linear control algorithm. It is characterized by a constant relationship between the controller input and output. The adjustable parameter of the proportional mode, K_c, is called the proportional gain. It is frequently expressed in terms of percent proportional band, PB, which is inversely related to the proportional gain as shown below:

$$K_c = 100/\mathrm{PB} \qquad \mathbf{2.2(3)}$$

"Wide bands" (high PB values) correspond to less "sensitive" controller settings, and "narrow bands" (low PB percentages) correspond to more "sensitive" controller settings.

As the name "proportional" suggests, the correction generated by the proportional control mode is proportional to the error. Equation 2.2(4) describes the operation of the proportional controller.

$$m = K_c e + b = (100/\mathrm{PB})(e) + b \qquad \mathbf{2.2(4)}$$

where m = the output signal to the manipulated variable (control valve); K_c = the gain of the controller; e = the deviation from set point or error; PB = the proportional band ($100/K_c$); b = the live zero or bias of the output, which in pneumatic systems is usually 0.2 bars and in analog electronic loops is 4 mA.

The proportional controller responds only to the present. It cannot consider the past history of the error or the possible future consequences of an error trend. It simply responds to the present value of the error. It responds to all errors in the same manner, in proportion to them.

When a small error results in a large response, the gain (K_c) is said to be large or the proportional band (PB) is said to be narrow. Inversely, when a large error causes only a small response, the controller is said to have a small gain or a wide proportional setting. The gain in digital control system (DCS) control packages is usually adjustable from 0 to 8, while in analog controllers it can usually be adjusted from 0.02 to about 25.

Proportional Offset

The main limitation of plain proportional control is that it cannot keep the controlled variable on set point.

Let us consider the level control system shown in Figure 2.2f. The float-type proportional controller can only respond to a load change (change in the outflow of water) because it must experience a change in level before it can respond. Therefore, the only condition when this process will be on set point is when the load is 50%. In all other cases the level will have to travel up or down on its "operating line" as a function of the load. The difference between the actual value of the level and set point is called the offset, because this is the amount by which the process is off set point.

The gain of this mechanical controller is the ratio of the lengths of the two arms around the pivot. $K_c = A/B = 2/14 = 0.143$, or PB = 700%. One can increase the gain of this controller by moving the pivot to the right. This would result in requiring a smaller change in level to fully stroke the valve, which in turn would narrow the throttling range of the system and would make the operating line more horizontal.

It is evident that by increasing the gain, one can reduce (but not eliminate) the offset. If the controller gain is very high, the presence of the offset is no longer noticeable, and it seems as if the controller is keeping the process on set point.

Unfortunately, most processes become unstable if their controller is set with such high gain. The only exceptions are the very slow processes. For this reason the use of plain proportional control is limited to slow processes that can tolerate high controller gains (narrow proportional bands). These include the float-type valves, thermostats, and humidostats. In other processes, the offset inherent in proportional control (Figure 2.2g) cannot be tolerated.

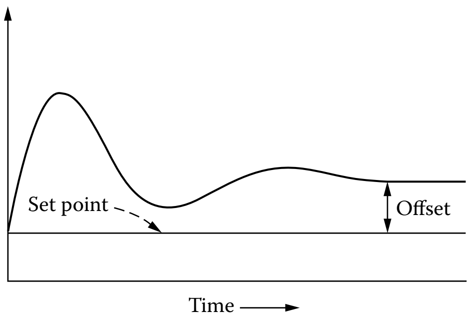


FIG. 2.2g
After a permanent load change, the proportional controller is incapable of returning the process back to the set point and an offset results. The smaller the controller's gain, the larger will be the offset.

THE INTEGRAL MODE

The integral (I) control mode is also sometimes called reset mode because after a load change it returns the controlled variable to set point and eliminates the offset, which the plain proportional controller cannot do. The mathematical expression of the integral-only controller is

$$m = \frac{1}{T_i}\int e\,dt + b \qquad 2.2(5)$$

while the mathematical expression for a proportional-plus integral (PI) controller is

$$m = K_c\left[e + \frac{1}{T_i}\int e dt\right] + b \qquad 2.2(6)$$

The term T_i is the integral time setting of the controller. It is also called reset time or repetition time, because with a step input, the output of a pure integrator reaches the value of the step input during the integral time, assuming $b = 0$.

The integral mode has been introduced in order to eliminate the offset that plain proportional control cannot remove. The reason proportional control must result in an offset is because it disregards the past history of the error. In other words, it disregards the mass or energy that should have been but was not added to (or removed from) the process, and therefore, by concerning itself only with the present, it leaves the accumulated effect of past errors uncorrected.

The integral mode, on the other hand, continuously looks at the total past history of the error by continuously integrating the area under the error curve and eliminates the offset by forcing the addition (or removal) of mass or energy that should have been added (or removed) in the past.

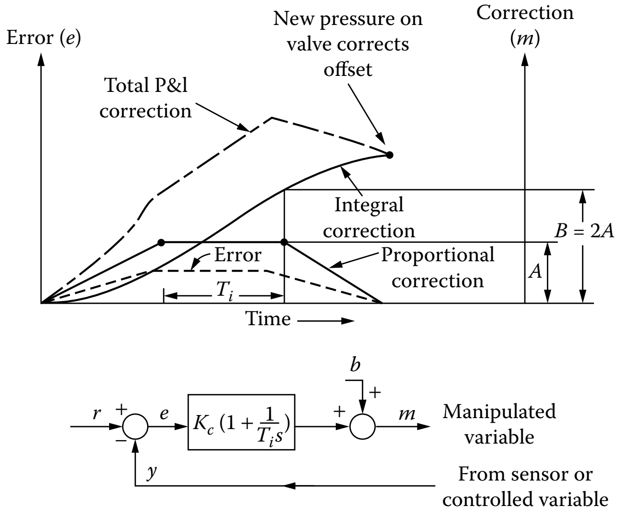


FIG. 2.2h
The contribution of the integral mode to the controller's output signal (m) is a function of the area under the error curve.

Figure 2.2h illustrates the correction generated by the integral mode in response to a supposed error curve indicated with a dotted line in the figure. It also shows the proportional and the combined (PI) correction.

Note that when the error is constant and therefore the proportional correction is also constant, the integral correction is rising at a constant rate because the area under the error curve is still rising. When the error and with it the proportional correction are both dropping, the integral correction is still rising because the area under the error curve is still rising. When the error reaches zero, the integral correction is at its maximum. It is this new signal level going to the control valve that serves to eliminate the offset.

The units of setting the integral time are usually given in "repeats/minute" or in "minutes/repeat." In DCS systems, the integral setting of control loops can usually be set from 0 to 300 repeats/minute, or from 0.2 seconds to about 60 minutes or more in units of minutes/repeat.

The meaning of the term "repeats/minute" (or its inverse) can be understood by referring to Figure 2.2h. Here, in the middle section of the error curve, the error is constant and therefore the proportional correction is also constant (A).

If the length of the duration of the middle section is one integral time (T_i), the integral mode is going to repeat the proportional correction by the end of the first integral time ($B = 2A$). The integral mode will keep repeating (adding "A" amount of correction) after the passage of each integral time during which the error still exists. The shorter the integral time, the more often the proportional correction is repeated (the more repeats/minute), and thus the more effective is the integral contribution.

Pure integral control (floating control) is seldom used, except on very noisy measurements as in some valve position or flow control systems, where the PI loop is usually tuned to have a low gain and lots of reset (integral). The proportional mode acts as a noise amplifier, while the integral mode

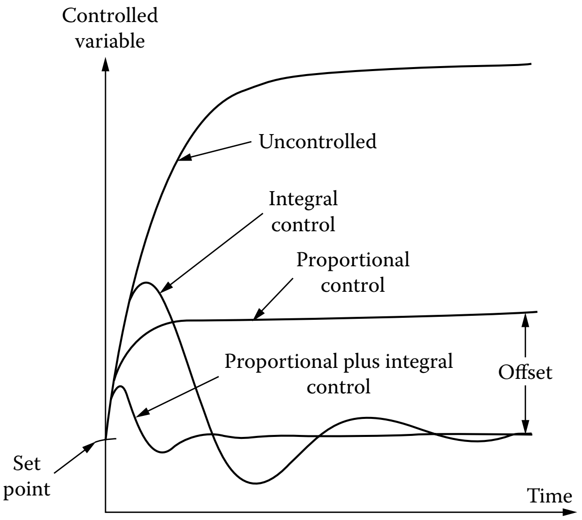


FIG. 2.2i
The different responses of a process variable to a load change in the process when it is uncontrolled and also when it is under P, I, and PI control.

integrates the area under the noisy error curve and gives a smooth average.

PI control is the most widely used control mode configuration and is used in all except the easiest applications, such as thermostats, and the most difficult applications, such as temperature or composition control. These difficult processes have large inertia that causes slow performance, and therefore the addition of rate (derivative) action accelerates the response of the control loop.

Figure 2.2i illustrates a process variable's response to a load change in the process when it is uncontrolled and also when it is under P, I, and PI control.

Reset Windup

It is often the case that the solving of one problem introduces another. While the introduction of the integral mode has solved the problem of offset, it has introduced another worry that has to do with the very nature of the integral mode. The fact that integral looks at the past history of the error and integrates it is good and useful when the loop is operational, but it can be a problem when the loop is idle.

When the plant is shut down for the night, the controllers do not need to integrate the errors under the error curves. Similarly, it would be desirable to "turn off" the integral mode in a surge controller when there is no surge or in a reactor temperature controller during heat-up. Other cases where integration of the error is undesirable include a selective control system, when a particular controller is not selected for control, or in a cascade master, if the operator has switched the loop off cascade, etc.

Under these conditions it is not desirable for the integral mode to stay active because if it does, it will eventually saturate, and its output will either drop to zero or rise to the maximum value of the air or power supply. Once saturated, the controller will not be ready to take control when called upon to do so but will actually upset the process by trying to introduce an equal and opposite area of error, which it has experienced during its idle state.

In all such installations the controller must be provided with either external reset, which protects it from ever becoming idle (see next section for details), or with an antireset windup feature, which protects it from saturating in its idle state. In selective and cascade control loops, external feedback is the most often applied solution. Here, instead of looking at its own output, which can be blocked, the integral mode of the controller looks at an external feedback signal (such as the opening of the valve), which cannot be blocked.

In surge control or reactor heat-up applications the chosen solution usually is to use the slave measurement as the external reset signal to prevent saturation. In some control systems, when a windup limit is reached, the integral (repeats/minute) is increased 8-, 16-, or even 32-fold, in order to speed the "unwinding" process and return the algorithm to normal operation. In DCS systems these functions are implemented in software.

THE DERIVATIVE MODE

The proportional mode considers the present state of the process error, and the integral mode looks at the past history of the error, while the derivative mode anticipates the future values of the error and acts on that prediction. This third control mode became necessary as the size of processing equipment increased and, correspondingly, the mass and the thermal inertia of such equipment also became greater. For such large processes it is not good enough to respond to an error when it has already evolved because the flywheel effect (the inertia or momentum) of these large processes makes it very difficult to stop or reverse a trend once it has evolved. The purpose of the derivative mode is to predict the process errors before they have evolved and take corrective action in advance.

Figure 2.2j describes the derivative response to the same error curve that has been used earlier. In the middle portion of the illustration where the error is constant, the derivative contribution to the output signal (to the control valve) is zero. This is because the derivative contribution given below by Equation 2.2(7) is based on the rate at which the error is changing, and in this region that rate of change is zero.

$$m = K_c\left[e + T_d \frac{d}{dt} e\right] + b \qquad \textbf{2.2(7)}$$

In a controller, derivative action is used not by itself, but in combination with the proportional (PD) or with the proportional and integral control modes (PID controller).

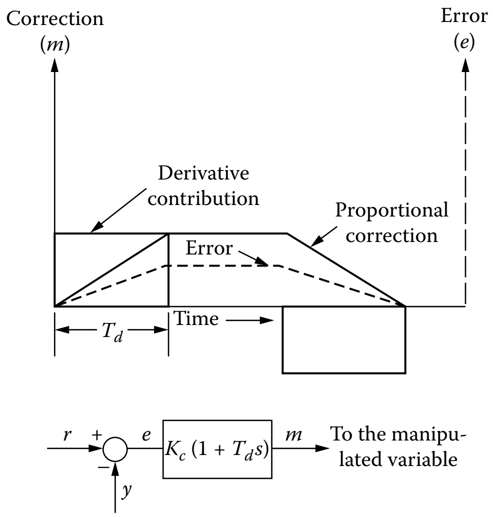


FIG. 2.2j
The contribution of the derivative mode to the total output signal (m) is a function of the rate at which the error is changing.

As illustrated in Figure 2.2j, when the error is rising (on the left of the figure), the derivative contribution is positive and is a function of the slope of the error curve.

The unit of the derivative setting is the derivative time (T_d). This is the length of time by which the D-mode (derivative mode) "looks into the future." In other words, if the derivative mode is set for a time T_d, it will generate a corrective action immediately when the error starts changing, and the size of that correction will equal the size of the correction that the proportional mode would have generated T_d time later. The longer the T_d setting, the further into the future the D-mode predicts and the larger its corrective contribution. When the slope of the error is positive (measurement is moving up relative to the set point), the derivative contribution will also rise, if the controller is direct-acting.

On the right side of Figure 2.2j, the error is still positive (measurement is above the set point), but the derivative contribution is already negative, as it is anticipating the future occurrence when the loop might overshoot in the negative direction and is correcting for that future occurrence now.

The derivative (or rate) setting is in units of time and usually can be adjusted from a few seconds to up to 10 hours or more. The applications for PD control loops (with proportional plus derivative effect) are few. They sometimes include the slave controller in temperature cascade systems, if the goal is to increase the sensitivity of the slave loop beyond what the maximum gain can provide.

Another application of PD control is batch neutralization, where the derivative mode protects against overshooting the target (pH = 7) while the P-mode reopens the reagent valve for a few droplets at a time as neutrality is approached.

PID control (with proportional plus integral plus derivative modes acting in) is more widely used, and its applications include most temperature and closed-loop composition control applications.

In most cases, when derivative action is used, the filtering of the measurement or error signal is also required, because otherwise it would respond to noise and to all abrupt input changes with excessively high changes in its output signal. The smaller the filter time constant is relative to the derivative time constant, the more effective will be the derivative mode.

Limitations of the Derivative Mode

Because the derivative mode acts on the rate at which the error signal changes, it can also cause unnecessary upsets because, for example, it will react to the sudden set point changes made by the operator. It will also amplify noise and will cause upsets when the measurement signal changes occur in steps (as in a chromatograph, for example).

In such situations special precautions are recommended. For example, it is necessary to change the control algorithm so that it will be insensitive to sudden set point changes. It is necessary to make sure that the derivative contribution to the output signal going to the control valve will respond not to them, but only to the rate at which the measurement changes. This change is aimed at making the derivative mode to act on the measurement only (Equation 2.2[8]), but *not* on the error (Equation 2.2[9]).

$$m = K_c\left[e + \frac{1}{T_i}\int e\,dt - T_d\frac{d}{dt}y\right] + b \qquad \textbf{2.2(8)}$$

$$m = K_c\left[e + \frac{1}{T_i}\int e\,dt + T_d\frac{d}{dt}e\right] + b \qquad \textbf{2.2(9)}$$

Some might prefer to eliminate the set point effect on the proportional contribution also. In that case, Equation 2.2(8) would be revised as follows:

$$m = K_c\left[-y + \frac{1}{T_i}\int e\,dt - T_d\frac{d}{dt}y\right] + b \qquad \textbf{2.2(10)}$$

Excessive noise and step changes in the measurement can also be corrected by filtering out any change in the measurement that occurs faster than the maximum speed of response of the process. DCS systems, as part of their software library, are provided with adjustable filters on each process variable.

Table 2.2k gives a summary of PID equations, while Figure 2.2l illustrates the responses to a step change in the set point of a pure P, a pure I, and a pure D control mode, and also gives the step responses of the PI, PD, and PID controllers.

The operation with a plain proportional controller (P) produces a large steady-state error. If an integral-only (I) controller is used, it continues to integrate the area under the error curve until the process variable is returned to set point. Adding the proportional mode to the integral mode provides a bit faster response. With a PD controller, the controller

TABLE 2.2k
Mathematical Descriptions of the Conventional Control Modes

Symbol	*Description*	*Mathematical Expression*
	One-Mode	
P	Proportional	$m = K_c e + b$
I	Integral (reset)	$m = \frac{1}{T_i}\int e\,dt + b$
	Two-Mode	
PI	Proportional-plus-integral	$m = K_c\left[e + \frac{1}{T_i}\int e\,dt\right] + b$
PD	Proportional-plus-derivative	$m = K_c\left[e + T_d\frac{d}{dt}e\right] + b$
	Three-Mode	
PID	Proportional-plus-integral-plus-derivative	$m = K_c\left[e + \frac{1}{T_i}\int e\,dt + T_d\frac{d}{dt}e\right] + b$

causes an initial overshoot and a steady-state error. A properly tuned PID controller reduces the initial overshoot and returns the controlled variable to set point.

Figure 2.2m separately describes the responses of the above controllers and control modes to a variety of disturbance inputs. Figure 2.2n shows the process responses of a closed loop with different control algorithms when they respond to a unit step change in set point. It can be seen that the integral mode returns the process to set point while the derivative amplifies transients.

Inverse Derivative Control Mode

A special-purpose control action used on extremely fast processes is the so-called inverse derivative mode. As the name implies, it is the exact opposite of the derivative mode. While the output of the derivative mode is directly proportional to the rate of change in the error, the output of the inverse derivative mode is inversely proportional to the rate of change in error.

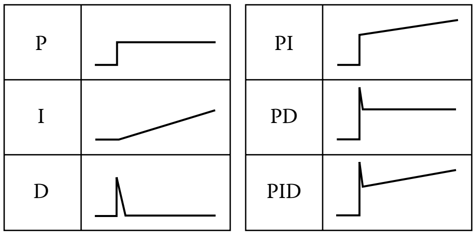


FIG. 2.2l
The responses to a step change in set point of the P, I, D control modes and of the PI, PD, and PID controllers.

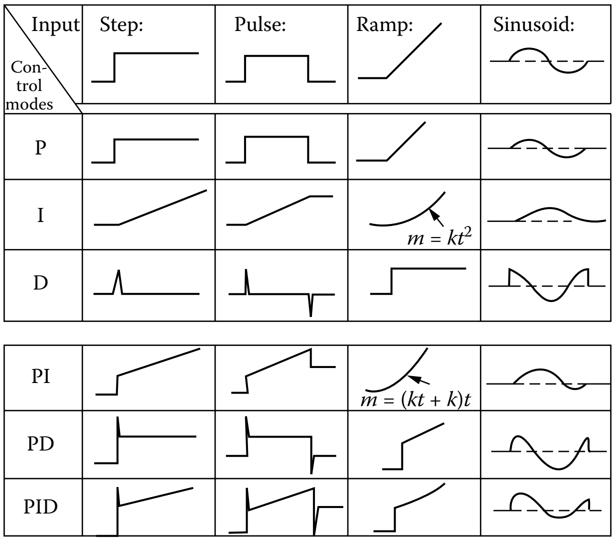


FIG. 2.2m
Responses of the P, I, D control modes and of the PI, PD, and PID controllers to a variety of input disturbances.

Inverse derivative is used to reduce the gain of a controller at high frequencies and is therefore useful in stabilizing a flow loop. The proportional-plus-inverse-derivative controller provides high gain to minimize offset at low frequency and low gain to stabilize at high frequency.

Inverse derivative can also be added to a proportional-plus-integral controller to stabilize flow and other loops requiring very low proportional gain for stability. One should add inverse derivatives only when the loop is unstable at the minimum gain setting of the proportional-plus-integral controller.

Since inverse derivative is available in a separate unit, it can be added to the loop when stability problems are encountered. Interestingly, the addition of the inverse derivative, if properly tuned, has little effect on the natural frequency of the loop.

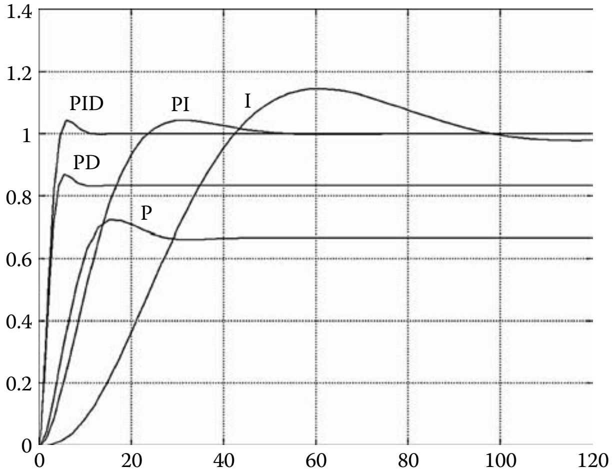


FIG. 2.2n
The unit-step responses of P, I, PI, PD, and PID controllers.

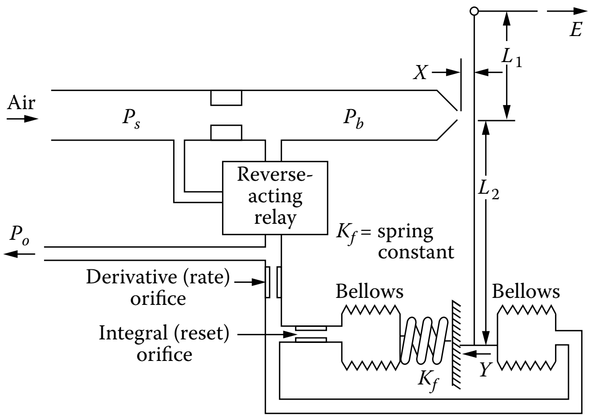


FIG. 2.2o
Interaction was unavoidable in the pneumatic PID controller because its components were physically interconnected, and a change in one tuning setting affected the others.

PID ALGORITHM VARIATIONS

Equations 2.2(8) and 2.2(9) are referred to as noninteracting PID equations. Such interacting behavior was unavoidable when pneumatic controllers were used because their three settings were physically interconnected (Figure 2.2o); if one setting changed, it affected all the others.

In these so-called interacting controllers, if we designate the PID settings on the controller knobs as PB_o (proportional band), T_{io} (minutes/repeat), and T_{do}, it can be shown that the actual "working" settings (PB, T_i, and T_d) that the process "sees" are as follows:

$$PB = PB_o/(1 + T_{do}/T_{io}) \qquad \mathbf{2.2(11)}$$

$$T_i = T_{do} + T_{io} \qquad \mathbf{2.2(12)}$$

$$T_d = (T_{do})(T_{io})/(T_{do} + T_{io}) \qquad \mathbf{2.2(13)}$$

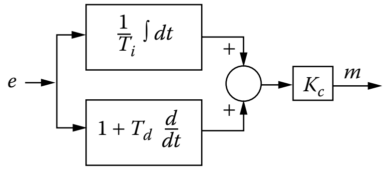
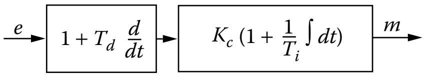
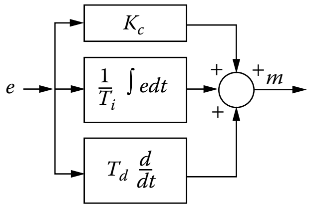
While interaction is somewhat confusing, the interacting controller has a major advantage: the actual working derivative (T_d) can never become more than one-quarter of the T_i. This feature contributes to the safety of using this controller because if it were possible to make $T_d > T_i/4$, the controller action would reverse, which could cause accidents.

Nowadays, the different DCS and PLC suppliers use a variety of PID control algorithms. Table 2.2p illustrates the working principles of the three most widely used versions, the interacting, the noninteracting, and the parallel algorithms. It should be understood that on the one hand all of these are usable, but on the other hand their settings will mean different things, and the noninteracting and parallel algorithms do need software protection against an operator accidentally setting T_d to exceed $T_i/4$.

Digital Algorithms

When PID algorithms are implemented in the digital world by DCS systems, what used to be integration in the analog world becomes summation, and what used to be time differential in the analog world becomes difference. The scan period of DCS systems is fixed at around 0.5 seconds or is

TABLE 2.2p
Conversions and Relationships Between the Different PID Algorithms

Types of Algorithms	*Standard or Noninteracting*	*Interacting*	*Parallel*
Equation for Output (m) =	$K_c\left(e+\frac{1}{T_i}\int e\,dt+T_d\frac{de}{dt}\right)+b$	$K_c\left(e+\frac{1}{T_i}\int e\,dt+T_d\frac{d}{dt}e\right)+b$	$K_c e+\frac{1}{T_i}\int edt+T_d\frac{de}{dt}+b$
Block Diagram Representation	 	 	
Working PB (100/K_c)	$PB = PB_0$	$PB=\frac{PB_0}{1+T_{do}/T_{io}}$	$PB = PB_0$
Working T_i (minutes/repeat)	$T_i = T_{io}$	$T_i = T_{do} + T_{io}$	$T_i = (T_{io})(K_c)$
Working T_d (minutes)	$T_d = T_{do}$	$T_d=\frac{(T_{do})(T_{io})}{T_{io}+T_{do}}$	$T_d = T_{do}/K_c$

Where: PB_o, T_{io}, and P_{do} are the proportional, integral, and derivative settings on the controller.

selectable for each loop from 0.1, 0.2, 0.5, 1.0, 5.0, 10, or 30 seconds.

The DCS system does not continuously evaluate the status of each measurement, but looks at them intermittently. This increases the dead time of the loop by two “scan periods” (the period between readings of a measurement) and bases its algorithm calculations on the “present” and the “previous” error.

There are two basic types of digital algorithms in use. One is called positional. This means that the full output signal to the valve is recalculated every time the measurement is looked at (Equation 2.2[14]).

$$m = K_c\left(e + \frac{1}{T_i}\sum_{o}^{n} e\,dt + \frac{T_d}{\Delta t}\Delta e\right) + b \qquad \textbf{2.2(14)}$$

where Δt = scan period; Δe = change between the previous and the present value of the error; and $\sum_0^n e$ = the sum of all previous errors between time zero and time n.

The other type of calculation is called a velocity algorithm. In that case, the value of the previous output signal (m) is held in memory, and only the required change in that output signal is calculated (Equation 2.2[15]).

$$\Delta m = K_c\left(\Delta e + \frac{1}{T_i}e\Delta t + \frac{T_d}{\Delta t}\Delta(\Delta e)\right) \qquad \textbf{2.2(15)}$$

where $\Delta(\Delta e)$ = the change in the change in the error between the previous and the present scan period.

The positional algorithm is preferred when the measurement is noisy because it works with the error and not the rate of error change when calculating its proportional correction. Velocity algorithms have the advantages of providing bumpless transfer, less reset windup (0 or 100%) and are better suited for controlling servomotor-driven devices. Their main limitations include noise sensitivity, likelihood to oscillate, and lack of an internal reference.

SAMPLE-AND-HOLD ALGORITHMS

When the dead time of a control loop exceeds the time-constant of that loop, the conventional PID algorithms cannot provide acceptable control. This is because the controller cannot distinguish between a nonresponding manipulated variable and one that is responding but whose effect cannot yet be detected because of the transportation lag (dead time) in the system.

This lack of response to a change in output during the dead-time period causes the controller to overreact and makes the loop unstable. For such applications the sample-and-hold type algorithms are used.

These algorithms are identical to the previously discussed PID algorithms except that they activate the PID algorithm for *only part of the time.* After the output signal (m) is changed to a new quantity, it is sealed at its last value (by setting the measurement of the controller equal to its set point) and it is held at that constant value until the dead time of the loop is exhausted.

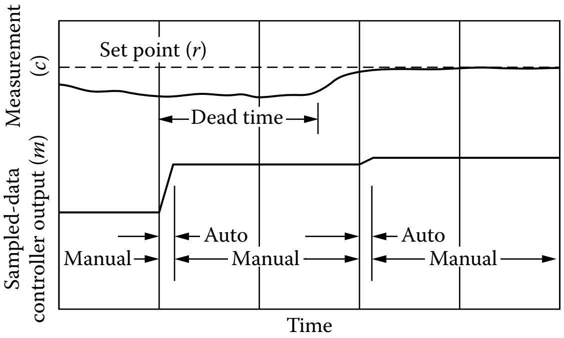


FIG. 2.2q
Sample-and-hold controllers are periodically switched from manual to automatic and back again. They are used mostly on dead time processes.

When the dead time has passed, the controller is switched back to automatic and its output (m) is adjusted based on the new measurement it “sees” at that time, using the PID algorithm. Figure 2.2q shows how these periods of manual and automatic operating modes are alternated. A timer sets the time period for which the controller is switched to manual. The timer setting is adjusted to exceed the dead time of the loop.

The only difference between conventional PID and its sample-and-hold variety is in the tuning of the loops: the sample-and-hold controller has less time to make the same amount of correction that the conventional PID would and therefore needs to do it faster. This means that the integral setting must be increased in proportion to the reduction in time when the loop is in automatic.

CONCLUSIONS

In this section the basic control modes and control algorithms were described. The next sections of this chapter will cover some of the more sophisticated analog and digital PID algorithm variations.

In summation one might say that on/off and plain proportional control is suited only to the easiest (fast with little dead time) processes, which do not need to be kept on a set point. The addition of integral action returns the controller to set point, but this PI controller must be protected from reset windup, usually by external reset. The addition of derivative provides anticipation into the future, but requires protection from noise and sudden set point changes.

Inverse-derivative is used to stabilize fast and noisy processes. Sample-and-hold algorithms are required if the dead time (transportation lag) of the process is long. Digital algorithms are discussed in more detail later, and it should suffice here to say that positional ones are better suited to the control of noisy processes, while velocity algorithms are preferred for the bumpless control of servomotor-driven devices.

Bibliography

Åström, K. J., and Hägglund, T., *PID Controllers: Theory, Design and Tuning,* Research Triangle Park, NC: ISA, 1995.

Butterworths, W. R., *Advanced Process Control,* New York: McGraw-Hill, 1981.

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: ISA, 1990.

Dorf, R. C., and Bishop, R. H., *Modern Control Systems,* 7th ed., Reading, MA: Addison-Wesley, 1995.

Eckman, D. P., *Automatic Process Control,* New York: John Wiley & Sons, 1958.

Franklin, G. F., Powell, J. D., and Emami-Naeini, A., *Feedback Control of Dynamic Systems,* 3rd ed., Reading, MA: Addison-Wesley, 1994.

Golten, J., and Werver, A., *Control System Design and Simulation,* New York: McGraw-Hill, 1991.

Goodwin, G. C., Graebe, S. F., and Salgado, M. E., *Control System Design,* Englewood Cliffs, NJ: Prentice Hall, 2001.

Hougen, J. O., *Measurements and Control Applications for Practicing Engineers,* Boston: Cahners Books, 1972.

Jones, B. E., *Instrumentation, Measurement and Feedback,* New York: McGraw-Hill, 1977.

Korn, G., *Interactive System Simulation,* New York: McGraw-Hill, 1989.

Kuo, B. C., *Automatic Control Systems,* New York: John Wiley & Sons, 2003.

Levine, W. S., (Ed.), *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Luyben, W. L., and Luyben, M. L., *Essentials of Process Control,* New York: McGraw-Hill, 1997.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* 4th ed., Englewood Cliffs, NJ: Prentice Hall, 1999.

Randhawd, R. S., "Understanding the Lazy Integral," *Instruments and Control Systems,* January 1983.

Rohrs, C. E., Melsa, J. L., and Schultz, D. G., *Linear Control Systems,* New York: McGraw-Hill, 1993.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* New York: John Wiley & Sons, 1989.

Skogestad, S., and Postlethwaite, I., *Multivariable Feedback Control,* New York: John Wiley & Sons, 1996.

Shinskey, F. G., *Process Control Systems: Application, Design and Tuning,* 4th ed., New York: McGraw-Hill, 1998.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

2.3 Control Modes—PID Variations

R. R. RHINEHART (1995, 2005) **F. G. SHINSKEY** (2005) **H. L. WADE** (2005)

Although the PID form represented by Equation 2.3(1) is presented as the standard form, many controller manufacturers offer a variety of modified versions. Some modifications are improvements; some are leftovers from early pneumatic implementations, and some are more traditional in certain industries than others. Regardless, the basic proportional, integral, and derivative actions are retained, so there are no conceptual barriers to using the various vendor offerings. The main issue is that the tuning behavior varies from one form to another.

$$m = K_C\left(e + \frac{1}{T_I}\int e\,dt + T_D\frac{de}{dt}\right) \qquad \textbf{2.3(1)}$$

An experienced operator who is accustomed to tuning a controller having a particular PID algorithm will be baffled when another controller does not respond as expected. An operator or software program that follows a recipe procedure to determine K_C, T_I, and T_D for the standard algorithm (Equation 2.3[1]) can be startled by the response when applying the procedure to a manufacturer-specific version. A manufacturer's equipment usually offers a number of alternative controller forms. An unknowledgeable user may choose some of them on the whim of the moment, with the results that a variety of algorithms may exist in controllers mounted side by side or in the same control system. When various algorithms are used in the same control room, the operators will not experience consistent behavior patterns, and as a result may never become competent at tuning. A controller that "cannot be tuned" will be placed in manual.

Controller tuning is an ongoing necessity. It is needed whenever process operating conditions change, when process gains and time constants change, and often when previous controller settings no longer give the desired response. Tuning is required to maintain good control, and a rational understanding of the controller form and its behavior is required for effective tuning.

An understanding of the various forms of control algorithms, plus the configuration options that are offered, is also necessary to properly design and apply process control strategies.

PID ALGORITHMS

PID algorithms that continuously (in analog control systems) or repetitively (in digital processor systems) calculate the required position of the valve or other final actuator are called position algorithms. On the other hand, algorithms that calculate the required change in position of the final actuator are called velocity, or incremental, algorithms. Velocity algorithms, which are limited to digital implementation, are covered in Section 2.4. There are three basic forms of the position PID algorithm, plus variations on each of these forms.

Noninteracting Form

The standard form is shown by the block diagram in Figure 2.3a and is given by Equation 2.3(1), where for a reverse acting controller, $e = r - c$, the difference between set point and measurement. (For a direct acting controller, $e = c - r$.) The units of both error e and output m are percentages of full-scale transmitter and controller outputs, respectively, and time is usually expressed in minutes. (Some manufacturers are beginning to express time in seconds, rather than minutes.) Accordingly, K_C is dimensionless, the units of T_I are minutes per repeat, and the units of T_D are minutes. In this form the controller gain attenuates each of the controller modes.

The integration is from the last time the controller was switched from manual to automatic until the present time. The integrator must be initialized, however, to prevent a "bump" in the controller output when the mode is switched

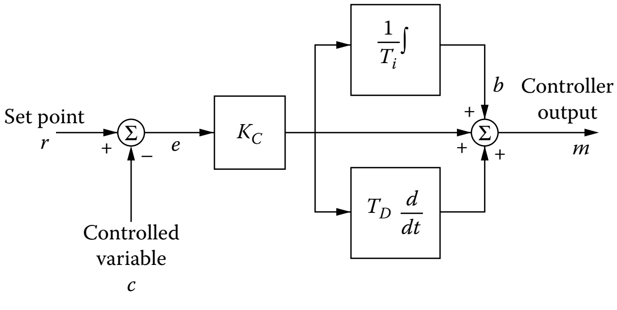


FIG. 2.3a
Block diagram, noninteracting PID controller.

from manual to automatic. If "*b*" represents the integral mode contribution to the controller output, that is,

$$b = \frac{K_C}{T_I}\int e\,dt \qquad \mathbf{2.3(2)}$$

then b_0, the initial value of b, must be set to

$$b_0 = m^* - \left(K_C e + T_D \frac{de}{dt}\right) \qquad \mathbf{2.3(3)}$$

where m^* is the operator-entered value of the controller output in the manual mode. Normally, b_0 tracks the right-hand side of Equation 2.2(3) continuously when the controller is in manual.

This form is often called the "ISA" form, although ISA has never sanctioned this or any other form of the PID. For reasons to be seen, it is also called the "noninteracting" form to distinguish it from the "interacting" form to be presented later. It is also called the "parallel" form, to distinguish it from the so-called "series" form, also to be presented later. Since there is no standard terminology among manufacturers, the user is cautioned to examine the mathematical expression representing the particular form, rather than relying on the manufacturer's nomenclature.

The proportional mode tuning parameter may be expressed as proportional band, PB, rather than controller gain, K_C. Proportional band is defined as the change in measurement required to produce a full range (100%) change in controller output due to proportional control action. The equations for conversion between PB and K_C are

$$K_C = \frac{100}{\text{PB}} \qquad \mathbf{2.3(4)}$$

$$\text{PB} = \frac{100}{K_C} \qquad \mathbf{2.3(5)}$$

The integral (reset) mode tuning parameter is often expressed as the reset rate, T_R, rather than as integral or reset time, T_I. The units of T_R are repeats/minute. The equations for conversion between reset time and reset rate are

$$T_R = \frac{1}{T_I} \qquad \mathbf{2.3(6)}$$

$$T_I = \frac{1}{T_R} \qquad \mathbf{2.3(7)}$$

Although there is no technological advantage in using either K_C or PB for the proportional mode tuning, or T_I or T_R for the integral mode tuning, the use of T_R has the psychological advantage that if faster integral action is desired, a larger number is set for T_R.

If both proportional and integral mode tuning parameters are changed, then Equation 2.3(1) becomes

$$m = \frac{100}{\text{PB}}\left(e + T_R\int e\,dt + T_D\frac{de}{dt}\right) \qquad \mathbf{2.3(8)}$$

Interacting Form

The interacting controller form is also called a "rate-before-reset" or "series" form algorithm. The "series" nomenclature arises from block diagram notation, in which the integral block is in series with the derivative block, whereas in the standard controller, the integral and derivative blocks are in parallel. The "parallel" nomenclature is also used in the next topic to describe an algorithm with independent gains on each mode. Therefore the "series/parallel" nomenclature can be misleading.

In the interactive controller, shown by the block diagram in Figure 2.3b and by

$$m = K'_C\left(e + T'_D\frac{de}{dt} + \frac{1}{T'_I}\int\left(e + T'_D\frac{de}{dt}\right)dt\right), \qquad \mathbf{2.3(9)}$$

integral action is taken on both the error and derivative terms. There is no technological advantage of Equation 2.3(9) over Equation 2.3(1), since with corresponding tuning, the performance of each is the same. Furthermore, if the derivative is not used ($T'_D = 0$), then the interacting and noninteracting controllers are identical in both form and performance. The interacting controller represents a physically realizable construction

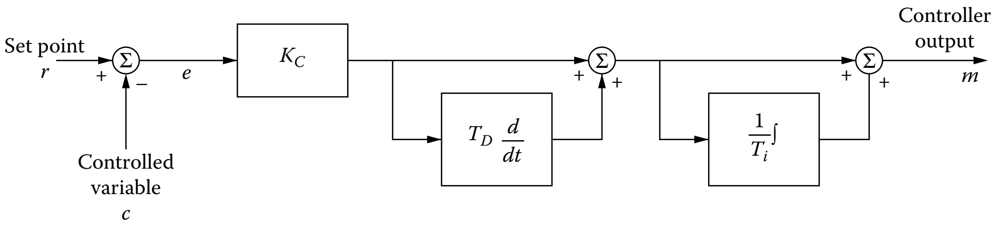


FIG. 2.3b
Block diagram, interacting PID controller.

of pneumatic devices, and is offered by vendors of electronic analog and digital systems so that new controllers will have a tuning response similar to certain pneumatic equipment that is being replaced.

Rearrangement of Equation 2.3(9) gives

$$m = K_C' \frac{T_I' + T_D'}{T_I'} \left(e + \frac{1}{T_I' + T_D'} \int e\,dt + \frac{T_I' T_D'}{T_I' + T_D'} \frac{de}{dt} \right) \tag{2.3(10)}$$

which shows exactly the same functional form as Equation 2.3(1). It also shows interactive behavior because adjustment of either T_I' or T_D' will affect the effective controller gain and both the effective integral and derivative times. K_C', T_I', and T_D' represent values actually entered for the noninteracting controller. The effective values of gain, integral, and derivative settings for an identical noninteracting, or standard, controller are given by

$$K_C = K_C' \frac{T_I' + T_D'}{T_I'} \tag{2.3(11)}$$

$$T_I = T_I' + T_D' \tag{2.3(12)}$$

$$T_D = \frac{T_I' T_D'}{T_I' + T_D'} \tag{2.3(13)}$$

If one has tuning parameters for a standard, or noninteracting, controller, Equation 2.3(1), then the corresponding tuning for Equation 2.3(10) is

$$K_C' = 0.5\,K_C\,(1 + \sqrt{1 - 4\,T_D/T_I}) \tag{2.3(14)}$$

$$T_I' = 0.5\,T_I\,(1 + \sqrt{1 - 4\,T_D/T_I}) \tag{2.3(15)}$$

$$T_D' = 0.5\,T_D\,(1 - \sqrt{1 - 4\,T_D/T_I}) \tag{2.3(16)}$$

Many conventional tuning rules define $T_D = T_I/4$. In this case, the corresponding tuning for the interacting controller is $K_C' = K_C/2$ and $T_I' = T_D' = T_I/2$. The interacting controller cannot duplicate a standard controller if $T_D > T_I/4$.

Noisy process signals should be smoothed or averaged before being applied to the control algorithm. A first-order filter (exponentially weighted moving average [EWMA]) is the standard offering. If the filter is used within the algorithm itself, it may be applied to the error signal, so that the filtered value of the error, e_f, replaces e in Equations 2.3(1) and 2.3(9). The filter equation is

$$T \frac{de_f}{dt} = e - e_f \tag{2.3(17)}$$

where T is the time constant of the filter. Since process noise is especially amplified by derivative action, some vendors make the filter time constant some multiple of the derivative time, T_D, so that

$$T = \alpha T_D \tag{2.3(18)}$$

Typical values for α range from 0.05 to 0.1. In pneumatic and electronic analog systems, α had a fixed value; in digital systems, α may or may not be accessible to the user. Some manufacturers refer to the reciprocal of α as “derivative gain.” Since filtering in essence averages new data with old data, the filtered value lags behind the process. The filtered variable dynamics are slower than the process, which means that (1) the controller must be tuned using the filtered data and (2) if the filter time constant is changed the controller may have to be retuned.

By combining Equations 2.3(17) and 2.3(9) and expressing the relationship in Laplace notation, the transfer function from error to controller output for an interacting controller with filter is

$$\frac{m(s)}{e(s)} = K_C \frac{(T_D s + 1)(T_I S + 1)}{(\alpha T_D s + 1) T_I S} \tag{2.3(19)}$$

Except for a modification that makes the derivative mode sensitive only to measurement changes, this transfer function is representative of most analog controllers and many implementations in digital systems.

Parallel Form

Whereas either the interacting or noninteracting forms of the PID are widely used in the process industries, the parallel form, also called the “independent gains” form, of PID is often used in power generation, drive control systems, robotics, and other applications. Figure 2.3c and Equation 2.3(20) presents the independent gains form of PID.

$$m = K_P\,e + K_I \int e\,dt + K_D \frac{de}{dt} \tag{2.3(20)}$$

K_P, K_I, and K_D are called the proportional gain, integral gain, and derivative gain, respectively. K_P is a dimensionless number, K_I is in units of time^{-1}, and K_D is in units of time.

There is no technological advantage of the parallel form of PID, since with corresponding tuning, the behavior of this form and either of the other forms will be identical. If one has tuning

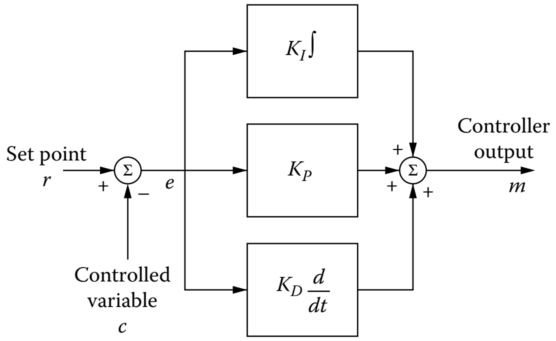


FIG. 2.3c
Block diagram, parallel PID controller.

parameters for a noninteracting controller, Equation 2.3(1), then the corresponding tuning parameters for Equation 2.3(20) are

$$K_P = K_C \tag{2.3(21)}$$

$$K_I = \frac{K_C}{T_I} \tag{2.3(22)}$$

$$K_D = K_C T_D \tag{2.3(23)}$$

A slight advantage of this form of configuration is that an integral-only controller can be obtained by simply setting K_C and K_D to zero. That cannot be done with either of the other forms if the integral tuning parameter is T_I, minutes/repeat.

SET-POINT RESPONSE SOFTENING

The algorithms presented previously may cause undesirable process effects whenever a set-point change is made. Each term acts on the error. Because the error is a deviation from set point, a sudden set-point change becomes a sudden change in error, which in turn causes an output spike — often an undesirable feature. Even if derivative is not being used, a set-point change will cause a "bump" in controller output; the size of this bump will be the magnitude of the set-point change times the controller gain. Another problem with set-point changes has to do with controller tuning. If a controller is tuned for optimum response to a set-point change, it will likely be too conservative for elimination of disturbances to the loop. The rate of recovery of the controlled variable will be much slower than its rate of departure. On the other hand, controller settings that produce optimum disturbance rejection tend to cause large overshoot on set-point changes. This is especially true of lag-dominant loops such as most liquid level, temperature, pressure, and composition loops. To avoid this undesirable behavior, several modifications that can soften the response to a set-point change can be used. In general, these modifications can be applied to any of the three forms of PID algorithm presented earlier.

Derivative and Proportional Actions

The algorithm can be modified so that either the derivative mode, or both the derivative and proportional modes, act only on the measurement. Output spiking due to the derivative mode and output bumping due to the proportional response are prevented, while the tuned controller response to a loop disturbance is unchanged. Derivative-on-error or -measurement and proportional-on-error or -measurement are often user-configuration choices. If both are chosen, the form of a noninteracting algorithm becomes

$$m = K_C\left(-c + \frac{1}{T_I}\int e\,dt - T_D\frac{dc}{dt}\right) \tag{2.3(24)}$$

Because of the elimination of output spikes and output bumps, the controller can be tuned for a more aggressive response to disturbances. Conventional tuning rules (e.g., Ziegler–Nichols) will produce a very sluggish response to set-point change if proportional-on-measurement is chosen. The response to a measurement disturbance will be the same if the proportional acts on the error or on measurement.

Set-Point Filtering

With proportional-on-measurement there may be a serious degradation in the response to a set-point change, since immediately following the set-point change, only the integral mode will act on the deviation. Since most modern PID controllers feature bumpless transfer between the manual and automatic modes, the response is the same as if the set point had been changed with the controller in manual, with the controller then switched to automatic.

To achieve the advantage of improved disturbance rejection due to more aggressive tuning of the PID yet avoid the degraded response following a set-point change, a lead-lag filter can be placed on the set point.

$$r'(s) = \left(\frac{\lambda T_I s + 1}{T_I s + 1}\right) r(s) \tag{2.3(25)}$$

Here, $r(s)$ is the (transform of) the actual set point; $r'(s)$ is the modified set point used by the PID algorithm; T_I is the integral time used by the PID, and λ is the lead-lag ratio. When λ is set to 1.0, there is no filtering; when set to zero, the filter is a first-order lag whose time constant is equal to the integral time. When Equation 2.3(25) is combined with the transform of Equation 2.3(1), using only the proportional and integral modes, the result is

$$m(s) = K_C\left(\lambda r(s) - c(s) + \frac{r(s) - c(s)}{T_I s}\right) \tag{2.3(26)}$$

Another way of viewing this modification is to consider taking a linear combination of proportional-on-error and proportional-on-measurement. From Equation 2.3(1), the proportional mode contribution to the output is $K_c(r - c)$. From Equation 2.3(24), the proportional mode contribution to the output is $-K_c c$. A linear combination of these becomes

$$\lambda K_C(r - c) - (1 - \lambda)K_C c = K_C(\lambda r - c) \tag{2.3(27)}$$

which, after transformation, is the same as the proportional mode contribution exhibited by Equation 2.3(26). A value for λ of 0.4 produces almost no set-point overshoot, whereas a value of 0.5 will produce a minimum integral of absolute error (IAE) response with about 10% set-point overshoot. These results will vary somewhat with the dynamics of the process and the tuning of the controller.

One manufacturer offers a "two-degree-of-freedom" controller in which linear combinations of both the proportional modes and derivative modes can be chosen.

$$m = K_C\left[(\lambda r - c) + \frac{1}{T_I}\int e\,dt + TD\frac{d(\delta r - c)}{dt}\right] \tag{2.3(28)}$$

Some manufacturers offer a set-point ramping option that prevents both output spikes and bumps by eliminating sudden set point changes.

Although any of the set-point softening techniques described here can be an advantage for single loop control or for the primary controller of a cascade configuration, neither set-point ramping nor proportional-on-measurement should be used on the secondary controller of a cascade loop, as that would degrade the responsiveness of the secondary loop.

WINDUP ACCOMMODATION

The integral is an accumulator that, in digital algorithms, continuously adds the deviation from set point to the previous sum. Whenever an error persists for a long time, the integral can grow to the largest value the assigned processor memory can hold. For analog systems, the integral can grow to the maximum signal value, usually beyond the limits of 0 or 100%. This is called integral or reset windup. Since the controller output is the sum of the three controller terms, the output will be dominated by the wound-up integral mode and will "max out" with the control valve either fully open or fully closed. When the error is subsequently removed or even changes sign, the integral windup continues to maintain the output saturated for a long time. This control action will persist until the negative error accumulation equals the previously accumulated positive error and thus permits the integral mode to "unwind."

Whenever a controller is not in control, the integral mode can wind up unless other provisions are made. Such situations include the controller being in the manual mode, a constraint being encountered (such as a valve stem at a limit), a primary controller of a cascade strategy being in automatic with its secondary in manual or local automatic, or the use of an override control strategy. Also, process conditions such as sensor failure, failure of control-loop communications, or a manual process bypass used for control or shutdown of a portion of the process can prevent a controller from being in control. Therefore it is desirable for every controller containing the integral mode to have antireset windup provisions. There are a variety of offerings.

Conceptually the simplest solution is to specify a limit on the integral contribution to the total output signal. The maximum and minimum reasonable values for the integral contribution correspond to the maximum and minimum controller outputs. For the standard controller:

$$I_{\max} = m_{\max}\ T_I/K_C \qquad \textbf{2.3(29)}$$

$$I_{\min} = m_{\min}\ T_I/K_C \qquad \textbf{2.3(30)}$$

Where integral limits are used, conventional practice uses 95% of the maximum:

$$I_{\text{U.L.}} = I_{\min} + .95(I_{\max} - I_{\min}) \qquad \textbf{2.3(31)}$$

$$I_{\text{L.L.}} = I_{\min} + .05(I_{\max} - I_{\min}) \qquad \textbf{2.3(32)}$$

However, the 95% value has been arbitrarily selected and can therefore be changed.

A closely associated strategy is to inhibit integral accumulation if the output hits a limit. With either of the mechanisms above, the integral windup will be limited to values corresponding to the fully open or fully closed valve positions. It is possible for the integral mode to wind up to fully open when the proper value should only be 30% open. In that case, the integral will keep the output excessively high, and until it unsaturates and has time to unwind, it will persist in introducing a process error in the opposite direction. To accelerate the return of the integral contribution to its proper value, some controllers will make the integral accumulation 16 times faster when coming off a limited value. These mechanisms generally work, but 95% limits, 16-times-faster recovery, and allowing some windup are not always the best solutions.

For digital systems, if a controller can detect when it is unable to control, then it can go into an "initialization" mode that forces the integral mode output to a value that will permit graceful recovery whenever the controller is again able to control. Situations when a controller can detect that it is unable to control include being in the manual mode, having the valve stem reach a limit (if valve stem position is detected and transmitted back to the controller) or, if the controller is a primary controller of a cascade configuration, having the secondary controller in some mode other than fully automatic cascade. If *b* represents the integral mode contribution to the output, then in the initialization mode, *b* can be reset using Equation 2.3(3), where m^* represents the operator-entered output value in the manual mode, the detected valve stem position, or the set point of the secondary controller.

RESET AND EXTERNAL FEEDBACK

Whereas Equations 2.3(1), 2.3(8), and 2.3(20) are good conceptual models for PID controller forms, they may not be representative of physical implementation. In essentially all analog controllers and in many digital control algorithms, the integral of the error is replaced with a first-order filter on the controller output; this is part of a positive feedback loop that computes the controller output. When the time constant of the filter is the same as the integral time, T_I, then the performance of the controller with reset feedback is the same as if it had pure integral mode. For the standard PI controller:

$$m(s) = K_C\ e(s) + \frac{1}{T_I s + 1} m(s) \qquad \textbf{2.3(33)}$$

which can be rearranged to

$$m(s) = K_C\left(1 + \frac{1}{T_I\, s}\right) e(s) \qquad \textbf{2.3(34)}$$

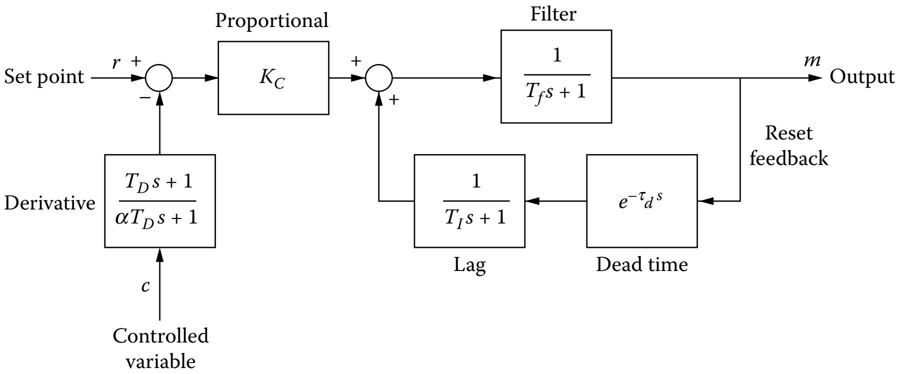


FIG. 2.3d
Block diagram, PIDτ_d controller.

Note that with reset feedback there is no integrator to wind up. The output of the filter, however, can wind up to the upper or lower limit of the controller output.

In addition to being a physically realizable method of construction, particularly for analog controllers, a key advantage of reset feedback is in override and cascade control structures, where the feedback is not directly from the controller output but from an external signal that would be identical to the controller output if the control loop were closed. With external reset feedback, the filter output “winds up” to the “right” value; when control is recovered there is no need for wind-down acceleration nor a need to guess at correct upper and lower integral limits.

PIDτ_d ALGORITHM

A relatively simple addition to a PID controller can significantly increase its performance, both to disturbances and to set-point changes. It involves delaying the integral mode of the controller in a manner similar to the delay or dead time that exists in the process being controlled. The structure of the PID-dead time (PIDτ_d) controller is not unlike that of model-based controllers—it may be considered a hybrid of PID and model-based technologies. A block diagram of the controller is shown in Fig. 2.3d. Without the dead time and filter blocks, this is a standard PID controller of the “interacting” or “series” type, with integration achieved by means of reset feedback as described earlier.

The PIDτ_d controller is capable of producing a phase lead, thereby substantially reducing the period of oscillation of a control loop. For lag-dominant processes, reducing the period of oscillation also reduces its dynamic gain, allowing a higher controller gain. This performance improvement comes with a price, however. Robustness is decreased, causing the loop to destabilize on rather small variations in process parameters unless the controller is precisely tuned. With deadtime-dominant processes, the dead time in the controller must match the process dead time. Any difference between process and controller dead times will produce a high-frequency oscillation with a period of twice their difference.

Robustness can be improved by the addition of the filter shown in Figure 2.3d. Its location lies in both feedback loops, making it doubly effective. However, performance is reduced by the ratio of filter time to dead time, so excessive filtering should also be avoided.

When the process contains lags as well as dead time, K_C can be increased further and derivative action added as well. In this role, it may be tuned much like a PID controller, with the dead time setting matched to the process dead time. But in the presence of noise, it may also be tuned as a PIτ_d controller, in which case optimum performance requires controller dead time to substantially exceed that of the process.

Bibliography

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: ISA, 1990.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Marlin, T. E., *Process Control—Designing Process and Control Systems for Dynamic Performance*, 2nd ed., New York: McGraw-Hill, 2000.

Ogunnaike, B. A., and Ray, W. H., *Process Dynamics, Modeling and Control,* New York: Oxford University Press, 1994.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* 2nd ed., New York: John Wiley & Sons, 2004.

Shinskey, F. G., “PID-Deadtime Control of Distributed Processes,” *Control Engineering Practice,* 9: 1177–1183, 2001.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

Wade, H. L., *Regulatory and Advanced Regulatory Control: System Design and Application,* 2nd ed., Research Triangle Park, NC: ISA, 2004.

2.4 Control Modes—Digital PID Controllers

H. L. WADE (2005)

In the days of analog instrumentation, the term "controller" had a definite connotation. It referred to a stand-alone piece of equipment that received a signal, either pneumatic (3 to 15 psi [20 to 100 KpA]) or electric (4 to 20 mA), performed some control computation using a pneumatic mechanism or electronic circuitry, and then produced an output signal of the same type as the input signal. The controller output was usually connected to a valve or other final actuator but was sometimes connected to another controller (another separately identifiable piece of equipment) as the set point for a "slave" control loop. The controller was probably mounted on a panel board and was self-contained in terms of providing its own HMI (human-machine interface), which consisted of indicating gauges and recording means, plus switches and knobs by which an operator entered a set-point value, switched between manual and automatic, and remotely set a desired valve position when intervention into the control loop was necessary. In essence, a controller was "something you could put your finger on."

In digital technology, the concept of "a controller" is probably a misnomer. Although stand-alone devices (single loop controllers) exist that perform control computation using their own microprocessors rather than a pneumatic mechanism or electronic analog circuitry, the usual situation is to have a multi-loop digital processor, such as a distributed control system (DCS) or a programmable logic controller (PLC). Here both the section of code performing the control computation and the provisions for human-machine interface (HMI) are shared by many loops. The only thing that is unique to a particular loop is a data record in memory that contains an identification of the loop, designation of the particular subroutine to use for the control calculation, pointers to other memory locations as sources for its input values, and memory locations for holding tuning parameters, manual-automatic status parameters, configuration options, and so on. In essence, the "controller" in digital technology exists as a concept only, not as a physical entity. Nevertheless, it is beneficial as well as the customary practice to think of a digital controller as if it were a specific device, performing a specified function, interacting with other controllers, providing HMI services upon request, and being represented on a piping and instrumentation drawing (P&ID) by a function block symbol that has been assigned a unique identifying tag. This is the conceptual model we will use when speaking of a digital controller.

The heart of a digital controller is the control algorithm; this is the section of code that performs the computation that mimics one of the PID forms described in the previous section. Because of the versatility of a digital processor, however, many functions that were impossible with analog controllers, such as automated bumpless transfer, wind-up prevention, and interblock status communications, can now be performed as a part of a digital controller.

PID POSITION ALGORITHMS

The noninteracting PID form was given by Equation 2.3(1) of Section 2.3 as:

$$m = K_C \left(e + \frac{1}{T_I} \int e\,dt + T_D \frac{de}{dt} \right) \qquad \textbf{2.4(1)}$$

We will develop an analogous digital algorithm that operates repetitively in a sampling environment. The sampling interval is ΔT (such as 1/60 minute); everything is related to the nth sampling instant.

Corresponding to the proportional mode term, e, is the error at the nth sampling instant, $e_n = r_n - c_n$. The term $\int e\,dt$ will be approximated by summing the errors at each sampling instant. If S_{n-1} represents the previously accumulated sum, then

$$S_n = S_{n-1} + e_n \qquad \textbf{2.4(2)}$$

and the integral mode term is:

$$\frac{\Delta T}{T_I} S_n$$

The derivative will be approximated by two-point differencing, so that the derivative mode term is:

$$\frac{T_D}{\Delta T}(e_n - e_{n-1})$$

Putting these all together, the digital algorithm corresponding to Equation 2.4(1) is:

$$m_n = K_C \left[e_n + \frac{\Delta T}{T_I} S_n + \frac{T_D}{\Delta T}(e_n - e_{n-1}) \right] \qquad \textbf{2.4(3)}$$

The terms K_C, T_I, and T_D represent the conventional tuning parameters: controller gain, reset time in minutes/repeat, and derivative time in minutes; ΔT represents the sampling interval.

A similar approach can be used for the "parallel" (independent gains) form of PID. (See Equation 2.3[20] in the previous section.) The error is summed using Equation 2.4(2), then the control output is computed from

$$m_n = K_C e_n + K_I S_n \Delta T + K_D \frac{e_n - e_{n-1}}{\Delta T} \qquad \textbf{2.4(4)}$$

where K_C, K_I, and K_D are the proportional gain, integral gain, and derivative gain, respectively.

Equations 2.4(3) and 2.4(4) are called "position" algorithms because they compute the required position of the final actuator, m_n, at the nth sampling instant.

The modifications mentioned in Section 2.3 may also be applied here. For instance, if both the proportional and derivative modes are on measurement, rather than error, Equation 2.4(3) becomes

$$m_n = K_C \left[-c_n + \frac{\Delta T}{T_I} S_n - \frac{T_D}{\Delta T}(c_n - c_{n-1}) \right] \qquad \textbf{2.4(5)}$$

To achieve bumpless transfer from manual to automatic, or from any detectable condition where the PID is unable to control, the stored value of the error sum, S_{n-1}, is initialized by the equation:

$$S_{n-1} = \frac{T_I}{\Delta T}\left[\frac{m^*}{K_C} - \left(\left(1 + \frac{\Delta T}{T_I}\right) e_n + \frac{T_D}{\Delta T}(e_n - e_{n-1}) \right) \right] \qquad \textbf{2.4(6)}$$

where m^* represents the operator-entered output value in manual, the set point of a cascaded secondary PID, or the actual valve stem position, if the actual valve stem position is detected. After application of Equations 2.4(2) and 2.4(3), this initialization of S_{n-1} causes the controller output, m_n, to equal m^* at the time of switching from manual to automatic; thus, no bump to the process is incurred. If either the proportional or derivative modes is on measurement, rather than error, then the appropriate *en* terms in 2.4(6) must be replaced with *cn* and appropriate sign changes made. Equation 2.4(6) can be invoked at the first calculation cycle following a manual- to-automatic switch; alternatively, it may be processed on each calculation cycle as long as the controller is in manual. In this case no special procedure for the "first calculation cycle" is required.

Although the concept of bumpless transfer is widely recognized, the concept of "bumpless tuning" is virtually unknown. Yet a little reflection will show the need. If the proportional mode is on error, and if controller gain is changed at a time when there is a deviation from set point, there will be a bump in the controller output. A change in the derivative setting at a time when the measurement is changing will also cause a bump in the controller output. To prevent this, whenever the controller detects that a controller tuning parameter, either K_C or T_D, has been changed (no adjustment is necessary if T_I is changed), then, before the next algorithm processing cycle, the stored value of the error sum must be "fixed up" so that the new tuning parameters will produce the same controller output as would have been produced by the old tuning parameters. Had there been no gain change, the controller output would have been calculated as

$$m_n = K_C^{\text{old}} \left[e_n + \frac{\Delta T}{T_I}\left(S_{n-1}^{\text{old}} + e_n\right) + \frac{T_D}{\Delta T}(e_n - e_{n-1}) \right] \qquad \textbf{2.4(7)}$$

If a new value has been entered for K_C and the value for S_{n-1} has been adjusted, the controller output will be calculated as:

$$m_n = K_C^{\text{new}} \left[e_n + \frac{\Delta T}{T_I}\left(S_{n-1}^{\text{new}} + e_n\right) + \frac{T_D}{\Delta T}(e_n - e_{n-1}) \right] \qquad \textbf{2.4(8)}$$

Since we want the controller output to remain the same in spite of the tuning parameter change, we equate the right-hand sides of Equations 2.4(7) and 2.4(8) and solve for S_{n-1}^{new} :

$$S_{n-1}^{\text{new}} = \frac{K_C^{\text{old}}}{K_C^{\text{new}}} S_{n-1}^{\text{old}} + \frac{T_I}{\Delta T}\left(\frac{K_C^{\text{old}}}{K_C^{\text{new}}} - 1 \right) \times \left[\left(1 + \frac{\Delta T}{T_I}\right) e_n + \frac{T_D}{\Delta T}(e_n - e_{n-1}) \right] \qquad \textbf{2.4(9)}$$

If a new value has been entered for T_D, Equation 2.4(10) can be used for updating S_{n-1}.

$$S_{n-1}^{\text{new}} = S_{n-1}^{\text{old}} + \frac{\left(T_D^{\text{old}} - T_D^{\text{new}}\right) T_I}{(\Delta T)^2}(e_n - e_{n-1}) \qquad \textbf{2.4(10)}$$

If the derivative or proportional mode is on measurement rather than error, then the values of e must be replaced by values of c and appropriate sign changes made.

Once a new value of S_{n-1} has been determined, use Equation 2.4(2) to update S_n and Equation 2.4(3) or 2.4(5) to calculate the next value for controller output.

With the interacting algorithm (see Equation 2.3[9] in Section 2.3), the integral mode can be created by using a first-order lag filter in a positive feedback loop. The time constant of the filter is the integral time, T_I. The input to the filter is called the "reset feedback." If the algorithm is being used as an ordinary PID controller in the automatic mode, the reset feedback is connected directly to the controller output. An important application is in override control strategies, where the reset feedback is not connected directly to the controller output; instead, it is connected to another source, such as the output of a high or low signal selector between this and other controller outputs. This configuration prevents windup in the nonselected controller.

The equations for a PID controller with reset feedback involve a two-step process. The filter output is calculated first from previous values of the filter output and the controller output itself.

$$b_n = \frac{\Delta T}{T_I} b_{n-1} + \left(1 - \frac{\Delta T}{T_I}\right) m_{n-1} \qquad \textbf{2.4(11)}$$

Then the new controller output is computed:

$$m_n = K_C\left(e_n + \frac{TD}{\Delta T}(e_n - e_{n-1})\right) + b_n \qquad \textbf{2.4(12)}$$

If the controller is used in an override strategy, m_{n-1} in Equation 2.4(11) should be replaced by the output of the high or low selector.

When the PID algorithm is formulated with reset feedback, the term b_{n-1} can be adjusted to achieve bumpless transfer or bumpless tuning, similar to the use of S_{n-1} in Equations 2.4(6), 2.4(9), and 2.6(10). For instance, at the time of manual-to-automatic switching, b_{n-1} is initialized by the equation:

$$b_{n-1} = m^* - K_C\left(e_n + \frac{T_D}{\Delta T}(e_n - e_{n-1})\right) \qquad \textbf{2.4(13)}$$

where m^* is the operator-entered value for the controller output.

PID VELOCITY ALGORITHMS

A capability unavailable in analog control systems is that of velocity algorithms. Previously discussed algorithms have been positional. They calculated m_n, the required position of the valve or other final actuator. Velocity algorithms, on the other hand, calculate Δm_n, the increment by which the position of the final actuator should change. ("Incremental" algorithms would be better terminology; however, the term "velocity" is in traditional usage.)

Velocity algorithms were developed during the time of supervisory control when a computer output set the set point of an analog controller. A stepping motor was often used to drive the analog controller set point. The controller output, Δm_n, was converted to a pulse train that incremented the stepping motor by the required amount. Even though that technology is rarely used today, velocity algorithms are retained and offer certain advantages as well as disadvantages.

To develop a velocity algorithm that is the counterpart of Equation 2.4(3), the equation will first be rewritten for the previous calculation cycle. That is, the subscript "n" will be replaced by "$n - 1$" wherever it appears:

$$m_{n-1} = K_C\left[e_{n-1} + \frac{\Delta T}{T_I}S_{n-1} + \frac{T_D}{\Delta T}(e_{n-1} - e_{n-2})\right] \qquad \textbf{2.4(14)}$$

The required incremental change is given by $\Delta m_n = m_n - m_{n-1}$. Hence, to develop an equation for Δm_n, the right-hand side of Equation 2.4(14) will be subtracted from the right-hand side of Equation 2.4(3). In doing so, the term $S_n - S_{n-1}$ is encountered. By Equation 2.4(2), however, this is simply equal to e_n. Hence the final results are

$$\Delta m_n = K_C\left[e_n - e_{n-1} + \frac{\Delta T}{T_I}e_n + \frac{T_D}{\Delta T}(e_n - 2e_{n-1} + e_{n-2})\right] \qquad \textbf{2.4(15)}$$

In digital control systems that use velocity algorithms today, the previous required position of the final actuator, m_{n-1}, is stored in a separate memory location. After the velocity algorithm computes the required incremental change, this increment is added to the previous position, and the new required position overwrites the previous value in the memory location. The new value is then converted to an analog value, such as 4–20 mA, for transmission to the final actuator. This operation is described by Equation 2.4(16),

$$m_n \leftarrow m_{n-1} + \Delta m_n \qquad \textbf{2.4(16)}$$

where " ← " means "replaces."

An examination of Equation 2.4(15) reveals that it contains no integrator that is subject to windup. Therefore, a purported advantage of the velocity algorithm is that "it doesn't wind up." Whereas that is true of the algorithm itself, it does not mean that there cannot be windup in the overall control system. Suppose there is a sustained error, the situation that causes windup in a position algorithm. Each time the velocity algorithm is processed, it calculates an output change, Δm_n. These incremental changes accumulate in the output memory location due to Equation 2.4(16), so the windup occurs in the memory location, not in the algorithm itself. If the incremental change were applied to an external stepping motor, then the stepping motor would wind up. Therefore a better description of the velocity algorithm is "it moves the windup to someplace else."

Some true advantages can be attributed to the velocity algorithm. Bumpless transfer between the manual and automatic modes is simplified. When the algorithm is switched from automatic to manual, it simply stops calculating an incremental change. The output memory location is then freed for the operator to manually enter a required position of the final actuator. When the controller is returned to automatic, the algorithm resumes by calculating an incremental change, which is added to the last position set by the operator; this becomes the new required valve position:

$$m_n \leftarrow m^* + \Delta m_n \qquad \textbf{2.4(17)}$$

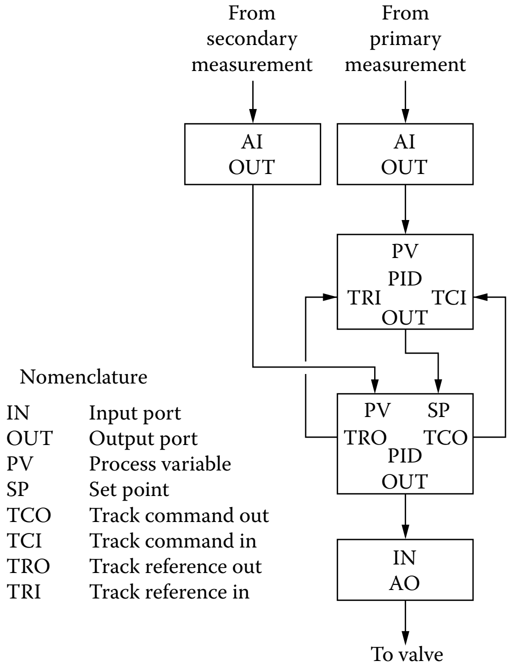
The problem of "bumpless tuning" is nonexistent. The velocity algorithm is inherently "bump free" when tuning parameters are changed.

Some true disadvantages of the velocity algorithm also exist. Reset feedback cannot be implemented with the velocity algorithm. Thus override control schemes that use reset feedback to prevent windup of the nonselected controller are not possible. Control systems that do use the velocity algorithm in override control strategies initialize the output memory location of the nonselected controller by the equation:

$$m = m^* + K_C\left(e_n + \frac{T_D}{\Delta T}(e_n - e_{n-1})\right) \qquad \mathbf{2.4(18)}$$

where, in this equation, m^* represents the output of the selected controller.

When operating in the automatic mode, and with a sampling rate that is relatively fast compared to the process dynamics (say, at least ten samples per dominant time constant), then there appears to be no performance advantage for either the velocity or position algorithms. With the DCS and PLCs in use today, sampling rate is rarely a problem. From the software developer's point of view, one form may have an advantage over the other due to the ease of developing the features to be included in a particular manufacturer's system. From an application developer's point of view, even though there may be no performance advantage to either, it is important to know which form is being used and the characteristics of that particular form.



Nomenclature

IN	Input port
OUT	Output port
PV	Process variable
SP	Set point
TCO	Track command out
TCI	Track command in
TRO	Track reference out
TRI	Track reference in

FIG. 2.4a
Function block structure for cascade control.

CONTROL STRUCTURES

Control strategies in digital systems are implemented by an organization of function blocks available from the manufacturer's function block library. Signals from one block to another are designated either graphically or in questionnaire format; other configuration choices are given in response to a questionnaire.

Using a cascade control strategy as an example, the signals that must be passed are:

- Primary process measurement from an analog input (AI) block to the primary PID
- Secondary process measurement from an AI block to the secondary PID
- Set point from the primary controller output to the secondary PID
- Controller output from the secondary PID to an analog output (AO) block

It is desirable to also pass status information between the blocks. For example, if the secondary block is manual or local automatic, the primary controller should be aware of this. Furthermore, the set point of the secondary should be passed back to the primary controller for initialization purposes; thus, when the secondary is returned to cascaded automatic, no bump will be created in the primary controller output.

This last requirement entails a form of interblock communication that was not present in analog control systems. The interblock communication can be implemented several ways, depending upon the manufacturer. One way is to configure specific forward and backward connections as shown in Figure 2.4a. The forward connections are:

Analog input (AI) "OUT" to PID "PV"—Floating point signal values
PID (primary) "OUT" to PID (secondary) "PV"—Floating point signal value
PID (secondary) "OUT" to analog output (AO) "IN"—Floating point signal value

TABLE 2.4b
PID Configuration Items (Minimum Set)

Block Identification	*Block Tag*
Block type	PID
Cascade?	Yes or No
If yes, set point source	Primary block tag
Set point track PV in manual?	Yes or No
Process variable source	AI block tag
Direct or reverse acting	DA or RA
Derivative mode on meas or error	M or E
Proportional mode on meas or error	M or E
Controller gain	K_C
Integral time, minutes/repeat	T_I
Derivative time, minutes	T_D

The backward connections are:

PID (secondary) "TCO" to PID (primary) "TCI" — Boolean signal
PID (secondary) "TRO" to PID (primary) "TRI" — Floating point signal value

Whenever the secondary controller is in manual or local set point, the status is communicated backward to the primary controller via the TCO–TCI communication link. At the same time, the set point of the secondary controller is communicated backward to the primary controller via the TRO–TRI communication link. The status notification puts the primary PID into an initialization mode. The primary is initialized using Equation 2.4(6). This forces the primary controller output to be the same as the secondary controller set point. (The secondary controller should be configured so that its set point tracks its measurement whenever it is in manual.)

A minimum set of configuration questions is given in Table 2.4b. Most systems would have many more questions covering loop description, engineering ranges, alarm limits, etc.

Bibliography

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: ISA, 1990.
Marlin, T. E., *Process Control — Designing Processes and Control Systems for Dynamic Performance,* 2nd ed., New York: McGraw-Hill, 2000.
Ogunnaike, B. A., and Ray, W. H., *Process Dynamics, Modeling and Control,* New York: Oxford University Press, 1994.
Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.
Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.
Wade, H. L., *Regulatory and Advanced Regulatory Control: System Design and Application,* 2nd ed., Research Triangle Park, NC: ISA, 2004.

2.5 Control Modes—Closed-Loop Response

C. F. MOORE (1985, 1995) **J. HETTHÉSSY, R. BARS** (2005)

INTRODUCTION

In general, the aim of a control system is to maintain some desired conditions (signals) in a process. In case of a processing plant the measured conditions (signals) that are to be controlled are the *output signals* of the plant and the desired performance is to be achieved by manipulating the input variables of the plant, also called *control signals* or *manipulated variables.* The manipulation of some input variables will disturb the process, while others will not.

For example the maintaining of a comfortable temperature in a room can be considered a process control task, where the disturbances can come from ambient conditions, opening of windows, or from the number of people in the room, and the manipulation to achieve control is by varying the heat input. Other control examples can include driving a car within a lane on the road, maintaining the level of water in a tank, or controlling the speed or the position of physical equipment under different conditions.

The aim of this section is to outline the fundamental structure and properties of closed-loop control systems. The discussion will be limited to linear, single-input–single-output systems.

Control systems can be analyzed in the time domain or in the frequency domain by the use of transfer functions. Thus the section will start with a description of the alternative means of system analysis. Then the fundamental benefits of using feedback systems will be presented and the open-loop and closed-loop techniques will be compared. The roles of various transfer functions characterizing closed-loop systems will be discussed, and then some design guidelines will be given. Finally, the approach to control system analysis will be illustrated through an example.

LINEAR SYSTEMS

The behavior of a linear system can be described in the time domain, in the complex frequency (Laplace) operator domain, or in the frequency domain. In the time domain differential equations or state-space equations, which are derived on the basis of the physical operation, define the relationship between the output and the input signals of the controlled process.

In the Laplace operator domain, the differential equations are transformed to algebraic equations. The transfer function is defined as the ratio of the Laplace transforms of the output and input signals, respectively, assuming zero initial conditions.

In the frequency domain, the frequency function characterizes the quasi-stationary response of the system to a sinusoidal input signal, supposing that the transients have already decayed. An important property of linear systems is that in case of a sinusoidal input the output in the steady-state will also be sinusoidal and of the same frequency, but the amplitude and the phase angle of the output signal will differ from those of the input signal.

Amplitude and phase angle depend on frequency and provide two components with real values for the complex valued frequency function: *magnitude function* and *phase function.* The frequency function is formally obtained from the transfer function by substituting $s = j\omega$.

EXAMPLE

Consider a simple electrical circuit consisting of a resistance R and an inductance L as shown in Figure 2.5a. The input signal is the voltage u, the output signal is the current i.

The differential equation:

$$u = iR + L\frac{di}{dt} \qquad 2.5(1)$$

The transfer function:

$$G(s) = \frac{i(s)}{u(s)} = \frac{1}{R + sL} = \frac{1/R}{1 + sL/R} = \frac{A}{1 + sT} \qquad 2.5(2)$$

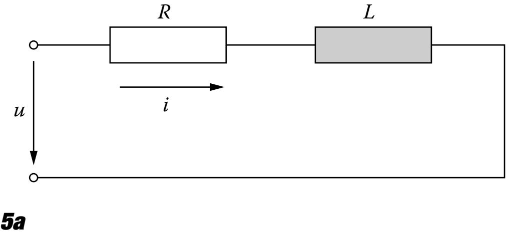


FIG. 2.5a
A simple electrical circuit.

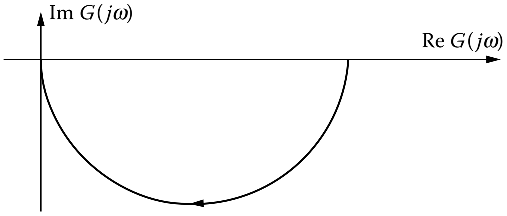


FIG. 2.5b
Nyquist diagram.

where s is the differential operator, A is the gain, and T is the time constant.

The unit step response is obtained by

$$i(t) = A(1 - e^{-t/T}), \qquad t \geq 0 \tag{2.5(3)}$$

while the frequency function becomes

$$G(j\omega) = \frac{A}{1 + j\omega T} \tag{2.5(4)}$$

The frequency function $G(j\omega)$ is a complex valued function, and it can be plotted in the complex plane, showing its values as ω changes from zero to infinity (Figure 2.5b). This polar plot is called the Nyquist diagram. The arrow indicates the direction of movement along the plot as the frequency increases.

Another possibility is to draw the magnitude and the phase angle of the frequency function versus the frequency separately. This technique is called *Bode diagram* (Figure 2.5c). Note that the Bode diagram uses a logarithmic scale for the frequency and *dB* values for the magnitude.

The performance of control systems can be evaluated in several ways. Analysis in the time domain requires solution of differential equations. Analysis in the complex frequency operator domain allows for much simpler mathematical calculations. However, it requires inverse Laplace transformation to obtain the results in the time domain. These same results in the time domain can also be obtained by simulation.

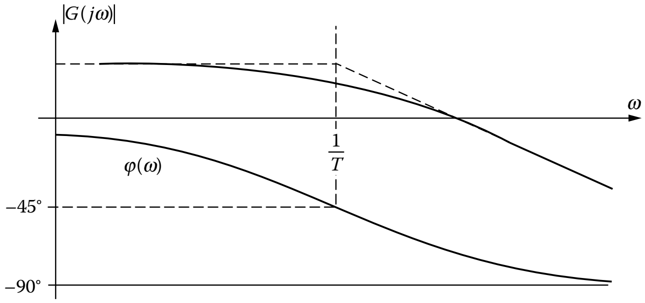


FIG. 2.5c
Bode diagram.

Analysis in the frequency domain — due to the relationship between the time and the frequency domain — gives an opportunity to draw conclusions for the essential system operation in the time domain.

Time and Frequency Domain Analysis

Control system signals can be analyzed by considering their frequency content. This means that they can be decomposed into an infinite number of sinusoidal signals of a given amplitude.

Periodic signals contain discrete frequencies (Fourier series), while aperiodic signals have a continuous frequency spectrum. The superposition theorem allows us to calculate the output signal of a linear system as a sum of the responses for the input signal components of individual frequencies.

This frequency response concept is illustrated by an example of a second-order system whose step response shows damped oscillations. The input u is a periodic rectangular signal. Figure 2.5d shows the input and its approximation versus time using only four of its sinusoidal components. The calculated accurate response y and the approximated response obtained as a sum of the individual responses for the four sinusoidal components are also given.

With more components, both the input and the output signals will be better approximated. Figure 2.5e shows the approximation with ten Fourier components.

Theoretically, if one knows the frequency response for all the frequencies in the whole frequency domain, the time

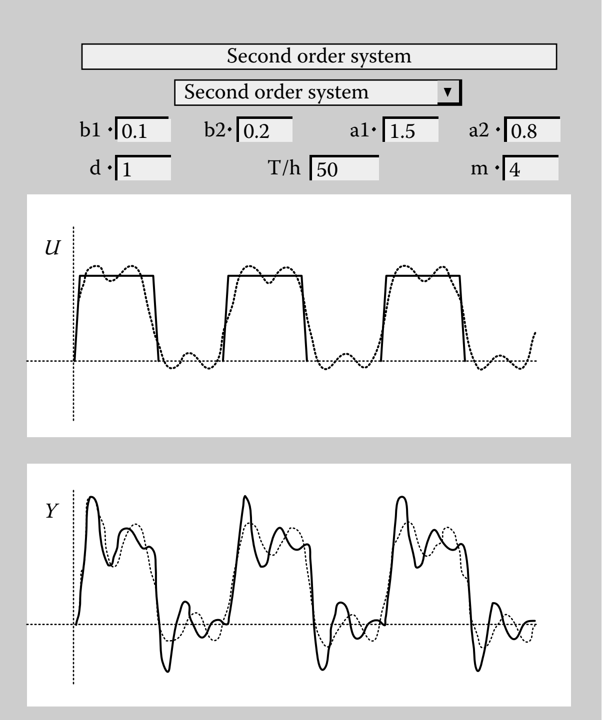
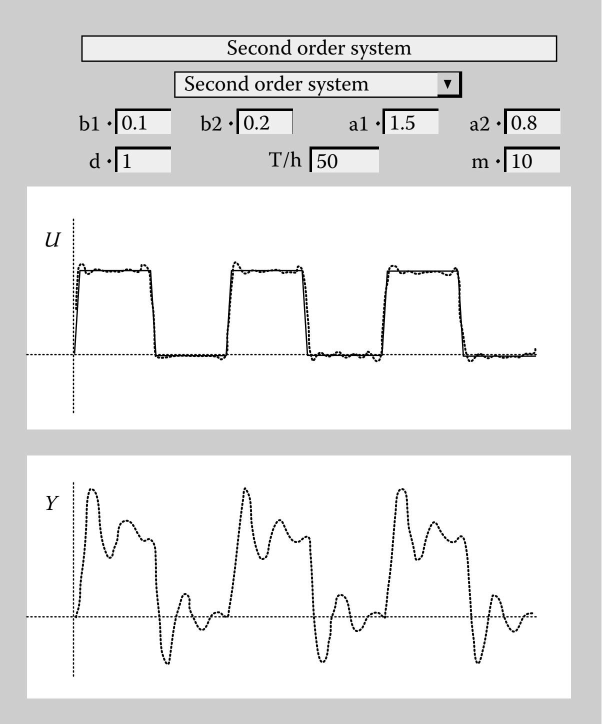

Second order system
Second order system ▾
b1 · 0.1
b2· 0.2
a1· 1.5
a2 · 0.8
d · 1
T/h 50
m · 4

FIG. 2.5d
Frequency response gives an approximation of the output signal.



Second order system

Second order system

b1 · 0.1 b2 · 0.2 a1 · 1.5 a2 · 0.8

d · 1 T/h 50 m · 10

FIG. 2.5e
Using more Fourier components the approximation of the input and output signals becomes more accurate.

response of the system for a given input can also be obtained. The problem is that an infinite number of output components must be taken into account.

Higher frequencies usually are strongly attenuated, so considering a finite, truncated frequency range still gives acceptable results. Approximating relationships between the time and the frequency domain provides a good orientation for evaluating the system behavior.

For instance, the steady-state value for the step response of a system can be obtained from the magnitude response taken at zero frequency. Assuming a step function input, the initial value of the output signal can be directly calculated by considering the frequency response at infinitely high frequencies. Similarly, the middle-range frequency behavior will characterize the transient response in the time domain.

OPEN- AND CLOSED-LOOP CONTROL

In designing a control system, the components of the control system first have to be selected considering the actual operating conditions, such as the range of the measured process variable and the available sensors and actuators.

In order to obtain a model of the process, the signal transfer properties have to be analyzed based on physical principles governing the operation of the individual elements, which are described by mathematical relationships. Another way to obtain the model is the identification of the static and dynamic behavior of the process based on the information represented by the available process input–output records. The result is a model, typically a transfer function for linear systems, that can serve as an analytical description of the control system.

In *open-loop* control systems, the output signal has no effect upon the controlled variable(s). The control is executed, for example, on a time basis. A practical example is the operation of a washing machine, where the duration of soaking, washing, and rinsing are set in advance and on the basis of experience with how these settings affect the output signal, which is the cleanness of the clothes. In such an open-loop system, the output signal is not automatically measured.

The main tool utilized by *closed-loop* controllers is the negative feedback. Here, a detector measures the output signal of the controlled process and its value is compared to a reference value (set point). The difference between the reference signal and the output signal is the error signal, which is driving a *controller.* The output of the controller generates the manipulated variable, which throttles the final control element that influences the input of the process. This is how the control loop is closed.

The difference between the reference signal (set point) and the measured output signal — the error — is the basis for the control action to eliminate that error. The control action — due to the dynamics of the individual parts — takes time. An appropriately tuned controller takes the process dynamics into account and ensures stable closed-loop control.

In practice, both open-loop and closed-loop controls are utilized. For instance, the complex start-up and shut-down sequences are often governed by open-loop control, provided by intelligent PLCs (Programmable Logic Controllers) and closed-loop control loops in combination.

Open-Loop vs. Closed-Loop Control

Open-loop control is simple and cheap because it does not require any sensors. The characteristics of the operating elements, the relationships between the input and output signals, and the expected disturbances have to be known in advance, as this knowledge is needed to set up the rules that govern the system operation.

Open loop control cannot tolerate changes in the loop parameters or in the disturbances, thus, errors in the assumed data will lead to improper operation. However, designing an open-loop control system is simple and does not require stability analysis.

Closed-loop control is more expensive, but in addition to being able to keep the process variable on set point, it is also capable of correcting for process disturbances and compensating for the errors in the model, which is implemented by the controller. Figure 2.5f shows the block diagram of a closed-loop control system.

The block diagram is composed of units, which are characterized by transfer functions. The *external inputs* of the closed-loop system are the *reference signal* (*set point*), the

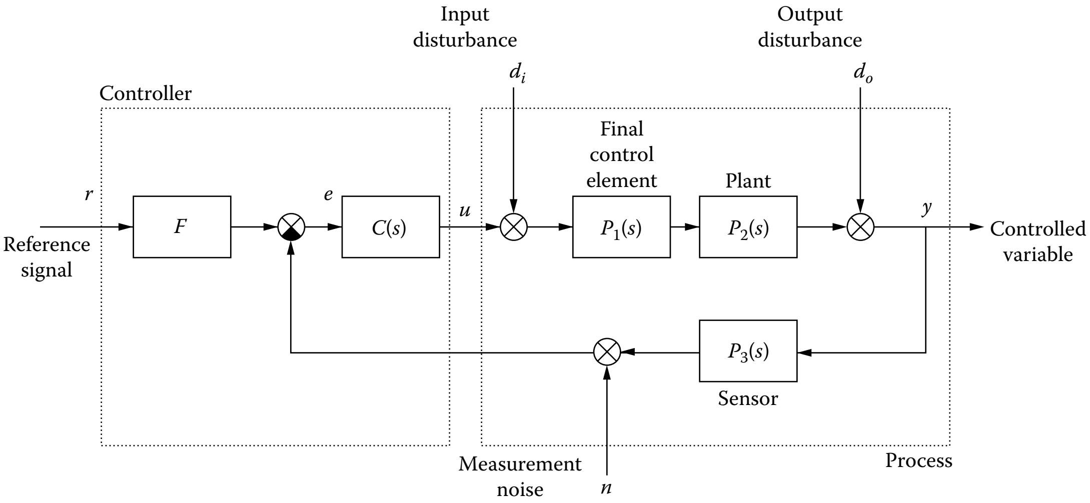


FIG. 2.5f
The structure of a closed-loop control system.

disturbances (sometimes called process load variations) and the *measurement noise.* The output is the *controlled variable,* measured by a sensor whose output (y) may be corrupted by measurement noise.

A simplified block diagram is shown in Figure 2.5g. The block diagram is an abstract model of a real closed-loop control system. In this model only the main system components are considered while others are neglected. At this point signal transfer properties are emphasized, and technology-dependent, practical realization aspects are ignored.

Stability and Dead Time

Closed-loop control loops can become unstable if not properly tuned. To ensure closed-loop stability is the most important goal. In general, feedback may lead to unstable closed-loop behavior due to delays, dead time, and high gains in the controlled process. The system is on the boundary of stability and instability if steady oscillations occur. Stability must be ensured for a feedback control system by all means.

Stability can be defined in various ways. Bounded input–bounded output (BIBO) stability means that all bounded inputs result in bounded outputs. Asymptotic stability means that when changing the reference signal or when disturbances occur, all the transient responses in the system will asymptotically approach zero.

To illustrate the problem of stability, consider the control system shown in Figure 2.5h. The controlled process is a dead time process, where the dead time can be caused by transportation lag.

It is assumed that the controller is proportional; its gain is K. Figure 2.5i shows the output and error signals that result from a step change in the set point (reference signal), if $K =$ 0.5, 1, and 2, respectively.

If the proportional gain is $K = 0.5$, the system is stable, but the steady-state value of the controlled variable (y, the output

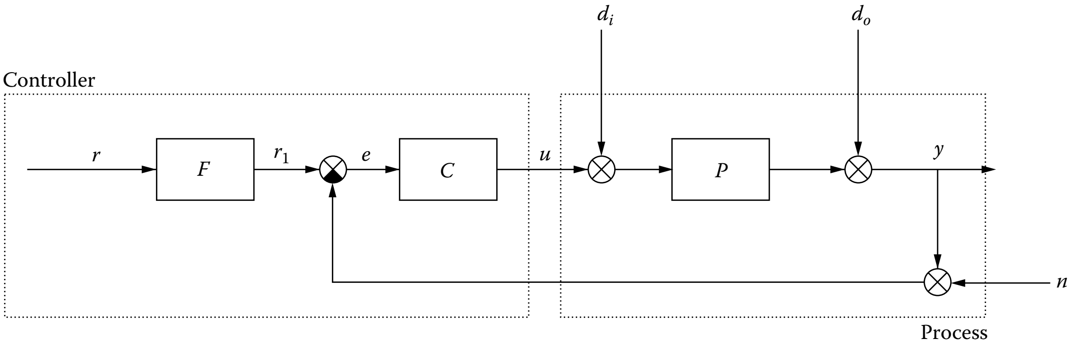


FIG. 2.5g
Simplified block diagram of a closed-loop control system. $P = P_1P_2P_3$.

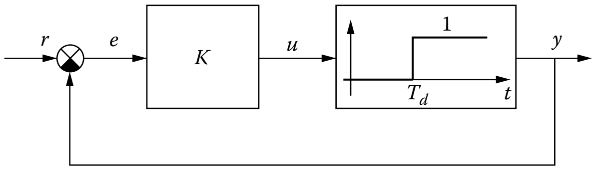


FIG. 2.5h
Proportional control of a dead-time process.

signal) is approaching 1/3 instead of 1. For $K = 1$ the system is on the boundary of stability, and with $K = 2$ the system becomes unstable. The controller has to be modified, such as replaced by a sample-and-hold controller (Section 2.2) in order to ensure both stability and acceptable steady-state response.

Advantages of Closed-Loop Control

One advantage of closed-loop control is that in the steady state, if the loop is properly tuned, the controlled variable is kept on set point within the accuracy of the system. This property is usually referred to as the *servo property.*

Another advantage of closed-loop control is the ability of a properly tuned controller to correct for process disturbances. As an example consider a controlled process that has two time lags, and review its response to a process disturbance under PI and under PID control. The process transfer function is:

$$P(s) = \frac{1}{(1+5s)(1+10s)} \qquad \textbf{2.5(5)}$$

The PI and PID controller transfer functions are:

$$C_1(s) = \frac{1+10s}{10s} \quad \text{and} \quad C_2(s) = 5\frac{1+10s}{10s}\frac{1+5s}{1+s} \qquad \textbf{2.5(6)}$$

Figure 2.5j shows the controlled variable (y, the output) and the manipulated variables (u, the input) control signals for both controllers as a function of time. At time $t = 80$ sec the same disturbance is introduced (a unit step). Both control loops return to set point, but with the PID controller the response is superior because it is faster.

Yet another advantage of feedback-based closed-loop controls is their ability to maintain control even when the process dynamics change. This is illustrated in Figure 2.5k, which shows the response of the previously discussed PID controller to changes in process dynamics. The line noted by "s" describes the loop's response when the gain and the time constants of the process drop by 20%, and the line identified by "h" shows the performance when the process gain and the time constants increase by 20%. It can be noted that the control system compensates for the by upsets in the process dynamics.

Closed-loop control can also be used to stabilize unstable open-loop control systems. To illustrate the stabilizing effect of feedback, consider the unstable process described by the transfer function $1/(s - 2)$. As the transfer function contains a pole on the right side of the complex plane, its step response goes to infinity.

If a plain proportional feedback controller with a gain of K is applied (Figure 2.5l), a proportional setting of $K > 2$

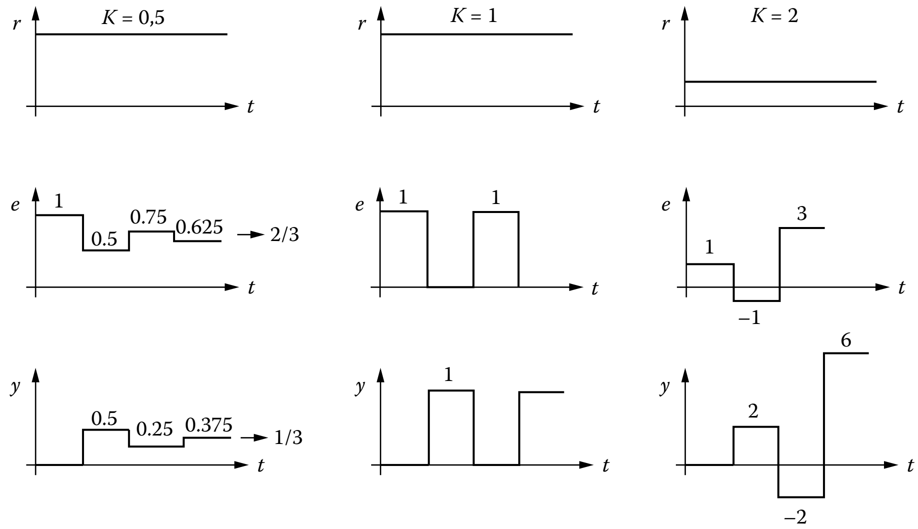


FIG. 2.5i
The response to a step change in set point (r) is shown when controlling a pure dead-time process with a plain proportional controller. The amounts of error (e) and the values of the controlled variable (y) are both shown in each dead-time period. It can be seen that the process becomes unstable when the gain setting exceeds K = 1.

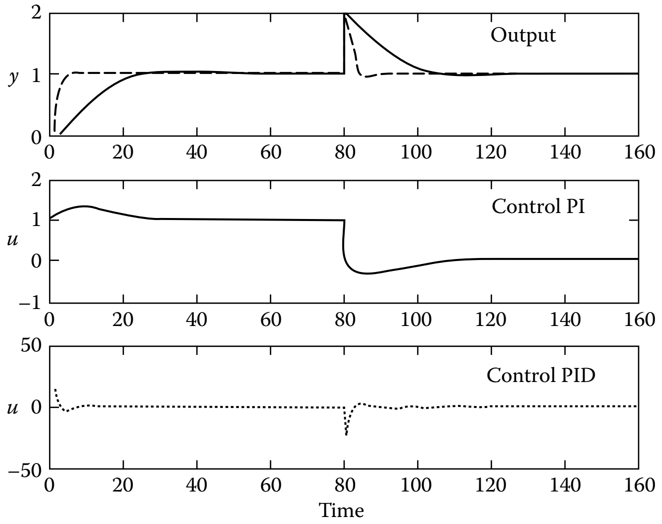


FIG. 2.5j
The controlled variable (y) recovers faster from the upset caused by a process disturbance if under PID control than with PI control.

will stabilize the loop. Figure 2.5m illustrates that with proportional gains of both $K = 5$ and $K = 10$, the response of the loop to a step change will be stable. Naturally, it also shows that proportional control cannot return the process to set point; a substantial offset remains.

CLOSED-LOOP TRANSFER FUNCTIONS

The relationships between external signals and response signals produced by the closed-loop control system with linear elements can be described by a set of transfer functions, called *overall transfer functions.*

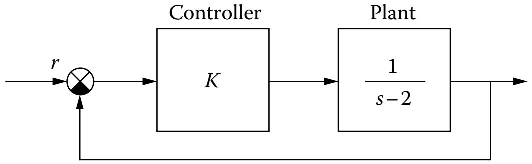


FIG. 2.5l
A high gain proportional controller can stabilize an otherwise unstable process loop but only if the gain is high (PB narrow).

When evaluating or designing a closed-loop control system, the response of the system to changes in the external signals is of interest. In the following paragraphs we will determine the overall transfer functions between the output signal (controlled variable — y) and the reference signal (set point — r), disturbances d_i and d_o and measurement noise n in Figure 2.2g. We will also derive the overall transfer functions related to the error signal e and the manipulated variable (control signal — u).

Block Diagram Algebra

Before calculating the transfer functions, let us consider some algorithms to handle the overall transfer functions of serial and parallel blocks (Figure 2.5n). For series connection the overall transfer function is obtained by multiplying the individual transfer functions, and for parallel connection the transfer functions are added.

Figure 2.5o shows the block diagram of a feedback control loop.

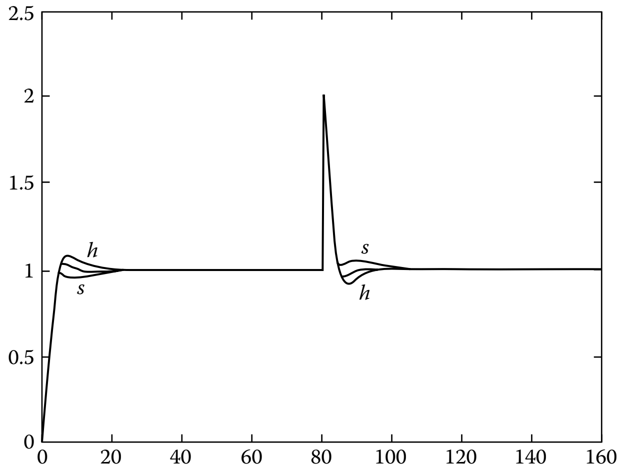


FIG. 2.5k
A closed-loop controller can correct for disturbances caused by changes in process dynamics. The process returns to set point both when the process gain and time constants drop (curve s) or when they rise (curve h).

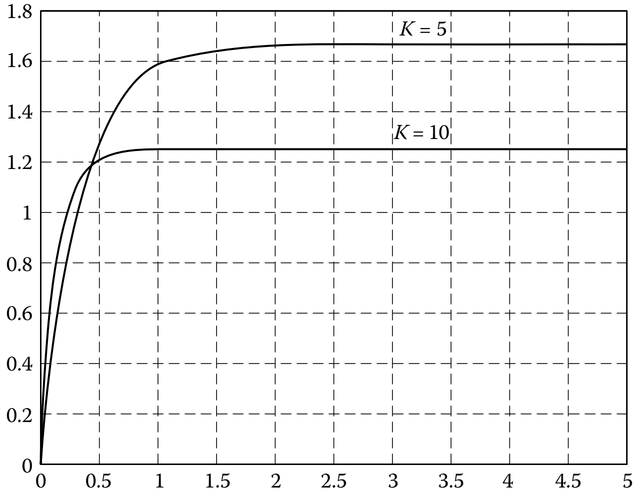


FIG. 2.5m
While the control loop in Figure 2.5l is stabilized, the plain proportional controller cannot return the controlled variable to set point.

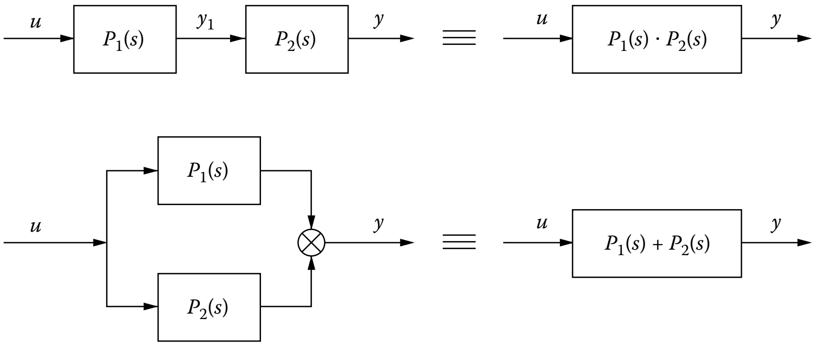


FIG. 2.5n
Overall transfer functions of serially and parallel connected blocks.

The transfer function that is obtained by walking around the loop is called the *loop transfer function L*(*s*)*:*

$$L(s) = C(s)P(s) \tag{2.5(7)}$$

The output signal, if for sake of simplicity the Laplace operator (*s*) is dropped, is calculated as

$$y = CPe = CP(Fr - y) \tag{2.5(8)}$$

The overall transfer function is

$$\frac{y}{r} = F\frac{CP}{1+CP} \tag{2.5(9)}$$

The numerator contains the transfer function along the forward path from a selected external input signal to the output signal of interest, while the denominator is simply (1 + *L*(*s*)). It is seen that the term 1 + *L*(*s*) is a rather important quantity. It is also called the *return difference*, a fictitious transfer function describing the difference between the signal entering the summation point (*F*(*s*)*r*(*s*)) and the signal returned by the control loop (−*L*(*s*)*e*(*s*)) , where *r*(*s*) is the Laplace transform of the reference signal and *e*(*s*) is the Laplace transform of the error signal.

Overall Transfer Functions

The overall transfer functions of the closed-loop control system described in Figure 2.5g are

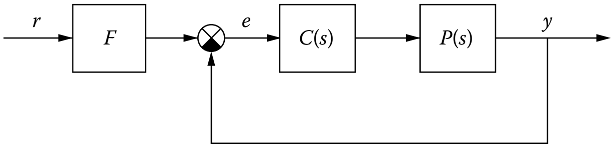


FIG. 2.5o
Overall transfer function of a feedback loop.

$$y(s) = \frac{F(s)C(s)P(s)}{1+C(s)P(s)}r(s) + \frac{P(s)}{1+C(s)P(s)}d_i(s) + \frac{1}{1+C(s)P(s)}d_o(s) - \frac{C(s)P(s)}{1+C(s)P(s)}n(s) \tag{2.5(10)}$$

$$e(s) = \frac{F(s)}{1+C(s)P(s)}r(s) - \frac{P(s)}{1+C(s)P(s)}d_i(s) - \frac{1}{1+C(s)P(s)}d_o(s) - \frac{1}{1+C(s)P(s)}n(s) \tag{2.5(11)}$$

$$u(s) = \frac{F(s)C(s)}{1+C(s)P(s)}r(s) - \frac{C(s)P(s)}{1+C(s)P(s)}d_i(s) - \frac{C(s)}{1+C(s)P(s)}d_o(s) - \frac{C(s)}{1+C(s)P(s)}n(s) \tag{2.5(12)}$$

Each component in the control loop can significantly contribute to the overall performance of the system. The denominators in all the transfer functions above are identical (return difference), and this denominator determines both the transient performance and the stability of the control loop.

The poles (the roots of the denominator polynomial) have to be at the left-hand side of the complex plain to ensure stability. Their location is essential as far as the properties of the transient response are concerned, too. Reference signal tracking in addition is influenced by the choice of the reference signal filter *F*. The numerators of the corresponding overall transfer functions also influence the set-point tracking, disturbance, and noise rejection properties.

One of the most important transfer functions among the above equations is the overall transfer function characterizing the servo property (set-point tracking), i.e., the way that the output of the closed-loop system follows the reference signal:

$$W(s) = \frac{y(s)}{r(s)} = \frac{F(s)C(s)P(s)}{1+C(s)P(s)} \tag{2.5(13)}$$

Bode introduced the term *sensitivity,* which describes the relative sensitivity of the overall transfer function (W) to the relative process model error:

$$S(s) = \frac{dW/W}{dP/P} = \frac{1}{1+C(s)P(s)} = \frac{1}{1+L(s)} \qquad \textbf{2.5(14)}$$

The ability of the closed-loop system to reject disturbances is also related to sensitivity. The sensitivity to the output disturbance is defined as

$$S_o(s) = \frac{e(s)}{d_o(s)} = \frac{-1}{1+C(s)P(s)} = \frac{-1}{1+L(s)} = -S(s) \qquad \textbf{2.5(15)}$$

Similarly, the sensitivity to the input disturbance is defined as

$$S_i(s) = \frac{e(s)}{d_i(s)} = \frac{-P(s)}{1+C(s)P(s)} = \frac{-P(s)}{1+L(s)} = -P(s)S(s) \qquad \textbf{2.5(16)}$$

and the sensitivity to the measurement noise is defined as

$$S_n(s) = \frac{e(s)}{n(s)} = \frac{-1}{1+C(s)P(s)} = \frac{-1}{1+L(s)} = -S(s) \qquad \textbf{2.5(17)}$$

Finally, the complementary sensitivity function has been introduced according to

$$T(s) = 1 - S(s) = \frac{L(s)}{1+L(s)} \qquad \textbf{2.5(18)}$$

Observe the straightforward relation

$$W(s) = F(s)T(s) \qquad \textbf{2.5(19)}$$

The complete set of the transfer function relations can be repeated as

$$y(s) = F(s)T(s)r(s) + P(s)S(s)d_i(s) + S(s)d_o(s) - T(s)n(s) \qquad \textbf{2.5(20)}$$

$$e(s) = F(s)S(r)r(s) - P(s)S(s)d_i(s) - S(s)d_o(s) - S(s)n(s) \qquad \textbf{2.5(21)}$$

$$u(s) = F(s)C(s)S(s)r(s) - T(s)d_i(s) - C(s)S(s)d_o(s) - C(s)S(s)n(s) \qquad \textbf{2.5(22)}$$

It is quite clear that the sensitivity function $S(s)$ plays a key role in the design procedure. For a one-degree-of-freedom system ($F(s) = 1$, when tracking and disturbance rejection properties are complementary), $C(s)$ is to be designed to fulfill the following closed-loop requirements: stability, good servo property, effective disturbance rejection, noise attenuation, and low sensitivity with respect to the process model uncertainties.

For two-degrees-of-freedom control systems ($F(s) \neq 1$) the servo property can be separated from the rest of the design goals by adding a filter to preprocess the reference signal. Some fundamental design aspects of this will be discussed in the next section, but first consider an introductory example.

EXAMPLE

The transfer function of the control system in Figure 2.5g is

$$P(s) = \frac{A}{(1+sT_1)^3}; \quad A = 1, \quad T_1 = 1; \quad C(s) = 0.3\frac{1+s}{s};$$

$$F(s) = \frac{1}{1+0.2s} \qquad \textbf{2.5(23)}$$

The step response and the magnitude curves for the overall transfer functions determining the output and control performance are shown in Figures 2.5p and 2.5q, respectively.

Figure 2.5r shows the controlled variable (y, the output) and the manipulated variable (u, the control response) to a unit-step change in the controller's set point (r, the reference signal). In this case, the disturbance is also a unit step, which occurs at $t = 30$ sec, and the measurement noise is a sinusoidal signal with a huge, unity amplitude and a frequency of 3 rad/sec.

The control system tracks the reference signal (returns to set point) without any steady-state error and also effectively corrects for the input disturbances. The effect of the measurement noise is significantly attenuated, as at this frequency the amplitude of the overall transfer function T is much below 1. The control effort to attenuate the effect of the noise therefore is considerable. Figure 2.5r also shows that at lower frequencies (below 1 rad/sec) the measurement noise is not filtered out effectively.

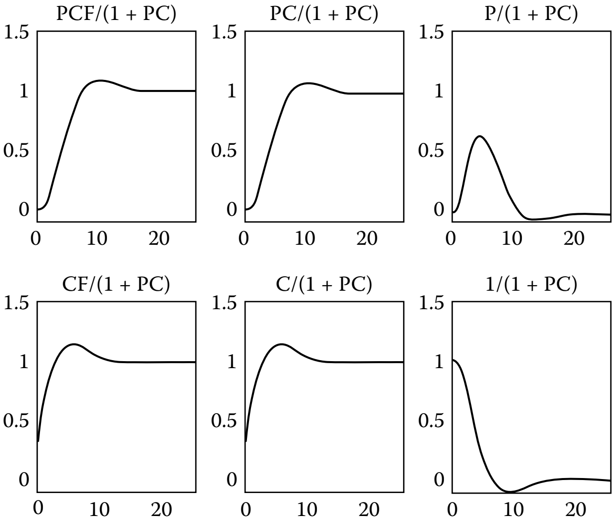


FIG. 2.5p
Step responses of the significant overall transfer functions in the control example.

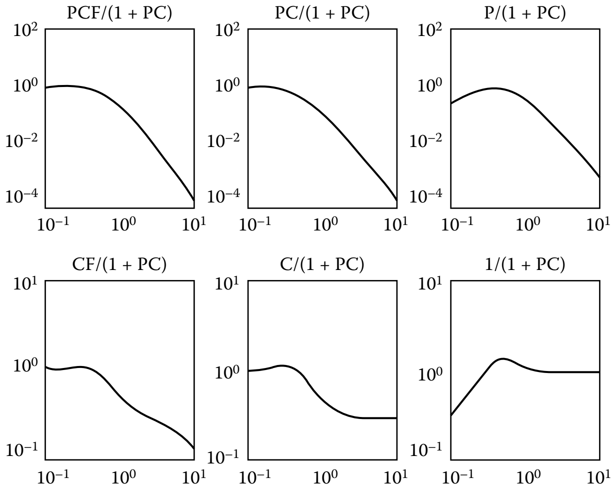


FIG. 2.5q
Gain curves of frequency responses of the significant overall transfer functions in the control example.

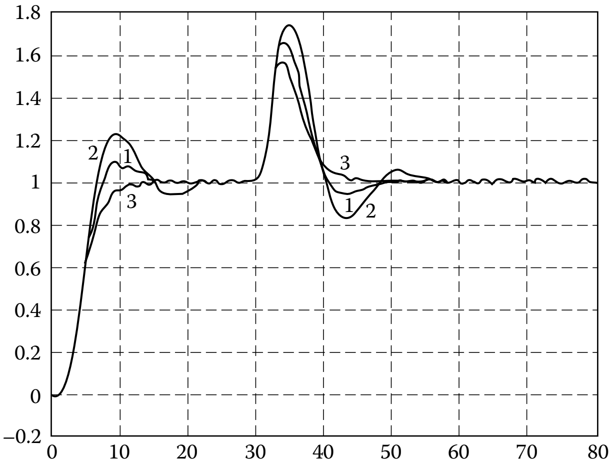


FIG. 2.5s
The response of the controlled variable (y, the process output) in response to variations in the gain and time constants of the process.

To evaluate the performance of a closed-loop control system, it is not enough to watch the controlled and manipulated variables (process output and control signals) of the model. The nominal design procedure assumes that the process model exactly matches the real process. It is also important to analyze how the system behaves if the plant parameters differ from their assumed nominal values.

When the model is inaccurate (has assumed a set of process dynamics, such as dead times and time constants, that are not accurate), Figure. 2.5s illustrates the effects of this inaccuracy on the controlled variable response to a disturbance.

Response curve 1 describes the situation where the transfer function $P(s)$ for the model was assumed to have a gain of $A = 1$ and time constant of $T_1 = 1$ sec and this was a correct assumption, i.e., it corresponds to the dynamics of the controlled process.

Response curve 2 describes the situation where the transfer function $P(s)$ for the model was assumed to have a gain of $A = 1$ and time constant of $T_1 = 1$ sec, while the real process had a gain of $A = 1.2$ and time constant of $T_1 = 1.2$ sec. Curve 3 describes the situation where the transfer function $P(s)$ for the model was as before, while the real process had a gain of $A = 0.8$ and time constant $T_1 = 0.8$ sec, respectively.

The response illustrated in Figure 2.5s shows that a 20% dynamic parameter uncertainty can be corrected for by the closed-loop controller.

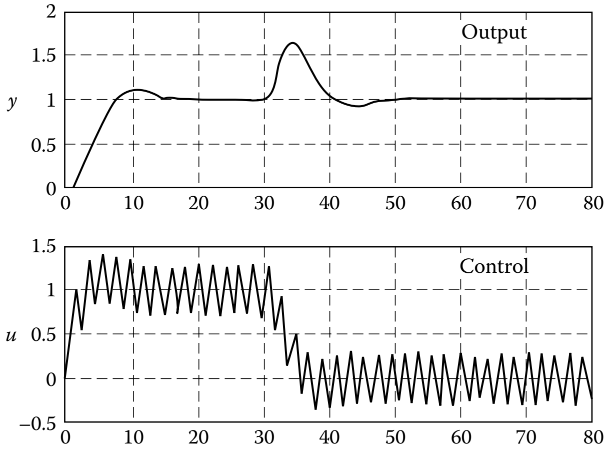


FIG. 2.5r
Controlled and manipulated variables (y and u, the output and input signals, respectively) of the controlled process with a measurement noise of 3 rad/sec frequency.

STABILITY AND DYNAMIC PERFORMANCE

The most important goal of control system design is to ensure *stability,* but some other goals are also important. These include steady-state accuracy (good set-point tracking), effective disturbance rejection along with noise attenuation, good transient performance (short settling time with no oscillations), and robustness with respect to modeling errors. In the paragraphs below, the goal of steady-state accuracy will be discussed first.

Following the Set Point

Let us first consider the set point (reference signal) tracking properties in steady state. In modeling set-point changes, one can consider the control system response to step changes and ramp and parabolic set-point adjustments.

Once a controller has been designed, the control loop can be described by the loop-transfer function $L(s) = C(s)P(s)$. The *type number* (number of pure integrators) of the control loop reflects the number of integrators in the loop transfer function. If no integrators are involved the type number is 0.

Assume that the loop-transfer function is given in the following form:

$$L(s) = \frac{K}{s^i} Q(s) \qquad \textbf{2.5(24)}$$

where i denotes the type number, $Q(s)$ is a rational transfer function with $\lim_{s \to 0} Q(s) = 1$, and K is the static loop gain.

Supposing a filter for the reference signal with unity steady-state gain ($F(0) = 1$), the error signal can be expressed as

$$e(s) = \frac{F(s)}{1 + L(s)} r(s) = \frac{F(s)s^i}{s^i + KQ(s)} r(s) \qquad \textbf{2.5(25)}$$

where $r(s) = \frac{1}{s^j}$, $j = 1$ for step, $j = 2$ for ramp, and $j = 3$ for parabolic reference input.

The steady-state value can be calculated by the final value theorem:

$$\lim_{t \to \infty} e(t) = \lim_{s \to 0} se(s) \qquad \textbf{2.5(26)}$$

Evaluating the above expression for j =1, 2, and 3, the results concerning the steady-state errors can be summarized in Table 2.5t.

According to the table, the Type 0 proportional control systems can track the set point (step reference signal) only with nonzero steady-state error, which error decreases as the proportional gain rises.

Controllers containing one integrator (Type 1) can track the step change in the set point accurately, without steady-state error, because an integrator keeps the output (controlled variable) constant only if its input (actually, the error signal) reaches zero.

TABLE 2.5t

Steady-State Errors Resulting from Different Types of Set Point Changes and for Different (Types 0, 1, 2) Processes

	Type 0	*Type 1*	*Type 2*
Step input $j = 1$	$\frac{1}{1+K}$	0	0
Ramp input $j = 2$	∞	$\frac{1}{K}$	0
Parabolic input $j = 3$	∞	∞	$\frac{1}{K}$

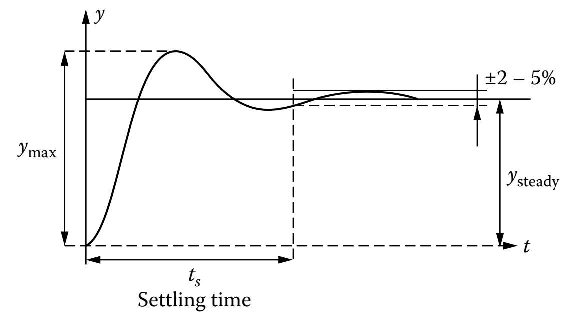


FIG. 2.5u

Dynamic response: overshoot and settling time.

With two integrators (Type 2), static accuracy is even better. Therefore, one may conclude that static accuracy is improved by increasing the number of integrators and/or the loop gain. This relation certainly holds; however, static accuracy and stability are conflicting requirements, in that too many integrators or high gains may cause instability. The goal of control system optimization is to find the best compromise.

Load Disturbance

A similar analysis can be performed to determine the impacts of load disturbances and measurement noise on the steady-state error. In the example discussed earlier in connection with Figure 2.5r, it was shown that a Type 1 control system accurately follows a step change in set point and completely eliminates the effect of a step disturbance.

Dynamic Response

The dynamic response of a closed-loop control system can be evaluated on the basis of the response of the controlled variable to a unit-step change in the set point (reference signal). Figure 2.5u describes such a response. The overshoot σ is defined by

$$\sigma = \frac{y_{\max} - y_{\text{steady}}}{y_{\text{steady}}} 100\% \qquad \textbf{2.5(27)}$$

The settling time t_s can be defined as time it takes to bring the controlled variable to within 2 to 5% of its steady-state value. Such dynamic properties as stability, overshoot, and settling time can also be evaluated from frequency domain analysis.

Frequency-Function-Based Evaluation

Figure 2.5v depicts a Bode amplitude–frequency diagram of an open-loop system $L(j\omega)$. The curve can be approximated

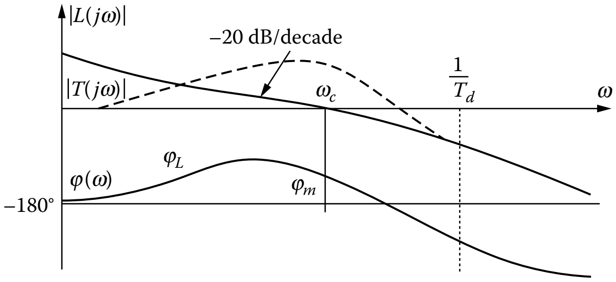


FIG. 2.5v
Typical open-loop Bode diagram of a stable system.

by its asymptotes. An important point on the curve is the cutoff frequency ω_c, where the absolute value is 1.

Stability properties can be obtained from the middle frequency performance of the open-loop frequency function. The system is stable if ω_c is at slope –20 dB/decade and is to the left of $1/T_d$, where T_d denotes the dead time of the plant. ω_c is also related to the settling time, which can be approximated as

$$3/\omega_c < t_s < 10/\omega_c \qquad \textbf{2.5(28)}$$

The shape of $|T(j\omega)|$ characterizing the relationship between the controlled variable (output) and the set point (reference signal) is also given in Figure 2.5v. Its magnification above 1 around ω_c indicates overshoot in the time response. This magnification must therefore be avoided.

The overshoot can be also estimated from the phase margin φ_m, which is calculated from the phase angle of the open-loop frequency function considered at frequency ω_c.

$$\varphi_m = 180° + \varphi_L(\omega_c) \qquad \textbf{2.5(29)}$$

The overshoot is expected to occur below 10% if $\varphi_c > 60°$.

Loop Shaping

When setting up a strategy to design a closed-loop control system to satisfy various control objectives one should consider Equations 2.5(20) to 2.5(22). Based on those, the fundamental design expectations can be summarized as follows:

- For good servo property $|L(j\omega)|$ has to be large.
- For good disturbance rejection $|L(j\omega)|$ has to be large.
- For good noise attenuation $|L(j\omega)|$ has to be small.
- To see moderate control $|L(j\omega)|$ has to be small.
- To compensate for the uncertainty in the process model $|L(j\omega)|$ has to be large.

Taking $S(s) = \frac{1}{1+L(s)}$ into account it is seen that requirements elaborated for $L(j\omega)$ can be transformed to a requirement for $S(j\omega)$. The sensitivity function $S(s)$ also plays a key role in the design. However, the challenge faced by the control engineers is that the above objectives are in conflict. To obtain a good design, the above listed control objectives should be carefully specified and a frequency region should be attached to each control objective.

In practice, the sinusoidal components of the reference signals and of the disturbances are in the low frequency region, while measurement noise is typically of high frequency. Consequently, most of the control objectives can be met by applying a high gain ($|L(j\omega)| > 1$) for low frequencies and a low gain ($|L(j\omega)| < 1$) at high frequencies. In addition, steady-state accuracy is also improved by high DC gain.

As far as the effect of the unmodeled process dynamics is concerned, uncertainty is usually reduced at the lower frequencies. If the model uncertainty exceeds a certain level even at low frequencies, there are two options to select from:

1. Conduct experiments to improve the process model by gaining more information about the process.
2. Lower the cutoff frequency and thereby decrease the demand for improved dynamic response.

The classical design technique performed in the frequency domain starts with the magnitude function of $P(j\omega)$ and evolves by gradually adding appropriate $C(j\omega)$ components to the loop of $L(j\omega) = C(j\omega)P(j\omega)$. This procedure is called *loop shaping*. When using this approach, special care should be taken when the controlled process is a nonminimum phase process or has substantial time delay.

Figure 2.5w shows the loop-shaping considerations related to the Bode amplitude diagram of the loop transfer function. A dotted line shows the complementary sensitivity function in the figure. At low frequencies it is approximately 1, while at high frequencies it is the same as $|L(j\omega)|$.

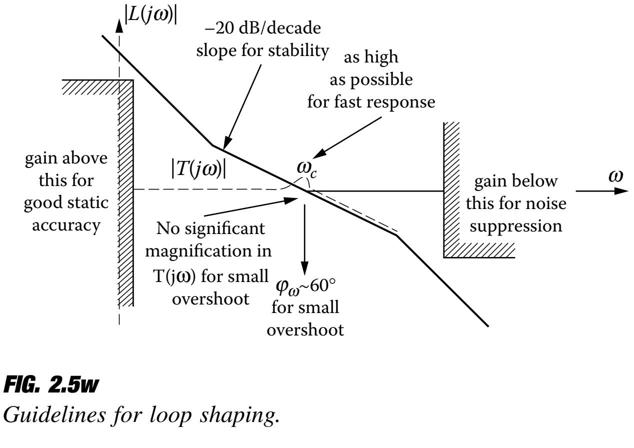


FIG. 2.5w
Guidelines for loop shaping.

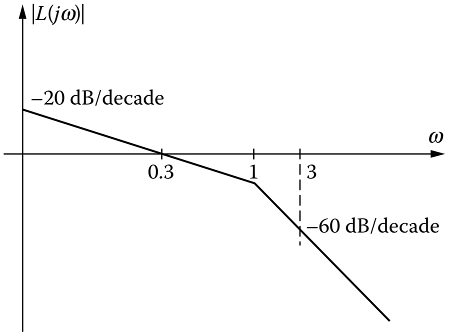


FIG. 2.5x
Bode amplitude–frequency diagram of the loop transfer function in the considered example.

Consider the third-order example analyzed previously. Assume that the process has three time lags and it is to be controlled by a PI controller. Figure 2.5x shows the Bode amplitude–frequency diagram of the loop's transfer function. The system is stable, as the Bode amplitude diagram crosses the zero dB axis with slope of –20 dB/decade (no dead time in the system).

The integrating effect in the PI controller ensures high gain at low frequencies. Therefore, as was shown in Figure 2.5s, steady-state error results from step changes in either set point or process load disturbances. Noise attenuation is very good, as at the noise frequency (3 rad/sec) the gain of the loop frequency function is very low. The phase margin is 58.8°, ensuring small overshoot. The settling time is about 15 sec.

CONCLUSIONS

The total performance of a control system depends on the overall response of the control loop and process acting together. The dynamic behavior of the total control system after an upset, which can be caused by a process disturbance, load change, or set point adjustment, is also a function of the type of controller used (Figure 2.5y).

The response is evaluated by the speed with which the controlled variable returns to the set point, by the amount of overcorrection or overshoot that occurs, and by the effects on the stability of the system during the upset condition. Depending on the nature of the process, different control modes may be required for optimum performance.

The system response utilizing transfer functions describing the controller and the process it controls has been analyzed. In case of uncertainties, sensitivity functions can be used to characterize the control system. To obtain guidelines for controller selection, loop-shaping considerations can be evaluated.

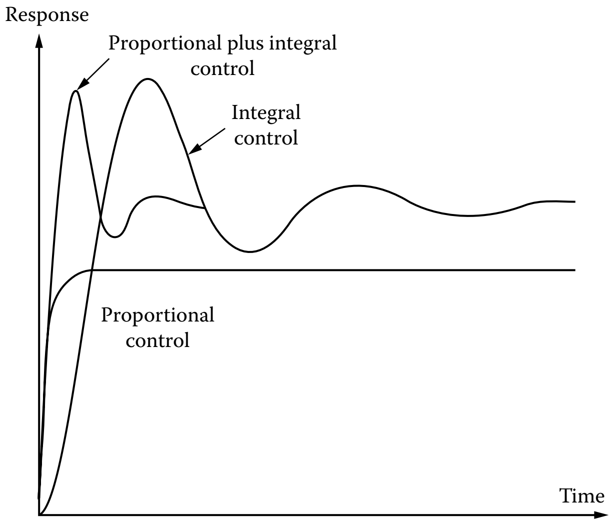


FIG. 2.5y
Response to a step change in set point with proportional, integral, and proportional-plus-integral controllers.

Bibliography

Åström, K. J., and Hägglund, T., *PID Controllers: Theory, Design and Tuning,* Research Triangle Park, NC: Instrument Society of America, 1995.

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: Instrument Society of America, 1990.

D'Azzo, J. J., and Houpis, C. H., *Feedback Control System Analysis and Synthesis,* New York: McGraw-Hill, 1960.

Dorf, R. C., and Bishop, R. H., *Modern Control Systems,* 7th ed., Reading, MA: Addison-Wesley, 1995.

Franklin, G. F., Powell, J. D., and Emami-Naeini, A., *Feedback Control of Dynamic Systems,* 3rd ed., Reading, MA: Addison-Wesley, 1994.

Golten, J., and Werver, A., *Control System Design and Simulation,* New York: McGraw-Hill, 1991.

Goodwin, G. C., Graebe, S. F., and Salgado, M. E., *Control System Design,* New York: Prentice Hall, 2001.

Kuo, B. C., *Automatic Control Systems,* New York: John Wiley & Sons, 2003.

Levine, W. S. (Ed.), *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Luyben, W. L., and Luyben, M. L., *Essentials of Process Control,* New York: McGraw-Hill, 1997.

Maciejowski, J. M., *Multivariable Feedback Design,* Reading, MA: Addison-Wesley, 1989.

Morari, M., and Zafiriou, E., *Robust Process Control,* New York: Prentice Hall, 1989.

Ogata, K., *Modern Control Engineering,* New York: Prentice Hall, 1997.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* 4th ed., New York: Prentice Hall, 1999.

Rohrs, C. E., Melsa, J. L., and Schultz, D. G., *Linear Control Systems,* New York: McGraw-Hill, 1993.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* New York: John Wiley & Sons, 1989.

Shinskey, F. G., *Process Control Systems: Application, Design and Tuning,* 4th ed., New York: McGraw-Hill, 1998.

Sinha, N. K., *Linear Systems,* New York: John Wiley & Sons, 1991.

Skogestad, S., and Postlethwaite, I., *Multivariable Feedback Control,* New York: John Wiley & Sons, 1996.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

Trentelman, H. L., and Willems, J. C. (Eds.), *Essays on Control: Perspectives in the Theory and its Applications,* Basel: Birkhäuser, 1993.

Van de Vegte, J., *Feedback Control Systems,* New York: Prentice Hall, 1994.

2.6 Control Systems — Cascade Loops

F. G. SHINSKEY (1970, 1985) **B. G. LIPTÁK** (1995) **R. BARS, J. HETTHÉSSY** (2005)

INTRODUCTION

Cascade control has a multi-loop structure, where the output of the controller in the outer loop (the "primary" or "master") is the set point of a controller in the inner loop (the "secondary" or "slave").

The slave measurement is an intermediate process variable that can be used to achieve more effective control of the primary process variable. In the cascade configuration the process is divided into two parts and therefore two controllers are used, but only one process variable (*m*) is manipulated.

In Figure 2.6a the primary controller maintains the primary variable y_1 at its set point by adjusting the set point r_2 of the secondary controller. The secondary controller, in turn, responds both to set point r_2 and to the secondary controlled variable y_2. This secondary controlled variable also affects the primary process and therefore the primary controlled variable (y_1), hence the loop is closed.

This cascade loop can also be shown in a more detailed block diagram form (Figure 2.6b). Here, the primary controller (C_1) generates the set point for the secondary controller (C_2), while the secondary controlled variable (y_2) also affects the primary process (P_1) and therefore it also affects the primary controlled variable (y_1).

Cascade control is advantageous on applications where the P_1 process has a large dead time or time lag and the time delays in the P_2 part of the process are smaller. Cascade control is also desirable when the main disturbance is in the secondary loop. This is because with the cascade configuration, the correction of the inner disturbance d_i occurs as soon as the secondary sensor (y_2) detects that upset.

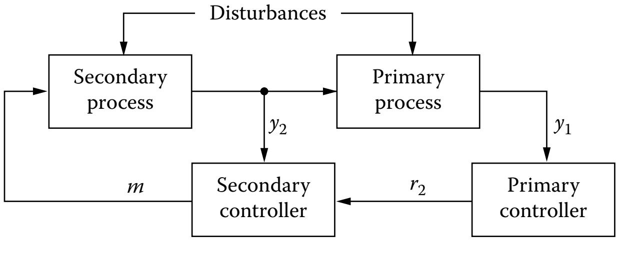


FIG. 2.6a
The cascade control system divides the process into two parts.

Cascade System Advantages

There are two main advantages gained by the use of cascade control:

1. Disturbances that are affecting the secondary variable can be corrected by the secondary controller before their influence is felt by the primary variable.
2. Closing the control loop around the secondary part of the process reduces the phase lag seen by the primary controller, resulting in increased speed of response. This can be seen in Figure 2.6c. Here, the response of the open loop curve shows the slow response of the secondary controlled variable when there is no secondary controller. The much faster closed loop curve describes the response of the secondary controlled variable when the cascade configuration adds a secondary inner control loop.

Other advantages of using cascade control include the ability to limit the set point of the secondary controller. In addition, by speeding up the loop response, the sensitivity of the primary process variable to process upsets is also reduced. Finally, the use of the secondary loop can reduce the effect of control valve sticking or actuator nonlinearity.

COMPONENTS OF THE CASCADE LOOP

The primary or outer control loop of a cascade system is usually provided with PI or PID control modes and is designed only *after* the secondary loop has already been designed. This is because the characteristics of the slave loop have an effect on the master loop.

For example, if the secondary measurement is nonlinear, such as the square root signal of an orifice-type flow sensor, it would produce a variable gain in the primary loop if the square root was not extracted.

For these reasons, the secondary loop requirements will be discussed first.

The Secondary Loop

Ideally, the secondary variable should be so selected as to split the process time delays approximately in half. This

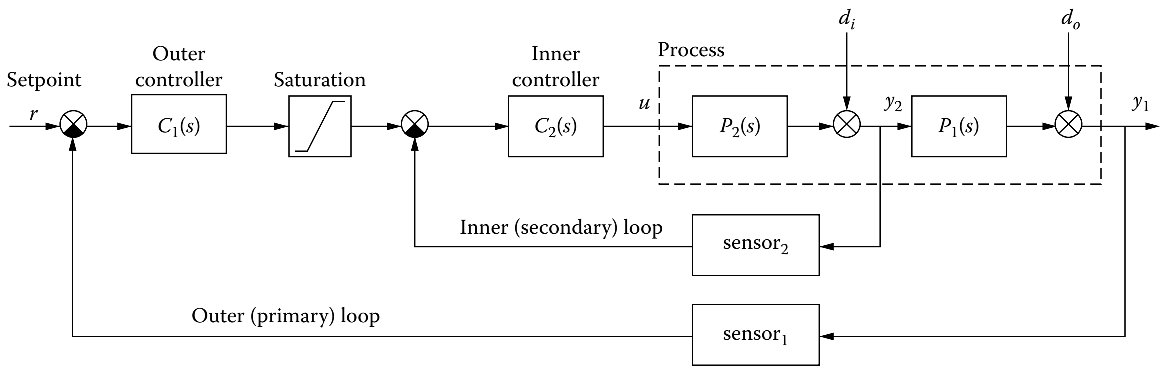


FIG. 2.6b
Block diagram of a cascade control system.

means that the secondary loop should be closed around half of the total time lags in the process. To demonstrate this need, consider two extreme cases:

1. If the secondary variable responded instantly to the manipulated variable (no time delay in the secondary loop), the secondary controller would accomplish nothing.
2. If the secondary loop was closed around the entire process, the primary controller would have no function.

Therefore, the dynamic elements of the process should be distributed as equitably as possible between the two controllers. When most of the process dynamics are enclosed in the secondary loop, that can cause problems. Although the response of the secondary loop is faster than the open loop configuration, in which a secondary controller does not even exist (Figure 2.6c), its dynamic gain is also higher, as indicated by the damped oscillation. This means that if stability is to be retained in a cascade configuration, the proportional band of the primary controller must be wider than it would be without a secondary loop; such de-tuning reduces responsiveness.

The right choice of the secondary variable will allow a reduction in the proportional band of the primary controller when the secondary loop is added because the high-gain region of the secondary loop lies beyond the natural frequency of the primary loop. In essence, reducing the response time of the secondary loop moves it out of resonance with the primary loop.

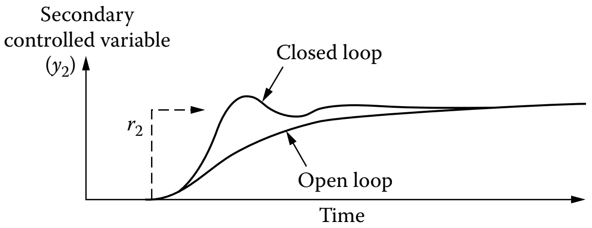


FIG. 2.6c
The response of the secondary variable is much improved when cascade control is used.

Secondary Control Variables

The most common types of secondary control variables are discussed next, discussed in the order of their frequency of application.

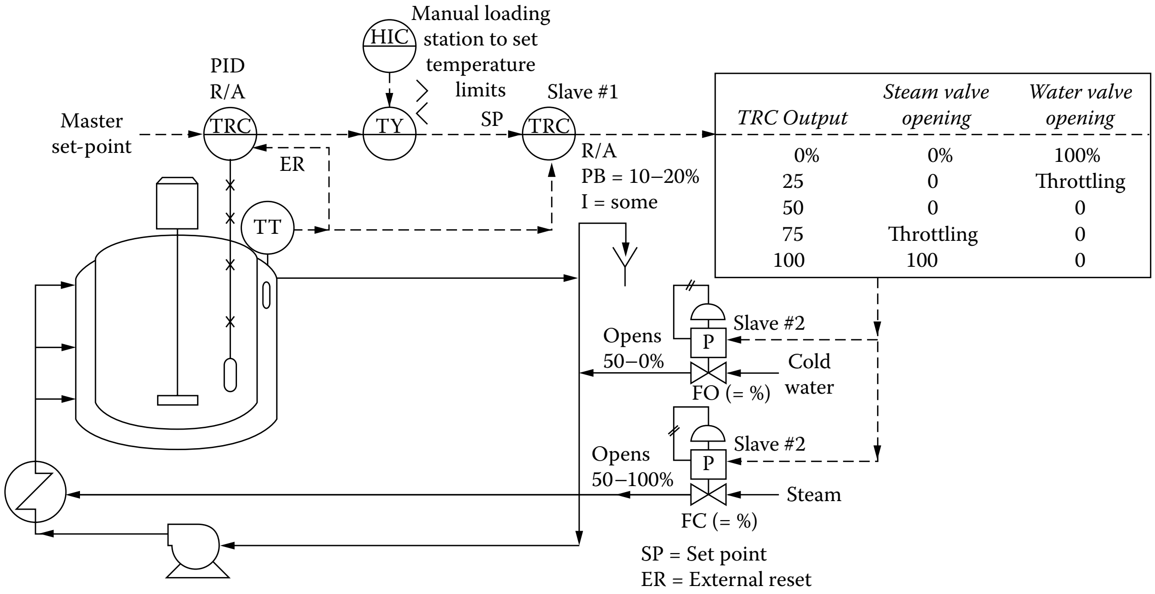
Valve Position Control (Positioner) The position assumed by the plug of a control valve is affected by forces other than the control signal, principally friction and line pressure. A change in line pressure can cause a change in the inner valve position and thereby upset a primary variable, and stem friction has an even more pronounced effect.

Friction produces hysteresis between the action of the control signal and its effect on the valve position. Hysteresis is a nonlinear dynamic element whose phase and gain vary with the amplitude of the control signal. Hysteresis always degrades performance, particularly where liquid level or gas pressure is being controlled with integral action in the controller.

The combination of the natural integration of the process, reset integration in the controller, and hysteresis can cause a "limit cycle" that is a constant-amplitude oscillation. Adjusting the controller settings will not dampen this limit cycle but will just change its amplitude and period. The only way of overcoming a limit cycle is to close the loop around the valve motor. This is what a positioner (a valve position controller) does; and this action can be considered to be the secondary loop in a cascade system.

Flow Control A cascade flow loop can overcome the effects of valve hysteresis as well as a positioner can. It also ensures that line pressure variations or undesirable valve characteristics will not affect the primary loop. For these reasons, in composition control systems, flow is usually set in cascade.

Cascade flow loops are also used where accurate manipulation of flow is mandatory, as in the feedforward systems shown in Section 2.8.



TRC Output	*Steam valve opening*	*Water valve opening*
0%	0%	100%
25	0	Throttling
50	0	0
75	Throttling	0
100	100	0

FIG. 2.6d

When cascade control is provided for a stirred reactor, it is recommended to provide the cascade master with a PID algorithm and with external reset from the measurement of the secondary controller. In this case, there are two inner loops; both can be plain proportional. The gain of the temperature controller is usually 5 to 10, while that of the valve positioner is about 20.

Temperature Control Chemical reactions are so sensitive to temperature that special consideration must be given to controlling the rate of heat transfer. The most commonly accepted configuration has the reactor temperature controlled by manipulating the coolant temperature in cascade. A typical cascade system for a stirred tank reactor is shown in Figure 2.6d.

Cascade control of coolant temperature at the exit of the jacket is much more effective than at the inlet because the dynamics of the jacket are thereby transferred from the primary to the secondary loop. Adding cascade control to this system can lower both the proportional band and reset time of the primary controller by a factor of two or more.

Since exothermic reactors require heating for startup as well as cooling during the reaction, heating and cooling valves must be operated in split range. The sequencing of the valves is ordinarily done with positioners, resulting in a second layer of inner control loops or cascade sub-loops in the system.

In Figure 2.6d, the secondary variable is the jacket temperature and the manipulated variables are the steam and cold water flows. Under control, the secondary variable will always come to rest sooner than the primary variable because initially the controller will demand a greater quantity of the manipulated variable than what is represented by the "equivalent" step change in the manipulated variable.

This is particularly true when the secondary part of the process contains a dominant lag (as opposed to dead time). Because with a dominant lag the gain of the secondary controller can be high, closing the loop is particularly effective. The dominant lag in the secondary loop of Figure 2.6d is the time lag associated with heat transfer across the jacket.

Secondary Control Modes Valve positioners are proportional controllers and are usually provided with a fixed band of about 5% (gain of 20). Flow controllers invariably have both proportional and integral modes. In temperature-on-temperature cascade systems, such as shown in Figure 2.6d, the secondary controller should have little or no integral. This is because reset is used to eliminate proportional offset, and in this situation a small amount of offset between the coolant temperature and its set point is inconsequential. Furthermore, integral adds the penalty of slowing the response of the secondary loop. The proportional band of the secondary temperature controller is usually as narrow as 10 to 15%. A secondary flow controller, however, with its proportional band exceeding 100%, does definitely require an integral mode.

Derivative action cannot be used in the secondary controller if it acts on set-point changes. Derivative action is designed to overcome some of the lag inside the control loop and if applied to the set-point changes, it results in excessive valve motion and overshoot. If the derivative mode acts only on the measurement input, such controllers can be used effectively in secondary loops if the measurement is sufficiently free of noise.

Cascade Primary Loop

Adding cascade control to a system can destabilize the primary loop if most of the process dynamics (time lags) are within the secondary loop. The most common example of this is using a valve positioner in a flow-control loop.

Closing the loop around the valve increases its dynamic gain so much that the proportional band of the flow controller may have to be increased by a factor of four to maintain

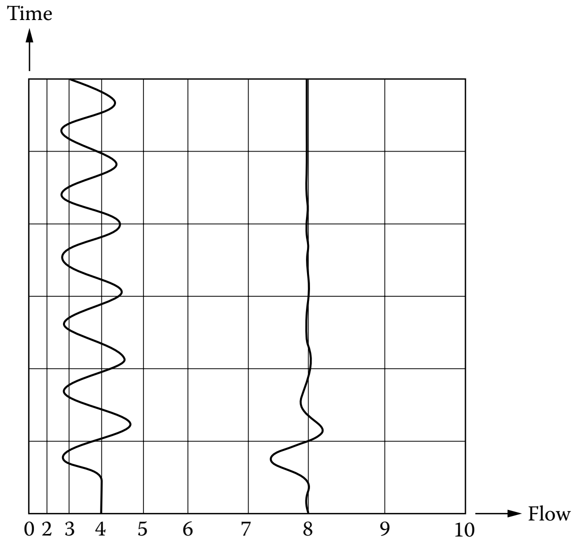


FIG. 2.6e
If cascade flow with a nonlinear orifice sensor is tuned at 80% of range, it will become unstable when the flow drops below 40%.

stability. The resulting wide proportional band means slower set-point response and deficient recovery from upsets. Therefore, in flow control loops, the existence of a large valve motor or long pneumatic transmission lines causes problems, and a volume booster rather than a positioner should be used to load the valve motor.

Instability Instability can also appear in a composition- or temperature-control system where flow is set in cascade. These variables ordinarily respond linearly to flow, but the nonlinear characteristic of a head flow meter can produce a variable gain in the primary loop.

Figure 2.6e compares the manipulated flow record following a load change that has occurred at 40% flow with a similar upset that took place at 80% flow. Differential pressure h is proportional to flow squared:

$$h = kF^2 \qquad \textbf{2.6(1)}$$

If the process is linear with flow, the loop gain will vary with flow, but when h is the manipulated variable, because flow is not linear with h:

$$\frac{dF}{dh} = \frac{1}{2kF} \qquad \textbf{2.6(2)}$$

Thus, if the primary controller is adjusted for 1/4-amplitude damping at 80% flow, the primary loop will be undamped at 40% flow and entirely unstable at lower rates.

Whenever a head-type flow meter provides the secondary measurement in a cascade system, a square-root extractor should be used to linearize the flow signal. The only exception to that rule is if flow will always be above 50% of full scale.

Saturation When both the primary and secondary controllers have automatic reset, a saturation problem can develop. Should the primary controller saturate, limits can be placed on its integral mode, or logic can be provided to inhibit automatic reset as is done for controllers on batch processes.

Saturation of the secondary controller poses another problem, however, because once the secondary loop is opened due to saturation, the primary controller will also saturate. A method for inhibiting reset action in the primary controller when the secondary loop is open (switched to manual) for any reason is shown in Figure 2.6f.

If the secondary controller has integral, its set point and measurement will be equal in the steady state, so the primary controller can be effectively reset by feeding back its own output or the secondary measurement signal. But if the secondary loop is open (in manual), so that its measurement no longer responds to its set point, the positive feedback loop to the primary controller will also open, inhibiting reset action.

Placing the dynamics of the secondary loop in the primary reset circuit is no detriment to control. In fact it tends to stabilize the primary loop by retarding the integral action. Figure 2.6f shows an application of this technique where the primary measurement is at the outlet of a secondary steam superheater, whereas the secondary measurement is at its inlet, downstream of the valve delivering water.

At low loads, no water is necessary to keep the secondary temperature at its set point, but the controllers must be prepared to act should the load suddenly increase. In this example the proportional band of the secondary controller is generally wide enough to require integral action.

When putting a cascade system into automatic operation, the secondary controller must first be transferred to automatic. The same is true in adjusting the control modes, insofar as the secondary should always be adjusted first, with the primary in manual.

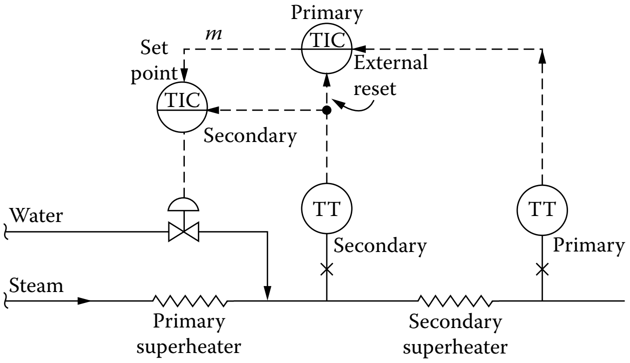


FIG. 2.6f
External reset is provided for the primary controller to prevent integral windup when the secondary controller is in manual.

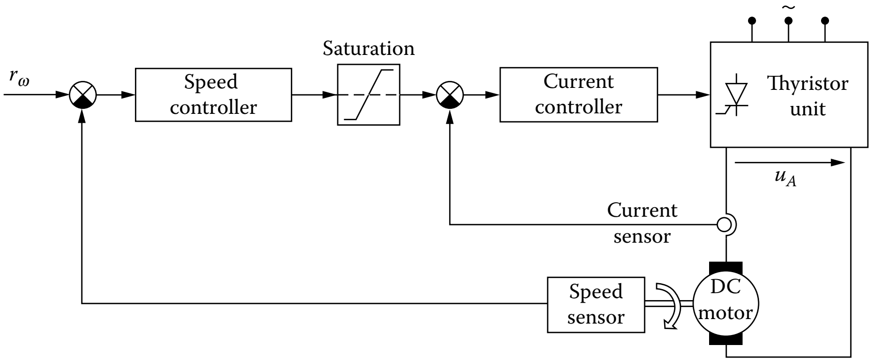


FIG. 2.6g
Cascade control of a DC motor.

Cascade Application Examples

Speed Control of a DC Motor When controlling the speed of a DC motor, the current can reach high peaks during the start-up and shut-down phases or when loading or breaking the motors. Therefore it is usual to apply cascade control with the armature current as the secondary variable. Saturation applied at the current set point (reference value) limits the maximum value of the current. The control system is shown in Figure 2.6g.

Another cascade control example is the control of electrical drives in controlling the position of servomechanisms with velocity feedback. Here the primary measured variable is the position, and the secondary measured variable is the velocity. The manipulated variable is the voltage applied to the servomotor.

Room Temperature Control A possible room temperature control configuration is shown in Figure 2.6h. The room is heated by a steam-heated heat exchanger, which warms the air supply to the room. The manipulated variable is the opening of the steam valve, which determines the steam flow through the heat exchanger. The primary variable is the room temperature. The variation in steam pressure can be the main source of upsets. The secondary variable is the inlet air temperature measured immediately after the heat exchanger, as the disturbances in the steam pressure affect it much sooner than they affect the room temperature.

Adding Valve Positioner to Tank Level Control When the liquid level in a tank is controlled by an outlet valve, placing a positioner on that valve forms a secondary cascade loop. This is desirable because the position of the valve's inner valve is affected by other factors besides the control signal, such as friction and inlet pressure. Variations in the line pressure can cause a change in position and thereby upset a primary variable. Stem friction can have an even more pronounced effect.

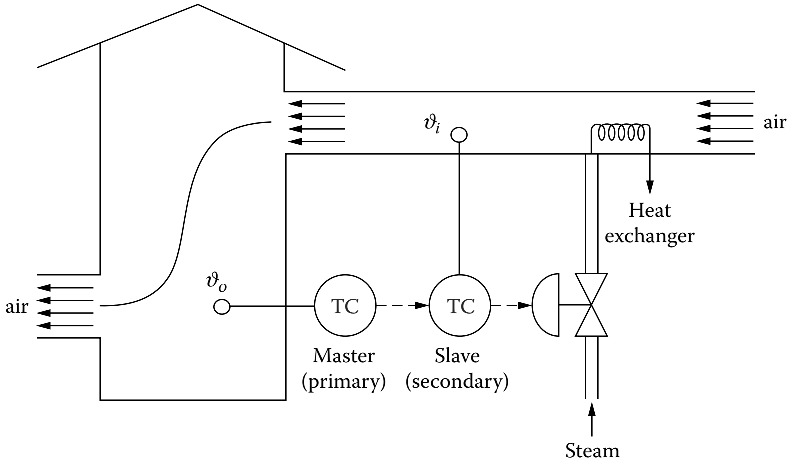


FIG. 2.6h
Cascade control of room temperature.

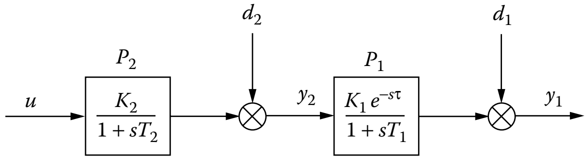


FIG. 2.6i
The process controlled by a cascade control system consists of two segments (P_1 and P_2) and has a measurable inner variable (y_2).

Inner valve sticking due to friction produces a square-loop hysteresis between the action of the control signal and its effect on the valve position. Hysteresis always degrades performance, particularly where liquid level is being controlled with integral action in the controller. The combination of the integrating nature of the process and the integral action in the controller with hysteresis causes a "limit cycle" that results in a constant-amplitude oscillation.

Adjusting the controller's tuning settings will not dampen this limit cycle but will only change its amplitude and period of oscillation. The only way of overcoming this limit cycle is to add a positioner and close the position-controlling slave loop around the valve motor.

Composition Control In composition control systems, flow is usually set in cascade. A cascade flow loop can overcome the effects of valve hysteresis. It also ensures that line pressure variations or undesirable valve characteristics will not affect the primary loop.

Cascade Controller Design and Simulation

The process controlled by a cascade system is shown in Figure 2.6i. The nominal values of the parameters and the uncertainties in their values are as follows:

$$K_{1n}=1,\quad 0.8<K_1<1.2 \quad K_{2n}=1,\quad 0.8<K_2<1.2$$
$$\tau_n=20,\quad 18<\tau<22 \quad T_{2n}=4,\quad 3<T_2<5$$
$$T_{1n}=10,\quad 8<T_1<12$$

The dual design objectives are:

1. First design a single-series PID controller, with feedback taken from the output signal y_1. Design objectives are stability and good reference signal (set point) tracking properties characterized by zero steady-state error and about 60° of the phase margin to keep the overshoot below 10%, when a step change is introduced in the set point.
2. Design a cascade control system with feedback taken from the primary and secondary (outer and inner) controlled variables (process outputs) y_1 and y_2, respectively. Both control loops have to be stable. The design objectives for tracking the set point (reference signal) are the same as before. Fast response is also required to any upsets in the secondary loop or measurement.

The design should consider the theoretical process model, and the effects of the possible errors or uncertainties in the model should be checked.

Single-Series Primary Controller The block diagram of the single control loop is shown in Figure 2.6j. To ensure accurate tracking of the set point (step reference signal), the controller must have an integral mode. The addition of rate action can accelerate the system response. A PID controller combines these two effects. With pole-cancellation technique the suggested PID controller transfer function is

$$C(s)=A\frac{(1+10s)(1+4s)}{10s(1+s)} \qquad \textbf{2.6(3)}$$

The ratio of the time constants of the PD part was chosen to be four.

The loop transfer function is

$$L(s)=\frac{A}{10s(1+s)}e^{-20s} \qquad \textbf{2.6(4)}$$

Gain A of the controller has to be chosen to ensure the required phase margin. In case of systems with dead time τ, a phase margin of about 60° is obtained if the cut-off frequency ω_c in the loop's Bode amplitude-frequency diagram is at slope of –20 dB/decade and is located at about $1/2\tau$.

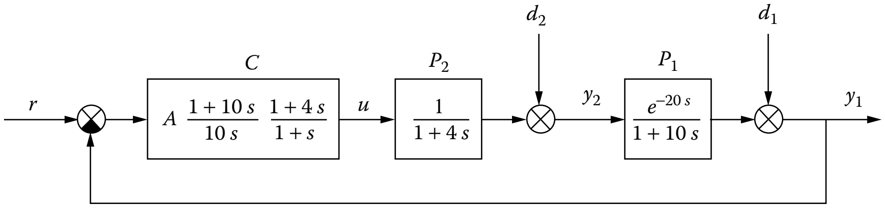


FIG. 2.6j
Series compensation.

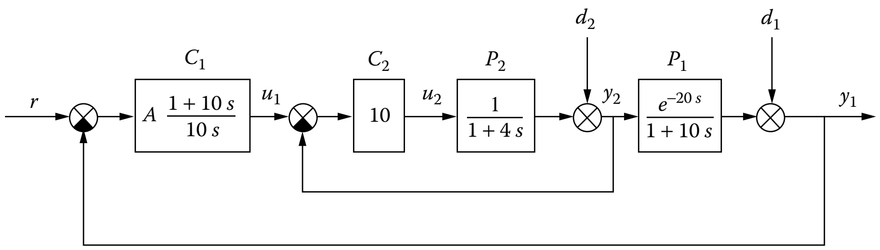


FIG. 2.6k
Cascade compensation.

In our case $\omega_c \sim 1/40$. At the cut-off frequency the absolute value of the frequency function is 1, thus $A/10\omega_c = 1$. Hence $A = 0.25$.

Cascade Controller System Design The block diagram of the control system is shown in Figure 2.6k. First the inner or secondary control loop is designed for fast correction of upsets that might occur in the secondary loop. The inner (secondary) controller can be proportional only. Its gain can be 10 and the cut-off frequency of the secondary loop can be 2.5. In that case the settling time is expected around $3/\omega_c$.

The resulting transfer function of the inner closed loop between signals y_2 and u_1 is

$$\frac{0.909}{1+0.36s} \qquad \mathbf{2.6(5)}$$

In the primary (outer) loop a PI controller is used, which will be responsible for accurate steady-state tracking of a step change in the set point (reference signal). Again, pole-cancellation technique is applied. The transfer function of the controller is

$$C_1(s) = A\frac{1+10s}{10s} \qquad \mathbf{2.6(6)}$$

The loop transfer function of the outer loop is

$$L(s) = A\frac{0.909}{10s(1+0.36s)}e^{-20s} \qquad \mathbf{2.6(7)}$$

The cut-off frequency is chosen again to be $\omega_c = 1/40$. At this frequency the absolute value of the frequency function is 1.

$$\frac{0.909A}{10\omega_c} = 1, \quad \text{hence } A = 0.275 \qquad \mathbf{2.6(8)}$$

Simulation Results The set-point (reference signal) change is a unit step that occurs at a time point of 10 sec. A unit step inner disturbance is added at a time point of 250 sec. The simulation ends at $t = 500$ sec.

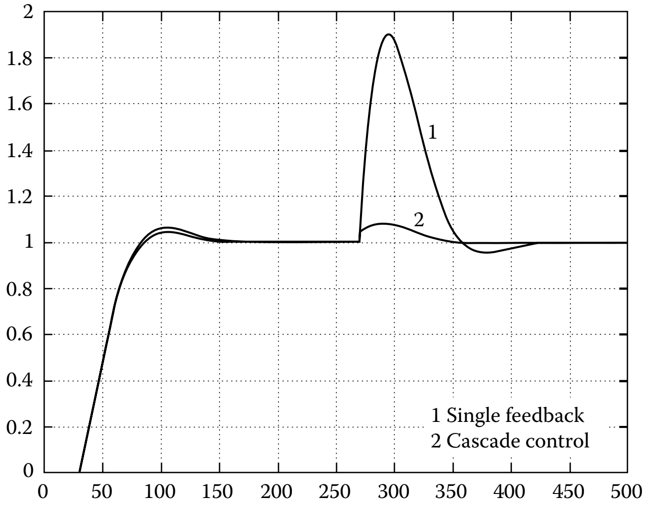


FIG. 2.6l
Tracking a unit change in set point that occurred at t = 10 sec and correcting for an upset of unit step size that occurred at t = 250 sec. Response 1 is that of a single feedback controller and 2 is that of a cascade system.

With the assumed nominal process model described earlier, the controlled variable responses (output signals) are shown in Figure 2.6l. It is seen that the set-point (reference signal) tracking is practically the same for both control configurations. On the other hand, the response to upsets (disturbance rejection) is much better with cascade control.

Figure 2.6m, in its upper part, shows the manipulated variable (u_2, controller output signal) and in the lower part, the controlled variable (y_2, inner variable) response of the secondary loop in the cascade configuration. The peak in the manipulated variable (secondary control signal) ensures the fast disturbance compensation. If the slave controlled variable (inner process variable) maximum value has to be under a given limit, this can be achieved by artificially saturating the output of the primary controller.

Figure 2.6n shows the effect of model parameter uncertainties in cases of the single loop and the cascade system. The controllers are designed based on the same nominal process model, and their responses to upsets are compared when the real process dynamics differ from the assumed

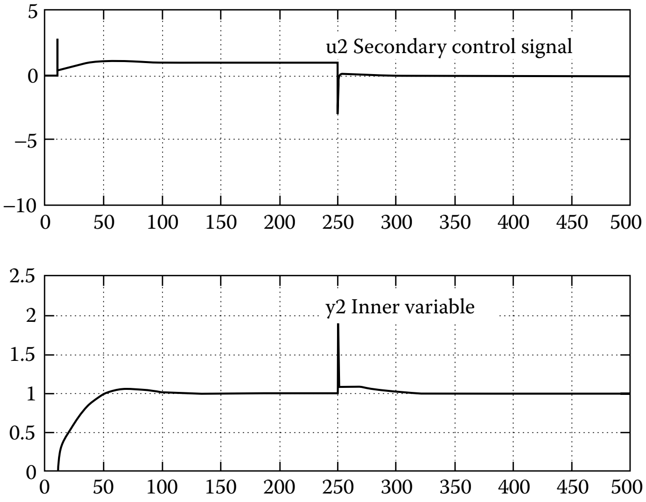


FIG. 2.6m
The response of the manipulated variable (u_2, controller output signal) and the controlled variable (y_2, inner variable) of the secondary loop in the cascade configuration.

nominal one that was used during tuning of the controllers (dynamics slower or faster) by various degrees.

Cascade control is more tolerant of model parameter uncertainties or errors in the model than a single controller and it gives particularly better performance when process upsets occur, as its disturbance rejection capability is superior to that of a single-loop controller.

Figure 2.6o describes the controlled variable responses of the single-loop and cascade control systems, assuming all dynamic parameters (time constants) of the process and its gain are both at their highest values.

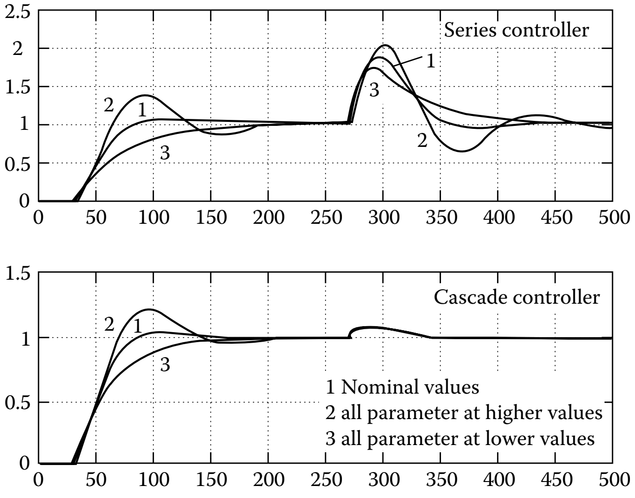


FIG. 2.6n
The response to both a set point step and a process upset of a single-loop controller (top) and a cascade system (bottom) to changes in the process dynamics. (1, correctly tuned; 2, process faster than at time of tuning; 3, process slower).

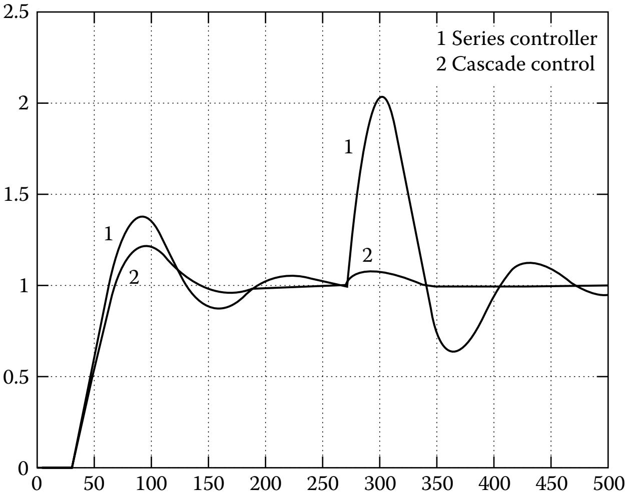


FIG. 2.6o
The controlled variable responses (output signals) of the single loop (series controller) and the cascade control system, when the time constants and the gain of the process are at their highest values.

SUMMARY

Cascade loops consist of two or more controllers in series and have only a single, independently adjustable set point, that of the primary (master) controller. The main value of having secondary (slave) controllers is that they act as the first line of defense against disturbances, preventing these upsets from entering and upsetting the primary process.

For example, in Figure 2.6d, if there were no slave controller, an upset due to a change in steam pressure or water temperature would not be detected until it had upset the master measurement. In this configuration, the cascade slave detects the occurrence of such upsets and immediately counteracts them, so that the master measurement is not upset and the primary loop is not even aware that an upset occurred in the properties of the utilities.

In order for the cascade loop to be effective, it should be more responsive (faster) than the master. Some rules of thumb suggest that the slave's time constant should be under 1/4 to 1/10 that of the master loop and the slave's period of oscillation should be under 1/2 to 1/3 that of the master loop.

The goal is to distribute the time constants more-or-less evenly between the inner (slave or secondary) and outer (master or primary) loops, while making sure that the largest time constant is *not* placed within the inner loop. When that occurs, such as in the case where a valve positioner is the slave and a very fast loop such as a flow or liquid pressure controller is the master, stability will be sacrificed because the valve has the largest time constant in the loop. In such configurations, stability can be regained only at the cost of reduced control quality (sluggish response to load or set-point changes).

Therefore, in such cases one would try to either speed up the valve or avoid the use of cascade loops. If the reason for using a positioner is to increase the air flow to the actuator, one can replace the positioner with a booster relay.

Providing external reset (Figures 2.6d and 2.6f) for the cascade master from the slave measurement is always recommended. This guarantees bumpless transfer when the operator switches the loop from automatic control by the slave to full cascade control. The internal logic of the master controller algorithm is such that as long as its output signal (m) does not equal its external reset (ER), the value of m is set to be the sum of the external reset (ER) and the proportional correction ($K_c e$) only. When m = ER, the integral mode is activated, and in case of a PI controller, the output is:

$$m = K_c(e + 1/T_i e\,dt) + b \qquad \textbf{2.6(9)}$$

Bibliography

Åström, K. J., and Hägglund, T., *PID Controllers: Theory, Design and Tuning,* Research Triangle Park, NC: Instrument Society of America, 1995.

Buckley, P. S., *Techniques of Process Control,* Englewood Cliffs, NJ: John Wiley & Sons, 1964.

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: Instrument Society of America, 1990.

Gibson, J. E., *Nonlinear Automatic Control,* New York: McGraw-Hill, 1963.

Goodwin, G. C., Graebe, S. F., and Salgado, M. E., *Control System Design,* Englewood Cliffs, NJ: Prentice Hall, 2001.

Harriott, P., *Process Control,* New York: McGraw-Hill, 1964.

Jones, B. E., *Instrumentation, Measurement and Feedback,* New York: McGraw-Hill, 1977.

Levine, W. S. (Ed.), *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Luyben, W. L., and Luyben, M. L., *Essentials of Process Control,* New York: McGraw-Hill, 1997.

Morari, M., and Zafiriou, E., *Robust Process Control,* Englewood Cliffs, NJ: Prentice Hall, 1989.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* New York: John Wiley & Sons, 1989.

Shinskey, F. G., *Process Control Systems: Application, Design and Tuning,* 4th ed., New York: McGraw-Hill, 1998.

Skogestad, S., and Postlethwaite, I., *Multivariable Feedback Control,* New York: John Wiley & Sons, 1996.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

2.7 Empirical Process Optimization

D. W. HOYTE (1995) **B. G. LIPTÁK** (2005)

Suppliers of Optimization Software: DMC; Gensym; Setpoint; Statistical Programs; Ultramax ($1000 to $20,000)

The optimization of control loops is discussed in Section 2.20, and the optimization of such unit operations as chemical reaction or distillation is covered in Chapter 8. Here the subject is the optimization of empirical processes, but before that discussion, it is appropriate to say a few words about optimization in general.

LEVELS OF OPTIMIZATION

The highest level of optimization is business- or enterprise-wide optimization, which includes not only the manufacturing process, but also the optimization of the raw material supply chain and of the packaging and product distribution chain. This is a higher and more important level of optimization than process optimization alone because it requires the simultaneous consideration of all three areas of optimization and requires the finding of enterprise-wide operation strategies that will keep all three areas within their optimum areas of operation (Figure 2.7a).

Plant-wide optimization also involves more than the optimization of the unit processes because it must also consider documentation, maintenance, scheduling, and quality management considerations. Plant-wide optimization involves the resolution of the sometimes conflicting objectives of the many unit operations and the envelope strategies required to optimize the entire plant.

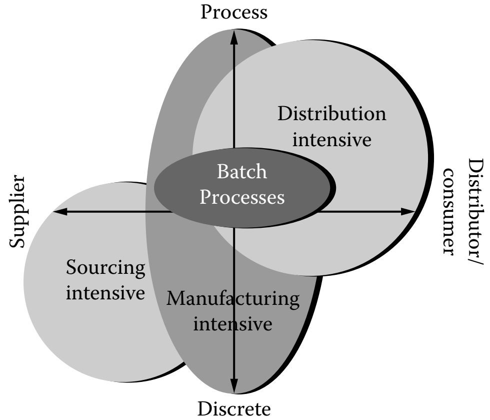


FIG. 2.7a
Enterprise-wide optimization requires the optimization of not only the manufacturing plant but also the raw material supply chain and the product distribution chain.

At the unit operations level, it goes without saying that multivariable optimization cannot be achieved when individual processing equipment is defective or when control loops are not properly tuned. In addition, it is important that measurements be sampled fast enough and that loops be tuned for sufficiently fast rates of recovery. Loop cycling must also be eliminated, which usually requires the elimination or correction for interactions between loops. When no mathematical model can describe a process and therefore the process can only be optimized experimentally, empirical optimization is required. This is the subject of this section.

EMPIRICAL OPTIMIZATION

The performance of some processes can be described by theoretical mathematical models. In these cases, mathematical techniques are available to determine the best operating conditions to satisfy selected performance criteria and thereby optimize the process.[1] There are, however, many processes for which no adequate mathematical model exists. These are known as *empirical* processes, and they can only be optimized experimentally. This section discusses approaches to finding the optimum process settings by experiment. The discussion concentrates on the Ultramax method, which is appropriate to a wide variety of processes and for which excellent user-friendly computer tools are available.

Optimization

The term "optimize" is used here as it refers to a measurable criterion or quality of the process (Q) that the user wishes to maximize or minimize. For example:

1. Plant production rate is to be maximized.
2. Cost-per-ton of output, calculated from many measurements, is to be minimized.

3. Flavor of a product, rated 1 to 10 by a group of tasters, is to be maximized.
4. Height/width ratio of a chromatograph peak is to be maximized.

Q is sometimes referred to as the target variable or the objective function. Several contending criteria may be given weights to produce a single measurement of process performance.[2] To simplify the discussion, the examples here will be limited to maximizing the value of Q.

Optimizing the process may include maximizing Q while other measured results of the process (R, S, etc.) remain within specified limits (e.g., maximizing plant production rate in making butter, paper, or sulfuric acid must be subject to a specified limit on water content in each product).

Information Cycle Times Some processes have inherently long information cycle times; resistance to weathering of different paint formulations or increase in farmland productivity from different mixtures of synthetic fertilizers are examples. In both cases, the search for an optimum would take an unacceptably long time if the results from one test were awaited before deciding the conditions for the next test. Therefore, multiple formulations of paints and fertilizers must be planned and then tested in parallel at the same time. This type of planning, and the subsequent analysis of the results, is often called parallel design of experiments.

Parallel Design Taking a specific example: A material is baked for time duration (D), at an oven temperature (T), to result in a quality (Q), which is to be maximized. This process could be, for example, the case-hardening of steel, the polymerization of resin, or the diffusion of ions in silicon chips. Only D and T are manipulated to affect Q.

Processing is performed with a selected array of D and T values, and the corresponding values of Q are measured. The optimum can be found graphically by drawing contours of Q on a graph showing the points ($D1$, $T1$), ($D2$, $T2$), and so on. Optimum operating conditions for the process are the values of D and T at the top of the "hill."

The contours of Q are sometimes called the response surface of the process and can be fitted to the data points mathematically, instead of sketching contours by eye.[3] This has clear advantages when there are many manipulated variables in the process. The math is simplified by selecting the array of experimental points to suit it. Also, the analysis is simplified by limiting the equation for the response surface to the second order.

Note that while statistical texts often refer to D and T as the *controlled* variables, an engineer will usually refer to D and T as *manipulated* variables and to Q as the *controlled* variable.[4]

Some commercially available computer programs for parallel designs of experiments are in the nature of a collection of routines that are useful to an experienced industrial statistician. The practicing engineer is likely to work faster with a user-friendly, menu-driven package such as Catalyst or Discover.[5]

Sequential Design Many processes have short information cycle times. In such cases, the results from using particular process settings become available in time to help in deciding the settings for the next test. This method is often called sequential design of experiments, or hill climbing.

Using the previous example, three different values for the settings D and T are evaluated in this process. The slope of the response surface defined by the values of $Q1$, $Q2$, and $Q3$ shows the direction "up the hill" to increase Q. The next values of D and T are selected to proceed in this direction. Repeating this procedure, the experimenter can "hill-climb" step by step, toward the maximum value for Q.

If the process has many independent variables and also has some noise that is affecting the results, it can be difficult to determine how each variable is affecting the quality of the product (Q). The Simplex search procedure works (finds the maximum value of Q) all the time, with one more process result than the number of independent variables. In the example used here, it would require three process results.

Figure 2.7b illustrates the method. To find the next settings for the process, a line is drawn from the settings with the lowest result, through the midpoint of the opposite face, to an equal distance beyond. In this example, the search reaches optimum with 20 data points.

The Simplex search was invented by William Spendley at ICI Billingham, England to automate the evolutionary operation (EVOP) procedures of G. P. Box and others, and the methods are amenable to implementation using paper and pencil or a desk calculator.[6,7] Reference 8 provides a practical treatment of the subject in which the authors review related computer programs.

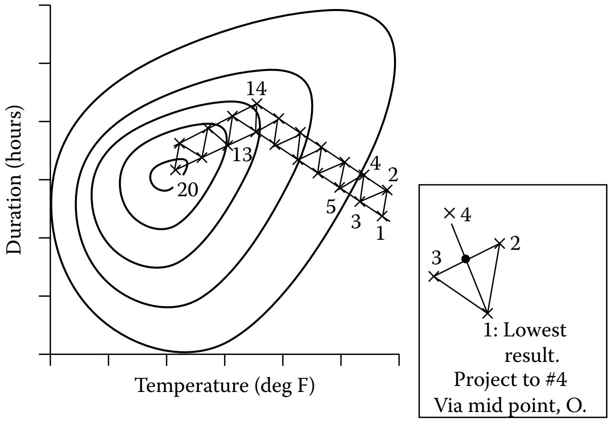


FIG. 2.7b

SIMPLEX search for a maximum of quality, Q. Contours of Q are shown, though usually unknown. Inset shows how the next process settings are found. Direction changes where #14 has lower Q than #13.

```
Main menu of ultramax functions.

    F  Formulate problem...
    E  Enter run data...
    L  Build models...
    A  Advice...
    W  What-if analysis...
    D  Data management...
    H  Historical data report...
    P  Plot data...
    Q  Quit ultramax.

      *** Enter function-->–

            Problem formulation:
   find duration and temperature settings
 to give max quality subject to limits on volume
```

VAR #	Name	Units	Type	Prior Region LO	HI	Constraints LO	HI
1	Temperature	Deg. F	1	450.	500.		
2	Duration	Hours	1	6.	7.		
3	Volume	Cu.In.	5	30.	31.	30.	31.
4	Quality	Score	6	5.	6.		

FIG. 2.7c

Main menu and a Problem Formulation for the Ultramax program. In the formulation, "Type" designates treatment of the variable. Note that all Ultramax reports have been simplified here to highlight concepts. (Courtesy of Ultramax Corp., Cincinnati, OH.)

Bayesian Methods Bayesian statistics are designed to match the human decision-making process in situations where few data are available. They are used to help in selecting the next process settings in a sequential hill-climbing program, Ultramax,[9] designed for the PC or mainframe computers. Ultramax is a menu-driven, engineer-oriented package.

The following illustration uses the baking process mentioned earlier, but with the addition of a constraint: volume (V) per unit of product. Here, volume and quality are both affected by duration and temperature of baking; the goal is to find settings of D and T to give maximum Q while holding V between specified limits.

Figure 2.7c shows the Main Menu of Ultramax. Keying F prompts the user to specify the "Problem Formulation" shown there. (Note that all examples shown have been simplified to clarify the concepts.) Salient points of this formulation are as follows:

1. Q is only acceptable when V is ranged from 30 to 31 cu. in.
2. TYPE code specifies how the program is to treat each variable, as follows:
   Types 0 and 4 are for data-recording only. (Types 0, 2, and 4 are not used in this example.)
   Type 1 is an independent (manipulated) variable.
   Type 2 is an independent variable that affects the target but cannot be manipulated (e.g., the time of day, the price of crude oil, a process disturbance).
   Type 5 is a dependent (*results*) variable with constraints.
   Type 6 is the target variable.
3. One Type 6 target variable must be specified in the formulation; several variables of other types may be specified. The standard version of Ultramax can deal with ten Type 1 variables, and *all* can be changed at will in each step of the search.
4. The "Prior Region" specifies the range where the user expects the optimum to lie and wishes the search to start. Note: A *parallel* design of experiments for two independent variables requires a dozen data points in this region and in every further region selected for exploration.

Providing Process Data

After the formulation of the problem, the menu function E (Enter Data) is used next. The user can enter, at one time, one or many data points: for example, sets of measurements of D, T, V, and Q from the process.

Process data can be keyed in or can be read in from a file. Although users often have extensive data to enter from prior process runs, they may wish to obtain and enter data points one at a time after the initial loading in order to obtain advice from Ultramax on the recommended process settings to use next. This way the optimum will be reached with the fewest runs of the process.

The search up through run #14, following the advice of Ultramax, is shown in Figure 2.7d. This starts with the same three data points used in Figure 2.7b and explores the contours

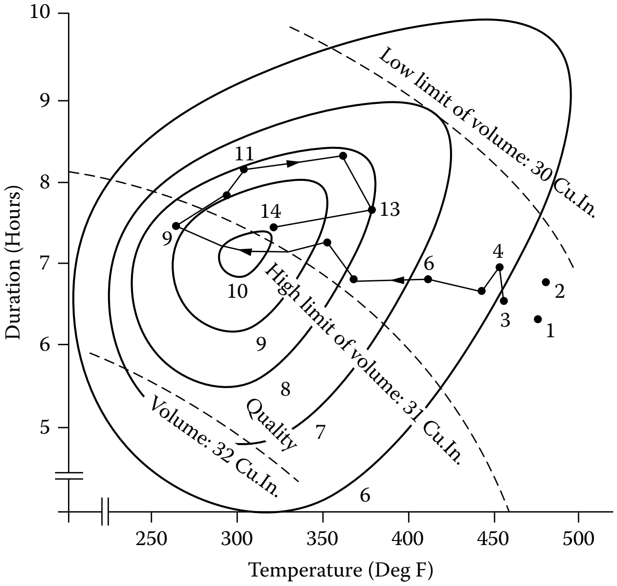


FIG. 2.7d

Ultramax search for process settings to give maximum quality (Q) without violating limits on volume (V). The contours are not known before the search.

of Q as it seeks the maximum. At the same time, the search is recognizing the contours of V and locating the barrier of $V = 31$ cu. in. By run #14, the search is close to the optimum settings for duration and temperature in this process. (The contours are shown for illustration only; they are not, of course, known while the search is being made.)

Search Procedure

After process data have been entered, menu function L (Build Model), is used to fit a second-order model to the data points. This is illustrated in the "Coefficients Report for Quality," which appears after the data for run #13 have been entered (see Figure 2.7e). (A similar report obtained for the model of V is not shown here.)

The user seeks process settings that give the largest value of Q and is only interested in fitting the model in that region. The data points used earlier in the search may detract from that fit; if they do, Ultramax automatically discards them. The actual points used in the model are listed at the bottom of Figure 2.7e together with their errors. "Minimum Prediction Error" at the top of Figure 2.7e gives a single value for goodness-of-fit.

Model coefficients for variable #4: quality.
problem: maximum quality subject to limits on volume

Model type: local fixed-point-centered quadratic

Minimum prediction error = 0.26 score

Model Equation:

$$F(x) = +12.28 - 0.0392 \ {}^*X(1) + 1.139 \ {}^*X(2) - 0.00006016 \ {}^*X(1)^{**}2 + 0.009428 \ {}^*X(2)^*X(1) - 0.289 \ {}^*X(2)^{**}2$$

Key	Variable	Units	Type
F(x)	Quality	Score	6
X(1)	Temperature	Deg F	1
X(2)	Duration	Hours	1

Datapoints Used and the Errors of the Model:

Run #	Actual quality	Model quality	Error
13	8.00	7.97	−0.03
12	8.10	8.06	−0.04
11	8.00	8.19	0.19
10	8.76	8.43	−0.33
9	8.62	8.70	0.08
8	8.09	8.16	0.07
7	7.78	7.70	−0.08
6	6.71	6.74	0.03

FIG. 2.7e
Example of a Model Coefficients Report: Run #13. Goodness-of-fit is shown by minimum prediction error and list of model errors. Ultramax fits the model in the region of the maximum quality, as desired by the user. It has discarded outlying data points numbers 1 to 5. (Courtesy of Ultramax Corp., Cincinnati, OH.)

Advice Report: Seeking Maximum Quality Subject to Limits on Volume

	Optimum	Advice	Advice	Advice
Advice #	14	15	16	17
Process Settings				
1 Temperature Deg. F	303	321	324	359
2 Duration Hours	7.44	7.52	7.86	9.27
Expected Results				
3 Volume Cu.In.	31.00	30.91	30.81	30.33
Expected Error	0.19	0.19	0.19	0.39
4 Quality Score	8.6	8.5	8.4	7.6
Expected Error	0.6	0.6	0.6	1.2
Standard distance ("Boldness")	1.1	1.0	1.0	2.8

FIG. 2.7f
Example of an "Advice Report" showing Advices for the next settings to run the process after run #13. "Optimum" gives the best process settings based on present data. "Standard Distance" is the estimated risk in the results from each Advice (similar to confidence limits). Maximum Standard Distance has been set by the user at 3.0. (Courtesy of Ultramax Corp., Cincinnati, OH.)

After the model-building step, the menu function A (Generate Advice) gives a report for a selected number of "Advices," as shown in Figure 2.7f. The column headed "Optimum" gives the best calculated settings for the process, based on the data at that stage. The user can then follow Advices for the next process settings to explore further for the optimum, in the priority order that Ultramax sees as the need for information. The user may choose to select several or only one of the Advices for the next process settings, or the user can use settings to explore the process based solely on the user's own knowledge and intuition about the process. Ultramax will use any data points it is given.

Finding the Optimum Sequential cycles consist of:

Data Entry
Build Model
Get Advices
Run Process
Data Entry

and will initially lead to data points in the general region of the process maximum. In this phase of the search, the size of steps made in changing the process settings is limited by a "Boldness-of-Search" parameter that the user specifies. Thus the user can choose either to search quickly or to make smaller and more cautious changes. This is somewhat similar to tuning one of the process control loops. The "boldness" of each Advice is ranged from 0 to 3 and is shown as "Standard Distance" in Figure 2.7f.

The search continues until the distinction between the best process settings is limited only by the repeatability (*noise*) of the process. At this point, changes in the "Optimum" value from one process-run to another for each manipulated variable (Figure 2.7f) will bottom out.

Process Sensitivity The model can be explored with menu function W (What-If Analysis). This evaluates Q and V from the models for any desired settings of D and T, showing the sensitivities of Q and V to changes in each independent variable.

Ultramax may be called periodically by a process control program such as The Fix or Onspec (see References 10 and 11) to calculate new values for the setpoints of manipulated variables. The program will perform a reasonableness check, then transmit these setpoints to its own process-control algorithm or to hardware controller setpoints.

Process Repeatability Where the repeatability of the process is not adequate for analyzing for optimum, the following steps can be considered:

1. Average many measurements of a variable to reduce "noise."
2. Improve repeatability of the measuring devices through recalibration and other means.
3. Seek further variables that affect the process and are not yet in the model.
4. Do a time-series analysis to investigate interactions and delayed actions.[12]

Programs are available to make this analysis.[13] A user-friendly interface was written for them, but it is not now available in the United States.[14] However, some control-system contractors, such as the ACCORD system, offer time-series analysis.[15,16]

CONCLUSIONS

The techniques of optimizing an empirical process owe much to the insights and work of the late James M. Brown.[17] These methods were once the sole preserve of the specialist statisticians but are now available in forms that are suitable for use by the practicing engineer.

Industrial user companies seldom advertise the successes they are achieving by the use of these tools, but information and assistance are readily available from the suppliers of optimization software.

References

1. Reklaitis, G. V., et al., *Engineering Optimization: Methods and Applications,* New York, Wiley, 1983.
2. Derringer, G., and Suich, R., "Simultaneous Optimization of Several Response Variables," *Journal of Quality Technology,* October 1980, pp. 214–219.
3. Myers, R. H., "Response Surface Methodology," Virginia Polytechnic Institute and State University, Blacksburg, 1976.
4. Cubberly, W. H. (Ed.), *Comprehensive Dictionary of Instrumentation and Control: Controlled Variable,* Research Triangle Park, NC: Instrument Society of America, 1988.
5. Catalyst and Discover, from BBN Software Products, 150 Cambridge Park Drive, Cambridge, MA 02140.
6. Spendley, W., et al., "Sequential Application of Simplex Designs in Optimization and Evolutionary Operation," *Technometrics,* November 1962, pp. 441–461.
7. Box, G. E. P., "Evolutionary Operation: A Method for Increasing Industrial Productivity," *Applied Statistics,* No. 6, 1957, pp. 3–23.
8. Deming, S. N., et al., *Sequential Simplex Optimization: A Technique for Improving Quality and Productivity in Research, Development and Manufacturing,* Boca Raton, FL: CRC Press, 1991.
9. Moreno, C. W., "Self-Learning Optimization Control Software," Instrument Society of America, ROBEXS '86 Conference, Houston, TX, June 1986. (Available from Ultramax Corp., 1251 Kemper Meadow Drive, Cincinnati, OH 45240.)
10. The Fix, from Intellution, Inc., 315 Norwood Park South, Norwood, MA 02062.
11. Onspec, from Heuristics, Inc., 9845 Horn Road, Mail Stop 170, Sacramento, CA 95827.
12. Findley, D. F., *Applied Time Series Analysis, II,* New York: Academic Press, 1981.
13. Akaike, H., et al., "Time Series Analysis and Control: TIMSAC-74, -78 and -84 Program Packages," Computer Science Monographs, Nos. 5, 6, 11, 22, and 23, Institute of Statistical Mathematics, 4–6–7 Minami-Azabu, Minato-ku, Tokyo, Japan 106, March 1985.
14. Schneider, P., private communication, January 1992.
15. Nakamura, H., and Uchida, M., "Implementation of Optimum Regulator at Supercritical Thermal Power Plants," Conference on Power Plant Controls and Automation, Miami, Florida, February 7–9, 1989. Proceedings available from Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, CA 94304.
16. Nakamura, H., et al., "Optimum Regulator Implementation by Microprocessor-Based Distributed Control System," Conference on Power Plant Controls and Automation, Miami, Florida, February 7–9, 1989. Proceedings available from Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94304.
17. Brown, J. M., University of Manchester Institute of Science and Technology, private communications, 1986–1989.

Bibliography

ARC Report, "Batch Process Automation Strategies," ARC Advisory Group, October 1999.

Box, G. E. P., et al., *Statistics for Experimenters,* New York: Wiley, 1978.

Draper, N. R., and Smith, H., *Applied Regression Analysis,* New York: Wiley, 1981.

Ghosh, A., "Maximizing the Potential of Batch Process Control," presented at WBF Conference, Brussels, October 2, 2000.

Luyben, W. L., et al., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Nachtsheim, C. J., "Tools for Computer-Aided Design of Experiments," *Journal of Quality Technology,* Vol. 19, No. 3, July 1987, pp. 132–159.

Ruel, M., "Stiction: The Hidden Menace," *Control Magazine,* November 2000.

Savas, E. S., *Computer Control of Industrial Processes,* New York: McGraw-Hill, 1965, pp. 55–77.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Smoak, R. A., "Control Design Methodology for Power Plants," EPRI Conference, 1989.

Storm, M. E., "Sequential Design of Experiments with Physically Based Models," Master's Thesis, Dept. Mech. Engineering, MIT, April 1989.

Wong, K. Y., et al., "Computer Control of the Clarksville Cement Plant by State Space Design Method," Cement Industry Conference of IEEE in St. Louis, Missouri, May 1968.

2.8 Expert Systems

B. A. JENSEN (1995) **B. JOSEPH** (1995) **B. G. LIPTÁK** (2005)

Artificial intelligence is the field in which computers model or make available nonnumeric human-like intelligence. Neural networks, like expert systems, fuzzy logic, robotics, natural language processing, and machine vision, are parts of an overall technology umbrella known as artificial intelligence (AI). Most of these techniques are discussed in this section, while model based control (MBC), model free control (MFC), multivariable (MIMO) control, and statistical process control are also covered in other sections of this chapter.

ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is a field of computer science concerned with the development and deployment of machines that mimic human behavior. Broadly speaking, AI can be divided into five major categories: expert systems, robotics, vision, natural language processing, and artificial neural networks (ANN). Of these five, the expert systems and ANNs have had the greatest impact on process control.

Expert Systems

Expert systems are computer software systems that mimic the tasks routinely carried out by experts. Usually, the domain of expertise of these systems is restricted to a well-defined area. Unlike humans, these systems lack common sense and hence cannot extrapolate their knowledge much beyond their narrow domain of expertise. Expert systems can be thought of as a storehouse of knowledge with efficient ways of retrieving this knowledge to solve the problem at hand.

Expert systems are valuable because they can serve as a permanent storehouse of knowledge that is widely accessible. Expert systems have been applied in process control in such areas as fault diagnosis, quality control, control system synthesis, statistical process control, alarm management, and process troubleshooting. Some companies specialize in the application of AI technology to process control.

Structure of Expert Systems Expert systems mimic the tasks carried out by experts. They encode the relevant knowledge that relates to a problem in a form that is suitable for manipulation. The structure of the expert system separates the knowledge from the mechanism for manipulating that knowledge. This has significant advantages because it makes the system easier to maintain and update. It is this separation of the knowledge base from the reasoning process that makes expert systems different from other software packages. Figure 2.8a shows the typical structure of an expert system.

The type of knowledge in the knowledge base depends on the problem domain. Knowledge is stored in the form of *facts* (e.g., as temperature goes up reaction rate increases), *rules* (e.g., if temperature in the reactor is high, then the cooling system is faulty), and *data frames* (frames are data structures for storing information in a hierarchical format). Knowledge may be also contained in *objects* that can perform specific tasks (e.g., do an energy balance). Objects may be coded in other programming languages. Both the knowledge and the data can be either qualitative or quantitative. This is another distinguishing feature of expert systems: ability to deal with qualitative data and knowledge.

The general problem-solving strategy of the expert system is contained in the inference engine. Common strategy here is to apply the rules in the knowledge base in a certain sequence until a conclusion is reached. This is called rule chaining. The user interface provides a mechanism for the expert to enter knowledge and the user to enter data for a specific problem.

Most expert systems are programmed using expert system shells. An expert system shell contains all of the structure shown in Figure 2.8a, but its knowledge base is empty. By programming the shell to contain knowledge about a domain, an expert system can be built.

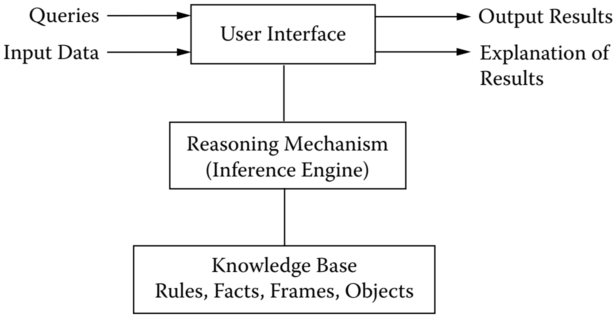


FIG. 2.8a
Structure of an expert system.

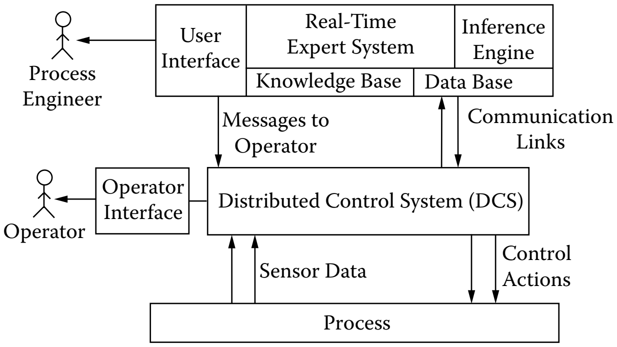


FIG. 2.8b
Structure of a real-time expert system.

Real-Time Expert Systems Real-time expert systems require additional features: e.g., the ability to include the time element in the reasoning process. Software vendors have recognized the unique requirements of real-time expert systems, and expert system shells are currently available on the market tailored to meet the requirements of process control applications. Naturally, it requires close integration with the underlying distributed control system (DCS) for exchange of sensor data and communication with the operator.

Figure 2.8b shows the structure of a real-time expert system. Examples of real-time expert system shells include G2 (from Gensym Corporation; see Reference 1) and Expert 90 (from ABB Process Automation, formerly Bailey Controls; see Reference 2).

Knowledge acquisition is widely recognized as the bottleneck in the implementation of expert systems. It had been thought that personnel trained in the technology of expert systems (called knowledge engineers) would interview the process control experts to elicit the knowledge and then incorporate these into an expert system. However, it has been found that it is easier for the expert to directly enter the knowledge into the expert system shell. The current shells on the market are built so that knowledge-base construction is highly automated.

Some shells can also generate the knowledge base from examples. At least one large chemical company (DuPont) espouses the philosophy that experts themselves are the best candidates for constructing the expert systems. Table 2.8c compares the advantages and disadvantages of using human experts vs. expert systems.

TABLE 2.8c
Expert Systems vs. Human Experts

Human Expert	*Expert System*
Knowledge is difficult to transfer	Knowledge is easily reproduced
Difficult to document	Self-documenting
Sometimes inconsistent	Consistent
Expensive	Affordable
Volatile	Permanent
Creative	Uninspired
Adaptive	Rigid
Broad focus	Narrow focus
Common-sense knowledge	Knowledge is limited to a narrow domain

Expert Systems in Control System Design Expert system applications to process control can be divided into two areas: control system design and real-time process control. The goal of an expert system for design is to capture the expertise of the design engineers and thereby reduce the design time that is required to produce an optional design strategy. For example, consider a control system engineer who is an expert on distillation control. He or she knows precisely what questions to ask and knows which control system configuration is appropriate for a given situation. For a typical design problem, the expert gathers all the necessary information on the particular problem at hand and uses experience to suggest a control system.

An expert system for distillation control would be built by first gathering the knowledge used by the expert, organizing this knowledge, and finally encoding it into a computer program. The user will interact with the program in an interview mode, supplying relevant data as needed for the specific problem. The expert system will suggest an appropriate control configuration. This was one of the early applications reported by Shinskey.[3]

The advantage of such an expert system is that it can be easily reproduced and distributed; hence, the expertise of one or a few experts becomes widely available. Many people view the expert system as a way of permanently capturing the knowledge of expert(s) that was gathered by years of experience while working in a specialized area.

This does not mean that experts can be replaced by expert systems. The knowledge contained in an expert system is never complete. The output of an expert system must still be validated by the design engineer to make sure that it conforms to the engineer's original design specifications.

Batch Process Control In the area of real-time control, expert systems have been applied in many areas, including batch process control, fault diagnosis, and statistical process control. Joseph et al.[4] discuss an application of expert systems to quality control of a batch process for making oil additives. The problem was to control the quality and production rate of a set of batch reactors.

The processing chemistry is not well understood and involves gas, liquid, and solid phases. Normally a team consisting of a process chemist, engineer, and operators is assembled to troubleshoot production problems. The objective is to capture the knowledge of these experts and encode that knowledge in a software package that could then be used by the operators to diagnose routine problems.

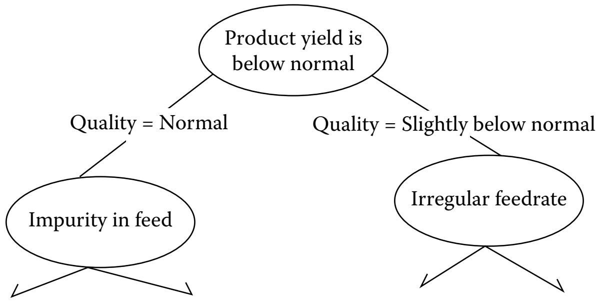


FIG. 2.8d
An example of a decision tree structure used for representing knowledge.

In such cases, the most convenient way to represent the knowledge is in the form of a decision tree with each node in the tree representing a decision to be made regarding the possible cause of the problem. Consider for example the following rules used by the experts:

Rule 1. "If production yield is low and the quality is within spec, then the problem is due to impurities in raw material A."

Rule 2. "If production yield is low and quality is slightly below target, then the problem is due to irregularity in the feed rate of raw material A."

These two rules can be represented in the decision tree shown in Figure 2.8d. The decision tree structure is a convenient way to extract and validate knowledge from experts.

The decision tree can consist of several levels and can provide for the exchange of business- and control-related data. The integration of business and control functions requires the definition of interfaces and the development of standards on nomenclature. ISA-SP95 Committee is working to develop an integration standard for enterprise-wide control (Figure 2.8e).

Fault Diagnosis In a process plant, fault diagnosis requires experts who are familiar with the operation. Although some failures can be diagnosed by a novice using an operations manual, often this is not the case, and experienced operators must be called in. Expert systems can serve as an additional tool for the inexperienced operator in diagnosis of faults.

Rich et al.[5] describe an application of expert systems to diagnose of faults in a whipped toppings process. The objective of the expert system was twofold: first, to diagnose the cause of the fault, and second, to suggest remedial actions to correct the situation and bring the process back to normal. The knowledge was acquired from two experts familiar with the day-to-day operations of the plant.

When multiple experts are involved, it is possible that their opinions may conflict. Care must be exercised to resolve such conflicts in a mutually agreeable fashion before the knowledge base is completed. It is preferable to first gather the knowledge from each expert using one-on-one interviews and then to bring them together as a group to resolve the conflicts. Completeness of the expert system (defined as the ability to resolve potential faults) may have to be sacrificed in favor of efficiency and speed of response.

Careful attention must be paid to the user interfaces provided. This includes the type of queries made by the expert system during the diagnostic session, the language and terminology used (explanation of terms should be incorporated), the effective use of graphics, and the response time of the system to entries made by the user.

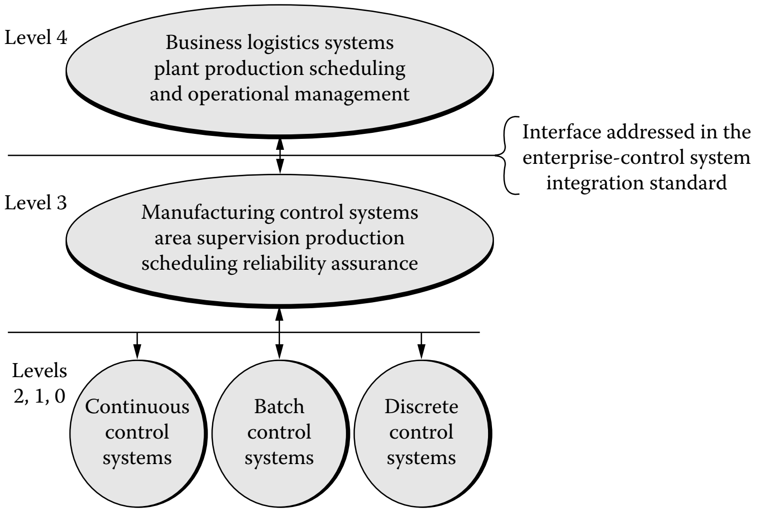


FIG. 2.8e
Enterprise-control domain. (Copyright 2001 ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission from ANSI/ISA-95.00.01-2000, "Enterprise-Control System Integration Part 1: Models and Terminology." For information, contact ISA at www.isa.org.)

Statistical Process Control (SPC)

Statistical process control (SPC) is discussed in more detail in a later section of this chapter. Online SPC is a very good tool for monitoring if the process is operating normally or if something unusual is going on. Many DCS systems have the ability to automatically generate quality control charts. The operator has to decide what action to take if some abnormality is detected. This can be done using some type of Pareto analysis.

Online expert systems (or DCS-embedded expert systems, as they are sometimes called) offer the capability of doing this diagnosis automatically. The DCS can feed the current state of the process to the expert system at prespecified time intervals. Using the knowledge contained in the expert system, further analysis of the data can be performed to provide advice and guidance to the operator so that the operator can take appropriate action. This advice can be in the form of messages that appear on the operator's CRT console or can be printed out on a printer. This can be thought of as an intelligent alarm device that not only points to the abnormality in the system but also provides advice on possible causes and remedies.

Unfortunately, problems still exist in the implementation of such online expert systems. Plant characteristics change from time to time, which often requires an updating of the knowledge base in the expert system. Often, the knowledge needed for diagnosis is not available. Research is underway to overcome these deficiencies. One approach is to build systems that can acquire the needed knowledge automatically.

This falls under the domain of machine learning. There are two major issues here. The first issue is the large amount of data that is usually generated on a process from online sensors; the second is the ability to assimilate these numbers and reduce them to a form that can be used by an expert system. Some progress has been reported in solving both of these issues, but it may be awhile before such products appear on the market.

Artificial Neural Networks

Artificial neural networks (ANNs) represent an important information abstraction in modeling intelligence in process control. In a simplistic sense, neural networks may be thought of as a functional mapping of inputs connected to outputs through an interconnected network of nodes. The weights given to the node interconnections are varied to achieve the desired mapping.

The advantages of neural networks lie in their ability to learn arbitrary function mapping with little or no prior knowledge about the function itself. Thus, they provide the capability to do "blackbox modeling" of a process for which only input/output data are given. In addition, these networks have been shown to be robust, resilient, and capable of adaptive learning. The disadvantage of this model is that the knowledge is implicit in the network connection weights.

Neural networks are excellent at doing pattern classification. Recent applications in industry involve the use of neural networks as "virtual" or soft sensors to infer quality from readily available sensor data on a process. Pivoso and Owens at DuPont have reported two applications of ANNs in this category.[6] The first is in the area of predicting the composition of a two-polymer mixture from the mixture's infrared spectrum. The inputs to the network are 39 measurements of the mixture spectrum at different wavelengths, and the output is the concentration of the primary component in the mixture.

To train the network (i.e., to determine the network connection weights), training data are generated from spectra of mixtures whose composition is known. Various numerical algorithms are available to train the network, and commercial software packages that implement these schemes are available.[7,8]

Another application reported in Reference 6 is the use of ANN as a virtual sensor to predict periodically measured concentrations. In many processes a critical quality variable to be controlled can only be measured intermittently. At DuPont, historical data on temperature, pressure, and concentration were used to train a network to predict the concentration from temperature and pressure measurements. This trained network could then be used to provide continuous, online prediction of concentration for process control purposes. The error in prediction was reduced to 1/1000 of a linear regression estimator. This may be attributed to the ability of ANNs to map nonlinear functions.

Future Trends

Expert systems open a new dimension in the way computers can be used for process control. Expert systems technology is being integrated into the architecture of distributed control systems and networked programmable controller systems. In some instances, such as in the self-tuning PID controllers, this technology is already incorporated.

Current and future applications include sensor testing and validation, control system design validation, performance evaluation, diagnostics, online tuning of multivariable control systems, and statistical process control and analysis. Many manufacturers are incorporating expert systems as part of the packaged machinery that they manufacture. For example, a boiler maker may incorporate a troubleshooting and diagnostic expert system as part of the control system for the boiler. Such enhancements could give the manufacturer a competitive edge.

Fuzzy Logic and Neural Networks

The common characteristics of fuzzy logic and neural networks is that one does not need to know anything about the mathematical model of the process in order to utilize them. In a way it is like the tennis player who can hit the ball without an in-depth knowledge of Newton's laws of motion and how these laws apply to the tennis process. A fuzzy logic controller just mimics the operator (the tennis player) in its responses.

Neural networks are similar to fuzzy logic insofar that the mathematical model which relates the inputs to the outputs of the process need not be known. It is sufficient to know the process response (as in tennis). The major difference between fuzzy logic and neural networks is that neural networks can only be trained by data, not with reasoning. In fuzzy logic each controller parameter can be modified both in terms of its gain (its importance) and also in terms of its functions.

Neural Networks

Artificial neural networks (ANN) can be looked upon as tools for examining data and building relationships. AAN has found applications in the process industry for capturing relationships among measurements and modeling. Compared to the technology of expert systems, the technology of ANN is immature and still evolving.

Neural networks attempt to mimic the structures and processes of biological neural systems. They provide powerful analysis properties such as complex processing of large input/output information arrays, representing complicated nonlinear associations among data and the ability to generalize or form concepts/theory.

Neural networks are a tool to solve many complex problems. Although they are often deployed as stand-alone applications, they can also be integrated into overall solutions. They may be embedded within databases or expert system applications, act as preprocessors or postprocessors to other systems, or be linked in a distributed architecture.

Advantages of neural networks include:

- Their good fit for nonlinear models
- Ability to adapt, generalize, and extrapolate results
- Speed of execution in recall mode
- Ease of maintenance

However, neural nets have the following inherent disadvantages:

- Cannot alone handle constraints
- Cannot optimize
- Need lots of data
- Need lots of horsepower (CPU) in training (learning) session
- Are unpredictable for utilization in "untrained" areas
- Are not well understood and therefore are not yet widely accepted

Terminology A definition for artificial neural networks (ANN) is "A cognitive information processing structure model based upon models of brain function. It is a highly parallel dynamic system that can carry out information processing by means of state response to inputs."[9,10] A cynical definition may refer to neural networks as the ultimate "fudge factor" algorithm, as the feedforward neural network has the potential to approximate any nonlinear function.[11]

Neural networks obtain their name and some of their associated terminology through the analogy to the biological system. For example, neural networks are built of "neurons." They are also called "nodes" or "processing elements." "Nodes" are usually arranged in layers or slabs and are often connected to many nodes in other layers.

A "layer" is a set of nodes whose weights are actively manipulated, serving as a buffer between input or output or other layers. A "slab" is a set of nodes that may be different in terms of internal specifications or connectivity but that share the same layer. A single layer may consist of multiple slabs. The comparison of a basic unit of a neural network model (processing element) is shown alongside a biological representation of a neuron[12] in Figure 2.8f.

Each node processes the input it receives via these connections and provides a continuous analog value to other

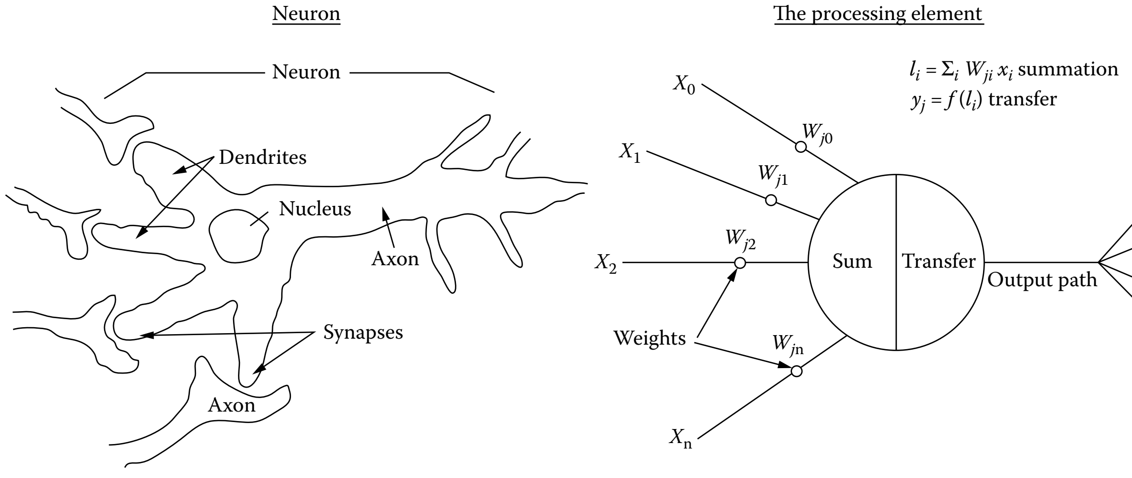


FIG. 2.8f
Neural net analogy with the brain.[12]

processing elements via its outgoing connections. As in biological systems, the strengths of these connections can change. Deciding how the processing elements in a network are connected, how the processing elements process their information, and how the connection strengths are modified all go into creating a neural network. Owing to the rapid progress of neural technology, today hundreds of different types of neural networks exist,[13] with more being added.

Traditional programming techniques, used for linear programs (LP) and other models, require that an algorithm be written to describe how the system should proceed. Data are fetched when needed for execution of the instructions, and the results are stored. All of these operations are performed in a deterministic and sequential manner. In contrast, neural networks consist of heavily interconnected processing elements that do not store information. The knowledge is stored by the way the processing elements are connected and by the importance of each connection, known as its weight. In contrast to being programmed, neural networks are trained. This means that examples are presented to the network, and the network adjusts itself by a specified learning rule (based on how correct the response is to the desired response). Therefore, many representative samples must be presented to a neural network in order for it to learn its rules for processing knowledge.

History The fundamental concepts of neural networks have existed since the 1940s,[14] with theoretical foundations laid down in the 1960s[15,16] and 1970s.[17] Research from the 1980s[18–22] and the early 1990s provides the basis of most neural network applications, (Section 2.1B discusses the most recent developments), as many groups are performing research into neural networks. These groups, each providing a different emphasis, insight, and motivation, include neuroscientists, cognitive psychologists, physicists, computer scientists, mathematicians, and engineers. The increasing power and inexpensive cost of computing has allowed development and deployment of systems for industrial applications.

Characteristics of Neural Networks A neural network processing element has many input paths. These inputs are individually multiplied by a weight and then summed. A nonlinear transfer function (also known as a squashing function) is applied to the result to calculate each processing element's output. The transfer function adds nonlinearity and stability to the network. It is supposed that the network has some global error function, E, associated with it, which is a differentiable function of all the connection weights. Common differentiable transfer functions include:

The sigmoid: where $f(z) = 1/(1 + e^{-z})$ **2.8(1)**

The hyperbolic tangent (tanh): where $f(z) = (1 - e^{-z})/(1 + e^{-z})$ **2.8(2)**

The linear: where $f(z) = z$ **2.8(3)**

The sine: where $f(z) = \sin(z)$ **2.8(4)**

Transfer functions used in other network types include:

The signum: where $$f(z) = 1 \quad \text{if } z > 0$$ $$= -1 \quad \text{if } z < 0 \qquad \textbf{2.8(5)}$$

The perceptron: where $$f(z) = z \quad \text{if } z > 0$$ $$= 0 \quad \text{if } z < 0 \qquad \textbf{2.8(6)}$$

Variations in how these transfer functions are fired determines the uniqueness of other types of networks. The sigmoid is the transfer function most often used in back-propagation networks. TanH is just a bipolar version of the sigmoid.

The output value of the transfer function is generally passed directly to the output path of the processing element (although exceptions to this exist). The output path is then connected to input paths of other processing elements through connection weights. These weights and connections form the "memory" or knowledge of the neural net. Since each connection has a corresponding weight, the signals on the input lines to a processing element are modified by these weights prior to being summed. Thus, the summation function is a weighted summation.

A neural network consists of many processing elements joined together. A typical network consists of a sequence of layers with full or random connections between successive layers. A minimum of two layers is required: the input buffer where data are presented, and the output layer where the results are held. However, most networks also include intermediate layers called hidden layers. Figure 2.8g shows a three-layer fully connected neural network.

Developing and Building Neural Networks The phases of a neural network life cycle[23,24] are concept, design, implementation, and maintenance.

Concept Phase The concept phase plans the approach to building the application. A strong emphasis is placed on experimentation. The user experiments under a multiple development environment in order to refine, redesign, and reformulate network parameters to achieve the best outcome. Successful applications for neural networks have several common characteristics. They include:

- Richness in historical data or examples
- Intensiveness of data and dependence upon several interacting parameters
- Incompleteness of data set, or a data set that contains errors or describes specific examples
- Unknown functions or expensive-to-discover solutions

The first step in implementing a neural network is to select a neural paradigm. Selection is based upon the traits of:

- Size (number of layers/number of slabs/number of nodes)
- Required output type (prediction/classification/association/conceptualization/filtering/optimization)

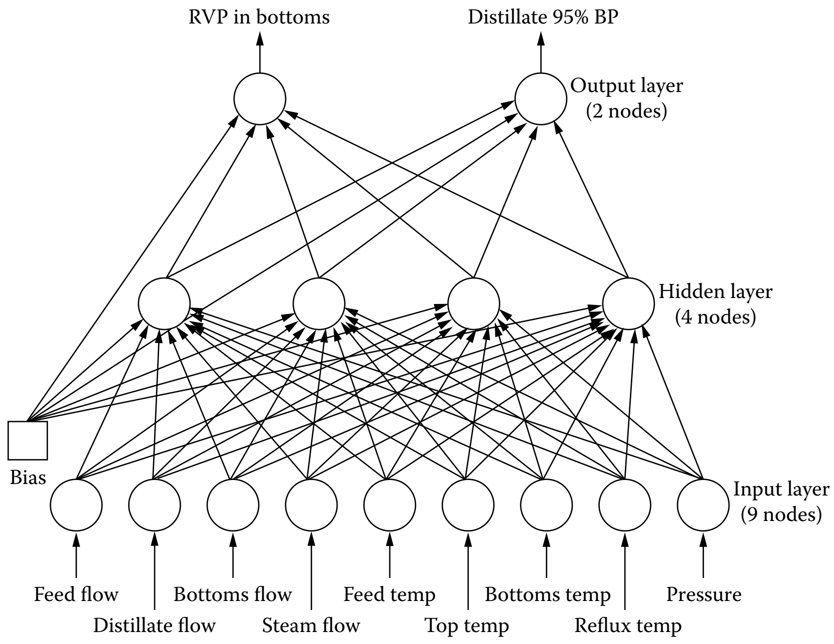


FIG. 2.8g
Example of a three-layer back-propagation network describing the application of physical property prediction.

- Associative memory classification (auto-/hetero-associative)
- Method of training (supervised/unsupervised)
- Time constraints (learning threshold limit)

Back-propagation neural networks have proven their applicability in chemical process control and have shown the ability to learn plant dynamics very well.[25] Two classic examples of a nonlinear relationship are the prediction of pH values in a stirred tank[26] and the predictions of physical properties (boiling points) of products from fractionators.[27] Other good applications include the use of neural networks to control systems,[28] including controller modeling[29] and process modeling.[30] Applications, therefore, should be based upon the following considerations:

- Conventional computing technology is inadequate.
- The problem requires qualitative or complex quantitative reasoning.
- The solution is derived from highly interdependent parameters that have no precise quantification.
- Data are readily available but multivariate and intrinsically noisy or error-prone.
- Project development time is short, but sufficient neural network training time is available.

Conversely, applications that are not suited for neural net solutions include:

- Mathematically accurate and precise applications (i.e., accounts receivable and inventory)
- Applications that require temporary data load, analysis, and reporting (i.e., sales data analysis and resource management)
- Applications that require deduction and stepwise logic

Design Phase The design phase specifies initial values and conditions for the selected neural paradigm. The type and number of processing elements, the number of slabs, the number of layers, the connectivity of the layers, the transfer function, the learning algorithm or rule, and other parameters such as momentum factor, learning threshold, learning coefficients, and learning schedules are design decisions that must be made.

The learning schedule is a breakpoint table that allows the learning coefficient to decay after a number of learning passes. The higher the value of the learning coefficient, the faster the learning. However, the higher the learning coefficient, the slower the convergence. Therefore, if the learning coefficient can be reduced as training proceeds, high learning and fast convergence can take place. The momentum factor acts as a low-pass filter on the weight differences that allows faster learning on low learning coefficients.

Other features to expedite convergence include algorithms for random weight generation, introduction of noise as the network learns, and algorithms to jog the weights randomly. Some neural paradigms constrain the choices or provide defaults, while other paradigms are very flexible and require numerous decisions.

Hidden layers act as layers of abstraction and help the network generalize and even memorize. Increasing the number of hidden layers augments the processing power of the neural network but also significantly increases and complicates training by requiring more training examples. Generally, most process control applications require only one hidden layer. Multiple parallel slabs within a layer may help to increase a network's processing power. This architecture forces the extraction of different features simultaneously. A common method of determining whether multiple layers or multiple slabs improve the performance of a network is experimentation.

Determining the number of nodes in a hidden layer is another experimental exercise. Rules of thumb give a starting point. One suggestion[23] is that if one hidden layer is chosen, the number of processing elements may be equal to 75% of the size of the input layer. Another suggestion is that the number of hidden nodes equals twice the square root of the sum of the number of input and output characteristics, rounded down to the nearest integer.[31] Using the amount of training data to set an upper bound for number of processing elements in the hidden layer has been promoted[12] according to the following equation:

$$h = \frac{\text{number of data sets}}{X(m+n)} \tag{2.8(7)}$$

where

h = number of processing elements in the hidden layer
m = number of processing elements in the output layer
n = number of processing elements in the input layer
X = noise factor—usually 2 to 10, where 10 is relatively noisy and 2 is very clean (extremely noisy data could require a number as high as 50 or 100)

In general, the more complex the relationship, the more processing elements are required in the hidden layer. Too few nodes in the hidden layer prevent the network from properly mapping inputs to outputs. On the other hand, too many nodes promote memorization and inhibit generalization. Memorization occurs when the patterns presented to the network are reproduced exactly without extracting the salient features. The network is then unable to process new patterns correctly because it has not discovered the proper relationships.

Depending upon the number of total nodes in a network, a sufficient number of training sets must be provided in order to train the system adequately. Otherwise, the situation is the same as trying to fit a third-order polynomial with two data points, where an infinite number of sets of polynomial constants can satisfy the equation. The situation of underspecification due to a lack of sufficient training examples is probably the single most common cause of problems for deployed networks.

Implementation Phase As many training data as possible need to be collected. Instead of volume, the quality and representativeness is important. A good training set will contain routine, unusual, and boundary condition cases. Preparing the data is an important consideration.[32] It includes:

- Transforming inputs into the proper form (ratios, classes) and data types (binary, bipolar, continuous)
- Preprocessing
- Converting the data to ASCII or binary vectors
- Normalizing (Euclidean, logarithmic) the data
- Scaling

The data may need to be converted into another form to be meaningful to the neural network. Likewise, how the data are represented and translated plays an important role in the network's ability to understand during training. Data may be continuous or binary, time-oriented or static. Data can be naturally grouped, may be represented as actual amounts or changes in amounts, or may be evenly or unevenly distributed over the entire range. When naturally occurring groups appear, binary categories are often the best method for making correlations.

Continuous data should not be used to represent unique concepts, however. When values are very continuous, artificially grouping them may hinder the network from distinguishing values on or near the border between two groups. At any rate, all data need to be normalized in order for the transform function to operate. Data may be scaled between minimum and maximum ranges or be set within a predefined range.

Implementing the neural network involves applying the design criteria using customized or commercial environments. Experimentation is carried out with learning coefficients, momentum factors, noise injection, etc., in order to train the network most efficiently. Because of the nature of some learning algorithms, the method of presenting the training data can also affect the training results.

Testing the neural network is required to verify the training process. Often the network is trained on one set of data (training data) and verified with a different set of data (recall data). Many statistical tools can help determine the degree to which the network produces the correct outputs. It can then be determined whether the sizes of the layers need to be changed and whether the learning parameters need to be adjusted.

Training three or four neural networks using the same paradigm is one common approach to verify training.

Another form of testing involves analyzing the actual weights of the network. Very small weights indicate that the processing elements are not influencing the outcome of the network. These nodes could then be eliminated and the network retrained in order to simplify the overall network configuration. Large weights may indicate too much dependence upon a particular input, indicating some degree of memorization.

Maintenance Phase Once a network is trained on a set of data, it is deployed to predict results based upon new sets of gathered data. Maintenance involves modifying the deployed neural network application because of decreasing accuracy due to conditions encountered for which the network was not trained. In such cases, the training set must either be modified or replaced. Next, the network must undergo retraining and reevaluation in an iteration of the implementation phase. If reevaluation shows that the network performance is not to specification with the new data set, the design phase must be repeated.

Back-Propagation Example The back-propagation network is the most popular feedforward predictive network deployed in the process industries. The back-propagation network assumes that all processing elements and connections are somewhat responsible for the difference between the expected output and the actual output. It uses the chain rule of derivative calculus to allocate the prediction error to every node in the feedforward network.[9] This error is propagated from the output layer back to the input layer through the connections to the previous layer. The process is repeated until the input layer is reached. The training algorithm is an iterative gradient descent algorithm designed to minimize the root mean square error (RMS) between the actual output and the desired output. It requires a continuous differentiable nonlinear search space.

Consider a fractionation example, where the product specifications are based on the Reid vapor pressure in the bottom product and on the 95% boiling point of the distillate. Process analyzers have inherent dead time in producing these measurements, or analyzers are not available to provide these measurements. Process models can be used to predict these properties, or other measurements can be used to infer them.

However, if data exist regarding the measurements around the fractionator based on collected laboratory or analyzer results, as commonly recorded on log sheets, a nonlinear neural network model can be built. Shown in Figure 2.8g is a back-propagation model with nine input nodes, four hidden nodes, and two output nodes with bias. Training a back-propagation network that uses the methodology represented in Figure 2.8h can be summarized as follows:

Step 1: All weights and processing element offsets are set to small random values.

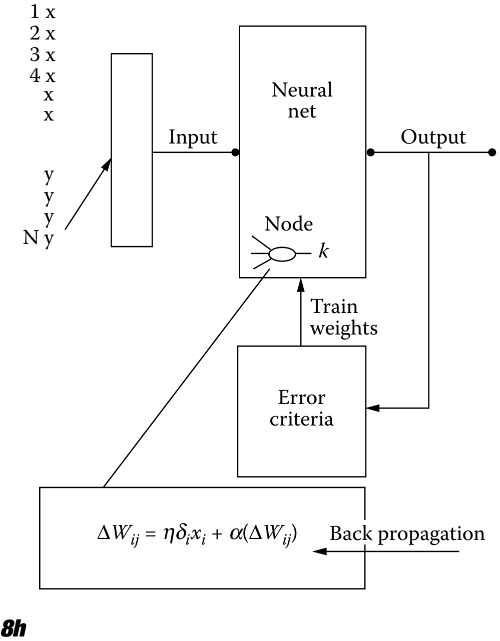


FIG. 2.8h
Back-propagation training.

Step 2: All inputs and outputs are normalized and presented to the network. Normalization is generally based upon expected minimum and maximum values.

Step 3: A transfer function (also known as a squashing function) is applied to the weighed sum of the normalized inputs at each processing element to calculate each processing element's output. The sigmoid is an often-used transfer function. TanH is a bipolar version of the sigmoid and is used when the output tends to saturate ("saturation" is when many transformed values congregate toward the upper or lower limit).

Step 4: A recursive algorithm starting at the output processing elements is used and repeated until the input processing elements are reached. The weights are adjusted by the equation below:

$$W_{ij}(t+1) = W_{ij}(t) + hd_j x_i \qquad \textbf{2.8(8)}$$

where

$W_{ij}(t)$ = the weight from hidden node i or from an input to node j at time t
x_i = either the output of node i or an input
h = a gain term
d_j = an error term for node j

Since the error term is

$$d_j = dE/dx_j \qquad \textbf{2.8(9)}$$

where E is the difference between the desired output and the actual output, using the sigmoid function gives

$$f(x) = 1/(1+e^{-x})^{-1} \qquad \textbf{2.8(10)}$$

which gives

$$dE/dx_j = x_j(1 - x_j) \qquad \textbf{2.8(11)}$$

If node j is an output node, then

$$d_j = y_i(1 - y_i)(d_j - y_i) \qquad \textbf{2.8(12)}$$

where d_j is the desired output of node j and y_i is the actual output.

If node j is a hidden node, then

$$d_j = x_j(1 - x_j)\sum d_k w_{jk} \qquad \textbf{2.8(13)}$$

where k is over all nodes in the layers before node j.

Internal node thresholds are adapted in a similar manner by assuming they are connection weights on links from auxiliary constant-valued inputs. Convergence is sometimes faster if a momentum term is added and weight changes are smoothed by the following equation:

$$W_{ij}(t+1) = W_{ij}(t) + hd_j x_i + a(W_{ij}(t) - W_{ij}(t-1)) \qquad \textbf{2.8(14)}$$

where $0 < a < 1$.

As with any gradient descent method, back-propagation could find a local minimum instead of the global minimum. The momentum term is designed to help the training algorithm overcome the small valleys of local minimums. The learning procedure requires that the change in weights be proportional to rate of change of error, with respect to changes in weights. The constant of proportionality is called the learning rate, h (or learning coefficient). The larger the value of h, the faster the learning rate.

Step 5: Repeat by going to step 2 until convergence is reached. Convergence will be when the RMS error reaches a user-set threshold value.

The initial choice of weights is very important to convergence. Small random numbers should be used and equal weights should be avoided. It is possible for the system to require many iterations before convergence: 10,000 to 100,000 iterations are not uncommon.

SUMMARY

As with expert systems, several good neural shells have come onto the market.[33] These shells offer a very good method of organizing and implementing an application. Neural networks are a tool for solving many complex problems. They can be deployed as stand-alone applications or integrated into overall solutions. They may be embedded within databases or expert system applications,[34,35] act as preprocessors or postprocessors to other systems, or be linked in a distributed architecture.

As a preprocessor, networks have been used to identify objects within an image and pass that information to an expert system. As a postprocessor, neural networks can remove noise, classify patterns, and make predictions from a database. Neural networks, combined with statistical validation, have also been introduced as viable solutions for the control of manufacturing facilities.[36]

References

1. Rowan, D., "On-line Expert Systems in Process Industries," *AI/Expert,* August 1989.
2. Frerichs, D., "Integrating Diagnostic Expert System with Statistical Process Control in a Modern Distributed Control System," Paper Presented at ISA-90, Instrument Society of America, 1990.
3. Shinskey, F. G., "Chemical Process Control II: Proceedings," T. J. McAvoy et al. (Eds.), Amsterdam: Elsevier, 1986.
4. Joseph, B., et al., "Knowledge Acquisition and Knowledge Representation in an Expert System for Control of a Batch Process," Paper Presented at AIChE National Meeting, Houston, 1991.
5. Rich, S., et al., "Development of a Diagnostic Expert System for a Whipped Toppings Process," *J. Loss Prev. Proc. Ind.,* Vol. 2, July 1989.
6. Pivoso, M. J., and Owens, A. J., "Sensor Data Analysis Using Artificial Neural Networks," in *Chemical Process Control IV Proceedings,* Y. Arkun (Ed.), New York: AIChE, 2001.
7. Neural Networks Explorer, Neural Ware, Pittsburgh, PA.
8. Brainmaker, California Scientific Software, Grass Valley, CA.
9. Rumelhart, D. E., and McClelland, J. L., *Parallel Distributed Processing,* Vol. 1, Cambridge, MA: MIT Press, 1986.
10. Ashenayi, K., "Neural Networks Overview," Proceedings of the Third Oklahoma Symposium on Artificial Intelligence, Tulsa, November 1989.
11. Hecht-Nielson, R., "Theory of Back-propagation Neural Networks," Proceedings of the International Conference on Neural Networks, I, New York: IEEE Press, 1989, pp. 593–611.
12. "Neural Computing," NeuralWare, Technical Publications Group, Pittsburgh, PA, 1991.
13. Lippmann, R. P., "An Introduction to Computing with Neural Nets," *IEEE ASSP Magazine,* April 1987, pp. 4–22.
14. McCulloch, W. H., and Pitts, W. S., "A Logical Calculus of the Ideas Immanent in Neural Nets," *Bulletin of Mathematical Biophysics,* Vol. 5, 1943, pp. 115–133.
15. Widrow, B., and Hoff, M., "Adaptive Switching Circuits," 1960 IRE WESCON Convention Record, Part 4, August 23–26, 1960, pp. 96–104.
16. Minsky, M. L., and Papert, S. S., *Perceptrons,* Cambridge, MA: MIT Press, 1969.
17. Grossberg, S., "Adaptive Pattern Classification and Universal Recoding: I. Parallel Development and Coding of Neural Feature Detectors," *Biological Cybernetics,* Vol. 23, 1976, pp. 121–134.
18. Hopfield, J. J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities," *Proceedings of the National Academy of Sciences, USA,* Vol. 79, April 1982, pp. 2554–2558.
19. Ackley, D. H., Hinton, G. E., and Sejinowski, T. J., "A Learning Algorithm for Boltzmann Machines," *Cognitive Science,* 9:147–169, 1985.
20. Kosco, B., "Bidirectional Associative Memories," *IEEE Transactions on Systems, Man, and Cybernetics,* 1987.
21. Anderson, J. A., and Rosenfeld, E., *Neurocomputing Foundations of Research,* Cambridge, MA: MIT Press, 1988.
22. Kohonen, T., *Self-Organization and Associative Memory,* 2nd ed., New York: Springer-Verlag, 1988.
23. Bailey, D. L., and Thompson, D. M., "How to Develop Neural-Network Applications," *AI Expert,* June 1990, pp. 38–47.
24. Bailey, D. L., and Thompson, D. M., "Developing Neural-Network Applications," *AI Expert,* September 1990, pp. 34–41.
25. Willis, M. J., Massimo, C. D., Montague, G. A., Tham, M. T., and Morris, A. J., "On the Applicability of Neural Networks in Chemical Process Control," Proceedings of the AIChE Winter Meeting, Chicago, Illinois, November 1990.
26. Bhat, N., and McAvoy, T. J., "Use of Neural Nets for Dynamic Modeling and Control of Chemical Process Systems," Proceedings of the American Control Conference, 1989, pp. 1342–1347.
27. Blaesi, J. E., and Jensen, B. A., "Comparison of a Neural Network With Theoretical Calculations of Inferred Properties," Proceedings of Industrial Computing Society, Vol. 2, Instrument Society of America, Toronto, April 1992.
28. Werbos, P. J., "An Overview of Neural Networks for Control," *IEEE Control Systems Magazine,* January 1991, pp. 40–41.
29. Hoptroff, R. G., Hall, T. J., and Burge, R. E., "Experiments with a Neural Controller," IEEE International Joint Conference on Neural Networks — IJCNN 90, Part 2, San Diego, June 1990.
30. Samad, T., "Neural Networks for Control — An Overview," Proceedings of the ISA/91, Vol. 46, Part 1, Instrument Society of America, Anaheim, CA, October 1991.
31. Hillman, D., "Software Review: Neuroshell v.3.2," *AI Expert,* September 1990, pp. 61–64.
32. Lawrence, J., "Data Preparation for a Neural Network," *AI Expert,* November 1991, pp. 34–41.
33. Reid, K., and Zeichick, A., "Neural Net Resource Guide to 50 Tools," *AI Expert,* June 1992, pp. 50–56.
34. Verduin, W. H., "Neural Network Software Assists Expert System," *Control Engineering,* July 1991, pp. 37–40.

35. Caudill, M., "Using Neural Nets: Hybrid Expert Networks," *AI Expert,* November 1990, pp. 49–54.
36. Minor, J. M., "Applied Neural Nets: Industrial Strength Solutions," Expert Systems Applications in Advanced Control, Proceedings of the Seventeenth Annual Advanced Control Conference Sponsored by The Purdue Laboratory for Applied Industrial Control, Purdue University and Control Engineering, October 1991, pp. 83–97.

Bibliography

Al-Kofahi, K., et al., "A Machine Learning Approach to Prior Case Retrieval," *Proceedings of the 8th International Conference on Artificial Intelligence,* New York: ACM Press, 2001.

American Association of Artificial Intelligence, 445 Burgess Dr., Menlo Park, CA 94025-3442, 2002.

Bhagat, P., "An Introduction to Neural Nets," *Chemical Engineering Progress,* August 1990, pp. 55–60.

Bullart, L. (Ed.), *Application of Artificial Intelligence in Process Control,* Science and Technology Books, June 1993.

Caudill, M., *Neural Network Primer,* San Francisco: Miller Freeman, 1990.

Caudill, M., "Using Neural Nets: Diagnostic Expert Nets," *AI Expert,* September 1990, pp. 43–47.

Caudill, M., "The View from Now," *AI Expert,* June 1992, pp. 24–30.

Coleman, K. G., and Watenpool, S., "Neural Networks in Knowledge Acquisition," *AI Expert,* January 1992, pp. 36–39.

Cox, E., "Integrating Fuzzy Logic into Neural Nets," *AI Expert,* June 1992, pp. 24–30.

Garson, D. G., "Interpreting Neural-Network Connection Weights," *AI Expert,* April 1991, pp. 47–50.

Gringrich, C. G., Kuespert, D. R., and McAvoy, T. J., "Modeling Human Operators Using Neural Networks," *ISA Transactions*, Vol. 31, No. 3, July 1992.

Grossberg, S., *Neural Networks and Natural Intelligence,* Cambridge, MA: MIT Press, 1988.

Guilandoust, M. T., Morris, A. J., and Tham, M. T., "Adaptive Inferential Control," *IEE Proceedings, Part D*, 1987.

Haesloop, D., and Holt, B. R., "Neural Networks for Process Identification," 1990 International Joint Conference on Neural Networks — IJCNN 90, Part 3, San Diego, June 1990.

Hastie, T., et al., *Elements of Statistical Learning*, Heidelberg: Springer-Verlag, August 2001.

Hecht-Nielsen, R., *Neurocomputing,* Menlo Park, CA: Addison-Wesley, 1989.

Jackson, P., *Introduction to Expert Systems,* 3rd ed., Harlow, England: Addison Wesley Longman, 1999.

Jameson, J., "A Neurocontroller Based on Model Feedback and the Adaptive Heuristic Critic," 1990 International Joint Conference on Neural Networks—IJCNN 90, Part 2, San Diego, June 1990.

Jones, T. M., "AI Application Programming," Charles River Media, March 2003.

Kaya, A., and Kumar, A., "A Real-Time Expert Algorithm for Process Fault Diagnostics," 1992 ISA Conference in Houston, October 1992.

Lawrence, S., et. al., "On the Distribution of Performance from Multiple Neural Network Trials," *IEEE Transactions on Neural Networks,* 1997.

Leahy, M. B., Jr., Johnson, M. A., Bossert, D. E., and Lamont, G. B., "Robust Model-Based Neural Network Control," Proceedings of the 1990 IEEE International Conference on Systems Engineering, Pittsburgh, PA, August 1990.

Miller, T. W., III, Sutton, R. S., and Werbos, P. J., *Neural Networks for Control,* Cambridge, MA: MIT Press, 1990.

Minor, J. M., "Introduction Neural Nets," Proceedings of the ISA/90, Vol. 45, Part 3, Instrument Society of America, New Orleans, October 1990.

Moorthy, S., and Iyer, P., "Neural Network Simulation for Plant Control Optimization," Proceedings of the Industrial Computing Conference, Vol. 1, Instrument Society of America, Anaheim, CA, October 1991.

O'Hare, G. M. P. (Ed.), *Foundations of Distributed Artificial Intelligence,* New York: Wiley, 1996.

Psichogios, D. C., and Ungar, L. H., "Nonlinear Internal Model Control and Model Predictive Control Using Neural Networks," Proceedings of the Fifth IEEE International Symposium on Intelligent Control, Philadelphia, September 1990.

Rehbein, D. A., Maze, S. M., and Havener, J. P., "The Application of Neural Networks in the Process Industry," Proceedings of the ISA/92, Vol. 23—Instrumentation in the Chemical and Petroleum Industries, Instrument Society of America, Toronto, April 1992.

Ripley, B. D., *Pattern Recognition and Neural Networks*, Cambridge: Cambridge University Press, 1996.

Rumelhart, D. E., and McClelland, J. L., *Explorations in Parallel Distributed Processing, a Handbook of Models, Programs, and Exercises,* Vol. 1, Cambridge, MA: MIT Press, 1988.

Russell, S., and Norvig, P., *Artificial Intelligence, A Modern Approach,* Englewood Cliffs, NJ: Prentice Hall, 2002.

Saerens, M., and Soquet, A., "Neural Controller Based on Back-propagation Algorithm," *IEE Proceedings, Part F: Radar and Signal Processing,* February 1991, pp. 55–62.

Sanders, F. F., "Continuous Control Gets On-Line Advisor," *InTech,* August 1992.

Schnelle, P. D., Jr., and Fletcher, J. A., "Using Neural Based Process Modeling for Measurement Inference," *Proceedings of the ISA*, Vol. 45, Part 2, 1990.

Simpson, P. K., *Artificial Neural Systems, Foundations, Paradigms, Applications, and Implementations,* New York: Pergamon Press, 1990.

Software for Neural Networks and Hybrid Systems, International Conference on Computational Intelligence and Neuroscience, Duke University, 1997.

Spinrad, M. D., "Self-Organizing Fuzzy Control for Non-Linear and Multiple-Input, Multiple-Output Processes," 1992 ISA Conference in Houston, October 1992.

Tsoukalas, L., and Reyes-Jimenez, J., "Hybrid Expert System-Neural Networks Methodology for Anticipatory Control in a Process Environment," Proceedings of the Third International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Charleston, South Carolina, July 1990.

Zeidenberg, M., *Neural Network Models in Artificial Intelligence,* New York: Horwood, 1990.

2.9 Feedback and Feedforward Control

M. F. HORDESKI (1985) **B. G. LIPTÁK** (1995) **F. G. SHINSKEY** (1970, 2005)

Feedback control is the action of moving a manipulated variable *m* in response to a deviation or error *e* between the controlled variable *c* and its set point *r* in such a way as to reduce, or if possible, to eliminate the error. Feedback control cannot anticipate and thereby prevent errors, because it can only initiate its corrective action *after* an error has already developed. Because of dynamic lags and delays in the response of the controlled variable to manipulation, some time will elapse before the correction will take effect. During this interval, the deviation will tend to grow, before eventually diminishing. Feedback control cannot therefore achieve perfect results; its effectiveness is limited by the responsiveness of the process to manipulation. By contrast, feedforward correction can be initiated as soon as any change is detected in a load variable as it enters the process; if the feedforward model is accurate and the load dynamics are favorable, the upset caused by the load change is canceled before it affects the controlled variable. Because feedforward models and sensors are both imperfect, feedforward loops are usually corrected by feedback trimming.

FEEDBACK CONTROL

The purpose of any process control system is to maintain the controlled variable at a desired value, the set point, in the face of disturbances. The control system regulates the process by balancing the variable load(s) with equivalent changes in one or more manipulated variables. For the controlled variable to remain stationary, the controlled process must be in a balanced state.

Regulation through feedback control is achieved by acting on the change in the controlled variable that was induced by a change in load. Deviations in the controlled variable are converted into changes in the manipulated variable and sent back to the process to restore the balance. Figure 2.9a shows the backward flow of information from the output of the process back to its manipulated input. The load *q* is a flow of mass or energy entering or leaving a process, which must be balanced by an equal flow leaving or entering. It may have more than one component—for example in a temperature loop, both the flow and temperature of a stream are components of its energy demand, but they may be balanced by a single manipulated variable such as steam flow. The steady-state gain *K* and dynamic gain vector **g** in the paths of the manipulated and load variables may differ, and therefore have been given different subscripts *m* and *q*.

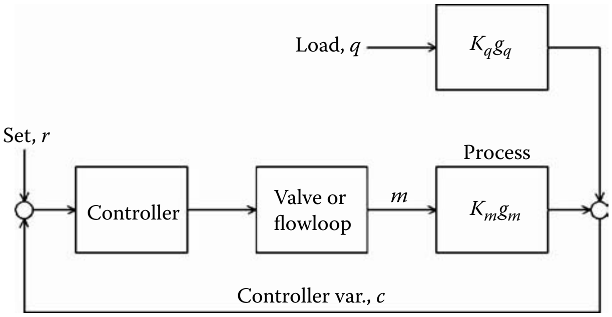


FIG. 2.9a
Load changes can enter through different gain and dynamics from the controller output.

Limitations of Feedback Control

Feedback, by its nature, is incapable of correcting a deviation in the controlled variable at the time of detection. In any process, a finite delay exists between a changing of the manipulated variable and the effect of the change on the controlled variable. Perfect control is not even theoretically attainable because a deviation in the controlled variable must appear before any corrective action can begin. In addition, the value of the manipulated variable needed to balance the load must be sought by trial and error, with the feedback controller observing the effect of its output on the controlled variable.

Best-Possible Feedback Control

An absolute limitation to the performance of a feedback control loop is imposed by any deadtime in the loop. Figure 2.9b describes the results of a step-load change applied to a process whose dynamics consist of deadtime and a single lag in both the load path $\mathbf{g}_q$ and the manipulated-variable path $\mathbf{g}_m$. The time scale is given in units of deadtime τ_d in the path of the manipulated variable. To simplify the illustration, the deadtimes in both paths are identical (this is not essential—*any*

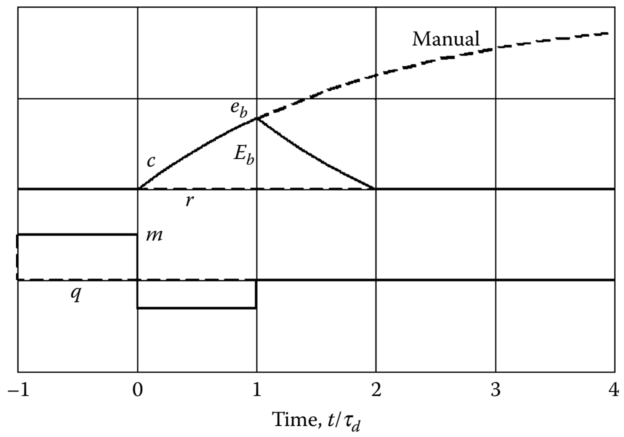


FIG. 2.9b
The best-possible load response for a process having deadtime and a single lag.

deadtime in the load path, or none, will produce the same response, simply shifting its location in time). Also for simplification, the steady-state gains in both paths are made equal.

At time –1, a step decrease in load q enters the process. After the deadtime in the load path expires, at time zero, the controlled variable c responds, beginning to rise along an exponential curve determined by the gain and lag in the load path. If the controller were left in Manual, it would continue to a new steady state. In this example, K_q is 2.0, leaving the final value of c in Manual having changed twice as much as load q; the time constant τ_q of the load lag in this example is $2.0\tau_d$.

Also shown is the possible return trajectory of c to set point r if the manipulated variable m were to step as soon as any deviation is detected, i.e., at time zero, by an amount

$$\Delta m = \Delta q(K_q/K_m)(1+\varepsilon^{-\tau_d/\tau_q}) \qquad \textbf{2.9(1)}$$

where ε is the exponential operator 2.718. This move turns out to be the best that is possible by a feedback controller,[1] as it causes the deviation to decay to zero during the next deadtime. The peak deviation reached during this excursion is

$$e_b = \Delta q K_q(1-\varepsilon^{-\tau_d/\tau_q}) \qquad \textbf{2.9(2)}$$

At the time the peak is reached, the controller output must be stepped to match the new steady-state load.

The leading edge of the curve, i.e., up to time 1.0, is determined completely by the load step and the gain and dynamics in its path; the trailing edge is determined by the controller and its tuning. The leading and trailing edges of the best response are complementary, so that the area that they enclose is

$$E_b = e_b\tau_d = \Delta q K_q \tau_d(1-\varepsilon^{-\tau_d/\tau_q}) \qquad \textbf{2.9(3)}$$

These values of e_b and E_b are the *best* that can be obtained for any feedback controller on a process whose dynamics consist of deadtime and a single lag. A real controller will yield a somewhat larger peak deviation and an area at least twice as great, depending on its modes and their tuning. If the process contains two or more lags, the value of E_b remains the same as estimated above but is more difficult to approach with a real controller.

Integrated Error

The actual effectiveness of feedback control depends on the dynamic gain of the controller, which is a function of its control modes and their tuning. Although a high controller gain is desirable, the dynamic gain of the closed loop at its period of oscillation must be less than unity if the loop is to remain stable. In effect, then, the dynamic gain of the process dictates the allowable dynamic gain of the controller.

For each process and controller, optimum settings exist that minimize some objective function such as Integrated Absolute Error (IAE). However, most controllers are not tuned optimally, for various reasons, such as process nonlinearities. Therefore, controller performance needs to be stated in terms of the actual settings being used. As an example, the integrated error produced by a load change to a process under ideal Proportional-Integral-Derivative (PID) control will be evaluated based only on its mode settings. The controller output m_1 at time t_1 is related to the deviation e_1 between the controlled variable and its set point by

$$m_1 = \frac{100}{P}\left(e_1 + \frac{1}{I}\int_{t_0}^{t_1} e\,dt + D\frac{de_1}{dt}\right) + C \qquad \textbf{2.9(4)}$$

where P, I, and D are the percent Proportional Band, Integral time and Derivative time, respectively, and C is the output of the controller at time, t_0 when it was first placed in Automatic. Let t_1 be a steady state, so that $e_1 = 0$ and its derivative is also zero.

Then a load change arises, causing the controller to change its output to return the deviation to zero. When a new steady state is reached, the controller will have an output m_2:

$$m_2 = \frac{100}{P}\left(e_2 + \frac{1}{I}\int_{t_0}^{t_2} e\,dt + D\frac{de_2}{dt}\right) + C \qquad \textbf{2.9(5)}$$

where e_2 and its derivative are again zero.

The difference in output values between the two steady states is

$$\Delta m = m_2 - m_1 = \frac{100}{PI}\int_{t_1}^{t_2} e\,dt \qquad \textbf{2.9(6)}$$

Solving for the integrated error:

$$E = \int_{t_1}^{t_2} e\,dt = \Delta m\frac{PI}{100} \qquad \textbf{2.9(7)}$$

For any sustained load change, there must be a corresponding change in the controller output, and in fact the current load is often reported as the steady-state controller output. The wider the proportional band and the longer the integral time, the greater will be the integrated error per unit load change. (While the derivative setting does not appear in the integrated-error function, its use allows a lower integral time than is optimum for a PI controller.) Therefore, loops with controllers having large PI products are candidates for feedforward control.

FEEDFORWARD CONTROL

Feedforward provides a more direct solution to the control problem than finding the correct value of the manipulated variable by trial and error, as occurs in feedback control. In the feedforward system, the major components of load are entered into a model to calculate the value of the manipulated variable required to maintain control at the set point. Figure 2.9c shows how information flows forward from the load to the manipulated variable input of the process. The set point is used to give the system a command. (If the controlled variable were used in the calculation instead of the set point, a positive feedback loop would be formed.)

A system, rather than a single control device, is normally used for feedforward loops because it is not always convenient to provide the computing functions required by the forward loop with a single device or function. Instead, the feedforward system consists of several devices if implemented in hardware or several blocks of software if implemented digitally. The function of these blocks is to implement a mathematical model of the process.

Load Balancing

A dynamic balance is required to keep the control variable at set point. It is achieved by solving the process material- and/or energy-balance equations continuously. When a change in load is sensed, the manipulated variable is automatically adjusted to the correct value at a rate that keeps the process continually in balance. Although it is theoretically possible to achieve such perfect control, in practice the system cannot be made to duplicate the process balance exactly.

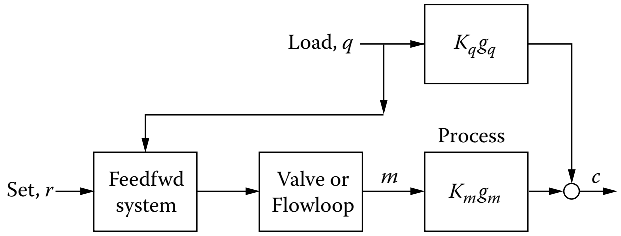


FIG. 2.9c
Feedforward calculates the value of the manipulated variable based on current load and set point.

The material- and energy-balance equations are not usually difficult to write for a given process. Variations in nonstationary parameters such as heat-transfer and mass-transfer coefficients do not ordinarily affect the performance of a feedforward system. The load components are usually inflow or outflow when level or pressure is to be controlled, feed flow and feed composition where product composition is to be controlled, and feed flow and temperature where product temperature is to be controlled. Flow is the primary component of load in almost every application because it can change widely and rapidly. Composition and temperature are less likely to exhibit such wide excursions, and their rate of change is usually limited by upstream capacity. In most feedforward systems these secondary load components are left out; their effects are accommodated by the feedback controller.

The output of a feedforward system should be an accurately controlled flow rate, if possible. This controlled flow rate cannot usually be obtained by manipulating the control valve directly, since valve characteristics are nonlinear and inconsistent, and the delivered flow is subject to such external influences as upstream and downstream pressure variations. Therefore, most feedforward systems depend on some measurement and feedback control of flow to obtain an accurate manipulation of the flow rate. Only when the range of the manipulated variable exceeds that which is available in flowmeters, or the required speed of response exceeds that of a flow loop, should one consider having the valves positioned directly, and in such cases care must be taken to obtain a reproducible response.

Steady-State Model

The first step in designing a feedforward control system is to form a steady-state mathematical model of the process. The model equations are solved for the manipulated variable, which is to be the output of the system. Then the set point is substituted for the controlled variable in the model.

This process will be demonstrated by using the example of temperature control in a heat exchanger (Figure 2.9d).[2] A liquid flowing at rate W is to be heated from temperature T_1 to controlled temperature T_2 by steam flowing at rate W_s. The energy balance, excluding any losses, is

$$WC_p(T_2 - T_1) = W_s\lambda \qquad \textbf{2.9(8)}$$

Coefficient C_p is the heat capacity of the liquid and λ is the latent heat given up by the steam in condensing. Solving for W_s yields

$$W_s = WK(T_{2\text{set}} - T_1) \qquad \textbf{2.9(9)}$$

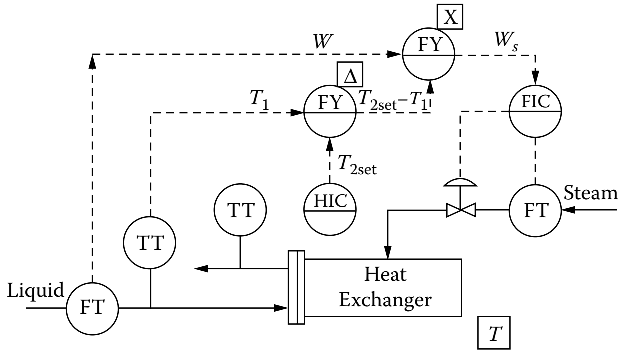


FIG. 2.9d
Steam flow is calculated here to satisfy the steady-state heat balance.

where $K = C_p/\lambda$, set point T_{2set} replaces T_2, and W and T_1 are load components to which the control system must respond.

Note that this energy balance is imperfect, because it disregards minor loads such as the heat losses to the environment and any variations in the enthalpy of the steam. Another factor is that even the two major loads cannot be measured perfectly, as there are no error-free sensors of flow and temperature. The net result is a model that is approximately an energy balance around this process.

An implementation of Equation 2.9(9) is given in Figure 2.9d. A control station introduces T_{2set} into the difference block [Δ] where input T_1 is subtracted. The gain of this block is adjusted to provide the constant K. Its output is then multiplied by the liquid flow signal W to produce the required set point for steam flow.

During startup, if the actual value of the controlled variable does not equal the set point, an adjustment is made to gain K. Then, after making this adjustment, if the controlled variable does not return to the set point following a change in load, the presence of an error in the system is indicated. The error may be in one of the calculations, a sensor, or some other factor affecting the heat balance that was not included in the system. Heat losses and variations in steam pressure are two possible sources of error in the model. Since any error can cause a proportional offset, the designer must weigh the various sources of error and compensate for the largest or most changeable components where practical. For example, if steam pressure variations were a source of error, the steam flowmeter could be pressure compensated. Because of all these errors, feedforward control as shown in Figure 2.9d is seldom used without feedback trim. Later in this section, a feedback controller will be added to automatically correct the loop for the errors in the model.

Dynamic Model

With the feedforward model as implemented in Figure 2.9d, a step decrease in liquid flow results in a simultaneous and proportional step decrease in steam flow. Transient errors following a change in load are to be expected in feedforward control systems. Two typical dynamic responses of the controlled variable are shown in Figure 2.9e; the offset in the dashed curve results from a 10% error in the static calculations.

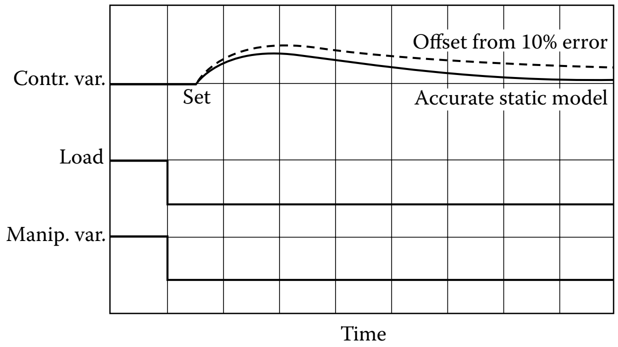


FIG. 2.9e
The dynamic response of the uncompensated feedforward system when the controlled variable responds faster to changes in load than to changes in manipulated flow.

If there is no static error, the temperature returns to set point eventually. However, the dynamic balance is missing. The decreased steam flow does not result in an instantaneous decrease in heat-transfer rate, because the temperature difference across the heat-transfer surface must first be decreased, and this requires a decrease in shell pressure. The lower shell pressure at lower loads means that the shell contains less steam as the load decreases. Since the static feedforward system does not overtly lower the steam inventory in the shell, it is lowered through a lagging heat-transfer rate, resulting in a transient rise in exit temperature on a falling load; conversely, a transient temperature drop follows a rising load. This transient error reveals a dynamic imbalance in the system—on a drop in load, the exchanger temporarily needs less heat than the steam flow controller is allowing. In order to correct the transient temperature error, a further cutback in the flow of energy must be applied to the exchanger beyond what is required for the steady-state balance.

A more general explanation is provided through the use of Figure 2.9f. This figure shows the load and the manipulated

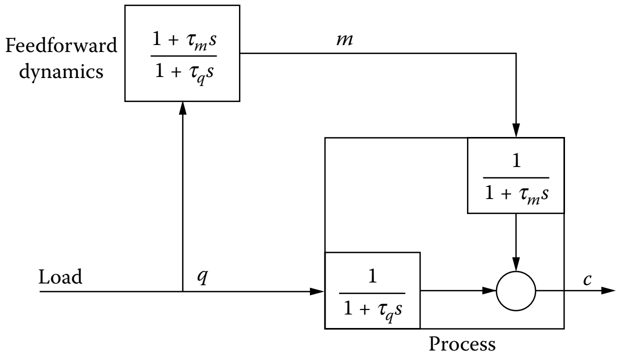


FIG. 2.9f
A feedforward dynamic model for a general process.

variable as entering the process at different points, where they encounter different dynamic elements. In the heat-exchanger example, the liquid enters the tubes while the steam enters the shell. The heat capacities of the two locations are different. As a result, the controlled variable (the liquid temperature) responds more rapidly to a change in liquid flow than to a change in steam flow. Thus, the lag time of the load input is less than that of the manipulated variable.

The objective of the feedforward system is to balance the manipulated variable against the load, providing the forward loop with compensating dynamic elements. Neglecting steady-state considerations, Figure 2.9f shows what the dynamic model must contain to balance a process having parallel first-order lags. The lag time τ_q of the load input must be duplicated in the forward loop, and the lag time τ_m in the manipulated input of the process must be cancelled. Thus the forward loop should contain a lag divided by a lag. Since the inverse of a lag is a lead, the dynamic compensating function is a *lead-lag*. The lead time constant should be equal to τ_m, and the lag time constant should equal τ_q. In the case of the heat exchanger, the fact that τ_m is longer than τ_q causes the temperature rise on a load decrease. This is the direction in which the load change would drive the process in the absence of control.

In transfer function form, the response of a lead-lag unit is

$$G(s) = \frac{1+\tau_1 s}{1+\tau_2 s} \qquad \textbf{2.9(10)}$$

where τ_1 is the lead time constant, τ_2 is the lag time constant, and s is the Laplace operator.

The frequency response of a feedforward loop is usually not significant, since forward loops cannot oscillate. The most severe test for a forward loop is a step change in load. The response of the lead-lag unit output y to a step in its input x is

$$y(t) = x(t)\left(1+\frac{\tau_1-\tau_2}{\tau_2}\varepsilon^{-t/\tau_2}\right) \qquad \textbf{2.9(11)}$$

Its maximum dynamic gain (at $t = 0$) is the lead/lag ratio τ_1/τ_2. The response curve then decays exponentially from this maximum at the rate of the lag time constant τ_2, with 63.2% recovery reached at $t = \tau_2$ as shown in Figure 2.9g. The figure shows some deadtime compensation added to the lead-lag function—however, it is not required for the heat exchanger.

When properly adjusted, the dynamic compensation afforded by the lead-lag unit can produce the controlled variable response having the form shown in Figure 2.9h. This is a signature curve of a feedforward-controlled process — since most processes do not consist of simple first-order lags, a first-order lead-lag unit cannot produce a perfect dynamic balance. Yet a major reduction in peak deviation is achieved over the uncompensated response, and the area can be equally distributed on both sides of the set point.

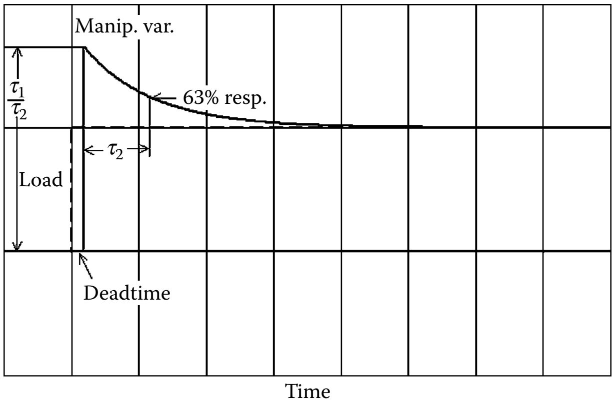


FIG. 2.9g
The step response of a lead-lag unit with deadtime compensation.

The lead-lag unit is applied to the flow signal as $[f(t)]$ in Figure 2.9i. Its settings are particular to that load input, and therefore it must be placed where it can modify that signal and no other. No dynamic compensator is included for T_1, as it is not expected to change quickly.

Tuning the Lead-Lag Unit First, a load change should be introduced without dynamic compensation by setting the lead and lag time constants at equal values or zero to observe the direction of the error. If the resulting response is in the direction that the load would drive it, the lead time should exceed the lag time; if not, the lag time should be greater. Next, measure the time required for the controlled variable to reach its maximum or minimum value. This time should be introduced into the lead-lag unit as the smaller time constant. Thus, if the lead dominates, this would be the lag setting. For example, in Figure 2.9e the temperature responds in the direction the load would drive it, and therefore the lead time setting should exceed the lag time setting. The time required for the temperature in Figure 2.9e to rise to its peak should be set equal to the smaller time constant, in this case equal to the lag time setting. Set the greater time constant at twice

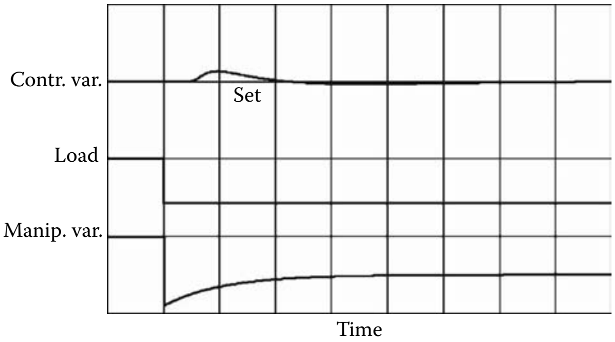


FIG. 2.9h
The typical step response of a dynamically compensated feedforward system.

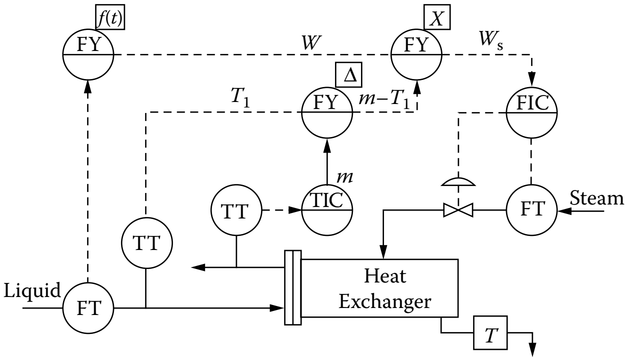


FIG. 2.9i
Feedforward–feedback control of the heat exchanger, with dynamic compensation.

this value and repeat the load change. If the peak deviation is reduced, but the curve is still not equally distributed around the set point, increase the greater time constant and repeat the load change.

The area (integrated error) of the response curve will be equally distributed about the set point if the difference between the lead and lag settings is correct. Once this equalization is obtained, both settings should be increased or decreased with their difference constant until a minimum error amplitude is achieved. When the controller is properly tuned, the steam energy equivalent of the transient (area between temperature and its set point) in Figure 2.9e should match the extra steam introduced (area above input step change) in Figure 2.9g.

If there is a significantly greater deadtime in the load path than in the manipulated-variable path, deadtime compensation may be added, which will reduce the peak deviation. However, if the deadtime in the manipulated-variable path is the longer of the two, exact compensation is impossible, and this is the case in the heat exchanger. Still, careful adjustment of the lead-lag compensator can result in zero integrated error.

Adding a Feedback Loop

Any offset resulting from steady-state errors can be eliminated by adding feedback. This can be done by replacing the feedforward set point with a controller, as shown in Figure 2.9i. The feedback controller adjusts the set point of the feedforward system in cascade, while the feedforward system adjusts the set point of the manipulated flow controller in cascade.

The feedback controller should have the same control modes as it would without feedforward control, but the settings need not be as tight. The feedback controller reacts to a disturbance by creating another disturbance in the opposite direction one-half cycle later. However, the feedforward system positions the manipulated variable so that the error in the controlled variable disappears. If acted upon by a tightly set feedback controller, the correct position calculated by feedforward will be altered, producing another disturbance that prolongs the settling time of the system.

The feedback controller must have the integral mode to eliminate any steady-state offset that might be caused by errors in the sensors, the model, or the calculations. However, the integrated error it sustains following a load change is markedly reduced compared to what it would be following a load change without feedforward. Now its output changes only an amount equal to the change in the *error* in the feedforward calculation. For small load changes, this change in error would approach zero, leading to an integrated error of zero, with or without dynamic compensation. In the presence of feedback, but without dynamic compensation, the transient in Figure 2.9e would be balanced by a following transient on the other side of set point, until the area is equalized. This usually prolongs the response without reducing the peak deviation. The dynamic compensator therefore needs to be tuned with the feedback controller in manual, to approach zero integrated error before adding feedback.

Linear and Bilinear Feedforward Systems

The relationship shown in the energy balance of Equation 2.9(8) is *bilinear*: steam flow is related to both liquid flow and to temperature rise in a linear manner, but as their product rather than their sum or difference. This distinction is crucial to the successful functioning of feedforward control because a bilinear system has variable gains. The feedforward gain—the ratio of steam flow to liquid flow—varies directly with the required temperature rise. This variable gain is accommodated in the system by the multiplier [×].

If inlet temperature were to vary so slowly that it did not require feedforward compensation or if it were not measured, then feedforward from flow alone would be required. Yet the required feedforward gain—the ratio W_s/W—would vary directly with $T_{2\text{set}} - T_1$. If either $T_{2\text{set}}$ or T_1 were to change, the feedforward gain should change with it. For a bilinear process, then, a linear feedforward system having a constant gain is a misfit—it is accurate at only one combination of secondary load and set point. To be sure, feedback control can trim any errors in the feedforward calculation and return the deviation to zero. But if the error is in the gain term, the feedback controller should adjust the feedforward gain, through a multiplier—otherwise the gain will be incorrect for all subsequent flow changes, resulting in only partial feedforward correction or even overcorrection.

Eliminating inlet temperature as a variable in Figure 2.9i also eliminates the difference block [Δ]. The controller output m then goes directly to the multiplier, and in reducing the temperature deviation to zero, m will assume a value representing $K(T_{2\text{set}} - T_1)$. If either $T_{2\text{set}}$ or T_1 were then to vary, the feedback controller would respond to the resulting deviation by changing its output until the deviation returns to zero. In so doing, it has changed the gain of the multiplier to the new value of $K(T_{2\text{set}} - T_1)$.

Control of composition is also a bilinear process. The required ratio of two flows entering a blender, for example, is a function of their individual compositions and that of the blend. Similarly, the ratio of product flow to feed rate or steam flow to feed rate in a distillation column or evaporator also varies with feed and product compositions. In general, temperature and composition loops are bilinear, and their flow ratios should always be adjusted through a multiplier. By contrast, pressure and level loops are linear, and their feedforward gain can be constant. An example of the latter is three-element drum-level control, where each unit of steam removed from the drum must be replaced by an equal unit of feedwater—the feedforward gain is constant at 1.0.

This distinction is crucial because many "advanced" multivariable control systems are based on a linear matrix with constant coefficients. The plant is tested to develop the coefficients that relate the variables, and the control matrix is built from these results. When the plant operating conditions move away from the original test conditions, the fixed coefficients in the control matrix may no longer represent the process relationships accurately, degrading the performance of the feedforward loops. This may require frequent retesting and recalibration of the matrix. Some multivariable systems contain multiple matrices, each corresponding to its own set of operating conditions and switched into service when those conditions develop.

Self-tuning feedforward control[3] is also available, applied either in a linear (additive) or bilinear (multiplicative) manner, as configured manually. The parameters that are tuned adaptively are the steady-state gain and a lag compensator. The steady-state gain is adjusted to minimize integrated error following a change in measured load, and the lag is adjusted to minimize integral-square error. Load changes must be sharp, clean steps, with settling time allowed between them, for tuning to proceed effectively. But once tuned, random disturbances can be accommodated. However, self-tuning is only recommended where precise modeling and flow control are unavailable.

Performance

The use of feedback in a feedforward system does not detract from the performance improvement that was gained by feedforward control. Without feedforward, the feedback controller was required to change its output to follow all changes in load. With feedforward, the feedback controller must only change its output by an amount equal to what the feedforward system fails to correct. If a feedforward system applied to a heat exchanger could control the steam flow to within 2% of that required by the load, the feedback controller would only be required to adjust its output to compensate for 2% of a load change, rather than the full amount. This reduction of Δm in Equation 2.9(7) by 50/1 results in the reduction of E by the same ratio. Reductions by 10/1 in errors resulting from load changes are relatively common, and improvements of 100/1 have been achieved in some systems.

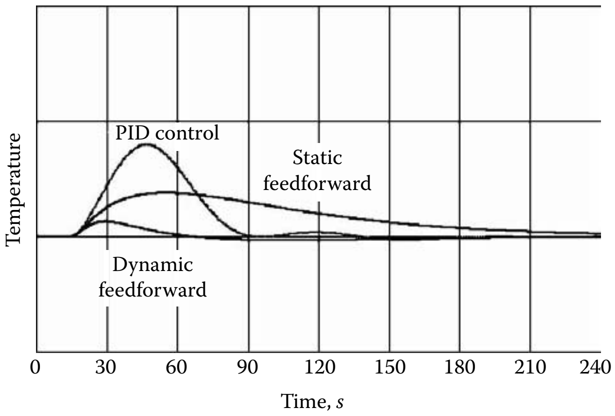


FIG. 2.9j
Comparison of feedforward control of a heat exchanger against optimally tuned PID feedback control.

Figure 2.9j illustrates the control performance of a steam-heated exchanger that has experienced a step decrease in process flow. The static feedforward model reduces the peak deviation substantially but extends settling time and does not improve on the integrated error. Dynamic compensation is seen to be essential in maximizing performance in this example. It is not perfect because the deadtime in the path of the manipulated variable is somewhat longer than that in the load path, and that lost time cannot be made up. Still, the lead-lag compensator has been tuned to eliminate any integrated error.

The feedforward system is more costly and requires more engineering effort than a feedback system does, so prior to design and installation, the control improvement it brings must be determined to be worthwhile. Most feedforward systems have been applied to processes that are very sensitive to disturbances and slow to respond to corrective action and to product streams that are relatively high in value. Distillation columns of 50 trays or more have been the principal systems controlled with this technology. Boilers, multiple-effect evaporators, direct-fired heaters, waste neutralization plants, solids dryers, compressors, and other hard-to-control processes have also benefited from feedforward control.

Variable Parameters

Heat exchangers are characterized by gain and dynamics that vary with flow. The settings of both the PID controller and the feedforward compensator that produced the results shown in Figure 2.9j were optimized for the final flow and would not be optimum for any other flow. Differentiation of the heat-balance equation, Equation 2.9(8), shows the process gain to vary inversely with process flow W:

$$\frac{dT_2}{dW_s} = \frac{\lambda}{WC_p} \qquad 2.9(12)$$

If PID control alone is applied, this gain variation can be compensated through the use of an equal-percentage steam valve, whose gain varies directly with steam flow. However, when the PID controller is combined with bilinear feedforward, its output passes through the multiplier shown in Figure 2.9i, where it is multiplied by process flow. This operation keeps the PID loop gain constant, while manipulating steam flow linearly.

Heat-exchanger dynamics—deadtime and lags—also vary inversely with process flow, as is typical of *once-through* processes, i.e., where no recirculation takes place. This problem is less easily solved. The PID controller must have its integral time set relative to the slowest response (lowest expected flow) and its derivative time set relative to its fastest response (highest expected flow), or ideally, programmed as a function of measured flow. Otherwise, instability may result at extreme flow rates. Ideally, lead and lag settings of the feedforward dynamic compensator should also be programmed as a function of measured flow. However, the penalty for not doing so is not severe. The lead-lag ratio is not subject to change because the lags on both sides of the process change in the same proportion with flow. The primary purpose of the dynamic compensator is to minimize the peak deviation following the load change, and this is accomplished by the dynamic gain of the compensator, which is its lead-lag ratio. Any subsequent variation in integrated error is eliminated by the PID controller.

Although parameter variations can also be accommodated by a self-tuning feedforward compensator and feedback controller, these methods are less accurate in arriving at optimum settings and are always late, having tuned for the last flow condition and not for the next one.

References

1. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, pp. 33–35.
2. Shinskey, F. G., "Feedforward Control Applied," *ISA Journal,* November 1963.
3. Hansen, P. D., Badavas, P. C., and Kinney, T. B., "New Adaptive Multivariable Controller," presented at the 3rd Intl. Control Eng. Conference, Chicago, 1994.

Bibliography

Butterworths, W. R., *Advanced Process Control,* New York: McGraw-Hill, 1981.

Gore, F., "Add Compensation for Feedforward Control," *Instruments and Control Systems,* March 1979.

Palmer, R., "Nonlinear Feedforward Can Reduce Servo Settling Time," *Control Engineering,* March 1978.

Seborg, D., et al., *Process Dynamics and Control,* New York: Wiley, 1989.

Shinskey, F.G., "Reflections on CPC-III," *Chemical Process Control-CPC III*, T. J. McAvoy, et al. (Eds.), Amsterdam: Elserier Sci. Pub., 1986.

Shinskey, F. G., "Smoothing Out Compressor Control," *Chemical Engineering,* February 1999.

2.10 Genetic and Other Evolutionary Algorithms

J. MADÁR, J. ABONYI (2005)

INTRODUCTION

The evolutionary algorithm (EA) uses the computational model of natural selection for optimization. EA has proved particularly successful in applications that are difficult to formalize mathematically and that are therefore not easily handled by the engineering tools of classical analysis.

EA does not rely on domain-specific heuristics and therefore is attractive for applications that are highly nonlinear, stochastic, or poorly understood; very little *a priori* information is required, although it can be utilized if so desired. For problems that are well understood, approximately linear, and for which reliable solutions exist, the EA is unlikely to produce competitive results.

It is possible for an EA to provide several dissimilar but equally good solutions to a problem, due to its use of a population. Hence, the EA is a robust search-and-optimization method that is able to cope with multimodality, discontinuity, time-variance, randomness, and noise.

A single control engineering problem can contain a mixture of "decision variable" formats (numbers, symbols, and other structural parameters). Since the EA operates on a "genetic" encoding of the optimized variables, diverse types of variables can be simultaneously optimized.

Each of the previously mentioned properties can prove significantly problematic for conventional optimizers. Furthermore, an EA search is directed and, hence, represents potentially much greater efficiency than a totally random or enumerative search.[1]

Real-time performance of an optimization algorithm is of particular interest to engineers. Unfortunately, there is no guarantee that the EA results will be of sufficient quality for use online. This is because EAs are very computationally intensive, often requiring massively parallel implementations in order to produce results within an acceptable timeframe.

In real-time applications there is also the matter of how individuals will be evaluated if no process model is available. Furthermore, EAs should be tested many times due to the algorithms' stochastic nature. Hence, online control of real-time processes, especially applications for safety-critical processes, is in most cases not yet feasible.

Applications

Evolutionary algorithms have been most widely and successfully applied to offline design applications. In the field of control systems engineering, these applications include controller design, model identification, robust stability analysis, system reliability evaluation, and fault diagnosis.

The main benefit of the EA is that it can be applied to a wide range of problems without significant modification. However, it should be noted that EA has several implementations: evolutionary programming (EP), evolutionary strategy (ES), genetic algorithm (GA), and genetic programming (GP), and the selection of the proper technique and the tuning of the parameters of the selected technique require some knowledge about these techniques.

Hence, the aim of this section is to introduce the readers to the basic concept of EAs, describe some of the most important implementations, and provide an overview of the typical process engineering applications, especially in the area of process control.

THE CONCEPT OF EA

Fitness Function: Encoding the Problem

EAs work with a population of potential solutions to a problem. In this population, each individual represents a particular solution, which can be described by some form of genetic code. The fitness value of the individual solution expresses the effectiveness of the solution in solving the problem. Better solutions are assigned higher values of fitness than the less well performing solutions.

The key of EA is that the fitness determines the success of the individual in propagating its genes (its code) to subsequent generations.

In practical system identification, process optimization, or controller design it is often desirable to simultaneously meet several objectives and stay within several constraints. For the purposes of the EA, these factors must be combined to form a single fitness value. The weighted-sum approach has proved popular in the literature because it is amenable to a solution by conventional EA methods, but Pareto-based

TABLE 2.10a
A Typical Evolutionary Algorithm

```
procedure EA;
{
    Initialize population;
    Evaluate all individuals;
    while (not terminate) do
    {
        Select individuals;
        Create offspring from selected individuals
            using Recombination and Mutation;
        Evaluate offspring;
        Replace some old individuals by some offspring;
    }
}
```

multi-objective techniques are likely to surpass this popularity in the future.

In some cases, the objectives and constraints of the problem may be noncommensurable and the objective functions explicitly/mathematically are not available. In these cases, interactive evolutionary computation (IEC) should be used to allow subjective human evaluation; IEC is an evolutionary algorithm whose fitness function is replaced by human users who directly evaluate the potential solutions.

Model of Natural Selection

The population has evolved over generations to produce better solutions to the task of survival. Evolution occurs by using a set of stochastic genetic operators that manipulate the genetic code used to represent the potential solutions. Most evolutionary algorithms include operators that select individuals for reproduction, produce new individuals based on the characteristics of those selected, and determine the composition of the population of the subsequent generation.

Table 2.10a outlines a typical EA. A population of individuals is randomly initialized and then evolved from generation to generation by repeated applications of *evaluation, selection, mutation* and *recombination.*

In the *selection* step, the algorithm selects the parents of the next generation. The population is subjected to “environmental pressure,” which means the selection of the fittest individuals. The most important automated selection methods are stochastic uniform sampling, tournament selection, fitness ranking selection, and fitness proportional selection:

- *Stochastic uniform sampling* (*SUS*) is the simplest selection method. In it, every individual has the same chance to be selected without considering individual fitness. This technique can be useful when the size of the population is small.
- The *tournament selection* method is similar to SUS, but the individuals that have higher fitness values have higher probabilities to be selected. The selection procedure is simple; in every tournament two individuals are picked randomly from the population, and the one with the higher fitness value is selected.
- *Fitness proportional selection* is the most often applied technique. In this selection strategy the probability of selection is proportional to the fitness of the individuals.[2]
- The *fitness ranking selection* method uses a rank-based mechanism. The population is sorted by fitness, and a linear ranking function allocates a rank value for every individual. The probability of selection is proportional to the normalized rank value of the individual.

After the selection of the individuals, the new individuals (offspring) of the next generation are created by recombination and mutation.

- *Recombination* (also called crossover) exchanges information between two selected individuals to create one or two new offspring.
- The *mutation* operator makes small, random changes to the genetic coding of the individual.

The final step in the evolutionary procedure is *replacement*, when the new individuals are inserted into the new population. Once the new generation has been constructed, the processes that resulted in the subsequent generation of the population are repeated once more.

GENETIC ALGORITHM

The GA, as originally defined by John Holland and his students in the 1960s,[3] uses bit-string representation of the individuals. Depending on the problem, the bit strings (chromosomes) can represent numbers or symbols (e.g., see Figure 2.10b).

Of course, this automatically raises the question as to what precision should be used, and what should be the mapping between bit strings and real values. Picking the right precision can be potentially important. Historically, genetic algorithms have typically been implemented using low precision, such as 10 bits per parameter.

Recombination means the swapping of string fragments between two selected parents (see Figure 2.10c), while mutation means the flip of a few bits in these strings.

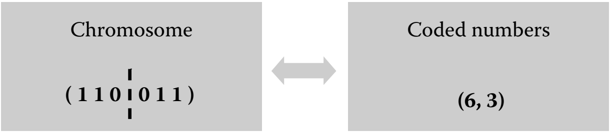


FIG. 2.10b
Example of binary representation of integer numbers.

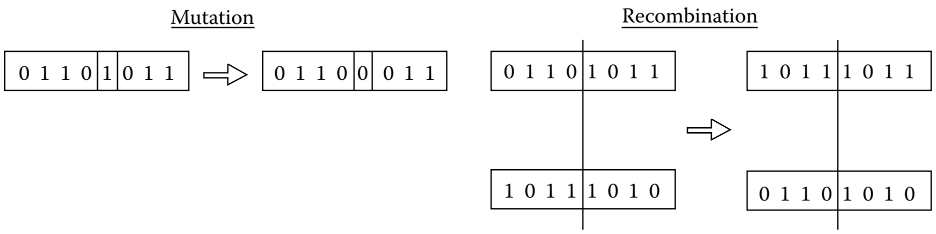


FIG. 2.10c
Mutation and recombination of binary strings.

Recombination has much higher probability than mutation, so recombination is often said to be the "primary searching operator."[3]

GA applies simple replacement technique: all the original individuals are replaced by the created offspring; except in case of the *elitist strategy* when some of the best individuals are also placed into the next generation.

GENETIC PROGRAMMING

GA is a powerful technique for numeric optimization, but because of the fixed length bit-string representation, it is not really suitable for solving structural and symbolic optimization problems. Genetic programming (GP), which has been developed by John Koza,[4] is based on tree representation. This representation is extremely flexible, since trees can represent computer programs, mathematical equations, or complete models of process systems.

A population member in GP is a hierarchically structured computer program consisting of functions and terminals. The functions and terminals are selected from a set of functions and a set of terminals. For example, function set F could contain the basic arithmetic operations: $F = \{*, -, +, /\}$. However, the function set may also include other mathematical functions, Boolean operators, conditional operators, or any user-defined operators.

The terminal set T contains the arguments for the functions, for example $T = \{y, x, p_i\}$ with x and y as two independent variables and p_i representing the parameters. The potential solution (program) can now be depicted as a rooted, labeled tree with ordered branches, using operations (internal points of the tree) from the function set and arguments (leaves of the tree) from the terminal set. An example of such a tree using the aforementioned function sets F and T is given in the top left corner of Figure 2.10d.

GP inherited the selection and replacement strategies from GA, but the mutation and crossover are adapted to the tree representation. The mutation exchanges a node in the tree, or deletes a sub-tree from the tree, or inserts a new sub-tree into the tree. Mutation can be implemented in two different ways:

- A random terminal or function is selected and replaced by another (randomly selected) terminal or function.

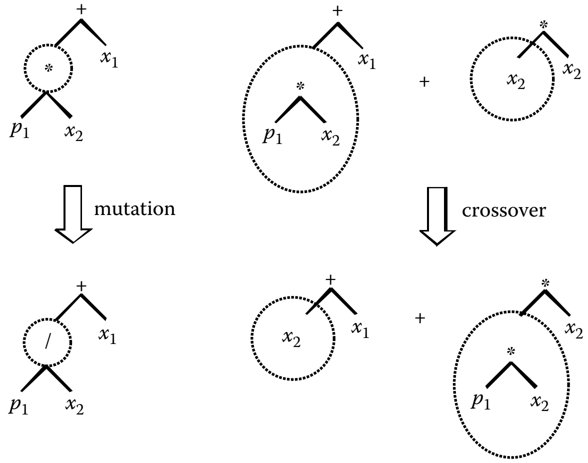


FIG. 2.10d
Mutation and crossover of trees in GP.

- A randomly selected sub-tree is replaced by a randomly generated sub-tree.

The crossover generates new children trees by jointing parental sub-trees. Crossover is determined by choosing two individuals based on fitness and generating for each tree the crossover point (node) at random.

For example: consider the trees shown at the top right of Figure 2.10d with crossover points 2 and 2. The sub-tree of the first solution (program) starting from crossover point 2 will be swapped with the sub-tree of the second program at crossover point 2, resulting in two new individuals shown at the bottom right of Figure 2.10d.

The size of tree can vary during mutation and crossover, which gives additional flexibility to GP. Hence, GP is more used for structure optimization than for parameter optimization. Koza[4.] has shown that GP can be applied to a great diversity of problems. He especially illustrated the use of GP with regard to optimal control problems, robotic planning, and symbolic regressions.

The key advantage of GP is the ability to incorporate domain-specific knowledge in a straightforward fashion. Another advantage is that the results can be readily understood and manipulated by the designer. However, GP structures can become complicated and may involve redundant pathways. Minimizing the complexity of a solution should be a specific objective, and the topic of “bloat” is a continuing area of study for GP researchers.[5]

Furthermore, GP can be quite processor intensive, especially for structural identification, where a parameter estimation procedure must be carried out for each individual structure at each generation. Due to the complexity of the structure, traditional (trusted and efficient) parameter estimation methods are often impossible to apply.

EVOLUTIONARY STRATEGY

Contrary to GP, ES searches in continuous space. The main distinction from GA is that ES uses real-valued representation of the individuals.

The individuals in ES are represented by n-dimensional vectors $(\mathbf{x} \in \Re^n)$, often referred to as object variables. To allow for a better adaptation to the particular optimization problem, the object variables are accompanied by a set of so-called strategy variables. Hence, an individual $\mathbf{a}_j = (\mathbf{x}_j, \boldsymbol{\sigma}_j)$ consists of two components, the object variables, $\mathbf{x}_j = [x_{j,1},\ldots,x_{j,i}]$, and strategy variables, $\boldsymbol{\sigma}_j = [\sigma_{j,1},\ldots,\sigma_{j,n}]$.

As in nature where small changes occur frequently but large ones only rarely, normally distributed random numbers are added as mutation operators to the individuals:

$$x_{j,i} = x_{j,i} + N(0, \sigma_{j,i}) \qquad \textbf{2.10(1)}$$

Before the update of the object variables, the strategy variables are also mutated using a multiplicative normally distributed process.

$$\sigma_{j,i}^{(t)} = \sigma_{j,i}^{(t-1)} \exp(\tau' N(0,1) + \tau N_i(0,1)) \qquad \textbf{2.10(2)}$$

with $\exp(\tau' N(0,1))$ as a global factor that allows an overall change of the mutability and $\exp(\tau N_i(0,1))$ allowing for individual changes of the mean step sizes $\sigma_{j,i}$. The parameters can be interpreted in the sense of global learning rates. Schwefel suggests to set them as

$$\tau' = \frac{1}{\sqrt{2n}} \qquad \tau = \frac{1}{\sqrt{2\sqrt{n}}} \qquad \textbf{2.10(3)}$$

Recombination in ES can be either sexual (local), where only two parents are involved in the creation of an offspring, or global, where up to the whole population contributes to a new offspring. Traditional recombination operators are discrete recombination, intermediate recombination, and geometric recombination, all existing in a sexual and global form. When *F* and *M* denote two randomly selected individuals from the μ parent population, the following operators can be defined:[6]

$$x_i' = \begin{cases} x_{F,i} & \text{no recombination} \\ x_{F,i} \quad \text{or} \quad x_{M,i} & \text{discrete} \\ (x_{F,i} + x_{M,i})/2 & \text{intermediate} \\ \sum_{k=1}^{\mu} \mathrm{x}_{k,i}/\mu & \text{global avarge} \end{cases} \qquad \textbf{2.10(4)}$$

$$\boldsymbol{\sigma}_i' = \begin{cases} \sigma_{F,i} & \text{no recombination} \\ \sigma_{F,i} \quad \text{or} \quad \sigma_{M,i} & \text{discrete} \\ (\sigma_{F,i} + \sigma_{M,i})/2 & \text{intermediate} \\ \sum_{k=1}^{\mu} \sigma_{k,i}/\mu & \text{global avarge} \end{cases} \qquad \textbf{2.10(5)}$$

The selection is stochastic in the ES. First we chose the best μ individuals to be parents, and then we select the parent-pairs uniformly randomly from these individuals.

The standard notations in this domain, $(\mu + \lambda)$ and (μ, λ), denote selection strategies in which the μ number of parents are selected to generate λ number of offspring. In case of (μ, λ) only the λ offspring are inserted into the subsequent generation (it means that λ is equal to the population size). In the case of $(\mu + \lambda)$ not only the λ offsprings but the μ parents are also inserted into the subsequent generation (it means that $\mu + \lambda$ is equal to the population size).

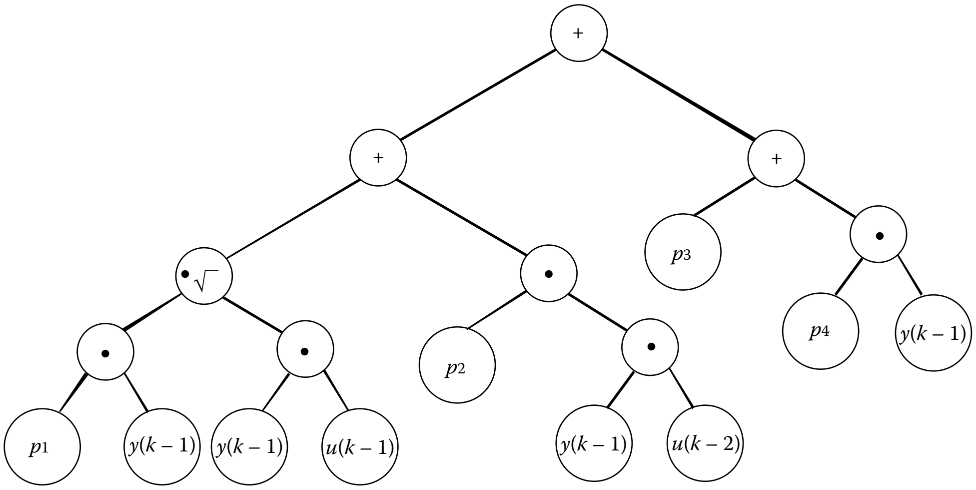
EVOLUTIONARY PROGRAMMING

Evolutionary programming (EP) was developed by Fogel et al.[7] independently from ES. Originally, it was used to achieve machine intelligence by simulated evolution. In the EP there are μ individuals in every generation. Every individual is selected and mutated (there is no recombination).

After the calculation of the fitness values of the new individuals, μ individuals are selected to form the next generation from the $\mu + \mu$ individuals, using a probabilistic function based on the fitness of the individuals.

SYSTEM IDENTIFICATION

System identification can be decomposed into two interrelated problems: selection of a suitable model structure and estimation of model parameters. Well-developed techniques exist for parameter estimation of linear models and linear-in-the-parameters nonlinear models. Techniques for nonlinear-in-the-parameters estimation and for the selection of the model structure are still the subject of ongoing research.



$$y(k) = -0.01568\, y(k-1)\sqrt{y(k-1)\, u(k-1)}$$
$$-6.404\, y(k-1)\, u(k-2) + 1.620\, y(k-1) + 9525$$

FIG. 2.10e
Tree representation of a solution.

The application of EAs to parameter identification of black-box and gray-box models has received considerable interest since the seminal papers by Kristinsson and Dumont[8] and Choi et al.[9]

One of the central problems in system identification is the choice of the input, output, and delay terms that are to contribute to the model. EAs provide a simple method for searching the structure space for terms that make the most significant contributions to process output.[10,11]

Multi-objective NARMAX (nonlinear autoregressive with exogenous input) polynomial model structure identification has been accomplished by the use of a multi-objective genetic programming (MOGP) strategy.[12] Here, seven objectives were optimized simultaneously: the number of model terms, model degree, model lag, residual variance, long-term prediction error, the auto-correlation function of the residuals, and the cross-correlation between the input and the residuals.

Polymerization Reactor Example

This example demonstrates how GP can be used for the development of a dynamic input–output model of a process from the input–output data of that process. The studied system is the dynamic model of a continuous polymerization reactor.[13]

The manipulated input variable in this process is the flow rate of the initiator, which influences the rate of the polymerization reaction. The controlled output is the mean molecular weight of the produced polymer (product quality).

The terminal set consists of the delayed process input and output terms. The function set simply consists of sum, product, and square root. The first two functions are sufficient to construct any NARMAX polynomial model, while the square root function is used on the basis of *a priori* knowledge about the reaction dynamics of the process.

In order to develop the model, input–output data were collected from the dynamic model of the process. Based on this process data, input–output models were identified. Figures 2.10e and 2.10f show the results.

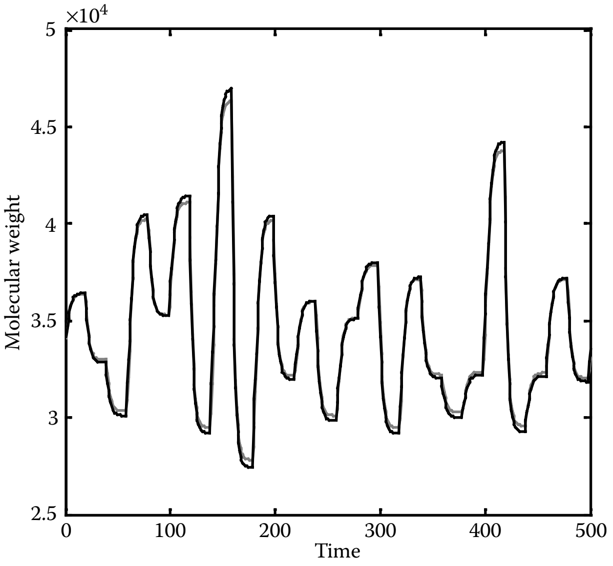


FIG. 2.10f
The results of model identification (GP). The solid line is the "real" output; the dotted line is the "estimated" output generated by the identified model.

Most of the resulting input-output models and estimates obtained by Rhodes and Morari[14] have identical input and output order. Hence, this result shows that GP is a useful tool for the data-driven identification of the model orders of unknown processes. In addition, the equations obtained by GP frequently had square-root terms, which were also in the original differential equations of the state-space model of the process.[13]

Hence, this example illustrates that the GP is not only good for the development of input–output models having good prediction capabilities, but because of the transparent model structure, these models are also useful tools to obtain additional knowledge about the process.

Complete Process Models

Gray et al.[15] performed nonlinear model structure identification using GP. They considered two representations: block diagrams (using Simulink) and equations (differential and integro-differential). A function library was constructed, which included basic linear and nonlinear functions and also specific *a priori* knowledge. The resulting scheme was applied to diverse systems of varying complexity, including simple transfer functions, a coupled water tank, and a helicopter rotor speed controller and engine.

In a paper by Marenbach et al., a general methodology for structure identification using GP with a block diagram library was developed.[16] Typical blocks included time delays, switches, loops, and domain-specific elements. Each block in the library was assigned a value, representing the subjective complexity of the block, and the method utilized an evaluation function that used this value. The developed methodology was applied to a biotechnological batch-type fermentation process, providing a transparent insight into the structure of the process.

PROCESS CONTROL APPLICATIONS

Controller Tuning

An early use of EAs was the tuning of proportional-integral-derivative (PID) controllers. Oliveira et al.[17] used a standard GA to determine initial estimates for the values of PID parameters. They applied their methodology to a variety of classes of linear time-invariant (LTI) systems, encompassing minimum phase, nonminimum phase, and unstable systems.

In an independent inquiry, Porter and Jones[18] proposed an EA-based technique as a simple, generic method of tuning digital PID controllers. Wang and Kwok[19] tailored an EA using inversion and preselection “micro-operators” to PID controller tuning. More recently, Vlachos et al.[20] applied an EA to the tuning of decentralized proportional-integral (PI) controllers for multivariable processes.

Onnen et al.[21] applied EAs to the determination of an optimal control sequence in model-based predictive control (MBPC), which is discussed in Section 2.17. Particular attention was paid to MBPC for nonlinear systems with input constraints. Specialized genetic coding and operators were developed with the aim of preventing the generation of infeasible solutions. The resulting scheme was applied to a simulated batch-type fermenter, with favorable results reported (compared to the traditional branch-and-bound method) for long control horizons.

The main problem of optimized controller tuning is the selection of the appropriate objective function. For this purpose classical cost functions based on the integral squared error (ISE) can be used. Usually, not only the control error is minimized but also the variation of the manipulated variable:

$$\min \sum_{k=1}^{n} e(k)^2 + \beta \cdot \sum_{k=1}^{n} \Delta u(k)^2 \qquad \textbf{2.10(6)}$$

The second term of this cost function is responsible for providing a smooth control signal, and in some control algorithms, such as model predictive control, stability.

The selection of the β weighting parameter, which balances the two objectives, is an extremely difficult design problem. Besides the selection of this weighting parameter, there are other freely selectable design parameters, e.g., the error can be weighted with the elapsed time after the setpoint change. As can be seen, the design of an appropriate cost function is not a straightforward task. The performance of the optimized controller does not always meet the expectations of the designer.

The following example will show that the application of IEC[22] is a promising approach to handle this problem; IEC is an evolutionary algorithm whose fitness function is replaced by human users who directly evaluate the potential solutions.

Application of IEC in Controller Tuning

The continuously stirred tank reactor (CSTR) control system model is taken from Sistu and Bequette.[23] The dynamic behavior of this system is interesting: it has multiple steady states as a function of the cooling water flow rate and reactor temperature. The reactor’s dynamic behavior is complex and nonlinear and as such, presents a challenging control problem.[24]

To control the reactor temperature, it is advantageous to use cascade-control, where the slave controls the jacket temperature, while the master controls the reactor temperature. According to the industrial practice, a PID controller in the master loop and a *P* (proportional) controller in the slave loop have usually been applied. If properly tuned, this selection of control modes should give good control performance.

This example demonstrates how ES can be used for the tuning of these controllers. The chromosomes consist of design and strategy variables. In this case study, the design variables ($\mathbf{x}_j$) are the three tuning settings of the master PID controller (gain, integral time and differential time) and the gain of the slave *P* controller.

Instead of using a fully automated ES optimization, interactive evolutionary strategy has been applied, where the user evaluates the performances of the controllers with the use of a process simulator incorporated into a graphical user interface tailored for this purpose. With this human–machine interface (HMI), the user can analyze the resulting plots and some numerical values, such as performance indices and parameters, and can select one or more individuals based on this information.

The prototype of this tool, which has been developed in MATLAB/Simulink, can be downloaded from www.fmt.vein.hu/softcomp/EAsy. The number of individuals is limited due to human fatigue; hence in this example the user can run, analyze, and evaluate eight independent tuning settings to provide feedback to the evolutionary process. Figure 2.10g shows the interactive figure that was used in this example.

The IEC converged quickly to good solutions; after only ten generations it resulted in well-tuned controllers (see Figure 2.10h).

For comparison, the direct optimization of the tuning settings by sequential quadratic programming (SQP) is considered. First, the cost function is based on the squared controller error and the variance of manipulated variable (Equation 2.10[6]). Although several β weighting parameters have been used, the SQP-technique led to badly tuned controllers.

Finally, the term that in the cost function contained the sum of the squared errors was replaced by a term based on the absolute value of the control errors. With the use of this new cost function, the SQP resulted in better controllers after a few experiments with different β weighting parameters. This example demonstrated that with the use of Interactive Evolutionary Computing we obtained the same controller performance as with a cost function–based controller tuning approach, but with much less effort and time.

Control Structure Design

Many EA applications simply optimize the parameters of existing controller structures. In order to harvest the full potential of EA, some researchers have experimented with the manipulation of controller structures themselves.

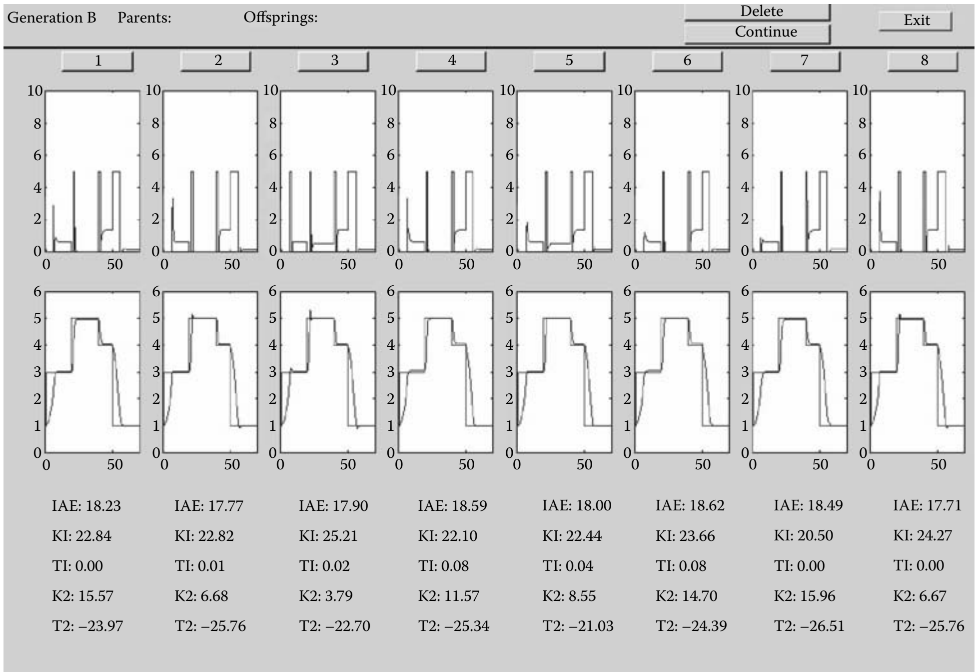
GP has been utilized for the automatic synthesis of the parameter values and the topology of controllers.[25] The system has reportedly duplicated existing patents (for PI and PID controllers) and rediscovered old ones (a controller making use of the second derivative of the error, which is the difference between the set point and the controlled variable, or in simulation language, the reference signal and the output signal).

Multi-objective evolutionary algorithms (MOEAs) have been utilized in controller structure optimization. For example, MOEAs have been used to select controller structure and suitable parameters for a multivariable control system for a gas turbine engine.[26]

Online Applications

Online applications present a particular challenge to the EA. The number of successful applications have been limited to date.

In online applications, it is important that at each sample instant, the controller must generate a signal to set the



```
IAE: 18.23   IAE: 17.77   IAE: 17.90   IAE: 18.59   IAE: 18.00   IAE: 18.62   IAE: 18.49   IAE: 17.71
KI: 22.84    KI: 22.82    KI: 25.21    KI: 22.10    KI: 22.44    KI: 23.66    KI: 20.50    KI: 24.27
TI: 0.00     TI: 0.01     TI: 0.02     TI: 0.08     TI: 0.04     TI: 0.08     TI: 0.00     TI: 0.00
K2: 15.57    K2: 6.68     K2: 3.79     K2: 11.57    K2: 8.55     K2: 14.70    K2: 15.96    K2: 6.67
T2: –23.97   T2: –25.76   T2: –22.70   T2: –25.34   T2: –21.03   T2: –24.39   T2: –26.51   T2: –25.76
```

FIG. 2.10g
IEC display.

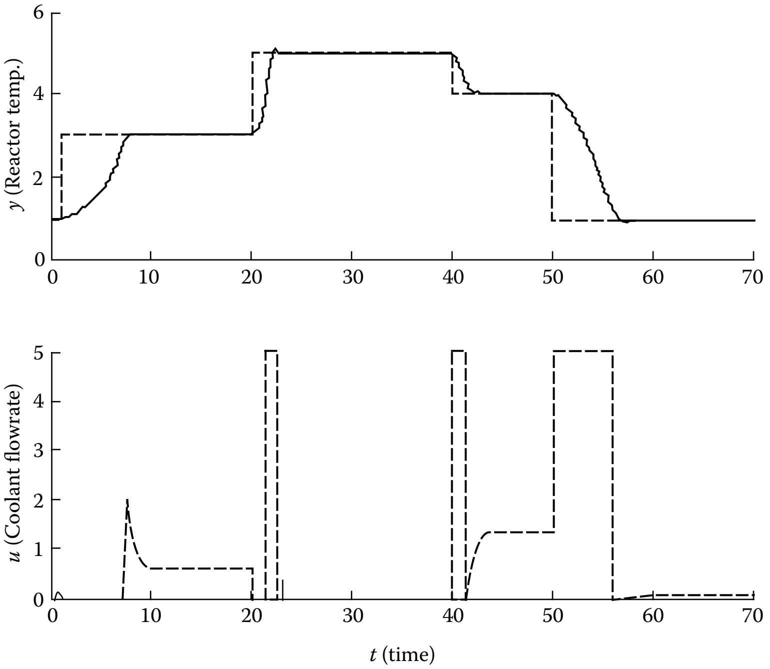


FIG. 2.10h

Control performance of cascade loop with controllers tuned on the basis of IEC. On the top: the dashed line represents changes in reactor temperature set point of the master controller and the solid line shows the actual temperature of the reactor. On the bottom: the cooling water flow rate is shown on the same time scale as on the top. This flow rate is manipulated by the slave controller.

manipulated variable(s). The actions of the "best" current individual of the EA may inflict severe consequences on the process. This is unacceptable in most applications, especially in the case of a safety- or mission-critical control system.

Given that it may not be possible to apply the values represented by any individual in an EA population to the system, it is clear that evaluation of the complete, evolving population cannot be performed on the actual process. The population may be evaluated using a process model, assuming that such a model exists, or performance may be inferred from system response to actual input signals.

Inference may also be used as a mechanism for reducing processing requirements by making a number of full evaluations and then computing estimates for the remainder of the population based on these results.

In a real-time application, there is only a limited amount of time available for an optimizer between decision points. Given current computing power, it is unlikely that an EA will execute to convergence within the sampling time limit of a typical control application. Hence, only a certain number of generations may be evolved. For systems with long sample times, an acceptable level of convergence may well be achieved.

In the case of a controller, an acceptable control signal must be provided at each control point. If during that period the EA evolved for only a few generations, then population performance may still be poor. A further complication is that the system, seen from the perspective of the optimizer, is changing over time. Thus, the evolved control signal at one instant can become totally inappropriate at the next.

EAs can cope with time-varying landscapes to a certain extent, but a fresh run of the algorithm may be required. In this instance, the initial population can be seeded with previous "good" solutions. Note that this does not guarantee fast convergence and may even lead to premature convergence.

There are three broad approaches to the use of EAs for online control:[27]

- Utilize a process model.
- Utilize the process directly.
- Permit restricted tuning of an existing controller.

The last approach can be used to ensure stability when combined with some form of robust stability analysis, while permitting limited exploration.

SOFTWARE TOOLS

Here an overview is provided of the EA implementation environments, based on the taxonomy of Filho,[28] which utilizes three major classes: application-oriented systems, algorithm-oriented systems, and tool kits (see Tables 2.10i and 2.10j).

Application-Oriented Systems

Most potential users of a novel computing technique, such as genetic algorithms, are not interested in the details of that

TABLE 2.10i
Programming Environments and Their Categories

	Algorithm-Oriented		*Tool Kits*	
Application-Oriented	*Algorithm-Specific*	*Algorithm Libraries*	*Educational*	*General Purpose*
EVOLVER	ESCAPADE			EnGENEer
OMEGA	GAGA	EM	GA Workbench	GAME
PC/BEAGLE	GAUCSD			MicroGA
XpertRule GenAsys	GENESIS	OOGA		PeGAsusS
	GENITOR			Splicer

technique, only in the contribution it can make to their applications. By using an application-oriented programming environment, it is possible to configure a particular application without having to know the encoding technique or the genetic operators involved.

Application-oriented systems follow many innovative strategies. Systems such as PC/BEAGLE and XpertRule GenAsys are expert systems using GAs to generate new rules to expand their knowledge base of the application domain. EVOLVER is a companion utility for spreadsheets, and systems such as OMEGA are targeted at financial applications.

EVOLVER is an add-on utility that works within the Excel, WingZ, and Resolve spreadsheets on Macintosh and PC computers. It is being marketed by Axcélis, Inc., which describes it as "an optimization program that extends mechanisms of natural evolution to the world of business and science applications."

The user starts with a model of a system in the spreadsheet and calls the EVOLVER program from a menu. After filling a dialogue box with the information required (e.g., cell to minimize/maximize) the program starts working, evaluating thousands of scenarios automatically until it is sure it has found an optimal answer.

The program runs in the background, freeing the user to work in the foreground. When the program finds the best result, it notifies the user and places the values into the spreadsheet for analysis. This is an excellent design strategy given the importance of interfacing with spreadsheets in business.

TABLE 2.10j
Programming Environment Lists

EM—Evolution Machine	http://www.amspr.gfai.de/em.htm
ESCAPADE	Frank Hoffmeister, iwan@ls11.informatik.uni-dortmund.de
EnGENEer	Logica Cambridge Ltd., Betjeman House, 104 Hills Road, Cambridge CB2 1LQ, UK
EVOLVER	http://www.palisade.com/html/evolver.html
GA Workbench	ftp://wuarchive.wustl.edu, mrh@i2ltd.demon.co.uk
GAGA	ftp://cs.ucl.ac.uk/darpa/gaga.shar, jon@cs.ucl.ac.uk
GAUCSD	http://www-cse.ucsd.edu/users/tkammeye/
GAME	ftp://bells.cs.ucl.ac.uk/papagena/, zeluiz@cs.ucl.ac.uk
GENESIS	ftp://ftp.aic.nrl.navy.mil/pub/galist/src/genesis.tar.Z, gref@aic.nrl.navy.mil
GENITOR	ftp://ftp.cs.colostate.edu/pub/GENITOR.tar, whitley@cs.colostate.edu
MicroGA	Steve Wilson, Emergent Behavior, 953 Industrial Avenue, Suite 124, Palo Alto, CA 94301, USA, e-mail: steve@emer.com
OMEGA	http://www.kiq.com/kiq/index.html
OOGA	The Software Partnership, P.O. Box 991, Melrose, MA 2176, USA
PC-BEAGLE	Richard Forsyth, Pathway Research Ltd., 59 Cranbrook Road, Bristol BS6 7BS, UK
PeGAsuS	http://borneo.gmd.de/AS/pega/index.html
Splicer	http://www.openchannelfoundation.org/projects/SPLICER/, bayer@galileo.jsc.nasa.gov
XpertRule GenAsys	http://www.attar.com

The OMEGA Predictive Modeling System, marketed by KiQ Limited, is a powerful approach to developing predictive models. It exploits advanced genetic algorithm techniques to create a tool that is “flexible, powerful, informative, and straightforward to use.” OMEGA is geared to the financial domain.

The environment offers facilities for automatic handling of data; business, statistical, or custom measures of performance; simple and complex profit modeling; validation sample tests; advanced confidence tests; real-time graphics, and optional control over the internal genetic algorithm.

PC/BEAGLE, produced by Pathway Research Ltd., is a rule-finder program that applies machine-learning techniques to create a set of decision rules for classifying examples previously extracted from a database.

XpertRule GenAsys is an expert system shell with embedded genetic algorithms, marketed by Attar Software. This GA expert system is targeted to solve scheduling and design applications. The system combines the power of genetic algorithms in evolving solutions with the power of rule-base programming in analyzing the effectiveness of solutions.

Some examples of design and scheduling problems that can be solved by this system include: optimization of design parameters in electronic and avionics industries, route optimization in the distribution sector, and production scheduling in manufacturing.

Algorithm-Oriented Systems

Algorithm-oriented systems are programming systems that support specific genetic algorithms. They subdivide into:

- Algorithm-specific systems—which contain a single EA; the classic example is GENESIS.
- Algorithm libraries—where a variety of EAs and operators are grouped in a library, as in Lawrence Davis’s OOGA.

Algorithm-oriented systems are often supplied in source code and can be easily incorporated into user applications.

Algorithm-Specific Systems Algorithm-specific environments embody a single powerful genetic algorithm. These systems typically have two groups of users: system developers requiring a general-purpose GA for their applications and researchers interested in the development and testing of a specific algorithm and genetic operators. The code has been developed in universities and research centers and is available free over worldwide computer research networks.

The best-known programming system in this category is the pioneering GENESIS system, which has been used to implement and test a variety of new genetic operators. In Europe, probably the earliest algorithm-specific system was GAGA. For scheduling problems, GENITOR is another influential system that has been successfully used.

GAUCSD allows parallel execution by distributing several copies of a GENESIS-based GA into UNIX machines in a network. Finally, ESCAPADE employs a somewhat different approach, as it is based on an evolutionary strategy.

Algorithm Libraries These systems are modular, allowing the user to select a variety of algorithms, operators, and parameters to solve a particular problem. Their parameterized libraries provide the ability to use different models (algorithms, operators, and parameter settings) to compare the results for the same problem.

New algorithms coded in high level languages, such as “C” or Lisp, can be easily incorporated into the libraries. The user interface is designed to facilitate the configuration and manipulation of the models as well as to present the results in different shapes (tables, graphics, etc.).

The two leading algorithm libraries are EM and OOGA. Both systems provide a comprehensive library for genetic algorithms, and EM also supports strategies for evolution simulation. In addition, OOGA can be easily tailored for specific problems. It runs in Common Lisp and CLOS (Common Lisp Object System), an object-oriented extension of Common Lisp.

Tool Kits

Tool kits are programming systems that provide many programming utilities, algorithms, and genetic operators, which can be used for a wide range of application domains. These programming systems can be subdivided into:

- Educational systems—to help the novice user obtain a hands-on introduction to GA concepts. Typically, these systems support a small set of options for configuring an algorithm.
- General-purpose systems—to provide a comprehensive set of tools for programming any GA and application. These systems may even allow the expert user to customize any part of the software, as in Splicer.

Educational Systems Educational programming systems are designed for the novice user to obtain a hands-on introduction to genetic algorithms concepts. They typically provide a rudimentary graphic interface and a simple configuration menu. Educational systems are typically implemented on PCs for reasons of portability and low cost. For ease of use, they have a nice graphical interface and are fully menu-driven. GA Workbench is one of the best examples of this class of programming environments.

General-Purpose Programming Systems General-purpose systems are the ultimate in flexible GA programming systems. Not only do they allow users to develop their own GA applications and algorithms, but they also provide users with the opportunity to customize the system to suit their own purposes.

These programming systems provide a comprehensive tool kit, including:

- A sophisticated graphic interface
- A parameterized algorithm library
- A high-level language for programming GAs
- An open architecture

Access to the system components is via a menu-driven graphic interface and a graphic display/monitor. The algorithm library is normally "open," allowing the user to modify or enhance any module. A high-level language—often object-oriented—may be provided to support the programming of GA applications, algorithms, and operators through specialized data structures and functions. Lastly, due to the growing importance of parallel GAs, systems provide translators to parallel machines and distributed systems, such as networks of workstations.

The number of general-purpose systems is increasing, stimulated by growing interest in the application of GAs in many domains. Examples of systems in this category include Splicer, which presents interchangeable libraries for developing applications, *Micro*GA, an easy-to-use object oriented environment for PCs and Macintoshes, and parallel environments such as EnGENEer, GAME, and PeGAsusS.

References

1. Rechenberg, I., *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution,* Stuttgart: Frommann-Holzboog, 1973.
2. Spears, W. M., De Jong, K. A., Back, T., Fogel, D. B., and Garis, H., "An Overview of Evolutionary Computation," European Conference on Machine Learning, 1993.
3. Holland, J. H., *Adaptation in Natural and Artificial Systems,* Ann Arbor: University of Michigan Press, 1975.
4. Koza, J. R., *Genetic Programming: On the Programming of Computers by Means of Natural Evolution,* Cambridge, MA: MIT Press, 1992.
5. Langdon, W. B., and Poli, R., "Fitness Causes Bloat," *Proceedings of the Second On-Line World Conference on Soft Computing in Engineering Design and Manufacturing*, London: Springer-Verlag, 1997.
6. Mandischer, M., "A Comparison of Evolutionary Strategies and Backpropagation for Neural Network Training," *Neurocomputing,* Vol. 42, 2002, pp. 87–117.
7. Fogel, L. J., Owens, A. J., and Walsh, M. J., *Artificial Intelligence through Simulated Evolution,* New York: Wiley, 1966.
8. Kristinsson, K., and Dumont, G. A., "System Identification and Control Using Genetic Algorithms," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 22, No. 5, 1992, pp. 1033–1046.
9. Choi, S. H., Lee, C. O., and Cho, H. S., "Friction Compensation Control of an Electropneumatic Servovalve by Using an Evolutionary Algorithm," *Proceedings of the Institution of Mechanical Engineers Part I,* 2000, pp. 173–184.
10. Fonseca, C. M., Mendes, E. M., Fleming, P. J., and Billings, S. A., "Non-Linear Model Term Selection with Genetic Algorithms," *IEE/IEEE Workshop on Natural Algorithms in Signal Processing*, Vol. 2, 1993, pp. 27/1–27/8.
11. Luh, G. C., and Wu, C. Y., "Non-Linear System Identification Using Genetic Algorithms," *Proceedings of the Institution of Mechanical Engineers Part I,* Vol. 213, 1999, pp. 105–118.
12. Rodriguez-Vasquez, K., Fonseca, C. M., and Fleming P. J., "Multiobjective Genetic Programming: A Nonlinear System Identification Application," *Late-Breaking Papers at the 1997 Genetic Programming Conference,* Stanford Bookstore, USA, 1997, pp. 207–212.
13. Doyle, F. J., Ogunnaike, B. A., and Pearson, R. K., "Nonlinear Model-Based Control Using Second-Order Volterra Models," *Automatica,* Vol. 31, 1995, pp. 697–714.
14. Rhodes, C., and Morari, M., "Determining the Model Order of Nonlinear Input/Output Systems," *AIChE Journal,* Vol. 4, No. 1, 1998, pp. 151–163.
15. Gray, G. J., Murray-Smith, D. J., Li, Y., Sharman, K. C., and Weinbrunner, T., "Nonlinear Model Structure Identification Using Genetic Programming," *Control Engineering Practice*, Vol. 6., No. 11, 1998, pp. 1341–1352.
16. Marenbach, P., Bettenhausen, K. D., Freyer, S., Nieken, U., and Rettenmaier, H., "Data-Driven Structured Modeling of a Biotechnological Fed-Batch Fermentation by Means of Genetic Programming," *Proceedings of the Institution of Mechanical Engineers Part I,* Vol. 211, 1997, pp. 325–332.
17. Oliveira, P., Sequeira, J., and Sentieiro, J., "Selection of Controller Parameters Using Genetic Algorithms," *Engineering Systems with Intelligence. Concepts, Tools, and Applications,* S. G. Tzafestas (Ed.), Dordrecht, The Netherlands: Kluwer Academic Publishers, 1991, pp. 431–438.
18. Porter, B., and Jones, A. H., "Genetic Tuning of Digital PID Controllers," *Electronics Letters,* Vol. 28, No. 9, 1992, pp. 843–844.
19. Wang, P., and Kwok, D. P., "Autotuning of Classical PID Controllers Using an Advanced Genetic Algorithm," *International Conference on Industrial Electronics, Control, Instrumentation and Automation (IECON 92),* Vol. 3, 1992, pp. 1224–1229.
20. Vlachos, C., Williams, D., and Gomm, J. B., "Genetic Approach to Decentralised PI Controller Tuning for Multivariable Processes," *IEE Proceedings—Control Theory and Applications*, Vol. 146, No. 1, 1999, pp. 58–64.
21. Onnen, C., Babuska, R., Kaymak, U., Sousa, J. M., Verbruggen, H. B., and Isermann, R., "Genetic Algorithms for Optimization in Predictive Control," *Control Engineering Practice,* Vol. 5, No. 10, 1997, pp. 1363–1372.
22. Takagi, H., "Interactive Evolutionary Computation — Cooperation of Computational Intelligence and Human KANSEI," *5th Int. Conf. on Soft Computing (IIZUKA'98),* Iizuka, Fukuoka, Japan, 1998, pp. 41–50.
23. Sistu, P. B., and Bequette, B. W., "Nonlinear Predictive Control of Uncertain Processes: Application to a CSTR," *AIChE Journal,* Vol. 37, 1991, pp. 1711–1723.
24. Cho, W., Lee, J., and Edgar, T. F., "Control System Design Based on a Nonlinear First-Order Plus Time Delay Model," *J. Proc. Vont.,* Vol. 7, 1997, pp. 65–73.
25. Koza, J. R., Keane, M. A., Yu, J., Bennett, F. H. III, and Mydlowec, W., "Automatic Creation of Human-Competitive Programs and Controllers by Means of Genetic Programming," *Genetic Programming and Evolvable Machines,* Vol. 1, 2000, pp. 121–164.
26. Chipperfield, A., and Fleming, P., "Multiobjective Gas Turbine Engine Controller Design Using Genetic Algorithms," *IEEE Transactions on Industrial Electronics,* Vol. 43, No. 5, 1996, pp. 1–5.
27. Linkens, D. A., and Nyongesa, H. O., "Genetic Algorithms for Fuzzy Control — Part 1: Offline System Development and Application" and "Part 2: Online System Development and Application," *IEE Proceedings—Control Theory and Applications,* Vol. 142, No. 3, 1996, pp. 161–176 and 177–185.
28. Filho, J. R., Alippi, C., and Treleaven, P., "Genetic Algorithm Programming Environments," *IEEE Comp. Journal,* Vol. 27, 1994, pp. 28–43.

Bibliography

Banzhaf, W., Nordin, P., Keller, R. E., and Francone, F. D., *Genetic Programming: An Introduction: On the Automatic Evolution of Computer Programs and Its Applications,* St. Louis, MO: Morgan Kaufmann, 1997.

Dasgupta, D., and Michalewicz, Z., *Evolutionary Algorithms in Engineering Applications*, Heidelberg: Springer-Verlag, 1997.

Fleming, P. J., and Purshouse, R. C., "Evolutionary Algorithms in Control Systems Engineering: a Survey," *Control Engineering Practice,* Vol. 10, 2002, pp. 1223–1241.

Goldberg, D. E., *The Design of Innovation*, Dordrecht, The Netherlands: Kluwer Academic Publishers, 2002.

Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning,* Reading, MA: Addison-Wesley, 1989.

Mitchell, M., *An Introduction to Genetic Algorithms (Complex Adaptive Systems),* Cambridge, MA: MIT Press, 1996.

Sette, S., and Boullart, L., "Genetic Programming: Principles and Applications," *Engineering Application of Artificial Intelligence,* Vol. 14, 2001, pp. 727–736.

Tang, K. S., Man, K. F., Kwong, S., and He, Q., "Genetic Algorithms and Their Applications," *IEEE Signal Processing Magazine,* November 1996, pp. 22–37.

Whitley, D., "An Overview of Evolutionary Algorithms: Practical Issues and Common Pitfalls," *Information and Software Technology,* Vol. 43, 2001, pp. 817–831.

2.11 Hierarchical Control

J. W. BERNARD (1970) **T. J. WILLIAMS** (1985, 1995) **B. G. LIPTÁK** (2005)

In modern, computer-controlled control systems (Figure 2.11a), several layers of control can be superimposed onto the basic controls, which are physically connected to the process. Such hierarchical control uses several levels of computer systems in an extended master–slave relationship to carry out not only the process control but also the supervisory, management and business functions in controlling industrial plants. This section describes the historical evolution of hierarchical control and some typical forms of such control systems, including the distribution of the potential duties among the several levels of computers involved.

HIERARCHICAL LEVELS

There are basically four levels in the hierarchy:

1. All direct connections to the process are uniformly referred to as Level 1. These basic controls include all sensors, final control elements, continuous PID controls, sequential interlocks, alarms, and monitoring devices. In some systems, where several levels of basic controls exist, they are referred to as Levels 1A and 1B.
2. Supervisory control is universally called Level 2. The control functions of this level are targeted to increased production, improved environmental protection, energy conservation, optimization, predictive maintenance, and plant-wide safety. The means used to achieve these goals include envelope control, model-based controls, fuzzy logic, statistical process control, and other types of expert systems, which are all described in this section.
3. In some plants the area controls or inter-area coordination is called Level 3, while in other plants it is included in Level 2. From this example it can be seen that not even the language has yet been standardized; this whole field is still in a process of evolution.

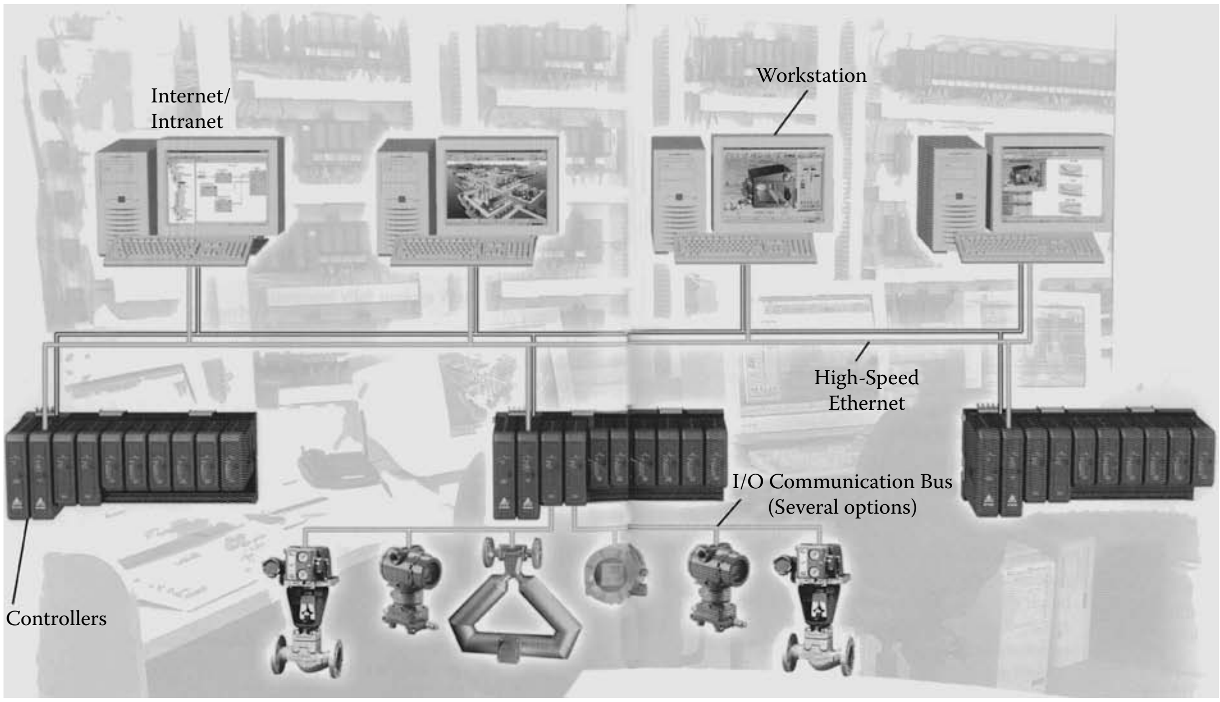


FIG. 2.11a
Modern control systems are well suited for hierarchical control because Level 1 controls, which consist of devices with direct connections to the process, can be cascaded to supervisory and management levels of control. (Courtesy of Control Associates, Inc.)

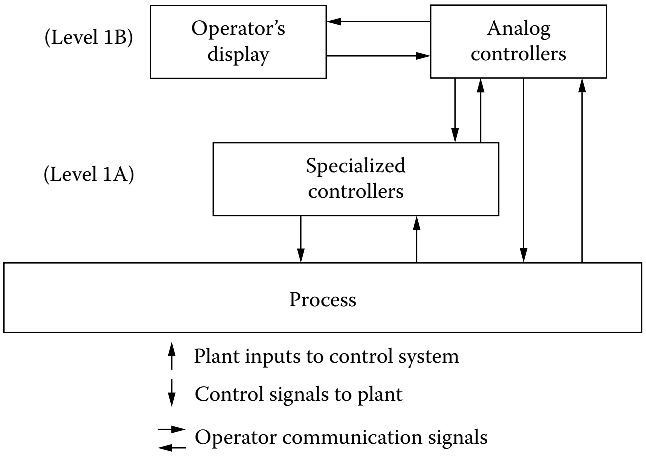


FIG. 2.11b
The basic old analog control system operated at Level 1 only.

4. In some plants the management functions are referred to as Level 3, while in others the scheduling and management information functions are called Level 4. Often this level is split into Levels 4A and 4B, where 4A covers contact with the plant (scheduling, etc.) and 4B covers contact with upper management. Management functions include marketing, financing, human resources, production scheduling, and inventory tasks. Some generalized software packages already exist for enterprise resource planning (ERP) and material requirement planning (MRP).

HISTORY

Prior to the introduction of the use of computers for industrial control applications, the standard industrial control system consisted of a large number of single-loop analog controllers, either pneumatic or electronic, as diagrammed in Figure 2.11b. Note in the figure that Level 1A refers to those instruments that have no signal readouts that are readily available to the operator (e.g., some field-mounted controllers, ratio instruments, etc.). Level 1B devices have transmitters and provide indication to the operator by a pointer or on a strip chart, etc. These latter devices include the standard PID (proportional, integral, and derivative control mode) controllers.

Computers were first used in industrial plants not as controllers but as monitors or data loggers (see Figure 2.11c). However, these installations were quickly followed by the first supervisory controllers (early 1960s), which used the computer to change the set points of the analog controllers (Figure 2.11d).

It was the desire of most control engineers who were working with early computer control systems to completely bypass the analog controllers of Figure 2.11d and have the computer directly control the valves. This resulted in direct digital control (DDC), as illustrated in Figure 2.11e and Figure 2.11f. Figure 2.11f includes specialized dedicated digital controllers to illustrate the relationship to Figure 2.11b. Although common today, these devices were not available in the earliest DDC systems. In these early systems, all computational work was carried out in the digital computer of Level 2 (Figure 2.11d).

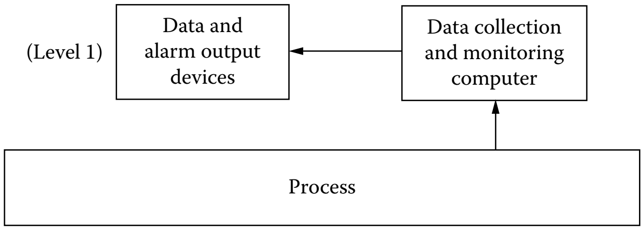


FIG. 2.11c
The early computers served only for data collection and monitoring.

Although direct digital control always had the potential for assuming an unlimited variety of complex automatic control functions, most early DDC systems were implemented as digital approximations of conventional three-mode analog controllers. Thus the computer system of Figure 2.11f is considered most often as a direct substitution for the analog control system of Figure 2.11b.

The Central Computer

The early digital control systems suffered from many drawbacks. Among these were:

1. The computers were all drum machines, i.e., the memory was located on a magnetic drum. As a result, they were very slow. A typical addition time for such machines was 4 milliseconds.
2. The memories were very small, usually 4000 to 8000 words of 12 to 16 bits each.
3. All programming had to be done in machine language.

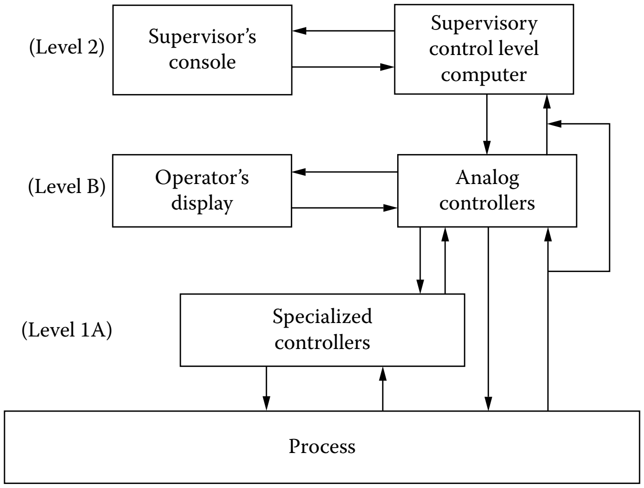


FIG. 2.11d
Level 2 supervisory computer used to adjust the set-points at Level 1.

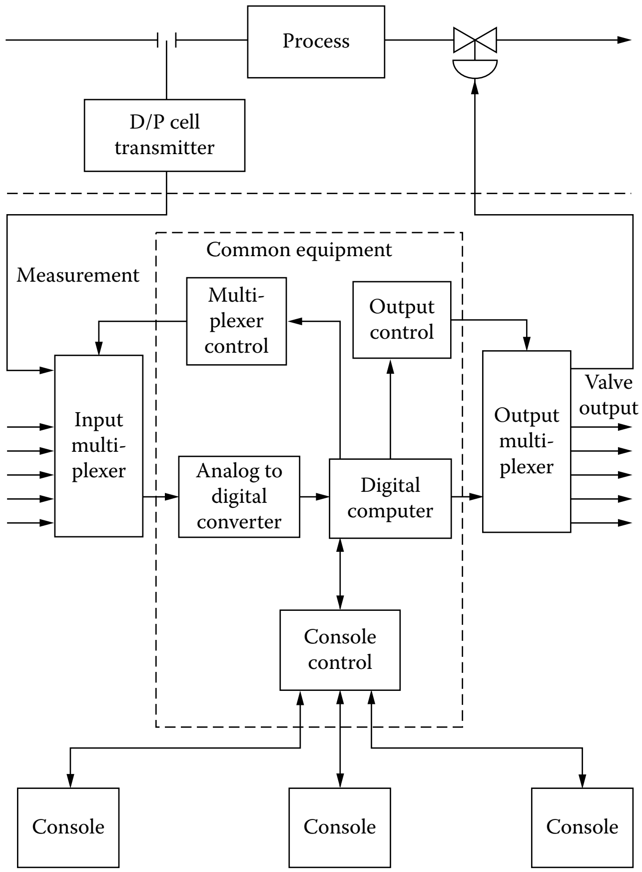


FIG. 2.11e
Block diagram of a direct control system.

4. Neither vendors nor users had any experience in computer applications. Thus, it was very difficult to size the project properly within computer capabilities. Most projects had to be reduced in size to fit the available machines.
5. Many of the early computer systems were very unreliable, particularly if they used temperature-sensitive germanium rather than silicon circuitry, as many did. In addition, these computers depended on unreliable mechanical devices, such as air conditioners, for their successful operation.

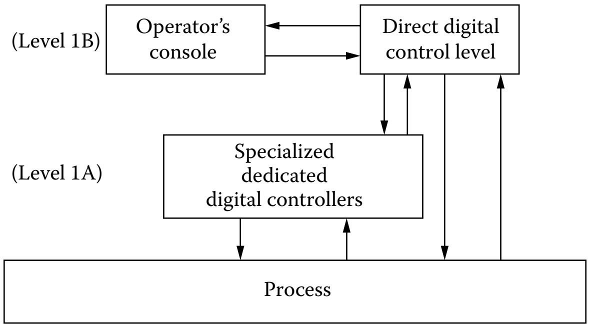


FIG. 2.11f
Digital substitution for the analog system in Figure 2.11b.

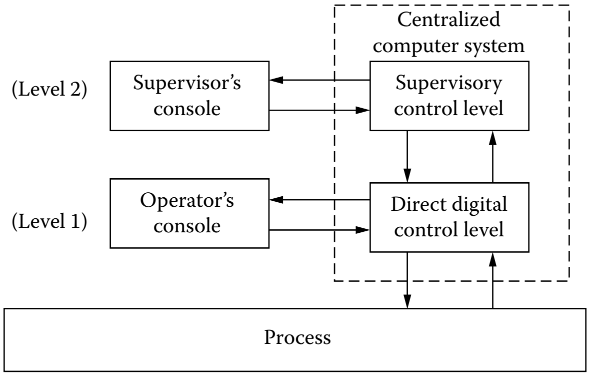


FIG. 2.11g
Two-level hierarchical control with supervisory control (Level 2) and direct digital control (Level 1).

In response to these problems, the computer manufacturers came out with a second-generation computer, which was much larger and had a magnetic core memory, wired-in arithmetic functions, etc., which made it much faster. But the high cost of core memories and the additional electronic circuitry made the system much more expensive. To help justify the cost, vendors advocated the incorporation of all types of computer functions, including both supervisory and DDC, into a single computer or mainframe in the central control room (Figures 2.11g and 2.11h.)

Although these computers had much greater speed and memory size compared to earlier systems, their use led to still further problems:

1. Most were marketed before their designs were thoroughly proven or their programming aids (compilers, higher-level languages, etc.) were fully developed. Thus there were many frustrating delays in getting them installed and running.
2. The vast amount of wiring and communication system required to bring the plant signals to the centralized computer location and return the control signals to the field was very expensive, and unless it was very carefully designed and installed, it was prone to electrical noise problems.
3. Because all of the control functions were located in one computer, the possibility that the computer might fail resulted in demands for a complete analog backup system. The resulting system, illustrated in Figure 2.11i, was a combination of the systems shown in Figures 2.11d and 2.11g, and it greatly increased the cost.
4. To compensate for these high costs, users and vendors alike attempted to squeeze as many tasks as possible into the computer system, thus drastically complicating the programming and aggravating the difficulties described in item 1 above.

As a result of these difficulties, management in many companies reacted disapprovingly to computer control, and there was a hiatus in computer system installations until the 1970s.

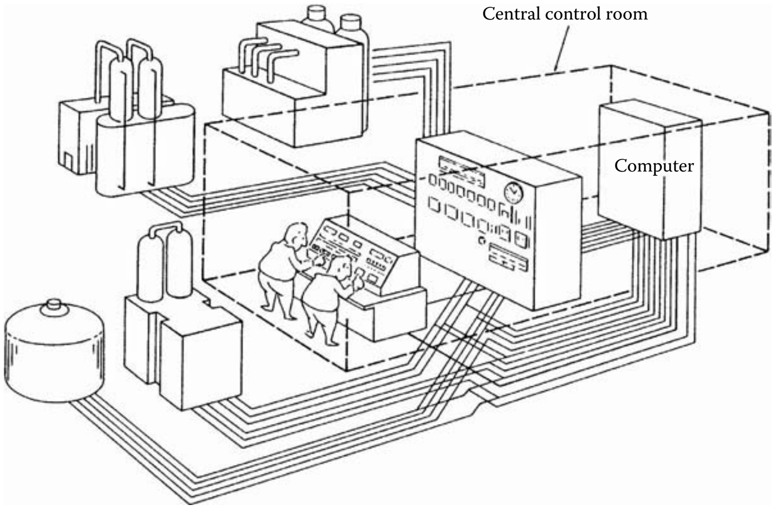


FIG. 2.11h
The use of centralized computer in a supervisory mode over analog controls located on an analog panel board and on a backup board (circa 1965–1975).

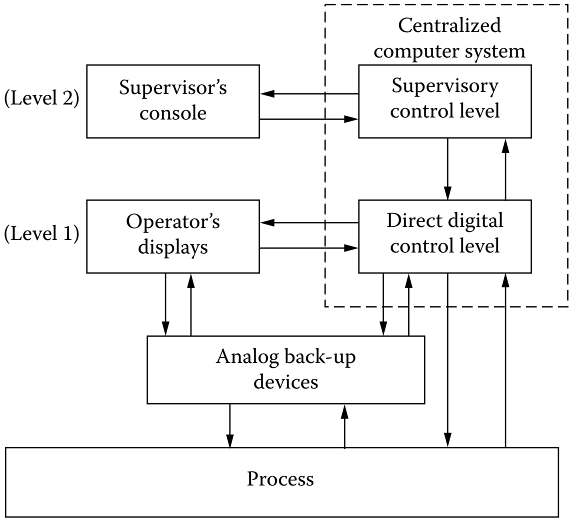


FIG. 2.11i
Complete secondary digital control: supervisory plus direct digital control with associated analog control backup.

Distributed Control

Because of the problems experienced with the centralized computer systems of the late 1960s, most of the new computer projects of the early 1970s were relatively small, specialized projects that took advantage of the capabilities of the newly arrived minicomputer. Many such projects flourished. They generally followed the lines of Figures 2.11d or 2.11i and differed from earlier undertakings mainly in their smaller size.

However, at this time, two other developments were under way that would forever change digital computer–based control and indeed all process control. The first of these was the rapid development of the integrated circuit and the resulting production of the all-important microprocessor or microcomputer. The second development was the courageous effort begun by the Honeywell Company in 1969 to design an alternative to the unwieldy and unreliable centralized computer control system. This alternative was the distributed computer system.

The idea behind the distributed computer system was to have a set of small, widely distributed computer "boxes" containing one or more microprocessors. Each of the boxes controlled one or a very few loops. All of them were connected by a single high-speed data link that permitted communication between each of the microprocessor-based "boxes" and with a centralized operator station or console.

This became the TDC 2000 system, the principles of which were widely followed by other process control system vendors. These systems solved reliability problems in two ways. First, these units controlled only a few process loops; thus any single failure involved only those few loops. Second, a digital backup capability was developed, that is, backup computer systems were included to take over the duties of any failed components.

Figure 2.11j illustrates this concept. Comparing this figure with Figure 2.11h shows what was accomplished with this new concept. Almost universally, the distributed computer systems offered the following features and capabilities, which greatly fostered their acceptance over electronic analog or centralized computer-based control systems:

1. A modular system development capability that is easy to use, particularly with the configuration aids available from the vendor.
2. A color-CRT–based, largely preprogrammed operator interface system that is easy to adapt to the individual plant situation.

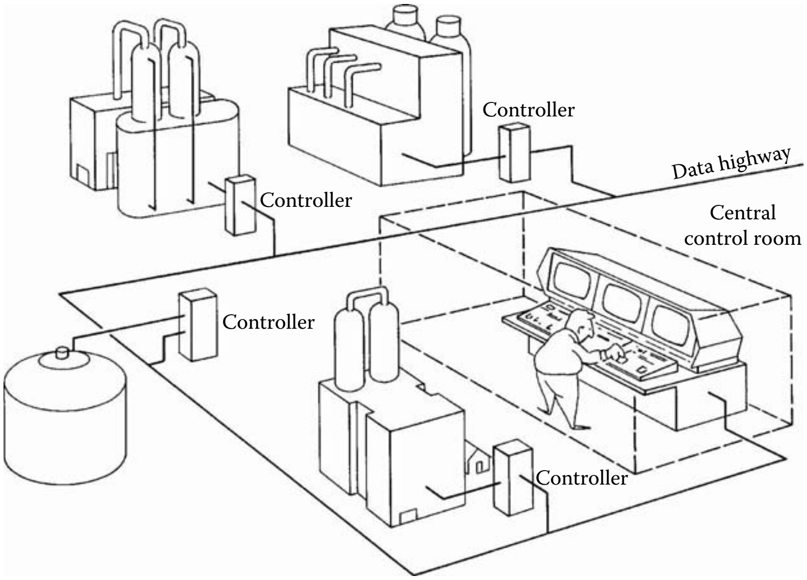


FIG. 2.11j
Microprocessor-based distributed direct digital computer control system (circa 1975).

3. A preprogrammed menu type of instruction system for the microcomputers of the controller box. This permits the final programming (configuration) of the total system to be done by pushing a few buttons on the keyboard.
4. A very wide selection of control algorithms or computational schemes within the preprogrammed menu, which permits easy selection and testing of alternate control schemes for a process.
5. Data highway based data transmission and communications capabilities between separate units of the system. Data highways provide very-wide-band communications and the possibility of redundancy for extra safety.
6. Relatively easy communications with mainframe computer systems for supervisory control or other higher-level process control or hierarchy control functions. However, these new control systems themselves are generally restricted to supplying the needs of the 1A and 1B level or dynamic control. Supervisory control is externally supplied.
7. Extensive diagnostic schemes and devices for easy and rapid maintenance through replacement of entire circuit boards.
8. Redundancy and other fail-safe techniques to help promote high system reliability. These features are often standard, but may be optional.

What was achieved with these systems is illustrated by Figures 2.11h and 2.11j, as noted above. These particular sketches are adapted from Honeywell drawings, but they are just as applicable to almost any other vendor's process control system.

HIERARCHICAL CONTROL

The development of the distributed digital control system greatly simplified the connection of the computer to the process. (Compare Figure 2.11k to Figure 2.11d.) Also, since redundancy or other backup devices could be incorporated directly into the digital system, the need for analog safety backup systems was minimized.

As illustrated in Figure 2.11k, there are three levels of control devices. Each has its distinct duties, as they form a hierarchical computer system in which upper-level computers depend on lower-level devices for process data, and the lower-level systems depend upon the higher ones for direction that serves more sophisticated control functions such as overall plant optimization. This configuration makes it possible to design a computer control system — one that combines the company's production scheduling and management information functions with the process control functions to form a total plant hierarchical control system, as illustrated in Figure 2.11l. This figure outlines most of the probable functions of such a hierarchical computer control system, but the magnitude of the tasks to be accomplished in the upper levels of the system is better indicated by the expanded version shown in Figure 2.11m. Such a system represents the ultimate dream of many process control system engineers. While what has already been achieved in the field of computer control fits easily within the framework shown in this diagram, this framework is only one of several possible structures for such a system and is not necessarily the optimal one for the task. Nevertheless, it is an excellent vehicle for our purpose here since it allows us to treat all the possible benefits of the computer system with one example.

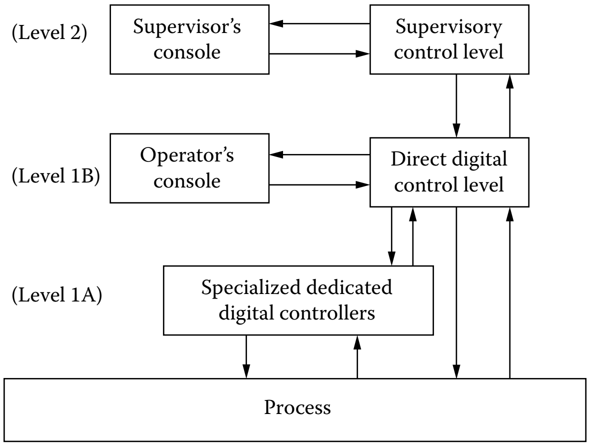


FIG. 2.11k
Complete secondary digital control: supervisory plus direct digital control.

It should be noted that the several levels shown in Figure 2.11m are operational levels and do not necessarily represent separate and distinct computer or hardware components. Although in large systems each level is likely to be handled by a separate computer, in small systems two or more operational levels might be combined into a single computer. The dedicated digital controllers at Level 1A require no human intervention since their functional tasks are completely fixed by systems design and are not altered online by the operator. All other levels have human interfaces as indicated in the figure. It should also be noted that each element of the hierarchy can exist as an individual element. Indeed, all of the earlier forms of industrial digital computer systems (Figures 2.11c, 2.11d, 2.11f, and 2.11i) still exist and will no doubt continue to be applied where their particular capabilities appear to best fit the application at hand.

The paragraphs below explain the basis for the organization of the six-level hierarchical system depicted in Figure 2.11m.

Overall Tasks of Digital Control Systems

Automatic control of any modern industrial plant, whether it is by a computer-based system or by conventional means, involves an extensive system for the automatic monitoring of a large number of variables, each having its own dynamics. The generation of control output signals requires the development of a large number of functions, some of which might be quite complex. These control corrections must be transmitted to a large number of widely scattered final control elements of various types. Because of the nature of the manufacturing processes involved, these control corrections often require the expenditure of very large amounts of material and energy.

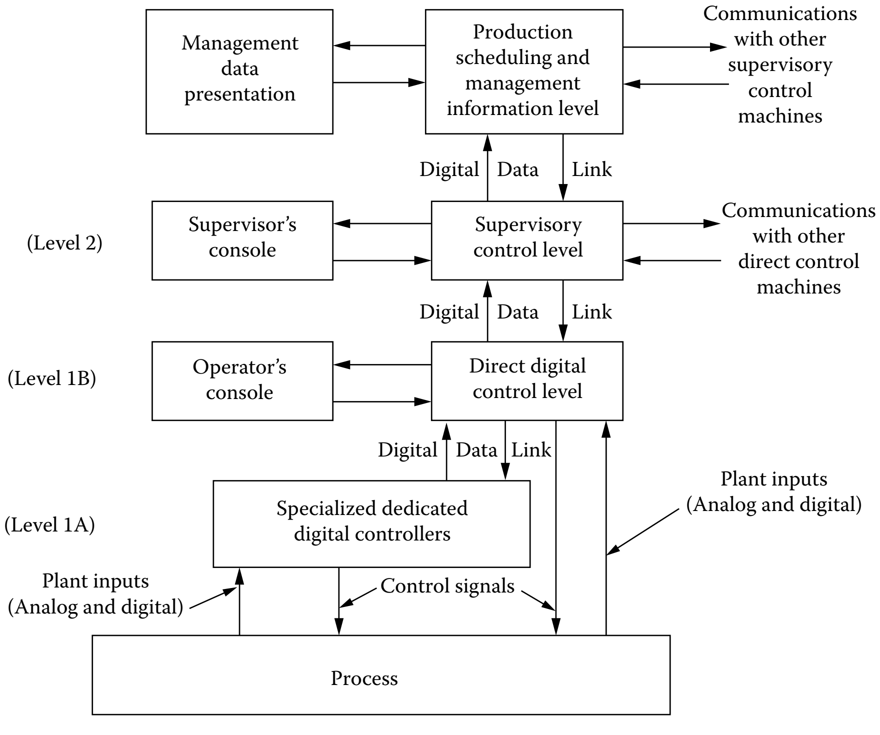


FIG. 2.11l
Hierarchical organization for a complete process computer control system.

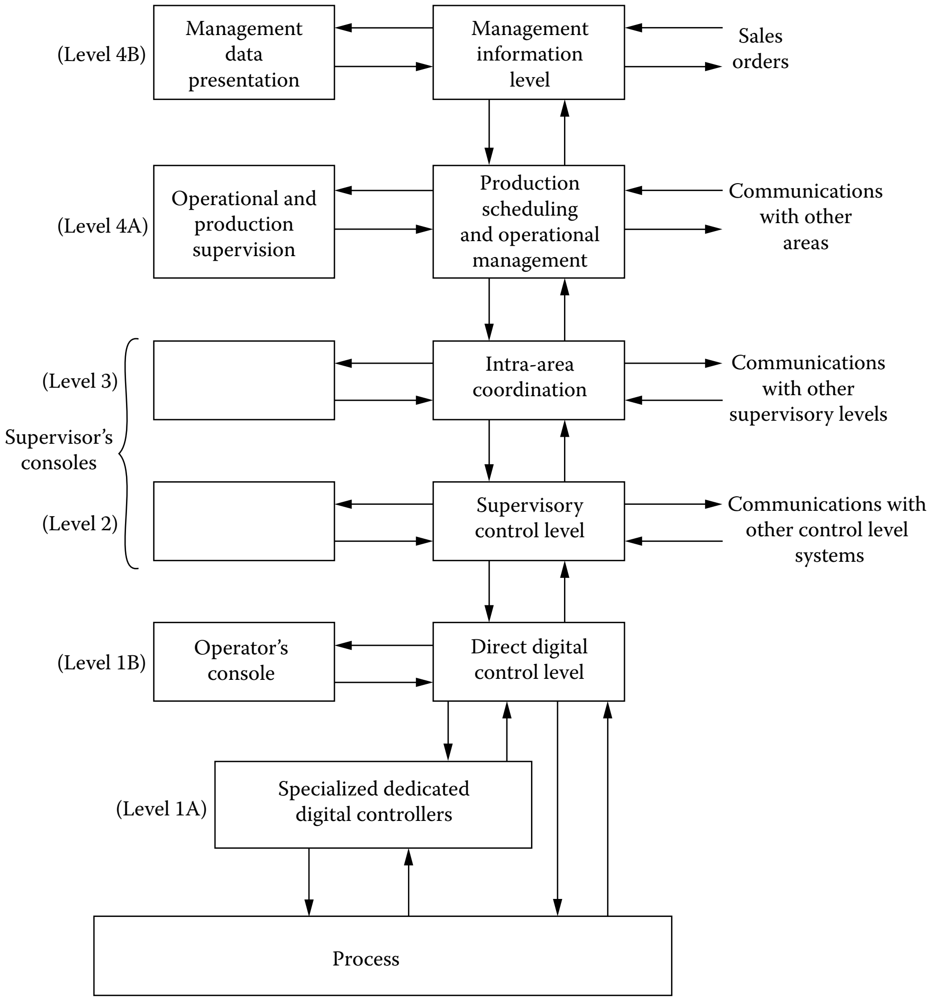


FIG. 2.11m
Hierarchical computer system structure for a large manufacturing complex.

Also, both operators and managers must be kept aware of the current status of the plant and of each of its processes.

In addition, an industrial plant is faced with the continual problem of adjusting its production schedule to match its customers' needs, as expressed by the order stream being continually received, while maintaining high plant productivity and the lowest practical production costs. Production scheduling today is usually handled through a manual, computer-aided production control system in conjunction with an in-process and finished goods inventory judged adequate by plant personnel.

Another role of digital computer control systems in industrial plants is as a "control systems enforcer." In this mode, the main task of the lower-level computers is continually to ensure that the control system equipment is actually carrying out the job that it was designed to do in keeping the units of the plant production system operating at some optimal level—that is, to be sure that the controllers have not been set on manual, that the optimal setpoints are being maintained, etc.

Often the tasks carried out by these control systems are ones that a skilled and attentive operator could readily do. But the automatic control systems offer a much greater degree of attentiveness over much longer periods of time.

All of these requirements must be factored into the design and operation of the control system that will operate the plant, including the requirements for maximum productivity and minimum energy use. As the overall requirements become more complex, more sophisticated and capable control systems are necessary. Thus we enter the realm of the digital computer-based control system.

System Capabilities To obtain the above described control responses, an overall system is needed that offers the following capabilities:

1. Tight control of each operating unit of the plant to ensure that it is using raw materials and energy at maximum efficiency and/or is operating at the most

efficient production level, based upon the production level set by the scheduling and supervisory functions of the control system. This control reacts directly to any emergencies that occur in its own unit. (These control functions are executed by Levels 1A and 1B in Figure 2.11m)
2. A supervisory and coordinating system that determines and sets the local production level of all units working together between inventory locations and optimizes their operation. This system ensures that no unit exceeds the general area level and thus ensures that no unit uses unnecessary amounts of energy or raw materials. This system responds to emergencies or upsets in any of the units under its control to shut down or systematically reduce the output in these and related units. (These control functions are executed by Levels 2 and 3 of Figure 2.11m.)
3. An overall production control system capable of carrying out the scheduling function for the plant based on customer orders or management decisions so as to produce the required products at the optimum combination of time, energy, and raw materials, suitably expressed as cost functions. (This is the specific task of Level 4A of the hierarchy.)
4. A method of assuring the overall reliability and availability of the total control system through fault detection, fault tolerance, redundancy, and other applicable techniques built into the system's specification and operation. (This task is performed at all levels of the control system.)

Because of their ever-widening scope of authority, control tasks 1, 2, and 3 above can effectively become the distinct and separate levels of a hierarchical control structure. Also, in view of the amount of information that must be passed back and forth among these three control tasks, it appears that some sort of distributed computational capability, organized in a hierarchical fashion, could be a logical structure for the required control system.

In Figure 2.11m, we can see that the right-hand elements of Levels 2 to 4A are all handled by computers of increasing capability as one goes up the hierarchy. Level 4B also contains a computer, but it is used mainly for communications and database management tasks. These elements are best handled by data processing or scientific computers since the noncontrol tasks at these levels far outnumber the process control tasks. These noncontrol computational tasks will thus determine the design and cost of these computers. As long as they are satisfactory for the purposes of process control, these established computer models should be used so that the economy gained from large-scale production can be capitalized on. This leaves Levels 1 and 2 as candidates for the application of digital process control hardware — the distributed, microprocessor-based systems discussed earlier.

TABLE 2.11n
Duties of Control Computer Systems

I. Production scheduling
II. Control enforcement
III. System coordination, reporting, and management information
IV. Reliability assurance

Detailed Task Listings

In the context of a large industrial plant, the tasks carried out at each level of the hierarchy are as described in Tables 2.11n to 2.11t. Note that in each table the tasks are subdivided into ones that are related to production scheduling, control enforcement coordination/reporting, and reliability assurance. This was described in items 1 through 4 previously. Note also that the duties listed in Table 2.11o for Levels 1 and 2 begin with Item II, Control Enforcement, since the lower-level machines do not do any production scheduling. Likewise, the upper-level machines do no control enforcement since they have no direct connection to the process actuators.

Finally, Level 4B does neither since its main task is management and staff function communications, with a production data file maintained by Level 4A, Production Scheduling. These tables outline the tasks that must be carried out in any industrial plant, particularly at the upper levels of the hierarchy. Details of how the operations are actually carried out will vary drastically, particularly at the lowest levels, because of the nature of the actual processes being controlled, but this does not change the basic definition of these tasks.

The general duties of the different levels in the hierarchical computer system are summarized in Figure 2.11t.

TABLE 2.11o
Duties of the Control Levels (Levels 1A and 1B)

II. Control enforcement
  1. Maintain direct control of the plant units under their cognizance.
  2. Detect and respond to any emergency condition in these plant units
III. System coordination and reporting
  3. Collect information on unit production raw material and energy use and transmit to higher levels.
  4. Service the operator's man/machine interface.
IV. Reliability assurance
  5. Perform diagnostics on themselves.
  6. Update and standby systems.

TABLE 2.11p
Duties of the Supervisory Level (Level 2)

II. Control enforcement
  1. Respond to any emergency condition in its region of plant cognizance.
  2. Locally optimize the operation of units under its control within limits of established production schedule; carry out all established process operational schemes or operating practices in connection with these processes.

III. Plant coordination and operational data reporting
  3. Collect and maintain data queues of production, inventory, and raw material and energy usage for the units under its control.
  4. Maintain communications with higher and lower levels.
  5. Service the man/machine interfaces for the units involved.

IV. System reliability assurance
  6. Perform diagnostics on itself and lower-level machines.
  7. Update all standby systems.

Lower-Level Computer Tasks

In the hierarchy shown in Figure 2.11m, all contact with the controlled process is maintained through the computers of Levels 1 and 2. The distributed, microprocessor-based systems, are all effectively stand-alone Level 1 and 2 systems.

TABLE 2.11q
Duties of the Area Level (Level 3)

I. Production scheduling
  1. Establish the immediate production schedule for its own area, including transportation needs.
  2. Locally optimize the costs for its individual production area as a basis for modifying the production schedule established by the production control computer system (Level 4A) (e.g., minimize energy usage or maximize production).

III. Plant coordination and operational data reporting
  3. Make area production reports.
  4. Use and maintain area practice files.
  5. Collect and maintain area data queues for production, inventory, raw materials usage, and energy usage.
  6. Maintain communications with higher and lower levels of the hierarchy.
  7. Collect operations data and off-line analysis as required by engineering functions.
  8. Service the man/machine interface for the area.
  9. Carry out needed personnel functions (such as vacation schedule, work force schedules, and union line of progression).

IV. System reliability assurance
  10. Diagnostics of self and lower-level functions.

TABLE 2.11r
Duties of the Production Scheduling and Operational Management Level (Level A)

I. Production scheduling
  1. Establish basic production schedule.
  2. Modify the production schedule for all units per order stream received, energy constraints, and power demand levels.
  3. Determine the optimum inventory level of goods in process at each storage point. The criteria to be used will be the trade-off between customer service (e.g., short delivery time) versus the capital cost of the inventory itself, as well as the trade-offs in operating costs versus costs of carrying the inventory level. (This is an offline function.)
  4. Modify production schedule as necessary whenever major production interruptions occur in downstream units, where such interruptions will affect prior or succeeding units.

III. Plant coordination and operational data reporting
  5. Collect and maintain raw material use and availability inventory and provide data for purchasing for raw material order entry.
  6. Collect and maintain overall energy use data for transfer to accounting.
  7. Collect and maintain overall goods in process and production inventory files.
  8. Collect and maintain the quality control file.
  9. Maintain interfaces with management interface level function and with area level systems.

IV. System reliability assurance
  10. Run self check and diagnostic routines on self and lower-level machines.

According to Table 2.11o, the tasks of these systems are to maintain direct control of the process; to detect and respond to timing signals, emergencies, and other events in the process; to collect process data for the plant operators or

TABLE 2.11s
Required Tasks of the Intracompany Communications Control System (Level 4B)

III. Plant coordination and operational data reporting
  1. Maintain interfaces with plant and company management, sales personnel, accounting and purchasing departments, and the production scheduling level (Level 4A).
  2. Supply production and status information as needed to plant and company management, sales personnel, and the accounting and purchasing departments. This information will be supplied in the form of regular production and status reports and in response to online inquiries.
  3. Supply order status information as needed to sales personnel.

IV. System reliability assurance
  4. Perform self check and diagnostic checks on itself.

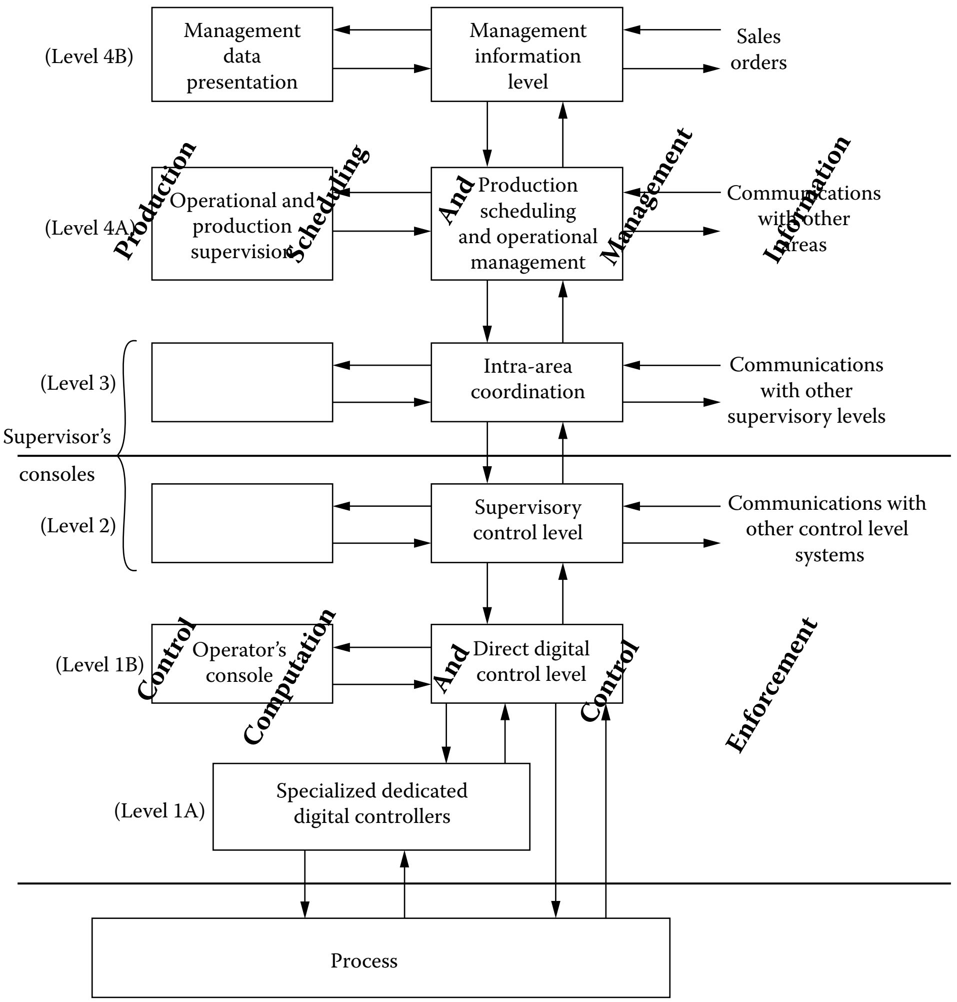


FIG. 2.11t
Summary of the tasks of the hierarchical computer.

for higher-level functions; and to ensure reliability by monitoring their own operation.

To do this, the system must monitor each important plant or process variable on a regular basis. That is, it must read the current value of the variable and, first, compare it with a set of alarm limits to detect the presence of any emergency situation and, second, compare it with the current operating set point to determine whether any correction to the current control output is necessary. Close check is also kept of the passage of time on the system real-time clock since most systems are time coordinated. That is, the monitoring program is reinitiated at fixed time intervals based upon the required sampling interval for the process variables.

As just indicated, plant variables are normally monitored either on a fixed time schedule, so that every variable is tested each second or fraction of a second (common in microprocessor-based distributed systems) or on a variable schedule depending upon the detected variable type. Table 2.11u lists the frequencies at which the various process variables are sensed. This second system has been popular for minicomputer-based digital control systems that commonly do not have the speed capability of the distributed systems. It is based on the dynamics or speed of response of the process being monitored and controlled.

TABLE 2.11u
Sampling Frequency for Process Variables

Variable	*Frequency*
1. Flow	Once each second
2. Level and pressure	Once each 5 seconds
3. Temperature	Once each 10 seconds
4. Queries from the operator's console	Once each second

Note: In some systems, process variables are sampled much more frequently than this. These are minimum rate values established from plant experience.

The timing of the process variable sampling is normally based upon a real-time clock included in the computer system.

If an emergency is detected, the computer's program contains routines for correcting the emergency or calling the operator's attention to it. The routines are selected according to a priority interrupt scheme. Control correction computations are carried out by means of a set of control algorithms or computational routines also stored in the computer's memory.

In addition to being used for emergency detection and control corrections, the values of the process variables are stored in the computer's memory in a process data base. These data are then available for readouts on the operators' consoles, for data logging, for historical records, for process efficiency calculations, and for readout to higher-level computers for optimization calculations, inventory monitoring, overall plant production status, historical file updates, and other necessary calculations. These calculations thus fulfill the systems coordination and reporting functions listed in Table 2.11o.

The computer uses any spare computational time to run test computations to check the validity of its own operation and of any companion computers in the system. These and other related tests fulfill the need for the reliability assurance functions listed in Table 2.11o.

Higher-Level Computer Tasks

As noted in the previous section, all contact with the controlled process is maintained through the input/output interfaces of Levels 1A and 1B. These interfaces may be part of a microprocessor-based distributed control system or of a minicomputer-based direct digital controller. Any other layout would require a reversion to the schemes of Figures 2.11d, 2.11g, or 2.11i, all of which have distinct drawbacks compared to Figures 2.11k or 2.11m.

The upper-level machines are connected to the lower-level systems and to each other through the communications system. Their major tasks, as outlined in Tables 2.11p to 2.11t, are to carry out the extensive computations involved in (1) the optimization of productivity and of raw material and energy use in each process; (2) the development of the best production schedule for the overall plant; and (3) the minimization of the plant's inventory.

An equally important task is the processing of the plant's production data as collected by the lower-level machines in order to supply the proper information to plant supervisory and management personnel and in order to maintain the plant's production database for the company's production, financial, and personnel reports.

Bibliography

Alexander, L., et al., *Hierarchical Non-Linear Switching Control Design,* Heidelberg: Springer-Verlag, 2000

Ash, R. H., "Manufacturing Systems Integrated by Information and Control," *Control Engineering,* Vol. 33, No. 5, May 1986, pp. 67–79.

Belick, F. M., "Computer Control," *Water & Wastes Engineering,* March 1975.

Bennett, C. E., "Evaluating Controls for Batch Processing," *Instrumentation Technology,* July 1977.

Chattaway, A. T., "Control System Integration," *InTech,* September 1983.

Dallimonti, R., "The Horizon Plant," *Control Engineering,* Vol. 33, No. 5, May 1986, pp. 72–76.

Fayssoux, J. O., "Planning and Design for Systemwide Control," Part 1, *Instrumentation Technology,* Vol. 39, No. 6, June 1992, pp. 20–22; Part 2, No. 7, July 1992, pp. 22–24.

Griem, P. D., "Security Functions in Distributed Control," *InTech,* March 1983.

"Hierarchical Control Problems," MCSD Seminar Series, Dept. of Operational Research, UNC, Boulder, CO, December 2001.

"Hierarchical Process Control of Chemical Vapor Infiltration," Technical Paper sponsored by Pentagon, May 1995.

Howard, J. R., "Experience in DDC Turbine Start-Up," *ISA Journal,* July 1966.

Kaiser, V. A., "Changing Patterns of Computer Control," *Instrumentation Technology,* February 1975.

Kovalcik, E. J., "Understanding Small Computers," *Instruments and Control Systems,* January 1976.

Larsen, G. R., "A Distributed Programmable Controller for Batch Control," *InTech,* March 1983.

Long, M. V., and Holzmann, E. G., "Approaching the Control Problem of the Automatic Chemical Plant," *Transactions of ASME,* October 1953.

Luyben, W. L., "Batch Reactor Control," *Instrumentation Technology,* April 1975.

Maurin, L. V., "Computing Control at Little Gypsy," *ISA Journal,* May 1966.

Merritt, R., "Personal Computers Move Into Control," *Instruments and Control Systems,* June 1983.

Miller, T. C., and Umbach, M., *Hierarchical Operations and Supply Chain Planning,* Heidelberg: Springer-Verlag, April 2002.

Miller, W., "Designing Reliability into Computer-Based Systems," *Instruments and Control Systems,* May 1982.

Miyazaki, S., "Development Project for DDC System," *CERR,* May 1970.

Morris, H. M., Kompass, E. J., and Williams, T. J. (Eds.), Advanced Control in Computer Integrated Manufacturing, Proceedings of the Advanced Control Conference, Purdue University, West Lafayette, Indiana, 1987.

Murphy, J. A., "Computer Control of an Aluminum Plant, " *Instrumentation Technology,* April 1972.

Prickett, M. J., "Microprocessors for Control and Instrumentation," *InTech,* March 1983.

Purdue Laboratory for Applied Industrial Control, *Tasks and Functional Specifications of the Steel Plant Hierarchy Control System* (Expanded Version), Report No. 98, Purdue University, June 1982.

Rispoli, L. M., "Hierarchical Computer Control Systems," *Instruments and Control Systems,* October 1970.

Roberts, D. B., "Optimal Control of Batch Processes," Ph.D. Research Project, School of Engineering, City University Northampton Square, London, 2000.

Sayles, J. H., "Computer Control Maximizes Hydrocracker Throughput," *Instrumentation Technology,* May 1973.

Stamphil, L. M., "Organizing In-House Systems Integrations Projects," *Instruments and Control Systems,* Vol. 65, No. 1, January 1992, pp. 45–48.

Stout, T. M., "Justifying Process Control Computers," *Chemical Engineering,* September 11, 1972.

Van Horn, L. D., "Crude Distillation Computer Control Experience," 1976 ISA Conference, Preprint No. 76-533.

Williams, D. L., and McHugh, A., "TDC-1000–An Overview," Minutes of the Eighth Annual Meeting, International Purdue Workshop on Industrial Computer Systems, Purdue University, September 22–25, 1980.

Williams, T. J., "Hierarchical Distributed Control," *InTech,* March 1983.

Williams, T. J., *The Purdue Enterprise Reference Architecture, A Technical Guide for CIM Planning and Implementation,* Research Triangle Park, NC: Instrument Society of America, 1992.

Williams, T. J. (Ed.), *A Reference Model for Computer Integrated Manufacturing, ICM, A Description from the Viewpoint of Industrial Automation,* Research Triangle Park, NC: Instrument Society of America, 1989.

2.12 Interaction and Decoupling

S. OCHIAI (1995, 2005)

INTRODUCTION

In this section, the interaction between loops, the measurement of interaction, and the design of the algorithms used are illustrated through a case study of level control on a chemical reactor and flasher system. The achievement of noninteracting control, or decoupling, is one of the goals of multivariable control. The decoupling algorithm is simple, and the physical meaning of the individual decoupling elements is clear to the designer and the plant operator.

During the last decade, advanced process control methods have been extensively used in processing industries. However, these methods have usually been applied to large processing systems because the cost of software and process model development is rather high.

In applications where the process is small or is only a small part of a larger process, the use of advanced process control may not be economically justified. When this is the case, process control engineers can develop and implement multivariable control on their own. Decoupling control is a tool that can be used to reach this goal.

INTERACTING PROCESS EXAMPLE

Figure 2.12a illustrates the conventional level controls of a process consisting of a chemical reactor and a flasher. Such systems are typical subprocesses in many petrochemical plants.

As shown in Figure 2.12a, feeds 1 and 2 enter the reactor, and the product from the reactor is sent to the flasher. The pressure in the flasher is much lower than that in the reactor, and therefore the low boilers contained in the product are vaporized and are sent from the flasher to a downstream distillation column for recovery.

The liquid from the flasher is returned to the reactor under level control (m_2). The product flow from the reactor to the flasher is manipulated to control the level in the reactor. Therefore, when the reactor level rises, the level controller increases the product flow, which in turn causes the flasher level to rise. Inversely, when the flasher level rises, the level controller on the flasher increases the recycle flow, which in turn raises the reactor level.

The left side of Figure 2.12b illustrates the oscillation, which is caused by the interaction between the two level control loops. On the right side of the figure, the response obtained by "partial decoupling" is shown. This technique will be described later, after developing some general equations and comments about noninteracting control.

Decoupling the Process

Figure 2.12c illustrates a noninteracting or decoupled version of the previously described process, in which the interaction between the level loops has been eliminated.

The reactor product flow is set as the sum of signals m_1 and m_2, which are the output signals of the level controllers on the reactor and the flasher. The recycle flow is set as $am_1 + m_2$, where "a" is the fraction of the product flow that stays in the flasher base as a liquid.

When the reactor level increases, causing the direct acting LC-1 controller output to also rise by Δm_1, the liquid flow to the flasher base will increase by $a\Delta m_1$. However, the recycle flow also increases by $a\Delta m_1$ if the adder #2 is correctly adjusted.

Therefore, the flasher level remains unchanged and the interaction has been "decoupled." A similar sequence of events occurs when the flasher controller output changes by Δm_2.

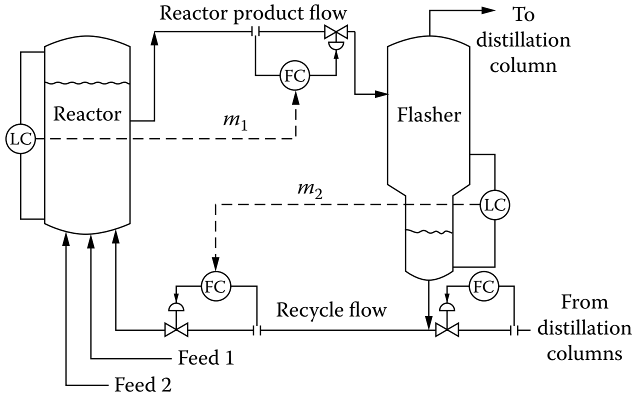


FIG. 2.12a
Level controls for reactor and flasher process without decoupling the interaction of the loops.

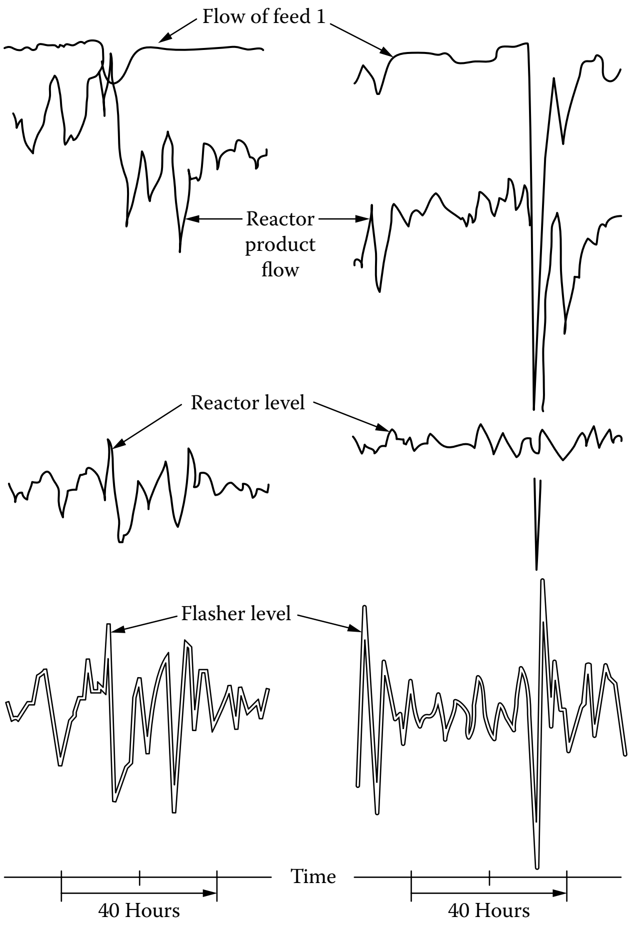


FIG. 2.12b
Response of the reactor–flasher system to upsets in feed flow under conventional (left) and partially decoupled controls.

In this example, the process dynamics were very simple, and a noninteracting control system could be configured by direct physical interpretation of the process. In the following discussion a generalized approach will be described.

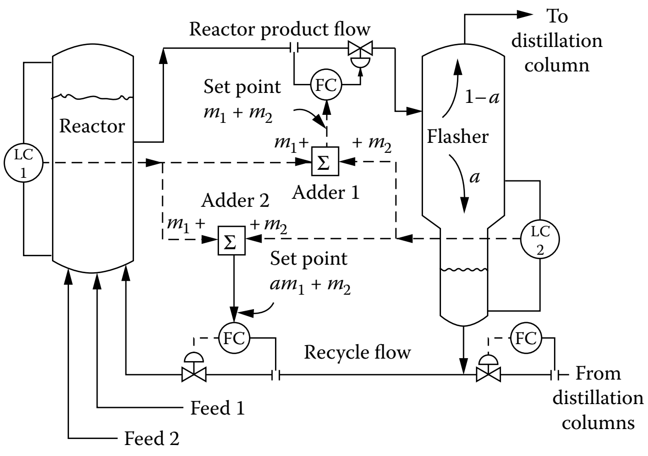


FIG. 2.12c
Noninteracting (decoupled) level control system for the reactor–flasher process.

Generalizing the Solution

Figure 2.12d represents a control system[1] having two controlled variables (such as levels of the reactor and the flasher in the previous example), $c_1(s)$ and $c_2(s)$, where s is a complex variable.

The transfer functions[1–3] of the processes are represented by $G(s)$ symbols, and the transfer functions[2,3] of the two controllers are denoted by $F_{11}(s)$ and $F_{22}(s)$. The controller outputs are respectively $m_1(s)$ and $m_2(s)$. The set points of the controllers are designated by $r_1(s)$ and $r_2(s)$.

In a noninteracting control system, the decoupler elements $F_{21}(s)$ and $F_{12}(s)$ are added so that they will compensate for and thereby eliminate the process interactions designated as $G_{21}(s)$ and $G_{12}(s)$.

In order to build a decoupled control system, one would first obtain F_{21}. The effect of m_1 to c_2 is transmitted from point "P" to "Q" via two (heavily drawn) paths, one through G_{21}, the other through F_{21} and G_{22}, and the sum of the two should be zero:

$$G_{21}m_1 + F_{21}G_{22}m_1 = 0 \qquad \textbf{2.12(1)}$$

Therefore,

$$F_{21} = -G_{21}/G_{22} \qquad \textbf{2.12(2)}$$

Similarly, considering the influence of m_2 on c_1,

$$F_{12} = -G_{12}/G_{11} \qquad \textbf{2.12(3)}$$

For a general case with n controlled variables and n manipulated variables, see Reference 4.

For the reactor and flasher system:

$$\begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} -1/(A_1 s) & 1/(A_1 s) \\ a/(A_2 s) & -1/(A_2 s) \end{bmatrix} \qquad \textbf{2.12(4)}$$

where A_1 and A_2 are, respectively, the cross-sectional areas of the reactor and the flasher. Consequently, the decoupler settings are

$$F_{21}(s) = -\frac{G_{21}}{G_{22}} = -\left(\frac{a}{A_2 s}\right)(-A_2 s) = a \qquad \textbf{2.12(5)}$$

$$F_{12}(s) = -\frac{G_{12}}{G_{11}} = -\left(\frac{1}{A_1 s}\right)(-A_1 s) = 1 \qquad \textbf{2.12(6)}$$

Note that the values of A_1 and A_2 in Equation 2.12(4) do not appear in Equations 2.12(5) and 2.12(6) because they

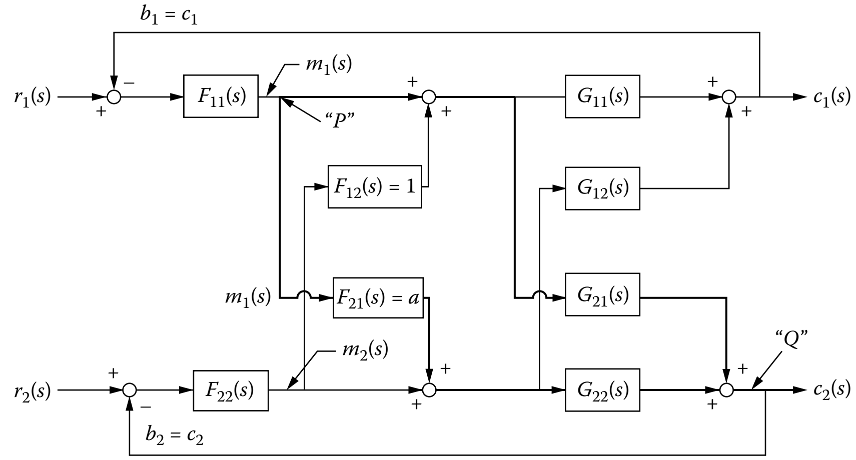


FIG. 2.12d
Generalized block diagram representation of a two-variable noninteracting control system.

cancel out. In general, it is easier to estimate the ratios than to estimate the individual transfer functions. This is particularly so when only crude estimates are available.

Measuring the Interactions

Before an effort is made to design a noninteracting control system, it is desirable to have an idea of the amount of interaction that is present. A relative gain calculation, described in Section 2.25, serves this purpose.

Relative gain is defined as the open-loop gain when all the other loops are in manual, divided by the open-loop gain when all the other loops are in automatic.

$$\lambda_{11} = \frac{\left.\dfrac{\partial c_1}{\partial m_1}\right|_{m_2}}{\left.\dfrac{\partial c_1}{\partial m_1}\right|_{c_2}} \qquad \mathbf{2.12(7)}$$

It can be seen from the block diagram in Figure 2.12d that if both $F_{12}(s)$ and $F_{21}(s)$ were nonexistent (i.e., if the single-variable control system of Figure 2.12a were implemented), the numerator in the relative gain calculation would be $G_{11}(s)$ and the denominator would be

$$(G_{11}G_{22} - G_{12}G_{21})/G_{22} \qquad \mathbf{2.12(8)}$$

because the control loop for c_2 would force m_2 to take on the value which would cause

$$G_{21}m_1 + G_{22}m_2 = 0 \qquad \mathbf{2.12(9)}$$

If the value of a is 0.5, then from Equations 2.12(5) to 2.12(8):

$$\lambda_{11} = 1/(1 - a) = 2.0 \qquad \mathbf{2.12(10)}$$

This means that when the recycle flow m_2 is manually fixed at a constant value (which is the numerator in Equation 2.12[7]), the effect of manipulating the reactor product flow m_1 to control the reactor level is twice as much as it is in the case when m_2 is automatically manipulated (the denominator in Equation 2.12[7]).

When the decoupler of Equation 2.12(5) is inserted into the loop (Figure 2.12d), c_2 does not vary when m_1 is changed. Therefore, output m_2 of controller F_{22} does not vary either, i.e., the denominator in Equation 2.12(7) is equal to the numerator and therefore the relative gain is 1.0. (As is shown in Section 2.25, $\lambda_{22} = \lambda_{11}$.)

Therefore, the effect of manipulating the recycle flow m_2 can be treated similarly. Interaction between the two loops is significant.

MERITS, DRAWBACKS, AND COMPROMISES

Decoupling algorithms such as Equations 2.12(5) and 2.12(6) can be simple, and therefore their physical meanings can be easily understood by the designer and by the plant operators. After noninteracting control is implemented, the control loops can be easily tuned based on single-variable control theory (Section 2.35).

Drawbacks

When noninteracting control is implemented, a set point change applied to one controlled variable will have no effect on the other controlled variable(s).

In some other industrial process control systems interaction between control loops can be exploited to achieve overall control. In other words, a noninteracting control system can be slower in recovery from an upset than an interacting one.

A more obvious drawback of decoupling can occur in the case when one of the G_{ij}'s in Figure 2.12d has a positive zero, that is, for example,

$$G_{11} = (s - b)/(s + c) \qquad \textbf{2.12(11)}$$

where b and c are positive constants. Such a process characteristic is called inverse response.[1] In this case decoupler $F_{12}(s)$ in Equation 2.12(3) will have a term $(s - b)$ in the denominator, and the control system will provide poor performance and could become unstable.

Partial and Static Decoupling

Because of these potential drawbacks, it is recommended to give consideration to partial decoupling, i.e., to implement only some of the available decouplers. Likewise, it may be an advantage to use a smaller gain than complete decoupling would allow. Partial decoupling can be more robust and less sensitive to modeling errors,[5] and in some cases it can achieve better control than complete decoupling does.

Another method of obtaining noninteracting control is through static decoupling.[6] Here the objective is to eliminate only the steady-state interactions between loops. The static decouplers are obtained by $\lim_{s\to 0} F_{21}(s)$ and $\lim_{s\to 0} F_{12}(s)$ in Equations 2.12(2) and 2.12(3). In case of the reactor and flasher systems, the decouplers used are static decouplers.

The Reactor–Flasher Example

The right half of Figure 2.12b was obtained by setting the decoupler F_{12} to 0.5 instead of using the 1.0 setting, which Equation 2.12(6) would have called for. The gain in the decoupler was halved to reduce the flow variation in the flasher overhead to the subsequent distillation column.

In addition, decoupler $F_{21}(s)$ was not implemented at all. It was left out because the response of the flasher level is much faster (the cross-sectional area of the tank is smaller) than that of the reactor, and therefore variation in the flasher level, within reasonable limits, is not detrimental to the operation, while changes in the reactor level are.

As can be seen in Figure 2.12b, even with only partial decoupling, the variation in the reactor level is much reduced in comparison to the case without decoupling. It should be noted that this improvement occurred even though a drastic change in reactor feed #1 did substantially upset the flasher level.

References

1. Marlin, T. E., *Process Control, Designing Process and Control Systems for Dynamic Performance*, 2nd ed., New York: McGraw-Hill, 2000, p. 630.
2. Ibid, p.111.
3. Chan, P. C., *Process Control, A First Course with Matlab,* New York: Cambridge University Press, 2002, pp. 21–24.
4. Ogunnaike, B. A., and Ray, W. H., *Process Dynamics, Modeling and Control,* New York: Oxford University Press, 1994, pp. 781–785.
5. Shinskey, F. G., *Process Control Systems, Application, Design and Tuning,* 4th ed., New York: McGraw-Hill, 1996, pp. 241–274.
6. Reference 4, pp. 790–798.

Bibliography

Marlin, T. E., *Process Control, Designing Process and Control Systems for Dynamic Performance,* 2nd ed., New York: McGraw-Hill Higher Education, 2000.

Ogunnaike, B. A., and Ray, W. H., *Process Dynamics, Modeling and Control,* New York: Oxford University Press, 1994.

Shinskey, F. G., *Process Control Systems, Application, Design and Tuning,* 4th ed., New York: McGraw-Hill, 1996.

2.13 Model-Based Control

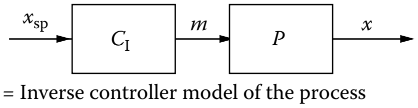
R. R. RHINEHART (1995, 2005)

In model-based control (MBC), computers use a process model to make control decisions. By comparison to PID-type approaches, the MBC action is "intelligent" and has shown benefits in uniformity, disturbance rejection, and setpoint tracking, all of which translate into better process safety and economics. On the other hand, MBC requires higher skill by the operator and engineer, and computational power. On the other hand, higher level of understanding of process models means better process understanding, better personnel training, diagnosis ability, and process management. There are many successful commercial controllers, and some distributed control system (DCS) suppliers offer MBC software packages with their equipment.

THE STRUCTURE OF MBC

The basic concept of MBC is illustrated in Figure 2.13a, where the "controller," C_I, uses a model of the process to calculate a value for the manipulated variable that should make the controlled variable (x) behave in a desired way. If it does, then that calculated value of m is then implemented. The controller does not have PID components and, in general, there is only one tuning parameter per controlled variable—the choice of how fast the controlled variable (x) should move to the setpoint (x_{sp}). The symbol C_I represents the model inverse. The "inverse" nomenclature arises from how the model is used. In a normal modeling approach, one specifies the process input, and the model predicts the process output response. By contrast, MBC determines the process input that causes a desired process output response; this is the model inverse.

If the model is perfect and no constraints exist, the open loop structure of Figure 2.13a will work. However, nearly all controller models are not perfect; therefore, they require some form of feedback correction. The conventional feedback approach is illustrated in Figure 2.13b, where the difference (d) between the model (C_M) and the process output (x) is monitored and used to adjust a controller feedback (c), which usually is either a bias to the setpoint (x_{sp}) or a model coefficient. The controller consists of the three functions C_I, C_M, and C_A, enclosed by the dashed line.



C_I = Inverse controller model of the process
P = Process

FIG. 2.13a
Model-based control (MBC) concept.

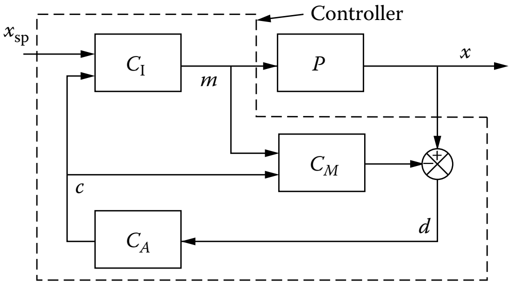


FIG. 2.13b
The three MBC functions: inverse, model, and adjustment.

Although the common aim of this strategy is to make intelligent control decisions, the choices of model structure, control objective, adjusting mechanism, and adjustable parameters have led to many variations in model-based controllers. The choice of model structure affects implementation, limits choices of other controller functions, and sets the accuracy of control decisions. Model structure, therefore, is of primary importance. Choices for control objective, adjustment mechanisms, and adjustable parameters, and methods for other features such as constraint handling, optimization, and data reconciliation are influential but secondary. Accordingly, MBCs are generally classified by model type.

MODELING APPROACHES

The three dominant modeling approaches of commercially offered MBC products—linear transfer function, linear time series, and nonlinear time series—are all empirical. However, phenomenological or fundamental, mechanistic models are used on a custom basis. Transfer function models are mostly based on open-loop Laplace transform descriptions of the process response to a step input and have been the traditional control modeling approach. Their familiarity and the simplicity of the resulting MBC are advantages which offset their limitations of linear and simplistic dynamic modeling. In addition, a controller based on a Laguerre polynomial model has demonstrated sustained commercial success.

Time series models represent the open-loop response of the process with a vector of impulses that are empirically determined and consist of 30 to 180, or so, elements. The precision of the performance of the modeled dynamic process is an advantage that offsets the limiting assumption that the process is linear, as well as the need for using matrix/vector algebra. This is the most common modeling approach in the industrial use of MBC today.

Some vendors offer nonlinear MPC versions that extend the valid operating range otherwise limited by the linear models. These include transitioning between multiple linear models when transitioning between operating regions, neural network–based models, or gain scaling of the linear model by any number of methods (fuzzy logic, expert rules, neural networks, process properties, etc.).

Nonlinear phenomenological models are design-type simulators. For markedly nonlinear or nonstationary process applications, their control intelligence can offset the disadvantage of their modeling and computational complexity.

In all cases, control is only as good as the model is a true representation of the process. Initializing the controller with a model that has been validated by process testing is the first and most critical implementation step.

INTERNAL MODEL CONTROL (IMC)

IMC uses open-loop step-response Laplace transfer function models. The model, C_M, may be the simple first-order-plus-deadtime (FOPDT) representation of a single-input–single-output (SISO) process:

$$\frac{\hat{x}}{\hat{m}} = C_M(s) = \frac{K_p}{\tau_p s + 1} e^{-\Theta_p s} \qquad \textbf{2.13(1)}$$

Ideally, the inverse is the reciprocal of the model, and if one wants the process to track the setpoint ($x = x_{\text{sp}}$) then

$$\hat{m} = C_I(s)\hat{x}_{\text{sp}} = C_m^{-1}(s)\hat{x}_{\text{sp}} = \left[\left(\frac{\tau_p}{K_p}s + \frac{1}{K_p}\right)e^{+\Theta_p s}\right]\hat{x}_{\text{sp}} \qquad \textbf{2.13(2)}$$

Equation 2.13(2) specifies how to calculate m, given x_{sp}, but its use presents some problems. First, the exp($+\Theta_p s$) term specifies that m must change before a change in x_{sp} has occurred. Since the future cannot be predicted, that action term is unrealizable. Second, the $(\tau_p/K_p)s$ term specifies taking the derivative of x_{sp}, and therefore, it responds to step changes in x_{sp} with the generation of an infinite spike in m. Third, if either the value of K_p is not exactly correct or if calibration errors have occurred, Equation 2.13(2) will result in a steady-state offset.

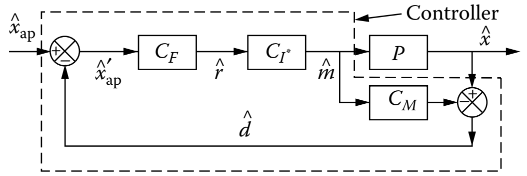


FIG. 2.13c
The basic IMC structure.

These problems are remedied in the IMC structure shown in Figure 2.13c. Here C_{I^*} is the realizable and stable part of the model inverse. For the SISO FOPDT example:

$$C_{I^*}(s) = \left(\frac{\tau_p}{K_p}s + \frac{1}{K_p}\right) \qquad \textbf{2.13(3)}$$

In Figure 2.13c, the variable d represents the model-to-process difference, a modeling error, regardless of error cause. If the model is in error, then steady-state offset can be removed by biasing the setpoint by that error, as shown. Setpoint biasing is the IMC feedback adjustment method. To eliminate spikes in m due to both measurement noise and setpoint changes, the biased setpoint is filtered (C_F) to obtain a reference trajectory, r. The inverse (C_I) calculates m, which would make the idealized process output (x) track r.

After some block diagram rearrangement, Figure 2.13d presents the SISO IMC function blocks for a process that can be modeled as FOPDT. Note that the functions are leads, lags, delays, summations, and gains, and they can be configured in almost any distributed control system (DCS), single-loop controller (SLC), or programmable logic controller (PLC). The single tuning parameter is τ_f, the filter time-constant. Lower values of τ_f make the controller more aggressive; higher values make it gentler.

One technical drawback to IMC is that the model is linear and stationary. If either the process gain or the time-constants change, the model-based calculations become either too aggressive or sluggish, and therefore, retuning is required.

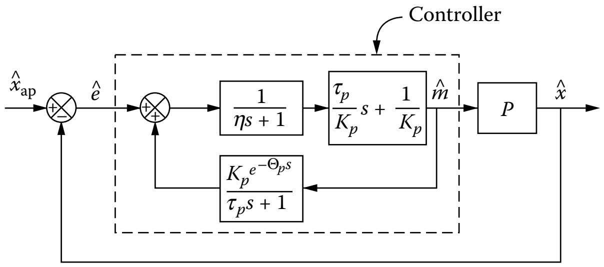


FIG. 2.13d
IMC for a FOPDT process. The C_F, C_{I^*}, and C_M blocks have been rearranged to show implementation.

This is not different from PID control, in which changes in process gain or dynamics also require retuning. If the process gain or dynamics change substantially, the model parameter values should be renewed. The other technical disadvantage of IMC is that for multi-input–multi-output (MIMO) processes the coupling causes the number of models to increase as the square of the number of manipulated variables, which makes finding the realizable inverse somewhat complicated, and there is no useful approach to solving for the MIMO control action in the case of constraints of an excess number of manipulated variables.

An advantage is that SISO IMC can be implemented with the skill of most control engineers using standard DCS or PLC functions. As a note, lambda tuning of PID controllers is based on the IMC approach.

MODEL PREDICTIVE CONTROL (MPC)

The MPC models, often called time series or convolution models, represent the open-loop process response as a vector. Figure 2.13e illustrates this. At i time increments after the single, unity input step change in the manipulated variable (m), the output (x) has changed by an amount h_i. A list of the values of the h_i elements in the vector $\underline{h}$ is the process model.

Any input behavior can be represented as a sequence of steps, and the output of a linear process as the cumulative effects of those steps. Figure 2.13f illustrates this for an input modeled as three steps after an extended period of steady conditions. Mathematically, this SISO time series process model is

$$\begin{aligned} x_1 &= x_0 + h_1 \cdot \Delta m_1 \\ x_2 &= x_0 + h_2 \cdot \Delta m_1 + h_1 \cdot \Delta m_2 \\ x_3 &= x_0 + h_3 \cdot \Delta m_1 + h_2 \cdot \Delta m_2 + h_1 \cdot \Delta m_3 \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \end{aligned} \qquad \textbf{2.13(4)}$$

where x_i is the ith value of x due to all preceding Δm's.

In compact vector/matrix notation,

$$\underline{\Delta x} = \underline{\underline{H}} \cdot \underline{\Delta m} \qquad \textbf{2.13(5)}$$

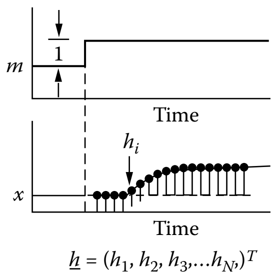


FIG. 2.13e
Dynamic vector, $\underline{h}$, describes an ideal unit step response.

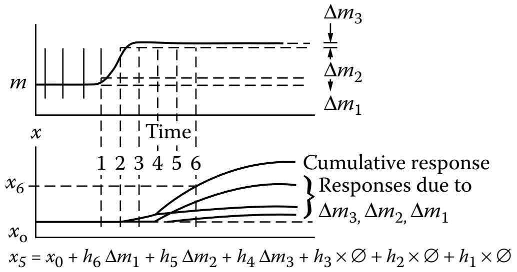


FIG. 2.13f
Illustration of time series modeling.

where $\underline{\Delta x} = \underline{x} - x_0 \cdot \underline{1}$ and $\underline{\underline{H}}$ is the lower diagonal matrix composed of staggered $\underline{h}$'s. $\underline{\underline{H}}$ is termed the dynamic matrix. $\underline{\Delta m}$ is the vector of past input steps.

If a desired behavior for x is specified (for instance, a path of return to setpoint), it can be expressed as a vector of Δr's and the necessary sequence of Δm's can be idealistically calculated from the SISO model inverse.

$$\underline{\Delta m}_f = \underline{\underline{H}}^{-1} \cdot \underline{\Delta r}_f \qquad \textbf{2.13(6)}$$

Here $\underline{\Delta m}_f$ is the required future sequence of m's. There are several problems, however, with this primitive concept. If the process has deadtime, leading elements of $\underline{\underline{H}}$ are zero and Equation 2.13(6) will result in a division by zero. Another problem is that Equation 2.13(6) ignores the residual influence of past Δm's that will affect x. If $\underline{h}$ is not a true model, it will result in steady-state offset. The methods for handling such issues vary but are conceptually expressed in Figure 2.13b. The function C_M is the model of Equation 2.13(5). The function C_A may simply be a filter to remove measurement noise from the model mismatch signal, d. For a SISO controller the function C_I would be

$$\underline{\Delta m}_f = \underline{\underline{H}}^{*-1} \cdot (\underline{\Delta r}_f + d \cdot \underline{1} - \underline{\underline{H}} \cdot \underline{\Delta m}_p) \qquad \textbf{2.13(7)}$$

where the subscripts f and p mean future and past, and d is the filtered value of the model mismatch, which is assumed to be constant into the future and is used to bias the reference trajectory. $\underline{\underline{H}}^*$ is the realizable model inverse, and the division-by-zero problem has been eliminated.

Exact tracking of $\underline{\Delta r}$ by Equation 2.13(7) may define extreme manipulated variable (MV) moves for high-order or inverse-acting processes. Accordingly, in practice, the MPC law, Equation 2.13(7), is usually restructured so as to determine the $\underline{\Delta m}_f$ values that minimize the sum of squared deviations of the model prediction from the desired path plus a weighted sum of squared changes in m, a least-squares optimization. There exist a variety of single-line equations that specify the linear algebra operations that would be used to calculate values for MVs for selected MIMO processes. However, there is one additional complication to be handled.

The preceding discussion was for a SISO process. The development is similar for MIMO processes. If the process is a "square system," in which the number of unconstrained manipulated variables (MV) equals the number of unconstrained controlled variables (CV), then the vector terms in Equations 2.13(5 to 7) become matrices of vectors, and the matrices in those equations become matrices of matrices. However, MIMO systems are rarely square. Manufacturing optimization usually places the operating point on a process constraint — either a maximum MV value or an uncontrolled process variable (PV) related to safety (pressure) or to proper operations (flooding). If not, there are often excess numbers of MVs. In either case, the square concept that leads to the closed form analytical expression of Equation 2.13(7) will not work. For nonsquare systems, we use numerical optimizers to minimize the sum of squared deviations from the setpoint, with a penalty for excessive manipulated variable changes, penalty for utilities cost, and subject to a variety of process constraints. An example is

$$\begin{aligned} &\frac{\min J}{\{m_{k,l}\}} = \sum_{i}^{n\mathrm{CV}} \left[\lambda_i \sum_{j}^{t} (e_{i,j})^2 \right] + \sum_{k}^{n\mathrm{MV}} \left[f_k \sum_{l}^{t} (\Delta m_{i,j})^2 \right] \\ &\mathrm{S.T} \left| \Delta m_{k,l} \right| < \Delta m_{\max k} \qquad \textbf{2.13(8)} \\ &\quad \mathrm{PV}_{\min i} \le \mathrm{PV}_{i,j} \le \mathrm{PV}_{\max i} \end{aligned}$$

where *n*CV and *n*MV represent the number of CVs and MVs, and subscripts *j* and *l* represent discrete future samplings.

MPC is used in standard industrial practice and is marketed by many control consulting and DCS vendor companies. Optimizing algorithms include linear programming and successive quadratic programming, and there are a variety of refinements to the objective function and constraints. Several major process companies have used teams of engineers to develop in-house controllers, but most users prefer to buy a tested externally developed package.

The common limitation of MPC is that $\underline{h}$ is a linear and stationary model; and therefore, substantial changes in operating conditions may require the development of a new model. To extend the range of linear MPC algorithms, some vendors are offering a range of improvements, which include transitioning between models, gain scheduling, or nonlinear time series models.

PROCESS-MODEL-BASED CONTROL (PMBC)

Process models can be developed from a mechanistic (or first-principles, or phenomenological) point of view, based on material and energy balances and thermodynamic relations. Design-type simulators and supervisory optimizers use the steady state form of these sorts of models. However, for control, full design rigor is not required. Simple, reduced, or tendency models are usually sufficient for control, but the model needs to include dynamic response. A dynamic model for a SISO process may be represented as

$$\frac{dx}{dt} = f(x, m, p) \qquad \textbf{2.13(9)}$$

where p is an adjustable process parameter.

A simple control objective would be to have the process start moving from x to x_{sp} at a rate such that it would reach x_{sp} within some time τ. This specifies the local derivative, and the control law becomes the model inverse which determines m from the desired control action. For the SISO process,

$$f(x, m, p) - \frac{x_{\mathrm{sp}} - x}{\tau} = 0 \qquad \textbf{2.13(10)}$$

Since x is measured, p is set, and x_{sp} and τ are user-defined, Equation 2.13(10) can be solved for m. τ is the single tuning parameter. Such equations are often nonlinear, and the solution may require iterative techniques such as Newton's or bisection. However, for nonlinear processes there is an advantage over linear model controllers. The PMBC model accounts for changes in process gain and dynamics; and, once tuned, the controller does not need to be returned when the process conditions change.

Expectedly, Equation 2.13(10) will result in steady-state offset if the model is not perfect. A variety of feedback adjustment approaches have been demonstrated; some bias the setpoint, some adjust a process parameter. SISO Generic Model Control (GMC) can be implemented relatively easily by the control engineer. It biases the setpoint with an integral of the past errors to yield the control law:

$$f(x, m, p) - \frac{x_{\mathrm{sp}} - x}{\tau_1} - \tau_2 \int (x_{\mathrm{sp}} - x)\,dt = 0 \qquad \textbf{2.13(11)}$$

This functional form is comfortable for many. It is like a PI controller with a nonlinear function on the controller output. And, comfortably, it is easily tuned by adjusting τ_1 like a proportional band to determine the aggressiveness of the initial response, then adjusting τ_2 to remove the offset at a comfortable rate.

A variety of other nonlinear model-based controllers result from a variety of choices about the feedback procedure. For instance, external reset feedback can be used to eliminate integral windup. Model parameter values can be adjusted to track process changes. Criteria for selecting the adjustable parameters are:

1. Choose a parameter that represents a real process mechanism (not a hypothetical disturbance).

2. Choose a mechanism that changes.
3. Choose a mechanism that has a substantial influence on the process output.

Choices for adjustable parameters might include heat exchanger fouling factors, ambient heat losses, catalyst activity, or tray efficiency. A secondary benefit of this feedback approach is that the adjustable parameter value can be used for process diagnosis and to trigger maintenance.

The PMBC approach extends to MIMO processes; but to accommodate nonsquare systems, the control law would be the same as Equation 2.13(8).

The same PMBC model can be used for supervisory economic process optimization throughout the process operating range. For nonlinear or nonstationary processes, the self-tuning, diagnostic, and economic optimization advantages of PMBC can offset the engineering effort required to develop an appropriate phenomenological model.

SUMMARY

Model-based controllers have demonstrated economic advantages over classical PID approaches, but they also have disadvantages. In order to develop a front-end model one must initiate substantial process disruptions (step tests) or initiate expensive engineering efforts. Often, additional or improved sensors are also required. Although SISO model-based controllers can be implemented in existing DCS or PLC microprocessors, often an additional microcomputer is also required for model-based controllers. The technology of the modeling, controller, tuning, and optimization approaches requires operator and process engineer training.

Accordingly, MPC methods are recommended for difficult-to-control, multivariable, constrained, and/or economically critical processes.

Bibliography

Cutler, C. R., and Ramaker, B. L., "Dynamic Matrix Control — A Computer Control Algorithm," *American Control Conference Proceedings,* San Francisco: American Automatic Control Council, 1980.

Garcia, C. E., and Morari, M., "Internal Model Control. 1. A Unifying Review and Some New Results," *Industrial and Engineering Chemistry Research,* Vol. 21, 1982, p. 308.

Lee, P. L., and Sullivan, G. R., "Generic Model Control (GMC)," *Computers and Chemical Engineering,* Vol. 12, No. 6, 1988, p. 573.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Marlin, T. E., *Process Control — Designing Processes and Control Systems for Dynamic Performance,* 2nd ed., New York: McGraw-Hill, 2000.

Ogunnaike, B. A., and Ray, W. H., *Process Dynamics, Modeling, and Control,* New York: Oxford Press, 1994.

Ou, J., and Rhinehart, R. R., "Grouped Neural Network Model Predictive Control," *Control Engineering Practice,* Vol. 11, No. 7, 2003, pp. 723–732.

Ou, J., Narayanswamy, G., and Rhinehart, R. R., "External Reset Feedback for Generic Model Control," *ISA Transactions,* Vol. 37, No. 3, 1998, pp. 189–199.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* 2nd ed., New York: John Wiley & Sons, 2003.

Subawalla, H., Paruchurri, V. P., Gupta, A., Pandit, H. G., and Rhinehart, R. R., "Comparison of Model-Based and Conventional Control: A Summary of Experimental Results," *Industrial and Engineering Chemistry Research,* Vol. 35, No. 10, October 1996, pp. 3547–3559.

2.14 Model-Based Predictive Control Patents

D. MORGAN (2005)

Patent Offices:	Canadian Patents Database (patents1.ic.gc.ca/intro-e.html) European Patent Office (EPO) (www.european-patent-office.org/espacenet/info/access.htm) German Patent Office (http://www.dpma.de/index.htm) Japan Patent Office (www.jpo.go.jp) United States Patent Office (USPTO) (www.uspto.gov) World Intellectual Property Organization (WIPO) (www.wipo.org/about-wipo/en/index.html)
Research Tools:	DPMApatentblatt (www.patentblatt.de/eng) INAS Patent Analysis Software (www.mayallj.freeserve.co.uk/inas.htm) IP Menu (www.ipmenu.com) Mospatent (www.mospatent.ru/en/links.htm) Thomson Derwant (www.ipr-village.com/index_ipr.html) Thomson Delphion (www.delphion.com)
Engineering Library:	Linda Hall Library (www.lhl.lib.mo.us)

INTRODUCTION

From optics to distillation to financial derivatives trading to navigation, the applications of model-based predictive control seem practically endless. Likewise, mathematical techniques vary from Model Algorithmic Control to Dynamic Matrix Control to Inferential Control to Internal Model Control. Diverse applications or functions, and techniques or forms, make model-based predictive control popular in the patent literature.

Patent Basics

For a patent to be granted, one has to "tell all," or as put forth by the U.S. Code of Federal Regulations:

> (a) The [Patent] specification must include a written description of the invention or discovery and of the manner and process of making and using the same, and is required to be in such full, clear, concise, and exact terms as to enable any person skilled in the art or science to which the invention or discovery appertains, or with which it is most nearly connected, to make and use the same.
>
> (b) The specification must set forth the precise invention for which a patent is solicited, in such manner as to distinguish it from other inventions and from what is old. It must describe completely a specific embodiment of the process, machine, manufacture, composition of matter, or improvement invented, and must explain the mode of operation or principle whenever applicable. The best mode contemplated by the inventor of carrying out his invention must be set forth.

Thus, the freely available patent databases of the world are a natural place to play out one's interest in any technical field. Whether one is interested in a particular application or the fundamental basics of the calculations involved, ample material exists to satisfy that interest.

The U.S. Patent Office (USPTO) describes a patent as a grant of property right to exclude others from making, using, offering for sale, selling, or importing the invention. Once a patent is issued, the patentee must enforce the patent on his or her own. Any person who "invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent," subject to the conditions and requirements of the law. The word "process" is defined by law as a process, act, or method, and primarily includes industrial or technical processes. Three general types of patents exist worldwide. In the U.S., they are described as follows:

> *Utility patents* may be granted to anyone who invents or discovers any new and useful process, machine, article of manufacture, or compositions of matters, or any new useful improvement thereof.

Design patents may be granted to anyone who invents a new, original, and ornamental design for an article of manufacture.

Plant patents may be granted to anyone who invents or discovers and asexually reproduces any distinct and new variety of plants.

All of the patents discussed here fall under the Utility patent classification.

Basic Patent History

Recent events on the global stage have made the patent process anything but "basic." Thus, a brief recap of the history of industrial intellectual property rights and a summary of recent developments is in order.

The United Kingdom claims to have the longest tradition of granting patents, beginning in 1449 with a "Letters Patent" issued by Henry VI to the manufacturer of stained-glass windows at Eton College. Elizabeth I granted around 50 open-ended term patents between 1561 and 1590 for the manufacture and sale of commodities such as soap, saltpeter, alum, leather, salt, glass, knives, sailcloth, starch, iron, and paper. She rejected the application for a water closet patent, however.

By 1610, concerns about the effect of monopolies led James I to revoke all patents. In 1624, the doctrine of the public interest was incorporated into the Statute of Monopolies. This statute banned monopolies except those "for the term of 14 years or under hereafter to be made of the sole working or making of any manner of new manufactures within this Realm to the true and first inventor." Such monopolies should not be "contrary to the law nor mischievous to the State by raising prices of commodities at home or hurt of trade."

The first U.S. patent law was enacted in 1790. The first U.S. utility patent was issued in 1836 to the inventor of a special type of wheel. Whether this is the source of the phrase "reinventing the wheel" is uncertain. As of this writing, there are over 6.6 million U.S. utility patents.

The first attempt to globalize industrial intellectual property rights was the 1883 Paris Convention. Further efforts led to the creation of the World Intellectual Property Office (WIPO) in 1967. WIPO is a specialized agency of the United Nations system of organizations with 179 member states. It administers 16 industrial property rights and 6 copyright treaties. Globalization efforts culminated with the Patent Cooperation Treaty (PCT). The PCT was concluded in 1970, amended in 1979, and modified in 1984 and 2001. There are well over 100 PCT contracting states at this time.

The PCT consolidates and streamlines the international patent application process for inventors. Although the patent offices of the individual states are still ultimately responsible for granting patents and collecting grant fees, inventors who use international applications under PCT only have to apply once.

They also pay a single application fee in the 1000 to 3000 USD range with the fee variation largely depending on the size of the application, the number of designated jurisdictions, whether they exercise a demand for preliminary examination, and the choice of authority (USPTO or EPO for example). In any event, the costs of multiple foreign filings dwarf the PCT fees. Moreover, while buying time for further commercial assessments, the PCT filing fees are often less than the time value of money incurred with multiple domestic and foreign filings. The process allows an inventor to establish global priority with one application.

WIPO promotes the value of the international patent search they sell. The results are set out in an "international search report." This report is generally made available to the inventor by the fourth or fifth month after the international patent application is filed. The international search report contains no comments on the value of an invention but lists prior art documents from the 30 million-plus universe of international patents (some of which are duplicates). Thus it comments on the possible relevance of the citations to the questions of novelty and inventive step (nonobviousness). This enables one to evaluate the chances of obtaining patents in the countries designated.

The most immediate effect the patent researcher will notice is the WO patent applications—the designation given to those applications filed under the PCT process. When a patent is finally granted by an individual state, the standard numbering for that state will be employed, prefixed by the state's two-letter code.

Online Patent Databases

Currently, several online patent databases that anyone can search are available. The approaches taken vary greatly. For example, some databases have global reach while others are restricted to the country that administers them. The USPTO is an example of the latter and the German database of the former. Whichever database you choose to search, you can retrieve four basic types of records:

Bibliographical—Information such as Title, Patent Number, Publication Date, Inventor(s), Applicant(s), Application Number, Priority Number(s), Classification, Literature References, and Prior Patents Cited. What you obtain depends on the country of origin and the date, as newer patents tend to be filed under the PCT.

Abstracts—Brief descriptions of the invention, usually available in English. The Japanese patent office has a "just in time" Japanese-to-English converter that is quite accurate.

Full Text—Text-only form of a patent, usually in HTML format. Full text formats differ from country to country. Some, such as the Canadian and the European Patent Offices, include drawings. There are some drawbacks to full text. For example, the U.S. full text patent database attempts to convey mathematical formulae using a plain English code page.

The USPTO provides full text records for patents back to 1976. Full text records often provide hyperlinks to other information, such as prior patents that were cited by the inventor. This is the type of information that allows for automated research.

Scanned Image — Practically all patents are now available or will be in the future in scanned form. PDF or TIFF file formats are usually employed. While not convenient for automated research, one often will ultimately need to obtain the original image in order to obtain diagrams and mathematical formulae in their original form. Patents in this form are usually in the language preferred in the country of origin. Unfortunately, at today's bandwidth, one can only download one page at a time.

Several countries and private organizations provide online patent databases that anyone can search. Some, such as Germany's DEPATISNet and the European Patent Office's esp@cenet, cover multiple foreign databases with their search engines. Others, such as the United Kingdom, rely on third parties, the European Patent Office (EPO), to provide such services. Currently the EPO's esp@cenet has the following coverage:

1. Abstracts, bibliographic data and full text:
   EPO, France, Germany, Switzerland, United Kingdom, United States, and WIPO
2. Abstracts and bibliographic data:
   China and Japan
3. Bibliographic data only:
   Argentina, ARIPO, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Croatia, Cuba, Cyprus, Czechoslovakia, Czech Republic, Denmark, Egypt, Ellas, Eurasian Patent Office, Finland, GDR, Hong Kong, India, Ireland, Israel, Italy, Kenya, Korea, Latvia, Lithuania, Luxembourg, Malawi, Mexico, Moldova, Monaco, Mongolia, The Netherlands, New Zealand, Norway, OAPI, Philippines, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, South Africa, Soviet Union, Spain, Turkey, Vietnam, Yugoslavia, Zambia, Zimbabwe

Most online databases are accessible via an epacenet.com link that is a combination of the country's two-letter abbreviation and "espacenet.com." Whenever a county also maintains its own freely searchable database, the link for that database is given. In the case of Germany, the German database itself is linked to other databases—much like espacenet—but in German.

Patent Contents

Besides the basic types of information noted above, patents contain a wealth of useful information for a researcher. At the beginning of U.S. patents in full text form, one first sees an abstract. This is followed by the bibliographic information. Next come links to patents cited by the inventor as prior art. Both patents that are useful to the present invention and those that will be improved upon by it are cited. Often what follows is a list of references to other, nonpatent resources that either aid in the understanding of the current patent, are to be incorporated by reference, or further describe the prior art to be improved upon.

Following the bibliographical and claims information usually comes a brief summary followed by a description of the prior art. Prior art descriptions often contain references to previous patents and the like, making it easy for one to quickly grasp an understanding of the field of invention and to understand what is to be improved upon. Next, the inventor will typically give a brief description of how the prior art is to be improved upon followed by a description of any drawings accompanying the patent. What then follows is a detailed description of the invention, its preferred embodiments, and any examples. Examples often include sample data and even source code in order to facilitate duplication of the work by anyone skilled in the art.

GUIDE

Patent literature concerning model-based control is rich in both form and function. Form represents the various core technologies that may or may not have developed into specific functions. Function represents specific implementations that may or may not have developed from established forms. The following attempt to produce a guide that characterizes patents according to their perceived form and/or function is based on a survey of over 280 U.S. patents pertaining directly to either the form and/or function of model-based control. These patents in turn cite over 1000 other patents and over 750 other references.

Such supporting citations either deal directly with the form of model-based control or some superseding technological function. The U.S. patent database Web site provides a very convenient means of surfing to both the patents that a given patent cites as well as those patents that cite a given patent. It also makes its own attempt at classification. Other references can be obtained online from sources such as the Linda Hall Library in Kansas City, Missouri, or in some other fashion via your own local university library system.

Forms

Prediction and adaptation are the business of Model-Based Predictive Control (MBPC). To those ends, a wide variety of approaches exist that fit fairly well into two major classes.

The first class involves predictors based on rigorous mathematical models of specific processes.

Generic controllers that can be used on any process form the other major class of MBPC patents. Early implementations focused on making controllers "adaptive." Over time and with the decrease in memory costs, multivariable predictive controllers gained in importance. One subclass of such controllers is known as extended horizon predictive and predictive adaptive controllers. This subclass is further divided according to the method used to supply the "extra" inputs that extend beyond the process horizon or delay time.

These controllers include the Extended Horizon Adaptive Controller (EHC), the Receding Horizon Adaptive Controller (RHC), the Control Advance Moving Average Controller (CAMAC), the Extended Prediction Self-Adaptive Controller (EPSAC), the Generalized Predictive Controller (GPC), the Model Predictive Heuristic Controller (MPHC), and the Dynamic Matrix Controller (DMC).

Neural net-based predictors, often called "model-free predictors" because very little understanding of the process is required, are also popular members of this class. Monte Carlo simulation, Kalman filters, hill climbing, constraint control, least squares, linear programming, quadratic programming, fuzzy logic, and partial differential equation-based approaches round out the generic predictor universe. Many of these known methods are represented in the U.S. patent literature.

U.S. Patents 4,349,869 and 4,616,308 and PCT patent WO03060614 all deal with the well-known Dynamic Matrix Control, or DMC, form of Model-Based Predictive Control (MBPC). The PCT version, published July 24, 2003, provides step-by-step guidance through an actual distillation unit example (Figure 2.14a), with actual input and output data. The U.S. patent versions, also highly detailed with respect to the calculation procedures, suggest methods for coupling DMC with an offline economic optimizer applied to a typical FCC.

U.S. Patent 4,358,822, "Adaptive-Predictive Control System," also provides detailed calculation procedures, two examples, and source code. The first example is a binary distillation unit. The second is for a single-input–single-output (SISO) process. The well-documented source code is for an adaptive control program capable of controlling SISO processes with time delays.

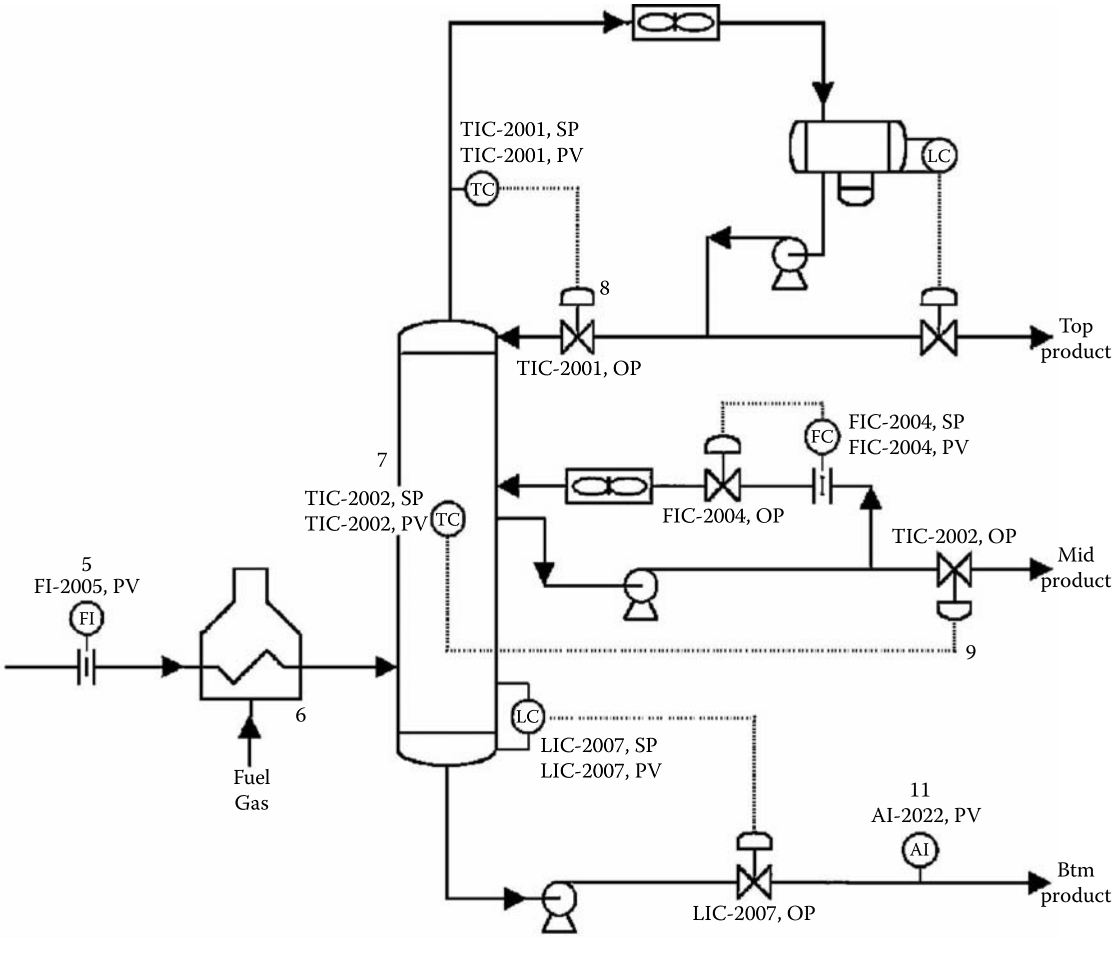


FIG. 2.14a

The Dynamic Matrix Control (DMC) form of Model-Based Predictive Control (MBPC), provides step-by-step guidance through an actual distillation unit example with actual input and output data.

Functions

As mentioned above, the applications for model-based control are practically limitless. As one would suspect, over time the range of applications has grown exponentially. While we tend to equate model-based predictive control with advanced digital control systems, inventors were applying analog systems well before the dawn of the digital age. The prime advantage that digital computing holds over analog systems is not that it is somehow easier for the inventor to perfect. Rather, it has more to do with the way in which computers facilitate dissemination of the art to those who actually apply it.

Auto-piloting aircraft were an early impetus for the art, and automation of advanced military craft remains a significant driving force. Recently, interest in automotive applications has grown. The art is actually applied in order to facilitate its application via computers. The literature on the use of MBPC in disk-drive tracking is extensive, for example. That the art itself is useful in bettering itself is rather unique.

Patents that are specific in form usually target a single process. However, those intended for generic use will often provide details, references, and examples of their use in specific processes. What follows is a guide to patents that are either specifically intended for a given field or those that cite a given field as an example.

Aircraft Aircraft experience large variations in the gains and time constants of the external disturbances they encounter, making them strong candidates for the application of model-based control. As time is compressed more and more for fighter pilots, model-based control is finding a niche in target identification and weapons targeting. One can even find patents related to missile flight path control. Today's stealth aircraft, such as the F-117A "Wobblin Gobblin," in which function is often sacrificed for form, would not be flyable save for MBPC.[1,2]

Computing MBPC finds application in two specific areas of computing—wafer fab and disk-drive positioning. In wafer fab, the goal is the reduction in run-by-run variability of the CVD, spin coating, and etching processes. CVD temperature control garners particular attention. Disk-drive accuracy and hence speed can be enhanced via MBPC, just as with any other servo process.[3,4]

Utility Power MBPC patents relating to power generation and distribution attack problems at both the system and equipment levels. For example, there are MBPC patents for better control of boiler chemicals as well as patents for optimal dispatching of multi-unit cogeneration facilities.

More and more plants designed for base load service are being used to meet peaking power demand. Steam turbines designed to operate in a continuous range of steady-state conditions are being cycled to meet peaking load. This can induce large thermal stresses in both the turbine and the boiler as a result of steep steam-to-metal temperature gradients that develop during rapid loading or unloading of the turbine. Thus, much attention is given to predicting loads based on extended load profiles. Figure 2.14b gives an example of a MBPC design

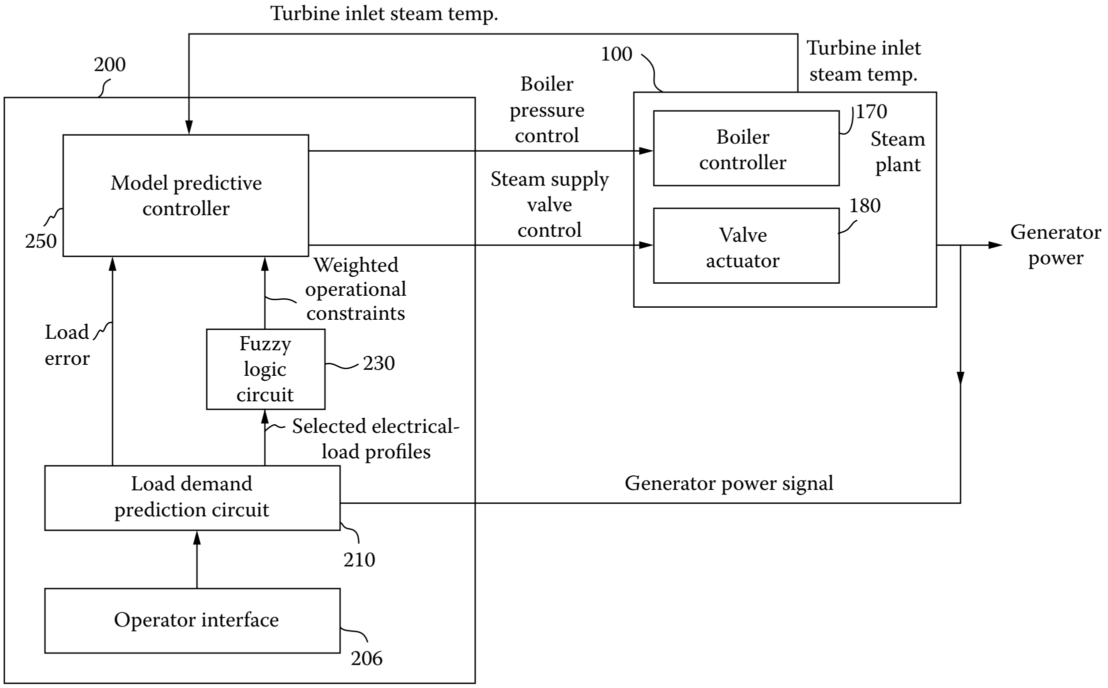


FIG. 2.14b
Patent for MBPC design used in peak-shedding applications to predict loads based on extended load profiles.

TABLE 2.14c
Steam Utility Patents and Their Areas of Application

Steam Power Patents

Patent No.	*Title*	*Year*	*Application*	*Approach*
3,758,762	Decoupled feedforward-feedback control system	9/11/1973	STEAM TURBINE	MODEL
3,939,328	Control system with adaptive process controllers especially adapted for electric power plant operation	2/17/1976	TURBINE	SELF ADAPTIVE
4,027,145	Advanced control system for power generation	5/31/1977	COGENERATION	ROM
4,110,825	Control method for optimizing the power demand of an industrial plant	8/29/1978	COGENERATION	ROM
4,258,424	System and method for operating a steam turbine and an electric power generating plant	3/24/1981	STEAM TURBINE	ROM
4,297,848	Method of optimizing the efficiency of a steam turbine power plant	11/3/1981	STEAM TURBINE	ROM
4,403,293	Control apparatus for use in multiple steam generator or multiple hot water generator installations	9/6/1983	COGENRATION	LP
4,628,462	Multiplane optimization method and apparatus for cogeneration of steam and power	12/9/1986	COGENRATION	MULTIPLANE
4,630,189	System for determining abnormal plant operation based on whiteness	12/16/1986	POWER PLANT	AR WHITENESS
4,768,143	Apparatus and method using adaptive gain scheduling algorithm	8/30/1988	WATER HEATER	LEAST SQUARES
4,802,100	Advanced cogeneration control system	1/31/1989	COGENRATION	ROM
4,805,114	Economical dispatching arrangement for a boiler system	2/14/1989	COGENRATION	ROM
4,922,412	Apparatus and method using adaptive gain scheduling	5/1/1990	WATER HEATER	LEAST SQUARES
4,969,408	System for optimizing total air flow in coal-fired boilers	11/13/1990	COAL FURNACE	ROM
5,040,725	Adaptive controller for forced hot water heating systems	8/20/1991	WATER HEATING	ADAPTIVE
5,159,562	Optimization of a plurality of multiple-fuel fired boilers using iterated linear programming	10/27/1992	BOILER	LP
5,268,835	Process controller for controlling a process to a target state	12/7/1993	STEAM GEN	ROM
5,305,230	Process control system and power plant process control system	4/19/1994	STEAM TURBINE	NEURAL NET
5,311,421	Process control method and system for performing control of a controlled system by use of a neural network	5/10/1994	STEAM TURBINE	NEURAL NET
5,347,466	Method and apparatus for power plant simulation and optimization	9/13/1994	STEAM TURBINE	ROM
5,442,544	Single input single output rate optimal controller	8/15/1995	ELECTRIC UTILITY	ROC
5,517,424	Steam turbine fuzzy logic cyclic control method and apparatus therefore	5/14/1996	STEAM TURBINE	FUZZY LOGIC
5,619,433	Real-time analysis of power plant thermohydraulic phenomena	4/8/1997	POWER PLANT	RELAP5/MOD3
5,696,696	Apparatus and method for automatically achieving and maintaining congruent control in an industrial boiler	12/9/1997	BOILER CHEMS	MACC
5,923,571	Apparatus and method for automatic congruent control of multiple boilers sharing a common feedwater line and chemical feed point	7/13/1999	BOILER CHEMS	ROM

from U.S. Patent 5,517,424, which is intended to address this issue. Table 2.14c provides a sample of steam utility patents and their application forms.[5–7]

Nuclear Energy Although regulatory issues tend to discourage the application of MBPC to direct control of nuclear power systems, MBPC still finds use in the industry. For example, motor operator valve (MOV) reliability is a subject of great concern. Hundreds of MOV failures have been investigated in different studies. In one study, electromechanical torque switches and limit switches were identified as the components at the root of approximately 32% of the documented MOV failures.

Mechanical failures (failure to operate, bent stems, damage to valve seats, gear binding and damage) accounted for 22% of the MOV failures. U.S. Patent 4,694,390 seeks to address this problem by instrumenting MOVs with a valve stem position sensor, a stem load sensor, and motor load sensor. Then, an MBPC-based monitor periodically tests the status of the motor-operated valve and turns off the power to the valve's motor if certain predefined criteria for the valve stem position and the valve load are not satisfied.

Pressurized and boiling water reactor core monitoring is an important area for MBPC. For example, U.S. Patent 4,770,843 discloses a MBPC application "for controlling the stability of a boiling water reactor using a digital computer to calculate online, from distributed steady-state values of only power, flow, enthalpy, and pressure; a stability index for selected fuel assemblies taking into account nuclear feedback as well as detailed hydrodynamic effects." U.S. Patent 4,234,925 applies linear programming (LP) to the problem of generating the optimum excitation waveform for excitation of the plasma in fusion reactors. U.S. Patent 4,982,320 discloses a self-adaptive controller for particle beam accelerator control.[8,9]

Motor Fuel Production Patents on processes ranging from gas lift oil production to hydrogen fluoride (HF) alkylation target the motor fuel production industry, where small optimizations can mean large returns. Hydrocracking and Fluid Catalytic Cracking (FCC) are especially well represented. U.S. Patent 4,349,869 provides a detailed example of the application of Dynamic Matrix Control coupled with offline optimization to FCC.[10,11]

Pulp and Paper MBPC patents target practically every aspect of pulp and paper production. Digestion, delignification, additive control, and final webbing all gain attention.[12,13]

Distillation Distillation is difficult to optimize because of its numerous degrees of freedom both controllable, such as reboiler and heat input, and uncontrollable, such as feed composition. Making matters worse are the strong couplings between variables and the nonlinearity encountered in high-purity towers. For example, as stated in U.S. Patent 5,343,407, "a 10% increase in boilup rate will result in only a moderate increase in the purity of the bottoms product, while a 10% decrease would cause a drastic decrease in the purity of the bottoms product. In addition, a 10% increase in boilup by itself will cause a drastic reduction in the purity of the overhead product."

Despite the inherent difficulties, high-purity distillation is a very common process. As noted again in U.S. Patent 5,343,407, "ethylene, propylene and styrene monomers of nearly 100% purity are required for their respective polymerization processes in order to produce polymers with the desired characteristics. Also, for the production of industrial grade acetic acid, levels of less than 200 ppm propanoic acid impurity must be maintained. In addition, chemical intermediate xylene products are typically produced as high purity products. Ethylene oxide and propylene oxide are separated industrially to produce products, each with about 200 ppm impurities."

Distillation's prominence coupled with its tricky nature makes it an ideal candidate for MBPC solutions. Table 2.14d

TABLE 2.14d
Patents Related to Distillation and Their Applications

Distillation-Related Patents				
Patent No.	*Title*	*Date*	*Application*	*Approach*
3,976,179	Controlling the temperature of a depropanizer tower by chromatographic analysis of feed and bottoms	8/24/1976	DEPROPANIZER	RIED VP MODEL
4,030,986	Control for maximizing capacity and optimizing product cost of distillation column	6/21/1977	DISTILLATION	COST FACTOR MODEL
4,070,172	Pressure responsive fractionation control	1/24/1978	DEMETHANIZER	ROM
4,252,614	Control of multiple feed fractional distillation column	2/24/1981	MULTI-FEED	ROM
4,230,534	Control of a fractional distillation column	10/28/1980	DISTILLATION	ROM
4,358,822	Adaptive–predictive control system	11/9/1982	DISTILLATION	DMC
4,367,121	Fractional distillation column control	1/4/1983	DISTILLATION	INTERNAL REFLUX MODEL
4,526,657	Control of a fractional distillation process	7/2/1985	DISTILLATION	ROM
4,560,442	Fractional distillation process control	12/24/1985	DISTILLATION	INTERNAL REFLUX MODEL
4,889,600	Fractionating column control apparatus and methods	12/26/1989	DISTILLATION	REID VP MODEL
5,132,918	Method for control of a distillation process	7/21/1992	DISTILLATION	ROM
5,260,865	Nonlinear model-based distillation control	11/9/1993	DISTILLATION	ROM
5,343,407	Nonlinear model-based distillation control	8/30/1994	DISTILLATION	MCCABE THIELE MODEL
5,396,416	Multivariable process control method and apparatus	3/7/1995	GAS FRACTIONATION	SELF ADAPTIVE DMC
5,477,449	Hierarchical model predictive control system based on accurate manipulation variable prediction	12/19/1995	DISTILLATION	DMC VARIANT
5,522,224	Model predictive control method for an air-separation system	6/4/1996	AIR SPEARATION	ROM

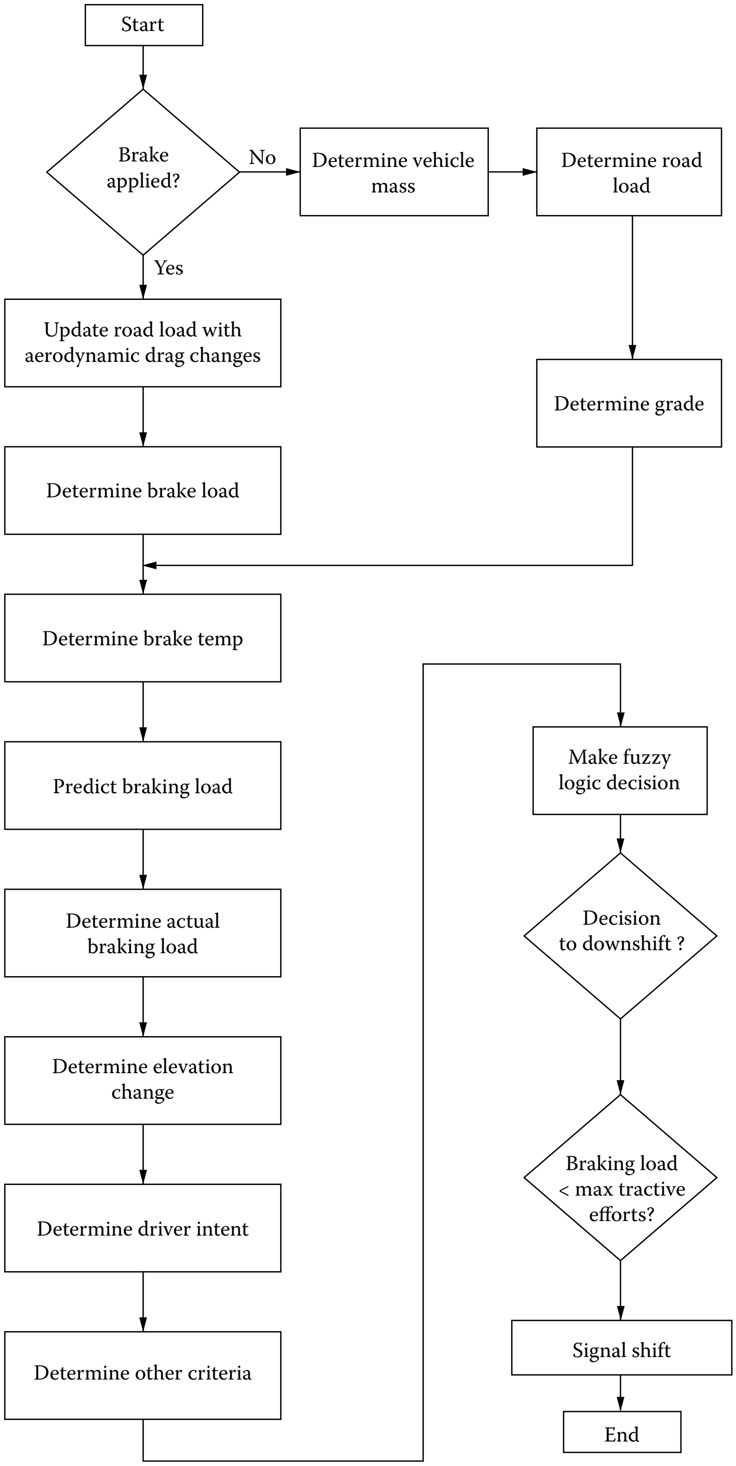


FIG. 2.14e
Patent describing a control strategy for braking.

provides a sample of distillation-related patents and their forms.[14,15]

Robotics It is unfortunate that something once labeled an "autonom"—which would imply independent action—has been more recently labeled a "robot," when that implies something that performs the same task in the same way, over and over again. The patent literature alone would beg to differ as it reflects a movement to make robots more adaptive and autonomous.

Today, hybrid robotic control incorporates adaptive force and position controllers based on highly rigorous models.

Automatic targeting systems can recognize moving targets and predict where they will be once fired upon. Unmanned aerial vehicles are able to perform missions unattended.[16,17]

Optics, Photography, and Astronomy Many of the technologies for the Keck telescope on the Mauna Kea volcano in Hawaii are fairly well represented in the patent literature, including relatives of its MBPC-based adaptive optics. Looking back at Earth, U.S. Patent 4,748,448 discloses a MBPC approach to gauging wind speed and direction from oceanic wave action observed from a satellite. Optimization and quality control of photographic emulsion products are also represented.[18,19]

Automotive The automotive field is one of the latest to which MBPC has been applied. Fuel-to-air ratio (lambda), braking, and cylinder disablement are some examples of areas where both rigorous and generic models are just some of the applications where inventions have been disclosed. U.S. Patent 6,625,535 discloses a control scheme for braking, whose logic is shown in Figure 2.14e.[20,21]

Polymers The patent literature describes MBPC-based inventions related to monomer production (for example, ethylene), polymerization reactions, and the production of polymer suspensions.

An area of particular interest is predicting polymer melt index from process data rather than having to rely on samples taken to the lab. As U.S. Patent 5,504,166 states,

> The desired control of the polymerization process is extremely difficult to attain because of the holdup time of polymerization reactors and the time involved in obtaining polymer samples and measuring the properties of those samples. Because of this time period, the polymerization conditions employed in the reactor at the time at which a property of a polymer sample withdrawn from the reactor is measured are not necessarily the same as the polymerization conditions employed in the reactor at the time at which such polymer sample was produced in the reactor and/or withdrawn from the reactor.
>
> This is especially the case when the attempted control of the polymerization process is based on the measurement of the melt flow rate — or in other words, the melt index — of the polymer product as determined according to the ASTM Test D-1238-62T. Although the melt flow rate or the melt index is a satisfactory control property for most solid polymers prepared from alpha-olefins, the time consumed in obtaining a polymer sample for measurement and in measuring the melt index of the sample combines with the aforesaid holdup time of the reactor to seriously hamper accurate control of the polymerization process.[22,23]

Financial Derivatives Trading Underlying the price of a commodity is a market process that can be modeled just like any other process. A trader makes a series of inputs to a continuous process based on analysis of feedback from a market.

"Financial engineering" was born in 1973 when a mathematician, Fischer Black, and an economist, Myron Scholes, devised one of the first mathematically accepted approaches for pricing options that can only be exercised at their expiration date ("European options"). Today there are well over a thousand "systems trading" applications on the market that can be used by anyone to "control" portfolios.

According to U.S. Patent 6,546,375: "What has become known as the Black–Scholes option formula was described first in 'The pricing of options and corporate liabilities,' *Journal of Political Economy* 81 (1973). The Black–Scholes option formula is presently of widespread use in financial markets all over the world. The price of such an option can be found by solving the Black–Scholes equation with the initial condition at expiration (i.e., the payoff of the option). The Black–Scholes equation is a reverse diffusion equation with parameters determined by the statistical characteristics of involved stocks and currencies such as risk-free interest rate, holding cost or expected dividends, and volatility."

Determining the forward pricing of "American options," which can be exercised before expiration, is difficult because it leads to an infinite-dimensional free boundary problem that cannot be solved explicitly nor finitely. In order to approximate a solution, U.S. Patent 6,546,375 employs a "discretized partial differential linear complementary problem (PDLCP)-based system." An "optimization problem in the form of a mathematical program with equilibrium constraints (MPEC)" is also used to derive implied volatilities of the assets underlying the subject derivatives.

Other patents integrate the MBPC side of the overall process with the requisite feedback information systems, much as a DCS workstation works with an operator and a process. Orders can be placed, real-time data can be analyzed, margins can be calculated, and buy and sell signals can be generated automatically via one's Internet-ready PC.[24,25]

References

1. U.S. Patent 4,258,545, "Optimal Control for a Gas Turbine Engine."
2. U.S. Patent 4,812,990, "System and Method for Optimizing Aircraft Flight Path."
3. U.S. Patent 5,646,870, "Method of Setting and Adjusting Process Parameters to Maintain Acceptable Critical Dimensions Across Each Die of Mass-Produced Semiconductor Wafers."
4. U.S. Patent 6,625,513, "Run-to-Run Control over Semiconductor Processing Tool Based Upon Mirror Image Target."
5. U.S. Patent 4,628,462, "Multi-Plane Optimization Method and Apparatus for Cogeneration of Steam and Power."
6. U.S. Patent 5,696,696, "Apparatus and Method for Automatically Achieving and Maintaining Congruent Control in an Industrial Boiler."
7. U.S. Patent 5,923,571, "Apparatus and Method for Automatic Congruent Control of Multiple Boilers Sharing a Common Feedwater Line and Chemical Feed Point."
8. U.S. Patent 5,392,320, "Core Automated Monitoring System."

9. U.S. Patent 4,694,390, "Microprocessor-Based Control and Diagnostic System for Motor-Operated Valves."
10. U.S. Patent 4,437,977, "Control of a Catalytic Cracking Unit."
11. U.S. Patent 6,595,294, "Method and Device for Gas-Lifted Wells."
12. U.S. Patent 5,301,102, "Multivariable Control of a Kamyr Digester."
13. U.S. Patent 6,328,851, "Method and Equipment for Controlling Properties of Paper."
14. U.S. Patent 4,560,442, "Fractional Distillation Process Control."
15. U.S. Patent 5,522,224, "Model Predictive Control Method for an Air-Separation System."
16. U.S. Patent 4,621,332, "Method and Apparatus for Controlling a Robot Utilizing Force, Position, Velocity, Spring Constant, Mass Coefficient, and Viscosity Coefficient."
17. U.S. Patent 4,725,942, "Controller for Multi-Degree of Freedom Nonlinear Mechanical System."
18. U.S. Patent 5,265,034, "Feedback Controlled Optics with Wave-Front Compensation."
19. U.S. Patent 5,067,067, "Method for Evaluating and Designing Lenses."
20. U.S. Patent 6,594,573, "Method for Regulating a Clutch or a Brake in a Transmission."
21. U.S. Patent 6,625,535, "Adaptive Power-Train Braking Control with Grade, Mass, and Brake Temperature."
22. U.S. Patent 4,668,473, "Control System for Ethylene Polymerization Reactor."
23. U.S. Patent 6,106,785, "Polymerization Process Controller."
24. U.S. Patent 6,061,662, "Simulation Method and System for the Valuation of Derivative Financial Instruments."
25. U.S. Patent 6,546,375, "Apparatus and Method of Pricing Financial Derivatives."

2.15 Model-Free Adaptive (MFA) Control

G. S. CHENG (2005)

Model-free adaptive (MFA) control, as its name suggests, is an adaptive control method that does not require process models. An MFA control system is defined to have the following properties:

1. No precise quantitative knowledge of the process is available.
2. No process identification mechanism or identifier is included in the system.
3. No controller design for a specific process is needed.
4. No manual tuning of controller parameters is required.
5. Closed-loop system stability analysis and criteria are available to guarantee the system stability.[1–3]

Derivations of the core MFA control technology address specific control problems as described here:[4–14]

- SISO MFA to replace PID so that manual controller tuning is eliminated
- Nonlinear MFA to control nonlinear processes
- MFA pH controller to control pH processes
- Feedforward MFA controller to deal with measurable disturbances
- Antidelay MFA to control processes with large time delays
- Robust MFA to protect the process variable from running outside a bound
- Time-varying MFA controller to control time varying processes
- Antidelay MFA pH controller for pH processes with varying time delays
- MIMO MFA to control multivariable processes

MFA controllers can be readily embedded into various control equipment and are becoming available on more and more platforms offered by multi-vendors including building controllers, single-loop controllers, programmable logic controllers (PLC), hybrid controllers, process automation controllers (PAC), control software, and distributed control systems (DCS).

SINGLE-LOOP MFA CONTROL SYSTEM

Figure 2.15a illustrates a single-loop MFA control system that includes a single-input single-output (SISO) process, a SISO MFA controller, and a feedback loop. The control objective is for the controller to produce an output $u(t)$ to force the process variable $y(t)$ to track the given trajectory of its setpoint $r(t)$ under variations of setpoint, disturbance, and process dynamics. In other words, the task of the MFA controller is to minimize the error $e(t)$ in an online fashion, where $e(t)$ is the difference between the setpoint $r(t)$ and the process variable $y(t)$. The minimization of error $e(t)$ is achieved by (i) the regulatory control capability of the MFA controller, and (ii) the adjustment of the MFA controller weighting factors that allow the controller to deal with the dynamic changes, disturbances, and other uncertainties of the control system.

MFA Controller Architecture

Figure 2.15b illustrates the core architecture of a single-input single-output MFA controller. A multilayer perceptron (MLP) artificial neural network (ANN) is used in the design of the controller. The ANN has one input layer, one hidden layer with N neurons, and one output layer with one neuron.

Within the neural network there is a group of weighting factors (w_{ij} and h_i) that can be updated as needed to vary the behavior of the dynamic block. The algorithm for updating the weighting factors is based on the goal of minimizing the error between the setpoint and process variable. Since this effort is the same as the control objective, the adaptation of the weighting factors can assist the controller in minimizing the error while process dynamics are changing. From another point of view, the artificial neural network–based MFA controller "remembers" a portion of the process data providing valuable information for the process dynamics. In comparison, a digital version of the PID controller remembers only the current and previous two samples. In this regard, PID has

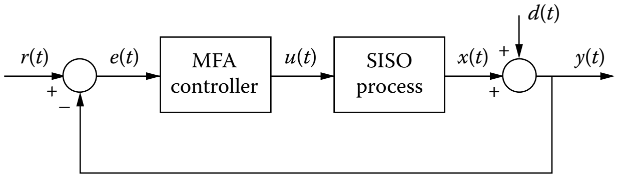


FIG. 2.15a
Single-loop MFA control system.

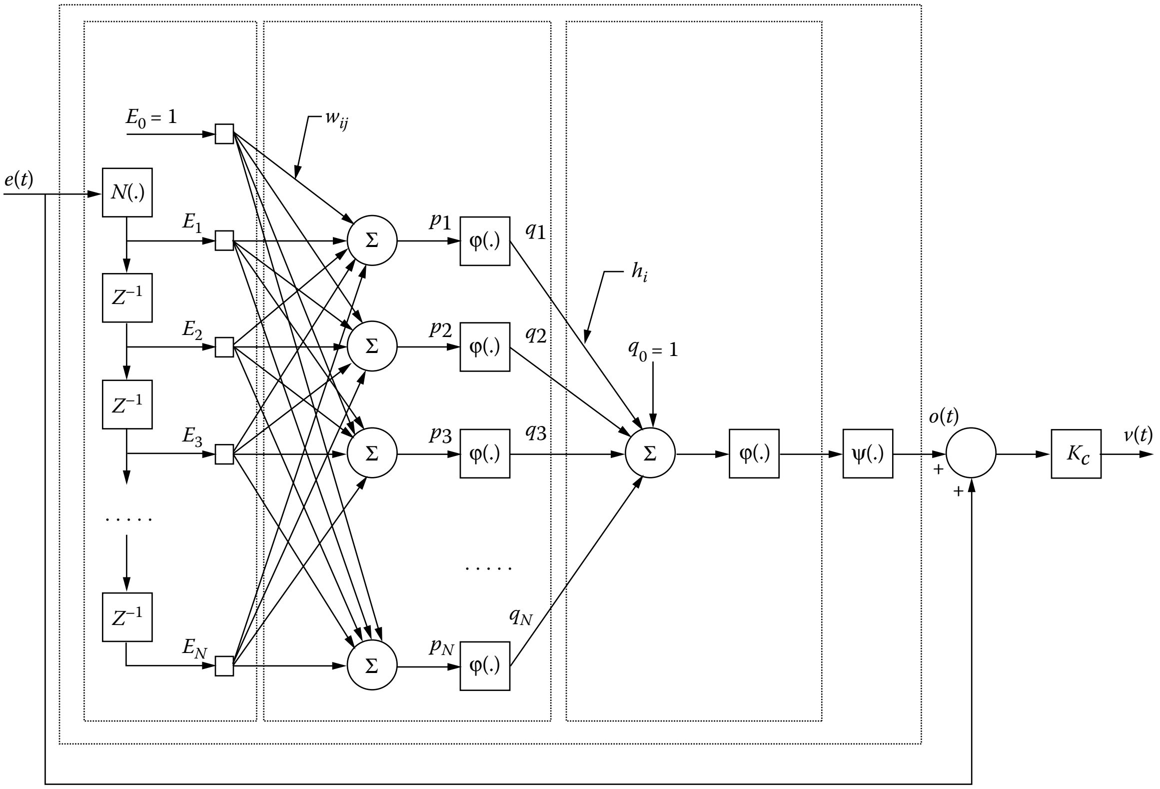


FIG. 2.15b
Architecture of a SISO MFA controller.

almost no memory, and MFA possesses the memory that is essential to a "smart" controller.

SISO MFA Control Algorithm

The core MFA control algorithm comprises the following difference equations:

$$p_j(n) = \sum_{i=1}^{N} w_{ij}(n)E_i(n) + 1, \qquad \textbf{2.15(1)}$$

$$q_j(n) = \varphi(p_j(n)), \qquad \textbf{2.15(2)}$$

$$\begin{aligned} o(n) &= \psi\left[\varphi\left(\sum_{j=1}^{N} h_j(n)q_j(n) + 1\right)\right], \\ &= \sum_{j=1}^{N} h_j(n)q_j(n) + 1, \end{aligned} \qquad \textbf{2.15(3)}$$

$$v(t) = K_c[o(t) + e(t)], \qquad \textbf{2.15(4)}$$

where n denotes the nth iteration, $o(t)$ is the continuous function of $o(n)$, $v(t)$ is the output of the MFA controller, $K_c > 0$ and is the MFA controller gain, and w_{ij} and h_j are weighting factors.

The weighting factors can be updated online at every sample interval using the following formulas:

$$\Delta w_{ij}(n) = \eta K_c e(n) q_j(n)(1 - q_j(n)) E_i(n) \sum_{k=1}^{N} h_k(n), \qquad \textbf{2.15(5)}$$

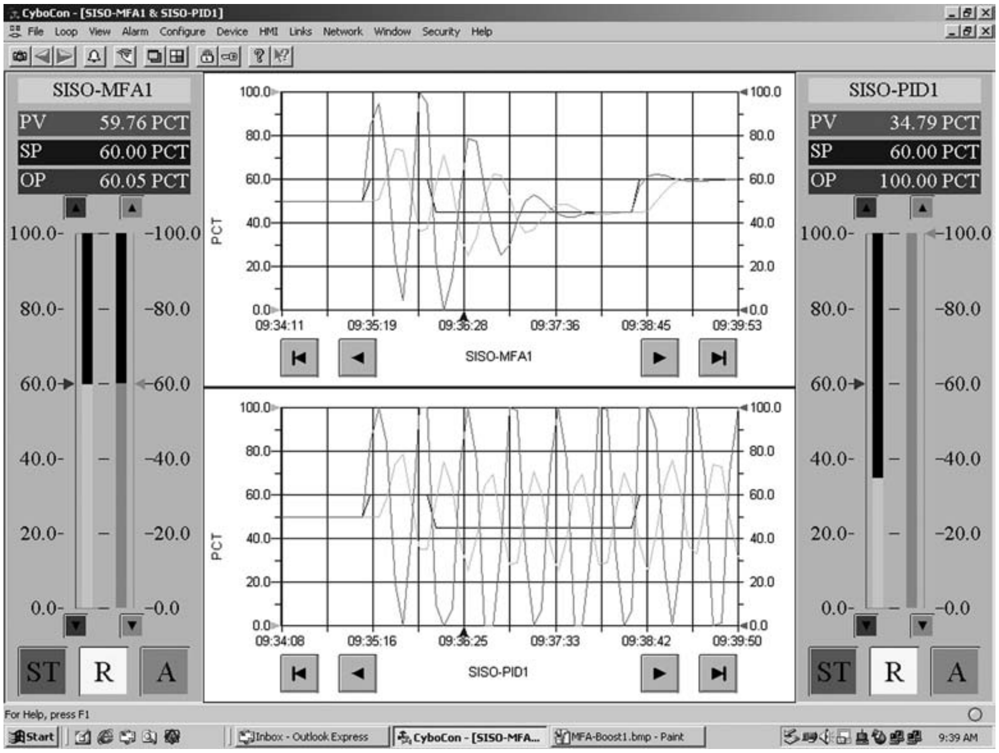
$$\Delta h_j(n) = \eta K_c e(n) q_j(n). \qquad \textbf{2.15(6)}$$

A more detailed MFA control algorithm and discussions can be found in references 1 and 2.

MFA and PID

Most industrial processes are still being controlled by PID (proportional–integral-derivative) controllers. PID is a simple general-purpose automatic controller that is useful for controlling simple processes. However, PID has major problems in controlling complex systems and also requires frequent manual tuning of its parameters when the process dynamics change. The performance of MFA (top) and PID (bottom) controllers is compared in Figure 2.15c to show how MFA adapts when process dynamics change.

Starting from the same oscillating control condition, the system will continue to oscillate under PID control, while the MFA system will quickly adapt to an excellent control



SISO-MFA1
PV 59.76 PCT
SP 60.00 PCT
OP 60.05 PCT

SISO-PID1
PV 34.79 PCT
SP 60.00 PCT
OP 100.00 PCT

FIG. 2.15c
Comparison of MFA and PID.

condition. If both controllers start from a sluggish situation, MFA will control the process faster and better.

MFA Control System Requirements

As a feedback control system, MFA requires the process to have the following behavior:

1. The process is controllable.
2. The process is open-loop stable.
3. The process is either direct or reverse acting (process does not change its sign).

If the process is not controllable, improvement of the process structure or its variable pairing is required. If the process is not open-loop stable, it is always a good practice to stabilize it first. However, for certain simple open-loop unstable processes such as a non-self-regulating level loop, no special treatment is required when using MFA. If a process changes its sign within its operating range, special MFA controllers are required.[1]

SISO MFA Configuration

A SISO MFA controller has only a few parameters to configure:

1. Sample interval—the interval between two samples or calculations in seconds. A high-speed MFA controller can run at a 1 millisecond rate.
2. Controller gain—use of a default value is recommended.
3. Time constant—a rough estimate of the process time constant in seconds.
4. Acting type—direct or reverse action of the process. If the process input increases and then its output increases, it is direct acting, and vice versa. However, MFA controllers embedded in various platforms always use the vendor's definition. Sometimes, controller acting type is used, which is different than the process acting type.

According to the principles in the information theory, it is required that sample interval be less than or equal to one third of the time constant. That is,

$$T_s \leq \frac{1}{3} T_c, \qquad 2.15(7)$$

where T_s is the sample interval, and T_c is the time constant.

Once the configuration is done, MFA can be launched at any time and will control the process immediately. MFA does not require process identification and is not a dynamic modeling-based controller; there is no need to first collect data to train the model. MFA controllers can be switched between automatic and manual at any time. No specific bumpless transfer procedure is required.

NONLINEAR MFA CONTROLLER

Nonlinear control is one of the most challenging topics in modern control theory. Although linear control system theory has been well developed, it is the nonlinear control problems that present the most headaches. The main reason that a nonlinear process is difficult to control is because there could be so many variations in process nonlinear behavior. Therefore, it is difficult to develop a single controller to deal with the various nonlinear processes. Traditionally, a nonlinear process has to be linearized first before an automatic controller can be effectively applied. This is typically achieved by adding a reverse nonlinear function to compensate for the nonlinear behavior so that the overall process input–output relationship becomes somewhat linear. It is usually a tedious job to match the nonlinear curve, and process uncertainties can easily ruin the effort.

The nonlinear MFA controller is a general-purpose controller that provides a more uniform solution to nonlinear control problems. The nonlinear MFA controller is well suited for nonlinear processes or processes with nonlinear sensors, actuators, and other elements.

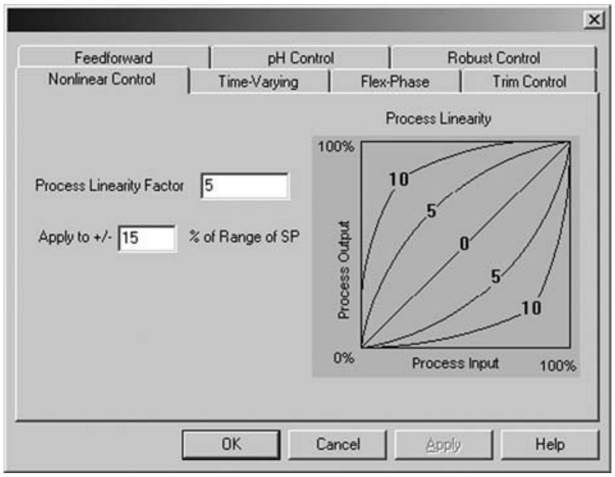
A flow or high-pressure loop is a typical nonlinear process that can cause the actuator to lose its authority in different operating conditions. Inevitable wear and tear on a valve typically makes a linear valve nonlinear. The dissolved oxygen in a bio-tech micro reactor to cultivate cells is another nonlinear process example. As cells grow, they suddenly start to consume much more oxygen. Since the number of bio-tech experiments is huge and the types of cells to grow can vary significantly, it is difficult and costly to deal with nonlinear characterization problems. The general-purpose nonlinear MFA controller is well suited for this application.

Nonlinear MFA Configuration

In addition to the parameters used in SISO MFA including sample interval, time constant, controller gain, and acting type, the nonlinear MFA has an extra parameter to enter: the process nonlinearity factor. As shown in Figure 2.15d, the graph on the menu shows how severe the nonlinear behavior is between the process input and process output.

The process linearity factor is a number between 0 and 10. A 10 represents an extremely nonlinear process while a 0 represents a linear process. Notice that the graph shows a nonlinear curve marked with 10 on both upper and lower positions. This means that a nonlinear MFA controller does not care what the nonlinear characteristics are for this process. For instance, the valve can be either "fast open" or "fast close," as represented by these two convex and concave curves.

When using nonlinear MFA, there is no need to worry about how the nonlinear curve is laid out. The curve can be concave, convex, or S-shaped. Simply advise the controller whether the process is extremely nonlinear (enter a 9 or 10), quite nonlinear (enter a 5 or 6), or somewhat nonlinear (enter a 1 or 2). The nonlinear MFA controller will be smart enough to handle the rest.



Process Linearity Factor 5

Apply to +/- 15 % of Range of SP

FIG. 2.15d
Configuration menu of nonlinear MFA.

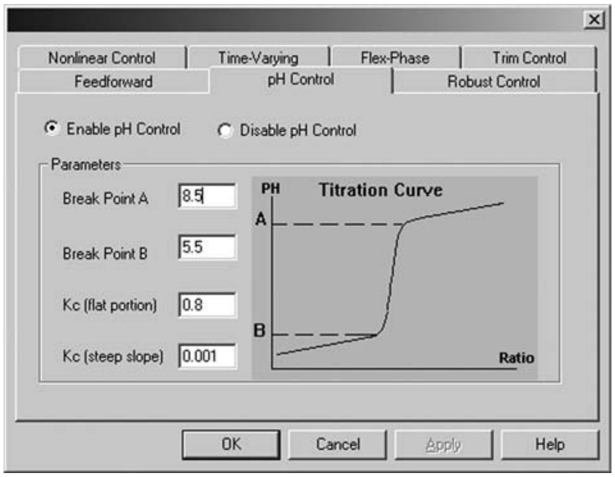
Simulations and real applications show that the nonlinear MFA controller can easily deal with a nonlinear process even if its gain changes hundreds of times. In a nonlinear MFA, there is no linearization calculation or process model. The MFA controller gain K_c is simply set at its nominal point and not retuned.

MFA pH CONTROLLER

Most process plants generate a wastewater effluent that must be neutralized prior to discharge or reuse. Consequently, pH control is needed in just about every process plant, yet a large percentage of pH loops perform poorly. The results are inferior product quality, environmental pollution, and material waste. With ever-increasing pressure to improve plant efficiency and tighter regulations in environmental protection, effective continuous pH control is highly desirable.

A strong-acid-strong-base pH process is highly nonlinear. The pH value vs. the reagent flow has a logarithmic relationship. Away from its neutrality, the process gain is relatively small. Near the neutrality where pH = 7, its process gain can be a few thousand times higher. There is no way for a fixed controller such as PID to effectively control this process. In practice, most pH loops are in a "bang-bang" type of control, with pumps turning on and off, resulting in large oscillations. Since acid and caustic neutralize each other, over-dosing acid and caustic is like continuously burning money. Statistics show that a poorly controlled pH process can cost tens of thousands of dollars in chemical usage each month, not counting the penalties imposed by violating Environmental Protection Agency (EPA) or local government discharge codes.

The MFA pH controller is able to control a wide range of pH loops because its adaptive capability allows it to compensate



● Enable pH Control ○ Disable pH Control

Parameters

Break Point A 8.5

Break Point B 5.5

Kc (flat portion) 0.8

Kc (steep slope) 0.001

FIG. 2.15e
Configuration menu of MFA pH controller.

for the large nonlinear gain changes. In addition, it can control the full pH range with high precision and enables automatic control of acid or alkaline concentration, both of which are critical quality variables for the chemical process industry.

MFA pH Controller Configuration

As shown in Figure 2.15e, one can easily enter Break Points A and B to define the estimated shape of the titration curve of the pH process. Then the MFA controller gain K_c for the flat portion and steep slope can be entered. For a strong-acid-strong-base pH process, if the controller gain for the flat portion is one, then the gain for the steep slope can be estimated as 0.001, which is 1000 times smaller. Due to the adaptive capability of the MFA controller, the titration curve does not have to be accurate and, in fact, its shape can vary in real applications. In addition, the flow rate and the pH value of the inflows may vary significantly. The MFA pH controller can effectively deal with these large disturbances.

The MFA pH controller has helped many users effectively control their tough pH loops. Return-on-investment in weeks or even days has been reported with savings on chemical reagents, no violation of discharge code, and smoother production operation.

FEEDFORWARD MFA CONTROLLER

Feedforward is a control scheme to take advantage of disturbance signals. If a process has a significant measurable disturbance, a feedforward controller can be used to reduce its effect before the feedback loop takes corrective action. A good feedforward controller can improve the control system performance economically.

Feedforward compensation can be as simple as a ratio between two signals. It could also involve complicated energy or material balance calculations. The feedforward MFA is a general-purpose feedforward controller. It does not attempt a perfect cancellation of the disturbances, which is very difficult to implement in industrial applications due to changing process dynamics and operating conditions. A feedback/feedforward MFA control system diagram is illustrated in Figure 2.15f, where G_{p1} is the main process and G_{p2} is the process with disturbance input and the process variable as output.

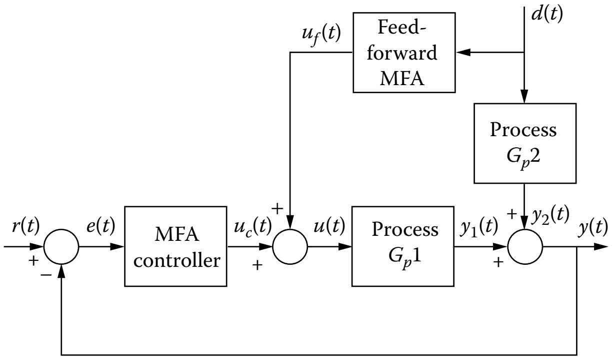


FIG. 2.15f
Feedback and feedforward MFA control system.

Feedforward MFA Controller Configuration

Since the feedback MFA controller has strong adaptive capabilities, the feedforward MFA can be designed in a simple form. There are two parameters to configure, the feedforward controller gain and time constant. The controller gain can be estimated based on the following formula:

$$K_{fc} = -\frac{K_{p2}}{K_{p1}}, \qquad 2.15(8)$$

where K_{p1} and K_{p2} are the estimated static gain for processes G_{p1} and G_{p2}, respectively. The rules for selecting the sign in order to ensure that the feedforward action rejects the disturbance can be summarized as follows:

- If processes G_{p1} and G_{p2} have the same sign, the feedforward gain should be negative.
- If processes G_{p1} and G_{p2} have different signs, the feedforward gain should be positive.

The feedforward MFA time constant can be an estimate of the time constant of G_{p2}. This is related to how fast the disturbance will affect the process variable (PV).

ANTIDELAY MFA CONTROLLER

Many processes have large time delays due to the delay in the transformation of heat, materials, and signals. No matter what control action is taken, its effect is not measurable during a period of time delay. This is equivalent to disabling

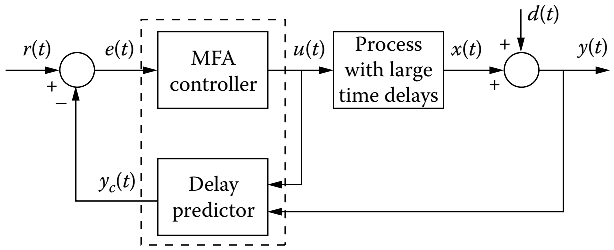


FIG. 2.15g
Antidelay MFA control system.

the feedback for a period of time, where feedback information is essential to automatic control.

If a PID is used to control a process with significant time delays, the controller output will keep growing during the delay time and cause a large overshoot in system responses or even make the system unstable. Typically, a PID has to be detuned significantly in order to stay in automatic but will sacrifice control performance. Generally speaking, a PID controller usually works for the process if its τ-T ratio (delay time/time constant) is smaller than one, unless it is detuned. When a controller is detuned, it loses the sharpness of its control capability, so the process cannot be tightly controlled. The Smith predictor is a useful control scheme to deal with processes with large time delays. However, a precise process model is usually required to construct a Smith predictor. Otherwise, its performance may not be satisfactory.

Figure 2.15g shows a block diagram for a SISO antidelay MFA control system with an antidelay MFA controller and a process with large time delays.

A special delay predictor is designed to produce a dynamic signal $y_c(t)$ to replace the process variable $y(t)$ as the feedback signal. The idea here is to produce an $e(t)$ signal for the controller and let it "feel" its control action without much delay so that it will keep producing proper control signals. In other words, the artificial dynamic signal $y_c(t)$ is able to keep the feedback loop working even when there is a large time delay. Since the MFA controller in the system has adaptive capability, the delay predictor can be designed in a simple form without knowing the quantitative information of the process.[1,2]

Compared to the traditional Smith predictor, the antidelay MFA controller does not need a precise process model. It only needs an estimated delay time as the basic information for its delay predictor. If the delay time used in the MFA delay predictor has a mismatch with the actual process delay time, the controller is robust enough to deal with the difference. Typically, it can deal with the situation where the delay time is two to five times larger or smaller than the actual delay time with satisfactory control performance. In addition, there is no real limitation on how large the τ-T ratio is, as long as an estimated delay time is provided.

The antidelay MFA controller is especially useful in controlling process quality variables since a quality variable is typically measured after the product or process material travels to a certain point, cools off, forms its shape, etc. Antidelay MFA makes it possible for process industries to achieve six sigma or zero defects quality control objectives.

In a semicontinuous production environment, the process line speed may change as many as 100 times or more, which will cause the delay time to change on a similar scale. Since the line speed is measurable, the delay time can be easily calculated and provided to the antidelay MFA controller in real time. In this way, the control performance will not sacrifice much even during large line speed changes.

On the other hand, if the delay time of a process changes on a scale of more than five times, and the delay time information cannot be provided to the controller, the time-varying MFA controller will be more suitable for this application.

ROBUST MFA CONTROLLER

In complex control applications, the following challenges may occur:

1. A large change in the system dynamics occurs, so that a prompt control action is required to meet the control performance criteria.
2. The dominant disturbance to the system cannot be economically measured, and therefore feedforward compensation cannot be easily implemented.
3. A controller purposely detuned to minimize the variations in its manipulated variable may lose control when a large disturbance or significant dynamic behavior change occurs.
4. The system dynamic behavior or load change does not provide triggering information to allow the control system to switch operating modes.

For instance, controlling the reaction temperature for a batch reactor is always a challenge due to the complex nature of the process, large potential disturbances, interactions between key variables, and multiple operating conditions. A large percentage of batch reactors running today cannot keep the reactor temperature in automatic control throughout the entire operating period, thus resulting in lower efficiency, wasted manpower and materials, and inconsistent product quality.

An exothermal batch reactor process typically has four operating stages:

1. Startup stage: ramps up the reactor temperature by use of steam to a predefined reaction temperature.
2. Reaction and holding stage: holds the temperature by use of cooling water while chemical reaction is taking place and heat is being generated.
3. No-reaction and holding stage: holds the temperature by use of steam after the main chemical reaction is complete and heat is not being generated.
4. Ending stage: ramps down the reactor temperature for discharging the products.

During the transition period from Stage 2 to Stage 3, the reactor can change its nature rapidly from a heat-generation process to a heat-consumption process. This change happens without any triggering signal because the chemical reaction can end at any time depending on the types of chemicals, their concentration, the catalyst, and the reaction temperature. Within a very short period of time, the reactor temperature can drop significantly. The control system must react quickly to cut off the cooling water and send in a proper amount of steam to drive the reactor temperature back to normal. A regular feedback controller is not able to automatically control a batch reactor during this transition. In practice, batch reactors are usually switched to manual control and rely on well-trained operators during critical transitions. It is a tedious and nerve-wracking job that can result in low product quality and yield.

The robust MFA controller is able to control the problematic processes described. Without the need to redesign a controller, using feedforward compensation, or retune the controller parameters, the robust MFA controller is able to keep the system in automatic control through normal and extreme operating conditions when there are significant disturbances or system dynamic changes.

Robust MFA Controller Configuration

As shown in Figure 2.15h, the robust MFA controller can be easily configured with these parameters:

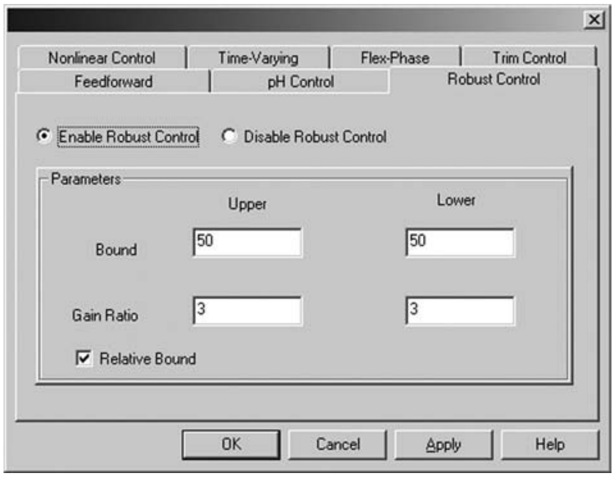
1. *Upper and lower bound*—the bounds for the process variable (PV) being controlled. They provide "intelligent" upper and lower boundaries for the PV. These bounds are typically the marginal values that the PV should not go beyond. PV is unlike the controller output (OP), where a hard limit or constraint can be set. PV is a process variable that can only be controlled by manipulating the OP. Therefore, the upper and lower bounds for PV are very different from the OP constraints.
2. *Gain ratio*—The coefficient to increase or decrease the MFA control action. Typically, you want to enter gain ratio = 3, which implies that the MFA gain working in abnormal situations is three times higher than the regular MFA gain setting. It is important to understand that this is not a gain scheduling approach, although it appears to be this way. Gain scheduling will not be able to resolve the complex problems described.

TIME-VARYING MFA CONTROLLER

The time-varying MFA controller is used to control a process with large time constant and/or delay variations. For instance, a temperature control loop usually has a shorter time constant when it heats up and a much longer time constant when it cools down because adding heat to the process is much faster than taking it away. Also, a line speed or flow rate change will cause the process delay time to vary significantly.

Time-Varying MFA Controller Configuration

As shown in Figure 2.15i, the time-varying MFA controller can be easily configured with an estimated minimum and maximum process time constant plus delay time. The controller is able to deal with the large time constant and/or delay time changes without having to retune any parameters.

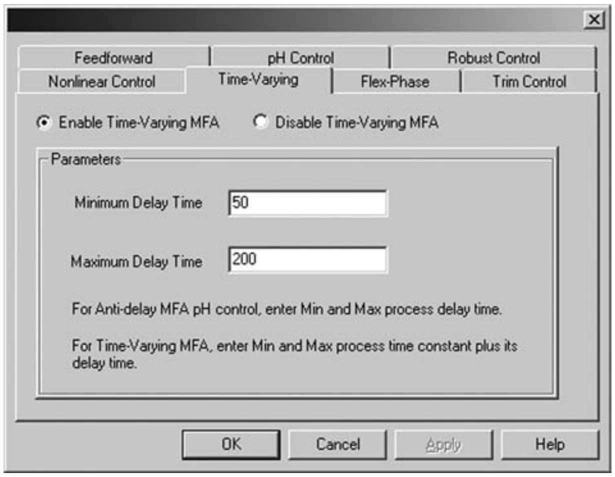


● Enable Robust Control ○ Disable Robust Control

	Upper	Lower
Bound	50	50
Gain Ratio	3	3

☑ Relative Bound

FIG. 2.15h
Robust MFA controller configuration menu.



● Enable Time-Varying MFA ○ Disable Time-Varying MFA

Minimum Delay Time 50

Maximum Delay Time 200

For Anti-delay MFA pH control, enter Min and Max process delay time.

For Time-Varying MFA, enter Min and Max process time constant plus its delay time.

FIG. 2.15i
Time-varying MFA controller configuration menu.

ANTIDELAY MFA pH CONTROLLER

When combining the time-varying MFA and MFA pH control functions, an antidelay MFA pH controller is generated that can control a pH process with large and varying time delays.

When a pH process has large varying time delays as well as large inflow rates and pH changes, the difficulty for this control loop quadruples. The extremely large gain changes with varying time delays make an already bad situation worse, causing the process to become almost "uncontrollable." Traditionally, a "bang-bang" type of control or batch-based pH neutralization would be the only solution.

The antidelay MFA type pH controller has the combined power of being predictive, adaptive, and robust. It is adaptive to compensate for the large gain changes, predictive to deal with large time delays, and robust enough to handle inflow changes, titration curve moves, and other uncertainties.

MULTIVARIABLE MFA CONTROL SYSTEM

Figure 2.15j illustrates a multivariable feedback control system with a model-free adaptive controller. The system includes a multi-input multi-output (MIMO) process, a set of controllers, and a set of signal adders, respectively, for each control loop.

Similar to a SISO system, the MIMO system has controller setpoints **r(t)**, error signals **e(t)**, controller outputs **u(t)**, process variables **y(t)**, and disturbance signals **d(t)**. Since it is a multivariable system, all the signals here are vectors represented in bold type.

TWO-INPUT TWO-OUTPUT MFA CONTROL SYSTEM

Without losing generality, we will show how a multivariable model-free adaptive control system works with a two-input two-output (2 × 2) system as illustrated in Figure 2.15k, which is the 2 × 2 arrangement of Figure 2.15j. In the 2 × 2 MFA control system, the MFA controller set consists of two controllers—C_{11} and C_{22}—and two compensators—C_{21} and C_{12}. The process has four subprocesses—G_{11}, G_{21}, G_{12}, and G_{22}.

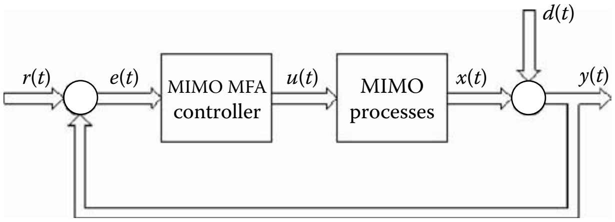


FIG. 2.15j
Multivariable MFA control system.

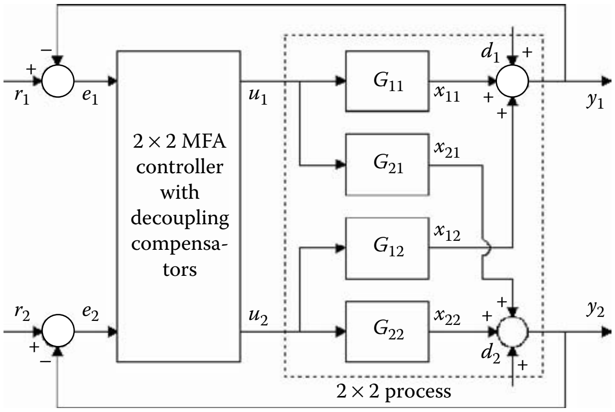


FIG. 2.15k
Two-input two-output MFA control system.

The measured process variables y_1 and y_2 are used as the feedback signals of the main control loops. They are compared with the setpoints r_1 and r_2 to produce errors e_1 and e_2. The output of each controller associated with one of the inputs, e_1 or e_2, is combined with the output of the compensator associated with the other input to produce control signals u_1 and u_2. The output of each subprocess is cross-added to produce measured process variables y_1 and y_2. Notice that in real applications the outputs from the subprocesses are not measurable and only their combined signals y_1 and y_2 can be measured. Thus, by the nature of the 2 × 2 process, the inputs u_1 and u_2 to the process are interconnected with outputs y_1 and y_2. The change in one input will cause both outputs to change.

The control objective for this 2 × 2 MFA control system is to produce control outputs $u_1(t)$ and $u_2(t)$ to force the process variables $y_1(t)$ and $y_2(t)$ to track their setpoints $r_1(t)$ and $r_2(t)$, respectively. The minimization of $e_1(t)$ and $e_2(t)$ is achieved by:

1. The regulatory control capability of the MFA controllers
2. The decoupling capability of the MFA compensators
3. The adjustment of the MFA weighting factors, which allow the controllers to deal with the dynamic changes, large disturbances, and other uncertainties

2 x 2 MFA Controller Configuration

A 2 × 2 MFA controller can be considered to have two main controllers, C_{11} and C_{22}. For each main controller, the parameters to configure are:

1. Sample interval—the interval between two samples or calculations in seconds. A high-speed MFA controller can run at a 1 millisecond rate.
2. Controller gain—use of a default value is recommended.

3. Time constant—a rough estimate of the process time constant in seconds.
4. Acting type—direct or reverse acting of the process.
5. Compensator gain—to deal with the interaction from the other loop.

MIMO MFA CONTROLLER APPLICATION GUIDE

A MIMO system can be much more complex than a SISO system; precautious have to be taken when applying a MIMO MFA controller. When designing a multivariable control system, the first step is to decide which process variable is paired with a manipulated variable. A MIMO MFA control system should follow these pairing rules:

1. Each process of the main loops has to be controllable, open-loop stable, and either reverse or direct acting.
2. A process with a large static gain should be included in the main loop as the main process (G_{11}, G_{22}), and a process with a small static gain should be treated as a subprocess (G_{21}, G_{12}).
3. A faster process should be paired as the main process and a slower process, and processes with time delays should be treated as subprocesses.
4. If pairing rules 2 and 3 should result in a conflict, a tradeoff is the only option.

In addition, an MFA control system should be designed based on the degree of interactions between the loops. Table 2.15l lists the control system design strategy based on the degree of interaction of a MIMO process.

TABLE 2.15l
MIMO System Design Strategy

Interaction Measure	Control Strategy
Small to no interaction	Tighten both loops with SISO MFA
Moderate interaction	Tighten important loops with SISO MFA and detune less important loops or use MIMO MFA for better overall control
Severe interaction	Use MIMO MFA to control the process; may need to de-tune less important loops

MFA CONTROL METHODOLOGY

"All roads lead to Rome." A problem usually has multiple possible solutions, and a process can usually be controlled using different controllers based on different control methods. Almost every control method has its merits and weakness. What is important is to use the right controller to fit the application at a minimum cost.

In natural science, the combination of physics, mathematics, and philosophy plays an integral part in developing a theory that is practically useful. Physics is the foundation for the study of the physical process or environment, mathematics provides the tools to precisely describe the physical process or phenomenon, and equally important is the philosophy that provides directions.

The development of model-free adaptive control technology started from a simple desire to develop a new controller that could easily and effectively solve various industrial control problems. The actual development process has evolved from a prolonged interest in the study of combined intelligence methodology. Since model-free adaptive control does not follow the traditional path of model-based adaptive control, the philosophy behind the combined intelligence has led the way up this long and rocky road.

SUMMARY

To see how the MFA control method is developed based on the combined intelligence methodology, we will relate MFA to each of four key points.

Simple Solution

PID control is simple since it is a general-purpose controller and its algorithm is easy to understand. However, PID is almost too simple to control complex systems. In this regard, PID cannot be considered an effective solution to the more difficult control problems. On the other hand, model-based advanced control methods have proven themselves too complex to launch and maintain since they depend on either a first principle or an identification-based process model. A dream controller has to be powerful enough to control various complex processes yet simple enough to use, launch, and maintain. MFA is a solution that fits these requirements.

Use All Information Available

Model-free adaptive control, as its name suggests, is a control method that does not depend on either first principle or identification-based process models. However, we do try to use all the process information available. For this reason, it can be considered an information-based controller.

For instance, process time constant defines how fast a dynamic system responds to its input. A slow process might have a 10-hour time constant and a fast process might have a 10-millisecond time constant. It would be unwise not to use this information for the controller. In addition, it is relatively easy to estimate the time constant by reading a trend chart. Other important yet easily obtained information about a process includes its acting type (either direct or reverse), static gain, and delay time if any. An MFA controller is designed to use the process parameters that can be easily estimated.

Information's Accuracy

A process can be classified as a white, gray, or black box. If its input–output relationship is clear, the process is a white box. We can easily use existing well-established control methods and tools to design a controller for this process.

When we are not sure if the process input–output relationship is accurate, or if the process has potential disturbances, dynamic changes, and uncertainties, the process is a gray box. In this case, MFA's adaptive capability is able to handle such changes and uncertainties. PID or model-based control methods will have a much tougher time or higher cost addressing these uncertainties.

Technique That Fits the Application

MFA is neither model based nor rule based. We might say that it is an information-based control method. If the argument is made that the process information used is equivalent to a process model, that is perfectly acceptable. The key to this approach is that we focus on delivering a simple, adaptive, and effective solution.

To extend this idea, a series of MFA controllers, many of which are described here, has been developed to address different difficult control problems. Users can simply select the appropriate MFA, configure its parameters, launch the controller, and reap the benefits.

References

1. Cheng, G. S., *MFA in Control with CyboCon,* Rancho Cordova, CA: CyboSoft, General Cybernation Group, Inc., March 2002.
2. VanDoren, V., *Techniques of Adaptive Control,* Burlington, MA: Butterworth-Heinemann, 2003.
3. VanDoren, V., "Model Free Adaptive Control—This New Technique for Adaptive Control Addresses a Variety of Technical Challenges," *Control Engineering Europe,* March 2001.
4. Major, M., "Model-Free Adaptive Control on an Evaporator," *Control,* September 1998.
5. Cheng, G. S., He, M., and Li, D. L., "Model-Free Coking Furnace Adaptive Control," *Hydrocarbon Processing,* December 1999.
6. Seiver, D., and Marin, O., "Air Separation Advances with MFA Control," *Control,* May 2001.
7. Cheng, G. S., and Huo, L. Q., "Control System Optimizes EOR Steam Generator Output," *Oil & Gas Journal,* September 2003.
8. Cheng, G. S., and Huo, L. Q., "Leading Edge Plant Control System—Model-Free Adaptive Control Puts Ling-Yuan Iron and Steel in League of Its Own," *InTech,* October 2003.
9. Cheng, G. S., *CyboCon CE User Manual*, Rancho Cordova, CA: CyboSoft, General Cybernation Group, Inc., 2001.
10. Cheng, G. S., "Tomato Process Improvements Using Model-Free Adaptive Controllers," *Tomato News*, April 2004.
11. Harris, S., "Model-Free Adaptive Control Improves Productivity and Efficiency," *Control*, June 2004.
12. Cheng, G. S., "Model-Free Adaptive (MFA) Control," *IEE Computing and Control Engineering*, June/July 2004.
13. Cheng, G. S., and Zhang, Z. W., "Model-Free Adaptive Control in Water Treatment," *Control Engineering Europe*, September 2004.
14. Cheng, G. S., and Zhang, W. D., "Model-Free Adaptive Technology Improves Distillation Column Chain Control," *Hydrocarbon Processing*, October 2004.

2.16 Modeling and Simulation of Processes

G. K. MCMILLAN (1995, 2005)

Presented below are a variety of simulation and modeling software categories and a partial list of their suppliers.

Steady State to Design Complex Chemical Processes:	ASPEN plus (Aspen Technology, Inc. /www.aspentech.com/) CHEMCAD (Chemstations, Inc. /www.chemstations.net/) Design II (WinSim Inc. /www.winsim.com/) gPROMS (Process Systems Enterprise Ltd. /www.psenterprise.com/) HYSYS (Hyprotech, Ltd. /www.hyprotech.com/) PEGS (CADcentre Ltd. /www.energyweb.net/OSO/) Pro II (SIMSCI-ESSCO division of Invensys /www.simsci.com/) ProChem (OLI Systems, Inc. /www.olisystems.com/) SIMSMART (Simsmart Inc. /www.ahtcopr.com/) VisualModeler (Omega Simulation Co, Ltd. /www.omegasim.co.jp/)
Steady State for Real Time Optimization of Chemical Processes:	DMO-PML and RTOpt (Aspen Technology, Inc. /www.aspentech.com/) HYSYS RTO$^+$ (Hyprotech, Ltd. /www.hyprotech.com/)
Dynamic to Design Complex Chemical Process Control Systems:	ASPEN Dynamics (Aspen Technology, Inc. /www.aspentech.com/) cc-Dynamics (Chemstations, Inc. /www.chemstations.net/) DynaChem (OLI Systems, Inc. /www.olisystems.com/) DYNSIM (SIMSCI-ESSCO division of Invensys /www.simsci.com/) gPROMS (Process Systems Enterprise Ltd. /www.psenterprise.com/) HYSYS Dynamics (Hyprotech, Ltd. /www.hyprotech.com/) VisualModeler (Omega Simulation Co, Ltd. /www.omegasim.co.jp/)
Dynamic to Design Specific Unit Operations in Batch Processes:	cc-Batch (Chemstations, Inc. /www.chemstations.net/) BatchCAD (Hyprotech, Ltd. /www.hyprotech.com/) Batchfrac and Batch plus (Aspen Technology, Inc. /www.aspentech.com/)
Dynamic for Custom Simulation:	acslXtreme (AEgis Technologies Group Inc. /www.acslxtreme.com/) ANSYS Multiphysics (ANSYS Inc. /www.ansys.com/) Custom Modeler (Aspen Technology, Inc. /www.aspentech.com/) MatLab-SimuLink (MathWorks Inc. /www.mathworks.com/)
Real Time for DCS Configuration Checkout and Operator Training Systems:	Autodynamics (Trident Computer Resources Inc. /www.tridentusa.com/) DYNSIM (SIMSCI-ESSCO division of Invensys /www.simsci.com/) HYSYS OTS$^+$ (Hyprotech, Ltd. /www.hyprotech.com/) MIMIC (MYNAH Technologies division of Experitec, Inc. /www.mynah.com/) XANALOG NL-SIM (Xanalog Corporation /www.xanalog.com/)
Dynamic to Optimize and Estimate Parameters for Continuous Systems:	ACSL optimize (AEgis Technologies Group Inc. /www.acslxtreme.com/) SIMUSOLVE (Dow Chemical Co.)
Dynamic to Optimize and Estimate Schedules for Discrete Systems:	GPSS/H (Wolverine Software Corp. /www.wolverinesoftware.com/) SLAM (Advanced Planning & Scheduling division of Frontstep /www.pritsker.com/)

Dynamic Evaluation of Business Decisions on Manufacturing and Inventory Control:
Arena (Rockwell Automation /www.software.rockwell.com/arenasimulation/)
WITNESS (Lanner Group /www.lanner.com/)

Real Time to Tune Controllers:
ExperTune (ExperTune /www.expertune.com/)
ProTuner (Techmation /www.protuner.com/)

Some of the benefits of simulation are listed in Table 2.16a. Simulation has an expanding role in many aspects of industrial production, including research, design, operations, maintenance, and regulatory compliance.

TYPES OF SIMULATIONS

Dynamic simulations are employed to obtain test results more rapidly, less expensively, and in a more controlled environment than through laboratory experimentation. Simulations can model both chemical and biological systems. They can be used to significantly narrow the range and detail of the types of experiments needed in chemical and biochemical research.[1] Custom simulations typically numerically integrate the differential equations that were set by the user.

Steady-State Simulation

Steady-state simulations are extensively used for process design and optimization and provide data for process flow sheets in terms of material and energy balances. They are also used to design process equipment such as heat exchangers, reactors, and distillation columns. These simulations usually consist of blocks of unit operations interconnected by the user and of physical property data for the chemical components of input streams specified by the user.

TABLE 2.16a
Simulation Functions, Types, and Benefits

Function	*Type*	*Benefits*
Experiment design	Dynamic iterative	Better, fewer experiments and faster research
Process design	Steady state	Faster design and better process performance
Equipment design	Steady state	Faster design and better equipment performance
Event optimization	Dynamic discrete	Better utilization, less equipment, more capacity
Controls design	Dynamic continuous	Faster design and better control-loop performance
Controls test	Dynamic real time	Safer and faster startup and greater onstream time
Sequence test	Dynamic real time	Safer and faster startup and greater onstream time
Interlock test	Dynamic real time plus verification	Safer and faster startup and greater onstream time
Operator training	Dynamic real time, custom or generic	Safer and faster startup and greater onstream time

Dynamic Simulations

Dynamic discrete simulations are run to optimize schedules to maximize throughput and minimize equipment requirements for batch processes, inventory management, and parts manufacturing and assembly. These simulations sequence an event list detailed by the user, provide random numbers for stochastic models, and accumulate statistics.

Dynamic continuous simulations enable users to improve control system design strategies to reduce control-loop errors, reduce startup time, and improve on-stream time. These simulations are similar in construction to those used for experimental design except that they provide a library of functional blocks or subroutines to model the more common types of process equipment and instrumentation. Consequently, iteration of the dynamic solution is usually unnecessary.

Real-Time Simulations

Real-time simulations are run to debug configurations and train personnel in the use of distributed control systems (DCS) and programmable logic controllers (PLC). These simulations are essential to smooth commissioning, continuing operation, and maintenance of the systems. They were first extensively used in utility power plants.[2,3] Their benefits are realized not only in improved process and control system performance but also in greater safety.

Operations, maintenance, and control system responses to various scenarios of disturbances or to equipment and instrument failures can be examined quickly and repetitively. Simulation can be used to reduce safety interlock trips through human factors engineering.[4] Simple or small-scale simulations have been configured into various DCS software packages to eliminate the need for an external computer, its associated interface, and wiring or serial communication.[5–8]

Expansion of simulations to develop the detailed response of large complex processes demanded the use of dedicated external computers and either the emulation of the controls or the development of serial communication with them. The increased power of personal computers and workstations has

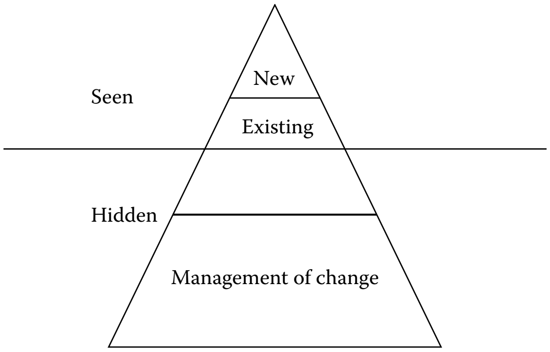


FIG. 2.16b
The iceberg of documentation and training.

facilitated the migration of simulation applications to PCs from general-purpose computers. Whatever computing hardware is used, the benefits are significant. It has been found that it is 20 times faster to correct a program error on a simulator than to correct an error in the field.[9]

New packages are emerging to systematically screen all possible discrete input and output combinations for valid patterns or exceptions and search for common mode failures in safety shutdown systems, as well as to provide dynamic simulation of plant operation. These packages detect both supplier and user software errors to provide a comprehensive verification of software integrity and reliability.

Although software cannot be proven to be bug-free, simulations can go a long way toward reducing software errors to a safe level and ensuring compliance with environmental and safety regulations.

The implications of the management of change dictated by the Occupational Safety Health Act's (OSHA) Process Safety Management Center are greater than realized by most participants in the process industry. The documentation requirements of the initial design are just the tip of the iceberg, as illustrated by Figure 2.16b. The identification and incorporation of changes in existing and new plants require a very large effort that will demand new tools, such as the integration of simulation and computer-aided design with DCS or PLC configurations.

Part of the attractiveness of DCS and PLC is their ability to readily accommodate changes to improve plant operations. Although heaviest during the test and commissioning of the system, the ongoing effort to improve the configuration increases as we proceed with more and more automation. The number of configuration parameters, which can be adjusted manually and automatically, rises exponentially as systems move toward the "Horizon" plant concept. In this system, routine operator actions have been eliminated and the operator's role is elevated to the higher technical role of process optimization.[10] The motivation for the "Horizon" plant is not manpower reduction but a step improvement in product quality, on-stream time, yield, and safety. Simulation is a key prerequisite for such a plant.

Generic Real-Time Simulations Generic real-time simulations are used to train operators for processes such as ammonia production. Packages have been developed for many of the more common hydrocarbon processes. The packages can be modified to include the user's instrument tag numbers, scales, and control system. For those systems that emulate the control system, the user's interface is also mimicked by the package. The emphasis is not on addressing the peculiarities of the plant or control system for an individual application but is more on the familiarization of operators with general process behavior.

These models represent years of process knowledge, simulation development, and improvement. The benefits are significant and have led to such statements as, "Two days on a simulation is worth several years of training in a control room."[11] Such models usually provide more realistic process interactions and simulation control options than do custom-built simulations.[12–14] An effective plant training center uses a variety of generic and project-specific simulations.[15]

STEADY-STATE SIMULATIONS

Steady-state simulations for process design are much more rigorous in the employment of chemistry and chemical engineering principles than are dynamic custom simulations written in software such as ACSL and MatLab, which are based on setting up and numerically integrating differential equations. Most of the newer steady-state simulation software packages now allow the user to graphically configure the model as a process flow diagram (PFD).

A new class of PFD-based dynamic simulation software, such as Aspen Dynamics, DYNSIM, CC-Dynamics, gPROMS, HYSYS Dynamics, and Visual Modeler, has also emerged that has a process fidelity as good as that of the steady-state models used to design processes. However, real-time simulations for configuration checkout and operator training and to optimize business decisions and scheduling generally lack the sophistication of physical property database integration found in the PFD simulators. Some exceptions are HYSYS OTS$^+$ and RTO$^+$, which are extensions of HYSYS Dynamics.

Steady-state simulators, while primarily for process design, can accurately estimate the steady-state process gain, which is important for control-loop analysis and tuning. Multiple runs are made to show changes in the controlled variable in response to changes in the manipulated variable. Since most processes exhibit nonlinearities in their operating points and direction, the simulation runs are made for changes in both directions around the projected set point.

For distillation column control, these runs are used to select the best location of the control tray for temperature control. The selected tray is the one that shows the largest and most symmetrical change in temperature for a change in distillate-to-feed ratio. For pH control, these runs are used to generate a titration curve (a plot of pH versus the ratio of reagent to influent flow). The titration curve is critical for assessment of system difficulty, control valve selection, signal linearization, set point selection, and controller tuning (for further discussion of pH control, see Chapter 8).

For column temperature control, the point with the highest process gain (sensitivity) is sought because the measurement sensitivity is normally low, and the temperature measurement error can translate to a significant composition error. For pH control, the point with the lowest process gain is sought because near neutrality the measurement sensitivity is normally extremely high, and large oscillations in pH can result from imperceptible fluctuations in reagent or influent flow or composition. The process gains are computed by taking the change in temperature and pH at set point and dividing these by the change in distillate and reagent flow, respectively.

Software Packages

The steady-state process flow diagram (PFD) simulators have integrated extensive physical property packages and/or provided links to the American Institute of Chemical Engineering (AIChE) DIPPR, the British PPDS, and the German DECHEMA data compilations. Some also provide the ability to model polymer and electrolyte systems.

Design II is a software package directed toward the hydrocarbon and petrochemical industries. It has essentially no rigorous electrolyte modeling capability to date. It runs conceptually in a batch mode.

Pro-II is widely distributed, has generic capabilities, and has more breadth than but not as much depth as other steady-state simulators. It has a connection with OLI, Inc., a firm specializing in the simulation of electrolyte systems.

ProChem extensively and rigorously models the equilibrium of ionic species in aqueous solutions (electrolyte systems). It is useful for accurately modeling the compositions and pH of waste streams to meet environmental regulations for surface discharges and deep wells. It is not well known that dissociation constants significantly change with temperature and are expressed in terms of activities and not concentrations.

ProChem can compute the change in dissociation constants with temperature and the change in activity from ionic interaction (data that is usually not available). Thus, ProChem can generate titration curves that include the effects of temperature and ionic strength for complex mixtures and are extremely useful for pH control system design.

ProChem is actually a package of programs such as *ElectroChem* (single-stage electrolyte model), *DynaChem* (dynamic model), *FraChem* (multistage electrolyte and reaction model), *ReaChem* (single-stage nonelectrolyte continuous stirred reactor model), *ScaleChem* (scale formation tendency model), *TransChem* (chemical upsets to natural geological formation model), and *DataChem* (data bank manager). Environmental Protection Agency (EPA) and international lists of chemicals of environmental concern and a process analysis section have been added to create an *Environmental Simulation Program* (ESP).

DYNAMIC SIMULATIONS

Dynamic simulations are used to model systems that are in transition to or from a steady state or that are in a state of flux and consequently never reach a steady state. Thus, they are used to investigate biological systems, upsets, failures, and startup/shutdown conditions for continuous processes, batch process conditions and scheduling, discrete manufacturing, fiber-spinning viscoelastic effects, and control system behavior.

The art of dynamic simulation is related to the need for economy. Since custom dynamic simulations numerically integrate and normally require significant original effort, the process detail must be less than that used in steady-state simulations. Similarly, the details of instrumentation and control system dynamics also need to be reduced.

The new class of PFD-based dynamic simulations and the virtual plant have as much or more process detail than do the steady-state simulations used to design the process and can be interfaced to the actual control system configuration. However, these are generally lumped models with perfectly mixed volumes, which means that transportation and mixing delays must be added.

The art is to know, *a priori*, what is and is not important to include. Since loop performance heavily depends on total loop dead time, it is critical to seek out all the major dead-time contributors.[16] The tendency is to concentrate on the dead-time contributions from process equipment and neglect the potentially larger sources of dead time shown in the block diagram for a pH loop in Figure 2.16c. For example, the dead time from reagent piping transportation delay and control valve dead band and stick-slip are significantly larger than the dead time from mixer turnover time for well-designed neutralizers (see the section on pH control in Chapter 8).

Differential equations set up for dynamic simulations are based on the principle of conservation of mass, energy, momentum, spin, and charge (the conversion of mass to energy for nuclear reactions will not be addressed). The net accumulation of these quantities at any point in the process is the integral of the net difference between the input and output, as shown in Figure 2.16d. For steady-state simulations, the output equals the input, so the accumulation and hence the process variable is constant.

There may be multiple inputs and outputs in various forms or phases. These need to be converted to common dimensions and summed. For example, inputs and outputs for a mass balance would be mass flow (e.g., pounds per hour) and would include vaporization, condensing, reaction, precipitation, and crystal growth rates besides stream flows.

Inputs and outputs for an energy balance would consider the energy flow (e.g., BTUs per hour), and the simulation model would include heat transfers, heats of reaction multiplied by reaction rates, heats of dilution and neutralization multiplied by reagent flows, heats of vaporization multiplied by vapor flows, and heats of condensation multiplied by condensing rates, plus the sensible heats of streams multiplied by flow rates. The various heat terms are functions of temperature.

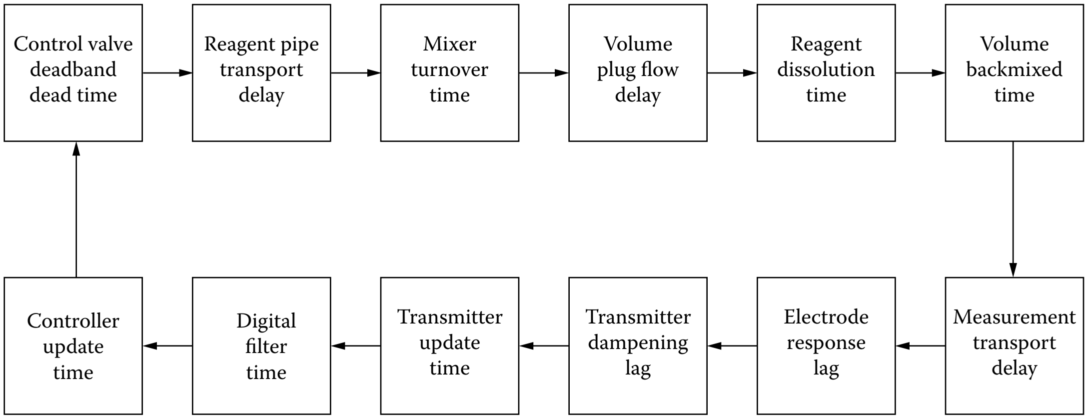


FIG. 2.16c
Block diagram of pH loop dynamics.

Temperature and Vaporization

Temperature processes are simulated by the segmentation of the equipment into volumes, each described by an individual differential equation for an energy balance. The temperature is the energy accumulation divided by the product of the mass and heat capacity of the contents. The smaller the volume, the greater the accuracy. The division into volumes should be greatest where the temperature changes are the largest.

For example, a heat exchanger would be split into volumes for the tube side fluid, the tube wall metal, and the shell side fluid. For counter flow exchangers, either the initial coolant temperature profile must be known or an iterative search should be done for the exit coolant temperature that gives the right inlet coolant temperature. For boiling mixtures, the boiling point as a function of pressure and composition is computed. When the temperature rises to the boiling point, vaporization starts in proportion to the net heat input (BTUs per hour) from streams and heat transfer surface areas divided by heat of vaporization (BTUs per pound).

Vaporization also occurs due to humidification. For scrubbers and textile air washers, the outlet air can be assumed to be saturated. The mole fraction water content of the exiting air can be estimated by dividing the partial pressure of water as a function of temperature by the total pressure of the air.[17] For the inlet air, the same calculation multiplied by the fractional relative humidity gives its moisture content. The mole fractions are converted to weight fractions and multiplied by the inlet and outlet air flows to get the net water loss out of the volume.

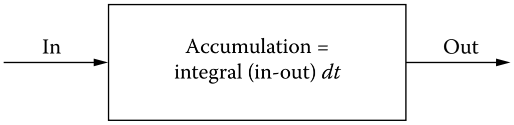


FIG. 2.16d
Accumulation for dynamic simulation.

Pressure and Water Hammer

Gas pressure processes are simulated by segmentation of the system into volumes, each with an individual differential equation for a mass balance. The pressure is computed from the mass accumulation via the ideal gas law and hence depends on the compressibility factor, molecular weight, temperature, and volume of the gas. For vapor spaces above solids and liquids, the gas volume changes with level. For furnace pressures of a few inches of water column, a temperature change can cause the measurement to go off scale, because the interaction between the energy and material balances is significant.

Liquid pressure processes are simulated by the use of pump curves and profiles of pressure drops through the system. The pump discharge pressure is computed from the curve for changes in speed or control valve position. The new pressure drops are proportional to the new flow rate squared, divided by the initial flow rate squared, and that ratio multiplied by the initial pressure drop. The new pressure drop determines a new intersection point of the system drop with the pump curve for variable speed pumps or a new pressure drop for a throttling control valve.

An algebraic loop is created, which is typically broken by the addition of a filter to simulate the gradual response due to impeller inertia, fluid inertia, and/or control valve actuator characteristics. Algebraic loops occur in a model when differential equations are omitted, and they create numerical instability unless a filter is added.

The simulation of water hammer, shock waves (i.e., pressure waves that travel at the speed of sound through the fluid), compressor surge, and the viscoelastic effect of fiber spinning requires the use of a momentum balance. The dynamics are extremely fast and require small integration step sizes. Momentum balances often create stiff systems (i.e., systems with relatively fast and slow dynamics) that require special variable order and step size integration methods to prevent excessively long run times.

pH and Solubility

The simulation of pH requires a component mass balance for each ionic species and for the solvent. The solvent and component concentrations and dissociation constants are substituted into an equation for a charge balance, and interval halving is used to search for the pH value that gives a net charge of zero for the volume. For greater accuracy, activities instead of concentrations should be used, the dissociation constants should be made a function of temperature, and electrode alkalinity and acid error should be included in the measurement response.

Precipitation processes require a solubility calculation based on temperature and concentrations in the liquid phase. Thus, they necessitate energy and component balances. When the solute concentration goes above or below the solubility limit, the precipitation and dissolution rate, respectively, proceed as fast as permitted for numerical stability.

Sensors, Transmitters, and Final Control Elements

The response of a variable speed drive or the stroke of a control valve is simulated by a velocity-limited filter. For the control valve there is also a prestroke dead time associated with the filling and exhausting of air from the actuator, and a dead time caused by dead band. The prestroke dead time is usually negligible, except for large actuators with accessories (e.g., air sets, solenoid valves, current-to-pneumatic transducers, or positioners) with small flow coefficients.

The dead time from dead band is often significant and is calculated by dividing the dead band by the rate of change of controller output. It is simulated by the use of a backlash functional block. Control valves with excessive friction, breakaway torque, or dynamic unbalance can exhibit stick-and-slip action. This is simulated by use of a stroke quantitizer block.

The dead times of chromatographs and digital transmitters and the dead times caused by the scan, cycle, and update times of digital controllers are simulated by delay blocks. The dead times from transportation delays for plug flow, conveyors, pipelines, and sampling lines are also simulated by use of delay blocks, but their delay times are a function of the throughput. A decrease in time delay can be handled bumplessly, but an increase does pose some computational difficulties. Most users pick a conservative fixed delay time.

The damping settings of sensors and transmitters are simulated by the addition of filters. For temperature and pH sensors, the time constant is variable. It depends upon such operating conditions as fluid velocity and process coatings, and the direction and magnitude of the change. Filter time constants can be bumplessly updated.

Simulation Languages

The *Advanced Continuous Simulation Language (ACSL)* will sort the equations in the dynamic section for integration and will identify any algebraic loops. It has many different integration methods and functional blocks. There are no blocks for unit operations, so users must develop their own library of macros and FORTRAN subroutines for process models. Also, the user must construct an extensive steady-state section for a starting point.

REAL-TIME SIMULATION

Real-time simulation connects a simulation of the process to the actual control system. Consequently, the control strategies, proportional–integral–derivative (PID) algorithms, filter time constants, and update, cycle, and scan times do not need to be simulated. It would be difficult to get enough information on proprietary PID algorithms with antireset windup protection and integral tracking options to do a good job of simulating their action. Thus, the use of the actual control system considerably reduces the simulation burden.

The wiring (documentation and installation) of a large simulation system to a control system can take as much time as the construction of the simulation. Most new systems use a high-speed serial communication link from the computer, which directs the simulation and the controller. Here the computer functionally replaces the input and output (I/O) wiring to the field, as shown in Figure 2.16e.

In other words, the same I/O channel assignments are used and the configurations are unaltered so that it is transparent to the controller whether it is connected to the actual plant or to a simulation. This setup as depicted for a combined DCS and PLC system is best from a standpoint of checking the entire control system as it will be used in actual operation.

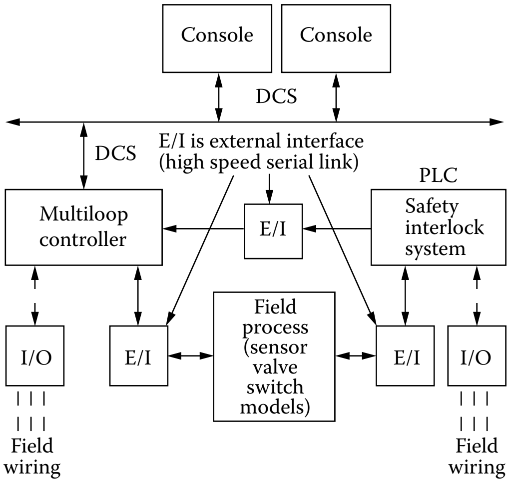


FIG. 2.16e
Real-time simulation system setup.

The use of the DCS data highway for the serial link is not advised because the data flow rate for fast loops will bog down the highway.

When the purpose of real-time simulations is to train operators, the emphasis is more on process rather than on familiarity with the control system. One can then emulate the control system to eliminate the need for communication links and control system hardware.

Real-time integration uses a fixed step size equal to the cycle time of the simulation. Consequently, the step size is too large to simulate fast pressure dynamics or use momentum balances. The gas volumes are increased or calculations similar to those for liquid pressure dynamics are used to achieve numerical stability. Compressor surge is simulated by following the negative slope and jumping over the positive slope portions of the characteristic curve and by the use of a filter to break the algebraic loop between compressor flow and discharge pressure.

The simulation is speeded up or slowed down by the multiplication of the derivative by a gain factor adjustable by the user. The filling and emptying of tanks are often speeded up to test batch operations. Also, the subroutine for integration can be set up to detect a change between the present and last accumulation that is larger than what is possible, and will substitute this as the new desired operating point. This gives the user the ability to make step changes in the operating point. This is particularly useful for batch operations where it is desired to change tank levels to start or retry sequences.

The running of a simulation connected to a control system other than real time will not show the true response of a control loop unless the PID algorithm speed is adjusted and the same ratio is maintained between loop dead time and time constant. This means that all delay times and time constants need to be scaled to simulation speed.

FUTURE TRENDS

The management of change and the desire for increased automation and design efficiency will be the impetus for integration of simulation, computer-aided design (CAD), and DCS or PLC configurations. The goal will be to enter a minimal amount of data (eventually this will be done graphically) and have an integrated system generate the documentation, simulations, control strategy, and interlock configurations required for implementation.

For example, a CAD drawing of the process flow diagram (PFD) with input streams might generate a steady-state simulation of the process that would fill in missing operating conditions and output stream compositions on the PFD. The PFD would in turn generate the skeleton for the piping and instrument diagrams (P&ID). The steady-state simulation would initialize a dynamic simulation whose controlled and manipulated variables would be set by instrumentation added to the P&ID.

The dynamic simulation would consist of two parts, the field and the control strategy. The field data source would be the P&ID, and the control strategy and the safety interlock system data source would be a graphical representation of the DCS or PLC configurations. The dynamic simulation of the strategy and interlocks would migrate to the actual configuration of the DCS or PLC used for real-time control. The dynamic simulation of the field would be designed to prove that the DCS or PLC configuration is wrong rather than to prove that it is right.

Specifically, the control system configuration would not generate the field portion of the dynamic simulation. Toward this end and to reduce common mode failures and to prevent frequent changes to the configuration from affecting the field simulation, the sources of the two parts of the simulations would be kept separate. Similarly, the interlock and control strategy configurations would be separated. These separations would be done by software rather than hardware because the continual increase in reliability and power of computers and the need for fast coordination will push the trend toward integration within a single computer for an operating unit.[18]

Thus, the control strategy and interlock documentation, dynamic simulation, and configuration before and after commissioning would be generated by graphical entry of the configuration. The dynamic simulation would be run faster than real time for control system studies and for accelerated control system testing or operator training. The flow of information for integration is shown in Figure 2.16f.

It is possible that neural networks (Section 2.18) will be used in conjunction with first-principle dynamic models to efficiently predict complex nonlinear effects and that rules and fuzzy logic will be used for more qualitative relationships. Equations from chaos theory may generate data that are not possible to obtain from a deterministic approach.[19] These simulations would run online faster than real time to project future failures and violations of operating

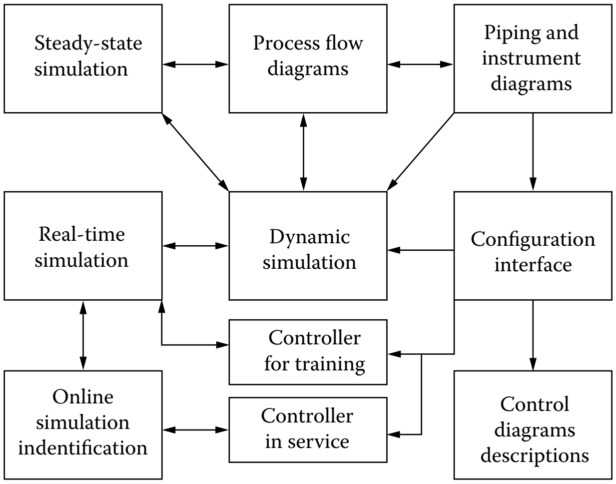


FIG. 2.16f
Future integration.

constraints that will facilitate global online optimization of processes.[20]

VIRTUAL PLANT

Virtual plants are the models and the seamless extension of the actual process and control system design. The dynamic process model must be the steady-state model used to design the process with just volumes and valves added, and the control system must be the actual configuration running in the DCS. The configuration from the DCS can be downloaded into the virtual plant and vice versa. This concept ensures maximum fidelity and eliminates errors from duplication, emulation, and simplification.

The virtual plant becomes a dynamic warehouse of process and control system knowledge. This knowledge can be used for fast prototyping and opportunity assessment in addition to the conventional uses of configuration checkout and operator training.[21] The virtual plant can be operated under conditions that might be considered hazardous for the actual plant. This provides an opportunity to explore new operating regions and methods for model predictive control and real-time optimization.[22]

References

1. Frazer, J. W., Balaban, D. J., and Wang, J. L., "Simulation as an Aid to Experimental Design," *Analytical Chemistry,* Vol. 55, No. 6, May 1983, pp. 904–910.
2. Butler, D. E., and Deaton, I. Y., "From Early Flight Trainers to Modern Utility Power Plant Simulation," ISA Power Division Symposium Proceedings, May 1983, pp. 243–260.
3. "Simulators Expedite Control System Checkout, Tuning, and Design," *Power,* June 1983, pp. 77–79.
4. Chiu, C., "Human-Factors Engineering Key to Trip Reduction," *Power,* July 1988, pp. 33–34.
5. McMillan, G. K., "Real Time Simulations in Distributed Control Systems for Training and Engineering," American Control Conference, June 1985, pp. 1718–1720.
6. Blevins, T., "Operator Training Using a Process Simulator," *Tappi Journal,* August 1985, pp. 128–130.
7. Dybeck, M., and Buzzard, W. S., "Turning Distributed Control Systems into Process Simulators," *Instruments and Control Systems,* September 1986, pp. 75–79.
8. Stuller, P. J., "Power Plant Simulations Using a Distributed Control System," *InTech*, April 1987, pp. 43–45.
9. Pritchard, K., "Applying Simulation to the Control Industry," *Control Engineering*, May 1989, pp. 70–72.
10. Dallimonti, R., "The Horizon Plant," St. Louis Engineer's Club Symposium on "Manufacturing in the 90's," April 5, 1990, pp. 1–9.
11. "Training Plant Operators," *Chemical Week*, September 21, 1983.
12. Clemens, M., "A Training Center for Process Control Engineers Using Simulation," *ISA,* October 1983, pp. 468–477.
13. Chou, A., "Operator Training Simulation—User's Viewpoint," *ISA,* October 1983, pp. 479–485.
14. Vervalin, C. H., "Training by Simulation," *Hydrocarbon Processing*, December 1984, pp. 42–49.
15. Kelly, K. P., "Should You Be Training Your Operators with Simulation Software?" *Control*, June 1992, pp. 40–43.
16. McMillan, G., *Tuning and Control Loop Performance,* 2nd ed., Research Triangle Park, NC: ISA, 1990, pp. 45–55.
17. McMillan, G., and Flores, A., "A Distributed Control System for Enthalpy Control of Air Washers," *ISA, Instrumentation in the Chemical and Petroleum Industries,* Vol. 19, May 1987, pp. 61–68.
18. Morely, R., "Computer Control for the 1990s," *Automation,* June 1991, pp. 28–31.
19. Gleick, J., *Chaos—Making a New Science,* New York: Penguin Books, 1987, pp. 24–30.
20. Morely, R., "Faster Than Real Time," *Manufacturing Systems,* May 1992, p. 58.
21. Mansy, M., McMillan, M., Gregory, K., and Sowell. M. S., "Step into the Virtual Plant," *Chemical Engineering Progress,* February 2002, pp. 56–61.
22. McMillan, G. K., and Cameron, R. A., *Models Unleashed—Virtual Plant and Model Predictive Control Applications,* Research Triangle Park, NC: ISA, 2003, pp. 1–8.

Bibliography

Franks, R., *Modeling and Simulation in Chemical Engineering*, New York: Wiley-Interscience, 1972.

McMillan, G. K., and Cameron, R. A., *Models Unleashed—Virtual Plant and Model Predictive Control Applications,* Research Triangle Park, NC: ISA, 2003.

2.17 Model Predictive Control and Optimization

W. K. WOJSZNIS (2005)

Software Products:

A. HIECON—Hierarchical constraint control
B. PFC—Predictive functional control
C. GLIDE—Identification package
D. DMCplus—Dynamic matrix control
E. DMCplus Model—Identification package
F. Aspen Target—Nonlinear model predictive control (MPC) package
G. Nova—NLC—Nonlinear controller
H. DeltaV Predict and DeltaV PredictPro — MPC controller and optimizer embedded into scalable DCS and integrated with identification and operator control applications
I. RMPCT—Robust model predictive control
J. Connoisseur—Control and identification package
K. Process Perfecter— Nonlinear controller
L. SMOC II—Shell multivariable optimizing control

Partial List of Suppliers:

Adersa — www.adersa.asso.fr/ (A, B, C)
AspenTech — www.aspentech.com/ (D, E, F)
DOT Products — www.dot-products.com/ (G)
Emerson Process Management—www.easydeltav.com/ (H)
Honeywell—www.honeywell.com/ (I)
Invensys—www.invensys.com/ (J)
Pavilion Technologies—www.pavtech.com/ (K)
Shell Global Solutions—www.shellglobalsolutions.com/ (L)

Costs:

The price for MPC usually based on: 1) Controller size, defined by the number of process inputs and outputs a given controller is capable of handling; 2) Number of controllers purchased; 3) Site wide or corporate licensing; 4) Run-time licenses vs. configuration licenses. In some cases the configuration tools are available at little or no cost; run-time licenses are always required; 5) Some DCS include embedded limited-size MPC software with no extra cost.

Model predictive control (MPC) was developed in the 1970s and 1980s to meet control challenges of refineries. Since then it has become the most effective advanced control technique for a wide range of industries. The advantages of MPC are most evident when it is used as a multivariable controller integrated with an optimizer.

The greatest MPC benefits are realized in applications with dead time dominance, interactions, constraints, and the need for optimization. Model predictive control takes into account the effect of past actions of manipulated and disturbance variables on the future profile of controlled and constraint variables and computes a sequence of controller moves to achieve the desired future profile of controlled variables.

The majority of MPC applications assume linear process models and apply linear programming (LP) for optimization. Techniques have been developed to also include nonlinear modeling and optimization.

The refining and petrochemical industries lead in MPC applications. MPC is also widely used in other industries including pulp and paper, chemicals, cement, power, food processing, automotive, metallurgy, and pharmaceuticals.

This section provides the basics of MPC operation and its theoretical background. It outlines also the following steps required for MPC implementation: process analysis and MPC configuration, process testing and model development, MPC simulation, and commissioning.

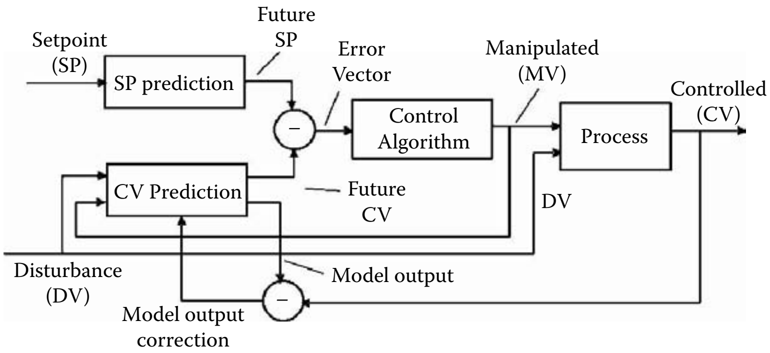


FIG. 2.17a

Model predictive control operation diagram. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society.)

MODEL PREDICTIVE CONTROL PRINCIPLES

One way to understand MPC is to develop an analogy with a traditional feedback control loop (Figure 2.17a). As opposed to a traditional control loop, where the controller applies the difference (error) between the set point and the recent values of the measurement as its input, the predictive controller uses the difference between the future trajectory of the set point and the predicted trajectory of the controlled variable as its input. The difference is expressed not by a single value as in a traditional feedback loop, but as a vector of error values from target for a time period from the present to some set time in the future, usually defined to cover the settling time of the process.

Figure 2.17a shows an MPC controller for a process with two inputs and one output that allows one to see the analogy to a typical feedback control loop. The process has a manipulated variable (MV) and a disturbance variable (DV) on the input and a controlled variable (CV) on the output. A simple MPC controller used in this configuration has three basic components:

- A process model that predicts the process output in the future up to the *prediction horizon* (typically, 120 or more scans)
- A future trajectory of the set point for the same number of scans as the trajectory of the predicted process output
- A control algorithm for computing a control action based on the error vector as the difference between the future trajectories of the set point and the predicted process output

The controller output is an MV that is applied to the inputs of the process and the process model, which is a part of every MPC controller. A measured load upset to the process input is also applied as a DV to the model input. The process model computes a predicted trajectory of the CV that is the process output. After correction of this trajectory for any mismatch between the predicted and an actual measured value of the controlled variable, the predicted trajectory is subtracted from the future trajectory of the set point to form an error vector (Figure 2.17b) applied on the predictive controller input. The predictive control algorithm develops outputs to minimize the sum of squared errors over the prediction horizon, taking into account several future MV moves in its calculations.

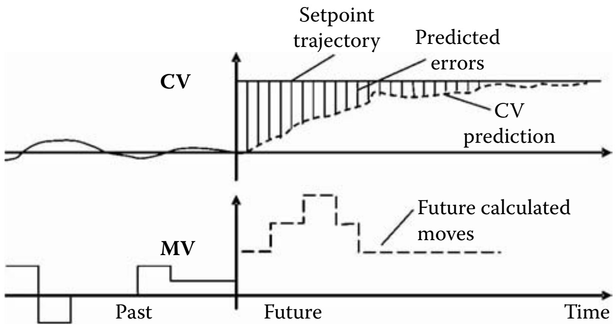


FIG. 2.17b

Illustration of the operation of an MPC controller. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society.)

MPC vs. Feedback Control Summary

The major differences between the operation of an MPC controller and a feedback controller are:

- A predicted error vector is applied to the MPC controller algorithm instead of the scalar values of recent errors that are used in a feedback controller.
- The error vector for an MPC controller is computed as the corrected model prediction subtracted from the future set-point values; whereas the error in a feedback

controller is the measurement subtracted from the set-point value.
- The increments in the MPC controller output required to bring the process output trajectory as close as possible to the set point trajectory are spread over several moves into the future over the *control horizon.* However, only the first move is implemented, and the computation procedure is corrected and repeated for the next scan.
- An MPC controller includes disturbances with the proper dynamics based on identified models from process step responses in the prediction of the process output.

The operation of an MPC controller relies on a good process model. If the model is accurate, the MPC controller can demonstrate superior performance for processes with long dead times, inverse responses, and higher order dynamics that are difficult to control with a classical PID feedback controller. For multivariable processes, the MPC controller takes into account process interactions. Additionally, the MPC controller handles constraints on the process output and the manipulated variables. Finally, MPC controllers are set up to integrate and manage optimization. Taken together, these capabilities make MPC the right choice for incorporating process knowledge and solving complex control and optimization problems on the unit operation control level.

PROCESS MODELING

A generic multivariable process controlled by MPC is presented as a black box in Figure 2.17c. In MPC, the process inputs are MVs and measured DVs. The process outputs are CVs and auxiliary or constraint variables (AVs).

Manipulated variables are managed by MPC controller outputs, predominantly by the manipulation of set points of the fast loops that stabilize the inputs to the process. In calculating the optimal solution, an MPC controller assumes in its calculations that it can make several future moves of the manipulated variables. Only the first move is implemented. The procedure for an optimal solution is repeated every scan. The *control horizon* is the number of future manipulated variable moves that are taken into account in developing the optimal MPC solution.

Manipulated variables have *hard constraints,* which means that no violations of the limits are allowed at any time. Both *absolute* and *rate* or *incremental* constraints are set for manipulated variables.

Measured disturbances are also inputs to the process; however, they are not managed by MPC.

Controlled variables are process outputs kept at specific set points (targets) or within specified ranges. Process models predict the future values of controlled variables for a number of scans ahead. The *prediction horizon* is the range of process output prediction in scans.

Finally, *constraint variables* are a type of controlled variables with only range control and no set points. Constraint variables have so-called *soft constraints*. Violation of this type of constraint may occur temporarily to satisfy other criteria or constraints of the system.

Some use the term *controlled variables* both for *controlled* and *constraint variables*, differentiating them by using the terms *controlled variables with set points* for *controlled variables* and *controlled variables with limits* for *constraint variables.*

The process model is the basis for MPC technology. Most MPC implementations to date use step response models, which provide an explicitly available future prediction of process outputs. The future prediction is used to compute the predicted error vector as an input to the MPC controller.

The step response is represented by a number of coefficients (40 coefficients in our example). Every coefficient corresponds to the value of the model step response at a specific time instance (scan) in the future (Figure 2.17d). In MPC terms, the step response is a prediction of the process output up to the prediction horizon, for a unit step input that was applied at scan zero.

An MPC controller uses an incremental model. It means that real process input and output values are assigned to the model at model initialization. Later on, increments of the

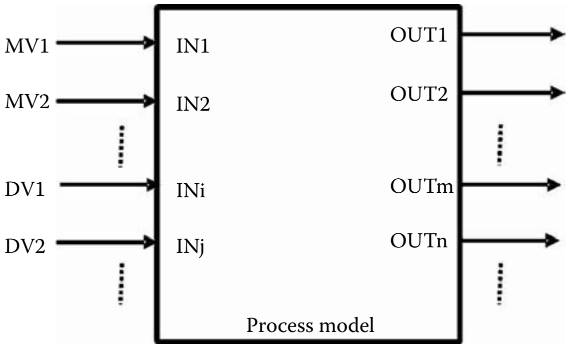


FIG. 2.17c
Multivariable MPC-controlled generic process configuration.

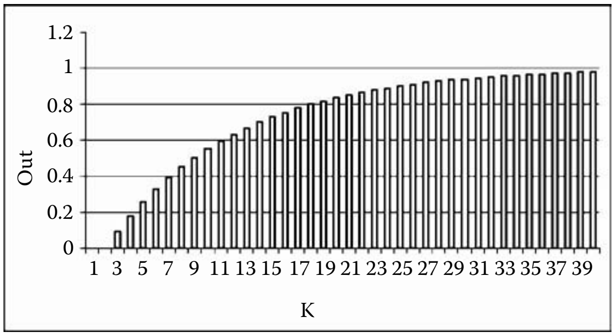


FIG. 2.17d
Step response with 40 coefficients.

inputs are accounted for and increments of the model outputs are calculated.

Process Modeling Equations

MPC application assumes a linear process model. Mathematically we can consider a predicted output trajectory of a process as a process state and use a modified state space form for process modeling. For single-input, single-output (SISO) process prediction, the equations are in the form:

$$\begin{aligned} \mathbf{x}_k &= \mathbf{A}\mathbf{x}_{k-1} + \mathbf{b}\Delta u_k + \mathbf{f}\mathbf{w}_k \\ y_0 &= \mathbf{C}\mathbf{x}_k \end{aligned} \qquad \textbf{2.17(1)}$$

where $\mathbf{x}_{k-1} = [y_0, y_1, \dots y_i, \dots, y_{p-1}]^T$ is the vector of process output prediction at a time $k - 1, 0, 1, 2, \dots p - 1$ steps ahead.

Matrix **A** is the shift operator defined for a self-regulating process as

$$\mathbf{A}\mathbf{x}_{k-1} = [y_1, y_2, \dots y_i, \dots, y_{p-1}, y_{p-1}]^T$$

$\mathbf{b} = [b_0, b_1, \dots, b_i, \dots, b_{p-1}]^T$ is the vector of p step response coefficients.

$\Delta u_k = u_k - u_{k-1}$ is the change in the process input/controller output.

$w_k = y^p - y^m = y^p - y_0$ is the process output measurement minus the model output (the mismatch between the process and the model that results from the noise, unmeasured disturbances, and model inaccuracy).

f is the p dimension filter vector with unity default values.

Matrix **C** is the operator for selecting the current model output defined as $y_0 = \mathbf{C}\mathbf{x}_{k+1}$.

The MPC controller updates prediction and control calculations every scan. This procedure is known as *receding horizon* control.

For n outputs and m inputs process, vector $\mathbf{x}_k$ has dimension $n * p$ and vector **b** becomes a matrix **B** with dimension $n * p$ rows and m columns. The graphical illustration of the equations in Figure 2.17e explains the prediction principles.

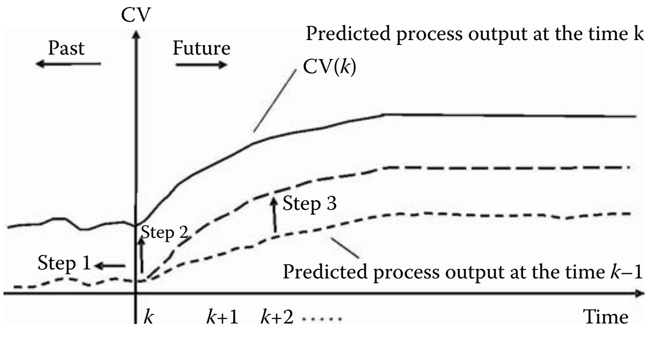


FIG. 2.17e
Illustration of linear process modeling. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society)

At any time instance k, the process output prediction (bottom curve) is updated in three steps:

1. The prediction made at the time $k - 1$ (the bottom dotted curve) is shifted one scan to the left.
2. The prediction curve is moved to the point to match the current measured process output, for filter coefficient = 1, or in general, prediction shift is $\mathbf{f}w_k$.
3. A step response, scaled by the current change on the process input, is added to the output prediction.

PROCESS MODEL IDENTIFICATION

MPC is used mainly for multivariable and highly interactive processes. Processes are tested by applying a special pulse test sequence instead of a single step. Then, a process model is built from process test data using mathematics.

Using test data, we can build equations for every output that equal the number of collected samples. The number of collected samples in test data is normally significantly higher than the number of unknown model coefficients. Such equations are solved using the least squares technique. This technique finds coefficients that may not fit perfectly in any equation but fit optimally for all equations, in such a way that the total squared error for all equations is minimal.

FIR and ARX Modeling

The form of the equations used for process modeling is defined by the identification modeling technique. There are a number of identification techniques used. The most common identification techniques are Finite Impulse Response (FIR) and Auto Regressive with eXternal inputs (ARX). FIR identifies pulse response coefficients, as in Equation 2.17(2) for a SISO process:

$$\Delta y_k = \sum_{i=1}^{p} h_i \Delta u_{k-i} \qquad \textbf{2.17(2)}$$

where p is prediction horizon, with a typical default value for MPC model 120; Δy_k is change in the process output at the time k; Δu_{k-i} is change in the process input at the time $k - i$; and h_i is the pulse response coefficient of the model.

Step response coefficients for the MPC controller are calculated directly from the pulse response as follows:

$$b_k = \sum_{i=1}^{k} h_i \quad k = 1, 2, \dots, p \qquad \textbf{2.17(3)}$$

An advantage of FIR is that it does not require any preliminary knowledge about the process. However, identifying step responses with full prediction horizon with as many as 120 or more coefficients results in low confidence levels of identified coefficient values. Therefore, a shorter horizon with about 60 points is more suitable for a FIR model.

A FIR model with a shorter horizon provides the initial part of the step response that is adequate for defining process dead times using a heuristic approach. On the other hand, the ARX model in Equation 2.17 (4) has fewer coefficients, which are defined with higher confidence, provided the process dead times are known.

$$y_k = \sum_{i=1}^{v} a_i y_{k-i} + \sum_{i=1}^{a} b_i u_{k-d-i} \qquad \text{for SISO process} \qquad \mathbf{2.17(4)}$$

where *a, v* are autoregressive and moving average equation orders of ARX; $a = 4$, $v = 4$ satisfy most applications; a_i, b_i are moving average and autoregressive coefficients of the ARX model; and *d* is dead time in scans.

Applying FIR first and defining dead times and then applying those dead times for ARX provide the best identification results. Step responses for the MPC controller with any prediction horizon are calculated directly from Equation 2.17(4).

For a MIMO process, superposition is applied from all inputs to every output both in FIR and ARX models. An identified model should be validated by applying real input data to the model inputs and comparing the model output with real output. An example of a validation plot is shown in Figure 2.17f.

The validation procedure also may include statistical techniques for calculating *confidence intervals.* Confidence intervals form an area around nominally defined step responses. The true value of the step response coefficients is found within confidence intervals with a predefined confidence (or probability). A 95% confidence interval is normally used.

The complete process model overview is presented in matrix form as in Figure 2.17g.

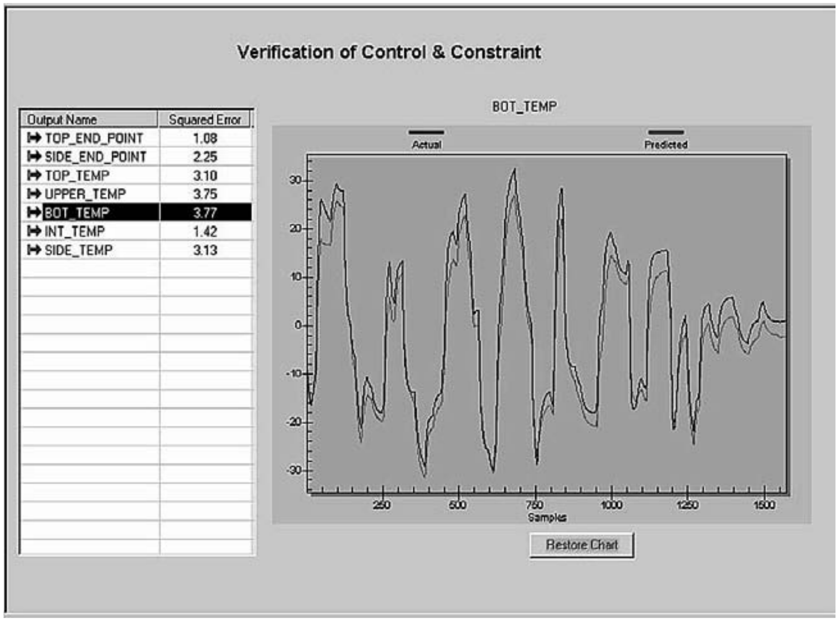
MODEL PREDICTIVE CONTROLLER

Dynamic matrix control (DMC) has historically been the most successful MPC implementation approach, based on a *dynamic matrix*. A dynamic matrix is used for developing an MPC controller. A dynamic matrix is built from step responses to predict the changes in the process outputs that result from moves of the manipulated variables over the control horizon. Dynamic matrix $\mathbf{S}^{\mathrm{u}}$ as in Equation 2.17(5), calculates prediction vector $\Delta\mathbf{x}_{\mathrm{k}}$ resulting from *c* future moves of MV, defined by the vector $\Delta\mathbf{u}_{(k)}$.

$$\Delta\mathbf{x}_k = \mathbf{S}^{\mathbf{u}}\Delta\mathbf{u}(k) = \begin{bmatrix} b_0 & 0 & 0 \cdots & 0 \\ b_1 & b_0 & 0 & 0 \\ b_2 & b_1 & b_0 & 0 \\ \vdots & \vdots & \vdots & \\ b_i & b_{i-1} & b_{i-2} & b_{i-c+1} \\ \vdots & \vdots & \vdots & \vdots \\ b_{p-1} & b_{p-2} & b_{p-3} & b_{p-c} \end{bmatrix} \begin{bmatrix} \Delta u_k \\ \Delta u_{k+1} \\ \Delta u_{k+2} \\ \vdots \\ \Delta u_{k+c-1} \end{bmatrix} = \begin{bmatrix} \Delta y_0 \\ \Delta y_1 \\ \Delta y_2 \\ \vdots \\ \Delta y_i \\ \vdots \\ \Delta y_{p-1} \end{bmatrix} \qquad \mathbf{2.17(5)}$$

MPC Controller Formulation

A general formulation of the MPC controller includes minimization of both the squared sum of predicted control error and the squared sum of calculated controller moves.



Output Name	Squared Error
➡ TOP_END_POINT	1.08
➡ SIDE_END_POINT	2.25
➡ TOP_TEMP	3.10
➡ UPPER_TEMP	3.75
➡ BOT_TEMP	3.77
➡ INT_TEMP	1.42
➡ SIDE_TEMP	3.13

FIG. 2.17f
Validation of identified model. (Courtesy of Emerson Process Management.)

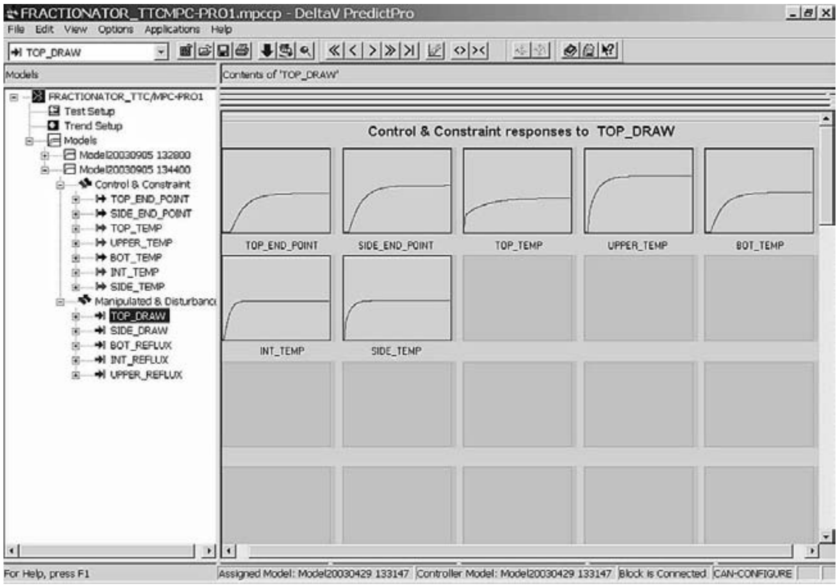


FIG. 2.17g
Step response model overview. (Courtesy of Emerson Process Management.)

To get the desired MPC controller performance, two tuning parameters called *penalty on moves* and *penalty on error* are used in the formulation.

The MPC controller objective for minimizing the squared error of the controlled variable includes the penalty on error over the prediction horizon. The objective for minimizing the squared changes in the controller output over the control horizon includes the penalty on moves in the following way:

$$\min_{\Delta \mathrm{MV}(k)} \{\| \boldsymbol{\Gamma}^{\mathbf{y}}[CV(k) - R(k)] \|^2 + \| \boldsymbol{\Gamma}^{\mathbf{u}} \Delta \mathrm{MV}(k) \|^2\} \qquad \mathbf{2.17(6)}$$

where CV(k) is the controlled output p-step ahead prediction vector; $R(k)$ is the p-step ahead reference trajectory (set point) vector; ΔMV(k) is the c-step ahead incremental control moves vector; $\boldsymbol{\Gamma}^{\mathbf{y}}$ is a diagonal penalty matrix on the controlled output error; $\boldsymbol{\Gamma}^{\mathbf{u}}$ is a diagonal penalty matrix on the control moves; p is the prediction horizon (number of scans); and c is the control horizon (number of scans).

MPC Controller Equations

The solution for the process with dynamic matrix $\mathbf{S}^{\mathbf{u}}$ satisfying Equation 2.17 is in the form:

$$\Delta \mathrm{MV}(k) = (\mathbf{S}^{\mathbf{u}T} \boldsymbol{\Gamma}^{\mathbf{y}T} \boldsymbol{\Gamma}^{\mathbf{y}} \mathbf{S}^{\mathbf{u}} + \boldsymbol{\Gamma}^{\mathbf{u}T} \boldsymbol{\Gamma}^{\mathbf{u}})^{-1} \mathbf{S}^{\mathbf{u}T} \boldsymbol{\Gamma}^{\mathbf{y}T} \boldsymbol{\Gamma}^{\mathbf{y}} E_p(k) \qquad \mathbf{2.17(7)}$$

where $\mathbf{S}^{\mathbf{u}}$ is the $p \times c$ process dynamic matrix built from the step responses of dimension $p \times c$ for a SISO model and $pn \times cm$ for a MIMO model with m manipulated inputs and n controlled outputs; and $E_p(k)$ is the error vector over prediction horizon.

The performance of the control algorithm is modified by the adjustable parameters: p, c, $\boldsymbol{\Gamma}^{\mathbf{u}}$, and $\boldsymbol{\Gamma}^{\mathbf{y}}$. From an implementation point of view, it is inconvenient to use p and c as tuning parameters. $\boldsymbol{\Gamma}^{\mathbf{u}}$ is a basic controller tuning parameter at the controller generation phase. Increasing $\boldsymbol{\Gamma}^{\mathbf{u}}$ elements makes control less aggressive, and while decreasing $\boldsymbol{\Gamma}^{\mathbf{u}}$ makes the control action more aggressive and the control response faster. It follows from experience that the dead time should be accounted for as a major factor in setting the penalty on moves.

The reference trajectory is applied for online tuning. One trajectory design acts in such a way that instead of penalizing any departure from the trajectory, only those deviations that are below the trajectory or above the set point value are penalized (area A and area C if range = 0, Figure 2.17h). In addition, the control error is considered zero if the controlled variable is within range (area C if range > 0).

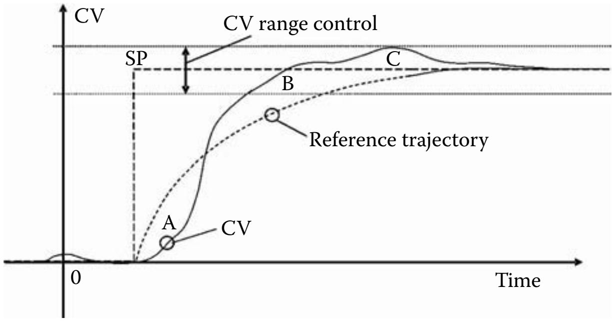


FIG. 2.17h
Reference trajectory for funnel control. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society.)

Presenting MPC control equation in the form

$$\Delta\text{MV}(k) = \mathbf{K}_{\mathbf{mpc}} E_p(k) \qquad \textbf{2.17(8)}$$

where

$$\mathbf{K}_{\mathbf{mpc}} = (\mathbf{S}^{\mathbf{u}T}\boldsymbol{\Gamma}^{\mathbf{y}T}\boldsymbol{\Gamma}^{\mathbf{y}}\mathbf{S}^{\mathbf{u}} + \boldsymbol{\Gamma}^{\mathbf{u}T}\boldsymbol{\Gamma}^{\mathbf{u}})^{-1}\mathbf{S}^{\mathbf{u}T}\boldsymbol{\Gamma}^{\mathbf{y}T}\boldsymbol{\Gamma}^{\mathbf{y}} \qquad \textbf{2.17(9)}$$

is the MPC controller gain, one can see that MPC uses integral action. This, however, does not mean that there is a direct analogy between an MPC controller and a conventional, integral-only feedback controller. The major difference is that the MPC controller uses predicted errors and integrates over the prediction horizon.

INTEGRATING MPC, CONSTRAINTS, AND OPTIMIZATION

A typical MPC configuration includes both controlled variables and constrained variables. The unconstrained MPC controller does not manage constrained variables directly. A supervisory constraint-handling algorithm or optimizer performs the task.

The objectives of optimization are usually to maximize a product value and minimize a raw material cost. For processes with an objective function dependent on several manipulated or controlled variables, optimization techniques are essential components of model predictive control technology. The proven optimization technique is linear programming (LP) with steady-state models.

Linear programming is a mathematical technique for solving a set of linear equations and inequalities in order to maximize or minimize an additional function called an *objective function.* Usually, objective functions express economic value, such as cost or profit.

MPC optimization uses incremental values of MV at the present time or the sum of increments of MV over the control horizon and incremental values of CV at the end of the prediction horizon instead of positional current values, as in typical LP applications. With the prediction horizon normally used in MPC, this assumption guarantees a future steady state for self-regulating processes and perfectly satisfies requirements for proper optimizer operation.

Optimization Equations

The LP technique uses steady-state models and therefore a steady-state condition is required for its application. The basic steady-state process equation in the incremental form is:

$$\Delta\text{CV}(t+p) = \mathbf{A}\Delta\text{MV}(t+c) \qquad \textbf{2.17(10)}$$

where

$$\Delta\text{CV}(t+p) = \begin{bmatrix} \Delta cv_1 \\ \dots \\ \Delta cv_n \end{bmatrix}$$

changes in controlled and constraint variables up to the end of prediction horizon

$$\mathbf{A} = \begin{bmatrix} a_{11}\dots\ a_{1m} \\ \dots \\ a_{n1} \quad a_{nm} \end{bmatrix}$$

process steady-state gains matrix

$$\Delta\text{MV}(t+c) = \begin{bmatrix} \Delta mv_1 \\ \dots \\ \Delta mv_m \end{bmatrix}$$

changes in manipulating variables to achieve the desired state at the end of control horizon.

ΔMV change should satisfy limits both on MV and CV.

$$\text{MV}_{\text{MIN}} \le \text{MV}_{\text{CURRENT}} + \Delta\text{MV}(t+c) \le \text{MV}_{\text{MAX}} \qquad \textbf{2.17(11)}$$

$$\text{CV}_{\text{MIN}} \le \text{CV}_{\text{PREDICTED}} + \Delta\text{CV}(t+p) \le \text{CV}_{\text{MAX}} \qquad \textbf{2.17(12)}$$

The objective function for both maximizing output product value and minimizing input raw material cost is defined in the following way:

$$Q_{\min} = -\text{UCV}^T * \Delta\text{CV}(t+p) + \text{UMV}^T * \Delta\text{MV}(t+c) \qquad \textbf{2.17(13)}$$

where UCV is the cost vector for a CV change of unit value and UMV is the cost vector for an MV change of unit value.

Applying 2.17(10), we can express the objective function through MV only:

$$Q_{\min} = -\text{UCV}^T * \mathbf{A} * \Delta\text{MV}(t+c) + \text{UMV}^T * \Delta\text{MV}(t+c) \qquad \textbf{2.17(14)}$$

The LP solution is always located at one of the vertexes of the region of feasible solutions. An illustration for a two-dimensional problem explains the point (Figure 2.17i).

The region of feasible solutions for two controlled variables and two manipulated variables is an area contained within MV1 and MV2 limits, represented by vertical and horizontal lines, and CV1 and CV2 limits, represented by straight lines (see Equations 2.17[15] and 2.17[16]:

$$\text{CV1}_{\min} = a_{11}\text{MV1} + a_{12}\text{MV2} \qquad \text{CV1}_{\max} = a_{11}\text{MV1} + a_{12}\text{MV} \qquad \textbf{2.17(15)}$$

$$\text{CV2}_{\min} = a_{21}\text{MV1} + a_{22}\text{MV2} \qquad \text{CV2}_{\max} = a_{21}\text{MV1} + a_{22}\text{MV2} \qquad \textbf{2.17(16)}$$

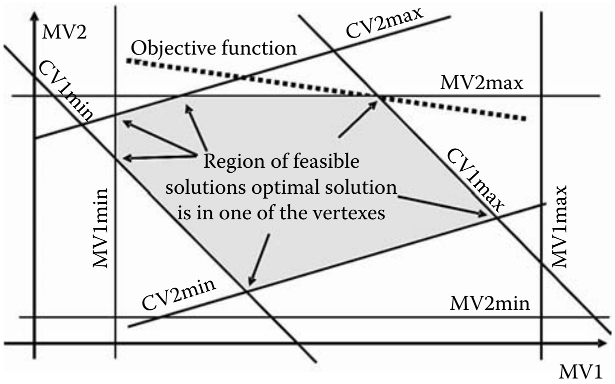


FIG. 2.17i
Optimization problem of a two-dimensional system. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society.)

The optimal solution is located at one of the vertexes marked by arrows. To find this solution, an LP algorithm calculates an objective function for an arbitrary vertex and improves the solution sequentially until the vertex with the maximum (or minimum) value of the objective function as an optimal solution is found.

MPC Optimal Controller

Optimal MV values are applied to the MPC algorithm as target values of MV to be achieved within the control horizon. If the MPC controller is squared, i.e., the number of MV is equal to the number of CV, then MV targets can be effectively achieved by a change in CV value.

$$\Delta \mathrm{CVT} = \mathbf{A} * \Delta \mathrm{MVT} \qquad \mathbf{2.17(17)}$$

where ΔMVT is the optimal target change of MVs and ΔCVT is the CVs target change to achieve the optimal MV. The CVs change is implemented by managing CV set points.

The MPC algorithm working with an optimizer therefore has two main objectives:

- Minimize CV control error with minimal MV moves within operational constraints.
- Achieve optimal steady-state MV values set up by the optimizer and target CV values calculated directly from MV values.

To satisfy these objectives, the original unconstrained MPC algorithm has been extended to include MV targets into the least square solution. The objective function for this MPC controller is

$$\min_{\Delta \mathrm{MV}(k)} \left\{ \|\mathbf{\Gamma^y}[\mathrm{CV}(k) - R(k)]\|^2 + \|\mathbf{\Gamma^u} \Delta \mathrm{MV}(k)\|^2 + \left\| \mathbf{\Gamma^o} \left[\sum \Delta \mathrm{MV}(k) - \Delta \mathrm{MVT} \right] \right\|^2 \right\} \qquad \mathbf{2.17(18)}$$

where $\mathbf{\Gamma^o}$ is a penalty on error of the sum of controller output moves over control horizon relative to the target optimal change of MV defined by the optimizer and $\sum \Delta MV(k)$ is the sum of the MV moves over control horizon. For simplicity of notation, the objective function is shown for the SISO control.

The first two terms are the objective function for the unconstrained MPC controller. The third term sets up an additional condition to make the sum of the controller output moves equal to the optimal targets. In other words, the first two terms set up objectives for controller dynamic operation, and the third term sets up steady-state optimization objectives. Graphically optimized MPC objectives are illustrated in Figure 2.17j.

The general solution for this controller, similar to that for the unconstrained MPC controller, can be expressed as

$$\begin{aligned} \Delta \mathrm{MV}(k) &= (\mathbf{S^{u}}^T \mathbf{\Gamma}^T \mathbf{\Gamma} \mathbf{S^u} + \mathbf{\Gamma^u}^T \mathbf{\Gamma^u})^{-1} \mathbf{S^u}^T \mathbf{\Gamma}^T \mathbf{\Gamma} E_{p+1}(k) \\ &= \mathbf{K}_{\mathrm{ompc}} E_{p+1}(k) \end{aligned} \qquad \mathbf{2.17(19)}$$

where ΔMV(k) is the change in MPC controller output at the time k and $\mathbf{K}_{\mathrm{ompc}}$ is the optimized MPC controller gain.

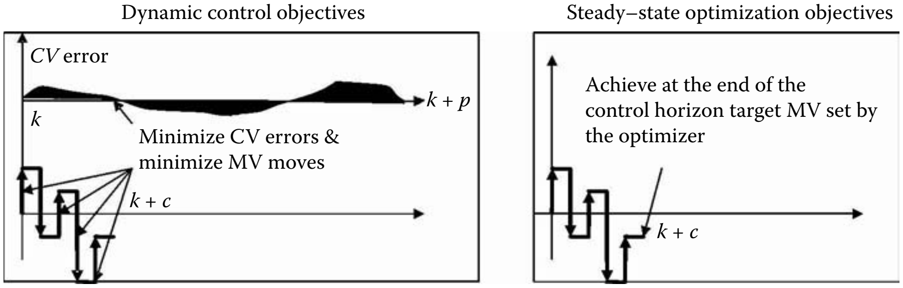


FIG. 2.17j
Optimized MPC control objectives. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society.)

Matrix $\mathbf{\Gamma}$ combines matrices $\mathbf{\Gamma^u}$ and $\mathbf{\Gamma^o}$. It is a square matrix of dimension $(p + 1)$ for the SISO controller and $[n(p + m)]$ for the multivariable controller. Superscript T denotes a transposed matrix.

MPC and Optimizer Integration

In operation, the optimizer sets up and updates the steady-state targets for the MPC unconstrained controller at every scan; thus, the MPC controller executes the unconstrained algorithm. Since the targets are set in a manner that accounts for constraints as long as a feasible solution exists, the controller works within the constraint limits. Optimization, therefore, is an integral part of the MPC controller. The integrated MPC controller and optimizer performs the following sequence of operations each scan:

1. Update CV and DV measurements.
2. Update CV predictions.
3. Determine optimal MV steady-state targets and calculate CV targets.
4. Provide MV outputs accounting for CV and MV targets.
5. Update MV.

Figure 2.17k illustrates communication and integrated operation of the optimizer and MPC algorithm. There is a typical situation when the optimizer cannot find an optimal solution because of too tight constraint limits, too many set points, or too severe disturbances. In these cases, the system constraints are relaxed by changing set points to range control and/or abandoning some constraints. This could be done by minimizing the squared error over a number of constraints or by abandoning constraints with the lowest priority sequentially.

Ill conditioning is another typical problem an optimizer has to deal with. As with the MPC algorithm, ill conditioning manifests itself as excessive changes in calculated MV targets even for minor corrections of constraints. Ill conditioning is removed dynamically by changing the configuration of active constraints (abandoning some constraints) or by removing the association between constraint or controlled variables and manipulated variables that have excessive moves.

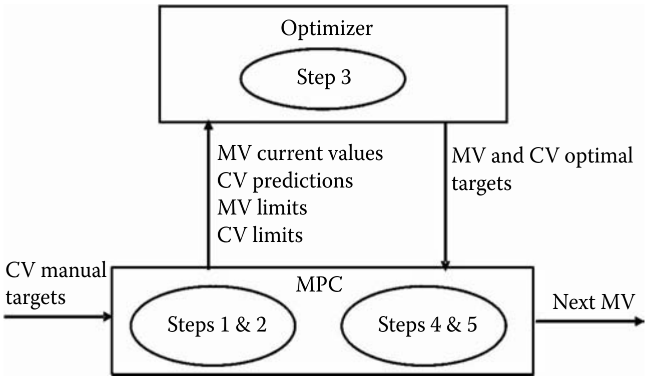


FIG. 2.17k
MPC and optimizer integration. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society.)

MPC APPLICATION DEVELOPMENT

Figure 2.17l illustrates a typical procedure of MPC application development consisting of the following basic steps:

1. Process analysis
2. MPC configuration development
3. Process testing
4. Process model development and controller generation
5. MPC simulation and tuning validation
6. MPC control evaluation and tuning adjustment

Process analysis should deliver a clear formulation of process control objectives and limitations as well as an understanding of how MPC control can achieve those objectives. Process analysis should result in *MPC configuration.* The process inputs and outputs are grouped into four different categories based on how they are utilized in the control of a process.

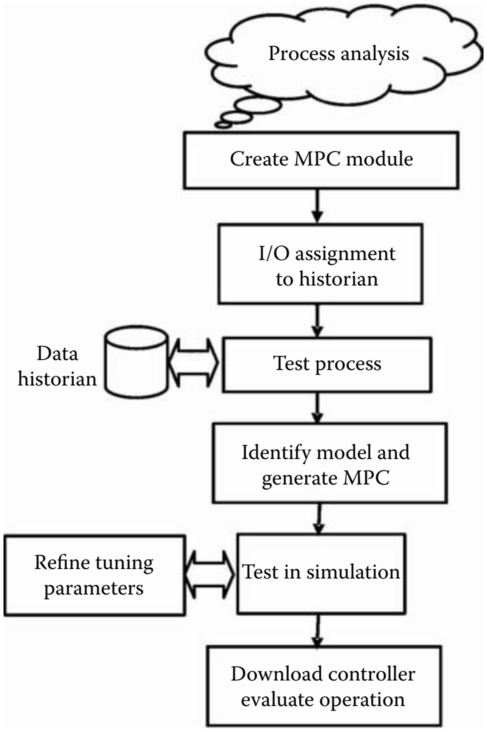


FIG. 2.17l
A typical procedure for MPC application. (Copyright © 2002, ISA—The Instrumentation, Systems, and Automation Society. All rights reserved. Used with permission of ISA—The Instrumentation, Systems, and Automation Society.)

The most common *process testing* procedure used today is pseudo random binary sequence (PRBS).

Process model development consists of test data review and model generation and validation. Controller generation computes the controller matrix defined by the process model and controller tuning parameters: *penalty on move* and *penalty on error.*

Penalty on move—Control sensitivity to changes in a process is determined by the controller robustness. The parameter used in controller generation that most impacts robustness is the penalty-on-move (PM) parameter. The PM defines how much the MPC controller penalizes changes in the manipulated output (MV). The penalty-on-move parameter is typically defined independently for every MV. High PM values result in a slow controller with a wide stability margin. With such settings, the control is relatively insensitive to a change in either the process or model errors. Low PM values result in a fast controller with a narrow stability margin. The penalty on move is also known as "move suppression" or an MV tuning weight.

Penalty on error—The penalty-on-error (PE) factor allows more importance to be placed on a specific controlled variable. In some systems an "equal concern" factor is used to account for differences in scale spans and engineering units, where a smaller equal concern places a greater importance on the error. The penalty on error is also known as a CV or AV tuning weight.

Testing in simulation—The user can influence the control behavior and robustness in online operation. A simulation environment permits operation in manual and automatic mode and allows for the introduction of noise and disturbance inputs and the acceleration of control execution.

Commissioning MPC Application

Before the new control is tried on the real process, the user must observe how well the control responds to set-point changes and load disturbances in simulation. If the process responds slowly to input changes, then it may take many hours in an operating plant to see the full control response. In a simulated environment, the process and control simulation can support faster than real-time execution to allow the response of even the slowest processes to be seen in a few minutes. Similarly, the ability to simulate the control and process response slower than real time is valuable when working with a very fast process. The ability to adjust the speed of execution allows the control response to be verified quickly for a variety of operating conditions. Through such testing, the user can gain confidence in the control performance.

In the *commissioning phase* of the MPC application, the user must first perform the obvious step of checking all communications between the process instrumentation and the MPC controller. Once the controller is operating in a stable manner, the performance elements that must be checked out during the commissioning phase include:

- Set-point changes
- Constraint control
- Steady-state giveaway
- Manipulated variable limitations
- Feedforward disturbances handling
- Optimizer functionality

Many issues of MPC application development have been addressed in modern MPC design. To provide a consistent environment for a configuration definition, MPC is implemented as a function block, and a simulation environment is automatically created based on the block definition, the identified process model, and the controller generated.

Once the controller has been generated, the MPC block is downloaded to the DCS controller to place the control online. Predefined applications are provided to allow each control, constraint, manipulated, or disturbance input to be placed on operator graphics. The trend display shows past plots and future predicted process outputs.

The tuning parameters associated with the generation of the MPC controller are automatically set based on the process dynamics. Thus, in most cases there is no need for users to make any adjustment to these tuning parameters.

CONCLUSIONS

Model predictive control and optimization are the primary techniques for achieving high performance unit operations. Modern MPC products, especially those integrated with DCS, are easy to apply and use. Good process understanding, however, is a major factor in setting the control objectives and in designing and commissioning an MPC application.

Bibliography

Badgwell, T. A., and Qin, S. J., "Review of Nonlinear Model Predictive Control Applications," in *Nonlinear Predictive Control: Theory and Practice,* B. Kouvaritakis, and M. Cannon, Eds., London: IEE Publishing, pp. 3–32, 2001.

Blevins, T. L., McMillan, G. K., Wojsznis, W. K., and Brown, M. W., *Advanced Control Unleashed,* Research Triangle Park, NC: ISA, 2002.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C., *Time Series Analysis, Forecasting and Control,* 3rd ed., Englewood Cliffs, NJ: Prentice Hall, 1994.

Camacho, E. F., and Bordens, C., *Model Predictive Control in the Process Industry,* London: Springer-Verlag, 1995.

Cutler, C. R., and Ramaker, D. L., "Dynamic Matrix Control—A Computer Control Algorithm," Proceedings JACC, San Francisco, 1980.

Froisy, J. B., "Model Predictive Control: Past, Present and Future," *ISA Transactions,* Vol. 33, pp. 235–243, 1994.

Garcia, C. E., Prett, D. M., and Morari, M., "Model Predictive Control: Theory and Practice—A Survey," *Automatica,* Vol. 25(3), pp. 335–348, 1989.

Lee, J. H., Morari, M., and Garcia, C. E., "State-Space Interpretation of Model Predictive Control," *Automatica,* Vol. 30(4), pp. 707–717, 1994.

Maciejowski, J. M., *Predictive Control with Constraints,* London: Pearson Education Limited, 2002.

McMillan, G. K., *Process Industrial Instruments and Control Handbook*, New York: McGraw-Hill, 1999, pp. 10.190–220.

McMillan, G., “A New Era for Model Predictive Control,” *Control,* Vol. 13(5), pp. 45–50, 2000.

Morari, M., and Ricker, N. L., *Model Predictive Control Toolbox,* Natwick, MA: The MathWorks, 1995.

Muskie, K. R. and Rowlings, J. B., “Model Predictive Control with Linear Models,” *AIChE Journal,* 39(2), pp. 262–287, 1993.

Prett, D. M., and Garcia, C., *Fundamental Process Control,* Boston: Butterworth, 1988.

Qin, S. J., and Badgwell, T. A., “An Overview of Industrial Model Predictive Control Technology,” in *Chemical Process Control—V,* pp. 232–256, Fifth International Conference on Chemical Process Control, AIChE and CACHE, 1997.

Rossiter, J. A., *Model-Based Predictive Control,* Boca Raton, FL: CRC Press, 2003.

Shinsky, F. G., *Process Control Systems: Application, Design, and Adjustment,* 3rd ed., New York: McGraw-Hill, 1988.

Wojsznis, K. W., and Wojsznis, P. W., “Robust Predictive Controller in Object-Oriented Implementation,” *ISA Transactions,* Vol. 32, pp. 339–344, 1993.

Wojsznis, K. W., Blevins, L. T., and Nixon, M., “Easy Robust Optimal Predictive Controller,” *Advances in Instrumentation and Control,* ISA TECH 2000, New Orleans.

Wojsznis, W., Gudaz, J., Mehta, A., and Blevins, T., “Practical Approach to Tuning MPC,” *ISA Transactions* Vol. 42(1), pp. 149–162, January 2003.

Abbreviations

ARX Auto regressive with eXternal inputs (model)
AV Auxiliary (constraint) variable
CV Controlled variable
DCS Distributed control system
DMC Dynamic matrix control
DV Disturbance variable
FIR Finite impulse response (model)
LP Linear programming
MIMO Multiple-input multiple-output
MPC Model predictive control
MV Manipulated variable
PE Penalty on error
PM Penalty on moves
PID Proportional, integral, derivative (control)
PRBS Pseudo random binary sequence

Definitions

Control horizon The number of future manipulated variable moves that are taken into account in developing the optimal MPC solution.

Dynamic matrix A matrix built from step responses to predict the changes in the process output that results from the moves of the manipulated variable over the control horizon.

Linear programming A mathematical technique for solving a set of linear equations and inequalities in order to maximize or minimize an additional function called an objective function.

Model predictive control (MPC) A model-based control technique that uses process output prediction and calculates several consecutive controller moves in order to satisfy control objectives.

Moving horizon control See **Receding horizon control**.

Penalty on error A parameter that ranks the importance of a specific controlled variable.

Penalty on move A parameter that defines how much the MPC controller penalizes changes in the manipulated variable.

Prediction horizon The range of process output prediction defined in scans.

Receding horizon control Control that is based on updating every scan output prediction over the prediction horizon.

Reference trajectory The future desired projection of the process output.

Set point trajectory See **Reference trajectory**.

2.18 Neural Networks for Process Modeling

B. A. JENSEN (1995) **J. ABONYI** (2005)

The design of control and process monitoring systems is currently driven by a large number of requirements posed by energy and material costs and by the demand for robust, fault-tolerant systems. These considerations introduce extra needs for effective process modeling techniques. Many systems are not amenable to conventional modeling approaches due to the lack of precise, formal knowledge about the system, due to strongly nonlinear behavior, high degree of uncertainty, or time-varying characteristics. Computational intelligence, the technical umbrella of artificial neural networks (ANNs), has been recognized as a powerful tool that is tolerant of imprecision and uncertainty and can facilitate the effective development of models by combining information from different sources, such as first-principle models, heuristics, and data.

Among the techniques of computational intelligence, ANNs attempt to mimic the structures and processes of biological neural systems. They provide powerful analysis properties such as complex processing of large input/output information arrays, representing complicated nonlinear associations among data, and the ability to generalize or form concepts/theory.

This section focuses on the promise of ANNs in the realm of modeling, identification, and control of nonlinear processes. The basic ideas and techniques of ANNs are presented in language and notation familiar to instrument engineers, and several examples are given of their industrial applications.

NEURAL NETWORKS FOR BLACK-BOX MODELING

Information for process modeling and identification can be obtained from different sources. According to the type of available information, three basic levels of model synthesis are defined:

- *White-box or first-principle modeling.* A complete mechanistic model is constructed from *a priori* knowledge and physical insight. Here, the dynamic models are derived based on mass, energy, and momentum balances of the process.
- *Fuzzy logic modeling.* A linguistically interpretable rule-based model is formed, which is based on the available expert knowledge and measured data.
- *Black-box model or empirical model.* No physical insight is available or used, but the chosen model structure belongs to families that are known to have good flexibility and have been “successful in the past.” The parameters of the models are identified based on measurement data.

This means that there may be situations where the most valuable information comes from input/output data collected during operation. In this case, the application of black-box models is the best choice. These black-box models are especially valuable when an accurate model of the process dynamics is needed. Therefore, nonlinear black-box modeling of process systems is a challenging and promising research field. The area is quite diverse and covers topics from mathematical approximation theory, via estimation theory and nonparametric regression, to algorithms and currently much-discussed concepts such as wavelets, fuzzy models and artificial neural networks (ANNs).

ANNs can learn complex functional relations by generalizing from a limited amount of training data; hence, they can serve as black-box models of nonlinear, multivariable static and dynamic systems and can be trained by the input/output data of these systems. Applications that are suitable for ANN solutions include those where:

- The solution is derived from highly interdependent variables that have no precise quantification.
- The problem requires qualitative or complex quantitative reasoning.
- Project development time is short, but sufficient data and neural network training time is available.
- Data are readily available but they are multivariate and intrinsically noisy or error-prone.

Conversely, applications that are not suited for neural net solutions include:

- Mathematically accurate and precise applications (i.e., accounts receivable and inventory)
- Applications that require deduction and stepwise logic

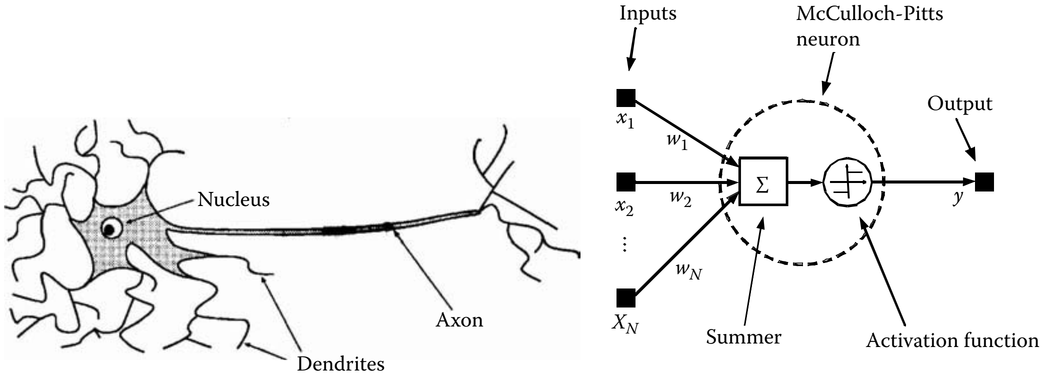


FIG. 2.18a
A biological neuron and its model (McCulloch–Pitts neuron).

STRUCTURE OF NEURAL NETWORKS

McCulloch–Pitts Neuron

It has long been known that learning in animals and humans can be achieved through observation of examples. The exact mechanism by which this learning takes place is still unknown, but science has yielded some clues.

The vertebrate brain consists of an enormous number of interconnected cells called neurons. It has become widely accepted that these neurons are the fundamental information processing elements of the brain. Neurons respond to electrical impulses collected from other neurons through connecting fibers called axons and dendrites. Prompted by studies in neuroscience, McCulloch and Pitts in 1943 developed a simple mathematical model for a neuron[1] (Figure 2.18a).

The McCulloch–Pitts neuron has multiple inputs and a single output. Each of the inputs has an associated weight. The weighed sum of the inputs is passed through a nonlinearity to the output of the neuron as follows:

$$y = f\left(\sum_{i=1}^{N} w_i x_i\right), \qquad f(z) = \begin{cases} 0 & z < 0 \\ 1 & z \geq 0 \end{cases} \qquad \mathbf{2.18(1)}$$

where x_i are the inputs, y is the output of the neuron, $f(z)$ is the nonlinear activation function in the form of the step function given above, and w_i are the strengths of the connections or weights. Given this particular neural model, it is important to understand what sorts of problems it can solve.

It can be used for pattern classification problems where input patterns have to be separated into two classes (Figure 2.18b). It is interesting that this simple model of a biological neuron is capable of emulating such a fundamental and important feature of natural intelligence. Although it demonstrates an example of how a simple element can perform a complex task, without a method of setting the connection weights and handling nonlinear separation problems it can be of little practical use. Hence, there was a period lasting from the late 1960s until the mid-eighties when interest of neural networks waned. It was not until 1986 that the vitality of the ANN field was restored by the publication by Rumelhart and McClelland.[2] This publication showed that the bane of neurons (perceptrons), the problem of linear inseparability, could be solved if perceptrons were bunched together in multiple layers to make multi-layer neural networks.

Feedforward Multi-Layer Neural Networks

The developed feedforward multi-layer neural network has one input layer, one output layer, and a number of hidden layers between them. For illustration purposes, consider an ANN with one hidden layer (Figure 2.18c).

The input-layer neurons do not perform any computations; they merely distribute the inputs x_k to the weights w_{jk}

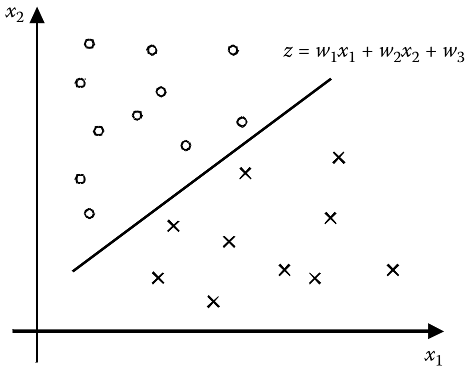


FIG. 2.18b
Graph showing how a line such as that defined by the weights in a McCulloch–Pitts neuron can be used to separate two classes of patterns.

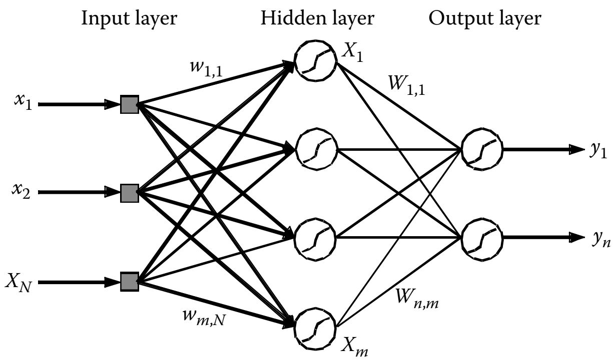


FIG. 2.18c
A multi-layer neural network with one hidden layer of neurons.

of the hidden layer. In the neurons of the hidden layer, first the weighted sum of the inputs is computed,

$$h_j = \sum_k w_{jk} x_k \qquad \mathbf{2.18(2)}$$

A nonlinear transfer function (also known as activation function and squashing function) is applied to the result to calculate each processing element's output,

$$X_j = f\left(\sum_k w_{jk} x_k\right) \qquad \mathbf{2.18(3)}$$

The transfer function adds nonlinearity and stability to the network. The most common differentiable transfer functions are listed in Table 2.18d.

TABLE 2.18d
Common Differentiable Transfer Functions

Name	*f(z)*
Sigmoid	$1/(1 + e^{-z})$
Hyperbolic tangent (tanh)	$(1 - e^{-z})/(1 + e^{-z})$
Linear	z
Sine	$\sin(z)$
Signum	1 if $z > 0$ -1 if $z < 0$
Percepton	z if $z > 0$ 0 if $z < 0$

The output value of the transfer function is generally passed directly to the output path of the processing element (although there are exceptions to this when a dynamical filter is used). The output path is then connected to input paths of other processing elements through connection weights. The outputs from the hidden units can then go through another layer of filters,

$$H_i = \sum_j W_{ij} X_j = \sum_j W_{ij} f\left(\sum_k w_{jk} x_k\right) \qquad \mathbf{2.18(4)}$$

and be fed through another layer of squashing functions to finally produce the outputs.

$$Y_i = f(H_i) = f\left(\sum_j W_{ij} f\left(\sum_k w_{jk} x_k\right)\right) \qquad \mathbf{2.18(5)}$$

If the network is being used for regression variables instead of classifying features, linear output functions should be used instead.

Contrary to the simplicity of this structure, this model is very general. It has been shown that with one hidden layer a network can describe any continuous function (if there are enough hidden units), and that with two hidden layers it can describe any function at all.[3] For local nonlinear basis functions, such as Gaussian RBFs, a single layer suffices.[4]

Example of an ANN

To illustrate the structure of ANNs, consider a fractionator example, where the product specifications are based on the Reid vapor pressure in the bottoms product and on the 95% boiling point of the distillate. Process analyzers have inherent dead time in producing these measurements, or the availability of analyzers to provide these measurements is limited.

Process models can be used to predict these properties, or other measurements can be used to infer them. However, if data exist regarding the measurements around the fractionator based on collected laboratory or analyzer results, as commonly recorded on log sheets, a nonlinear neural network model can be built. Shown in Figure 2.18e is a model with nine input nodes, four hidden nodes, and two output nodes with bias.

TRAINING OF FEEDFORWARD NEURAL NETS

Previously it was shown that neural networks consist of heavily interconnected processing elements that do not store information. The knowledge is stored by the way the processing elements are connected and by the importance of each connection, known as its weight. Training of the network means the adaptation of these weights such that the error between the desired output and the network output is minimized. Two steps are distinguished in this training procedure:

Feedforward Computation

From the network inputs $\mathbf{x}_n$, the outputs of the first hidden layer are first computed (by Equation 2.18[2]). Then using these values as inputs to the second hidden layer, the outputs

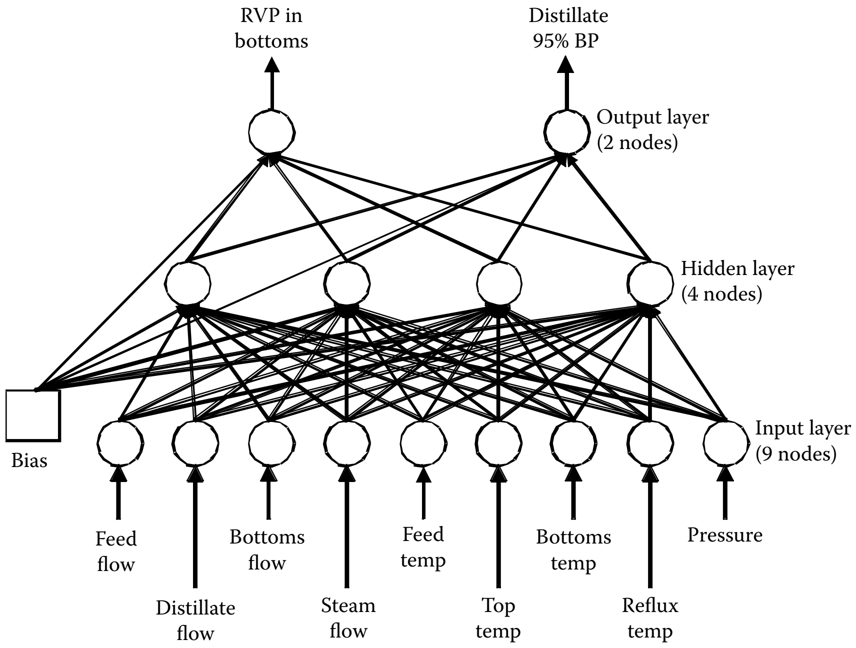


FIG. 2.18e
Example of a three-layer fully connected neural network describing the application of physical property prediction.

of this layer are computed (by Equation 2.18[3]), etc. Finally, the output of the network is obtained, $Y_i(\mathbf{x}_n)$ (Equation 2.18[5]).

Weight Adaptation

The output of the network is compared to the desired output, $y_i(\mathbf{x}_n)$. The difference between these two values, the error, is then used to adjust the weights first in the output layer, then in the layer before, etc., in order to decrease the error:

$$\begin{aligned}\min_{w,W} E^2 &= \frac{1}{2}\sum_n\sum_i [y_i(\mathbf{x}_n) - Y_i(\mathbf{x}_n)]^2 \\ &= \frac{1}{2}\sum_n\sum_i \left[y_i(\mathbf{x}_n) - f\left(\sum_j W_{ij} f\left(\sum_k w_{jk} x_{k,n}\right)\right)\right]^2\end{aligned} \qquad \mathbf{2.18(6)}$$

According to gradient-descent optimization, the updates step in the output weights can be found by differentiating the cost function given by Equation 2.18(6):

$$\begin{aligned}\Delta W_{ij} &= -\eta \frac{\partial E^2}{\partial W_{ij}} \\ &= \eta \sum_n \left[y_i(\mathbf{x}_n) - Y_i(\mathbf{x}_n)\right]^2 f'(H_i) X_j \equiv \eta \sum_n \Delta_i X_j\end{aligned} \qquad \mathbf{2.18(7)}$$

with the definition $\Delta_i = \eta[y_i(\mathbf{x}_n) - Y_i(\mathbf{x}_n)]^2 f'(H_i)$

The update in the input weights can be found from the chain rule:

$$\begin{aligned}\Delta w_{jk} &= -\eta \frac{\partial E^2}{\partial w_{jk}} \\ &= \eta \sum_n\sum_i [y_i(\mathbf{x}_n) - Y_i(\mathbf{x}_n)]^2 f'(H_i) W_{ij} f'(h_i) x_{n,k} \\ &= \eta \sum_n\sum_i \Delta_i W_{ij} f'(h_i) x_n \equiv \eta \sum_n \delta_j x_n\end{aligned} \qquad \mathbf{2.18(8)}$$

defining $\delta_j = \eta f'(h_i)\sum_i W_{ij}\Delta_i$.

The deltas for the input layer are found in terms of the deltas for the output layer by running them backwards through the network W_{ij}s. This is straightforward to generalize to networks with more than one hidden layer. Training a network by gradient descent, feeding the errors backwards through the network in this way, is called back propagation.

DEVELOPING AND BUILDING NEURAL NETWORKS

The phases of a neural network life cycle are concept, design, implementation, and maintenance.[5,6] In all phases, a strong emphasis is placed on experimentation. The user experiments under multiple development environments in order to refine, redesign, and reformulate network parameters to achieve the best outcome.

Concept

The concept phase plans the approach to building the application. The first step in implementing a neural network is to select a neural paradigm. Selection is based upon the traits of the problem (e.g., prediction, classification, association, conceptualization, filtering), size of the problem (number of input and output variables, size of the identification data), method of training (supervised/unsupervised), and time constraints.

Design, Implementation and Maintenance

The *design phase* specifies initial parameters and conditions for the selected neural paradigm. The type and number of processing elements (neurons), the number of layers, the connectivity of the layers, the transfer function, and the learning algorithm are design decisions that must be made.

Implementing the neural network involves applying the design criteria using customized or commercial environments. Experimentation is carried out with learning coefficients, momentum factors, noise injection, etc., in order to train the network most efficiently. Because of the nature of some learning algorithms, the method of presenting the training data can also affect the training results.

Maintenance involves modifying the deployed neural network application because of decreasing accuracy due to conditions encountered for which the network was not trained. In such cases, the training set must either be modified or replaced. Next, the network must undergo retraining and reevaluation in an iteration of the implementation phase. If reevaluation shows that the network performance is not to specification with the new data set, the design phase must be repeated.

In the following paragraphs, the most critical issues of this whole procedure are presented from the viewpoint of the identification data and the design parameters of ANNs.

Data Collection and Preparation

For a successful application of ANNs, as much training data as possible should be collected. Even more than volume, the quality and representativeness is important.[7] A good training set will contain routine, unusual, and boundary condition cases. In case of the identification of dynamic processes, a point worth noting is that unlike with linear models, data collected during pseudo random binary signal (PRBS) testing are unsuitable for nonlinear identification purposes.

It is important to consider the effects of feedback controllers when collecting data. If the model is to be used for control purposes then using data collected under closed-loop operation may introduce problems. If, however, the model is to be used for monitoring purposes then the process data should be collected with the system in its standard configuration. For example, if the system typically operates in closed loop then the data should be collected in closed-loop operation.

Preparing the data is an important consideration. The data may need to be converted into another form to be meaningful to the neural network. Hence, the preparation includes *transforming inputs* into the proper form (ratios, classes) and data types (binary, bipolar, continuous), and *normalizing* and *scaling* the data in order for the transform function to operate.

Following the cleaning of the process data, it is useful to determine the cause and effect variables and the major time constants and time delays present in the systems. The tools used for this analysis can include cross-correlation and multivariable techniques such as principal component analysis and principal component regression.

The results from these analyses can be subsequently discussed and validated with process operators and engineers. Such discussions usually prove extremely useful and provide greater insight into the operation of the process than is possible with the data analysis techniques alone.[7]

Applying the Data for Validation

It is important to note that the criterion function used for the training of the model accounts only for modeling errors during the training. This means that the obtained model depends on the training set used. Because of this dependence, the minimization of this criterion function does not guarantee a good neural model. To avoid problems, a common practice is to evaluate the generalization error in parallel with the iterative minimization process. The generalization error is obtained by testing the neural model with fresh data, i.e., data not used for the training.

Usually, the training error is reduced when the number of hidden neurons is increased (degree of freedom is increased) or when more training iterations are performed. "Over-training" may cause problems in both cases. In case of over-training, the fit on the training data becomes almost perfect, but the generalization error becomes worse. A plot of this phenomenon is presented in Figure 2.18f.

To avoid overtraining, most often methods are used that compare the training results with the test results. These methods

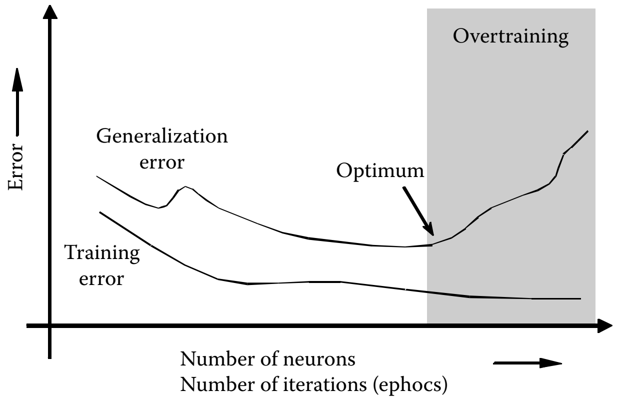


FIG. 2.18f
Training error and generalization error.

include “bootstrapping” or “cross validation.” The cross validation uses two data series, one for training the neural net, and the other to find the optimum, at the minimum of the generalization error. Hence, when the generalization error at a certain number of training iterations increases, the training is stopped and the weights are stored.

Hidden Layers, Nodes, and Algorithms

Determining the Number of Hidden Layers Hidden layers act as layers of abstraction and help the network generalize and even memorize. Increasing the number of hidden layers augments the processing power of the neural network but also significantly increases and complicates training, by requiring more training examples. Generally, most process modeling and control applications require only one hidden layer.

Determining the Number of Nodes Determining the number of nodes in a hidden layer is another experimental exercise. Rules of thumb give a starting point. One suggestion is that if one hidden layer is chosen, the number of processing elements may be equal to 75% of the size of the input layer.[5] Another suggestion is that the number of hidden nodes equals twice the square root of the sum of the number of input and output characteristics, rounded down to the nearest integer.[8]

In general, the more complex the relationship, the more processing elements are required in the hidden layer. Depending upon the number of total nodes in a network, a sufficient number of training sets must be provided in order to train the system adequately. Otherwise, the situation is the same as trying to fit a third-order polynomial with two data points, where an infinite number of sets of polynomial constants can satisfy the equation.

The situation of underspecification due to a lack of sufficient training examples is probably the single most common cause of problems for deployed networks.

Besides the previously presented cross-correlation analysis, many statistical tools can help determine the degree to which the network produces the correct outputs and to determine the number of hidden nodes.[9,10]

Tuning the Training Algorithm The initial choice of weights is very important to convergence. Small random numbers should be used, and equal weights should be avoided. It is possible for the system to require many iterations before convergence: 10,000 iterations are not uncommon.

Increasing the learning coefficient speeds up the learning. However, the higher the learning coefficient, the slower will be the convergence. Therefore, if the learning coefficient can be reduced as training proceeds, high learning and fast convergence can take place. Hence, a breakpoint table can be used to define the learning schedule. This table allows the learning coefficient to decay after a number of learning passes.

Gradient descent slows down near the minima and therefore the Levenberg–Marquardt method can be used to improve convergence. Although the searching process can easily get stuck in local minima, the situation can be improved by adding momentum so that the search rolls out of small wells. Another option is to add some random fluctuations to help kick the search out of minima.

APPLICATIONS

Neural networks have made an unusually rapid transition from laboratory experiments into the marketplace, overtaking other, better-established technologies. They are mostly used under human supervision or integrated with expert systems and fuzzy-logic systems. An indication of the commercial potential of neural networks for process control can be gauged from Boolean search of the U.S. Patents Database.[11] Among the 4848 neural network patents, 411 are related to process control, 121 to system identification, 62 to process monitoring, and 1596 to sensors.

In the next paragraphs, some of these applications of ANNs will be reviewed.

Identification of Dynamic Systems

ANNs can learn complex functional relations by generalizing from a limited amount of training data; hence, they can thus serve as black-box models of nonlinear, multivariable static and dynamic systems. Yet, the application of ANNs for the representation of dynamic systems is not so straightforward. While fundamental physical process models are mostly developed in continuous time, computer-based process control systems operate in discrete time; measurements are made and control actions are taken at discrete instants of time.

In addition, the available input/output data used for model identification are only available at discrete instants of time. Hence, neural network models of dynamical systems are mostly identified based on observed inputs $u(k)$ and outputs $y(k)$. In some cases simulated outputs $\hat{y}_u(k|\Theta)$, prediction errors $\varepsilon(k-n_e) = y(k-n_e) - \hat{y}(k-n_e \mid \Theta)$, and simulation errors $\varepsilon_u(k) = y(k) - \hat{y}_u(k \mid \Theta)$ are also used. According to the applied regressors, following the nomenclature of linear models, it is natural to coin similar nonlinear model structures:[12]

- NFIR, *Nonlinear Finite Impulse Response models*; in this case the regressor vector is composed as $\mathbf{x}_k = [u(k-1),\ldots,u(k-n_b)]$.
- NARX, *Nonlinear AutoRegressive with eXogenous input models*, which use regressors $\mathbf{x}_k = [y(k-1),\ldots,\hat{y}(k-n_a),\ u(k-1),\ldots,u(k-n_b)]$.
- NOE, *Nonlinear Output Error models,* which use $\mathbf{x}_k = [\hat{y}_u(k-1 \mid \Theta),\ldots,\hat{y}_u(k-n_b \mid \Theta),u(k-1),\ldots,u(k-n_b)]$.
- NARMAX, *Nonlinear AutoRegressive Moving Average with eXogenous input models,* where $\mathbf{x}_k = [y(k-1),\ldots\ y(k-n_a),\ u(k-1),\ldots,u(k-n_b),\ \varepsilon(k-1),\ldots,\varepsilon(k-n_e)]$.

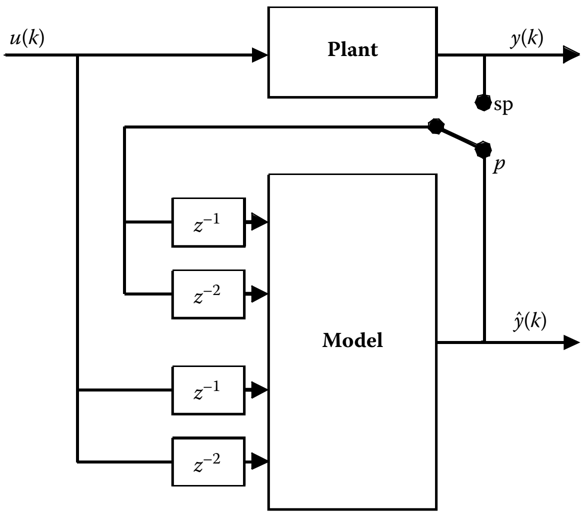


FIG. 2.18g
Series–parallel (sp) or parallel (p) identification methods. z^{-1} *is a backward shift operator.*

- NBJ, *Nonlinear Box-Jenkins models;* where the regressors are past inputs, past estimated outputs, estimation errors using past outputs, and the estimation errors using past estimated outputs $\mathbf{x}_k = [\hat{y}(k-1\,|\,\Theta),\ldots,\hat{y}(k-n_a\,|\,\Theta),\ u(k-1),\ldots,u(k-n_b),\varepsilon_u(k-1),\ldots,\varepsilon_u(k-n_u),\varepsilon(k-1),\ldots,\varepsilon(k-n_e)]$

The NARX model is often called a *series–parallel model*, while the NOE is referred to as a *parallel model*[13] (Figure 2.18g). The NARMAX, NOE, and NBJ models are recurrent models, because they use the estimated output that constitutes feedback. This makes the identification of these models difficult.

Because the NARX model structure is nonrecursive, its parameters are easier to estimate. Therefore, the NARX model is frequently used for identification. The identified NARX (series–parallel) model is often used and tested as a NOE (parallel) model, when the past outputs of the nonlinear model are used instead of the real plant outputs. By using this approach, a multi-step-ahead prediction can be made, using the former predictions of the system.

This procedure is often called "free run" simulation. The "free run" is a very rigorous test of the predictive power of the model because in this way small errors can accumulate and become major ones.

So far, only regressors that consist of previously measured input/output data have been used. However, based on *a priori* knowledge, semiphysical regressors can also be defined.[14] A special type of these regressors exists when the filtered input is used as a regressor.

Special dynamic networks can be realized by using recurrent networks in which neurons are arranged in one or more layers and feedback is introduced either internally in the neurons, to other neurons in the same layer, or to neurons in preceding layers.[15] Examples of these networks are the Elman network (Figure 2.18h) or the Hopfield network. This is a good thing to do if the network should learn to model such feedback behavior.

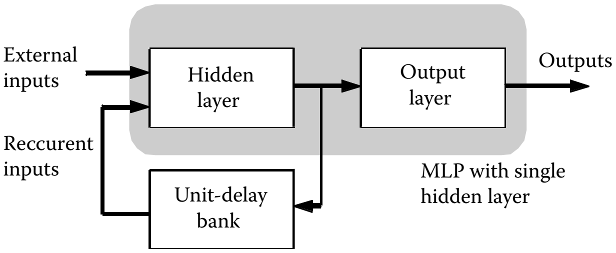


FIG. 2.18h
Example of a recurrent neural network model.

Gray-Box Modeling

Although neural networks are often deployed as stand-alone applications, they can also be integrated into overall solutions. They may be embedded within databases or expert system applications, act as preprocessors or postprocessors to other systems, or be linked in a distributed architecture. From the process modeling point of view, two main integration approaches can be distinguished: bias modeling and semimechanistic modeling.

The bias or parallel modeling approach assumes that a first-principle (white-box) model of the process can be obtained, but it is not possible to identify all the sources and statistical characteristics of disturbances. In this case, a neural network is trained to predict the difference (residual) between the process and its first-principles model.[16]

The semimechanistic model is based on a first-principle model.[17] Because the process is not perfectly known, the unknown parts of the white-box model are represented by black-box elements. To illustrate the structure of these models, consider the modeling of general chemical processes based on macroscopic balances, for instance, mass or energy balances. The following conservation principle describes these balances:

$$\frac{\begin{bmatrix}\text{accumulation of } S\\ \text{within the system}\end{bmatrix}}{\text{time period}} = \frac{\begin{bmatrix}\text{flow of } S\\ \text{into the system}\end{bmatrix}}{\text{time period}} - \frac{\begin{bmatrix}\text{flow of } S\\ \text{out of the system}\end{bmatrix}}{\text{time period}}$$
$$+\frac{\begin{bmatrix}\text{amount of } S\\ \text{generated}\\ \text{within the system}\end{bmatrix}}{\text{time period}} - \frac{\begin{bmatrix}\text{amount of } S\\ \text{consumed}\\ \text{within the system}\end{bmatrix}}{\text{time period}} \qquad \mathbf{2.18(9)}$$

where S is a certain quality, e.g., mass or energy. As in the modeling phase it becomes evident which parameters of this model structure are easier and which are more laborious

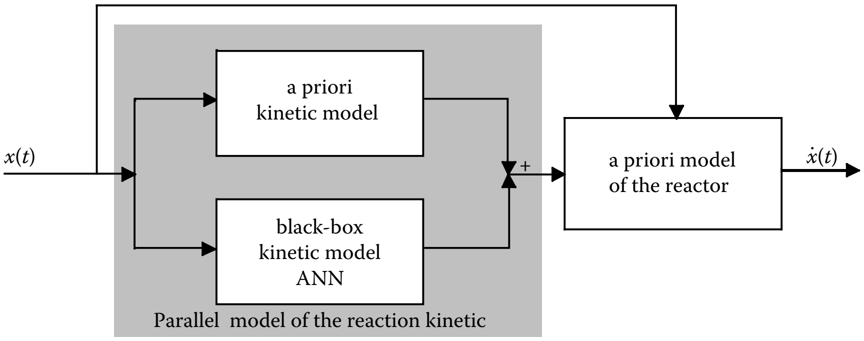


FIG. 2.18i
Series–parallel combination of white- and black-box models.

to obtain. Black-box models can be used to model the otherwise difficult-to-calculate parameters. For example, Hubert te Braake used this type of approach to model a continuously stirred tank reactor (CSTR).[18]

Thompson and Kramer combined the parallel and the hybrid modeling approaches into a series–parallel model (Figure 2.18i).[17] Of course, more complex combinations of white and black boxes can also be obtained. The key issue of this approach is the method to train these black boxes, which are based on input/output data. For this purpose, direct training data obtained from well-designed experiments or an extended Kalman filter can be used.

Neural Networks in Process Control

A neural controller is an automatic controller in which an artificial neural network is utilized in the process of calculating the control signal.[19] As described previously, the functionality of an ANN is usually wholly defined through learning by example. This implies that an important part of the design of a neural controller is the design of a mechanism by which learning exemplars can be collected and used to adjust the parameters of the ANN. This mechanism is very similar to that of adaptive controllers, so most neural controllers in fact are adaptive in their nature.

Figure 2.18j shows the hierarchical classification of ANN-based control algorithms. Each of the groups is briefly described in Table 2.18k.

The ability of neural networks to represent nonlinear relations leads to the idea of using networks directly in a model-based control strategy[20] (see Section 2.10 on Genetic Algorithms).

The concept is based on the capability of neural networks to learn both a system input/output relationship and the corresponding inverse relationship. In these solutions artificial neural networks are used for the construction of plant models and their inverses, and it is intended that they be used directly within the internal model control (IMC) structure[21] (Fiure 2.18l).

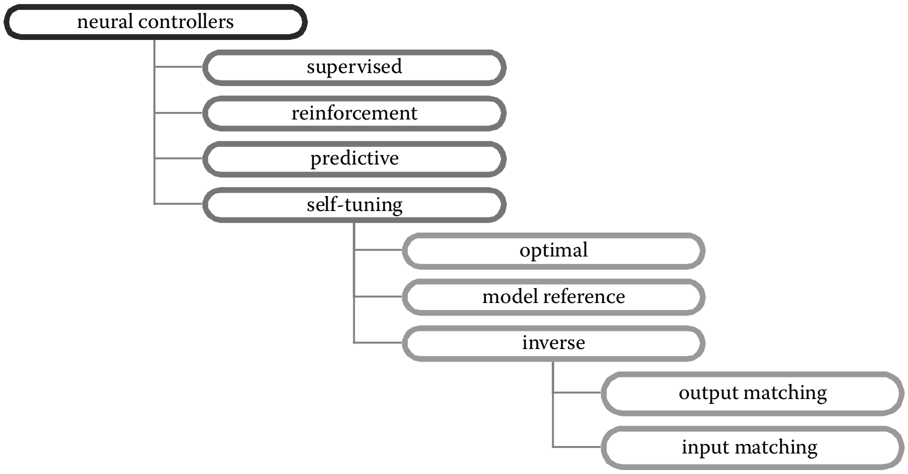


FIG. 2.18j
Neural controller classification scheme based on the method by which the error signal used for training the ANN is derived.

TABLE 2.18k

Brief Descriptions of a Variety of Neural Controller Classes

Class	*Description*
Supervised	ANNs trained to mimic another controller (e.g., a human operator)
Adaptive	Neural controllers defined by use of minimization of cost functions
Reinforcement	Neural controllers trained by means of reinforcement learning
Predictive	Neural controllers trained to match the output of an optimization routine operating on a plant emulator
Optimal	Nontrivial cost functions, i.e., ones that are not based purely on the current plant output or control signal error, are minimized
Model reference	Neural controllers trained to make the system track a reference model
Inverse	Inverse models of the plant are used as a controllers
Output matching	Methods aimed at minimizing the error between the plant output and the reference signal
Indirect output matching	Controller errors are calculated from plant output error signals by back propagating them through plant models
Direct output matching	Controller errors are calculated directly from plant output error signals without use of plant models
Input matching	Methods aimed at minimizing the error between a known control signal and the output of the neural controller
Indirect input matching	Controller errors are calculated using inverse plant models
Direct input matching	Controller errors are calculated using the controllers as inverse models

Following the standard IMC practice, the controller is selected as the plant inverse model. The model uncertainty is included in the feedback information; thus, the effect of the uncertainty on the system output will be reduced. The difference between the outputs of the actual plant and the internal model is caused not only by the model uncertainty, but also by disturbances acting on the plant. Therefore, internal model control can also reduce the disturbance effect on the system output.

Recent developments have resulted in theoretical frameworks to demonstrate stability for ANN-based nonlinear control, and these have arisen from combined developments in control theory and in numerical analysis. It is a telling development that traditional control expertise is even more necessary for the safe deployment of neural network controllers than it was for traditional three-term controllers.

Adaptive control (discussed in the previous section) applications are seen as one area where neural networks have much to offer. Again, further research is required in this area to increase the credibility of adaptive neural network solutions.

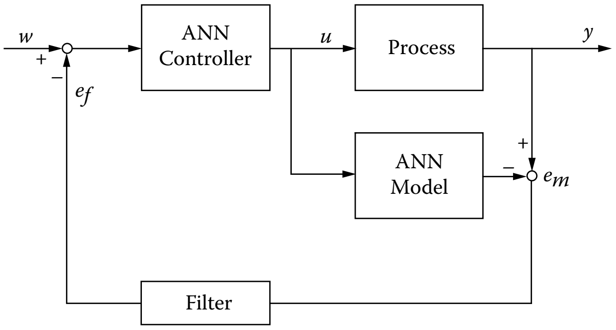


FIG. 2.18l

Application of ANNs in the internal model control structure.

Plant Monitoring and Fault Detection

The aim of plant monitoring is to compare actual plant operation with the desired one. Any deviation is indicative of plant malfunction, which can trigger a diagnostic search for the most likely fault, in real time. The benefit here is in obtaining earlier warning about incipient faults, as well as the potential for providing the operator with a ranked list of possible hazards (see Section 1.6).

The difference between using neural networks rather than purely knowledge-based systems for this application lies in the ability to integrate complex signals at a low level, yielding more complex advice than may be possible by representing the activity of individual sensors directly in symbolic form.[22]

The application of ANNs for nonlinear principal component analysis is an interesting example of information integration. While traditional PCA has been widely studied and successfully applied to process engineering, it has the limitation of being a linear operation.

Kramer presented a nonlinear principal component analysis method based on auto-associative neural networks, which was further improved by Dong and McAvoy.[23] The architecture of this network is shown in Figure 2.18m.

The bottleneck layer of the network performs the dimension reduction because the number of neurons in this layer is smaller than that in the input and output layers, so that the network is forced to develop compact representation of the input data.

Pham and Oztemel used an integrated neural network-based pattern recognizer and expert system for analyzing and

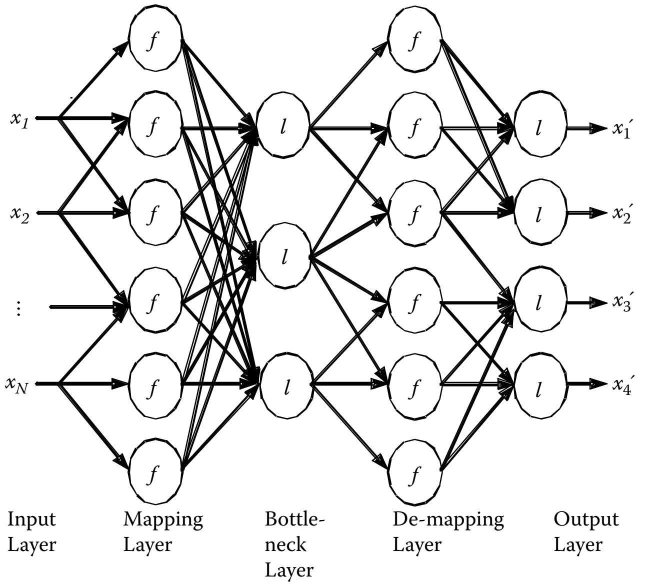


FIG. 2.18m
An auto-associative nonlinear PCA network with five layers. Transfer functions f are nonlinear and transfer functions l are linear.

interpreting control charts.[24] In another application, intelligent process monitoring functions based on a unique combination of rule-based expert systems and neural networks were developed to control a complex pulp and paper process.[25]

Fault detection in process systems has been the subject of numerous studies in the past few decades. Initial work in the area employed a variety of paradigms to both detect and characterize faults, including signal-based, model-based, and knowledge-based approaches. Model-based approaches traditionally employed linear state-space modeling techniques such as Kalman filters.

Fault detection is achieved in these methods through the analysis of the residuals produced by the models. As with model-based controllers, improvements to the accuracy of the monitoring procedure can be made through the utilization of ANNs if the system exhibits nonlinear characteristics. Examples of applications are fault classification; turning, tapping, and welding comparisons; and acoustic emission for tool condition monitoring.[26]

Intelligent Soft Sensors

Further gains are likely to result from the development in soft sensor technology (see Section 7.18 in Volume 1 of this handbook). Industrial demand for soft sensors is substantial. Safety benefits will arise from access to estimates of hidden state variables. An example of this would be estimating the amount of pulverized fuel in a coal mill, when wet coal can become trapped in the rolling chains of a coal feeder. In effect, what is needed is a sophisticated nonlinear observer that is capable of estimating the coal content of the mill for air temperature and differential flow measurements.

There are many other instances in the process industries where the direct measurement of variables that are key to either the detection of performance or to safety are unavailable. This is either because they are too expensive to measure accurately or are impossible to measure fast enough for closed-loop control. Examples of such cases include some pH and other measurements in loop reactors.

Industrial Applications

Although a number of process companies have applied ANNs, few have reported their results publicly. Some application examples involve Texas Eastman, which has used neural networks for steady-state modeling and has significantly improved the utilization of some costly intermediate chemical species.[27]

Neural network modeling was also used to identify operating cost savings in a municipal wastewater treatment facility.[28] Texaco has reported a feedforward control application on a refinery cooker, which greatly improved drum switchover.[29] DuPont has used neural networks as soft sensors, and for model-based control of an acid plant.[30]

Like expert systems, several good neural shells have come onto the market.[31] These shells offer a very good method of organizing and implementing an application. With these tools ANNs can be deployed either as stand-alone applications or as integrated overall solutions. They may be embedded within databases or expert system applications,[32,33] act as preprocessors or postprocessors to other systems, or be linked in a distributed architecture, combined with statistical validation. They have also been introduced as viable solutions for the control of manufacturing facilities.[34]

Among the biggest vendors of control systems, Emerson (former Fisher-Rosemount) offers an *Intelligent Sensor Toolkit* for creating virtual sensors. This tool can be also used for process analysis and knowledge discovery to help determine pieces of process model in a variety of control applications. NeuCOP II from NeuralWare (an AspenTech subsidiary in Pittsburgh, Pennsylvania) combines neural networks, statistics, and multivariable modeling techniques to create dynamic, nonlinear models from process data. The NeuCOP II controller incorporates these nonlinear models into a model-predictive control strategy with an embedded optimizer. It can compute optimal control actions without violating the operating constraints. The results can then be displayed to an operator for manual implementation or fed to a real-time control system, such as DMCPlus, for automatic implementation.

With this tool, neural networks can be installed in existing multivariable control applications to improve our ability to calculate inferential properties and to provide tighter control of highly nonlinear processes.

The Process Perfector from Pavilion Technologies (Austin, Texas) also combines neural networks with model-predictive control technology. Unlike traditional model-predictive controllers, however, the Process Perfector uses a nonlinear model that is applicable to a wider range of processes.[35]

CONCLUSIONS

The potentials of neural network technology in the process industries are vast. The ability of neural networks to capture and model process dynamics and severe process nonlinearities makes them powerful tools in model-based control and monitoring.

The theoretical and practical issues associated with using neural networks for industrial applications were reviewed in this section. Readers were also provided with an insight into the problems and benefits encountered when exploiting this technology.

In the future, neural networks will be used not only on their own, but also in conjunction with other advanced technologies. The fusion of neural networks and fuzzy logic in the form of neuro-fuzzy techniques is seen by many as the most promising way ahead for advanced process monitoring and control applications. An alternative field that also offers great potential is hybrid modeling, an identification methodology that complements simplified mechanistic models with either linear or nonlinear data-based models.

References

1. McCulloch, W. H., and Pitts, W. S., "A Logical Calculus of the Ideas Immanent in Neural Nets," *Bulletin of Mathematical Biophysics,* Vol. 5, 1943, pp. 115–133.
2. Rumelhart, D. E., and McClelland, J. L., *Parallel Distributed Processing,* Vol. 1, Cambridge, MA: MIT Press, 1986.
3. Cybenko, G., "Approximation by Superposition of a Sigmoidal Function," *Mathematics of Control, Signals, and Systems,* Vol. 2, 1989, pp. 303–314.
4. Hartman, E. J., Keeler, J. D., and Kowalski, J. M., "Layered Neural Networks with Gaussian Hidden Units as Universal Approximations," *Neural Computation,* Vol. 2, 1990, pp. 210–215.
5. Bailey, D. L., and Thompson, D. M., "How to Develop Neural-Network Applications," *AI Expert,* June 1990, pp. 38–47.
6. Bailey, D. L., and Thompson, D. M., "Developing Neural-Network Applications," *AI Expert,* September 1990, pp. 34–41.
7. Lennox, B., Montague, G. A., Frith, A. M., Gent, C., and Bevan, V., "Industrial Application of Neural Networks — An Investigation," *Journal of Process Control,* Vol. 11, 2001, pp. 497–507.
8. Hillman, D., "Software Review: Neuroshell v. 3.2," *AI Expert,* September 1990, pp. 61–64.
9. Hsiao, T. C. R., Lin, C. W., and Chiang, H. K., "Partial Least-Squares Algorithm for Weights Initialization of Backpropagation Network," *Neurocomputing,* Vol. 50, 2003, pp. 237–247.
10. Islam, M., Yao, X., and Murase, K., "A Constructive Algorithm for Training Cooperative Neural Network Ensembles," *IEEE T Neural Networks,* Vol. 14(4), 2003, pp. 820–834.
11. U.S. Patents Database, http://www.uspto.gov/patft/index.html.
12. Sjoberg, J., Zhang, Q., Ljung, L., Benveniste, A., Deylon, B. Glorennec, P.-Y., Hjalmarsson, H., and Juditsky, A., "Nonlinear Black-Box Modeling in System Identification: a Unified Overview," *Automatica,* Vol. 31, 1995, pp. 1691–1724.
13. Narendra, K. S., and Parthasarathy, K., "Identification and Control of Dynamical Systems Using Neural Networks," *IEEE Transactions on Neural Networks,* Vol. 1, 1990, pp. 4–27.
14. Abonyi, J., *Fuzzy Model Identification for Control,* Boston: Birkhauser, 2003.
15. Chovan, T., Catfolis, T., and Meert, K., "Neural Network Architecture for Process Control Based on the RTRL Algorithm," *AIChE Journal,* Vol. 42(2), 1996, pp. 493–502.
16. Psichogios, D. C., and Ungar, L. H., "A Hybrid Neural Network — First Principles Approach to Process Modeling," *AIChE Journal,* Vol. 38, 1992, pp. 1499–1511.
17. Thompson, M. L., and Kramer, M. A., "Modeling Chemical Processes Using Prior Knowledge and Neural Networks," *AIChE Journal,* Vol. 40, 1994, pp. 1328–1340.
18. te Braake, H. A. B., van Can, H. J. L., and Verbruggen, H. B., "Semi-Mechanistic Modelling of Chemical Processes with Neural Networks," *Engineering Applications Artificial Intelligence,* Vol. 11(4), 1998, pp. 507–515.
19. Hunt, K. J., Sbarbaro, D., Zbikowki, R., and Gawthrop, P. J., "Neural Networks for Control Systems — A Survey," *IEEE Transactions on Neural Networks,* Vol. 28, 1992, pp. 1083–1112.
20. Psichogios, D. C., and Ungar, L. H., "Direct and Indirect Model-Based Control Using Artificial Neural Networks," *Ind. Eng. Chem. Process Des. Dev.,* Vol. 30, 1991, pp. 2564–2573.
21. Hunt, K. J., and Sbarbaro, D., "Neural Networks for Nonlinear Internal Model Control," *IEE Proceedings — D,* Vol. 138, 1991, pp. 431–438.
22. Stephanopoulos, G., and Han, C., "Intelligent Systems in Process Engineering: A Review," *Computers and Chemical Engineering,* Vol. 20, 1996, pp. 743–791.
23. Dong, D., and McAvoy, T. J., "Nonlinear Principal Component Analysis — Based on Principal Curves and Neural Networks," *Computers and Chemical Engineering,* Vol. 20(1), 1996, pp. 65–78.
24. Pham, D. T., and Oztemel, E., "An Integrated Neural Network and Expert System Tool for Statistical Process Control," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* Vol. 209(B2), 1995, pp. 91–97.
25. Beaverstock, M. C., "It Takes Knowledge to Apply Neural Networks for Control," *ISA Transactions,* Vol. 32, 1993, pp. 235–240.
26. Stich, T. J, Spoerre, J. K., and Velasco, T., "The Application of Artificial Neural Networks to Monitoring and Control of an Induction Hardening Process," *Journal of Industrial Technology,* Vol. 16, 2000, pp. 1–11.
27. Samdani, G., "Neural Nets. They Learn From Examples," *Chemical Engineering,* 1990, pp. 37–45.
28. Miderman, P. A., and McAvoy, T. J., "Neural Net Modeling and Control of Municipal Waste Water Process," *Proceedings of American Control Conference,* 1993, pp. 1480–1484.
29. Hobson, G., "Neural Net Applications at PSP," *National Petroleum Refiners Association Meeting,* 1990, pp. CC-90-137
30. Piosovo, M., and Owens, A., "Sensor Data Analysis Using Artificial Neural Networks," *Proceedings of Chemical Process Control IV,* 1991, pp. 101.
31. Reid, K., and Zeichick, A., "Neural Net Resource Guide to 50 Tools," *AI Expert,* June 1992, pp. 50–56.
32. Verduin, W. H., "Neural Network Software Assists Expert System," *Control Engineering,* July 1991, pp. 37–40.
33. Caudill, M., "Using Neural Nets: Hybrid Expert Networks," *AI Expert,* November 1990, pp. 49–54.
34. Minor, J. M., "Applied Neural Nets: Industrial Strength Solutions," *Expert Systems Applications in Advanced Control, Proceedings of the Seventeenth Annual Advanced Control Conference,* 1991, pp. 83–97.
35. VanDoren, V. J., "Advanced Control Software Goes Beyond PID," *Control Engineering,* January 1, 1998.

Bibliography

Grossberg, S., *Neural Networks and Natural Intelligence,* Cambridge, MA: MIT Press, 1988.

Haykin, S. S., *Neural Networks: A Comprehensive Foundation,* 2nd ed., Englewood Cliffs, NJ: Prentice Hall, 1998.

Hecht-Nielsen, R., *Neurocomputing,* Menlo Park, CA: Addison-Wesley, 1989.

Jang, J. S. R., Sun, C.-T., and Mizutani, E., *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence,* Englewood Cliffs, NJ: Prentice Hall, 1996.

Miller, T. W. III, Sutton, R. S., and Werbos, P. J., *Neural Networks for Control,* Cambridge, MA: MIT Press, 1990.

Nelles, O., *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models,* London: Springer-Verlag, 2000.

Omatu, S., Khalid, M., and Yusof, R., *Neuro-Control and Its Applications,* London: Springer-Verlag, 1996.

Pham, D. T., and Liu, X., *Neural Networks for Identification, Prediction, and Control,* London: Springer-Verlag, 1995.

Simpson, P. K., *Artificial Neural Systems, Foundations, Paradigms, Applications, and Implementations,* New York: Pergamon Press, 1990.

Suykens, J. A. K., Vandewalle, J. P. L., and De Moor, B. L. R., *Artificial Neural Networks for Modelling and Control of Non-Linear Systems,* New York: Kluwer Academic, 1996.

Swingler, K., *Applying Neural Networks: A Practical Guide,* New York: Academic Press, 1996.

Zeidenberg, M., *Neural Network Models in Artificial Intelligence,* New York: Horwood, 1990.

2.19 Nonlinear and Adaptive Control

F. G. SHINSKEY (1970) **P. D. SCHNELLE** (1985) **B. G. LIPTÁK** (1995)
R. D. ROJAS (2005)

INTRODUCTION

This section discusses the conceptual approaches and tools of adaptive control and their applications to nonlinear processes. Adaptive control can be thought of as a system that automatically designs a control system. This technique has considerable potential for applications where the processes being controlled are not well understood or have significant nonlinearities or time-varying parameters.

Adaptive control utilizes the natural steps of control system design. These include process modeling and identification (Section 2.16), relative gain calculations (Section 2.25), controller design (for example, self-tuning controllers, Section 2.29), and optimization (Section 2.20).

At present, these approaches and ideas are being extended to include such new tools as artificial intelligence and as expert systems (Section 2.8), neural networks (Section 2.18), fuzzy logic (Section 2.31), and genetic algorithms (Section 2.10), among others. Therefore, it is suggested that all of these sections be studied if a thorough understanding of the subject matter is desired.

DEFINITIONS AND TERMINOLOGY

Most processes are nonlinear in some respects. The process gain can vary with load (for example, the gain of heat transfer processes will drop with rising load) or can change with time (for example, because of dirt or coating buildup). The dead time and with it the period of oscillation can also vary (for example, the transportation lag or dead time represented by a reactor jacket increases as the water flow rate through the jacket drops).

Feedback control loops are designed so that the controlled variable (the process output) is maintained at its set point even if disturbances occur or the dynamics of the process changes. When the process is operating close to the upper or lower boundaries of its performance range, the total loop gain will approach one of its limits. If it approaches its upper limit, it will cause undamped oscillation; if it drops near the lower limit, the loop becomes sluggish and cannot keep its controlled variable on set point. Similarly, because the process dead time constrains the oscillations of the feedback loops, when the dead time of the process changes, the control settings also need to be revised.

Adaptive control involves the automatic detection of changes that occur in the process parameters or in the set point and the readjustment of the controller settings, thereby adapting the tuning of the loop to the changing conditions. Adaptation can be based on the inputs entering the process (known operating conditions or measurable disturbances) or on the closed-loop performance (behavior of the controlled variable, c, or tracking error, e). In the first case, the system is adjusted on the basis of a measurement of the disturbance variable, which is called *feedforward adaptation* (or *gain scheduling*). In the second case, the system uses a measurement of its own performance, which is called *feedback adaptation* (or *self-adaptation*).

Steady-State and Dynamic Adaptation

The adaptation criteria can be specified in the steady state: an example is the goal of setting the air–fuel ratio of a boiler so that it will maintain the process to maximum efficiency (the minimum total-loss) point on its performance curve. This is called *steady-state adaptive* (or *optimizing*) *control.*

Another way to specify the adaptation criteria is on the basis of the process transient response. An example of this approach is the modification of tuning constants based on the damping of the controlled variable after an upset. In this case the adaptation method is called *dynamic adaptation.*

This method can be used, for example, to stabilize a PID-based pH control system. Without adaptation, when the process gain (K_p) rises as the process approaches neutrality, the increased process gain results in the loop becoming unstable (cycling). In this application the adaptive controller maintains the damping ratio (usually set for $^1/_4$). Therefore, when the controller realizes that K_p has risen, it starts to lower the controller gain (K_c) so that the total gain product of the loop ($K_pK_cK_vK_t$, where K_v and K_t are valve and transmitter gains, respectively) will remain at the desired value (usually 0.5).

An adaptive control system is one whose parameters are automatically adjusted to meet the corresponding variations in the behavior of the process being controlled in order to optimize the response of the loop. The significant difference is that adaptive control changes the controller *parameters* (*tuning settings*), which in PID control are normally fixed, as opposed to changing the outputs of a system, which are expected to vary.

The parameters set into controllers and control systems naturally reflect the characteristics of the processes they control. To maintain optimum performance, these parameters should change as the associated process characteristics change. If these changing characteristics can be directly related to the magnitude of the controller input or output, it is possible to compensate for their variation by the introduction of suitable nonlinear functions at the input or output of the controller.

Examples of this type of compensation would be the introduction of a square-root extractor in a differential pressure type flow-measurement signal to linearize it, or the selection of a particular valve characteristic to offset the effect of line resistance on flow. This step is not considered to be part of adaptive control, in that the controller functions remain fixed.

If the process nonlinearities are compensated by controller functions (i.e., by feedback linearization or by changing PID controller gain, integral, or derivative tuning settings) the controller is nonlinear. An example is a pH controller, where the controller gain is adjusted to compensate for the nonlinearities of the pH system titration curve. Changing controller gain as a function of the pH controller input (measured variable) is not strictly adaptive, but it is nonlinear.

APPROACHES TO ADAPTIVE CONTROL

Adaptive control has been an important field of theoretical research, but now it is also a good tool to solve real-world problems. Classical adaptive control techniques have shown to provide simple solutions to the control of nonlinear and time-variant systems; several texts and articles pursue the details of the approaches presented here.

Some good examples are the books of Astrom and Wittenmark,[1] Sastry and Bodson,[2] and Ioannou and Sun;[3] electronic versions of the latter two books can be obtained by free download from the authors' Web pages. They can be complemented with the Tao[4] book, which summarizes the recent fast growth of adaptive control applications and theory development.

With the increasees in the complexities of processes, including the effects of system component imperfections and nonsmooth nonlinearities, the adaptive control system is expected to provide greater robustness. These needs justify the emergence of such expert and artificial intelligence techniques as neural networks, fuzzy logic, and genetic algorithms.

Feedforward or Gain Scheduling

Where a measurable process variable produces a predictable effect on the gain of the control loop, or a known nonlinear behavior depends on the operating conditions, compensation for its effect can be programmed into the control system. This is the basis for *gain scheduling.* The most notable example of gain scheduling is the variation of dynamic gain with flow in pipes and other longitudinal equipment where no back-mixing takes place. This variation of dynamic gain is commonly seen in heat exchangers, but it causes a particularly severe problem in once-through boilers.

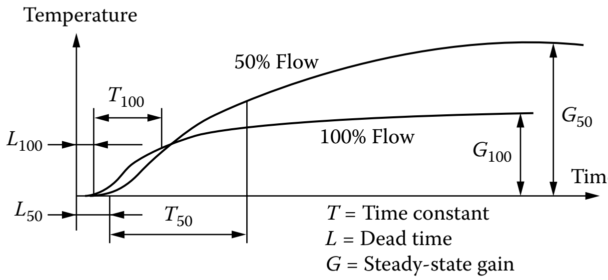


FIG. 2.19a

Step response of steam temperature to firing rate in a once-through boiler.

In a once-through boiler, feed water enters the economizer tubes, passes directly into the evaporative tubing and then into the superheater tubing, and leaves as superheated steam whose temperature must be accurately controlled. No mixing takes place as in drum-type boilers, so a sizable amount of dead time exists in the temperature control loop, particularly at low flow rates.

Figure 2.19a shows the response of steam temperature to changes in firing rate at two different feed water flow rates. At 50% flow, the steady-state gain is twice as high as at 100% flow, because only half as much water is available to absorb the same increase in heat input. The dead time and the dominant time constant are also twice as great at 50% flow.

The effect of these variable properties on the dynamic gain of the process is evidenced in Figure 2.19b. The same size load upset produces a larger excursion in temperature at 50% flow, indicating a higher dynamic gain. The difference in damping between the two conditions also reveals the

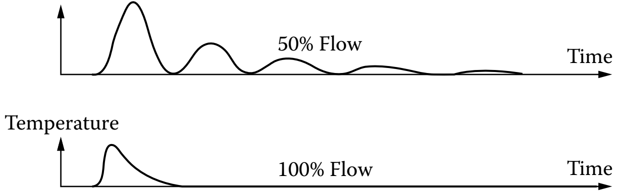


FIG. 2.19b

Response of the steam temperature loop to step changes in load without adaptation.

change in dynamic gain and so does the response, which is twice as fast at 100% flow. If oscillations were evident in the 100% flow response curve, their period would be much shorter than for 50% due to the difference in dead time. At 25% flow, oscillation periods would be still shorter and damping would disappear altogether.

If the only feedback mode used in this application were proportional, the period of oscillation would vary inversely with flow, as would the dynamic gain at that period. The only adjustment that could be made would be that of the controller gain (K_c), which should vary in direct proportion to flow. This would make the damping uniform, but with a plain proportional controller nothing can change the increased sensitivity of the process to upsets at low flow rates.

Since reset and derivative modes are normally also used to control temperature, some consideration should also be given to their adjustment as a function of changes in flow. The process dead time—hence the period of oscillation under proportional control—varies inversely with flow. Therefore, in order to complete the adaptation, the reset and derivative time constants should also be varied inversely with flow.

The equation for the flow-adapted three-mode controller is

$$m = K_c w\left(e + \frac{w}{T_i}\int e\,dt + \frac{T_d}{w}\frac{d}{dt}e\right) + b \qquad \mathbf{2.19(1)}$$

where w is the fraction of full-scale flow and K_c, T_i, and T_d are the proportional, integral, and derivative settings at full scale flow. Equation 2.19(1) can be rewritten to reduce the adaptive terms to two:

$$m = K_c\left(we + \frac{w^2}{T_i}\int e\,dt + T_d\frac{d}{dt}e\right) + b \qquad \mathbf{2.19(2)}$$

Other parameters can be substituted for flow in instances where the adaptation is based on variables other than flow.

Dead-Band Control Dead-band control is a frequently used form of nonlinear control. The dead-band action is typically a programmed nonlinearity or adjustment depending on whether dead band is set by a process variable or by a disturbance factor.

Dead-band control is not a stand-alone control function. Proportional, integral, and derivative modes can still be used. If this function is used, a dead band is placed around the controller error such that no control action occurs unless the error exceeds the dead-band range. If the error is within the dead band, the error is set to zero.

Dead-band control is also used to stabilize very fast or sensitive processes (guidance of missiles or other projectiles) where stability is gained by not making changes (in direction) unless the limits of the gap (control tunnel) are approached. Dead-band control has found application in pH control and in systems requiring two control valves—one large, one small—to overcome the rangeability problem.

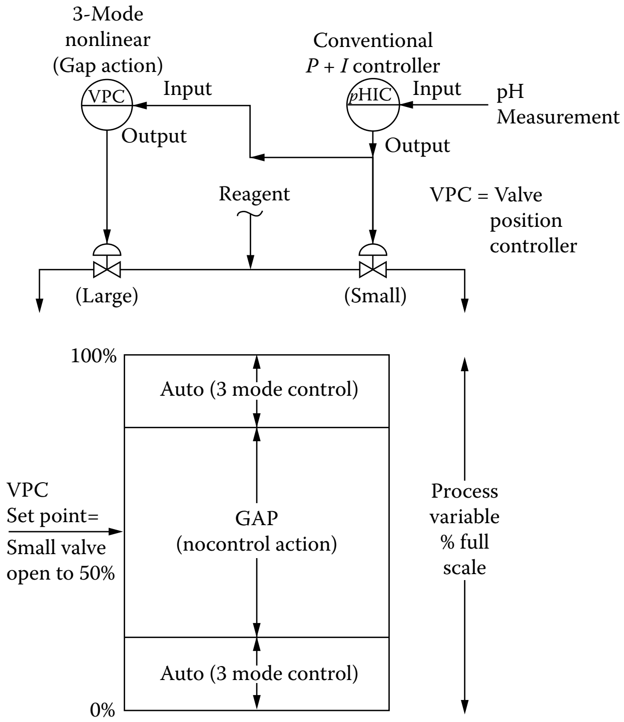


FIG. 2.19c
The valve position controller (VPC) measures the opening of the small valve and prevents it from fully opening or fully closing by becoming active only when such extreme positions are evolving.

For example, consider the control scheme shown in the top portion of Figure 2.19c. Here the dead-band controller is used to drive the large valve when the output of the conventional proportional and integral controller reaches a preset limit (top or bottom). The large valve makes a rough adjustment when the error is outside of its dead band or gap. Returning the pH back within the gap allows the small control valve to once again take control.

Figure 2.19c shows the regions where the dead-band control action is in operation. Depending on the relative size of the small (trim) valve and the response desired, the size of the dead band can vary widely. Commercial controllers and DCS or computer control packages are available with this dead-band feature.

Switching Controller Gains Another form of programmed adaptation or nonlinear control is referred to as variable breakpoint control. The proportional gain for this type of controller is changed at certain predefined values of the process variable or controller error (process variable − set point). In this type of application it is usually desired to set the controller gain at a low value at or around the set point, so that not much controller action would occur near the set point.

If the process moves to some preset distance from set point, the controller gain is increased to drive the process back into the set point region at a faster rate. This type of control action may also be used to compensate for very nonlinear process gain characteristics. In this case the controller gain is set high when the process gain is low and set low when the process gain is high in order to maintain a relatively constant (0.5 or so) closed-loop gain (i.e., consistent closed-loop performance throughout control range).

The pH control problem is a good example of nonlinear control. Consider the process gain characteristic of the neutralization system illustrated in Figure 2.19d. This is a typical strong acid, strong base neutralization process. The control set point is at pH = 7. The process has extremely high gain at this point (i.e., a small amount of reagent change results in a dramatic change in pH). The required controller gain between a pH of 3 and a pH of 9 must be small for stability. If the pH measurement were to move above 9 or below 3, this low gain would result in a very sluggish response because at those pH values the process gain is low. Therefore, above 9 and below 3 the controller gain is switched to a high value. The required controller gain is shown as a dashed line in Figure 2.19d.

Such nonlinear controllers are available both in the form of electronic hardware controllers or as DCS algorithms. They can also be implemented with more than two line segments, and in more sophisticated systems the controller gains can correspond to the mirror image of the process gain curve.

It is important to point out that the implementation of this algorithm does require more than just a simple switching of gain numbers at breakpoints. Care must be taken to ensure that the gain change does not result in a bump or discontinuity in control action.

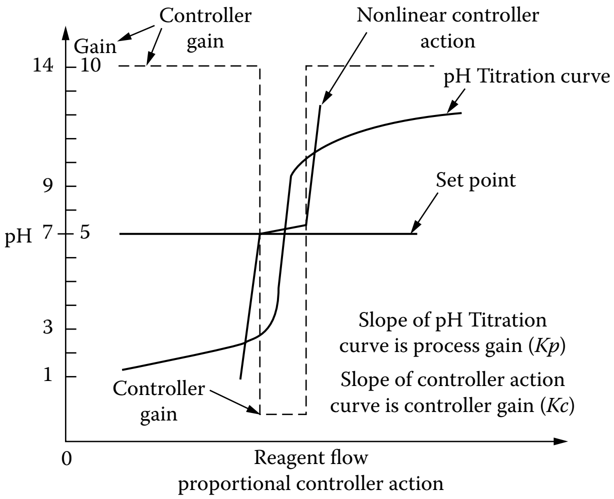


FIG. 2.19d
Variable breakpoint nonlinear control action is used for strong acid, strong base neutralization.

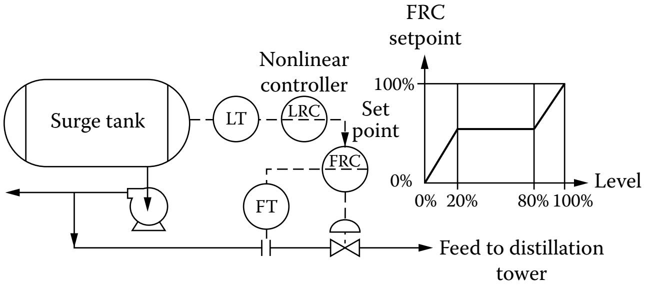


FIG. 2.19e
Nonlinear level control keeps the set point of its cascade slave (FRC) unchanged most of the time while allowing the level to float, but it still prevents flooding or draining of the surge tank.

An example of a process application where the discontinuity in controller gain can be tolerated, while the use of the nonlinear controller improves overall performance, is the case of surge tank level control. As the purpose of installing a surge tank is to absorb the differences in flows between plant sections, the intent is to let the level in these tanks float. When this level is cascaded to a flow controller (Figure 2.19e), it is even more desirable not to react to small level variations (which have no harmful consequence) by changing the flow controller set point, which if changed would upset the material balance of the downstream process.

Therefore, the nonlinear controller is an ideal cascade master for such an application. As shown in Figure 2.19e, the nonlinear controller can be adjusted to make no change at all in the slave (FRC) set point as long as the level is between 20 and 80%. This control system still protects the tank from draining or flooding by becoming active when the level goes outside this gap.

Continuous Adjustment of Controller Gains Just as in the programmed adaptation algorithm of Equation 2.19(1), in which the gain was multiplied by (and thereby was made a function of) the load variable, one can also adjust the controller gain (K_c) as a function of other parameters. One obvious option is to multiply the control gain (K_c) by some function of the error (e). For example, if K_c is substituted with $(K_c)(e^2)$, this will make the controller gain nonlinear and would result in the following PID algorithm:

$$m = K_c(e^2)\left(e + \frac{1}{T_i}\int e\,dt - T_d\frac{dc}{dt}\right) + b \qquad \mathbf{2.19(3)}$$

In Equation 2.19(3), the derivative mode acts on the measurement (c) only, instead of acting on the error (e) in order to make it insensitive to set-point changes. Note that this algorithm has a lower gain when the errors are small, and therefore this controller is less responsive to small errors.

This characteristic makes the nonlinear controller well suited for use in applications involving noisy measurements, such as flow or level, because instead of amplifying the noise, its response is minimal when the error is small. Naturally, it should be remembered that with large errors the controller gain rapidly gets very high and can cause stability problems and cycling unless some reasonable limit is placed on it.

When using "error squared" PID algorithms, a variety of configurations is available. In one configuration, the controller gain (K_c) is substituted with XK_c, where the value of X is found from Equation 2.19(4) below:

$$X = 1.0 + 90\ E/Y + 3{,}045\ E/Y^2 + 145{,}675\ E/Y^3 \qquad \textbf{2.19(4)}$$

where E = the normalized absolute value of the error (E is between 0 and 1.0) and Y = a constant that sets the severity of the desired nonlinearity in the controller action.

For example, if Y = 14%, it means that X will double after every 14% increase in the error. In other words, if X is 2 when the error is 14%, it will be 4 when the error is 28%, and it will be 8 when the error is 42%.

The multiplier X can be applied to the controller gain (K_c) and/or to its integral time (T_i). When both are chosen the algorithm is particularly suited to level control. For more detail on the control technique see Reference 5.

Gain scheduling obtained its name because it was originally used to accommodate changes in the process gain. Today it is also used, based on measurements of the operating conditions of the process, to compensate for the variations in process parameters or for known nonlinearities in the process.

Feedback Adaptation or Self Adaptation

Where the cause of changes in control-loop response is unknown or unmeasurable, *feedforward adaptation* or *gain scheduling* cannot be used. If an adaptive system is to be applied under these circumstances, it must be based on the response of the loop itself, i.e., it must be part of a *feedback adaptive scheme. Feedback adaptation* is a more difficult problem than *feedforward adaptation* because it requires an accurate evaluation of loop responses, ideally without the knowledge of the nature of the disturbance input.

The block diagram of a *feedback adaptive system* is given in Figure 2.19f. This system has all the problems that are associated with implementing *feedforward adaptation* plus the problems of evaluating the response and making a decision on the correct adjustment. Several *feedback adaptive techniques* that are currently being used in the process industry are briefly discussed below. These include the self-tuning regulator, the model reference controller, and the pattern recognition adaptive controller.

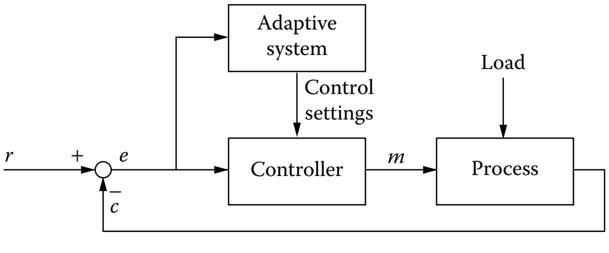


FIG. 2.19f
A self-adaptive system is a control loop around a control loop.

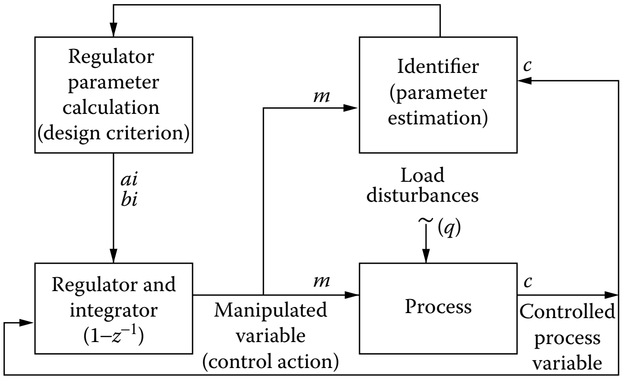


FIG. 2.19g
Self-tuning regulator.

Self-Tuning Regulator The self-tuning regulator (STR) is a name given to a large class of self-adaptive systems. The block diagram in Figure 2.19g shows the common structure of these STR systems. The figure indicates that all STRs have an identifier section, typically consisting of a process parameter estimation algorithm. Also common to all STRs is a regulator parameter calculation section. This section calculates the new controller parameters as a function of the estimated process parameters.

The methods used in these two blocks distinguish the type of STR being used. Popular varieties of STR include minimum-variance, generalized minimum-variance, detuned minimum-variance, dead-beat, and generalized pole-placement controllers. Reference 6 is a summary paper on basic STR technology. For a more detailed treatment of self-tuning controllers, refer to Section 2.29.

Model Reference Adaptive Controls Model reference adaptive control (MRAC) offers another method for the self-adaptive control problem. The controller is composed of a reference model, which specifies the desired performance; an adjustable controller, whose performance should be as close as possible to that of the reference model; and an adaptation mechanism. This adaptation mechanism processes the error between the reference model and the real process in order to modify the parameter of the adjustable controller accordingly.

Figure 2.19h schematically shows how the parts of a model reference controller are organized. Model reference adaptive controllers were originally designed to solve the deterministic servo problem, that is, control of the process-to-variable set

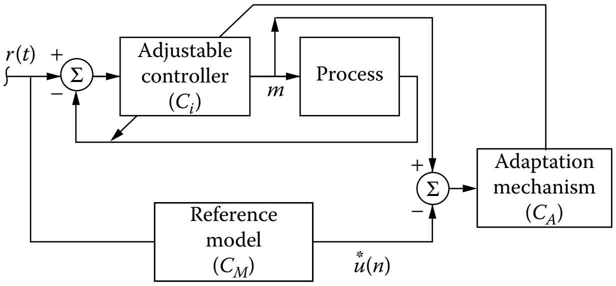


FIG. 2.19h
Model reference adaptive controller.

point reference signals. The design of MRAC has been mostly based on stability theory. For more information, see Reference 7 and also refer to Sections 2.13 and 2.16.

Pattern Recognition Controllers Other self-adaptive controllers exist that do not explicitly require the modeling or estimation of discrete time models. These controllers adjust their tuning base on the evaluation of the system's closed-loop response characteristics (i.e., rise time, overshoot, settling time, loop damping, etc.). They attempt to "cut and try" the tuning parameters and recognize the pattern of the response, thus the name "pattern recognition."

Pattern recognition controllers are commercially available from several instrument vendors. They are microprocessor based and usually heavily constrained with regard to the severity of allowable tuning parameter adjustments. They are gaining fair acceptance in operating plants.

INTELLIGENT ADAPTIVE TECHNIQUES

Process changes, such as flow disturbances and sensor noise, are common sources of inaccurate process information; because they degrade the control loop performance, the use of adaptive control is justified as the complexity of the process (nonlinearities and time-varying parameters) increases.

Adaptive control capability to deal with unknown or time-variant dynamics can be extended by using intelligent control (*intelligent identification and/or tuning*). Furthermore, economically optimal strategies have emerged to improve profitability. Many of these utilize switching operation points that require multiple models or multiple controllers or both (*multiple model adaptive control*).

To maintain optimal performance, the control system has to adapt continuously to the changes in the process and must perform well while it is adapting and while set points are also changing. In such cases, a supervised switching control strategy can be considered, where a variety of control algorithms can be chosen from a *supervisor* or *safety net* that is driven by a switching logic-based criterion, often realized using artificial intelligence or expert systems. Some of these will be discussed in the next paragraphs.

Intelligent Identification and/or Tuning

The overall goal of this approach is to use intelligent structures (fuzzy logic, neural networks, genetic algorithms) to substitute for either the identification, control, or adaptation tasks in the control scheme (see Figure 2.19i). As is discussed

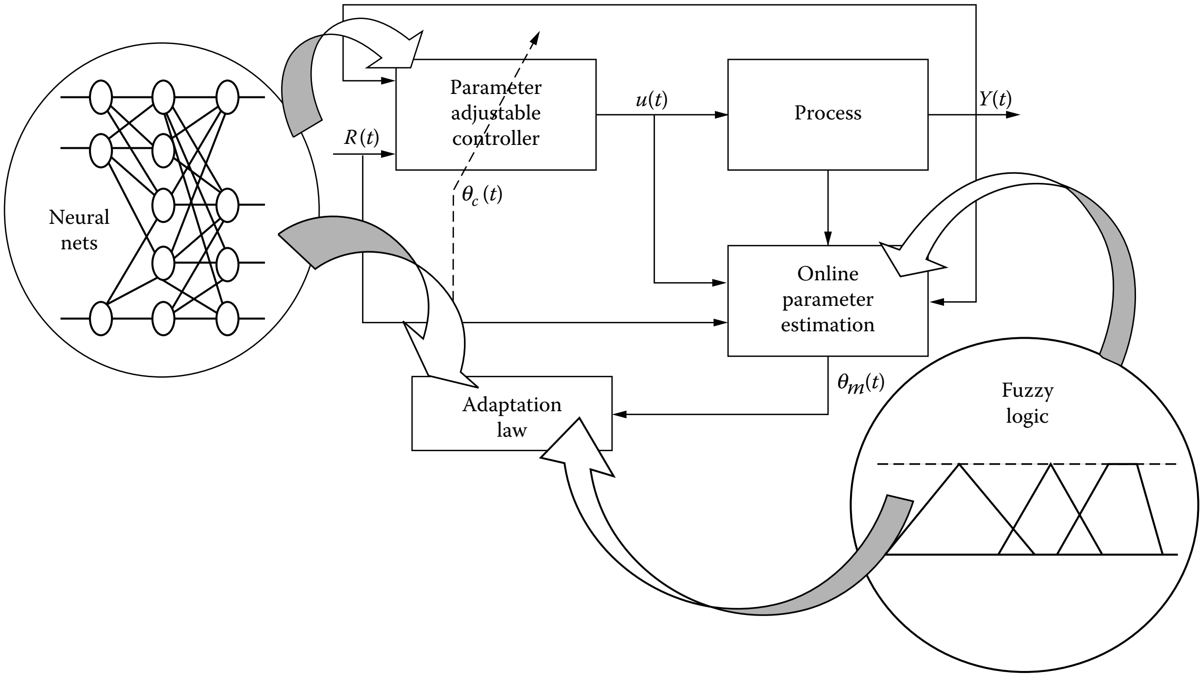


FIG. 2.19i
Intelligent identification and tuning in adaptive control.

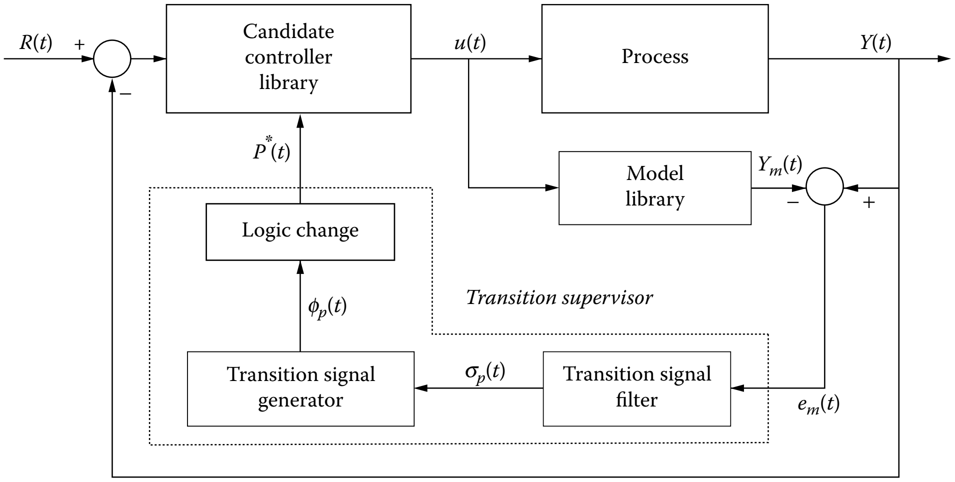


FIG. 2.19j
Adaptive control system for matching multiple process models to multiple process algorithms and selecting the best pairing in a supervised adaptive scheme.

in other sections of this chapter, control schemes combining neural nets with model predictive control (MPC) can tolerate inaccuracy and uncertainty in the model and can apply online training to continuously improve the model.[8]

Multiple Model Adaptive Control (MMAC)

Figure 2.19j illustrates the concept of multiple model adaptive control (MMAC).[9–11] This technique is based on the use of a library of process models, a library of controller algorithms, and an intelligent switching logic.

The MMAC system operates by identifying a process model in its library that closely resembles that of the actual process and, based on that knowledge, selecting a control algorithm (from the *Candidate Controller Library*) that is best suited for controlling such a process. The process model is usually selected on the basis of controlled variable (process output) pattern recognition, from the *Model Library*.

Once the process model is identified, a switching law is applied (*Logic Switching Supervisor* or *Transition Supervisor*) to adaptively select the correct matching between the model and its corresponding controller. The switching laws can roughly be divided into two categories: those based on process estimation and those based on a direct performance evaluation of each candidate controller.

MMAC Limitations and Suppliers MMAC is like a consultant in a box. Most consultants have hours, sometime weeks, to think about the best way of controlling a difficult process, while in the MMAC system, such decisions are automated and the options are those available in the libraries. Therefore, the selected process model will be as good as the library and its selection algorithm, and the same holds true for the controller library.

The main limitation of MMAC is that it is discontinuous; every time a new computer algorithm is selected from the library, it will upset the process. Another limitation of MMAC is its adaptation to start-up and shut-down conditions, where the process model is changing very fast. It should also be noted that at this time few adaptive controllers on the market can update their control strategies while the process is running.

References 12 and 13 provide further information on the suppliers and their products, and the reader can also visit their Web sites:

- QuickStudy, www.quickstudy.com, from Adaptive Resources (Pittsburgh, PA)
- Exact and Connoisseur, www.foxboro.com, from the Foxboro Co. (Foxboro, MA)
- BrainWave, www.brainwave.com, from Universal Dynamics Technologies (Vancouver, BC, Canada)
- Intune, www.controlsoftinc.com, from ControlSoft (Highland Heights, OH)
- KnowledgeScape, www.kscape.com, from KnowledgeScape Systems (Salt Lake City, UT)
- CyboCon, www.cybocon.com, from CyboSoft (Rancho Cordova, CA)

Adaptive model-based controllers such as Exact, BrainWave, QuickStudy, and Connoisseur generate their models automatically while the controller is online, using previously recorded historical process data. This is convenient when compared to classical controllers because in theory, such adaptive controllers can predict process trends and future behavior changes over time.

CyboCon skips the modeling step. Instead, CyboCon looks for patterns in the recent errors and the *learning* algorithm produces a set of gains or *weighting factors* that are then used as parameters for the control law.

CONCLUSIONS

Classically, the performance difference between gain scheduling and self-adaptive systems is analogous to that which exists between feedforward and feedback control.

A self-adaptive controller (comparable to feedback) cannot make an adjustment to correct its settings until an unsatisfactory response is encountered. Two or more cycles must pass before an evaluation can be made upon which to base an adjustment. Therefore, the cycle time of the adaptive loop must be much longer than the natural period of the control loop itself. Consequently, the self-adaptive system cannot correct for the present poor response but can only prepare for the response to the next upset, assuming that the control settings presently generated will also be valid then.

On the other hand, the gain scheduling system should always have the correct settings because it responds automatically to changes in process variables in a manner analogous to feedforward control. It does not have to "learn" the new process dynamics via adaptive loops.

MMAC with supervised switching offers the possibility of good performance and robustness over a wider range of operating conditions than do the traditional approaches, but it will take some time for it to mature. Its potential is best for those processes that are poorly understood and nonlinear with time-varying dynamics (dead times and time constants). There is a great deal of activity in the evolving field of AI-based model adaptive control.[14,15] Some believe that they will eventually eliminate the need for tuning and re-tuning of controllers, but we will have to wait for the fifth edition of this handbook to be sure.

References

1. Astrom, K., and Wittenmark, B., *Adaptive Control*, Reading, MA: Addison-Wesley, 1995.
2. Sastry, S., and Bodson, M., *Adaptive Control: Stability, Convergence, and Robustness,* Englewood Cliffs, NJ: Prentice Hall, 1989. www.ece.utah.edu/~bodson
3. Ioannou, P.A., and Sun, J., *Robust Adaptive Control.* Englewood Cliffs, NJ: Prentice Hall, 1996. http://www-ref.usc.edu/~ioannou/Robust_Adaptive_Control.htm
4. Tao, G., *Adaptive Control Design and Analysis.* New York: John Wiley & Sons, 2003.
5. Shunta, J. P., and Fehervari, W., "Nonlinear Control of Level," *Instrumentation Technology,* January 1976.
6. Soderstrom, T., Ljung, L., and Gustausson, I., A Comparative Study of Recursive Identification Methods, Division of Auto. Cont., Lund Institute of Technology, Report 7427, 1974.
7. Goodwin, G. C., *Adaptive Filtering, Prediction, and Control,* New York: Academic Press, 1981.
8. Nikolaou, M. "Identification and Adaptive Control," *Comp. Chem. Eng.,* special issue on NSF/NIST Vision 2020 Workshop, 23/2, pp. 215–225 (1998).
9. Fu, M., "Minimum Switching Adaptive Control for Model Following," 35th IEEE Conference on Decision and Control, Kobe, December, 1996.
10. Anderson, B., Brinsmead, T., De Bruyne, F., Hespanha, J., Liberzon, D., and Morse, A., "Multiple Model Adaptive Control. Part 1: Finite Controller Coverings," *International Journal of Robust and Nonlinear Control,* 10, pp. 909–929, 2000.
11. Hespanha, J., Liberzon, D., Morse, A., Anderson B., Brinsmead, T., and De Bruyne, F., "Multiple Model Adaptive Control. Part 2: Safe Switching," *International Journal of Robust and Nonlinear Control,* 11, pp. 479–496, 2001.
12. Vandoren, V., *Techniques for Adaptive Control,* Oxford: Butterworth-Heinemann, 2002.
13. Vandoren, V., "Adaptive Controllers Work Smarter, Not Harder" *Journal of Control Engineering,* October 1, 2002.
14. Conradie, A., Miikkulainenm, R., and Aldrich, C., "Adaptive Control Utilising Neural Swarming," Proceedings of the Genetic and Evolutionary Computation Conference, New York, 2002.
15. Miller, D., "A New Approach to Adaptive Control: No Nonlinearities," *Systems & Control Letters,* 49, adaptive control special issue, 67–79, 2003.

Bibliography

Abonyi, J., Andersen, H., Nagy, L., and Szeifert, F., "Inverse Fuzzy-Process-Model Based Direct Adaptive Control," *Mathematics and Computers in Simulation,* 51(1–2):119–132, 1999.

Anderson, B., Brinsmead, T., Liberzon, D., and Morse, A., "Multiple Model Adaptive Control with Safe Switching," *International Journal of Adaptive Control and Signal Processing,* 15, 445–470, 2001.

Balchen, J. G., and Mummè, K. I., *Process Control Structures and Applications,* New York: Van Nostrand Reinhold, 1988.

Borisson, U., "Self-Tuning Regulators for a Class of Multivariable Systems," *Automation,* July 1977.

Box, G. E. P., and Jenkins, G. M., *Time Series Analysis: Forecasting and Control,* rev. ed., San Francisco: Holden-Day, 1976.

Bristol, E. H., "Adaptive Process Control by Pattern Recognition," *Instruments and Control Systems,* March 1970.

Bristol, E. H., "Pattern Recognition Adaptive Control," *IEEE Transactions,* January 1962.

Buckley, P. S., *Techniques of Process Control,* New York: John Wiley & Sons, 1964.

Cutler, C. R., and Ramaker, B. L., "Dynamic Matrix Control — A Computer Control Algorithm," *ACC Conference Proceedings,* San Francisco, 1980.

Fabri, S. G., and Kadirkamanathan, V., *Functional Adaptive Control: An Intelligent Systems Approach,* Heidelberg: Springer-Verlag, 2001.

Harriott, P., *Process Control,* New York: McGraw-Hill, 1964.

Hausman, J. F., "Applying Adaptive Gain Controllers," *Instruments and Control Systems,* April 1979.

Hespanha J. P., Liberzon, D., and Morse, A., "Overcoming the Limitations of Adaptive Control by Means of Logic-Based Switching," *Systems & Control Letters,* 49, adaptive control special issue, 49–65, 2003.

Kelly, K. P., "Should You Be Training Your Operators with Simulation Software?" *Control,* June 1992, pp. 40–43.

Keviczky, L., and Hetthessy, J., "Self-Tuning Minimum Variance Control of MIMO Discrete Time Systems," *Automatic Control Theory and Applications,* May 1977.

Koenig, D. M., *Control and Analysis of Noisy Processes,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Magdalena, L., and Monasterio, F., "Evolutionary-Based Learning Applied to Fuzzy Controllers," Proceedings of Fourth IEEE International Conference on Fuzzy Systems and Second International Fuzzy Engineering Symposium (FUZZ-IEEE/IFES'95), pp. 1111–1118, Yokohama, March 1995.

McCauley, A. P., “From Linear to Gain-Adaptive Control,” *Instrumentation Technology,* November 1977.

McMillan, G., *Tuning and Control Loop Performance,* 2nd ed., Research Triangle Park, NC: ISA, 1990, pp. 45–55.

Morely, R., “Faster Than Real Time,” *Manufacturing Systems,* May 1992, p. 58.

Roffel, B., and Chin, P., *Computer Control in the Process Industries,* Chelsea, MI: Lewis Publishers, 1989.

Shinskey, F. G., *Process-Control Systems,* New York: McGraw-Hill, 1967.

Shinskey, F. G., “A Self-Adjusting System for Effluent pH Control,” paper presented at the 1973 ISA Joint Spring Conference, St. Louis, April 1973.

Shunta, J. P., “Nonlinear Control of Liquid Level,” *Instrumentation Technology,* January 1976.

Trevathan, V. L., “pH Control in Waste Streams,” ISA Paper #72-725, 1979.

Vervalin, C. H., “Training by Simulation,” *Hydrocarbon Processing,* December 1984, pp. 42–49.

2.20 Optimizing Control

F. G. SHINSKEY (1970) **C. G. LASPE** (1985) **B. G. LIPTÁK** (1995) **M. RUEL** (2005)

INTRODUCTION

Optimization of an industrial process usually means an increase in profitability while maintaining safety and product quality. The criteria for optimization vary with the process. Optimization in materials transportation systems (pumps, compressors, fans) usually means the full opening of all throttling devices so that all the energy introduced is used to transport materials and none of it is spent on overcoming artificially introduced friction elements, such as throttling valves and dampers.

The global trend is for higher product quality at a lower cost. The link between product quality and performance is not clearly defined in all cases. An in-depth analysis might be needed to determine the optimum performance for each piece of equipment, then each process, and finally for each control loop. What we do know is that for higher product quality, users must tighten their specifications.

The reader will discover that optimizing a process is similar to a conductor directing an orchestra: coordinate, accelerate here, slow down there, synchronize other parts, etc. The "conductor" must know about control but must also know about the process itself. We should move to multivariable systems when simple single-loop techniques fail, even if properly tuned. On the other hand, we should not make the common mistake of using advanced multivariable control with bad equipment or with control loops that are not tuned to deliver the expected performance.

When optimizing, it is important to consider design, equipment performance, control strategies, operation procedures, performance monitoring, logic, and special control strategies for startup, shutdown, grade changes, and abnormal conditions.

Defining Optimum Performance

The optimization of a batch reactor might mean to operate it at minimum cycle period, which in turn would imply a maximum but safe rate of heat-up, accurate determination of endpoint, and maximum safe rate of stripping. In continuous chemical reactors, optimization might mean maximized rate of conversion (maximum exothermic heat) and therefore an operation where the chilled water valve opening is "pushed" up to and maintained at a safe maximum opening.

The optimization of distillation towers might include the goal of minimizing the heat of separation. As the heat of separation is reduced when the liquids are boiled at lower temperatures, this aspect of optimization usually means that the columns are operated at a safe minimum pressure level.

The optimization of other processes is achieved by operating them at the minimum or maximum point of a cost or efficiency curve. For example, the point where one would like to operate all combustion processes is their minimum total loss point (Figure 2.20a). By plotting all the different forms of losses in a boiler or similar combustion process and adding up these losses to arrive at a total loss curve, it becomes possible to search for the minimum point on that curve. This minimum provides the optimum excess air set point for this process.

Similar total operating cost curves can be constructed for chillers and other types of cooling systems. Figure 2.20b shows the method of arriving at a total cost curve for a cooling tower (CT) system and using it to continuously determine the optimized set point for the approach controller. As the approach to the ambient wet bulb temperature increases (the cooling tower water gets warmer), the pumping costs increase (because more water is needed to provide the same cooling). At the same time, the cost of operating the CT fans drops (because less air circulation is needed as the CT water gets warmer).

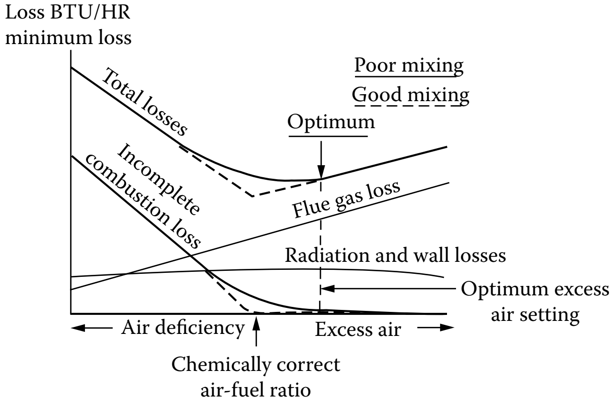


FIG. 2.20a
The optimum operating point for a combustion system is the minimum point of the total loss curve.[1]

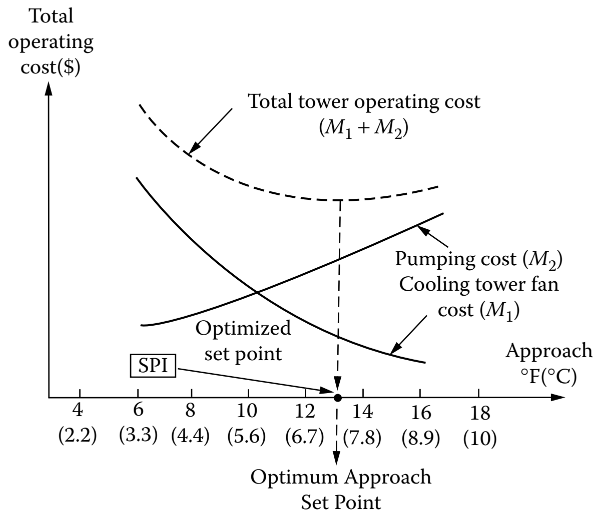


FIG. 2.20b
The set point for the optimum approach controller of a cooling tower is the point where the total system operating cost is the minimum.

Therefore, the sum of these two curves has a minimum point, which corresponds to the minimum cost of operation for the system.

Finding Minimum and Maximum Points

Any process where the operating cost has two components with opposite slopes relative to the load will have a total operating cost curve, with a minimum point. Chillers and heat pumps, for example, are in this category. The optimization of such devices always involves the searching for that minimum point and then making that point the set point of the associated controller.

The slope of the cost curve is zero at the minimum point on the curve. Because the optimization is often performed by differentiation and because the slope of the tangent at the minimum point is zero, one must be very careful to protect such optimization systems from noise and to take into account the low accuracy of differentiators.

The methods of searching for minimum and maximum points include both continuous and sampling techniques. This section will also discuss multidimensional optimization, for use when the optimum condition of a process cannot be described by a curve (two dimensions) but instead requires the climbing and evaluation of multiple peaks.

Optimizing control is essential in industrial processing today if an enterprise expects to retain viability and competitive vigor. The growing scarcity of raw materials aggravated by the dwindling of energy supplies has created a need to achieve the utmost in productivity and efficiency (Figure 2.20c).

"Optimization" means different things to different people. To some, it means the application of advanced regulatory control. Others have obtained impressive results from applying multivariable noninteracting control. To be sure, these latter techniques have merit and may actually comprise a subset of optimization.

It is the purpose of this section to guide the practicing process control engineer in a suitable course of action when

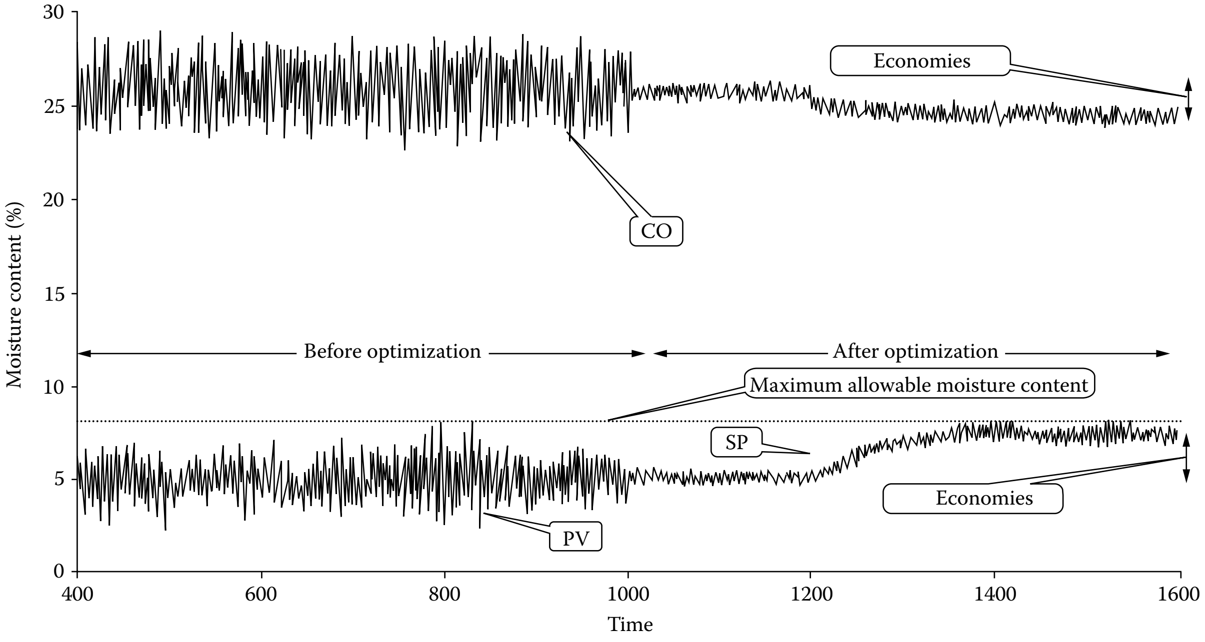


FIG. 2.20c
Minimizing energy consumption and increasing set point closer to the specification in a drying process.

confronted with a problem involving optimization. Topics emphasized include:

1. Defining the problem scope
2. Choosing between feedback or feedforward optimization
3. Mathematical tools available
4. Multivariable non-interacting control
5. Identifying and handling constraints

An extensive reference list and bibliography will be useful in amplifying the concepts and supplying details for the topics covered.

OPTIMIZATION CONSIDERATIONS

Optimization provides a management tool for achieving the greatest possible efficiency or profitability in the operation of any given production process. Changes in the operational environment, consisting of the current constraints and values for the disturbance variables, will inevitably alter the optimal position. Hence, optimizing control must be able to cope with change.

Perhaps the most difficult task in the design of an optimization control system is in the definition of the problem scope and in the subsequent choice of optimizing tactics. The need for an online optimizing system can only be ascertained following an in-depth feasibility study. It is the purpose of this section to provide some insights into the resolution of the above problems.

Processes that are the best candidates for optimizing control have these characteristics:

1. Multiple independent control parameters
2. Frequent and sizeable changes in disturbance variables affecting plant profitability
3. A required and necessary excess in the number of degrees of freedom in the independent control parameters

Figure 2.20d attempts to illustrate, in a simplistic fashion, the interrelationships that exist among the process system variables. The independent variables, $X(i)$, represent those variables that can be controlled by conventional regulatory systems. The disturbance variables, $Y(i)$, represent variables over which little or no control can be exercised, such as ambient conditions, feedstock quality, and possible changes in market demand.

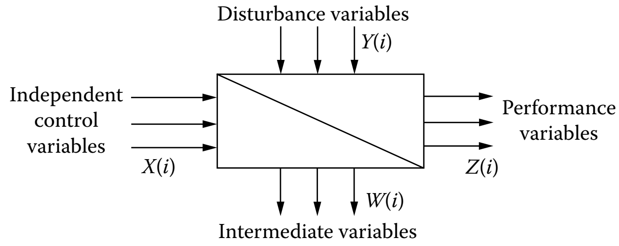


FIG. 2.20d
Process variable interrelationships.

The intermediate variables, $W(i)$, are those that describe certain complex and calculable process conditions, such as internal reflux rates, rates of conversion in reactors, and tube metal temperatures. The intermediate variables can be, and often are, constraining factors in the process operation. The performance variables, $Z(i)$, represent the objective or target values for the process, such as yields, production rates, and quality of product.

Once the scope of the problem has been defined, a choice between the strategies of either local or global optimization can be made. The local optimization problems, such as optimal boiler fuel/air ratio control, can be implemented using simpler control algorithms and hardware. On the other hand, for complex assemblages of equipment with a high degree of interaction in their operation, global optimization may be required.

Feedback or Feedforward Optimization

The feasibility study will normally indicate whether feedforward or feedback control will be more effective in optimizing the process operation. As a general rule, feedback optimization is used only in connection with local optimization problems. Feedforward, on the other hand, is eminently suited for larger, more global control problems.

Feedback optimization control strategies are always applied to transient and dynamic situations. Evolutionary optimization is a good example. Steady-state optimization, on the other hand, is widely used on complex processes, if they have long time constants and infrequently changing disturbance variables. Hybrid strategies are also employed in situations involving both long-term and short-term dynamics. Obviously, hybrid algorithms are more complex and require custom tailoring for a truly effective implementation.

Feedback control implies transitory, and the response time is mostly related to the dead time. Also, in feedback systems, when a disturbance occurs, during the time while the process is being brought back to set point, the errors could become unacceptable. Feedforward control, on the other hand, requires modeling and advanced mathematical tools, but the resulting control is better and tighter.

Reducing Set Point Variability

Variability is a measure of the range of variance in the controlled variable. It is expressed as a percentage of the mean value of the controlled variable, and this percentage allows one to compare the different processes. (For further information refer to the section on Statistical Process Control in this chapter.)

The equations for determining mean and standard deviation and to calculate variability based on them are given below:

Mean: $$\mu = \sum_{i=1}^{i=n} \frac{x_i}{n} \qquad \textbf{2.20(1)}$$

Variance: $$\sigma^2 = \sum_{i=1}^{i=n} \frac{(\mu - x_i)^2}{n-1} \qquad \textbf{2.20(2)}$$

Standard deviation: $$\sigma = \sqrt{\sigma^2 = \sum_{i=1}^{i=n} \frac{(\mu - x_i)^2}{n-1}} \qquad \textbf{2.20(3)}$$

Variability: $$\text{var} = \frac{2\sigma}{\mu} * 100 \qquad \textbf{2.20(4)}$$

Based on a normal distribution, the variability of data represents the area, which includes 95% of all points, the average ±2 standard deviations.

In a process control system, the goal is not to eliminate the variability of the controlled variable but to move it so that it can be directed. For example, consider a continuous drying process, where the energy used is steam. The goal is to deliver the end product at constant moisture content using the minimum energy and operating within the limits specified.

If the production rate varies, the total amount of drying energy needed also has to change. The total amount of drying energy needed also has to change if the temperature of the process material varies, or if disturbances occur in ambient conditions, steam quality (energy content depends on steam pressure and temperature), etc. The tuning of the moisture controller must consider all of these disturbances in order to ensure that they will not cause excessive variations in the moisture content of the product.

As shown in Figure 2.20c, if the variability of the product moisture content is reduced by optimization, the moisture controller's set point can be increased and the product quality will still be acceptable. As the variability is reduced, not only the product uniformity but also the amount of energy needed to produce it will improve. As illustrated in Figure 2.20c, if the price per ton of product remains the same, the profitability of the operation will be improved by reducing variability.

Evolutionary Optimization (EVOP)

Feedback control in certain situations can achieve optimal plant performance. Evolutionary optimization, or EVOP, is one such technique using feedback as the basis for its strategy. EVOP is an online experimentor. No extensive mathematical model is required, since small perturbations of the independent control variable are made directly upon the process itself. As in all optimizers, EVOP also requires an objective function.

EVOP also suffers from certain limitations. These include the need for the process to be able to tolerate some small changes in the major independent variable. Second, it is necessary to apply EVOP or feedback control to perturb a single independent variable at a time.

If in a particular process, two independent variables have major influences on the objective of optimization, then it may be possible to configure the controller to examine each variable sequentially in alternate sampled-data periods. This latter approach is feasible only if the process dynamics are rapid in comparison with the frequency of expected changes in the disturbance variables.

EVOP has been successfully used to maximize the thermal efficiency of industrial boilers when the fuel/air ratio was controlled by an oxygen-trim control system. In that application, EVOP was used to perturb the oxygen set point in evaluating the actual operation. The objective function consisted of an online calculation of the thermal efficiency (Figure 2.20a).

Feedforward Optimization

If numerous independent variables that affect the process performance of multivariable processes exist, they can best be optimized by the use of feedforward optimization. An absolute necessity for doing this is an adequate predictive mathematical model of the process. Such models are not used to perturb the process, but only to evaluate the consequences of changes in process variables.

In order to serve process optimization, the mathematical model must be an accurate representation of the process. To ensure a close to one-to-one correspondence with the process, the model must be updated prior to each use. Model updating is a specialized form of feedback operation in which the predictions of the model are compared with the actual operation of the plant. Any variances noted are then used to adjust certain key coefficients in the model to make it more representative of the actual process.

Figure 2.20e is a signal-flow block diagram of a computer-based feedforward optimizing control system. Process

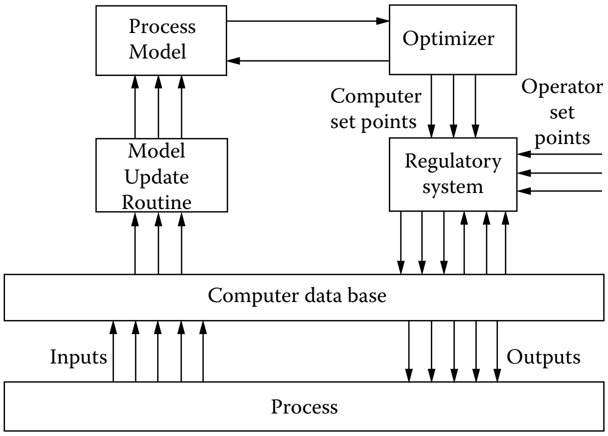


FIG. 2.20e
Block diagram of a feedforward optimizer.

variables are measured, checked for reliability, filtered, averaged, and stored in the computer database. A regulatory system is provided as a front-line control to keep the process variables at a prescribed and desired slate of values. The conditioned set of measured variables is compared in the regulatory system with the desired set points. Errors detected are then used to generate control actions that are transmitted to final control elements in the process.

The set points for the regulatory system are derived either from operator inputs or from the outputs of the optimization routine. Note that the optimizer operates directly upon the model in arriving at its optimal set-point slate. Also note that the model is updated by means of a special routine just prior to use by the optimizer.

The feedback update feature ensures adequate mathematical process description in spite of minor sensor or instrumentation errors and, in addition, will compensate for discrepancies arising from simplifying assumptions incorporated into the model.

The mathematical techniques and operating characteristics of some feedback optimization systems will be discussed more fully in the paragraphs that follow.

OPTIMIZING TOOLS

Calculus of variations is a classical mathematical approach that can be used to optimize the operation of dynamically changing processes. Though this technique has not found widespread use in industry, it should be considered in any batch chemical reactor problem in which the time–temperature–concentration relationship must be optimized. Theoretical considerations and some application data may be found in References 2 through 5.

Economic dispatch or optimal load allocation is a technique directed to the most effective use of the capabilities of parallel multiple resources to satisfy a given production requirement. As an example, the problem of steam load allocation among several boilers in an industrial utilities plant is quite frequently encountered.

A typical boiler efficiency curve is shown in Figure 2.20f. For all practical purposes, the curve may be represented by a quadratic equation.

$$\eta = \eta_0 - K(S - S_0)^2 \qquad \textbf{2.20(5)}$$

where η = efficiency at given steam load, S; η_0 = maximum efficiency at S_0; K = constant; and S, S_0 = steam loads (any consistent units, such as tons/hour).

The cost of producing a given amount of steam is defined by

$$C = \frac{F \times S}{\eta} \qquad \textbf{2.20(6)}$$

where F = cost per unit of steam.

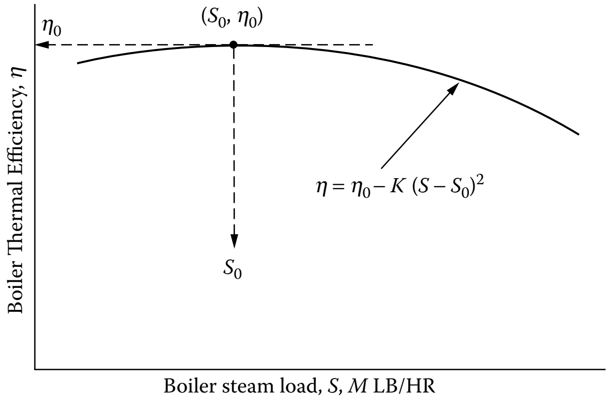


FIG. 2.20f
Typical boiler efficiency curve.

If the steam can be generated by multiple sources, use of calculus indicates an optimum when the incremental costs are equal for all boilers involved. Mathematically,

$$\frac{dC_1}{dS_1} = \frac{dC_2}{dS_2} = \frac{dC_3}{dS_3} \cdots \qquad \textbf{2.20(7)}$$

Also, to satisfy the total demand in the plant for steam, the following constraint equation must be satisfied:

$$S_1 + S_2 + S_3 + \cdots = S \qquad \textbf{2.20(8)}$$

Substituting Equation 2.20(1) into Equation 2.20(2) and then differentiating with respect to steam load S gives, for a given boiler,

$$IC = \frac{FK\left[\left(\eta_0 K - S_0^2\right) + S^2\right]}{[\eta_0 K - (S - S_0)^2]^2} \qquad \textbf{2.20(9)}$$

where IC = incremental cost.

In the above equation, the parameters η_0, K, and S_0 must be evaluated on the basis of experimental test data for each boiler in the system. Figure 2.20g is a plot describing a hypothetical three-boiler system showing a graphical solution to the load allocation problem. The horizontal line intersecting the incremental cost curves satisfies the plant load demand expressed in Equation 2.20(8).

Linear Programming

Linear programming is a mathematical technique that can be applied provided that all the equations describing the system are linear. In the field of industrial process control, the number of truly linear systems is limited, but one can occasionally be encountered. Simple physical ingredient blending problems such as in cement, glass, and certain alloy manufacturing plants are examples of these. A linear program can be

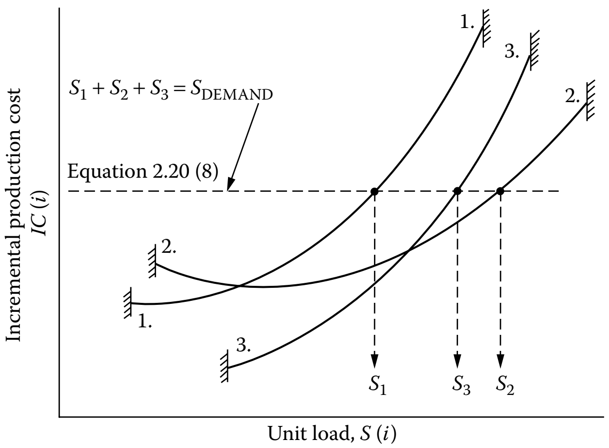


FIG. 2.20g
Graphical solution of an optimal load allocation problem.

used as an optimizing method if the problem can be reduced to the following set of relationships:

$$C = C_1X_1 + C_2X_2 + \cdots C_nX_n \qquad \mathbf{2.20(10)}$$

$$\left|\begin{array}{l} a_{11}X_1 + a_{12}X_2 + \cdots a_{1n}X_n = b_1 \\ a_{21}X_1 + a_{22}X_2 + \cdots a_{2n}X_n = b_2 \\ \quad\cdot \qquad\quad \cdot \qquad\quad \cdot \quad \cdot \\ \quad\cdot \qquad\quad \cdot \qquad\quad \cdot \quad \cdot \\ a_{m1}X_1 + a_{m2}X_2 + \cdots a_{mn}X_n = b_m \end{array}\right| \qquad \mathbf{2.20(11)}$$

Subject to:

$$\begin{array}{ll} X_i \geq 0 & i = 1, 2, 3, \cdots n \\ b_i \geq 0 & i = 1, 2, 3, \cdots n \end{array} \qquad \mathbf{2.20(12)}$$

where X = independent variables; C= associated cost factors; A = linear coefficients; and b = specified variables.

The b variables can represent specific desired concentrations, mass balance, or heat balance requirements. Note that the constants a or C can assume any value, including zero.

The objective function will either be maximized or minimized depending upon whether it represents profit or costs. Many excellent textbooks outline the procedures and algorithms used in the solution of a linear program.[6–9]

Nonlinear Programming

Most industrial processes, especially in the chemical and petroleum industries, cannot be described by linear equations over their complete operating range. These processes therefore require some type of nonlinear optimizer to achieve maximum profitability. Two types of nonlinear optimizers — the sectionalized linear program and the gradient search — have been successfully implemented in advanced computer control schemes.

The sectionalized linear program is especially useful for those processes that exhibit slight-to-moderate nonlinearities in their variable relationships. By restricting the permissible range of excursion for each independent variable, the small amount of nonlinearity can be assumed to be zero. In essence, the coefficients c and a in Equation 2.20(10) and Equation 2.20(11) are replaced by their partial derivatives. Thus, in general, the following approximations are made:

$$a(i, j) = \left(\frac{\delta b(i)}{\delta x(j)}\right)_{x(j)} \qquad \mathbf{2.20(13)}$$

Note that the partial derivatives are evaluated at the current value for the independent parameter $x(j)$. The resultant matrix or tableau of partials is used in the linear program to arrive at an interim solution. If the current or interim optimum is greater than the last value obtained, the whole procedure is repeated by "relinearizing" the process at the newly defined operating point. Examples utilizing the above technique may be found in References 10, 11, and 12.

An alternative to the sectionalized linear program is the use of a gradient or "hill-climbing" approach. The technical literature is replete with descriptions of the mathematical bases for numerous gradient optimization methods.[13–16] Likewise, many successful industrial applications have been reported and reviewed.[17–19]

In contrast to the linear program approach, which considers only one variable at a time in its serial and sequential search for the optimum, the gradient methods generally simultaneously perturb all the independent variables. The magnitude of perturbation applied to each variable is directly proportional to the direction cosine of that variable. Mathematically,

$$\frac{U \cdot \dfrac{\delta c}{\delta x(i)}}{\sqrt{\displaystyle\sum_{i=1}^{i=n}\left(\frac{\delta c}{\delta x(i)}\right)^2}} \qquad \mathbf{2.20(14)}$$

where U = unit step magnitude; c = value of objective function, such as profit; and $x(i)$ = the ith independent control variable.

Most gradient methods, once a path of steepest ascent has been established, continue to move along that path until no further improvement in the objective function is obtained. At this juncture, another gradient is established, and the entire system proceeds in this new direction until another ridge is encountered.

CONSTRAINT HANDLING METHODS

Optimization of a constrained nonlinear process requires a mathematical algorithm for searching along a constraint boundary. The two successful approaches that have been implemented are "hemstitching" and the use of penalty functions.

Figure 2.20h is a hypothetical representation of a two-dimensional process whose permissible operating range is

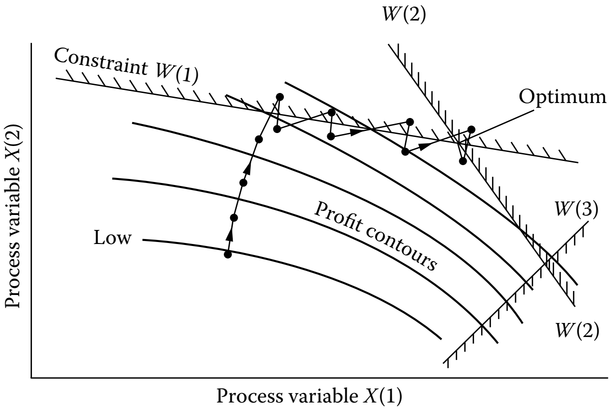


FIG. 2.20h
Gradient search using hemstitching for constraint handling.

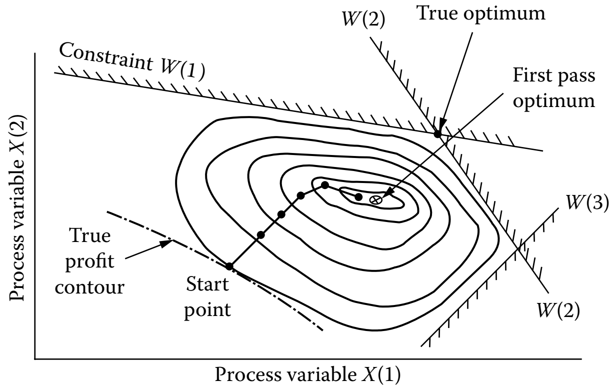


FIG. 2.20i
Gradient search using created response surface (first pass).

constrained by three variables $w(1)$, $w(2)$, and $w(3)$. Parametric lines of the objective function (profit) are also plotted. Although the lines of constraint are portrayed here as linear, in most circumstances they will be markedly curved.

Gradient Search Method

Beginning at the start point, the gradient search method proceeds up the hill in the direction of the steepest ascent. As soon as a constraint function $w(1)$ is violated, the algorithm must return the computed operating position to some point within the feasible region. Generally, this is accomplished by moving perpendicularly or normal to the constraint function. The slope of the constraint is evaluated from the model by two perturbations of $x(1)$ and $x(2)$. Thus:

$$\text{slope} = \left(\frac{\delta x(2)}{\delta x(1)} \right)_{w(1)} \qquad \mathbf{2.20(15)}$$

where the partials are evaluated at the particular constraint boundary. Now the quickest way to return to the feasible region is in the direction of the normal:

$$\text{normal} = \frac{-1}{\text{slope}} \qquad \mathbf{2.20(16)}$$

Once the constraint has been recognized, the algorithm must attempt to move along this boundary.

The method of a created-response surface suggested by C. W. Carroll[20,21] avoids many of the constraint searching problems. In this approach, the objective function is multiplied by a series of penalty terms, one for each active constraint. Thus, along each constraint boundary, the profit contour assumes a value of zero.

A typical response surface is shown in Figure 2.20i. Note that a fictitious hill is created whose summit is within the feasible operating range. Standard gradient search methods are then used to find the peak. Obviously, the peak is not at the true optimum as defined by the intersection of the constraints. If the true value of the objective function is greater than the value at the starting point, then a recentering is employed.

The profit contours passing through the peak are assigned a value of zero. This, along with the zero-valued constraints, permits the creation of a second hill. See Figure 2.20j. Note that the summit of the second hill is much closer to the true optimum. The foregoing recentering is repeated until the improvement in the true objective function is less than a desired and arbitrary value.

MULTIVARIABLE NONINTERACTING CONTROL

Many industrial processes exhibit considerable interaction among the control variables when attempting to regulate the values of the dependent variables. In general, a marked interaction problem will exist if the process can be described by the following set of mathematical relationships:

$$\left| \begin{matrix} Z(i) = fi(Xj,Yj) \\ W(i) = gi(Xj,Yj) \end{matrix} \right| \qquad \mathbf{2.20(17)}$$

for $j = 1, 2, \ldots n$ and $i = 1, 2, \ldots, m$.

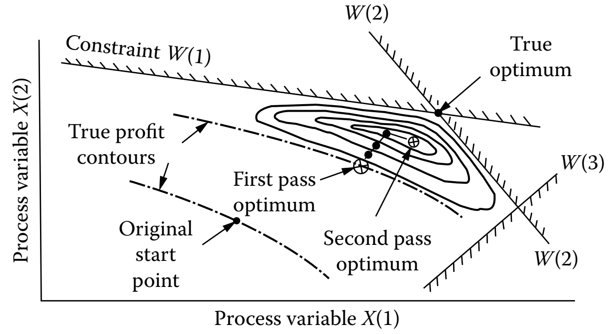


FIG. 2.20j
Gradient search using created response surface (second pass).

In the general case, the functions *fi* and *gi* are nonlinear in form and may be implicit in the independent variables *Xj* and *Yj*. This situation is very complex and requires a computer iterative solution in order to decouple the interaction effects.

However, for many applications the process can be linearized around the current operating point, thus providing a much more tractable solution to the control problem. Linearization reduces the variable interrelationships to the following mathematical form:

$$\begin{vmatrix} \frac{\delta W_1}{\delta X_1}\ \frac{\delta W_1}{\delta X_2} \cdot\cdot \frac{\delta W_1}{\delta X_n} \\ \frac{\delta W_2}{\delta X_1}\ \frac{\delta W_2}{\delta X_2} \cdot\cdot \frac{\delta W_2}{\delta X_n} \\ \frac{\delta W_3}{\delta X_1}\ \frac{\delta W_3}{\delta X_2} \cdot\cdot \frac{\delta W_3}{\delta X_n} \end{vmatrix} \times \begin{vmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_n \end{vmatrix} = \begin{vmatrix} W_1 \\ W_2 \\ W_3 \end{vmatrix} \tag{2.20(18)}$$

In the above equation set, the *Xs* are the independent control parameters and the *Ws* represent the dependent controlled variables. The matrix containing the partial derivatives is known as the sensitivity matrix. The noninteracting control problem arises when it is desirable to change the value of one dependent variable without affecting the current values of the remaining dependents. Or, alternatively, how should the independent *X* variables be set in order to achieve a desired slate of *W* values?

Using matrix algebra, the sensitivity matrix can be inverted, giving rise to the so-called control matrix:

$$\begin{vmatrix} \frac{\delta X_1}{\delta W_1} & \frac{\delta X_1}{\delta W_2} & \frac{\delta X_1}{\delta W_3} \\ \frac{\delta X_2}{\delta W_1} & \frac{\delta X_2}{\delta W_2} & \frac{\delta X_2}{\delta W_3} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\delta X_3}{\delta W_1} & \frac{\delta X_3}{\delta W_2} & \frac{\delta X_3}{\delta W_3} \end{vmatrix} \times \begin{vmatrix} \varepsilon W_1 \\ \varepsilon W_2 \\ \varepsilon W_3 \end{vmatrix} = \begin{vmatrix} \Delta X_1 \\ \Delta X_2 \\ \Delta X_3 \\ \cdot \\ \cdot \\ \Delta X_n \end{vmatrix} \tag{2.20(19)}$$

The error terms of the variable W are indicated by εWi and the required changes in control variables X are given by ΔXj.

Constraint Following

Unfortunately for plant managers, but fortunately for computers, many nonlinear constrained processes can be subject to two or more constraints at the same time. As a result of the latest optimization calculation, a slate of set points is provided to achieve the stated objective, such as maximizing profit. The optimizer will at the time identify which of the control variables should be at their limiting values.

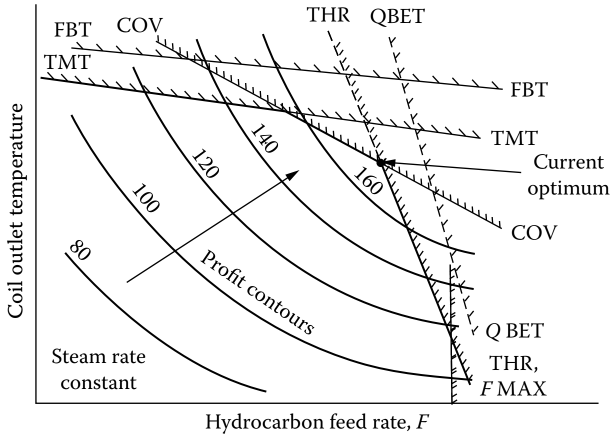


FIG. 2.20k
Typical profit surface for ethylene plant furnace cracking naphtha.

A practical example of the multi-constraint situation is depicted in Figure 2.20k, which shows the profit contours bounded by six possible constraints. For an ethylene plant pyrolysis furnace, where TMT = tube metal temperature; FBT = fire box temperature; COV = coil outlet velocity; THR = total heat release; QBET = quench boiler exit temperature; and FMAX = maximum feed availability.

As shown in Figure 2.20k, there are four active constraints that define the feasible operating region. In addition, several more constraints may become active if process loading or some other disturbance variable undergoes a change.

For the above example, visual inspection of Figure 2.20k indicates the optimum occurring at the intersection of the THR and COV constraints. To keep the process at optimum conditions in the interval between optimization calculations, a dual constraint–follower regulation scheme would be set in motion. The set points for these two controllers would be the maximum permitted values for THR and COV. Since interaction between these two constraints will occur, the multivariable noninteraction techniques discussed in the previous paragraph would be employed.

References

1. Lipták, B. G., *Optimization of Unit Operations,* Radnor, PA: Chilton, 1987.
2. Akhiezer, N. I., *The Calculus of Variations,* New York: Blaisdell, 1962.
3. Elsgole, L.E., *Calculus of Variation,* Reading, MA: Addison-Wesley, 1962.
4. Savas, E. S., *Computer Control of Industrial Processes,* New York: McGraw-Hill, 1965.
5. Bolza, O., *Lectures on the Calculus of Variations,* New York: Dover, 1961.
6. Dorfman, R., et al., *Linear Programming and Economic Analysis,* New York: McGraw-Hill, 1958.

7. Garvin, W. W., *Introduction to Linear Programming,* New York: McGraw-Hill, 1960.
8. Gass, S. I., *Linear Programming,* 2nd ed., New York: McGraw-Hill, 1964.
9. Kuehn, D. R., and Porter, J., "The Application of Linear Programming Techniques in Process Control," IEEE General Meeting Paper CP63-1153, Toronto, Ontario, 1963.
10. Corbin, R. L., and Smith, F. B., "Mini-Computer System for Advanced Control of a Fluid Catalytic Cracking Unit," ISA Paper No. 73-516, presented in Houston, October 1973.
11. Laspe, C. G., Smith, F. B., and Krall, R. A., "An Examination of Optimal Operation of an Industrial Utilities Plant," ISA Paper No. 76-524, presented at ISA 76 International Conference and Exhibit, Houston, October 1976.
12. Horn, B. C., "On-Line Optimization of Plant Utilities," *Chemical Engineering Progress,* June 1978, pp. 76–79.
13. Gabriel, G. A., and Ragsdell, K. M., "The Generalized Reduced Gradient Method: A Reliable Tool for Optimal Design," *Journal of Engineering for Industry,* Trans. ASME, series B, Vol. 99, No. 2, May 1977, pp. 394–400.
14. Kelley, H. J., "Method of Gradients," in Leitman, G., Ed., *Optimization Techniques with Applications to Aerospace Systems,* New York: Academic Press, 1962, pp. 205–254.
15. Rosenbloom, P. C., "The Method of Steepest Descent," *Numerical Analysis, Proceedings of the 6th Symposium on Applied Mathematics,* New York: McGraw-Hill, 1956.
16. Schuldt, S. B., et al., "Application of a New Penalty Function Method to Design Optimization," *Journal of Engineering for Industry,* Trans. ASME, series B, Vol. 99, No. 1, February 1977, pp. 31–36.
17. Roberts, S. M., and Lyvers, H. I., "The Gradient Method in Process Control," *Industrial Engineering Chemistry*, November 1961.
18. Schrage, R. W., "Optimizing a Catalytic Cracking Operation by the Method of Steepest Ascents and Linear Programming," *Operations/Research,* July-August 1958.
19. Laspe, C. G., "Recent Experiences in On-Line Optimizing Control of Industrial Processes," paper presented at the 5th Annual Control Conference, Purdue Laboratory, Lafayette, IN, April 1979.
20. Carroll, C. W., "The Created Response Surface Technique for Optimizing Restrained Systems," *Operations Research,* 1961.
21. Carroll, C. W., "An Approach to Optimizing Control of a Restrained System by a Dynamic Gradient Technique," Preprint 138-LA-61, ISA Proceedings of the Instrumentation–Automation Conference, Fall, 1961.

Bibliography

Badgwell, T. A., and Qin, S. J., "Review of Nonlinear Model Predictive Control Applications," in *Nonlinear Predictive Control: Theory and Practice,* Piscataway, NJ: IEE Publishing, 2001.

Balakrishnan, A. V., and Neustadt, L. W. (Eds.), *Computing Methods in Optimization Problems,* New York: Academic Press, 1964.

Balchen, J. G., and Mummè, K. I., *Process Control Structures and Applications,* New York: Van Nostrand Reinhold Co., 1988.

Bellman, R., *Dynamic Programming,* Princeton, NJ: Princeton University Press, 1957.

Blevins, T. L., Brown, M., McMillan, G., and Wojsznis, W. K., *Advanced Control Unleashed,* Research Triangle Park, NC: ISA, 2003.

Bryson, A. E., and Denham, W. F., "A Steepest-Ascent Method for Solving Optimum Programming Problems," *Trans. ASME*, *Journal of Applied Mechanics,* February 1962.

Chatterjee, H. K., "Multivariable Process Control, Proceedings of the First IFAC Congress," Moscow, Vol. 1, Butterworth, 1963.

Chestnut, H., Duersch, R. R., and Gaines, W. M., "Automatic Optimizing of a Poorly Defined Process," Proceedings of the Joint Automation Control Conference, Paper 8-1, 1962.

Chien, G. K. L., "Computer Control in Process Industries," in C. T. Leondes (Ed.), *Computer Control Systems Technology,* New York: McGraw-Hill, 1961.

Cutler, C. R., and Ramaker, B. L., "Dynamic Matrix Control — A Computer Control Algorithm," *ACC Conference Proceedings,* San Francisco, 1980.

Douglas, J. M., and Denn, M. M., "Optimal Design and Control by Variational Methods," *Industrial Engineering Chemistry,* November 1965.

Forsythe, G. E., and Motzkin, T. S., "Acceleration of the Optimum Gradient Method," Preliminary report (abstract), *Bulletin of the American Mathematical Society,* 1951.

Gore, F., "Add Compensation for Feedforward Control," *Instruments and Control Systems,* March 1979.

Harriott, P., *Process Control,* New York: McGraw-Hill, 1964.

Hestenes, M. R., *Calculus of Variations and Optimal Control Theory,* New York: John Wiley & Sons, 1966.

Hestenes, M. R., and Stiefel, E., "Method of Conjugate Gradients for Solving Linear Systems," *Journal of Research of the National Bureau of Standards,* 1959.

Himmelblau, D. M., *Applied Nonlinear Programming,* New York: McGraw-Hill, 1972.

Horowitz, I. M., "Synthesis of Multivariable Feedback Control Systems," *IRE Trans. Autom. Control*, AC-5, 1960.

Huang, H., and Shah, S. L., *Performance Assessment of Control Loops,* London: Springer-Verlag, 1999.

Kane, E. D., et al., "Computer Control of an FCC Unit," National Petroleum Refiners' Association Computer Conference, Technical paper 62-38, 1962.

Kelly, K. P., "Should You Be Training Your Operators with Simulation Software?" *Control,* June 1992.

Koenig, D. M., *Control and Analysis of Noisy Processes,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Laspe, C. G., "Optimal Operation of Ethylene Plants," *Instrumentation Technology,* May 1978.

Lefkowitz, I., "Computer Control," in E. M. Grabbe, S. Ramo, and D. E. Wooldridge, Eds., *Handbook of Automation, Computation, and Control,* Vol. 3, New York: John Wiley & Sons, 1961.

Leitmann, G. (Ed.), *Optimization Techniques,* New York: Academic Press, 1962.

Lipták, B. G., "Envelope Optimization for Coal Gasification," *InTech,* December 1975.

Lipták, B. G., "Optimizing Plant Chiller Systems," *InTech,* September 1977.

Lipták, B. G., "Save Energy by Optimizing Boilers, Chillers and Pumps," *InTech,* March 1981.

Lipták, B. G., "Envelope Optimization for Clean Rooms," *Instruments and Control Systems,* September 1982.

Liptak, B. G. (Ed.), *Instrument Engineers' Handbook: Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002.

Lloyd, S. G., "Basic Concepts of Multivariable Control," *Instrumentation Technology,* December 1973.

Luntz, R., Munro, N., and McLeod, R. S., "Computer-Aided Design of Multivariable Control Systems," IEE 4th UKAC Convention, Manchester, 1971.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Luyben, W. L., Luyben, M. L., and Bjorn, T. D., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Mayne, D. Q., "Design of Linear Multivariable Systems," *Automatica,* Vol. 9, 1973.

McMillan, G., *Tuning and Control Loop Performance,* 2nd ed., 1990, ISA, pp. 45–55.

Merriam, C. W. III, *Optimization Theory and the Design of Feedback Control Systems,* New York: McGraw-Hill, 1964.

Morely, R., "Faster Than Real Time," *Manufacturing Systems,* May 1992.

Munro, N., and Ibrahim, A., "Computer-Aided Design of Multivariable Sampled-Data Systems," IEE Conference on Computer Aided Control System Design, Cambridge, 1973.

Roberts, S. M., and Mahoney, J. D., "Dynamic Programming Control of Batch Reaction," *Chemical Engineering Progress Symposium Series,* Vol. 58, No. 37, 1962.

Roffel, B., and Chin, P., *Computer Control in the Process Industries,* Chelsea, MI: Lewis Publishers, 1989.

Ruel, M., "PID Tuning and Process Optimization Increased Performance and Efficiency of a Paper Machine," 87th Annual Meeting, PAPTAC, Book C (February 2001), pp. C63-C66.

Sandgren, E., and Ragsdell, K. M., "The Utility of Nonlinear Programming Algorithms: A Comparative Study—Part 1 and 2," *ASME Journal of Mechanical Design,* July 1980.

Shinskey, F. G., "Feedforward Control Applied," *ISA Journal,* November 1963.

Shinskey, F. G., "Adaptive Nonlinear Control System," U.S. Patent 3,794,817, February 26, 1971.

Shinskey, F. G., "Process Control Systems with Variable Structure," *Control Engineering,* August 1971.

Shinskey, F. G., "Adaptive pH Controller Monitors Nonlinear Process," *Control Engineering,* February 1974.

Van Doren, V. (Ed.), *Techniques for Adaptive Control,* Boston: Butterworth Heinemann, 2003.

Vervalin, C. H., "Training by Simulation," *Hydrocarbon Processing,* December 1984.

Wells, C. H., "Industrial Process Applications of Modern Control Theory," *Instrumentation Technology,* April 1971.

Wellstead, P. E., and Zarrop, M. B., *Self-Tuning Systems,* New York: Wiley, 1991.

Wilde, D. J., *Optimum Seeking Methods,* Englewood Cliffs, NJ: Prentice Hall, 1964.

Zahradnik, R. L., Archer, D. H., and Rothfus, R. R., "Dynamic Optimization of a Distillation Column," Proceedings of the Joint Automation Control Conference, Paper 13-3, 1962.

Zellnik, H. E., Sondak, N. E., and Davis, R. S., "Gradient Search Optimization," *Chemical Engineering Progress,* August 1962.

2.21 PID Algorithms and Programming

M. D. WEISS (1970) **T. J. WILLIAMS** (1985) **B. G. LIPTÁK** (1995, 2005)

Types of Software Packages:
A. Control system design
B. Process monitoring and control
C. Process optimization
D. Expert systems

Partial List of Software Suppliers:
ABB (C)
Allen-Bradley (B)
DMC (C)
Emerson (C)
Exper-Tune (C)
Gensym (D)
Honeywell (C)
Icom (A, B)
Iconics (B)
Invensys (C)
Matricon (C)
Setpoint (C)
T/A Engineering (B)
Techmation (C)
U.S. Data (B)
Wonderwere (B)

This section describes some of the basic PID algorithms and the early developments of process control-oriented computer languages. The more advanced expert systems utilizing artificial intelligence and other, more sophisticated methods are discussed in other sections of this chapter.

PID CONFIGURATIONS

The output of a PID controller is a function of the size of the error (proportional contribution), the time period during which the error existed (integral contribution) and of the rate at which the error is changing (derivative contribution). The standard or ideal PID configuration (also called noninteracting) is illustrated in Figure 2.21a and is described by Equation 2.21(1):

$$m = K_c\left(e + \frac{1}{T_i}\int e\,dt + T_d\frac{de}{dt}\right) + b \qquad \textbf{2.21(1)}$$

where m = controller output; e = error signal; K_c = proportional gain = 100/Proportional Band; T_d = derivative time constant (from seconds to hours); T_i = integral time constant (minutes/repeat) also called reset; and b = bias.

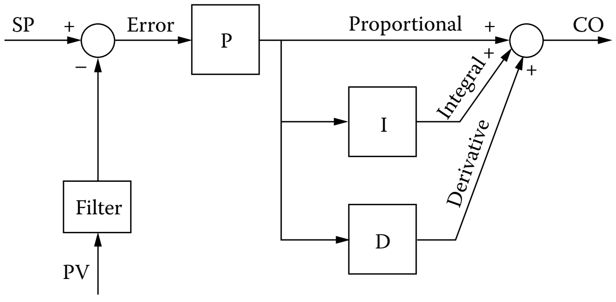


FIG. 2.21a
The standard or ideal PID configuration, which is also called noninteracting or the ISA version.

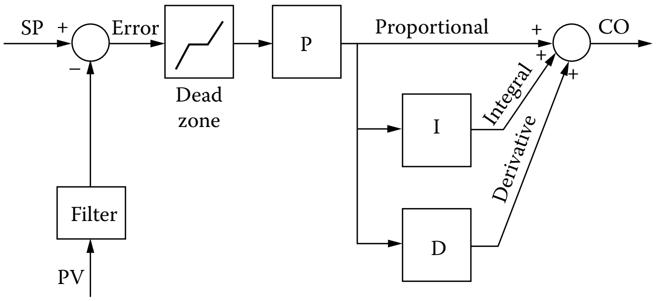


FIG. 2.21b
The gap-type PID configuration.

Equation 2.21(2) describes the interacting (also called series) and Equation 2.21(3) the parallel PID configuration:

$$m = K_c\left(e + \frac{1}{T_i}\int e\,dt\right)\left(1 + T_d\frac{de}{dt}\right) + b \qquad \mathbf{2.21(2)}$$

$$m = K_c e + \frac{1}{T_i}\int e\,dt + T_d\frac{de}{dt} + b \qquad \mathbf{2.21(3)}$$

There are a number of variations of these basic PID configurations. One option is to let some of the control modes (D or P&D) act on the process variable (PV in Figure 2.21a) and not on the error. These configurations eliminate the overshoot that otherwise can occur on set-point change. When frequent set-point changes are expected, another way to reduce overshoot is to use set-point filtering. Both of these methods allow for increasing the proportional gain (reducing the proportional band) and thereby make the controller more effective.[1,2]

Error-squared algorithms become more aggressive as the error rises. For that reason they are widely used on level control. Integral squared algorithms are effective on all variable dead time processes and cause no overshoot because when the variable is on set point, the integral contribution goes to zero. Figure 2.21b illustrates the gap-type PID controller, which does not change its output at all as long as the error is within the dead zone or gap. This configuration is well suited for controlling noisy process variables, which otherwise would cause the controller output to be unstable, reducing the life expectancy of the control valve or other final control element.

ALGORITHMS

An algorithm, by definition, is the description of a particular equation by the digital computer. It has also come to mean an equation for the implementation of a particular function on the computer. We will use the term in the second sense since we wish to describe several equations that have become particularly important in the process control field, especially for the implementation of PID control.

In DCS control, the digital computer calculates the required movement of the control valve or other final actuator. In supervisory control, on the other hand, the computer merely computes the required change in the variable's set point, and the actuation of the final control element is left to an analog controller.

Digital Algorithms

It should also be understood that when these algorithms are implemented in the digital world, what used to be integration in the analog world becomes summation, and what used to be time differential in the analog world becomes difference. The scan period of DCS systems is around 0.5 seconds or is selectable for each loop from 0.1, 0.2, 0.5, 1.0, 5.0, 10, or 30 seconds. As the DCS system does not continuously evaluate the status of each measurement, but looks at them intermittently, it increases the dead time of the loop by two "scan periods" (the time it takes to return to that measurement) and it makes its algorithm calculations on the basis of the "present" and the "previous" error. Two basic types of digital algorithms are in use. One is called positional. This means that the full output signal to the valve is recalculated every time the measurement is examined (Equation 2.21[4]).

$$m = K_c\left(e + \frac{1}{T_i}\sum_o^n e\,dt + \frac{T_d}{\Delta t}\Delta e\right) + b \qquad \mathbf{2.21(4)}$$

where Δt = scan period; Δe = change between the previous and the present value of the error; and $\sum_o^n e$ = the sum of all previous errors between time zero and time n.

The other calculation is called a velocity algorithm, and when it is used, the value of the previous output signal (m) to the valve is held in memory, and only the required change in that output signal is calculated (Equation 2.21[5]).

$$\Delta m = K_c\left(\Delta e + \frac{1}{T_i}\sum_o^n e\Delta t + \frac{T_d}{\Delta t}\Delta(\Delta e)\right) \qquad \mathbf{2.21(5)}$$

where $\Delta(\Delta e)$ = the change in the change in the error between the previous and the present scan periods.

The positional algorithm is preferred when the measurement is noisy because it works with the error and not the rate of error change when calculating its proportional correction. Velocity algorithms have the advantages of bumpless transfer and less reset windup (0 or 100%) and are better suited for controlling servomotor-driven devices. Their main limitations include noise sensitivity, likelihood to oscillate, and lack of an internal reference.

These two major approaches were initially introduced in the 1960s for digitally representing the action of conventional three-mode controllers: position and velocity algorithms.[3] They differ principally in the place where integration occurs. For the position algorithm, the computer output is the corrected valve (or other final element) position, because the integration is performed in the computer. For the velocity algorithm, the computer output is the change that the valve should undergo between sampling periods; therefore, integration must be performed by the final element via a stepping motor, integrating amplifier, or other similar device.

The Position Algorithm

The three-mode controller can be represented by

$$P_n = K_P e + K_D \frac{\Delta e}{\Delta t} + K_I \sum_0^n e\Delta t + P_M \qquad \textbf{2.21(6)}$$

$$e = S - V \qquad \textbf{2.21(7)}$$

where

P_n = valve position at time n
P_M = a median valve position
K_P = proportional-mode gain (K_c)
K_I = integral-mode gain (K_c/T_i)
K_D = derivative-mode gain (K_c/T_d)
e = error
S = setpoint
V = variable

In the usual representation of this controller by analog hardware, there is usually some interaction among the three modes as follows:

$$P_n = K_P\left(e + T_D\frac{\Delta e}{\Delta t} + \frac{1}{T_R}\sum_0^n e\Delta t\right) + P_M \qquad \textbf{2.21(8)}$$

Thus, $K_d = K_P T_D$, and $K_I = K_P/T_R$. Individual mode gains cannot be changed independently. In the process industry, the terms are

K_P = PB/100, where PB is proportional band (%)
T_D = rate time (minutes)
T_R = reset time or Ti, integral time (minutes)
$\frac{1}{T_R}$ = reciprocal time (repeats per minute)

Equations 2.21(6) and 2.21(7) are convenient to use with digital systems. However, reference must frequently be made to Equation 2.21(8) because of the nomenclature built around it.

Because of the vagaries of digital data and particularly due to the sampling of noisy data, the derivative expression in Equations 2.21(6) and 2.21(8) must be chosen carefully. A four-point difference technique, chosen because of convergence and accuracy considerations, is defined as follows:

$$\text{Let } V^* = \frac{V_n + V_{n-1} + V_{n-2} + V_{n-3}}{4} \qquad \textbf{2.21(9)}$$

where V is a variable and $n - 1$, $n - 2$, $n - 3$ denote times previous to time n. Then

$$\frac{\Delta V}{\Delta t} = \frac{\frac{V_n - V^*}{1.5\Delta t} + \frac{V_{n-1} - V^*}{0.5\Delta t} + \frac{V^* - V_{n-2}}{0.5\Delta t} + \frac{V^* - V_{n-3}}{1.5\Delta t}}{4} \qquad \textbf{2.21(10)}$$

$$\frac{\Delta V}{\Delta t} = \frac{\Delta e}{\Delta t} \text{(if the setpoint is constant)} \qquad \textbf{2.21(11)}$$

$$\frac{\Delta e}{\Delta t} = \frac{1}{6\Delta t}(V_n - V_{n-3} + 3V_{n-1} - 3V_{n-2}) \qquad \textbf{2.21(12)}$$

$$= \frac{1}{6\Delta t}(e_n - e_{n-3} + 3e_{n-1} - 3e_{n-2}) \qquad \textbf{2.21(13)}$$

The position algorithm requires that the computer recalculate the full value of the controller output at each time increment. In addition, if this value is to be transmitted to the valve positioner as an analog signal, this requires a digital-to-analog converter or the equivalent for each output.

The Velocity Algorithm

If the computer output is sent to a stepping motor or an integrating amplifier, the computer will only calculate the required change in valve position. The output is a digital pulse train. Where a stepping motor is used, the stepping motor can drive a slide wire, and the slide wire output is proportional to the correct valve position. Therefore, the combination of stepping motor (digital correction) and slide wire (analog signal to the final control element) acts as a digital-to-analog converter, and no separate digital-to-analog converter is required. An integrating amplifier will perform a similar action in an electronic loop.

The velocity algorithm output (ψ_n) is found by subtracting the outputs of two successive position algorithm calculations:

$$\psi_n = P_n - P_{n-1} \qquad \textbf{2.21(14)}$$

$$\begin{aligned}\psi_n &= K_p(e_n - e_{n-1}) + K_I e_n \Delta t \\ &\quad + \frac{K_D}{6\Delta t}(e_n + 2e_{n-1} - 6e_{n-2} + 2e_{n-3} + e_{n-4})\end{aligned} \qquad \textbf{2.21(15)}$$

The stepping motor, or its equivalent, serves to sum or integrate the position value, and

$$P_n = P_0 + \sum_0^n \psi_t = P_{n-1} + \psi_n \qquad \textbf{2.21(16)}$$

The derivative mode is optional, but some small amount of integration is always required for the velocity algorithm. The algorithm with derivative included can be written as

$$\begin{aligned}\psi_n &= K_p(V_n - V_{n-1}) + K_I(S - V_n)\Delta t \\ &\quad + \frac{K_D}{6t}(V_n - 2V_{n-1} - 6V_{n-2} + 2V_{n-3} + V_{n-4})\end{aligned} \qquad \textbf{2.21(17)}$$

Since the set-point appears only in the integral term, severe controller drift could occur unless this latter term is always included.[5]

In the discussion that follows, Equation 2.21(15) will be used as the basic control equation for the system. It can be represented as

$$\psi_{n_i} = F_i(e_i) \qquad \mathbf{2.21(18)}$$

where the subscript *i* is the index to the control loop in question. In general, where the *i* and *j* loops are considered in a cascade or a feedforward configuration, the higher index refers to the inner or downstream loop.

The velocity algorithm has a unique property. At any given sample time, the proportional and integral terms do not necessarily have the same signs as do the corresponding terms in the position algorithm. Examine the proportional and integral terms of the velocity algorithm:

$$\psi_n = K_P(e_n - e_{n-1}) + K_I e_n \Delta t \qquad \mathbf{2.21(19)}$$

These terms have the same sign when the control variable is moving away from the set point but have opposite signs when the control variable moves toward the set point (except for the first step of the variable in the direction of the set point). See Figure 2.21c for a sketch analyzing this action. As can be seen from the figure, the proportional terms can cause an oscillation as follows:

$$\begin{aligned} \text{Output } 1 &= K_p(e_1 - e_0) \\ &= -|\Delta e_1| K_p \\ \text{Output } 2 &= K_p(e_2 - e_1) \\ &= +|\Delta e_2| K_p \end{aligned} \qquad \mathbf{2.21(20)}$$

Thus, unless strongly overbalanced by the integral term, an uncorrected velocity algorithm can give the oscillatory response illustrated in the figure. An obvious remedy to this situation is to define the proportional term as follows:

$$\begin{aligned} \text{Output } P &= K_p(e_n - e_{n-1})(\text{Modified}) \\ &= K_P|\Delta e_n|(\text{Sign of the integral term}) \end{aligned} \qquad \mathbf{2.21(21)}$$

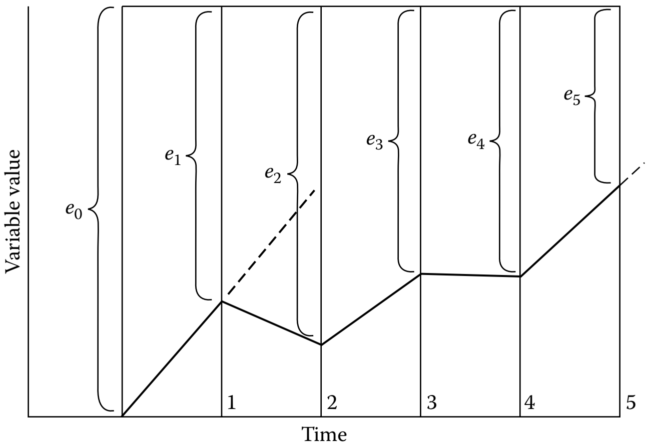


FIG. 2.21c
The action of the velocity algorithm without sign correction of the proportional term.

This property may also be used to help prevent integral overshoot and subsequent integral term oscillation. As the set point is approached, the opposing proportional term is desired for damping the integral action. When the control variable is some distance from the set point, all proportional terms are modified to have the same sign as the integral term just described.

This limited proportional action can be achieved by placing a band around the set point. Within the band all proportional contributions are accepted as calculated; outside the band all proportional terms have the same sign as the integral term. The control action can be further improved by having high proportional gain constants outside the band and much lower gain constants within the band. This allows a faster recovery from a process upset than with the conventional algorithm.

This modified velocity algorithm makes possible better control action than the positional algorithm. Analog values are scaled so that values from the analog-to-digital converters (ADC values) occupy the full computer variable range. This makes it possible to define the set-point bandwidth (S.B.) as some fraction of full scale. The following definitions and expressions can be used to modify the velocity algorithm in the described manner:

$$\begin{aligned} & e_n = \text{ADC}_{s_n}(\text{setpoint}) - \text{ADC}_n \\ & \Delta e = e_n - e_{n-1} \\ & \text{S.B.} \sim 0.07 \text{ full scale} \\ & |e_{n-1}| \le \text{S.B., add proportional term to velocity algorithm } \psi \\ & |e_{n-1}| > \text{S.B., do not add proportional term to } \psi \\ & \qquad (\text{if } \Delta e \text{ has sign different from } e_n) \end{aligned} \qquad \mathbf{2.21(22)}$$

(*Note:* ADC as used here refers to the raw count obtained by the analog-to-digital converter without correction to engineering units.)

The set-point band of 7% full scale gave good results on simulation tests of systems with first- and second-order time constants. Better control or finer tuning could be obtained if the set-point band could be adjusted for each loop since high inertia systems need wider bands to get more damping of integral action. Figure 2.21d illustrates Equation 2.21(22).

Advantages and Disadvantages

Position and velocity algorithms each have some specific advantages. The major benefit of the position algorithm is that it maintains its own reference value in the term P_M of Equation 2.21(8). Thus it need not depend upon an external

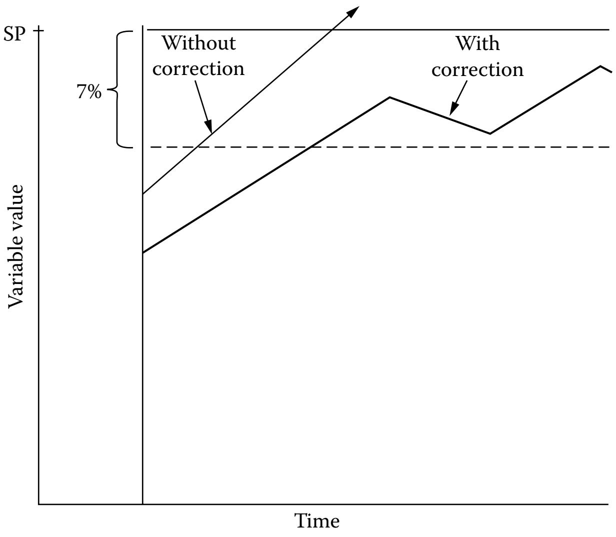


FIG. 2.21d
Effect of set-point band adjustment on velocity algorithm response.

device such as a stepping motor or integrating amplifier for a reference value. The position algorithm has two major drawbacks, however, which are thus advantages for the velocity algorithm.

These are its susceptibility to reset windup and its lack of bumpless transfer. Reset windup is the saturation of the integral mode of the position algorithm if the input is connected and the output is disconnected (such as in the test mode). Then a repeatedly computed correction term, without a corresponding process reaction to correct the error, will eventually cause the integral term to increase to its limit (saturated or "windup"). This will result in a major upset when the output is finally connected. Since it has no internal integration, the velocity algorithm does not exhibit this problem. This problem can, of course, be corrected for the position algorithm by activating the integral term only when the output is connected. This latter modification does impair the testing capabilities of the system, however.

Whenever a sudden and major change is made in the external set point of the process, a large error is detected by the algorithm. Unless special precautions are taken, this will cause a correspondingly large correction to be computed, with a resulting "bump" to the process. This can occur when the loop is initially closed if care is not taken to adjust the set point to be close to the output value of the loop. The velocity algorithm is not susceptible to this problem since its output is a pulse train that can be readily limited to some value that is within the capabilities of the process to respond.

Unlike the position algorithm, the velocity algorithm does not have an internal reference. A second major drawback to the velocity algorithm is its oscillatory response to large inputs.

An important point to keep in mind when using these or any other control algorithm is the effect of process sampling rate on systems gain, i.e., the effective value of K_P, K_I, or K_D, for example. This can be easily understood if one considers the following example:

Suppose that one has effectively tuned a control loop at some specified sampling rate. Suppose then that the sampling rate were doubled. Since the process response rate has not changed, the amount of error detected by the controller has doubled (twice as many readings of the same average values). If the same computations are made (i.e., same controller gains), the amount of correction sent to the valves will also be doubled. Therefore we can say that overall control loop gain is a direct function of process sensor sampling rate.

Cascade and Ratio Control

In analog control the primary or master controller computes the set point of the secondary controller in essentially the same manner as if its output signal were being sent to the final control element (see Figure 2.21e). This configuration

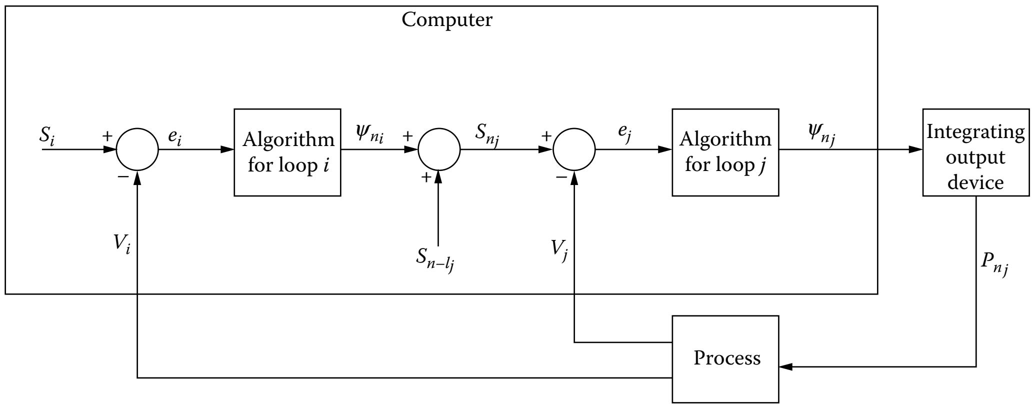


FIG. 2.21e
Cascade control.

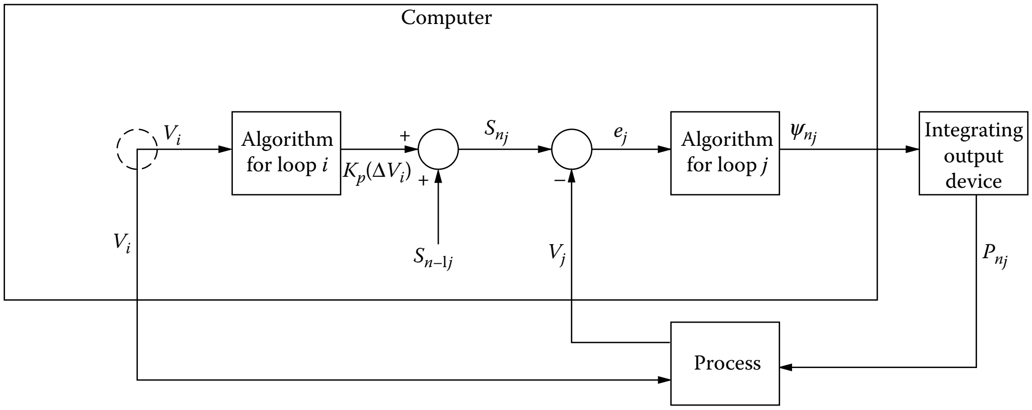


FIG. 2.21f
Ratio control.

is called cascade control. In terms of the developed nomenclature it may be expressed as follows:

$$S_{n_j} = S_{n-1_j} + \psi_{n_i} \tag{2.21(23)}$$

Another related method of process control is called ratio control, expressed as

$$S_{n_j} = RV_{n_i} \tag{2.21(24)}$$

where R is the ratio constant. However, this is also equivalent to

$$S_{n_j} = K_p V_{n_i} \tag{2.21(25)}$$

for the positional and to

$$S_{n_j} = S_{n-1_j} + K_P(V_{n_i} - V_{n-1_i}) \tag{2.21(26)}$$

for the velocity algorithm. Therefore, ratio control can be considered as cascade control where the set-point value S_i is numerically equal to zero (see Figure 2.21f).

Feedforward Control

Feedforward control looks at load changes, anticipates their effect on the process, and compensates for these effects before they have evolved. Feedforward improves control by inserting a function into the control loop that closely duplicates the process response to load change and final control element position.[4,5] Corrections, if applied too early, will introduce too much anticipation, resulting in over-control and oscillation.

Process models must be expressed in the form of transfer functions or differential equations because of the dynamic functions. Exact duplication of analog feedforward on the control computer would require integration, imposing a severe load on computer time. This makes it desirable to devise an approximation method that will adequately describe the process but will not overload the computer. The method described below approximates the open-loop step response of the process by a time delay followed by a ramp.

The response of chemical process systems to a step input can be approximated by one of the following transfer functions:[4,6]

$$\frac{Ke^{-T}DT^{s}}{(\tau_1 s + 1)(\tau_2 s + 1)} \tag{2.21(27)}$$

$$\frac{Ke^{-T}DT^{s}}{(\tau s + 1)^n} \tag{2.21(28)}$$

where K is the process gain constant, T_{DT} is its dead time, and τ, τ_1, and τ_2 are time constants of the process.

Other Digital Algorithms

While the algorithms described above are the basic ones, many others are used in implementing microprocessor-based systems. Some of these are described below, and many others are described in other sections of this chapter.

Dead-Band Control Often the establishment of a dead band about the set point (Figure 2.21b) is desirable to reduce the effects of noisy computer inputs. For example, consider level control of a reboiler on a distillation column. Surface turbulence caused by boiling makes the level signal quite erratic within a band around the set point. Dead-band control would be implemented as follows:

If $|e| = |(S_j - V_j)| < D$ (dead band), then $\psi_{n_j} = 0$

If $|e| >$ D, then use Equations 2.21(6) or 2.21(13)

Emergency Response In many plant emergency situations the correct valve action can be specified ahead of time. If a process alarm system or the computer detects an emergency, the necessary corrective action can be taken automatically.

This is easily handled by holding in memory a group of alternate set points for the controlled process variables, including zero and full open. When the emergency situation is detected, the computer immediately substitutes a new set point for the previous "normal operational" value, taking corrective action through the standard control algorithm.

Error Squared A method of effectively accomplishing the higher gain when away from the set point, as discussed earlier, is to use a square value of the error as part of the calculation of either the proportional or integral modes. This effectively achieves Equation 2.21(21) since squared values are always positive, as is the absolute value. At the same time, the higher value effectively acts as a higher gain on the original error value.

The technical literature from digital control system suppliers is a useful source of additional information on available digital control algorithms.

PROCESS CONTROL PROGRAMMING

The entire third volume of the *Instrument Engineers' Handbook* is devoted to the subject of control software and programming. For this reason, the subject is discussed only briefly here. The purpose of computer programming is to individualize the functions carried out by a particular computer system.

Programming for process control tended to lag somewhat behind the state-of-the-art programming for scientific and business applications because of number of process control applications is smaller. Nevertheless, the high cost of developing programs for specific applications from scratch has led to many developments. New programs tend to be developed from old ones that have been used previously for similar functions.

There is also a trend to make all programs as general as possible in order to enhance their use for more applications. By reducing the individuality of a program, suppliers also reduce the cost of development. In fact, a major segment of the computer industry does nothing else but produce these general programs.

Types of Programs

The desire for minimum individuality in process control computer programs has resulted in a program organization that consists of only three major parts:

1. The *operating system* or *executive program*
2. The *application program*
3. The *system support* software

The executive program consists of all the programs that supervise the overall operations of the computer system. The functions it performs are:

1. Scheduling and actually starting the execution of the application programs
2. Operating all hardware (e.g., allocating main memory to specific functions, loading programs into main memory from external bulk memory)
3. Supervising input/output operations
4. Servicing the priority interrupt system
5. Loading analog and digital inputs into memory
6. Controlling outputs to the plant actuators

Applications programs handle all of the specialized functions required for that particular installation. These functions may include:

1. Conversion of plant input data to engineering units
2. Scheduling, optimization, and control correction computations
3. Operator's console displays, logging, and other management presentations

Even though applications programs are individualized for each installation, every effort is made to use prewritten, general programs whenever possible and to minimize the contribution that is required by the user.

The system support software items are those programs that help the user to prepare the application programs. They include the following:

1. Assemblers and compilers that convert programs written in a specific assembly language or a higher-level language such as FORTRAN into the machine language of the computer.
2. Editors, linking-loaders, and similar packages, which allow segments of programs written separately to be incorporated into a single program in the computer.
3. Programs that help debug applications programs. These programs may not be carried in the computer's memory at all times but may be entered when needed from external storage (disk or tape).

One technique is *host compiling* or *host programming.* Here a large computer contains in its programming system an *emulator* of the smaller computer for which software is being developed. The emulator lets the large computer behave just like the smaller one. Since the larger computer is faster, has a larger memory, and is capable of utilizing more complex and capable system support software, it can develop better programs for the smaller computer than the latter could by itself. The machine language program developed by the large computer is then read into (downloaded to) the small computer as its object program. Thus the small computer need never have any system support software.

Features of Process Control Programs

Computer systems used for process control differ from those used for general scientific and business purposes in that they contain features that allow them to operate in a real-time environment. That is, process control systems perform time-relative operations that are governed by a real-time clock, and they respond to other externally generated occurrences through an external or priority interrupt system. They must also be able to read the values of external variables and transmit signals to external devices, including to human-interpretable systems such as an operator's console.

Obviously, the process control program used in the system must allow for these functions. Accommodating these functions comprises the major difference between programming in the process control field and programming in general scientific and business fields. Because the capability for priority interrupt vitally affects the overall management of the computer system, this function is usually included within the executive program of the process control computer. The other time-based and externally stimulated functions are included in the application program of the system.

A second major difference between process control programs and typical scientific or business programs is the great dependence upon *multiprogramming* or *parallel execution* functions. The time-based operation of the control computer and the large amount of time required to complete many functions (such as printing reports and reading process variables), and the necessity to check the status of many tasks means that the computer must shift its attention between many different functions while completing only small parts of any one function at any one time.

To keep track of the ongoing status of each of the several tasks under way at any one time requires a very sophisticated executive program. All successful process control executive programs must have this capability, while only the most sophisticated scientific and business systems require the computer sizes that are used for process control.

Figure 2.21g illustrates the operations that are executed by programming in the process computer system. The diagram shows the overall system as carried out by a single computer containing all functions. The modular system shown in the diagram allows any particular module to be modified without affecting any of the other modules. This greatly simplifies both the initial programming effort and any later program modifications. This is made possible by the use of the data tables indicated in the diagram. A further advantage of such a program system is that programs developed by others for any of the modules can be readily integrated into the overall program. The chance of finding a suitable existing module is obviously much greater than that of finding an existing overall program for any particular application.

The Executive Program

The executive program apportions computer time to tasks that must be carried out by the computer according to their relative urgency.[7] Response-critical events occurring online must be taken care of at once, resulting in priority tasks execution. When no such urgency exists, the executive program assigns time to background operations.

The executive software function is basically a timesharing one. It resolves process situation conflicts, switching computing facilities back and forth between critical and noncritical demands. Figure 2.21h is a schematic representation of this timesharing function, in which available computer time is

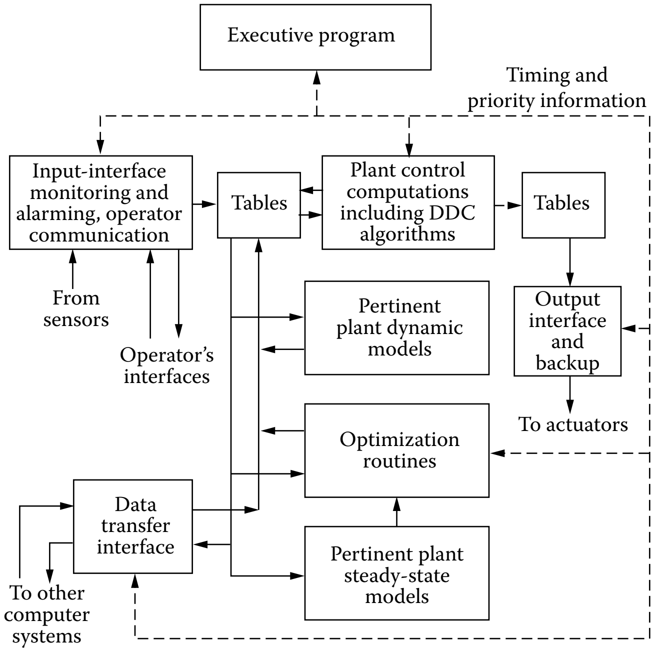


FIG. 2.21g
Modular process control programming system.

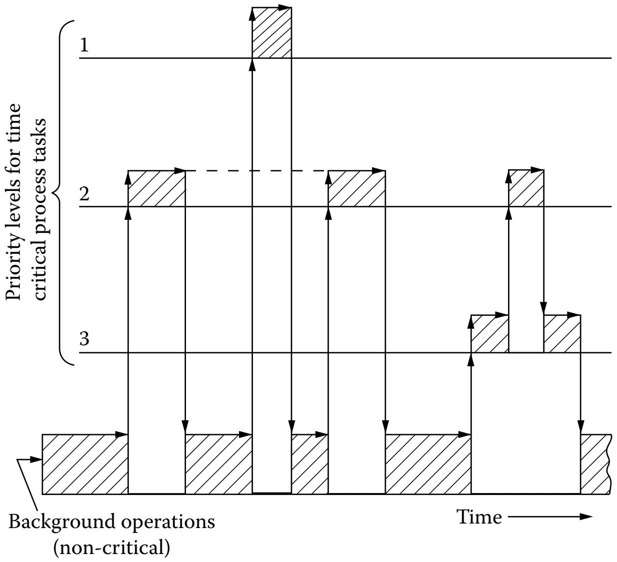


FIG. 2.21h
Operation of the executive program's timesharing function.

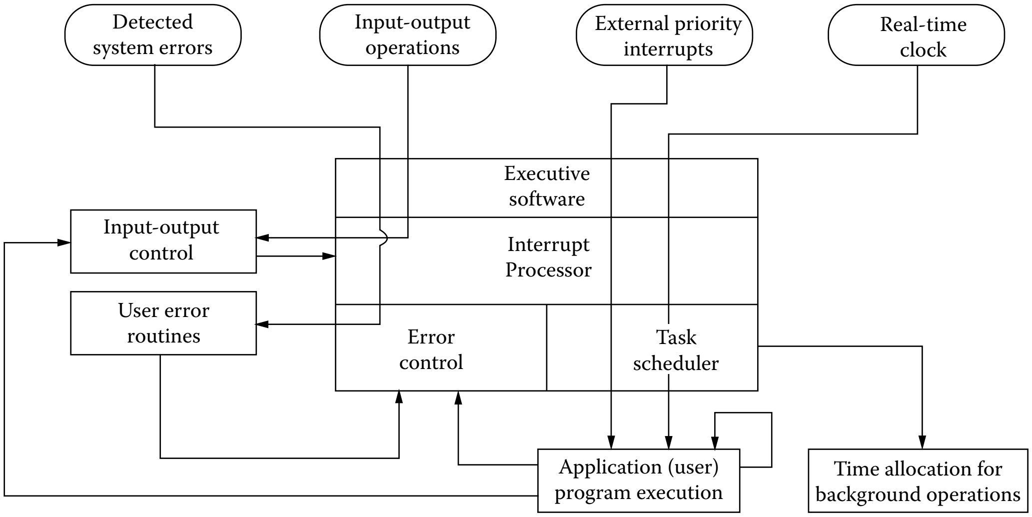


FIG. 2.21i
Principal executive software functions and lines of communication with external events or other software.

allotted among events of varying priority. Preestablished priority rules are used to determine which situation is attended to at any point in time. As process tasks and problems are resolved, computer time is returned to noncritical operations.

The principal programs contained in the executive software are the interrupt processor, task scheduler, and error control. Figure 2.21i shows these functions, as well as other sources of data or routines with which the executive program regularly communicates.

Interrupt Processor and Task Scheduler The *interrupt processor* plays a central role in the control system. Events in the outside world that require response usually appear as external interrupts. The interrupt processor must be capable of sensing the degree of urgency and assigning computer time in proper order of priority. At the same time, the internal registers and the status of an interrupted program must be "saved" to properly restore service.

The *tasks scheduler* must accommodate random interrupts as well as the periodic monitoring of live input values to determine if present limits have been violated or if control corrections are necessary. Thus the executive software must control the use of hardware timer interrupts (the real-time clock) and provide a periodic response to the user program. Aside from controlling periodic monitoring, task scheduling is usually carried out in response to a random demand for program execution. Simple algorithms are often used to assign priorities to these demands. But more sophisticated applications require that the user vary priorities. In some cases user software may dynamically assign priorities that take into account complex combinations of process events.

Error, I/O and Diagnostics *Error control* is probably the most difficult function in the executive software. This function is responsible for the proper handling of errors originating either in hardware or created by software. Some errors can be handled with no response from the user; others require user intervention to minimize the effect on the process being controlled. In extreme cases, error handling requires computer control backup to prevent a catastrophe. In general, error handling routines initiate actions that are unique to each application.

The principal purpose of the *input-output* (*I/O*) *software* is to optimize the use of peripheral devices, resolve conflicts arising from concurrent requests for their use, and provide a simple interface with peripheral and process equipment for programming ease. I/O devices may be peripheral to the computer, such as printers and tape readers, or they may be functional units in the process.

The best *hardware diagnostics* can often be built into the user program. For example, input A/D can be verified for reasonableness prior to being applied to the actuation of an output device. Again, based on availability of output member feedback, the program can require verification that the output motion has actually taken place.

In many process control applications, down time for maintenance is not practical. This is true, for example, of I/O devices that are on call for random service. But devices such as typewriters, card readers, punches, and disks can be diagnosed online if there is a backup capability. Online diagnosis is also possible if the application is designed to be carried out under degraded conditions — an important consideration in control programming.

Programming Languages for Process Control

Because they are easier to use than assembly languages, the so-called higher-level languages are used for all types, including process control programming. Assembly languages permit the knowledgeable programmer to take advantage of programming tricks and of the design details of the particular computer to optimize the operating speed of the system while minimizing the need for memory. In other words, assembler languages permit more efficient use of the computer.

However, the ever-increasing speed and capacity and the ever-shrinking cost of computers make such sophisticated programming techniques and the extensive experience required to apply them unnecessary. Thus, the use of the assembler languages is disappearing, except perhaps for operating systems and for a few very large microcomputer applications where any possible savings resulting from optimum use of the computer's capacity are critical. Thus our emphasis in this section will be on higher-level programming languages.

Higher-Level Languages Higher-level languages for process control can be divided into two major branches: general-purpose languages and problem-oriented languages. Figure 2.21j shows these two branches and their further subdivision into several different categories. Higher-level languages have two major advantages over an assembler or any other so-called lower-level language. First, the same program can be run on any type of computer for which the corresponding translator program has been written. That is, the higher-level languages are not usually machine architecture–dependent as are assemblers.

The translators, called compilers or interpreters, are discussed below. The second advantage is that their symbology is usually more readily interpreted by the average user than are the assembler or machine languages they replace because they are more like normal English or mathematical statements.

A compiler takes the code (program) as written by the programmer and converts it completely to the machine language of the particular computer. This translation may include converting the program into an assembler language as an intermediate step.

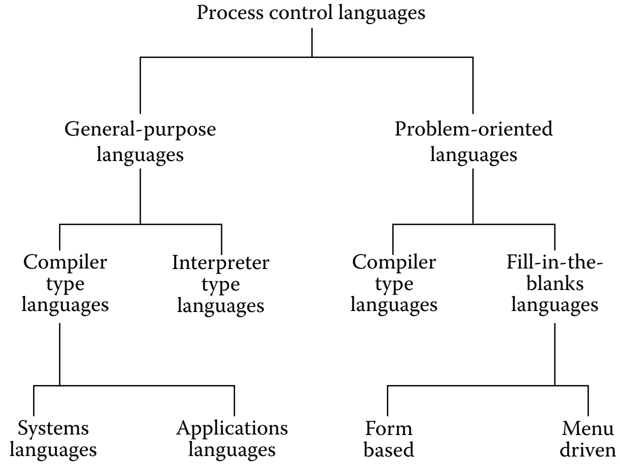


FIG. 2.21j
Organization of higher-level languages for process control.

An interpreter, on the other hand, stores the program in the computer in approximately the form developed by the programmer. The computer then executes the program a line at a time by interpreting, or converting, each line to machine language and carrying out the instructions included therein before going on to the next line.

It can be readily understood that programs to be used with an interpreter must be somewhat simpler than programs to be used with a compiler. When compiling a program, the computer can sense the complexity of the entire program and incorporate some means to handle it. This "look-ahead" capability is, by definition, not available with the interpreter.

Compiler Languages Compiler languages differ in terms of their capabilities. Those that permit complex programs to be readily written in them are called "systems programming languages." Some of these can be used in the preparation of very difficult programs, such as large computer operating systems or executive programs. Otherwise, these programs must be written in the assembler language of the particular computer.

Problem-oriented languages are those that have been especially developed for a particular type of application. Process control is a particularly important application category in this regard. In fact, many problem-oriented languages have been especially developed for relatively narrow types of applications within process control, such as the control of batch reactors.

As indicated in Figure 2.21j, problem-oriented languages can be further divided into compiler-type languages and "fill-in-the-blanks" languages. One of the major types of "fill-in-the-blanks" programs originated with the General Process Control Programming system (GPCP), developed by Humble Oil and Refining Company.[8] The system was later picked up by IBM as PROSPRO[9] and by General Electric as BICEPS.

Compiler-type languages are usually modified or extended versions of the general-purpose compiler languages. FORTRAN and BASIC have been especially popular as the basis for such efforts. A well-known example of this kind of program is AUTRAN (Automatic Utility TRANSlator), which was developed originally by the Merck, Sharp and Dohme Automation Department.

The fill-in-the-blanks system is a largely prewritten, interpreter-type program that offers a large number of choices in its execution in addition to a well-developed skeleton database. By means of responding to a series of questions or a "menu" displayed on a CRT screen, the system developer or programmer is able to make a proper choice of algorithms for each of the control loops from the available options. The programmer can also supply all of the necessary system parameters to the computer's database.

The result is a completely operative control program for the process, developed from the CRT-displayed options. Program development may also be handled by means of a set of preprinted forms, one for each control loop, one for each analog or digital input, one for each process function, etc.

These forms are then converted and are read into the computer's memory.

The prewritten fill-in-the-blanks program must be interpretively executed since it is developed with no knowledge of how the final programmer will organize the execution of the final program or how many and what type of functions will be involved.

Deficiencies of Compiler Programs FORTRAN and BASIC are popular scientific computation languages because programs written in these languages strongly resemble normal mathematical expressions. As a result they are widely used as a basis for university teaching, and a great number of people are familiar with their use.

This same mathematical formula-like representation makes these languages viable candidates for process control programming. However, because they were both originally developed only for batch scientific programs, they were incapable of readily expressing many of the operations vital to process control. These operations are primarily related to the process computer's need to interface with its surroundings and to carry out logic functions. These operations can be listed as follows:

1. Bit string manipulation (i.e., all the logic functions, bit counts, bit testing, and bit movements within a computer word)
2. Input/output data handling (analog and digital input and output)
3. Task scheduling and priority interrupt functions (mainly executive program functions)
4. Some aspects of file handling

Because of the popularity of FORTRAN and its obvious potential as a basis for a process control language, nearly every process control computer vendor provides a FORTRAN compiler as part of its programming systems. The deficiencies of FORTRAN for process control (as listed above) have been circumvented by providing nonstandard extensions to the language that mechanize the missing functions in a FORTRAN-like manner.

Standardization of Process Control Languages Almost since the appearance of the first high-level languages and language extensions for real-time applications, there has been a demand—especially from users—to standardize the languages in the same way as languages for "conventional" applications have been standardized. There are several good reasons for this demand, but the most important one is the need for portability of user programs. Portability of user programs has been made possible in principle by higher-level languages, but it has been thwarted repeatedly by the existence of numerous high-level languages and by the appearance of "language-dialects" of otherwise standard languages.

Standardization should therefore serve two purposes:

1. It should keep the number of different programming languages for one problem area as small as possible.
2. It should discourage the development of dialects.

The benefits of standardization to users are obvious, but vendors, too, will find it profitable because they will have to provide fewer different compilers for any one specific computer.

Much work has been done over the years to develop a standard programming language for process control. Fortunately, this work has now come to fruition and has achieved wide acceptance. The first of these efforts was directed toward the possible standardization of the FORTRAN language for process control applications. This work was carried out by the International Purdue Workshop on Industrial Computer Systems, its predecessor, and associated organizations throughout the world.

In the early days, FORTRAN standardization efforts were carried out by ISA[10–12] and by international groups,[13] while international BASIC standardization efforts were initiated by ECMA[14] (European Computer Manufacturers Association). The major part of future programming by user groups will be carried out using problem-oriented languages (POLs). These can be developed specifically for individual projects. Standardization would be achieved by having their own compilers written in the overall standard language. Thus even though the application of the language is specific, its compilation could be handled on any computer. Thus, transportability would still be achieved.

In the area of process control, separate computer languages been developed for batch type processes,[15] for tuning PID algorithms,[16] for large-scale process control systems,[17,18] and for batch manufacturing,[19] and DCS software was developed for the design of multivariable control systems.[20]

As is discussed in other sections of this chapter, an explosion of new programming is in progress today. These efforts aim at artificial intelligence, neural networks, hiararchical optimization, multivariable control, statistical process control, model-based and model-free optimization, fuzzy logic, envelope control, and many other expert system applications.

References

1. Ruel, M., "How Valve Performance Affects the Control Loop," *Chemical Engineering,* Vol. 107, No. 10, 2000.
2. Ruel, M., "PID Tuning and Process Optimization Increased Performance and Efficiency of a Paper Machine," Paper presented at 87th Annual Meeting of PAPTAC, February 2001.
3. Cox, J. B., et al., "A Practical Spectrum of DDC Chemical Process Control Algorithms," *ISA Journal,* October 1966, pp. 65–71.
4. Lupfer, D. E., and Parsons, J. R., "A Predictive Control System for Distillation Columns," *Chemical Engineering Progress,* September 1962, pp. 37–42.
5. Bollinger, R. E., and Lamb, D. E., "The Design of a Combined Feedforward–Feedback Control System," paper presented at the Joint Automatic Control Conference, University of Minnesota, June 19–21, 1963.

6. Moczek, J. S., Otto, R. E., and Williams, T. J., "Approximation Models for the Dynamic Response of Large Distillation Columns," *Process Control and Applied Mathematics — Chemical Engineering Progress Symposium Series,* Vol. 61, 1965, pp. 136–146.
7. Weaver, S., "Importance of Manufacturer Software," *Control Engineering,* January 1968, pp. 60–64.
8. Ewing, R. W., et al., "Generalized Process Control Programming Systems," *Chemical Engineering Progress,* January 1967, pp. 104–110.
9. Markham, G. W., "Fill-in-the-Form Programming," *Control Engineering,* May 1968, pp. 87–91.
10. ISA S61.1, Standard: "Industrial Computer System FORTRAN Procedures for Executive Functions and Process Input–Output, Bit Manipulation and Data and Time Information," Instrument Society of America, 1972. Revised February 1976.
11. ISA S61.2, Standard: "Industrial Computer System FORTRAN Procedures for Handling Random Unformatted Files," Instrument Society of America, 1978.
12. ISA S61.3, Proposed: "Industrial Computer System FORTRAN Procedures for Task Management," Instrument Society of America and Purdue Laboratory for Applied Industrial Control, Purdue University, 1979.
13. "Industrial Real-Time FORTRAN," European Workshop on Industrial Computer Systems, Submitted to the International Standard Organization (ISO) by the Netherlands Standards Organization (NNI), 1981.
14. "Industrial Real-Time BASIC," European Workshop on Industrial Computer Systems, Budapest, Hungary, Central Research Institute for Physics, September 1981.
15. Bristol, E. H., "A Design Tool for Batch Process Control: Terminology and a Structural Model," *Instrumentation Technology,* Vol. 35, No. 10, October 1985, pp. 154–161 (precursor to ISA SP 88 Standards).
16. Chia, T. L., and Lefkowitz, I., "Robust PID Tuning Using IMC Technology," Part I, *Instrumentation Technology,* Vol. 39, No. 10, October 1992, pp. 36–38; Part II, No. 11, November 1992, pp. 36–39.
17. Elzer, P. F., Ed., *Proceedings of the IFAC Conference on Software for Computer Control,* London: Pergamon Press, 1986; also later meetings of this conference, same publisher.
18. Kompass, E. J., Whitlock, S. K., and Williams, T. J., Eds., *Software Engineering for Large Scale Process Control Systems, Proceedings of the Advanced Control Conference,* West Lafayette, IN: Purdue University, 1988.
19. Kompass, E. J., Whitlock, S. K., and Williams, T. J., *Generic Control for Batch Manufacturing, Proceedings of the Advanced Control Conference,* West Lafayette, IN: Purdue University, 1990.
20. Ochiai, S., "Designing Multivariable Control Systems with DCS Software," *Instrumentation Technology,* Vol. 39, No. 7, July 1992, pp. 35–37.

Bibliography

ARC Report, "Batch Process Automation Strategies," ARC Advisory Group, October 1999.

ARC Report, "The Real Deal Present and Future of Hybrid Manufacturing Integration," ARC Advisory Group, June 2001.

Astrom, K. J., *PID Controllers: Theory, Design, and Tuning,* 2nd ed., Research Triangle Park, NC: ISA, 1995.

Butterworths, W. R., *Advanced Process Control,* New York: McGraw-Hill, 1981.

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: ISA, 1990.

Johnson, C. D., *Process Control Instrumentation Technology,* New York: Wiley & Sons, 1998.

Korn, G., *Interactive System Simulation,* New York: McGraw-Hill, 1989.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed, New York: McGraw-Hill, 1990.

MacFarlane, A. G. J., and Hung, Y. S., "Quasi-Classical Approach to Multivariable Feedback Systems Design," 2nd IFAC Symposium, Purdue University, September 1982.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Nyquist, H., "Regeneration Theory," *Bell System Technical Journal,* Vol. 13, 2001.

Ruel, M., "How Valve Performance Affects the Control Loop," *Chemical Engineering,* Vol. 107, No. 10, 2000.

Ruel, M., "PID Tuning and Process Optimization Increased Performance and Efficiency of a Paper Machine," presented at 87th Annual Meeting, PAPTAC, February 2001.

Seborg, D., et al., *Process Dynamics and Control,* New York: Wiley, 1989.

Shinskey, F. G., *Process Control Systems,* New York: McGraw-Hill, 2000.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice-Hall, 1984.

Svoboda, H. C., "Planning and Coordination of Computer Projects," ISA/94 Technical Conference, Anaheim, 1994.

2.22 Process Gains, Time Lags, Reaction Curves

P. S. SCHERMANN (1995) **W. GARCÍA-GABÍN** (2005)

INTRODUCTION

An understanding of a process can be obtained by developing a theoretical process model using energy balances, mass balances, and chemical and physical laws; however, this complicated and time-consuming effort can be omitted in many cases. A good approximation of process dynamics can be obtained by using simplified calculation methods, thus reducing the work required to design control systems.

The dynamic behavior of typical processes in industrial applications (pipe–vessel combinations, heat exchangers, transport pipelines, furnaces, boilers, pumps, compressors, turbines, and distillation columns) can be described using simplified models composed of process gains, dead times, and process dynamics.

In this section the concepts of process gain and dead time and their roles in control system design are explained. Practical guidelines are given for developing experimental models using reaction curves for various equipment categories that are frequently used in the process industry. The effect of instrumentation on loop dynamics is also examined. Although mathematical derivations are minimized in this section, full descriptions and applications of the mathematical tools can be obtained from the bibliography.

PROCESS GAINS

Process gain is determined by a number of factors. These include the physical properties of the equipment such as the vessel volumes, compressor characteristics, and connecting piping dimensions; process variables such as pressure, temperature, and fluid velocity; and various laws of physics and chemistry. While some of the associated gains are a function of the process, considerable improvement can be made by revising the control systems, because the instrumentation has a strong effect on the loop dynamics. A transmitter gain, for example, can be modified by adjusting the measurement range.

The process gain indicates how much a process property (output) changes per unit of input change. The input can be a flow that is set by a valve opening; it affects the output, which can be a process property, such as temperature. The gain defines the sensitivity relating the output and input variables. It can be calculated as follows:

$$K = \frac{\Delta \text{ Output [Transmitter units]}}{\Delta \text{ Input [Controller units]}} \qquad \textbf{2.22(1)}$$

The transmitter and controller signals can be expressed in milliamperes [mA], pounds per square inch [psig] or percentage [%] units, if pneumatic or electronic analog instrumentation or DCS systems are used, respectively.

The process gain frequently depends on the load or operating point. For example, the gain changes when the cross-section of a vessel varies (this is the case with horizontal cylindrical tanks), or if the load is modified (for example, the gain of heat transfer processes will drop with rising load).

These variations introduce nonlinearities, making the process gain highly dependent on the operating point. This means that the process gain is a function of the operating point, which must be taken into account in designing the control system. The control system engineer should be aware of the phenomenon of process gain variation, and the variations in loop gains must be compensated for. The use of adaptive (Section 2.19) or robust (Section 2.26) control can sometimes compensate for gain variations.

DEAD TIMES AND TIME CONSTANTS

Time lag, transportation lag, time delay, or dead time (t_d) is commonly present in industrial processes. Dead time can result from measurement lag, analysis and computation time, communication lag, or the transport time required for a fluid to flow through a pipe.

For example, if a concentration analyzer is located downstream of a reactor, it takes a certain dead time before the fluid leaving the reactor arrives at the concentration measurement point. Dead time is also a result of analysis time and communication time between field (instrument and final control element: valve) and the control room.

The effect of these lags in the process response is shown in Figure 2.22a. This figure illustrates the performance of a loop, when at time = 1, a step change was applied to the set point of a controller (an input step change) in the control

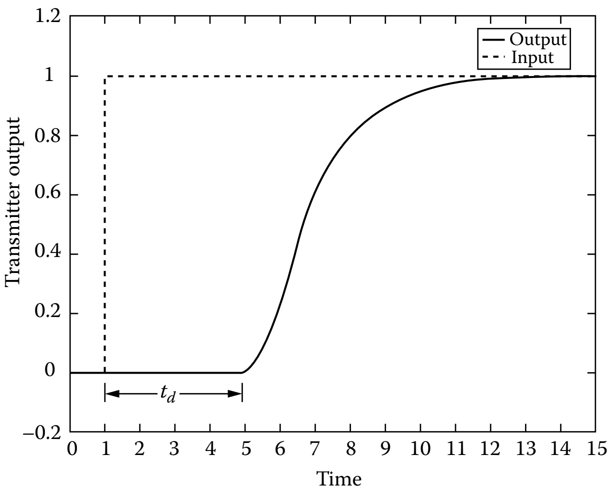


FIG. 2.22a
The response of a process having both dead time and a time constant.

room. This input step change resulted in the noted response of the controlled variable (the output response), which started after a dead time (t_d) had passed and reached a new steady state after the passing of the time constant (τ).

The presence of dead time significantly complicates the analysis and design of feedback control loops. It degrades the control loop's performance because it introduces an unstable behavior, which makes it difficult to achieve a satisfactory control. Special care should be taken when the dead time–to–time constant ratio (t_d/τ) exceeds 1, because the PID controller cannot handle such processes and such control structures as the Smith predictor could be required.

Instrumentation Effect on Process Dynamics

The instrumentation and final control element are not part of the process but are part of the control loop. Yet they affect the dynamic behavior of the process, and the location of the sensor and/or final control element also changes the total dead time.

The process control engineer must know both the speed of detection and the time it takes to transport that information. These times can change considerably as a function of measurement variable and type of instrumentation used.

The time constants of pressure and differential pressure measurements are on the order of 0.1 s. Temperature measurements are slower. Their time constants are usually between 1 and 10 s. Composition measurements (analyzers) are even slower, varying from 5 s to 10 min.

Measurement signal propagation delays are negligible with electronic instrumentation, but signal processing delays can occur in digital systems because the loops are processed sequentially, one by one. The scan period of inputs and outputs normally varies between 0.1 and 1 s, and processing delays inside DCS systems are of the same order of magnitude.

Depending on the length and diameter of tubing between I/P converter and control valves, pneumatic control valves can introduce considerable delays. The use of valve positioners can considerably reduce this delay. The stroking times of conventional (pneumatic) valves vary from 2 s to 1 min, while high-speed electric and hydraulic actuators can reduce them to the millisecond range. As an example, a 10 in. (0.254 m) valve without positioner and with 1/4 in. (6.35 mm) pneumatic tubing of 300 ft (91 m) length between controller and valve has a time constant of about 30 s.

Transportation Lag

When material or energy is physically moved in a process plant, there is a dead time associated with that movement. For instance, if a temperature change travels through a pipe with the fluid, without mixing, the resulting dead time is a function of the length of the pipe. If the pipe is considered to be long (its length is much greater than its diameter), "plug flow" occurs. The resulting dead time between two points on the pipeline can be calculated by

$$t_d = \frac{V}{F_v} = \frac{L}{v_v} \qquad \textbf{2.22(2)}$$

where
t_d is the deadtime
V is the volume of the pipe
F_v is the volume flow rate
L is the length of pipe
v_v is the fluid velocity

This dead time equals the residence time of the fluid in the pipe. Note that the dead time is inversely proportional to the flow rate. The fact that the dead time drops as the production rate of the plant rises must be taken into account when designing the control system.

Even when the plug flow assumption is not valid, transportation lag usually can approximately be modeled by a pure dead time. For liquid flow in a pipe, the plug flow assumption is most accurate when the axial velocity profile is flat, a condition that occurs when Newtonian fluids are transported in turbulent flow. For non-Newtonian fluids and for laminar flow, the fluid transport lag can still be modeled as a pure dead time, which is calculated on the basis of the average fluid velocity.[1]

EXAMPLE

Suppose that a process stream is flowing through a pipe in plug flow. The stream is an incompressible fluid. The pipe is perfectly insulated and has a length of $L = 200$ ft (60.96 m). The fluid flows at a constant velocity of $v = 20$ ft/s (6.096 m/s) and is initially at a constant temperature of $T_0 = 100$°F (37.778°C) throughout the length of the pipe. In this case, the transportation lag of a

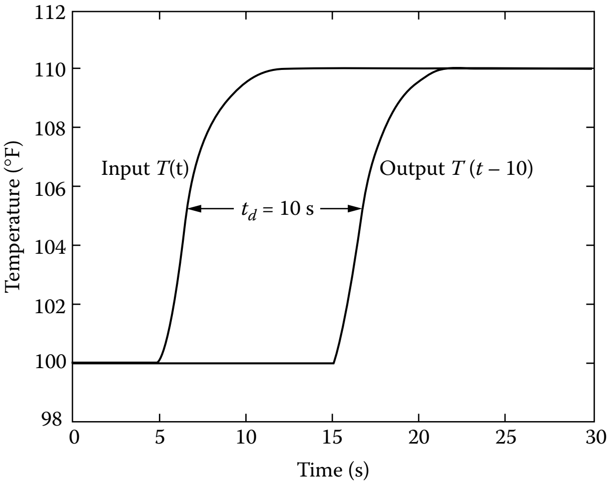


FIG. 2.22b
Transportation lag resulting from plug flow in a pipeline.

temperature change between some points of input and output is given by

$$t_d = \frac{L}{v_v} = \frac{200 \text{ ft}}{20 \text{ ft/s}} = 10 \text{ s} \qquad \textbf{2.22(3)}$$

Figure 2.22b illustrates this input/output transportation lag.

Dead Time Representation

Dead time representation varies according to the mathematical model. Table 2.22c shows different ways to represent the dead time.

Dead Time Approximations

The frequency (Laplace) domain is frequently used to study control systems. In this case the dead time is represented by $e^{-t_d s}$. Because the methods used to analyze and design control systems cannot incorporate exponential functions, polynomial functions are used to represent such exponential terms. The most common representations are Padé approximation, Taylor series expansion, and first-order systems series approximations.[2]

TABLE 2.22c
Dead Time Representation

Domain	*Input*	*Output*
Time	$T(t)$	$T(t - t_d)$
Laplace	$T(s)$	$T(s)e^{-t_d s}$
Discrete t_m : sample time	$T(z)$	$T(z)z^{-\frac{t_d}{t_m}}$

Padé Approximation Padé approximation represents the $e^{-t_d s}$ term by the ratio of low-order polynomials in the Laplace domain,

$$e^{-t_d s} = \frac{M_k(s)}{N_k(s)} \qquad \textbf{2.22(4)}$$

where k is the order approximation and $M_k(s)$ and $N_k(s)$ are polynomials:

$$M_k(s) = m_0 + m_1 t_d s + m_2 (t_d s)^2 + \cdots + m_k (t_d s)^k \qquad \textbf{2.22(5)}$$

$$N_k(s) = n_0 + n_1 t_d s + n_2 (t_d s)^2 + \cdots + n_k (t_d s)^k \qquad \textbf{2.22(6)}$$

They are calculated by

$$M_k(s) = \sum_{i=0}^{k} \frac{(2k-i)!k!}{2k!i!(k-i)!}(-1)^i (t_d s)^i \qquad \textbf{2.22(7)}$$

$$N_k(s) = \sum_{i=0}^{k} \frac{(2k-i)!k!}{2k!i!(k-i)!}(t_d s)^i \qquad \textbf{2.22(8)}$$

Taylor Series Expansion The Taylor series expansion of the exponential term $e^{-t_d s}$ is defined as

$$e^{-t_d s} = 1 - t_d s + \frac{(-t_d s)^2}{2!} + \frac{(-t_d s)^3}{3!} + \cdots \frac{(-t_d s)^k}{k!} + \cdots \qquad \textbf{2.22(9)}$$

However, simulations require the polynomial numerator degree to be equal or smaller than that of the denominator; for this reason, instead of Equation 2.22(9), the following approximation is used in practical applications:

$$e^{-t_d s} = \frac{1}{e^{t_d s}} = \frac{1}{1 + t_d s + \frac{(t_d s)^2}{2!} + \frac{(t_d s)^3}{3!} + \cdots \frac{(t_d s)^k}{k!} + \cdots} \qquad \textbf{2.22(10)}$$

First-Order Systems Series (FOSS) An exponential term is $e^{-t_d s}$ equivalent to a system composed of an infinite sequence of first-order systems series (FOSS):

$$e^{-t_d s} = \text{Lim}_{k \to \infty} \frac{1}{\left(\frac{t_d}{k} s + 1\right)^k} \qquad \textbf{2.22(11)}$$

Table 2.22d summarizes the transfer functions of the Padé, Taylor, and FOSS approximations. It presents only first-, second-, and third-order approximations because they are the most common in practical control applications.

EXAMPLE

Using the following dead time system:

$$G_1(s) = e^{-s} \qquad \textbf{2.22(12)}$$

TABLE 2.22d
Transfer Function Approximations of the Exponential Term ($e^{-t_d s}$) of Dead Time

Order	*Padé*	*Taylor*	*FOSS*
1	$\dfrac{1-\frac{t_d s}{2}}{1+\frac{t_d s}{2}}$	$\dfrac{1}{1+t_d s}$	$\dfrac{1}{1+t_d s}$
2	$\dfrac{1-\frac{t_d}{2}s+\frac{t_d^2}{12}s^2}{1+\frac{t_d s}{2}+\frac{t_d^2}{12}s^2}$	$\dfrac{1}{1+t_d s+\frac{t_d^2}{2}s^2}$	$\dfrac{1}{\left(1+\frac{t_d}{2}s\right)^2}$
3	$\dfrac{1-\frac{t_d}{2}s+\frac{t_d^2}{10}s^2-\frac{t_d^3}{120}s^3}{1+\frac{t_d s}{2}+\frac{t_d^2}{10}s^2+\frac{t_d^3}{120}s^3}$	$\dfrac{1}{1+t_d s+\frac{t_d^2}{2}s^2+\frac{t_d^3}{6}s^3}$	$\dfrac{1}{\left(1+\frac{t_d}{3}s\right)^3}$

Different approximations have been compared to show how close the various approaches are to the exponential term. Figures 2.22e and 2.22f show Bode diagrams for first- and second-order approximations. Padé approximation is too close to the exponential term in the magnitude diagram, but the phase lag is inadequately approximated at higher frequencies. Taylor and FOSS produce a similar approximation; they match reasonably the exponential term in magnitude and phase at low frequencies, but they are unsuccessful at high frequencies. The Padé approach provides better results for the same approximation order than Taylor and FOSS.

In order to illustrate how the Padé, Taylor, and FOSS approximations modify the time response, a step is applied to a pure dead time system (Equation 2.22[12]). In this case, the ideal output would be the same input with a dead time (t_d). This ideal output has been compared with first- and second-order dead time approximations, respectively, in Figures 2.22g and 2.22h.

Taylor and FOSS approaches produce better approximations in the time domain for a unit step response than does the Padé approximation. The Padé approximation has right half plane zeros (RHPZ); also, the number of RHPZ increases with the order of approximation. The presence of RHPZ in the transfer function is responsible for the inverse response. It is also the source of a considerable amount of difficulty in controller design. Some control algorithms have stability problems when controlling a process with RHPZ. This occurs

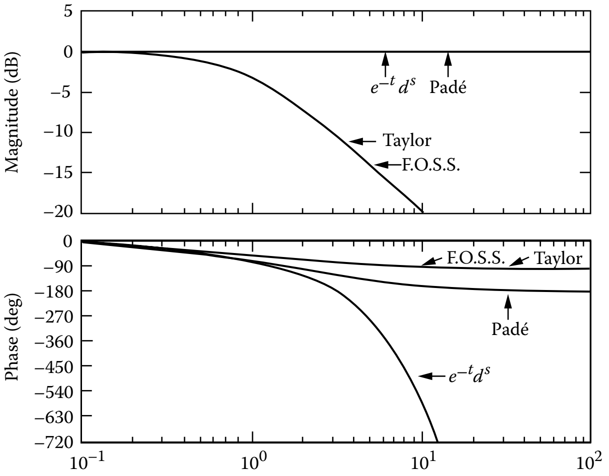


FIG. 2.22e
Bode diagram of a dead time system (Equation 2.22[12]) using first-order approximations.

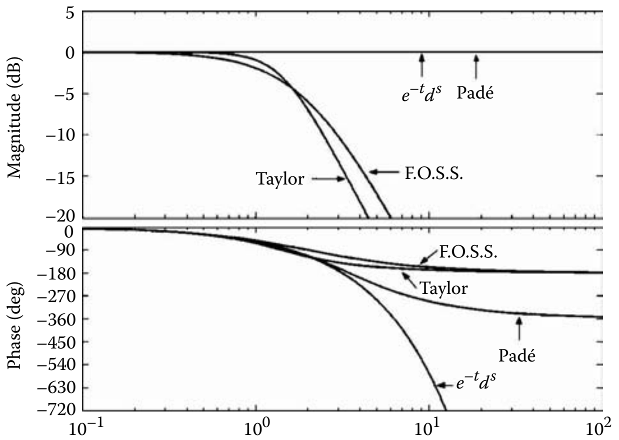


FIG. 2.22f
Bode diagram of a dead time system (Equation 2.22[12]) using second-order approximations.

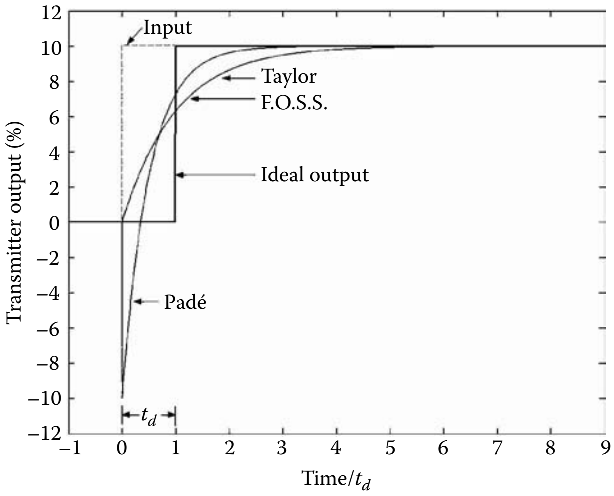


FIG. 2.22g
Step response of a dead time system (Equation 2.22[12]) using first-order approximations.

when the controller contains an inverse model of the process and high performance is desired.

In summary, Padé is a better approximation than Taylor and FOSS, but Padé approximation presents RHPZ, which can cause the controller to become unstable. In such cases the Taylor or FOSS approximations can result in a better choice to approximate the exponential term.

REACTION CURVES

A complete mathematical description of the process can be constructed using mass and energy balances, engineering relations, valve equations, etc. It is a difficult and time-consuming job to develop a dynamic process model. An approximation of the dynamic process model can be obtained by using the reaction curve method. This provides a simple and fast procedure to determine an approximated linear process model.

This experimental technique is based on applying a change in the manipulated variable (input of the process) with the loop open (controller in manual) and recording the response of the controlled variable (output of the process). The parameters of the linear model are calculated on the basis of the locations of some specific points in the output response.

The linear model represents an approximate model that is adequate for many engineering purposes; however, it is only valid in an operating point and it does not take into account the high-order behavior and nonlinearity of the process. Most of the industrial processes can be represented by a first-order plus dead time model. Other processes can be approximated by a second-order underdamped plus dead time system or by an integrating plus dead time system.

The following procedures can be used to approximate linear models for these kind of processes.

First-Order Plus Dead Time Processes

A high percentage of all chemical process can be modeled by using first-order plus dead time systems.

$$G(s) = \frac{K}{\tau s + 1} e^{-t_d s} \qquad \textbf{2.22(13)}$$

where
K is the steady-state gain
τ is the time constant
t_d is the dead time

The parameters of the process model, K, t_d, τ are obtained by using the following procedure:

1. The control loop is opened by switching it to manual.
2. This is done when the controlled variable (system output) is at a constant value and no disturbances or other upsets are allowed to occur, while the reaction curve is developed.
3. A step change is applied to the manipulated variable (controller output, which is an input to the process). The step changes are usually 5 to 10%. The step time should be long enough for the manipulated variable (system input) to reach a new steady state.
4. The manipulated variable (system input) response is recorded to provide good visibility on both the amplitude and time scales.

Two slightly different graphical techniques are utilized in these procedures. These methods are explained in the next paragraphs.

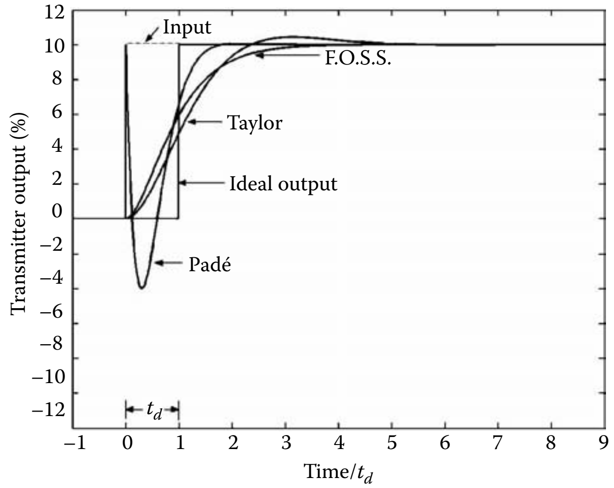


FIG. 2.22h
Step response of a dead time system (Equation 2.22[12]) using second-order approximations.

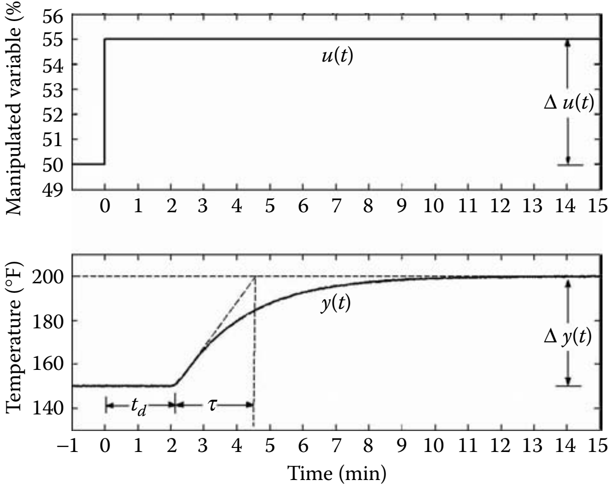


FIG. 2.22i
Reaction curve evaluation using the "first method" to determine dead time and time constant of a first-order plus dead time process.

First Method The first method uses a tangent to the initial process response curve, drawn at the point of maximum slope; the point where it intercepts the final steady-state value is noted. In this case, the process time constant is the time between the tangent intercepting the original and the new steady-state lines. The dead time is measured as the time it takes between applying the step change and the beginning of the time constant.

EXAMPLE

In Figure 2.22i, a 5% step change to the manipulated variable (controller output) has been applied, and process response in terms of the controlled temperature is recorded. If the first method is utilized, the model parameters are

$$K = \frac{\Delta y(t)}{\Delta u(t)} = \frac{50°\text{F}}{5\%} = 10\frac{°\text{F}}{\%} \qquad \textbf{2.22(14)}$$

$$\tau = 2.4\ \text{s}, \qquad \textbf{2.22(15)}$$

$$t_d = 2.1\ \text{s} \qquad \textbf{2.22(16)}$$

Second Method The second method eliminates the need to draw a tangent. This approach proposes that the values of τ and t_d are so selected that the model will coincide with the process response at two points.[3] On the time scale, these two points are at times when the controlled variable (system output) reaches 28.3 and 63.2% of its final steady-state value. Once these points match, the parameters of the process model are calculated using Equations 2.22(17) to 2.22(19).

EXAMPLE

As shown in Figure 2.22j, a step change in the manipulated variable is applied to the process, and the response of the controlled variable is evaluated. The model parameters can be calculated using the following equations:

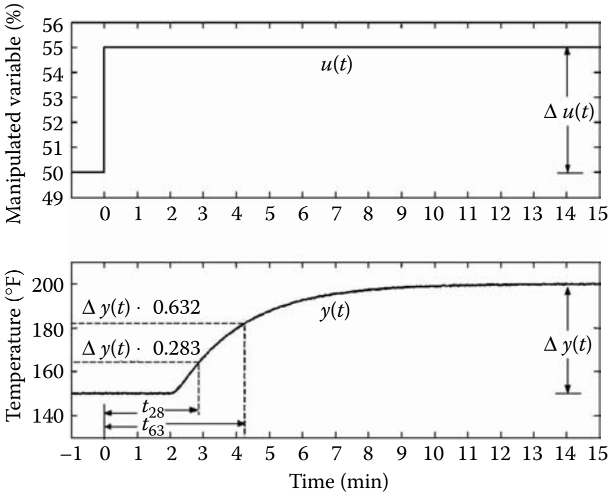


FIG. 2.22j
Reaction curve evaluation using the "second method" to determine dead time and time constant of a first-order plus dead time process.

$$K = \frac{\Delta y(t)}{\Delta u(t)} \qquad \textbf{2.22(17)}$$

$$\tau = \frac{3}{2}(t_{63} - t_{28}) \qquad \textbf{2.22(18)}$$

$$t_d = t_{63} - \tau \qquad \textbf{2.22(19)}$$

thus,

$$K = \frac{50°\text{F}}{5\%} = 10\frac{°\text{F}}{\%} \qquad \textbf{2.22(20)}$$

$$\tau = \frac{3}{2}(4.2 - 2.9) \approx 2.0\ \text{s} \qquad \textbf{2.22(21)}$$

$$t_d = t_{63} - \tau = (4.2 - 2.0) = 2.2\ \text{s} \qquad \textbf{2.22(22)}$$

The first and second methods are compared with the process response in Figure 2.22k. The second method results in a better approximation than the first one. One limitation of the first method is that drawing the tangent line at the point of maximum slope is not an easy task. The other reason why the second method is superior is because it matches the process response curve at two points in the region of maximum slope, instead of one.

Figure 2.22l shows how the second method can be used to obtain a rude approximation of a nonminimum phase system by a first-order plus dead time approximation. The principal idea is to approximate the inverse response by a dead time.[3]

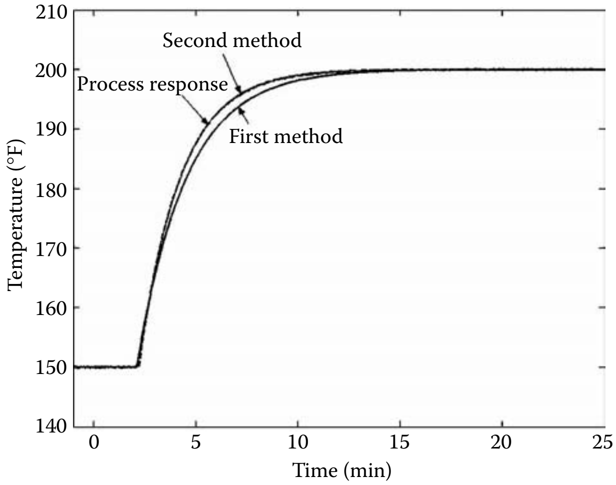


FIG. 2.22k
Comparison between the first and second methods of approximating the process responses.

Underdamped Processes

Chemical processes rarely are underdamped; however, electrical–mechanical systems with time lag (due to communication lags or signal analysis lags) typically are, when an input step change is applied. A reaction curve can be used to obtain the parameters for a second-order underdamped with dead time process in the same way as was done with a first-order plus dead time system.

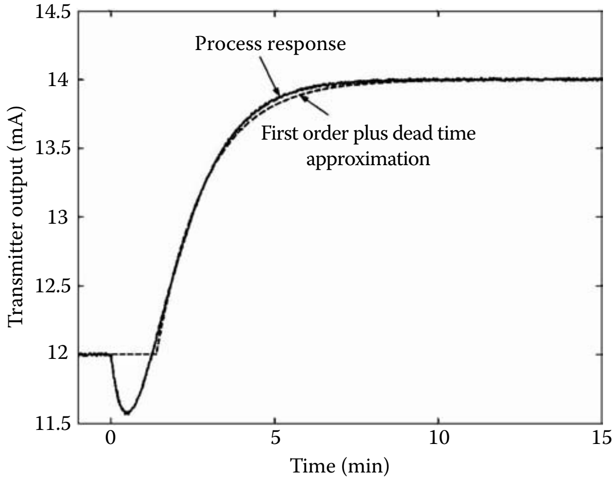


FIG. 2.22l
Approximation of a nonminimum phase process by a first-order plus dead time model using the "second method."

Closed-loop chemical processes are often tuned to have an underdamped response. In that case, a process model is calculated using the online step response.[4] A process that is underdamped with dead time can be approximated by the equation below:[5]

$$G(s) = \frac{K}{\tau^2 s^2 + 2\tau\xi s + 1} e^{-t_d s} \qquad \textbf{2.22(23)}$$

and the parameters of this model can be estimated from the following expressions:

$$\mathrm{OR} = \frac{\Delta y_{\mathrm{OR}}}{\Delta y} \qquad \textbf{2.22(24)}$$

$$\theta = \arctan\left(\frac{-\pi}{\ln \mathrm{OR}}\right) \qquad \textbf{2.22(25)}$$

$$\tau = t_{\mathrm{OR}} \frac{\sin(\theta)}{\pi - \theta} \qquad \textbf{2.22(26)}$$

$$\xi = \cos\theta \qquad \textbf{2.22(27)}$$

where

K is the steady-state gain
t_d is the dead time
OR is the overshoot ratio
t_{OR} is the time to reach the maximum overshoot
Δy_{OR} is the maximum peak overshoot
τ is the time constant
ξ is the damping factor

EXAMPLE

A step test signal, $u = 10\%$, is applied at $t = 0$ min to the manipulated variable (process input). The output response (velocity) is shown in Figure 2.22m. The figure also identifies the readings that have to be made to calculate the second-order underdamped model parameters.

$$\mathrm{OR} = \frac{\Delta y_O}{\Delta y} = \frac{125\ \mathrm{rpm}}{100\ \mathrm{rpm}} = 0.25 \qquad \textbf{2.22(28)}$$

$$\theta = \arctan\left(\frac{-\pi}{\ln 0.25}\right) = 1.16\ \mathrm{rad} \qquad \textbf{2.22(29)}$$

$$\tau = 4.3\frac{\sin(1.16)}{\pi - 1.16} = 2.0\ \mathrm{s} \qquad \textbf{2.22(30)}$$

$$\xi = \cos 1.16 = 0.4 \qquad \textbf{2.22(31)}$$

$$K = \frac{\Delta y}{\Delta u} = \frac{100\ \mathrm{rpm}}{10\%} = 10\frac{\mathrm{rpm}}{\%} \qquad \textbf{2.22(32)}$$

$$t_d = 1\ \mathrm{s} \qquad \textbf{2.22(33)}$$

Finally, the approximated model is as follows:

$$G(s) = \frac{10}{s^2 + 0.8\ s + 1} e^{-s} \qquad \textbf{2.22(34)}$$

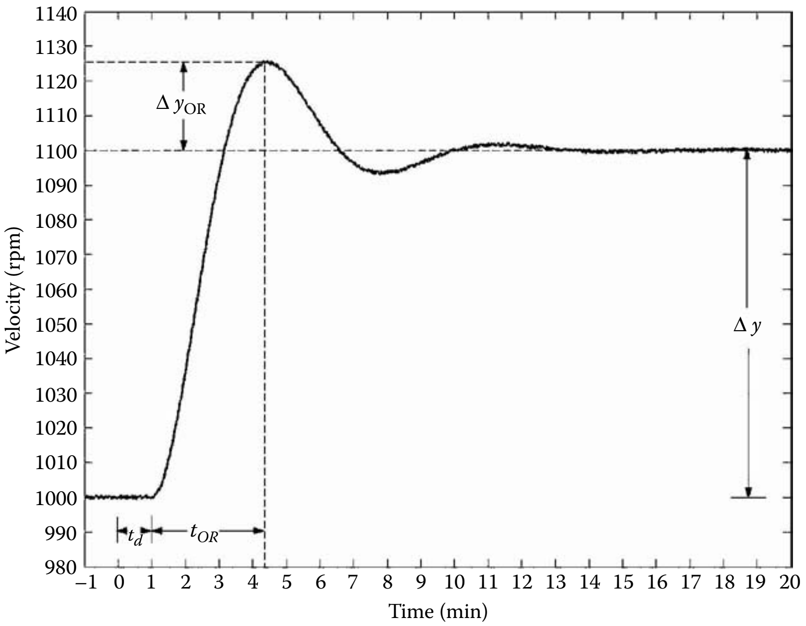


FIG. 2.22m

The approximation of an underdamped, second-order plus dead time process requires the measurement of the noted four values from the process response curve.

Integrating Plus Dead Time Processes

Sometimes industrial processes contain pure integration elements. They are not self-regulating processes and do not have a steady state. Therefore, any step change will cause the controlled variable (process output) to increase or decrease linearly with time.

This is the case with controlling level in a tank, when the manipulated variable is feed flow rate into the tank. The following model in Laplace domain can approximate an integrating system with dead time:

$$G(s) = \frac{K}{s} e^{-t_d s} \qquad \textbf{2.22(35)}$$

where

K is the process gain
t_d is the dead time

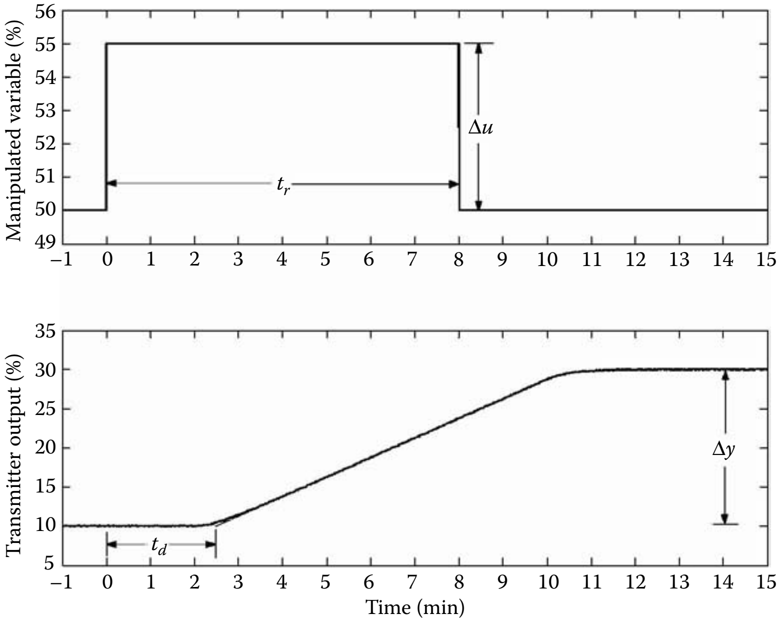


FIG. 2.22n

A pulse test is used to establish the dynamic characteristics of an integrating plus dead time process.

The process model of an integrating process is obtained by applying a pulse test and not a step test because the latter would produce an output curve that will change linearly, without reaching a steady-state value. The gain of the integration element can be calculated by[5]

$$K = \frac{\Delta y}{\int_0^{t_r} u(t)dt} \qquad \textbf{2.22(36)}$$

When a pulse step is applied, Equation 2.22(36) can be rewritten as

$$K = \frac{\Delta y}{\Delta u t_r} \qquad \textbf{2.22(37)}$$

The readings that need to be obtained are shown in Figure 2.22n, consequently

$$K = \frac{\Delta y}{\Delta u t_r} = \frac{20\%\ \text{MV}}{5\%\ \text{TO}\ 8\ \text{s}} = 0.5\frac{\%\ \text{MV}}{\text{TO s}} \qquad \textbf{2.22(38)}$$

$$t_d = 2.5\ \text{s} \qquad \textbf{2.22(39)}$$

CONCLUSION

The dynamic behavior of most chemical processes can be approximated by the use of first-order plus dead time models. Other methods, also described in this section, included the methods of modeling second-order underdamped and integrating plus dead time processes.

The reaction curve methods were used to model the dynamic characteristic of the processes. These are input/output (black-box) linear models that do not take into account nonlinearities or high-order dynamics. These models are usually valid only in a region near the operation point of the process, but they are powerful tools in the design and tuning of control loops.

References

1. Seborg, D., Thomas, E., and Mellichamp, D., *Process Dynamics and Control,* New York: John Wiley & Sons, 1989.
2. Ogunnaike, B. A., and Ray, W. H., *Process Dynamics, Modeling, and Control,* New York: Oxford University Press, 1994.
3. Smith, C., and Corripio, A., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1997.
4. Yuwana, M., and Seborg, D., A new method for on-line controller, *AIChE Journal,* May 1982.
5. Luyben, W., *Process Modeling, Simulation, and Control for Chemical Engineers,* New York: McGraw-Hill, 1990.

Bibliography

Aris, R., and Amundson, N. R., *Mathematical Models in Chemical Engineering,* Englewood Cliffs, NJ: Prentice Hall, 1973.

Bischoff, K. B., and Himmelblau, D. M., *Process Analysis and Simulation of Deterministic Systems,* New York: John Wiley & Sons, 1968.

Friedly, J. C., *Dynamic Behavior of Processes,* Englewood Cliffs, NJ: Prentice Hall, 1972.

Gould, L. A., *Chemical Process Control: Theory and Applications,* Reading, MA: Addison-Wesley, 1969.

Kecman, V., *State-Space Models of Lumped and Distributed Systems,* Heidelberg: Springer-Verlag, 1988.

Koppel, L. B., *Introduction to Control Theory with Applications to Process Control,* Englewood Cliffs, NJ: Prentice Hall, 1968.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* New York: McGraw-Hill, 1973.

Marlin, T., *Process Control: Designing Processes and Control Systems for Dynamic Performance,* New York. McGraw-Hill, 2000.

Peczkowski, J. L., and Sain, M. K., "Nonlinear Control by Coordinated Feedback Synthesis with Gas Turbine Applications," Proceedings of the American Control Conference, June 1985.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Rademaker, O., Rijnsdorp, J. E., and Maarleveld, A., *Dynamics and Control of Continuous Distillation Units,* Amsterdam: Elsevier, 1975.

Rijnsdorp, J. E., MacGregor, J. F., Tyreus, B. D., and Takamatsu, T., *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes,* Oxford: Pergamon Press, 1990.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

2.23 Ratio Control

F. G. SHINSKEY (1970, 1985) **B. G. LIPTÁK** (1995) **J. E. JAMISON, A. ROHR** (2005)

INTRODUCTION

Ratio control systems maintain a relationship between two variables to provide regulation of a third variable. Ratio systems are used primarily for blending ingredients into a product or as feed controls to chemical reactors. An example would be the addition of a gasoline additive under ratio control in order to maintain the required octane number of the product, which number may or may not be measured.

Ratio systems portray an elementary form of feedforward control (Section 2.9). In the above example, if the load input to the system (gasoline flow) changed, it would cause a variation in the controlled variable (octane number). This variation can be eliminated by an appropriate adjustment of the manipulated variable (additive flow rate).

The load, or wild flow, as it is called, may be uncontrolled, controlled independently, or manipulated by other controllers that respond to the variables, such as pressure or level.

FLOW RATIO CONTROL

Ratio control is applied almost exclusively to flows. Consider maintaining a certain ratio R between ingredients A and B:

$$R = B/A \qquad \mathbf{2.23(1)}$$

There are two ways to accomplish this. The more common method is to calculate and manipulate the set point of a flow loop as follows:

$$\text{Set Point} = B = RA \qquad \mathbf{2.23(2)}$$

This system is shown in Figure 2.23a. The set point for the flow ratio controller (FFC-2) is calculated by an adjustable-gain device, which is known as the ratio station. If the ratio station (FY in Figure 2.23b) is outside the control loop, it does not interfere with the secondary loop's response.

The second method is to calculate the ratio R from the individual measurements of flows A and B (Equation 2.23[1]) and use this calculated ratio as the measurement input (controlled variable) into a manually set ratio controller (RIC). Such a scheme is shown in Figure 2.23c. The principal disadvantage of this configuration is that it places a divider inside a closed loop. If flow B responds linearly to the opening of valve B, the gain of the loop will vary because of this divider. The differentiation of Equation 2.23(1) explains why this is true:

$$\frac{dR}{dB} = \frac{1}{A} = \frac{R}{B} \qquad \mathbf{2.23(3)}$$

Equation 2.23(3) shows that the loop gain varies both with the ratio R and the flow B. In most applications, the ratio R would not often be subject to change, but the flow B would. Because the loop gain varies inversely with flow B, this can cause instability at low rates. Therefore the use of an equal-percentage valve characteristics is essential to overcome this danger. If the ratio were inverted,

$$R = A/B \qquad \mathbf{2.23(4)}$$

then

$$\frac{dR}{dB} = -\frac{A}{B^2} = -\frac{R}{B} \qquad \mathbf{2.23(5)}$$

and the result is essentially the same.

The square-root extractors in Figures 2.23c are shown in broken lines to indicate that the control system will also function without them, as it can use the flow-squared signals. If that is the case, the controlled variable is

$$R^2 = B^2/A^2 \qquad \mathbf{2.23(6)}$$

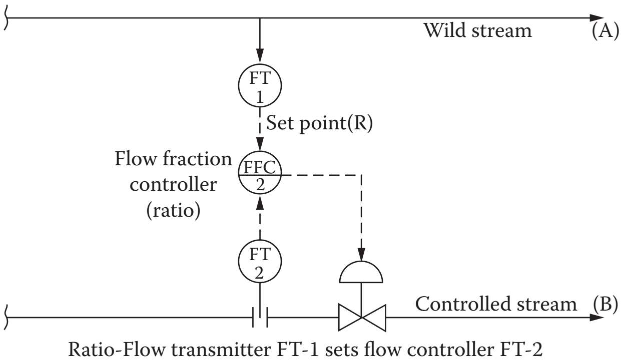
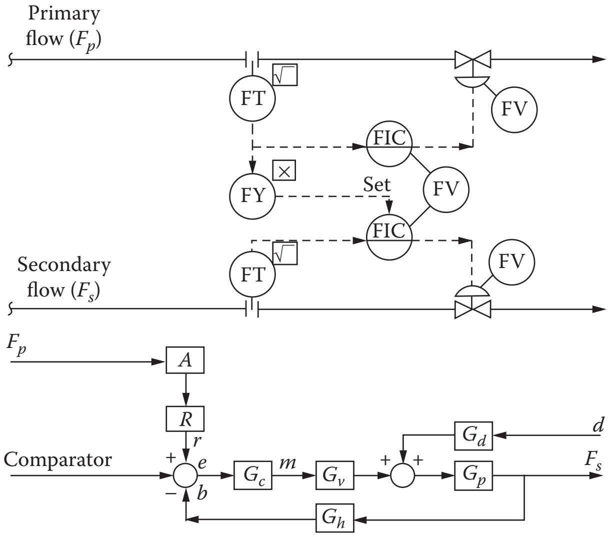


FIG. 2.23a
Controlled stream is manipulated to maintain a constant ratio (R) of controlled flow (B) to wild flow (A).



A = Input element.
b = Feedback variable.
d = Disturbance or load Variable.
e = Error (deviation) signal.
F_p = Primary flow.
F_s = Secondary flow.
G_c = Controller transfer function.
G_d = Disturbance or load transfer function.
G_h = Feedback sensor transfer function.
G_p = Process transfer function.
G_v = Control valve transfer function.
m = Manipulated variable.
r = Reference (set-point) Input.
R = Desired ratio.

FIG. 2.23b
If the ratio calculation is made outside the secondary flow control loop, its setting does not change the dynamics or response of the loop.

As a consequence, the scale of the ratio controller or ratio station has to be nonlinear. Differentiating Equation 2.23(6):

$$\frac{d(R^2)}{dB} = \frac{2B}{A^2} = \frac{2R}{A} = \frac{2R^2}{B} \qquad \textbf{2.23(7)}$$

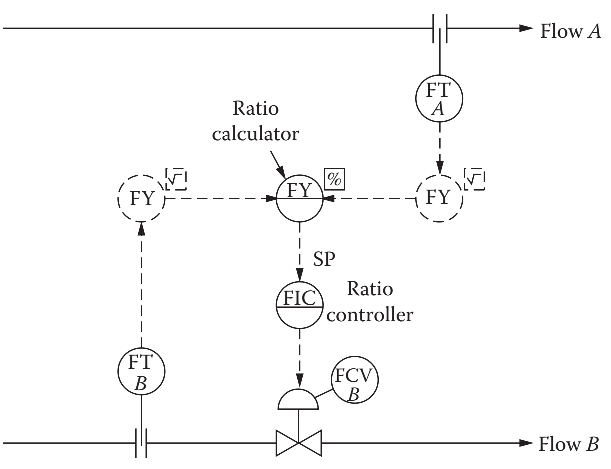


FIG. 2.23c
Here the ratio of flows A and B is calculated by a divider within the closed ratio control loop (FIC). The disadvantage of this approach is that the loop gain changes every time the ratio setting of flow B changes.

In this case too, for the configuration shown in Figure 2.23c, the loop gain varies inversely with flow.

Advantages and Disadvantages

The principal advantage of using the ratio control system is that the controlled variable — flow ratio — can be kept constant and can be directly recorded to verify control performance (Figure 2.23c). In contrast, using the loop configurations in Figures 2.23a or 2.23b would require two records and their evaluation for ratio verification.

The disadvantage of using analog transmitters in ratio control is that it has no memory. For example, in a blending system, if the controlled flow for any reason cannot match the required value, even temporarily, the resulting incorrect composition cannot be automatically corrected because there is no memory of the amount of past error.

One way to try to correct this situation is to add a totalizer per stream to verify the overall correctness of total blended product. If such information is available, subsequent operator intervention can bring the average composition of the blend to the desired value. Naturally, this assumes that a correct mean ratio is acceptable for the purposes of the process, rather than continuously maintaining a correct ratio.

While head-type flow meters were used in the earlier examples, the general comments that were made are equally valid when positive displacement (PD) or turbine-type flow detectors are used. When the plant is controlled by microprocessor-based digital controllers or DCS systems, the reconfiguration of such ratio control loops requires much less effort. If the flow measurement signals are transmitted utilizing bus-based technologies, the associated dead time should be considered in the tuning of the flow ratio control loops.

RATIO STATIONS

No matter what portion of the ratio control loop is implemented in hardware or in software, a computing element must be used whose scaling requires some consideration. The ratio station (FY in Figure 2.23b) normally has a gain range of about 0.3 to 3.0. The primary flow signal *Fp*, in percent of scale, is multiplied by the gain setting to produce a set point for the secondary flow controller (*Fs*), in percent of scale. The true flow ratio must take into account the scales of the two flowmeters. The setting of the ratio station is related to the true flow ratio by

$$R = (\text{true flow ratio})(\text{Scale of } Fp)/(\text{Scale of } Fs) \qquad \textbf{2.23(8)}$$

As an example, assume that the true flow ratio of the additive to gasoline is to be 2.0 cc/gal (0.53 cc/ℓ). If the additive flow scale is 0 to 1200 cc/min, and the gasoline flow scale is 0 to 500 gal/min (0 to 1890 ℓ/m), then:

$$R = 2.0 \text{ cc/gal} \frac{500 \text{ gal/min}}{1200 \text{ cc/min}} = 0.833 \qquad \textbf{2.23(9)}$$

TABLE 2.23d
Comparing the Actual Ratio Settings Required for Linear and Head-Type Flowmeters to Obtain the Same Flow Ratio

	Actual Ratio (Gain) Setting Required to Achieve the Desired Ratio	
Flow Ratio Desired	If Signals Are Linear	If Flow-Squared Signals Are Used and the Scale Is Linear
0.6	0.6	0.36
0.8	0.8	0.64
1.0	1.0	1.0
1.2	1.2	1.44
1.4	1.4	1.96
1.6	1.6	2.56

When head-type flowmeter signals are used and the scale is linear, such scale should show the square root of R in order to be meaningful. Table 2.23d compares the gain of a ratio station (corresponding to linear flow ratio) with the ratio setting for head-type flowmeters. As shown by Table 2.23d, the available range of ratio settings is seriously limited when using squared flow signals.

A ratio controller (FFC-2 in Figure 2.23a) combines the ratio function and the controller in one unit. If implemented in hardware, it saves cost and panel space. Such algorithms are also available in digitally based control systems.

Since ratio stations are used with remote-set controllers, some means must be available for setting flow locally during start-up, or during abnormal operation. An auto-manual station is sometimes provided for this purpose. With the ratio controller, this feature requires two scales on the set point mechanism—one reading in ratio for remote-set operation, the other reading in flow units for local-set operation.

When using a divider with linear flow signals as the ratio calculator, the scale factor for the divider should be 1/2. This places a ratio of 1.0 at midscale:

$$R = \frac{1}{2}\frac{B}{A} \qquad \textbf{2.23(10)}$$

Equal flow signals will produce a 50% output from the divider, and the full ratio range is then 0 to 2.0, linear. If flow-squared signals are used, the divider should have a scale factor of 1/3 to provide a full range of 0 to 1.73, with a square root scale. This places a ratio of 1.0 ($A = B$) at 0.58 on the square root scale.

$$R^2 = \frac{1}{3}\frac{B^2}{A^2} \qquad \textbf{2.23(11)}$$

If a scale factor of 1/2 were used, the ratio range would be restricted to 0 to 1.41.

SETTING THE RATIO REMOTELY

When a divider is used to calculate the flow ratio (Figure 2.23c), the set point for this ratio can come from other, related segments of the plant's process. In order to do this correctly, a multiplier must replace the ratio station (Figure 2.23b). If the flow measurement signals are linear, the usual choice of scaling factor for such a multiplier is 2.0:

$$B = 2RA \qquad \textbf{2.23(12)}$$

In this way, a ratio of 1.0 appears at mid-scale of the ratio input ($R = 50\%$), when $A = B$. If squared flow signals are used, a scaling factor of 3.0 provides a ratio range of 0 to 1.73:

$$B^2 = 3R^2A^2 \qquad \textbf{2.23(13)}$$

One reason for using a multiplier to set the flow ratio (Figure 2.23b) is the availability (even in hardware) of very narrow ratio ranges for those applications in which the need for precision, not rangeability, is paramount. Ratio ranges of 0.9 to 1.1, 0.8 to 1.2, etc. are possible.

A typical application would be the accurate proportioning of ammonia and air to an oxidation reactor for the production of nitric acid. If the compositions of the individual feeds are constant, only a very fine adjustment of the ratio is required.

RATIO CONTROL APPLICATIONS

Blending

Figure 2.23e shows a combination of cascade and ratio controls of blending fluids A and B and sending the mix into a blend tank. In this process, the liquid level in the blend tank is affected by total flow, hence the liquid level controller sets flow A, which in turn sets flow B proportionately. (Whenever adjusting the set point of a head-type flow controller in a

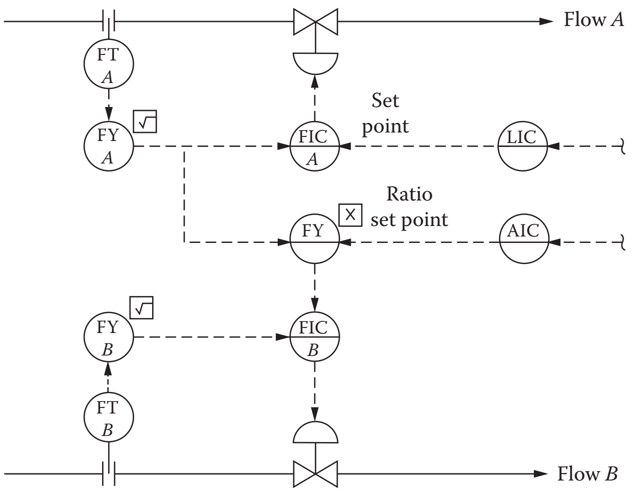


FIG. 2.23e
The level controller manipulates flow rate and the composition controller manipulates flow ratio.

cascade manner, in order to prevent instability at low flows, the square root is removed from the flow measurement signal and linear flow signals are used.)

Conversely, composition is not affected by the absolute value of either flow but only by their ratio. Therefore, to make a change in composition, the AIC controller must adjust the ratio set point of the multiplier (FY). To minimize the interaction of the composition controller with blend tank level control (through its manipulation of flow *B*), flow *B* should be the smaller of the two streams.

Distribution Controls A special case of ratio control is distribution control. When a common supply of some material is to be distributed among three or more destinations, an additional degree of freedom is created. Therefore, in addition to setting the individual flow controllers to some percentage of the total flow, one can also make sure that the system pressure drop is minimized. Because equal flow distribution can be achieved with all the valves nearly open, nearly closed, or anywhere in between, one must introduce an additional control variable.

This new controller (VPC in Figure 2.23f) is usually selected to be the opening of the most-open control valve. If the most-open valve is held nearly fully open while the flows are equally distributed, the control system is fully defined and is also efficient because it provides the minimum resistance to flow. This approach is applicable to all liquid or gas distribution controls, and its use is illustrated by the example of distributing the returning cooling tower water among several cooling cells.

The purpose of the control system in Figure 2.23f is not only to distribute the returning water correctly between cells but also to make sure that this is done at minimum cost of operation. The operating cost in this case is pumping cost, and it will be minimal when the pressure drop through the control valves is minimal. This is the function of the valve position controller (VPC) in Figure 2.23f. As long as even the most open valve is not nearly fully open, the VPC adds a positive bias to all the set-point signals of all the flow fraction controllers (FFICs). As a result, all valves open and keep opening until the most open valve reaches the desired 95%. This technique enables the meeting of the dual goals

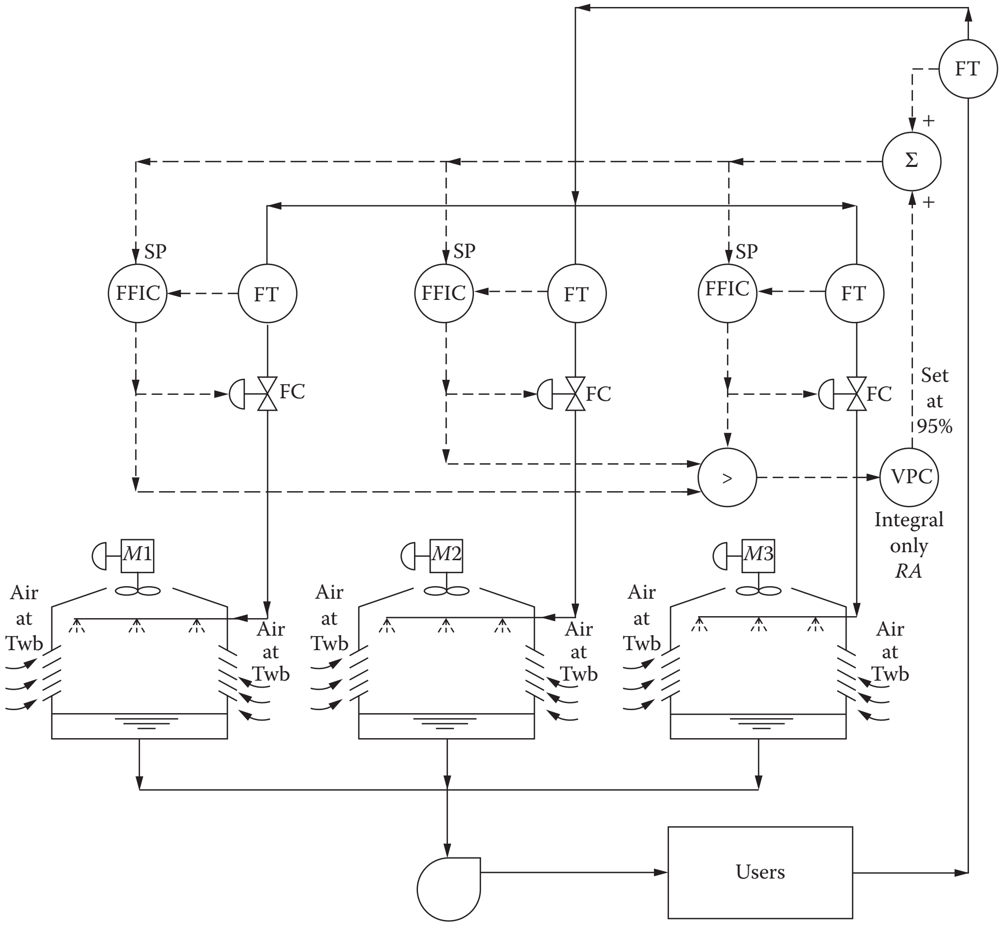


FIG. 2.23f
Pumping costs can be minimized automatically by operating water distribution systems at minimum pressure drop and therefore with the most open control valve nearly fully open.

of providing the correct flow distribution and doing it at a minimum cost in pumping energy.

When cooling tower cells are manually balanced, it is not unusual to find all balancing valves throttled to some extent or to see the same water flow when the fan is on or off. Both of these conditions are undesirable because they will increase operating costs. The savings from automatic balancing can more than justify the instruments required for implementing it.

Surge Control of Compressors

In compressor surge protection applications, the ratio that is controlled is not between flows. The surge line of variable speed compressors closely approximates a parabola if the pressure rise across the machine (*Pd* − *Pi*) is plotted against the volumetric inlet flow rate (Figure 2.23g). If the pressure rise across the compressor (*Pd* − *Pi*) is plotted against the orifice differential (*h*) of a head-type flow sensor located in the inlet to the compressor (Figure 2.23h), the surge line becomes an almost straight line.

Because surge occurs at low flows, the control system is so configured as to keep the compressor operations to the right of the surge line. Therefore, the set point for the surge protection FIC is the surge line itself ($m(\Delta P)$) plus a safety bias (b). This in effect locates the control line parallel with and to the right of the surge line.

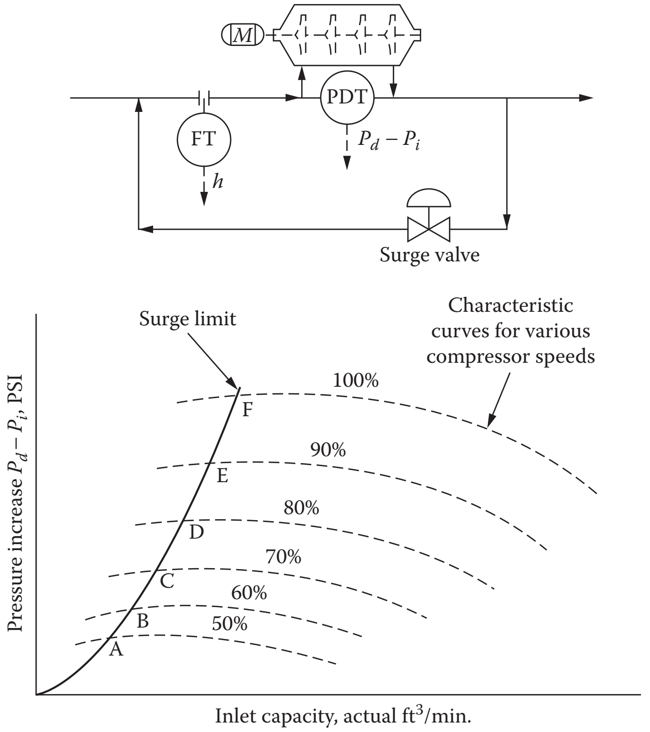


FIG. 2.23g
The surge line of a variable speed centrifugal compressor is parabolic if the pressure rise across the machine (Pd − Pi) is plotted against inlet volumetric flow rate.

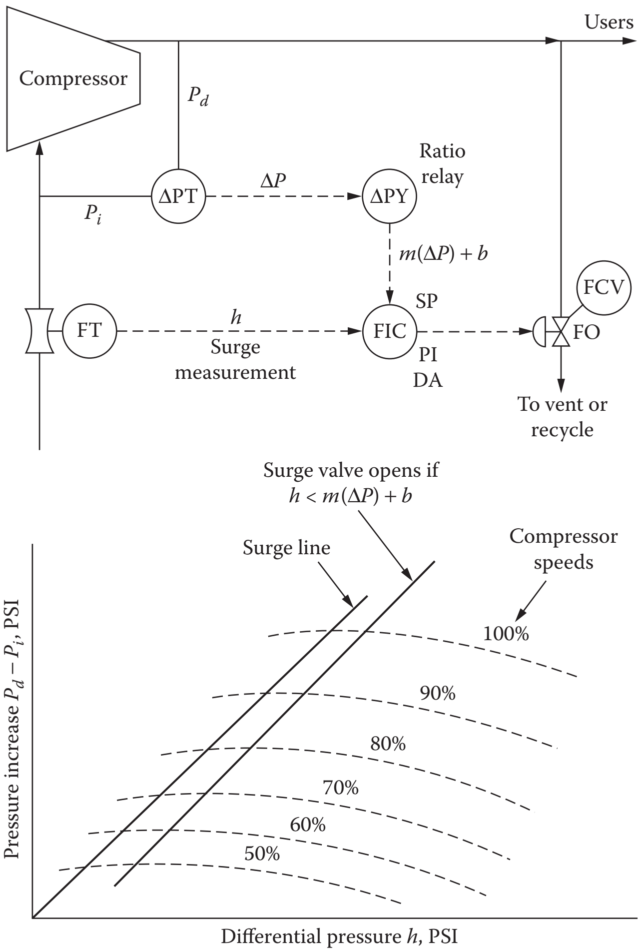


FIG. 2.23h
The parabola shape of the surge line changes to a straight line if the pressure rise across the compressor (Pd − Pi) is plotted against the orifice differential (h) of a head-type flow meter.

The slope *m* of the surge line is

$$m = (Pd - Pi)/h \qquad \textbf{2.23(14)}$$

where (*Pd* − *Pi*) is the head of the compressor and *h* is the differential pressure across the flow element. A safe set point for the flow controller (FIC) is

$$h = (Pd - Pi)/m + \text{bias} \qquad \textbf{2.23(15)}$$

The complete surge control loop is shown in Figure 2.23h. The figure shows both the surge line and the control line at which the surge valve opens.

Bibliography

Balchen, J. G., and Mummè, K. I., *Process Control Structures and Applications,* New York: Van Nostrand Reinhold, 1988.

Beard, M. J., "Analog Controllers Develop New Wrinkles," *Instruments and Control Systems,* November 1977.

Buckley, P. S., "Dynamic Design of Pneumatic Control Loops: Parts I and II," *InTech,* April and June 1975.

Buckley, P. S., *Techniques of Process Control,* New York: John Wiley & Sons, 1964.

Eckman, D. P., *Automatic Process Control,* New York: John Wiley & Sons, 1958.

Erickson, K. T., and Hedrick, J. L., *Plant-Wide Process Control,* New York: John Wiley & Sons, 1999.

Harriott, P., *Process Control,* New York: McGraw-Hill, 1964.

Jones, B. E., *Instrumentation, Measurement and Feedback,* New York: McGraw-Hill, 1977.

Korn, G., *Interactive System Simulation,* New York: McGraw-Hill, 1989.

Luyben, W., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Luyben, W. L., and Luyben, M., *Essentials of Process Control,* New York: McGraw-Hill, 1997.

Marlin, T., *Process Control: Designing Processes & Control Systems for Dynamic Performance,* New York: McGraw-Hill, 2000.

McAvoy, T. J., *Interaction Analysis,* Research Triangle Park, NC: ISA, 1983.

Merritt, R., "Electronic Controller Survey," *Instrumentation Technology,* June 1977.

Murrill, P., *Application Concepts of Process Control,* Research Triangle Park, NC: ISA, 1988.

Murrill, P., *Fundamentals of Process Control,* 2nd ed., Research Triangle Park, NC: ISA 1991.

Nachtigal. C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Roffel, B., and Chin, P., *Computer Control in the Process Industries,* Chelsea, MI: Lewis Publishers, 1989.

Rowan, D., "On-Line Expert Systems in Process Industries," *AI/Expert,* August 1989.

Seborg, D., and Mellichamp, D. A., *Process Dynamics and Control,* 2nd ed., New York: McGraw-Hill, 2004.

Shinskey, F. G., "Control Topics," Publ. no. 413-1 to 413-8, Foxboro Company, 1976.

Shinskey, F. G., *Distillation Control,* 2nd ed., New York: McGraw-Hill, 1984.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Smith, C. A., *Automated Continuous Process Control,* New York: John Wiley & Sons, 2002.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

White, M. H., "Surge Control for Centrifugal Compressors," *Chemical Engineering,* Dec. 25, 1972.

2.24 Real-Time Performance Assessment

J. P. GERRY (2005)

INTRODUCTION

To optimize the regulatory control loops in a plant, one first needs to assess the current performance. How are the loops operating now? Could they be performing better? By assessing the performance in real time, a number of benefits can be obtained:

- Knowledge, on a real-time basis, of the health of the plant, unit operations, and loops can be acquired.
- Poor-performing loops can easily be flagged.
- Continual performance assessment allows one to record the history of performance, so that differences in performance between time periods can be compared. For example, one can evaluate the effect of the following variables on performance:
  - Seasonal changes
  - Grade changes
  - Feed changes
  - Shift changes
  - Equipment changes

The tools of optimization should be readily accessible in the performance monitor.

PLANT ASSESSMENT AND MONITORING

Process variable, set-point, and controller output data should be collected from the plant for every control loop. In order to maintain accuracy, these data should be collected while all software compression is turned off. Ideally, the data should be collected at intervals of roughly 1/10 of the dead time of the loop. Generally once per second is adequate, but the collection frequency of some loops such as temperature can be longer. As of this writing, standard connection methods are available using OPC or Ole for Process Control, which is a standard developed by the OPC Foundation (www. opcfoundation.org).

Loop Performance and Health Assessment

The overall performance of a control loop or loop health may be different for different businesses. It can also vary from plant to plant depending on the business objectives. For example, in paper mills, the variability of the paper is often a key variable to be assessed, whereas in refineries the average temperature error can be a key of assessment.

The definition of loop health can also vary from loop to loop. For example, a flow loop may have different performance objectives than the level in a surge tank. A flexible definition is the best since it allows for improvements and refinements over time.

Loop health can be evaluated by considering the key assessments, baselines, thresholds, and economic significance.

Assessment Criteria of Loop Health Assessments are calculated over "assessment intervals." This interval should be defined for each unit operation and should be at least 6 times longer than the longest cycle in the unit or at least 100 times the longest process dead time. Generally, the assessment interval should be 4 hours or longer.

Loop health can be evaluated and assessed on the basis of many criteria, and the reader should be familiar with the related terminology, described below in alphabetic order. Any of the following criteria or their combinations can be used as key assessments, which refer to criteria that are relevant to the health of a control loop:

- *Average error* is the average of the absolute error that has existed between the setpoint and measurement during the assessment period.
- The *Harris Index* (normalized) is defined as the ratio between the error (PV – SP) variance during this assessment over the variance achievable by a minimum variance controller. The larger the value of the Harris Index, the poorer the performance of the loop. The Harris Index will vary as a number between one and infinity. A value of one is perfection or minimum variance control. As the quality of loop performance drops, the index rises. However, sluggish tuning usually causes the Harris Index to drop. Cyclic response or overly aggressive tunings will make the Harris Index rise. The Harris Index is considered to be high when it is between 50 and 100, but it can get much higher. However, these numbers may vary from loop to loop and also depend on the occurrence of set-point changes and load upsets. The amount and pattern of noise can greatly influence the Harris Index.

- *Noise band* is determined on the basis of statistical analysis of the noise, which during the assessment period is superimposed on the process variable, expressed in percentage. The noise band is the standard deviation (sigma) of the process variable during the quietest time. The minor variations in the process variable that are not caused by the controller, by set point changes, or by load upsets are considered to be normal process noise. Such noise may be caused by electrical interference, magnetic fields, flow turbulence loops, waves in a tank in level loops, etc. Generally, the larger the noise band, the poorer the performance of the loop. The cycling of a process variable in an unstable and fast control loop might also appear as noise to this assessment.
- *Oscillation detection confidence value* is a value between 0 and 100%, where 100% indicates full confidence that the loop is continuously oscillating. This assessment is performed by analyzing the pattern of the error signal.
- *Oscillation diagnosis confidence value* is between 0 and 100%, where 100% indicates full confidence that the oscillation is caused by one of the three causes: valve stiction or hysteresis, bad tuning, or causes external to this loop.
- *Output standard deviation* is the standard deviation in the controller output during the assessment period.
- *Output at a limit* is the percentage of time that the controller output was at a limit.
- *Probable Performance Increase Index* is a measure of the performance increase, which is possible by using the fastest PID tuning with a safety factor of 2.5. It is the "percent probable" performance increase in integrated absolute error (IAE) terms, when the loop is responding to a load upset, assuming the controller does not overshoot with either settings. This index shows the improvement in process response to upsets with the new tuning. This index is often proportional to the savings that the plant can achieve from improved control loop tuning.
- *Set-point crossings* is the number of times the process variable (PV) crosses the set point in one day.
- *Time in Normal* is the term for the time period in percentage that the loop spends in the Normal mode. Most often the Normal mode is the Automatic mode, but not always.
- *Valve travel* is the total distance the controller output moves during the assessment period, which is normalized to be on a 24-hour basis. It is found by summing the absolute value of the controller output changes that occur between each sample.
- *Valve reversals* is defined as the number of times the controller output changes direction during an assessment period normalized to a 24-hour basis. It is a value between zero and infinity. This value is of interest because more reversals usually result in more valve wear.
- *Variability* is the relative value of variance (defined in the next paragraph) in the process variable during the assessment period. It is expressed as a percentage of the mean and so allows comparison between the level of variability in different processes.

$$\text{Variability} = 2(\text{Standard Deviation})(100/\text{Abs(Mean)}) \qquad \textbf{2.24(1)}$$

where Abs = absolute value; Standard Deviation = sqrt(Variance); and Variance = $\text{sum}(\text{Mean} - x(i))^2/(\text{npts} - 1)$.

- *Variance* measures the spread or dispersion of the process variable during the assessment period. "Variance, normalized" is dimensionless because it has been normalized with 100 divided by the span of the PV, squared:

$$\text{Variance} = [\text{sum}(\text{Mean} - x(i))^2\,(100)^2]/[(\text{npts} - 1) \times (\text{span of PV})^2] \qquad \textbf{2.24(2)}$$

Related Assessments Criteria

The following assessments provide beneficial diagnostics but are not contributors to loop health.

- *Three largest oscillation periods and strengths.* This value is obtained by spectral analysis of the error data during the assessment time. The three oscillations of the process variable having the most power (and their amplitudes) provide the six assessments. The power should be evaluated using a cluster of peaks and not just a single peak, and the zero frequency component should be ignored. This assessment can provide some very beneficial diagnostics but is not viewed as a contributor to loop health. For example, if the cause of loop oscillation is unknown, by sorting the loops by oscillation period, one can identify the loops having the same period of oscillation — even when it is "hidden" in other information. This can help in pinpointing the cause of oscillation.
- *Process model parameters.* These are parameters of the process model — dead time, gain, lag times, and integration time, including second-order coefficients that allow for imaginary roots. The process model is usually not a key assessment of control loop health.
- *Model quality.* This is an assessment of the accuracy of the fit between the actual process and its model. The quality of the process model is usually not a key assessment criterion in the evaluation of the health of the control loop.
- *Robustness.* Robustness is a relative number giving an indication of how robust or sensitive the loop is to process upsets or changes. The larger this number, the more robust the loop. There is always a tradeoff

between fast tuning and robustness. The robustness number is discussed in detail in Section 2.26 in this handbook. The robustness number is normalized since it uses gain and dead time ratios. It can be used as a comparison between loops.

- The *relative response time* (*RRT*) is a relative indicator of the speed of the control loop. The smaller the relative response, the faster the loop. The higher the relative response value, the slower the loop. This number is used for comparison purposes. The RRT is dependent on the controller's speed of response and can be changed by adjusting the PID settings. If one uses fast tuning, the controlled process variable often will oscillate around the set point after it is changed. The period of this oscillation in the time response is roughly the same as the RRT. The value of the RRT is found by:
  1. Calculating the closed loop frequency response to a load upset
  2. Finding the frequency where the amplitude ratio peaks
  3. Converting this to a time period

  Interacting loops and RRT. In order to prevent interaction, the PID tuning settings should be so adjusted that the RRTs of the loops differ by a factor of three. For example, if there are three interacting loops and the fastest one has a period of 1 second, the tuning of the other two controllers should be readjusted so that their RRTs are at least three and nine.

  Cascade Loops and RRT. For cascade loops, the master (outer) loop should have an RRT that is three times slower than that of the inner loop. This relationship is obtained by first tuning the slave (inner) loop, and then the master. If the cascade master loop has an RRT that is less than 3 times that of the slave (inner) loop, the master should be de-tuned so that its RRT will be 3 times that of the slave.

Baseline and Threshold Values The baseline time is a window or period of time when the plant is running well. The baseline time can be used as a reference for comparison. The length of the baseline period should exceed six and should preferably equal 30 assessment periods. For statistical methods of evaluation, at least 30 assessments are suggested.

Every assessment can have a baseline value. A baseline value represents either the ideal value or the value from a time when the plant was running well. For example, a baseline value for average error might be zero, since zero error is ideal. The baseline for process dead time might be the average or minimum dead time that is present during the baseline time period.

Thresholds represent values of the assessed criteria that one would prefer not to exceed. Thresholds represent limits or boundaries between which the assessment criteria would remain if the plant is running well. All thresholds should be bidirectional. For example, a threshold may be set above the baseline and/or below the baseline.

For example, a threshold for variability might be calculated by looking at a number of assessments from a period when the plant was running well. In that case, one should find the standard deviation and average value of the variability during these assessments. The standard deviation is also called the sigma value. X number of sigma values over the average value is called the X sigma value. So setting the threshold of variability to six standard deviations over the average would be setting it to the six sigma value. This is one strategy for setting thresholds.

Thresholds could also be set to the minimum or maximum value of an assessment over the baseline time.

Loop Health Calculation The calculation of loop health should be flexible enough to allow for different interpretations depending on the nature of the plant or process. It should allow for different categories of loops to have different health requirements. For example, a temperature loop may need to be fast and responsive whereas a level loop may need to allow for large level variations before causing a change in flow.

Loop health calculation should also allow different loops to have different priority ratings based on the relative importance of the loops.

Key Assessment Criteria

In order to quantitatively evaluate the performance of a loop, one has to select assessments that in combination define the loop's health. These are called key assessments. The key assessments should be those that are important to the operation of the plant, and any assessment can therefore be selected to be a key assessment.

Percent towards Threshold Every assessment can potentially provide a contribution to describing the health of the loop. For each assessment during every assessment period, a "Percent towards Threshold" is found:

$$\text{Percent towards Threshold} = (\text{Current Value} - \text{Baseline})/(\text{Threshold} - \text{Baseline})100 \qquad \mathbf{2.24(3)}$$

Next, an "Average Percent towards Threshold" is found. The Average Percent towards Threshold is the average of all the Percent towards Threshold values for all the key assessments during an assessment period. Using this calculation, lower values of Average Percent towards Threshold mean a better-performing loop. Larger values mean a poorer-performing loop. A patent is pending for this technology, which is used in performance assessment.

$$\text{Loop Health} = (\text{Average Percent towards Threshold})/(\text{Economic Significance}) \qquad \mathbf{2.24(4)}$$

Economic Significance is a divisor that incorporates a factor of economic importance to the health of each loop. For example, if two loops are exhibiting the same "Average Percent towards Threshold" the one with the higher economic significance (lower divisor) will bubble to the top of a priority list of loops.

Historization of Assessments The assessments for every loop should be historized. This allows for performance comparisons at different time periods. This information can be very valuable in troubleshooting and optimizing operations.

Accessibility and Presentation of Results Currently, the most accessible method for distribution of the results seems to be via a browser-based interface. This means the performance monitoring system should be a Web server; it should be able to serve up the performance results in an easy-to-read format for plant personnel. Most plant personnel are already familiar with navigation in browser interfaces and Web pages. Any personnel with access to a computer that is connected to the plant local area network (LAN) will have access to such information.

A network bridge allows this same connectivity on the office LAN. A network bridge uses two network interface cards in the server machine running the plant monitoring software. Using two cards enables the plant LAN to be totally separate from the office LAN. A Web server running the monitoring software can send assessment results to the office LAN.

Results should be shown in both graphical and tabular form. All of these results should be up to date on a real-time basis as much as possible.

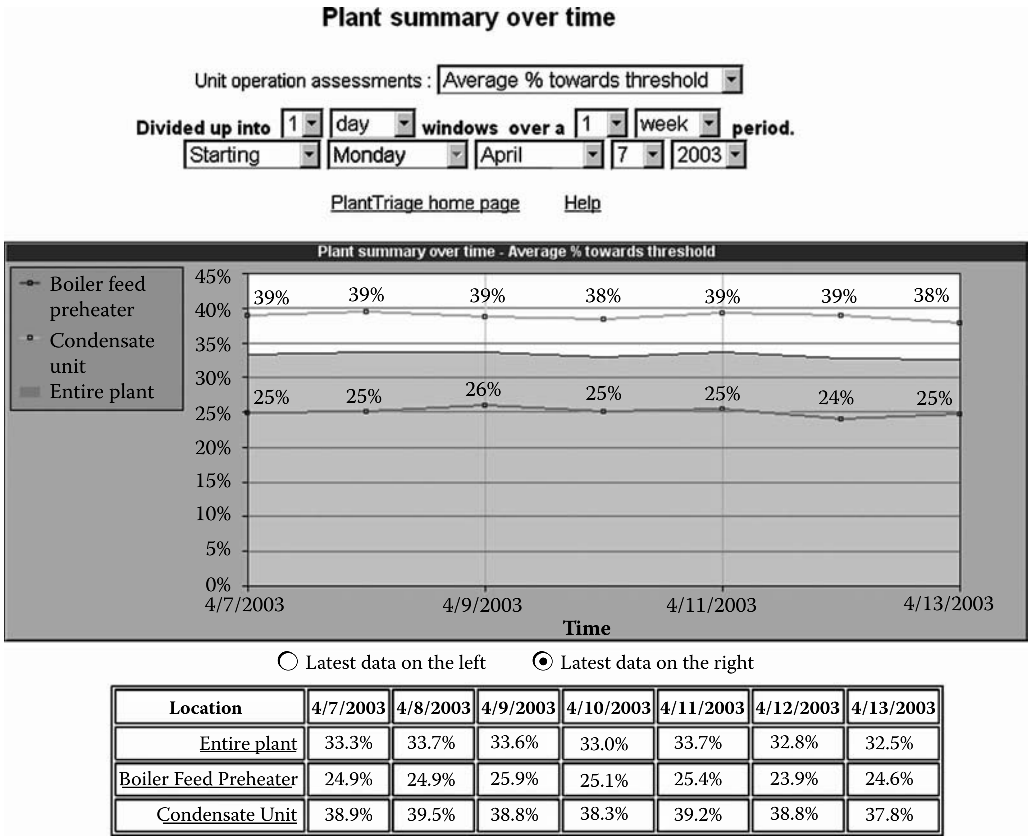
Plant Overview

An overview window should provide an assessment of loop health in the various sections of the plant and also in the overall plant. Each plant area should have a history of assessments in that area. For example, Figure 2.24a shows a table and plot of loop health for 1 week. The two lines at the top and bottom represent two plant sections (unit operations), while the line in the middle describes the health of the entire plant. In the table below the chart, the health of the two unit operations making up the plant is listed as the third and fourth rows.

Prioritization of Problem Areas

The interface for users of assessment software should allow for default and for the generation of priority lists, which are configured by the user. The user should be able to sort the control loops in the plant on the basis of any combination of assessment criteria chosen by the user.

Priority List of Loops Having Most Impact One of the lists that is generated includes, in their order of priority, the loops needing attention. Both the engineering and plant maintenance



Plant summary over time

Unit operation assessments : Average % towards threshold

Divided up into 1 day windows over a 1 week period.

Starting Monday April 7 2003

○ Latest data on the left ● Latest data on the right

Location	4/7/2003	4/8/2003	4/9/2003	4/10/2003	4/11/2003	4/12/2003	4/13/2003
Entire plant	33.3%	33.7%	33.6%	33.0%	33.7%	32.8%	32.5%
Boiler Feed Preheater	24.9%	24.9%	25.9%	25.1%	25.4%	23.9%	24.6%
Condensate Unit	38.9%	39.5%	38.8%	38.3%	39.2%	38.8%	37.8%

FIG. 2.24a
Visual and tabular overview of the health of unit operations in a plant.[1]

Biggest payback loops

Refresh PlantTriage home page Snoozed loops Help

Loop name	Unit operation	Description	Average economic assessment	Snooze loop
FIC100	Condensate unit	Condensate return	34.0%	Snooze
TIC101	Boiler feed preheater	Preheat temperature	33.5%	Snooze
LIC100	Condensate unit	Knockout drum level	23.0%	Snooze
FIC101	Boiler feed preheater	Preheat steam condensate flow	14.7%	Snooze
PIC100	Condensate unit	LP Steam makeup	0.5%	Snooze

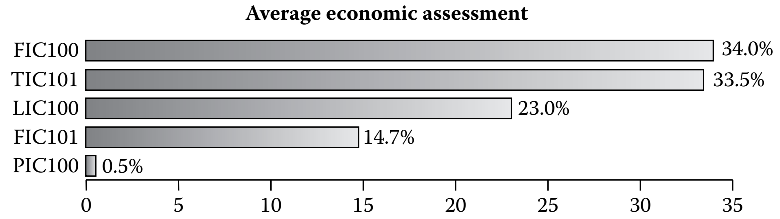


FIG. 2.24b
A sample list with the control loops listed in order of their need of engineering or maintenance attention.[1]

departments should find this list useful. The order of the loops in the list should reflect their potential for improving the productivity of the operation. To allow the user to obtain more information about the various unit operations, a clickable "drill down" can be provided, with more detail about the area in question. Figure 2.24b shows an example of such a page.

Listing Oscillating Loops The system should also include a list of oscillation assessments. Time of oscillation detection, diagnosis, period, and strength are some of the valuable data collected. Sorting by "oscillation detection" shows all loops that are oscillating.

Sorting by oscillation period groups the loops by their periods of oscillation. If the oscillation of several loops is caused by the same problem, they are likely to have similar oscillation periods. The top three oscillation periods for each loop should be included in the sort so that any secondary oscillations can be picked up.

Model and Quality Loop List A list containing columns of Model Quality and Probable Performance Increase Index makes it possible to improve the tuning of many PID loops all at once. In the list, high model quality means that good identification exists and hence good PID tuning is possible and is readily available for that loop. For those loops with a high model quality and where the Probable Performance Increase shows that better tuning is available, the improvements can be made to several loops simultaneously.

In such cases, the user, using the analysis software for each loop, would request the data used by the performance monitoring software to arrive at the model. As is discussed in more detail later, the analysis software displays the optimal PID tuning along with time simulations and robustness analysis.

Custom Loop Lists The loop performance assessment software allows users to create their own lists of loops. The rows and columns are selected by the user and can contain both custom loop lists and custom assessment lists as they are chosen by the user. Once set up, the software allows the user to recall the custom loop lists.

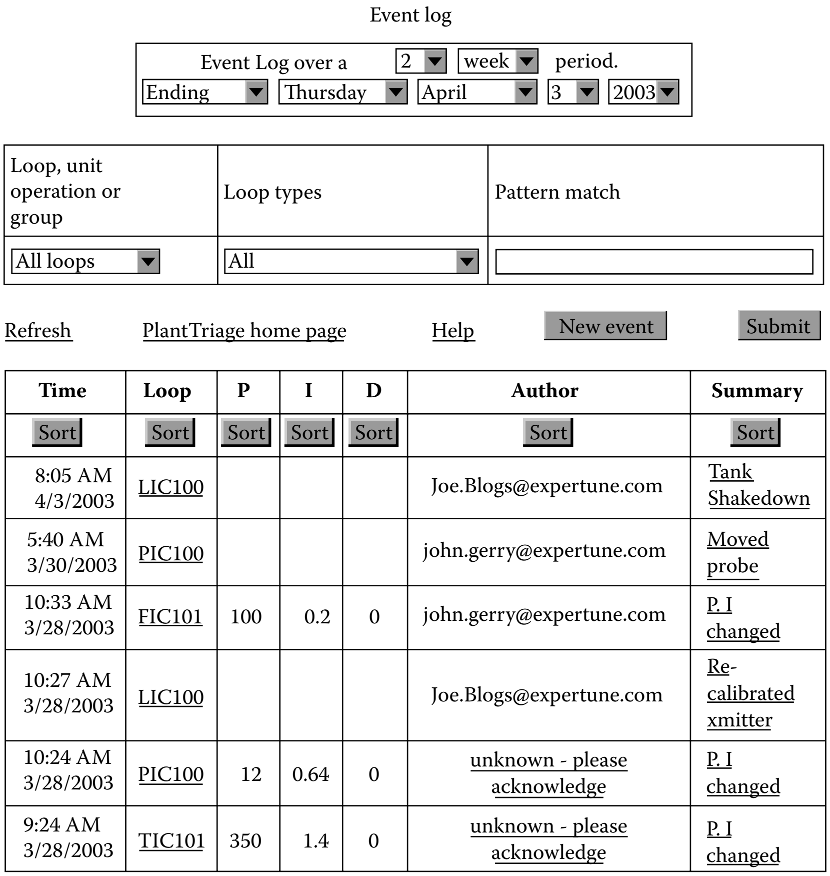
Recording Tuning Changes PID tuning values and changes are recorded by the performance monitor. A time window of the history of these changes is available for viewing in the browser-based Web interface. This feature also allows the user to enter notes or events. When several tuning adjustments are made to a PID loop, this information can be used as an indication that a problem exists with that loop.

Figure 2.24c shows events that were automatically picked up by the performance monitor and also shows some user-entered events. The data in the rows containing PID settings were detected and automatically entered by the system.

Performance Monitoring Software Capabilities

A brief listing is provided here of the types of testing and reporting that should be available as part of the performance monitoring software:

- *Stiction and hysteresis testing and reporting.* Many plant oscillation problems are caused from poorly operating valves
- *Valve wear analysis.* This analysis compares valve travel and reversals with differing PID tuning values and with and without the use of various types of process variable filters.



Event Log over a 2 week period.
Ending Thursday April 3 2003

Loop, unit operation or group	Loop types	Pattern match
All loops	All	

Time	Loop	P	I	D	Author	Summary
Sort	Sort	Sort	Sort	Sort	Sort	Sort
8:05 AM 4/3/2003	LIC100				Joe.Blogs@expertune.com	Tank Shakedown
5:40 AM 3/30/2003	PIC100				john.gerry@expertune.com	Moved probe
10:33 AM 3/28/2003	FIC101	100	0.2	0	john.gerry@expertune.com	P. I changed
10:27 AM 3/28/2003	LIC100				Joe.Blogs@expertune.com	Re-calibrated xmitter
10:24 AM 3/28/2003	PIC100	12	0.64	0	unknown - please acknowledge	P. I changed
9:24 AM 3/28/2003	TIC101	350	1.4	0	unknown - please acknowledge	P. I changed

FIG. 2.24c
Part of an event log listing.[1]

- *PV filter analysis.* This analysis finds the largest possible filter that will reduce valve movement but will not inhibit speed of response or reduce robustness.
- *Unfiltering capability.* The software should have the ability to remove the effects of a PV filter so the unfiltered PV can be viewed graphically by the user and the analysis software.
- *Model identification and display.* The model is used by the analysis software for time simulations, robustness analysis, and valve wear analysis. The model should be identified with minimal prior user knowledge of the dynamics; the software should find the best model structure and dead time without the user estimating these. The model can be used by model predictive controllers.
- *Optimal PID tuning.* A variety of methods including set point or lambda tuning should be used to minimize load upsets (minimum IAE).
- *Robustness*. There is always a tradeoff between tuning for fast response and sensitivity to process load changes. Robustness analysis is critical to providing practical PID tuning settings for the loop. Robustness analysis should reflect the changes that occur in process dead time and process gain.
- *Simulating the loop performance using a variety of tuning settings and PV filter values.* Simulations must include both set-point response simulations and responses to load upsets. The simulations should display and compare the performance with both the existing and the proposed settings. Simulations should describe the response of the loop to noise and should compare the response of the model of the process to that of the real process.
- *Linearization analysis including characterization and characterizer builder.* This allows linearization of nonlinear loops through the use of a characterizers. This software should be suitable to both the types of processes whose linearity changes with load and those whose nonlinearity depends on setpoint (pH).
- *Power spectral density.* This allows a detailed analysis plot of power in a signal at all frequencies or periods.
- *Statistical analysis including a histogram.* This allows for the generation of statistical comparisons to evaluate and document the benefits of work done.
- *Time line analysis.* This analysis describes time relationships between components.
- *Integrated reporting.*

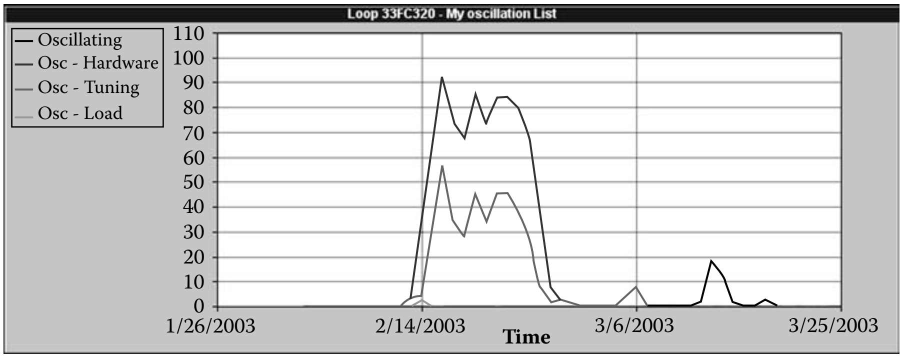


FIG. 2.24d
The oscillation of a flow loop on a paper machine was detected by the assessments of "oscillation detection" and "oscillation from tuning," signaling a need to re-tune the PID when the paper grade is changed.[1]

CASE STUDIES

Model Predictive Controlled Refinery

In a refinery, model predictive controllers (MPC) tend to move the process toward its constraints as far as possible to maximize the resulting economic gain. How close the controller can push the process to its constraints is related to setpoint crossings, average absolute error, and variability. These criteria of assessments infer the ability of the control loop to closely approach the constraint. High-quality performance is indicated when this assessment value is low and if the baselines are lower than thresholds.

In order for an MPC controller to work, the loops tied to it need to be in their normal operating mode and not at a limit. The assessments "time in Normal" and "output at a limit" assess the performance of these. Time in Normal should be maximized and should have a high baseline and low threshold whereas output at a limit should be minimized: it should have a low baseline.

In a fluidic catalytic cracker the side valves are very expensive—costing upwards $100,000 each, and the entire unit must be shut down when the valve needs service. Therefore, the assessments of "valve reversal" and "valve travels" can be used to measure the wear and tear on these valves. The lower the assessment values the better. Baselines should be low and thresholds high for these assessments.

Grade Change in a Pulp Mill

Figure 2.24d describes the assessment of the oscillation of a flow loop in a pulp mill. The two lines reflect the assessments of "oscillation detection" and "oscillation from tuning" or a diagnosis that the oscillation was caused by the PID controller with the wrong tuning settings.

Later, it was learned that the plant changed the paper grade, which would have required re-tuning and resulted in oscillation. It was noted that oscillation can be avoided if the tuning settings are automatically changed whenever that grade of paper is run.

CONCLUSIONS

Assessing the performance of a plant in real time requires the careful selection of the appropriate software, but the benefits are also compelling. Process plants will benefit from considering this option.

Reference

1. These visual aids are provided courtesy of ExperTune Inc.

Bibliography

Gerry, J., "Prioritizing and Optimizing Problem Loops Using a Loop Monitoring System," ISA, Chicago, October 2002.

Gerry, J., "Process Monitoring and Loop Prioritization Can Reap Big Payback and Benefit Process Plants," ISA, Chicago, October 2002.

Mansfield, E., *Basic Statistics with Applications,* New York: W.W. Norton, 1986.

Paulonis, M., and Cox, J., "A Practical Approach for Large-Scale Controller Performance Assessment, Diagnosis, and Improvement," *Journal of Process Control,* Vol. 13, 2003.

Ruel, M., "Learn How to Assess and Improve Control Loop Performance," ISA, Chicago, October 2002.

2.25 Relative Gain Calculations

B. D. CAMPBELL, P. G. FRIEDMANN (1995) **F. G. SHINSKEY** (2005)

Most processes require multiple control loops to operate efficiently, and these loops usually do not act independently. Changes made in the output of one controller can be a source of upset to another loop. That interaction may pass in one direction only; for example, changing the feed rate to a process upsets the composition of the product leaving it. However, there are also processes where the disturbed controller responds by sending another disturbance back to the first loop. For example, a top-composition controller on a distillation column manipulating reflux flow can upset bottom composition, whose controller manipulating heat input reacts by upsetting top composition. This *bidirectional* interaction comprises an additional feedback loop passing through both controllers in series and is a cause of instability to the process.

When two or more manipulated variables affect a single controlled variable, or when a single manipulated variable affects two or more controlled variables, a question arises of how best to pair the manipulated and controlled variables together in closed loops. With a two-loop process, only two possible single-loop structures exist: variable c_1 can be controlled by manipulated variable m_a and variable c_2 controlled by m_b, or vice versa. If one pairing, say c_1-m_a and c_2-m_b, performs well, the other pairing will not and may even be unstable. It is also possible that both pairings will perform equally poorly. In some cases, the more-effective pairing may be obvious, but in others it will not. The number of possible single-loop structures increases factorially with the number of loops, reaching 120 for a five-loop distillation column.

The relative gain method of interaction analysis solves this problem, providing a quantitative measure of loop interactions that is surprisingly easy to use and universally applicable. The relative gains calculated are dimensionless numbers arranged in an array, which by their values indicate the best single-loop pairings and where decoupling is advisable. The method was first proposed by Bristol,[1] and it was then named and applied by Shinskey.[2,3] Its applications include integrating processes[4,5] and those with unequal numbers of manipulated and controlled variables.[6] It has been extended to include dynamic considerations[7] in the analysis of interactions.

Much of the Relative Gain Array (RGA) literature is devoted to distillation,[8–10] particularly to top and bottom composition control. Relative gain calculation for distillation applications is covered in Chapter 8 of this book. Other applications include power systems, blending, combustion, and heat transfer controls.

This section presents the computation, interpretation, and application of relative gain arrays. It also provides examples that illustrate RGA application to simple processes.

THE RELATIVE GAIN ARRAY

The relative gain for a selected pair of variables is defined as the ratio of the open-loop gain for that pair with all other loops open to their open-loop gain when all other loops in the process are closed, with their variables held at set point by their controllers:

$$\lambda_{ij} = \frac{\left.\dfrac{\partial c_i}{\partial m_j}\right|_m}{\left.\dfrac{\partial c_i}{\partial m_j}\right|_c} \qquad 2.25(1)$$

where λ_{ij} is the relative gain of any pair of variables c_i controlled by manipulating m_j. The numerator is the open-loop gain determined with all other manipulated variables constant, and the denominator is the open-loop gain determined with all other controlled variables constant, which literally applies to a steady state. Therefore the above applies strictly to steady-state gains.

Because the same dimensions appear in both numerator and denominator, the relative gain numbers are dimensionless. They are also independent of scale and of common terms appearing in both open-loop gains. They are unaffected by nonlinearities at any given operating point. If the process is nonlinear, however, the calculated relative gains can be expected to change with the operating conditions.

The RGA is a table of λ_{ij} values arranged in rows c against columns m:

$$\Lambda = \begin{array}{cccccc} & m_1 & . & . & m_n \\ c_1 & \lambda_{11} & . & . & \lambda_{1n} \\ . & . & . & . & . \\ . & . & . & . & . \\ c_n & \lambda_{n1} & . & . & \lambda_{nn} \end{array} \qquad 2.25(2)$$

where n is the number of variable pairs in the system. An important property of the RGA is that the numbers in each of the columns and each of the rows add to 1.0. To fill out a 2×2 array, then, only one λ needs to be calculated; the others in the same row and the same column are complementary, and the diagonal value is identical. To fill out a $n \times n$ array then, $(n-1)^2$ values must be calculated. An important corollary to this property is that for every λ that exceeds 1.0, there must be a negative number in the same row and in the same column, and control-loop pairs with negative numbers, as described below, must be avoided.

Properties of λ

The values of λ cover all possible numbers from –infinity to + infinity. They can be separated into six distinct regimes, each with its own set of properties.

λ = 1.0 The most desirable value of λ is 1.0, which indicates that numerator and denominator of Equation 2.25(1) are equal. This means that the closing of other loops has no effect on the behavior of the loop being examined. So in evaluating an RGA, it is advisable to connect those loops having relative gains close to 1.0. This is not an absolute guarantee of best performance because it does not include dynamic considerations, but it does guarantee steady-state stability. One-way interaction could still exist: if either m_j does not upset other controlled variables, or c_i is not affected by other manipulated variables, the relative gain will be 1.0. But without two-way interaction, no hidden feedback loops exist, and hence the stability of the loop in question is not affected by the closure of other loops.

λ = 0 If λ is zero, it is because the numerator in Equation 2.25(1) is zero, but the denominator is not. In this case, the loop in question cannot be closed successfully alone, but could possibly control quite well when the other loops in the system are closed. This is usually the case when one of the other controlled variables is liquid level. A continuous process must have its level loops closed before any of the others can function properly. In a complex multivariable process such as distillation, it is usually necessary to assign the level loops first and then evaluate the relative gains of the composition loops. This eliminates zeros from the RGA. This procedure is described in detail in Section 8.20.

λ = ±∞ As the denominator in Equation 2.25(1) approaches zero, λ will approach infinity, and a plus-infinity anywhere in the array will require a minus-infinity in the same row and in the same column. These loops cannot be controlled independently because the variables have essentially the same effects on each other. While calculating a result of infinity is not always achievable, any relative gain whose absolute value exceeds 100 qualifies — the variables are essentially dependent and the loops should not be closed.

λ < 0 Negative relative gains indicate the presence of conditional stability because the numerator and denominator in Equation 2.25(1) have opposite signs. If the loop in question has been configured to be a negative feedback loop while other loops are open, it will become a *positive feedback loop* when the others are closed. Positive feedback causes the controlled variable to run away from set point rather than approach it and so must be avoided in a regulatory system.

It is possible to reconfigure the loop in question to have negative feedback when the other loops are closed by reversing the action of one controller. But its dynamic response will be poor because it contains additional process elements, along with other controllers. However, the principal danger is that the opening of another loop, either by placing its controller in manual or by its output reaching a limit, will leave the reconfigured loop as positive feedback. Therefore, loops with negative relative gains should never be connected.

0 < λ < 1.0 Relative gains in the zero to one range indicate the existence of an additional *negative* feedback loop formed by interaction. If $\lambda = 0.5$, that loop is as strong as the intended single loop, and lower values indicate it is even stronger. To return the intended loop to the same stability it had when operating alone, the proportional gain of the controller would have to be multiplied by λ. The additional feedback loop contains more dynamic elements as well, requiring an increase in integral time. Two-by-two systems where all the relative gains are 0.5 are common; an example is control of boiler pressure and flue-gas oxygen by fuel and air flows. Since neither configuration of single loops is effective, decoupling is required. In the case of the boiler, flue-gas oxygen content is controlled by the fuel-to-air ratio.

λ > 1.0 Relative gains > 1.0 indicate the existence of an additional *positive* feedback loop formed by interaction. As that loop increases in strength, λ also increases. (When the positive feedback loop dominates, λ is negative.) The positive feedback loop has 180 degrees less phase shift than the negative feedback loop and therefore does not significantly slow the dynamic response of the negative feedback loop. In the case of two interacting loops having similar dynamics and $\lambda = 10$ or more, both controller gains need to be halved when both loops are closed, but integral and derivative settings may be left as they were for single-loop operation.

CALCULATION METHODS

The RGA can be evaluated in a number of different ways, with the choice depending on the availability of information about the process to be controlled.

The Ratio of Partial Derivatives

This method solves Equation 2.25(1) directly for each of the needed $(n-1)^2$ elements in the array. The numerator is obtained by a partial derivative of c_i with respect to m_j in

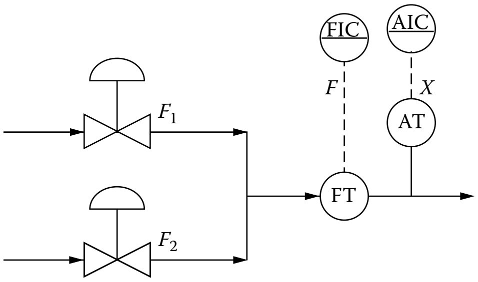


FIG. 2.25a
A simple two-component blender has two interacting loops.

terms of the other manipulated variables held constant. As an example, consider the blending of two streams into a single product of composition x flowing at rate F, as shown in Figure 2.25a. The total flow is the sum of the two manipulated flows:

$$F = F_1 + F_2 \tag{2.25(3)}$$

where the individual flows are expressed as linear with controller output. (Relative gain is independent of nonlinearities and common factors, allowing such simplifications.) Then,

$$\left.\frac{\partial F}{\partial F_1}\right|_m = 1.0 \tag{2.25(4)}$$

To formulate the denominator, the partial differentiation has to be repeated in terms of the other controlled variable, in this example, x. Let stream 2 contain none of the measured ingredient, and stream 1 have a concentration of $x_{\max}$. Then the concentration in the blend will be

$$x = \frac{x_{\max} F_1}{F} \tag{2.25(5)}$$

Now the partial may be evaluated in terms of x:

$$\left.\frac{\partial F}{\partial F_1}\right|_x = \frac{x_{\max}}{x} \tag{2.25(6)}$$

Dividing the first partial by the second gives the relative gain:

$$\lambda_{F1} = x/x_{\max} = x' \tag{2.25(7)}$$

The RGA then looks like this:

$$\begin{array}{ccc} & F_1 & F_2 \\ \Lambda = F & x' & 1-x' \\ x & 1-x' & x' \end{array} \tag{2.25(8)}$$

This method gives the advantage of a solution in terms of the variables themselves, in this case x. The solution then applies generally to a class of processes, e.g., two-component blending, where it can be used again.

Two Loops Open

Process models are often formulated as a set of linear equations with gains K:

$$\begin{aligned} c_1 &= K_{11}m_1 + K_{12}m_2 \\ c_2 &= K_{21}m_1 + K_{22}m_2 \end{aligned} \tag{2.25(9)}$$

If the process is limited to two pairs of variables, a simple formula applies:

$$\lambda_{11} = \frac{1}{1 - K_{21}K_{12}/K_{11}K_{22}} \tag{2.25(10)}$$

This formula gives some important insights. If there are an odd number of negative signs among the gains, λ_{11} will fall between zero and 1.0. Otherwise, λ_{11} will fall outside of that range and be negative if the product of the off-diagonal gains exceeds that of the diagonal gains — dominant positive feedback.

Two Loops Closed

Some process models solve for values of the manipulated variables that will produce designated values of the controlled variables:

$$\begin{aligned} m_1 &= H_{11}c_1 + H_{12}c_2 \\ m_2 &= H_{21}c_1 + H_{22}c_2 \end{aligned} \tag{2.25(11)}$$

Again, for two pairs of variables, we have the same formula:

$$\lambda_{11} = \frac{1}{1 - H_{21}H_{12}/H_{11}H_{22}} \tag{2.25(12)}$$

The Two-Slope Method

Open-loop testing can provide still another way to calculate relative gains in a two-loop system. From a steady state, m_1 can be stepped with m_2 constant, and the resulting changes in the two controlled variables divided. The ratio of their changes $(\Delta c_1/\Delta c_2)_{m2}$ is also the ratio K_{11}/K_{21}. Then m_2 is stepped with m_1 constant, with these changes divided, $(\Delta c_1/\Delta c_2)_{m1}$ being the same as the ratio K_{12}/K_{22}. Substitution of those ratios into Equation 2.25(10) gives

$$\lambda_{11} = \frac{1}{1 - (\partial c_1/\partial c_2)_{m1}/(\partial c_1/\partial c_2)_{m2}} \tag{2.25(13)}$$

This formula applies particularly well to distillation composition loops, where the product compositions can be plotted against one another while various manipulated variables are held constant. This is the method used in Section 8.20.

The Matrix Method

Equations 2.25(9) and 2.25(11) are sets of linear simultaneous equations that can be expanded to any number of variables. When the number exceeds two pair, however, it is easier to express the relationships using matrix algebra. Accordingly, Equation 2.25(9) could be expressed as

$$\mathbf{c} = \mathbf{K}\mathbf{m} \qquad \mathbf{2.25(14)}$$

where **c** is a vector of n controlled variables, **m** is a vector of n manipulated variables, and **K** is the $n \times n$ matrix of open-loop gains relating them. Similarly, Equation 2.25(11) could be expressed as

$$\mathbf{m} = \mathbf{H}\mathbf{c} \qquad \mathbf{2.25(15)}$$

Matrix **H** is the inverse of matrix **K** (but its elements are not the inverse of the elements in the **K** matrix):

$$\mathbf{H} = \mathbf{K}^{-1} \qquad \mathbf{2.25(16)}$$

If one matrix is available, the other may be obtained by inverting it using standard matrix inversion procedures. Then each relative gain element λ_{ij} may be found by multiplying the element K_{ij} by the corresponding element H_{ji}:

$$\lambda_{ij} = K_{ij} \cdot H_{ji} \qquad \mathbf{2.25(17)}$$

The subscripts are reversed because H_{ji} relates c_i and m_j, as does K_{ij}.

Where a zero appears in the **K** matrix, one will also appear in the RGA. Other than that observation, the calculations need to be done to determine the RGA. A numerical example follows. Given that

$$\mathbf{K} = \begin{bmatrix} 2.2 & 0.80 & 0.40 \\ 1.2 & 1.6 & 0 \\ 0 & 0.70 & 0.50 \end{bmatrix}$$

Then

$$\mathbf{H} = \begin{bmatrix} 0.495 & -0.074 & -0.396 \\ -0.371 & 0.681 & 0.297 \\ 0.520 & -0.953 & 1.584 \end{bmatrix}$$

and

$$\Lambda = \begin{bmatrix} 1.089 & -0.297 & 0.208 \\ -0.089 & 1.089 & 0 \\ 0 & 0.208 & 0.792 \end{bmatrix}$$

The best pairing appears to be along the diagonal—not surprising, given the two off-diagonal zeros.

Reducing the Array

Arrays larger than 3 × 3 may present misleading information, principally due to dynamic effects. In general, fast loops can upset slow loops, but slow loops cannot upset fast ones. In this light, the pairing of the fast loops in a system can be arbitrarily assigned, reducing the array to include only the slower loops. This is the approach used to evaluate relative gains for distillation in Section 8.20. A typical column has five loops: two compositions, two levels, and one pressure. Arbitrary assignment of the level and pressure loops reduces the RGA to the two composition loops. However, there are several of these 2 × 2 RGAs to be evaluated, one for each assignment of the level and pressure loops. Each of these RGAs needs to be evaluated to determine which has the most favorable relative gains.

DECOUPLED RELATIVE GAINS

Relative gain analysis is not limited to evaluating interactions among single loops. Controller *outputs* can drive mathematical combinations of manipulated variables, such as their sums, differences, or ratios. Controller *inputs* can also be mathematical combinations of controlled variables. Selection of these functions should be logical and meaningful to those who operate the loops.

Partial Decoupling

As an example, return to the blending system of Figure 2.25a. The composition of the blend is obviously determined by the ratio of the two feed flows, F_1/F_2. The output of the composition controller then ought to manipulate that ratio, rather than one of the individual flows. This is done by inserting a multiplier [×] in the output of the composition controller, as shown in Figure 2.25b. The other input to the multiplier is

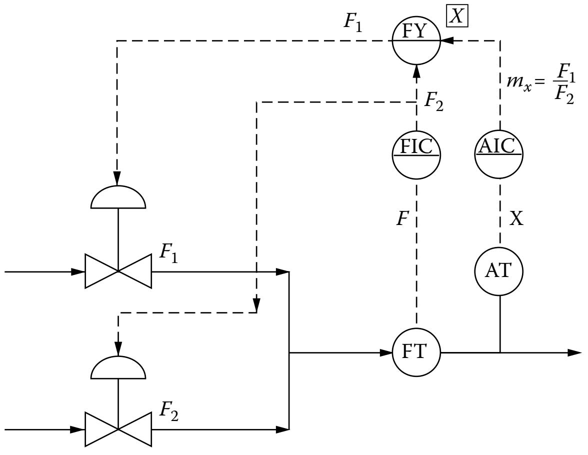


FIG. 2.25b
Partial decoupling is enough to drive relative gains to one and zero and protect the costly variable against upset.

F_2, and the output of the multiplier is the driven value of F_1. Calculating F_1 from F_2 in this way means that the controller output $m_x = F_1/F_2$.

To evaluate the effectiveness of this configuration, it is only necessary to calculate one gain. Beginning with Equation 2.25(5), we substitute for F_1—which is no longer an independent manipulated variable—the product m_xF_2:

$$x = x_{\max}\frac{F_1}{F_1+F_2} = x_{\max}\frac{m_xF_2}{m_xF_2+F_2} = x_{\max}\frac{m_x}{m_x+1} \tag{2.25(18)}$$

As can be seen, the second manipulated variable F_2 drops out of the equation, leaving composition x unaffected by it:

$$\left.\frac{\partial x}{\partial F_2}\right|_{mx} = 0 \tag{2.25(19)}$$

With its numerator at zero, $\lambda_{x2} = 0$. Then the RGA becomes

$$\Lambda = \begin{array}{ccc} & m_x & F_2 \\ F & 0 & 1 \\ x & 1 & 0 \end{array} \tag{2.25(20)}$$

The system is now decoupled.

However, the configuration in Figure 2.25b is only a *partial* decoupler. While composition is no longer upset by changes in the output of the flow controller, total flow *is* changed every time the composition controller moves its output. This is not usually considered a problem because the flow loop is so much faster than the composition loop. The output of the composition controller will tend to move so slowly that any upset in total flow will be small and easily corrected by the faster flow controller.

Partial decouplers are usually preferred to full decouplers in that they accomplish the principal objectives of eliminating the third feedback loop—which created the stability problem—and of protecting the slower and more costly variable from upsets. A full decoupler, which would only provide the (unnecessary) protection of the faster variable, also adds a potentially dangerous feedback loop[11] and complicates operation of the system.

It will be observed that a partial decoupler is implemented in the same way as a feedforward loop. There is no real distinction between them. Feedforward uses an independent load variable and the output of a feedback controller to calculate its manipulated variable, whereas a decoupler uses another manipulated variable in place of the load. Both protect the controlled variable from disturbances, but the decoupler also breaks a hidden feedback loop.

References

1. Bristol, E. H., "On a New Measure of Interaction for Multivariable Process Control," *IEEE Transactions on Automatic Control,* January 1966, pp. 133–134.
2. Shinskey, F. G., *Process Control Systems,* New York: McGraw-Hill, 1967, pp. 189–198.
3. Shinskey, F. G., *Process Control Systems*, 4th ed., New York: McGraw-Hill, 1996, pp. 264–278.
4. Wolverton, P., "How to Use Relative Gain Analysis in Systems with Integrating Variables," *InTech,* September 1980, pp. 63–65.
5. Arkun, Y., and Downs, J., "A General Method to Calculate Input/Output Gains and the Relative Gain Array for Integrating Processes," *Computers Chem. Engng.,* 10, 1990, pp. 1101–1110.
6. Chang, J.-W., and Yu, C.-C., "The Relative Gain for Non-Square Multivariable Systems," *Chemical Engineering Science,* 5, 1990, pp. 1309–1323.
7. McAvoy, T. J., *Interaction Analysis,* Research Triangle Park, NC: ISA, 1983.
8. Shinskey, F. G., *Distillation Control,* 2nd ed., New York: McGraw-Hill, 1984.
9. Skogestad, S., Lundstrom, P., and Jacobsen, E. W., "Selecting the Best Distillation Control Configuration," AIChE, May 1990, pp. 753–764.
10. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, pp. 347–350.
11. Shinskey, *ibid.,* pp. 267–270.

Bibliography

McAvoy, T. J., *Interaction Analysis,* Research Triangle Park, NC: ISA, 1983.

Perron, S., "The New Generation Language for Controls Design," *Control Engineering,* February 1992.

2.26 Robustness: A Guide for Sensitivity and Stability

J. P. GERRY (2005)

In process control there is always a tradeoff between good, stable performance and speed of response or sensitivity to process or set point changes. This sensitivity of the control system can be called the robustness of the loop. Understanding the tradeoff between control loop robustness and stable (optimal) controller performance is important to achieving good control system design.

The system robustness is quantifiable and can be shown graphically, which makes it easier to examine the tradeoff between sensitivity and stability of the loop.

ROBUSTNESS PLOTS

Figure 2.26a shows process dead time on the vertical axis plotted against process gain on the horizontal axis. If the control loop or the controlled process is located on this plot, that can be called the robustness plot of the loop or process.

On a plot of robustness, a line separates the region in which the control loop will be stable and the area where it will not. As shown in Figure 2.26b, the two regions are separated by a line, which is the curve of marginal stability. A control loop is marginally stable if it responds to an upset by the controlled variable drawing a sine wave and the loop sustaining that oscillation.

An increase in either the gain or the dead time of the control loop will make it less stable; the curve shows the values of various combinations of loop gain and dead time at which the control loop will become marginally stable. As was shown in Figure 2.1x, the gain of the loop is the product of the process and the controller gains, while the dead time of the loop is the sum of the process and controller dead times.

In the lower left region of the plot, the process gains and dead times are the smallest. Hence the lower left area of the plot represents a region where the control loops are likely to be stable. Conversely, in the upper right area of the plot, they are likely to be unstable. Areas above and to the right of the marginal stability curve define the unstable region. Areas below and to the left of the marginal stability curve represent a stable region.

On Figure 2.26b, a cross identifies the control loop in its present state. The location of the cross corresponds to the dead time and gain of the controlled process or of the loop. In order to easily compare the effects of the various control loop configurations that are being evaluated and their corresponding robustness, the scales of the two axes of the plots are adjusted

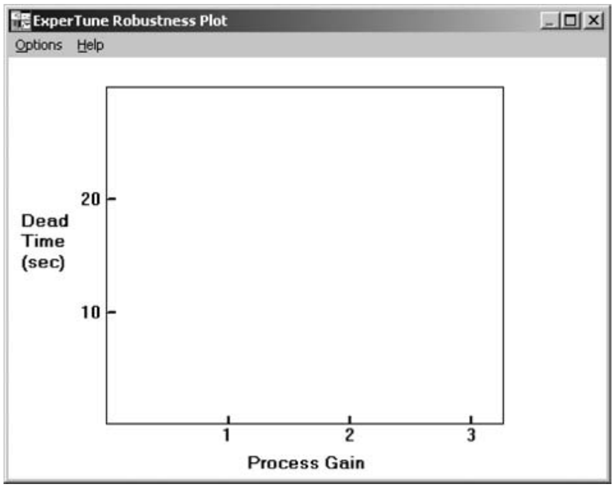


FIG. 2.26a
The process robustness plot locates the particular control loop on a plot of process dead time against process gain.

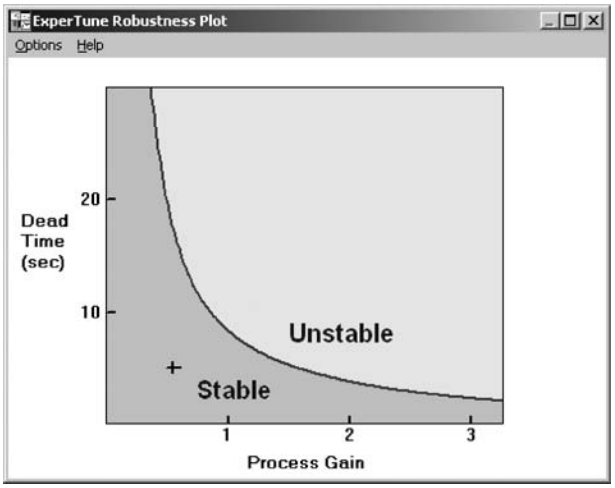


FIG. 2.26b
The marginal stability line separates the stable area from the unstable region on a robustness plot.

so that the current system gain and dead time (see cross in Figure 2.26b) occur at the same location on every plot.

Dynamic Parameters

Dead time is the limiting element of good control performance and robustness, yet almost all control loops contain some dead time. The small dead times and transportation lags of the control loop add up and combine to form the total loop dead time.

Changes in time constants (lags) change the shape of the marginal stability curve. Large lags and integration times affect the low-frequency stability of the loop; hence, the upper left portion of the marginal stability curve will change the most. Changing high-frequency components of the controller, such as the derivative mode settings, will change mostly the shape of the lower right portion of the curve.

Gain is used as the other axis of the robustness plot. System gain is the product of all the gains in the loop, including the process gain and controller gain. Changing any of the gains in the loop will shift the marginal stability curve to the left or right, but the shape of the curve will not change. Increasing the loop gain will move its operating point (cross in Figure 2.26b) closer to the marginal stability curve and closer to instability.

As the dynamics of a loop move down and to the right on the marginal stability curve, the frequency of sustained oscillation will increase.

Design Guidelines

Figure 2.26c shows a recommended region for design stability. The four-sided object in Figure 2.26c has four corners at a factor and divisor of two in both the gain and dead time axis. The top corner of the object represents twice the system dead time with the gain unchanged, and the bottom corner of the object represents half the system dead time. The right corner of the object represents twice the system gain with the dead time unchanged. The left corner represents half the system gain. The four corner points are connected with curving lines that on a log–log plot would be straight.

This recommended safe operating area is used because a factor of two is often considered a reasonable safety margin for operating process control systems.

With a design goal of a stability margin of two in any direction, the noted region represents the area that the marginal stability line should not touch or cross. If the marginal stability line moves inside this region, the margin of stability would drop below two. The purpose of identifying such a safe operating area is not to set a hard rule but only to provide a design aide.

EXAMPLE

In Figure 2.26c, the control loop shown by the cross has a dead time of 5 seconds and an overall system gain of 0.5. If the dead time of this control loop remained constant while the gain of the loop increased, this would shift the cross that represents the loop dynamics horizontally toward the right. If the system gain were to increase by a factor of three (to about 1.5), the cross would reach the curve of marginal stability, and the control loop would become marginally stable.

If the gain of this control loop remained constant while its dead time increased, this would cause the cross, representing the control loop, to move vertically upwards. If the dead time of the loop increased by a factor of three (to about 15 seconds) the cross would reach the curve of marginal stability, and the loop would become marginally stable.

If both the gain and dead time of the loop increased, this would result in moving the cross, which represents the control loop dynamics up and to the right; with enough of an increase the system would become unstable when the marginal stability line was crossed.

Figure 2.26d illustrates the condition when the control loop gain has increased, shifting the curve to the left, reaching marginal stability. When this happens, the control loop itself becomes unstable. This means that the

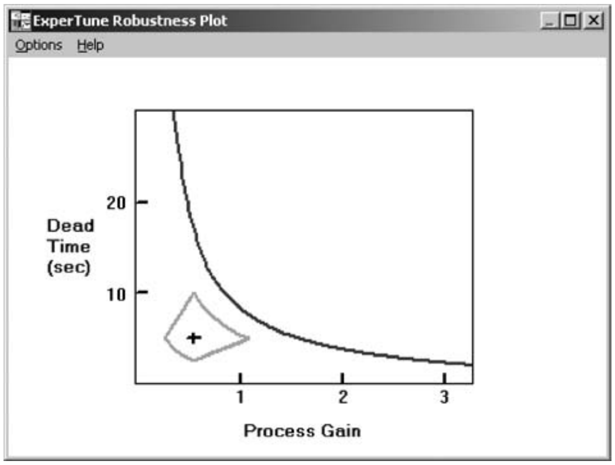


FIG. 2.26c
Robustness plot showing the recommended safe operating area for a control loop that will maintain a stability margin of two.

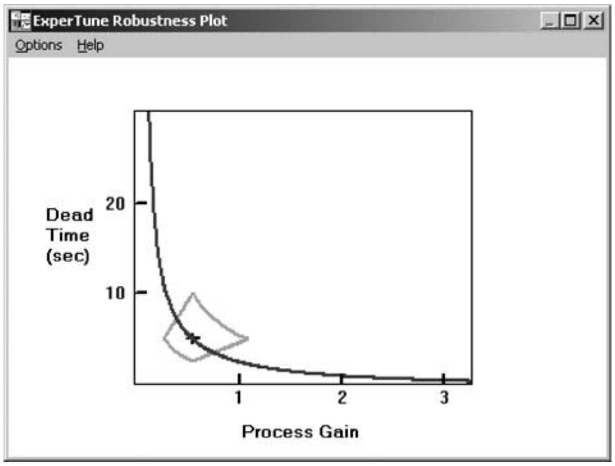


FIG. 2.26d
Robustness plot illustrating a condition where the control loop is in sustained oscillation.

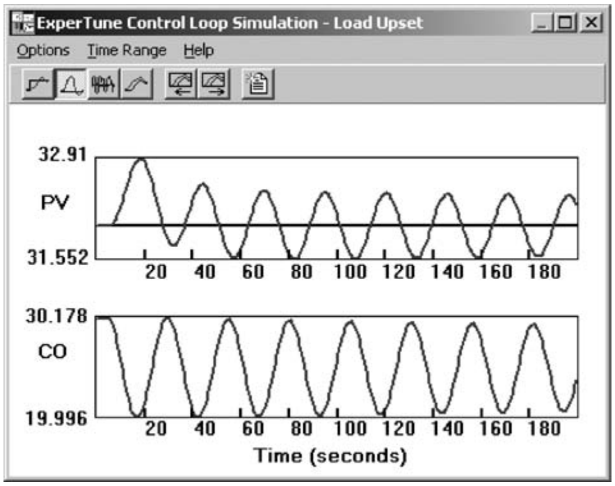


FIG. 2.26e
When the control loop has reached the state of marginal stability, both the controlled (PV) and the manipulated (CO) variables are in sustained and undampened oscillation.

controller output (manipulated variable) starts to cycle and as a consequence, the controlled process variable (PV) will also start to cycle.

Figure 2.26e illustrates this marginal stability condition where both the controlled and the manipulated variables (PV and CO) are in continuous, undampened oscillation. If the control loop dynamics worsened and moved further into the unstable region in Figure 2.26d, the sustained, undampened oscillation would turn into runaway oscillation. Under these conditions, if some safety constraint does not shut the system down, accidents can occur.

CONCLUSION

A balance between loop sensitivity and stability always exists. When tuning a control loop (see the later sections in this chapter on tuning), the goal should always be to reach a reasonable balance between the two goals. Robustness plots can be convenient graphical tools because they can show the safety margin that the control system has before it would become unstable.

Bibliography

Gerry, J. P., “How to Control Processes With Large Dead Times,” *Control Engineering,* March 1998.

Gerry, J. P., “Tune Loops For Load Upsets vs. Setpoint Changes,” *Control,* September 1991.

Shinskey, F. G., *Process Control Systems,* New York: McGraw-Hill, 1996.

2.27 Sampled Data Control Systems

J. HETTHÉSSY, R. BARS (2005)

The main characteristic of digital control is that it utilizes and operates on sampled data. This section discusses the nature and characteristics of SDCS (sampled data control systems), which in this age of the computer, tend to dominate process control. The section also describes the nature and the mathematical derivation of the various digital control algorithms.

Although simple examples are used to illustrate the theoretical background, the mathematical steps involved in the development of these control algorithms can be useful to those readers who are involved in the design of digital algorithms. The section concludes with a discussion of Smith predictors, used to control dead time processes and with such practical issues as the selection of the sampling frequency in digital control.

SAMPLED DATA CONTROL SYSTEMS

The configuration of a sampled data control system is shown in Figure 2.27a. In this section it will be assumed that the controlled processes are continuous and that both the manipulated variables sent to the process and the measurement signals received from the process are also continuous. These continuous (analog) signals are made available for digital processing by A/D converters to provide the data in discrete form as a sequence of numbers. Consequently, the analog world of the process is interfaced to the world of digital computations.

A/D and D/A converters implement this interface. As shown in Figure 2.27a, a real-time clock is governing the operation of the digital computer by controlling the sampling and holding of data.

Symbols Used

In this section, in order to distinguish the analog and discrete versions of the signals involved in sampled data control systems, the following notations will be used:

- $x(t)$: continuous-time signal
- $x[k] = x(kh)$: discrete-time signal, where h denotes the sampling time and $k = 0, 1, 2, \ldots$ assigns the sampling instant

Sampling according to $x[k] = x(kh)$ is also called mathematical sampling. The heart of the control scheme is a digital control algorithm realized in a real-time software environment. The interfaces applied are to accomplish the following functions:

- Sampling the analog process variables to deliver discrete-time information for the control algorithm
- Delivering control signals calculated by the control algorithm for the actuators
- Receiving set-point requests from the man–machine interface (MMI) or via the communication network

In the field of process control, as elsewhere, there is a general trend toward digital implementations. The advantages of sampled data control systems over their analog counterparts can be listed as follows:

- The digital technology applied is more reliable and cheaper.
- Flexibility is superior, considering both the implementation and the variety of the control algorithms.
- Possible modifications and/or extensions are easier to accomplish.
- Accuracy is kept constant over a long period of time.
- It is simple to deliver the set-point value for the controller, to overwrite the controller parameters, as well as to monitor the controller operation.

On the other hand, special care is required in the following areas:

- Between two samples, the control system is left to operate in open loop.
- The sampling rate should be carefully selected to be in harmony with the dynamics of the process and to comply with the capabilities of the real-time environment (performance and the number representation applied).
- The output of the digital controller (also called the manipulated variable) must be interpolated from a digital sequence to a continuous-time function, thus the waveform of the control signal is limited.
- Sampling introduces additional difficulties when the process dynamics are nonlinear or contain substantial dead time.

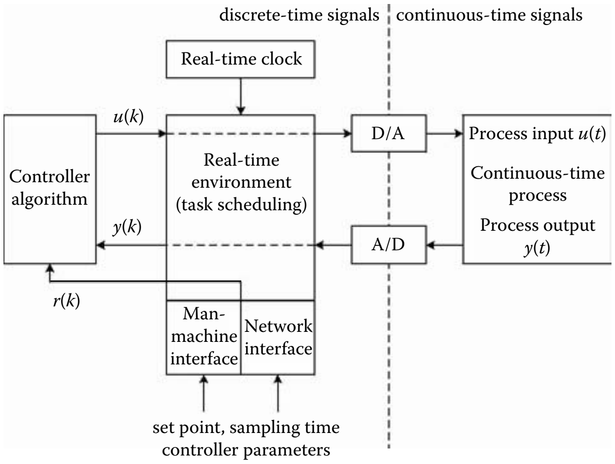


FIG. 2.27a
The main components of sampled data control systems.

Properties of SDCS

Figure 2.27a shows a simplified block diagram of a digital control system. Note that the holding device is an absolutely necessary component in sampled data control systems because the continuous process needs to be controlled by a continuous signal. To generate the analog signals, D/A converters are used. They are updated in each sampling period and this results in a staircase-type output (see Figure 2.27b).

The digital computer provides wide flexibility in the availability of control algorithms. In this section only linear, single-input single-output control systems having one degree of freedom are discussed.

EXAMPLE

In this discussion, the terms and concepts developed in the age of analog control will be used; however, none of the results will automatically be reused. To illustrate, consider the following continuous process:

$$P(s) = \frac{1}{(1+5s)(1+10s)} \qquad \textbf{2.27(1)}$$

which is closed-loop controlled and controlled by a proportional controller $C(s) = K$. According to the classical control theory, any $K > 0$ will stabilize the system because the loop transfer function $L(s) = KP(s)$ exhibits a *structurally stable* system.

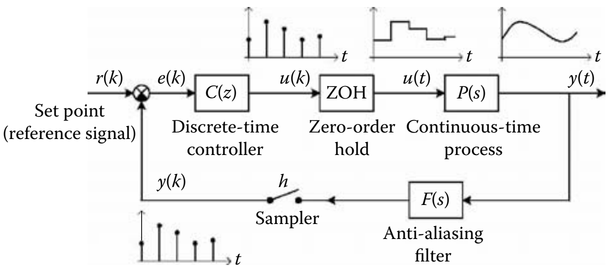


FIG. 2.27b
The components and signals used in sampled data control systems.

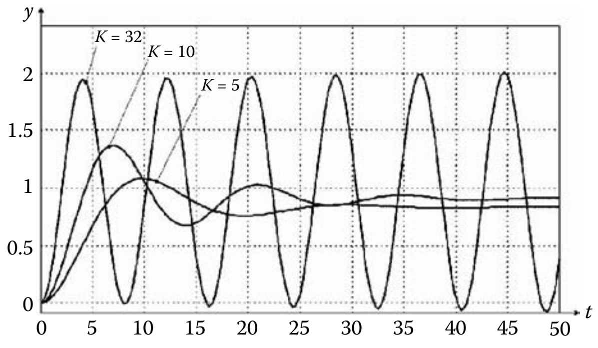


FIG. 2.27c
Step responses of the closed-loop system at various controller gains. The controlled variable (y) becomes unstable when the gain of the plain proportional controller reaches K = 31.2.

Now consider the sampled data version of the control loop using a sampling time of $h = 1$ sec. Figure 2.27c shows the results obtained as a function of the gain settings of the proportional-only controller. Unlike the analog control loop, the digital sampled data version will go unstable when the gain is $K > 31.6$. A quick stability analysis can verify this, but first a conceptual explanation is desired.

Figure 2.27d illustrates the effect of sample and hold (S&H) on detecting a sinusoidal input and generating a staircase-type output. A phase delay exists between the analog sinusoidal signal and the first harmonic component of the output generated by the holding device. The size of this delay depends on the sampling time and the frequency of the sinusoidal signal applied; in this example, it is around the half of the sampling time. The consequence is that the zero-order holding (ZOH) device introduces a time delay; thus, the structural stability is lost.

Mathematical Aspects of SDCS

Sampling In this section, for the sampling of analog signals, *periodic sampling*, where the sampling instants are equally spaced, will be considered. The frequency of the sampling is considered to be correct if the spectrum represented by the samples is identical to the spectrum of the converted analog signal. Shannon's sampling theorem says that if an analog signal contains no frequencies above $\omega_{\max}$, then it can be

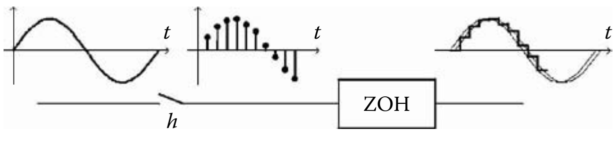


FIG. 2.27d
The effects of sample and hold (S&H) and of zero-order hold (ZOH) in signal conversion.

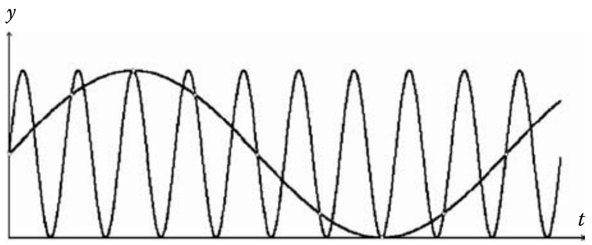


FIG. 2.27e
When sampling a high-frequency sinusoidal analog signal at a slow sampling rate, the analog signal will appear to be of low frequency due to the aliasing (frequency folding) effect.

completely reconstructed from its samples if $\omega_s \geq 2\,\omega_{\max}$ holds for the sampling frequency ω_s.

On the other hand, if the sampling rate is not high enough, a high-frequency analog signal will be interpreted as a low-frequency signal. This phenomenon is called *aliasing* or *frequency folding.* Figure 2.27e, where samples of a high-frequency sinusoidal signal appear to be a low-frequency signal due to aliasing, illustrates this phenomenon. To avoid aliasing, a low-pass filter (also called an antialiasing filter) must be placed between the analog measurement signal and the A/D converter. The purpose of the low-pass filter is to filter out the signal components above the frequency of $\omega_{\max}$.

Zero-Order Hold (ZOH) The function of the ZOH unit is to generate an output at the instant of sampling that is identical to its input and to hold this value constant for the rest of the cycle, until it is time for the next sample to arrive. Figure 2.27f verifies that a parallel combination of an integrator by $1/s$ and a delayed integrator by $e^{-\mathrm{sh}}/s$ satisfies this required functionality. Thus the transfer function of the zero-order hold unit is

$$G_{\mathrm{zoh}}(s) = \frac{1-e^{-\mathrm{sh}}}{s} \qquad \textbf{2.27(2)}$$

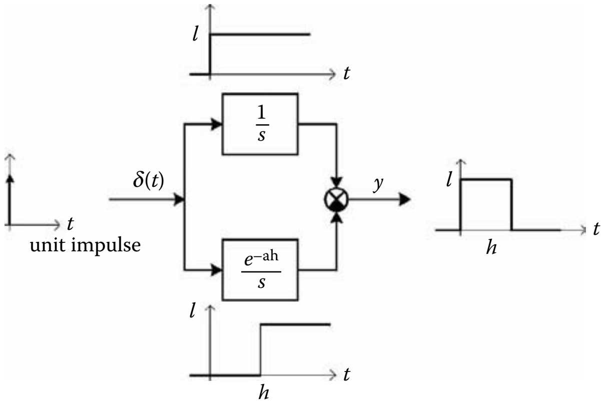


FIG. 2.27f
The transfer function of the zero-order hold (ZOH) unit.

Discrete-Time Models One way to describe discrete-time systems in the time domain is to express the input/output relation by a *difference equation,* e.g.,

$$\begin{aligned} &y[k+2]-1.7236y[k+1]+0.7408y[k] \\ &\quad = 0.0091u[k+1]+0.0082u[k] \end{aligned}$$

To derive a more compact form, one may introduce the forward shift operator q as follows:

$$qx[k] = x[k+1]$$

then

$$q^2y[k]-1.7236qy[k]+0.7408y[k]=0.0091qu[k]+0.0082u[k]$$

is obtained.

The most widely used tool to handle discrete-time systems is the Z-transformation defined by

$$Z\{x[k]\} = X(z) = \sum_{k=0}^{\infty} z^{-k}x[k] \qquad \textbf{2.27(3)}$$

where $z = e^{\mathrm{sh}}$ is the operator of the Z-transformation (s is the Laplace operator). Taking the Z-transform of the input and output samples, respectively, *the discrete or pulse transfer function*:

$$G(z) = \frac{Y(z)}{U(z)} \qquad \textbf{2.27(4)}$$

uniquely determines the input/output behavior at the sampling instants, for zero initial conditions.

Discrete-Time Process Models The components of the closed loop shown in Figures 2.27b and 2.27g included both analog and digital (both continuous-time and discrete-time) signals. To set up a closed-loop model with purely discrete transfer functions, the continuous elements between the hold and sample units in the loop should be converted from a continuous to a discrete model. Note that the process and the ZOH are combined and taken into account in the following transformation (see Figure 2.27g):

$$P(z) = Z\{L^{-1}[G_{\mathrm{zoh}}(s)P(s)]_{t=\mathrm{kh}}\} = (1-z^{-1})Z\left\{\frac{P(s)}{s}\right\} \qquad \textbf{2.27(5)}$$

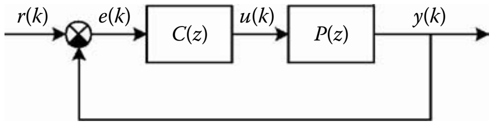


FIG. 2.27g
Components of the discrete-time closed-loop model.

As an example, consider a continuous-time process of

$$P(s) = \frac{1}{(1+5s)(1+10s)}$$

sampled by $h = 1$ sec. Straightforward substitution to Equation 2.27(5) (for those using MATLAB, the c2dm command) gives

$$P(z) = \frac{0.0091z + 0.0082}{z^2 - 1.7236z + 0.7408} = \frac{0.0091(z + 0.9048)}{(z - 0.9048)(z - 0.8187)}.$$

This is exactly the example discussed earlier to illustrate a difference equation form. Note that the numbers in this example indicate a numerical difficulty: the magnitude of the zeros and poles of the discrete transfer function are all close to unity. This is especially true if h is small, selecting, e.g., $h = 0.1$ sec, the discrete-time model turns out to be

$$P(z)\big|_{h=0.1} = \frac{(9.901z + 9.802)\cdot 10^{-5}}{z^2 - 1.9702z + 0.9704} = \frac{9.9006(z + 0.99)\cdot 10^{-5}}{(z - 0.99)(z - 0.9802)}.$$

With $P(z)$ having been introduced according to Equation 2.27(5), now the stability condition referred to earlier can be verified. The closed-loop characteristic equation is

$$1 + K \cdot P(z) = 0$$

or

$$z^2 + (0.0091K - 1.7236) + 0.0082K + 0.7408 = 0.$$

According to $z = e^{sh}$, the left side of the s-plane (i.e., the stability region for the continuous-time systems) is mapped into the unit circle in the z-plane. Consequently, the unit disc represents the stability region for discrete-time systems. For $K = 0$ the poles are located at $p_1 = 0.9054$ and $p_2 = 0.8182$ in the z-plane. For $K > 0$ the poles will move toward the unit circle, and $K = 31.6$ results in closed poles $p_{1,2} = 0.718 \pm j0.696$ just sitting on the unit circle as $|p_{1,2}| = 1$. For $K > 31.6$ the closed-loop poles will be outside of the unit disc leading to an unstable system.

Historically, the *delta transformation* was first introduced as a tool to handle discrete-time systems by

$$D\{x[k]\} = X_\delta(\gamma) = \sum_{k=0}^{\infty}(1+\gamma h)^{-k} x[k]h \qquad \textbf{2.27(6)}$$

where

$$\gamma = \frac{e^{sh} - 1}{h},$$

which involves

$$z = 1 + \gamma h.$$

A key property of delta transforms is that they converge to the associated Laplace transforms as $h \to 0$:

$$\lim_{h\to 0} X_\delta(\gamma) = X(s)_{s=\gamma}.$$

This is why design techniques based on the delta transforms are very close to continuous-time design concepts, especially in case of fast sampling. Also, the numerical difficulties outlined for the Z-transforms are not experienced here. One way to see the numerical difficulties associated with the Z-transforms is to consider the stability region (i.e., the unit disc), which is much smaller than the left side in the s-plane.

Therefore poles and zeros need a precise number representation. Using Z-transforms, the high sensitivity with respect to the changes in the parameters becomes even more serious if a fast sampling rate has been selected. The delta transformation overcomes the numerical difficulties outlined above; here, the stability region expands as the sampling rate becomes faster.

Design Aspects

Recalling the introductory block diagram of sampled data control (Figure 2.27b), the closed-loop system is a hybrid system in the sense that it contains both continuous and discrete components. Consequently, a digital controller can be designed in several ways:

- *Hybrid pole-placement.* This technique is based on the classical pole placement of analog technology. The discrete-time controller is developed in a step-by-step way, where the replacement of a selected pole of the analog process is accomplished by expanding the controller with a discrete-time PI or PD term in each step.
- *Approximate continuous design.* Here too the classical design methods learned from analog practice are utilized such that a continuous controller is designed and then converted to a discrete form.
- *Direct discrete-time design.* In this case a discrete-time model of the process is first developed, and then a discrete-time controller is designed for that model.

Within each of the above classes several approaches are used. It should be emphasized that the actual design method may combine a number of underlying methodologies, such as loop shaping, pole placement, internal model control, and state-variable feedback. In this section, four methods will be described:

- Hybrid pole placement
- Approximate continuous-time design using w or delta transforms
- Dead-beat controller design
- Smith predictor

Some numerical examples will also be described in order to illustrate the characteristic steps involved.

Hybrid Pole Placement In the low-frequency region, sampled data systems can be approximated with their continuous counterparts modified by the addition of some extra dead time. While restricted to manipulations in the low frequency region, this is a convenient and easy method to directly design a discrete-time controller. The design is based on the classical *pole cancellation technique* and evolves by sequentially applying discrete-time PI and PD elements.

In this design, the unfavorable poles of the plant are cancelled by the zeros of the controller. Integrators are introduced, depending on the steady-state accuracy requirements. Finally, a controller gain is chosen to achieve the desired phase margin. Assuming that T_1 is a time constant of the continuous process (typically, T_1 is the largest time constant of the process), which is to be cancelled and to be replaced by a PI element:

$$C_1(z) = \frac{z - e^{-h/T_1}}{z-1} \qquad \mathbf{2.27(7)}$$

Similarly, to cancel a continuous time process pole associated with a time constant T_2, a PD controller element by

$$C_2(z) = \frac{z - e^{-h/T_2}}{z} \qquad \mathbf{2.27(8)}$$

is added (more precisely serially connected) to the PI element.

As more elements are added in a similar way, a controller gain A is determined such that the phase margin exhibited by the frequency function of the loop transfer function $L(z) = C(z)P(z)$ is at a prescribed value (typically ~60°). Note that when using MATLAB for the design, a simple *dbode* MATLAB command delivers the required gain value. The complete discrete-time controller then becomes

$$C(z) = AC_1(z)C_2(z) = A\frac{(z - e^{-h/T_1})(z - e^{-h/T_2})}{(z-1)z} \qquad \mathbf{2.27(9)}$$

EXAMPLE

A continuous-time process can be described by the equation:

$$P(s) = \frac{e^{-s}}{(1+10s)(1+5s)} \qquad \mathbf{2.27(10)}$$

The goal is to design a discrete-time controller by applying a sampling time of $h = 1$ sec to achieve a phase margin $\cong 60°$ and to ensure $j = 1$ (type number) in order to be able to follow a unit step change in the set point with no steady-state error. The discrete-time process model is

$$P(z) = z^{-1}(1-z^{-1})Z\left\{\frac{P(s)}{s}\right\} = 0.0091\frac{(z+0.9048)}{\left(z - e^{-\frac{h}{10}}\right)\left(z - e^{-\frac{h}{5}}\right)z}$$

$$= 0.0091\frac{(z+0.9048)}{(z-0.9048)(z-0.8187)z}$$

Designing the discrete-time controller to cancel the continuous pole at $p_1 = -0.1$ requires that

$$C_{\mathrm{PI}}(z) = \frac{z-0.9048}{z-1},$$

while the pole at $p_2 = -0.2$ is handled by

$$C_{\mathrm{PD}}(z) = \frac{z-0.8187}{z}.$$

Then the controller will have the following form:

$$C(z) = A\frac{z-0.9048}{z-1}\frac{z-0.8187}{z}.$$

Finally, to ensure a phase margin of 60°, assuming a unit step in the set point, A should be set as 15.13. The process input/output plots are shown in Figure 2.27h.

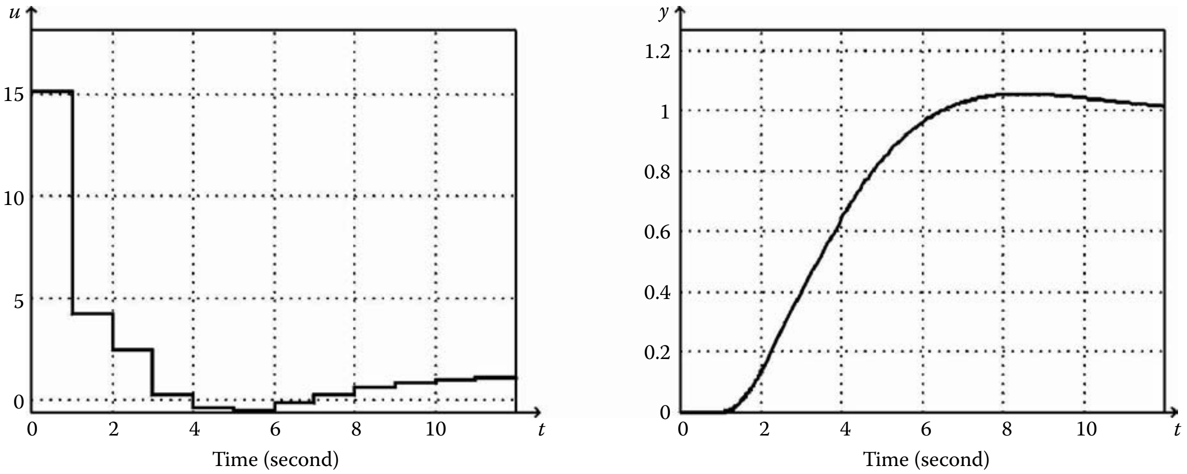


FIG. 2.27h
Records of the manipulated variable (u) and controlled variable (y) when the control algorithm was obtained by the hybrid pole placement design.

Using w or Delta Transforms Another way to utilize the experience with analog design is to combine a continuous-time design technique with a discrete-to-continuous conversion technique.

Design with w-Transforms The concept behind the *w*-transformation (also called Tustin or bilinear transformation) is a numerical integration method performing trapezoidal approximation for each sample period. The approximation leads to the following formula to convert the discrete system to a closely equivalent continuous one and vice versa:

$$w = \frac{2}{h} \cdot \frac{1 - z^{-1}}{1 + z^{-1}}$$

$$z = \frac{1 + \frac{wh}{2}}{1 - \frac{wh}{2}} \qquad \mathbf{2.27(11)}$$

Here *w* stands for a complex frequency operator. The notation $P(w)$ is used for a transfer function of a continuous-time system, indicating that this transfer function has been transformed from a discrete-time system by *w*-transformation. The steps of the design can be summarized as follows:

- Find the discrete-time transfer function of the process and the zero-order hold by using

$$P(z) = (1 - z^{-1}) Z \left\{ \frac{P(s)}{s} \right\} \qquad \mathbf{2.27(12)}$$

- Find the *w*-transform of $P(z)$ by

$$P(w) = P(z)\Big|_{z = \frac{1 + \frac{wh}{2}}{1 - \frac{wh}{2}}}.$$

- Design a continuous-time controller $C(w)$ for the process given by $P(w)$. The design may be based on the frequency function of $P(w)$, which is

$$P(j\nu) = P(w)\Big|_{w = j\nu}.$$

- Find the discrete-time controller by

$$C(z) = C(w)\Big|_{w = \frac{2}{h} \cdot \frac{1 - z^{-1}}{1 + z^{-1}}}.$$

- To complete the design, realize the discrete-time controller and check the closed-loop system properties.

Checking will reflect the fact that the frequency axis used while completing the design has been distorted according to the relation between frequency ν and the actual frequency ω:

$$\nu = \frac{2}{h} \cdot \tan\left(\frac{\omega h}{2} \right) \qquad \mathbf{2.27(13)}$$

The frequency distortion is small if the selected sampling time is sufficiently small. If the sampling rate cannot be chosen to be sufficiently fast, the design steps should be repeated using a scaled version of the *w*-transformation, which avoids the frequency distortion at the expected cutoff frequency (ω_c) of the loop transfer function:

$$w' = \frac{\omega_c}{\tan(\omega_c h/2)} \cdot \frac{1 - z^{-1}}{1 + z^{-1}} \qquad \mathbf{2.27(14)}$$

The transformation according to the above relation is called *w*-transformation with *prewarping*. Another problem with the *w*-transformation–based design is that $P(w)$ is a nonminimum phase transfer function requiring special care to design $C(w)$.

Design with Delta Transforms The direct application of the delta transformation may lead to results similar to those obtained by the sophisticated *w*-transform method. The concept is that a continuous-time controller $C(s)$ is to be designed for the continuous-time process $P(s)$. Then derive the delta form of the controller by substituting

$$\gamma = \frac{e^{sh} - 1}{h}$$

for *s* in $C(s)$: $C(\gamma) = C(s)\big|_{s=\gamma}$.

The key point then is to apply (γ^{-1}) as a building block to implement the control algorithm without converting back to the *Z*-form. The delta building blocks can be manipulated in much the same way as integrators are handled in the implementation of continuous-time models.

The various approximately continuous design techniques, in general, lead to rather similar closed-loop control properties. In deciding which to use, careful analysis of expected modeling error, parametric sensitivity, CPU performance, etc., is recommended.

Dead-Beat Controller Algorithm

In sampled data control systems, the software provides the freedom to implement sophisticated control algorithms. Some discrete-time algorithms can achieve performance that is not achievable by analog control. One class of these algorithms ensures that after an upset, the controlled variable (the continuous output signal from the process) will accurately settle within a finite settling time.

It should be noted that this zero error is guaranteed only at the sampling instants; in the intersampling period (i.e., in continuous-time), error should still be checked. In the literature these algorithms are referred to as dead-beat control algorithms.

To understand the essence of this method — for the sake of simplicity — consider a stable control loop that experiences a step change in set point. A design procedure will be derived in three steps. In the first step a control algorithm will be designed to achieve the fastest possible transient response, i.e., the output signal will be settled in the minimum number of sampling steps. The controller output signal can be very high, and oscillations (also called ripples) can occur between the sampling points.

In the second step the design will be modified to avoid the unwanted intersampling oscillations. It is known, however, that cancellation of discrete-time process zeros outside of a well-defined area within the unit circle is the reason for the oscillations. Separating and not cancelling those zeros, a modified version of the fundamental control algorithm will be derived. The modified algorithm will moderate the control effort, but it will increase the settling time.

The design will be executed in the z operator domain. A remarkable feature of the design method is that the undesired time-domain properties (oscillations, large excursions of the control signal) can be avoided.

The first step is to transform the hybrid control problem to a pure discrete-time problem. To do so a sampling time h should be selected, then the pulse transfer function $P(z)$ of the plant combined with the D/A converter should be derived. Once the discrete-time controller algorithm $C(z)$ has been determined, the closed-loop system performance is to be checked.

To illustrate the design procedure, consider the following continuous process:

$$P(s) = \frac{e^{-s}}{(1+10s)(1+5s)} \tag{2.27(15)}$$

According to the nature of the process, the earliest time instant to see $y[k] = 1$ (i.e., zero error) is $k = 2$. This requirement can directly be expressed by the overall pulse transfer function:

$$\frac{C(z)P(z)}{1+C(z)P(z)} = z^{-2} \tag{2.27(16)}$$

Hence, the pulse transfer function of the control algorithm is expressed as

$$C(z) = \frac{1}{P(z)(z^2-1)} \tag{2.27(17)}$$

Handling now $P(z)$ as a ratio of two polynomials

$$P(z) = \frac{B(z)}{A(z)} = \frac{0.0091(z+0.9048)}{(z-0.9048)(z-0.8187)z}$$

and the control algorithm turns out to be

$$C(z) = \frac{A(z)}{B(z)(z^2-1)} = \frac{109.9(z^3-1.7236z^2+0.7408z)}{z^3+0.9048z^2-z-0.9048}$$

Figure 2.27i shows the result, which is clearly unacceptable. The oscillations are introduced by the control algorithm itself since its poles are responsible for the ripples between the samples. If unstable or lightly damped zeros are not cancelled by associated poles, no oscillations will occur in the continuous output signal. The area of the well-damped process zeros can be determined analytically.

Let us analyze the contour of a conjugate complex pair with a given damping factor in the z domain. Recall that in the s domain the constant ζ lines are straight lines by $s = \sigma + j\omega$. For a given σ value

$$\omega = \frac{\sigma}{\zeta}\sqrt{1-\varsigma^2}$$

along a constant ζ. Then $z = e^{sh}$ maps this line to the z-plane.

As an example, Figure 2.27j shows the area where the damping is constant at $\zeta = 0.4$. If one separates the zeros of the process pulse transfer function according to

$$B(z) = B_1(z)B_2(z) = 0.01733(0.525z + 0.475)$$

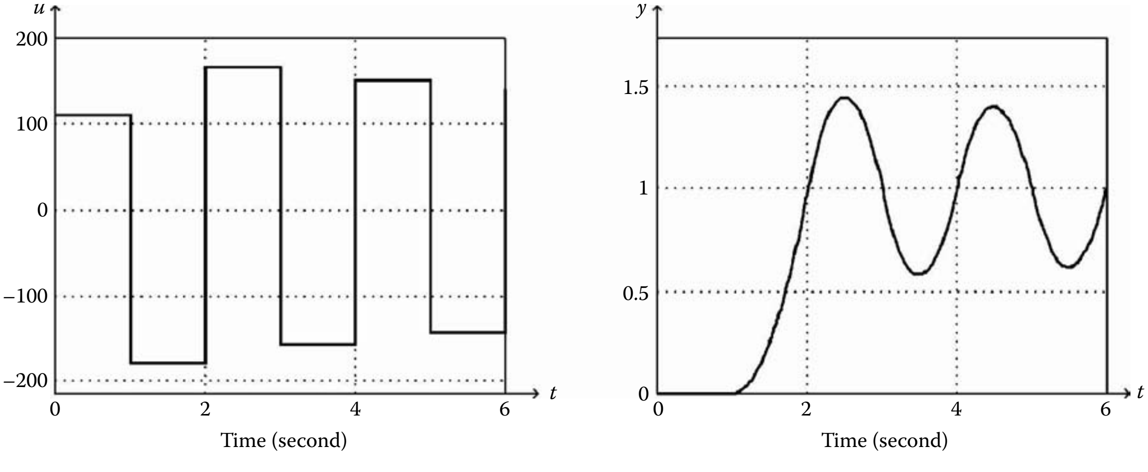


FIG. 2.27i
Records of the manipulated variable (u) and controlled variable (y) when the control algorithm was obtained by a dead-beat algorithm designed with minimum settling time.

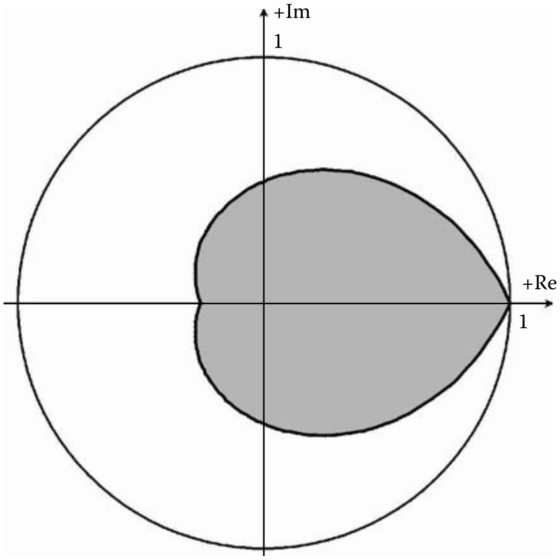


FIG. 2.27j
Region of $\zeta \geq 0.4$ in the z-plane.

where $B_1(z) = 0.01733$ contains the well-damped zeros (no such zero exists here) and $B_2(z) = 0.525z + 0.475$ contains the lightly damped zeros (one such zero exists at $z = -0.9048$). Observe that $B_2(z)$ is normalized by $B_2(1) = 1$. The overall transfer function is planned to be

$$\frac{C(z)P(z)}{1+C(z)P(z)} = B_2(z)z^{-2-\deg(B_2)} = B_2(z)z^{-3} \qquad \mathbf{2.27(18)}$$

where $\deg(B_2)$ is the degree of polynomial $B_2(z)$. Expressing $C(z)$, the controller is

$$C(z) = \frac{A(z)}{B_1(z)[z^3 - B_2(z)]} = \frac{57.7(z^3 - 1.7236z^2 + 0.7408z)}{z^3 - 0.525z - 0.475}.$$

The input/output plots are shown in Figure 2.27k. No oscillations are present, and the overexcitation in the manipulated and controlled variables is also reduced. The control system becomes slower, and the manipulated variable (controller output signal) needs three steps to reach its steady-state value.

On the other hand, the maximum value of the manipulated variable (u, the control signal at 57.7 maximum value) is still too high, and the average control valve actuator would be unable to respond to it. For these reasons, it is necessary to modify the design in order to reduce the step size in the controller output signal (by a factor of five), while keeping the settling time reasonably short (corresponding to five samples).

The following design polynomial can be added to the original algorithm:

$$T(z) = 0.2z^2 + 0.3z + 0.5$$

It increases the finite settling to five samples, while the initial size of the step in the manipulated variable (the controller output signal) is decreased by a factor of five.

The closed-loop control equation with the design polynomial $T(z)$ is

$$\frac{C(z)P(z)}{1+C(z)P(z)} = T(z)B_2(z)z^{-2-\deg(B_2)-\deg(T)} = T(z)B_2(z)z^{-5}$$

2.27(19)

and therefore, the controller algorithm becomes

$$C(z) = \frac{A(z)T(z)}{B_1(z)[z^k - B_2(z)T(z)]} = \frac{11.54(z^5 - 0.2235z^4 + 0.6555z^3 - 3.198z^2 + 1.852z)}{z^5 - 0.105z^3 - 0.2525z^2 - 0.405z - 0.2375}.$$

Observe that $T(1) = 1$, so the design polynomial does not affect the zero steady-state error. The input/output plots for this control algorithm are shown in Figure 2.27l.

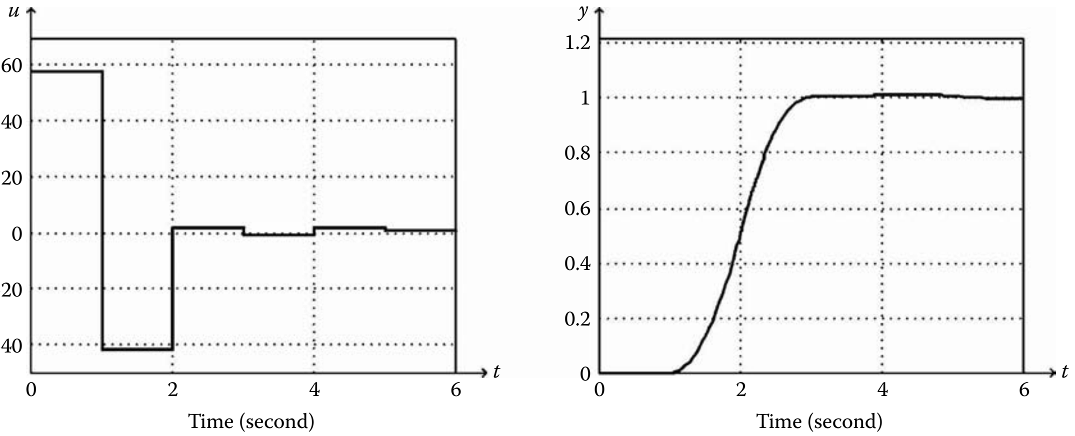


FIG. 2.27k
The proces input–output records obtained when the dead-beat control was revised by not canceling purely damped process zeros.

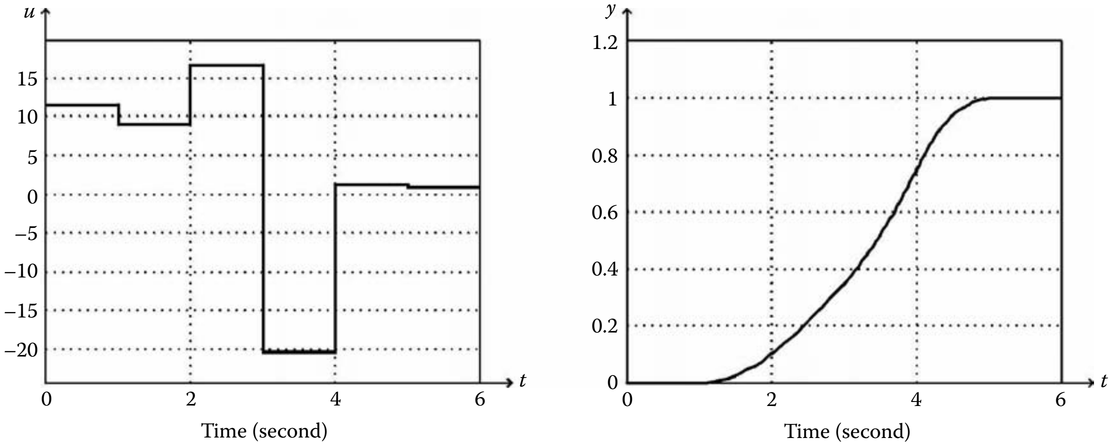


FIG. 2.27l
The process input–output records obtained when the dead-beat control was revised by adding a design polynomial of $T(z) = 0.2z^2 + 0.3z + 0.5$.

Smith Predictor Design

In the 1950s, Smith suggested a control algorithm that can be used to control processes that have significant dead times. Using discrete-time design, the pulse transfer function of the process has the following form:

$$P(z) = z^{-d}P'(z) \qquad \textbf{2.27(20)}$$

where $P'(z)$ is the part of the process with no dead time, and d corresponds to the dead time as an integer multiple of the sampling time. The main idea is that the operating feedback control loop is converted into an equivalent fictitious control loop, in which the dead time is outside the feedback loop (Figure 2.27m). In this case the control algorithm can be designed for the process without dead time.

This control strategy ensures a much faster response than does the PID controller, which is tuned to control a process with dead time. Applying simple block diagram algebra, the equivalence of the transfer functions of the two closed-loop systems leads to

$$C(z) = \frac{C'(z)}{1 + (1 - z^{-d})C'(z)P'(z)} \qquad \textbf{2.27(21)}$$

The first step in developing the control algorithm is to design $C'(z)$ for a process with no dead time and then calculate the discrete-time controller by the Equation 2.27(21) relation. Theoretically, the Smith predictor can be used both for continuous and sampled data control systems. However, as the dead time appears directly in the controller algorithm, the discrete-time applications are far more preferred because the dead time term is easier to generate.

CONCLUSIONS

Considering the practical applications of sampled data control systems, the hardware issues include selecting the right resolution for the D/A and A/D converters. The resolutions of the D/A and A/D converters should be in the range of 12 to 16 and 8 to 12 bits, respectively.

As to the controller algorithm to be used, 1) the errors from converter quantization, 2) round-off errors in the arithmetic operations, and 3) the errors due to quantization of the controller parameters will have an impact on the control algorithm selection.

Bumpless Transfer and Controller Windup

Bumpless transfer is essential so that when the control loop is switched from manual to automatic or back to manual, the controlled variable is not upset. This goal is served by an implementation algorithm in the form of setting appropriate initial conditions for the iterative program schemes based on the measurement signal (control input).

Reset or integral windup occurs, for example, if an error exists while the controller is in manual and the PID algorithm is active. As the integral mode keeps integrating the area

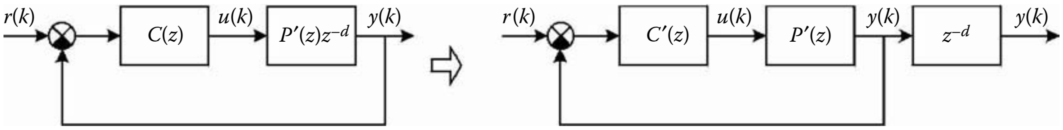


FIG. 2.27m
The concept of the Smith predictor.

under the error curve, when the loop is switched back to automatic, the controller output signal (manipulated variable) can be saturated, meaning that it has reached its maximum limit (constraint).

One should prevent the controller from such windup, because it can cause overshoots followed by cycling or other forms of instability. Integrators are components that can wind up, and *antireset windup* refers to the stopping of integration once the control signal has been saturated.

For digital control applications a number of antireset windup algorithms have been developed. A common characteristic of each is the use of a feedback signal from the manipulated variable (the controller output signal), to keep that signal from reaching saturation.

Selection of the Sampling Rate

The selection of the sampling time depends on the application. The selection will set conditions for the real-time computing power requirements. If the sampling rate is too slow, fast-occurring process disturbances might not be detected and can cause upsets. Too-fast sampling, on the other hand, can increase the computing demand without substantially improving control quality.

Rules of thumb suggest that the sampling frequency should be set as a function of the dead time and should correspond to 1/10 of it. Other criteria suggest to have four to ten samples in each transient response period or to ensure a sampling rate at least five times higher than the bandwidth to be achieved by the closed-loop control.

Bibliography

Åström, K. J., and Wittenmark, B., *Computer-Controlled Systems: Theory and Design,* 3rd ed., Englewood Cliffs, NJ: Prentice Hall, 1997.

Clarke, D. W., "PID Algorithms and Their Computer Implementation," O.U.E.L. Report No. 1482/83, University of Oxford, Department of Engineering Science, 1983.

Franklin, G. F., Powell, J. D., and Workman, M. L., *Digital Control of Dynamic Systems,* 3rd ed., Englewood Cliffs, NJ: Prentice Hall, 1998.

Goodwin, G. C., Graebe, S. F., and Salgado, M. E., *Control System Design,* Englewood Cliffs, NJ: Prentice Hall, 2001.

Kuo, B. C., *Digital Control Systems,* New York: Holt, Rinehart and Winston, 1980.

Levine, W. S. (Ed.), *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.

Middleton, R. H., and Goodwin, G. C., "Improved Finite Word Length Characteristics in Digital Control Using Delta Operators," *IEEE Transactions on Automatic Control*, Vol. AC-31, No. 11, November 1986.

Morari, M., "Some Control Problems in the Process Industries," in *Essays on Control: Perspectives in the Theory and its Applications,* Trentleman, H. L., and Willems, J. C., Eds., Basel: Birkhäuser, 1993.

Ogata, K., *Discrete-Time Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1987.

Smith, O. J., "Closer Control of Loops with Dead Time," *Chemical Engineering Progress,* Vol. 53, pp. 217–219, 1957.

Tuschák, R., "Automatic Control Systems" (lecture notes in Hungarian), Müegyetemi Kiadó (Technical University of Budapest Press), 1994.

2.28 Selective, Override, and Limit Controls

F. G. SHINSKEY (1970) **J. P. SHUNTA** (1985, 1995) **J. E. JAMISON, A. ROHR** (2005)

INTRODUCTION

In this section some examples are given of control configurations where more than one controller can manipulate the same process and where the selection of the active controller is dictated by some prearranged logic.

In some control systems, more variables need to be controlled than there are variables that can be manipulated. When this is the case, logic must be provided to decide which controllers should have access to the manipulated variables and which should be temporarily blocked. Switching controller outputs can be easily and smoothly accomplished by a variety of hardware and software signal selectors.

In selective control applications, the signal selectors choose the lowest, highest, or median signal from among two or more signals. Such selectors are available both as analog hardware or as digital software in DCS control packages.

In most applications, selective control is a form of multivariable control where the selectors facilitate the online modification of control strategies as a function of changing operating conditions. The selectors allow the control strategies to be changed smoothly and without disturbing the process. Selective control applications include:

- Protecting process equipment by keeping operating variables within their design limits
- Automatic startup and shutdown
- Protection against instrument failures
- Selection of one from among several signals

OVERRIDES

In override configurations one controller can take command of a manipulated variable away from another controller when otherwise the process would exceed some process or equipment limit or constraint.

Selective control is less abrupt than the use of interlocks, which usually shut down equipment in order to avoid exceeding a limit or constraint. Overrides usually keep some process variable from reaching an unsafe condition, and therefore the interlock trip points are not reached. Thus, selective control keeps the equipment running although perhaps at a suboptimal level.

This concept is illustrated in Figure 2.28a. The "hard" constraint denotes the point at which the interlock trips. Overrides come into play at some point before the interlocks are actuated and therefore are sometimes called "soft" constraints.

Signal selectors can facilitate the overriding of one controller by another. Often, overrides are preferable and therefore are activated before the safety interlocks would be, but in most applications the overrides are backed up by interlocks. Overrides can be defined as:

> Controllers that remain inactive until a constraint is about to be reached or exceeded, at which point they take over control of the manipulated variable from the normal controller through a selector and thereby prevent the exceeding of that constraint.

Some common applications of override include the prevention of:

- Flooding in distillation columns, by throttling boil-up or feed flow rates
- Exceeding level ranges by draining or flooding
- High pressure or temperature caused by a runaway reaction when heat input is reduced
- The development of low oxygen levels in furnace off-gas streams, by reducing fuel flow

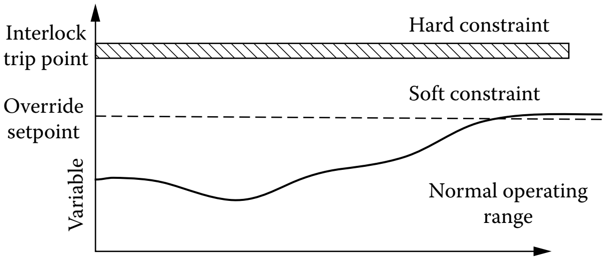


FIG. 2.28a
Illustration of the nature of "hard" and "soft" constraints. An interlock is a "hard" constraint because it usually shuts down the protected equipment, while selective overrides are called "soft" because they keep the process variable from reaching such "hard" limits.

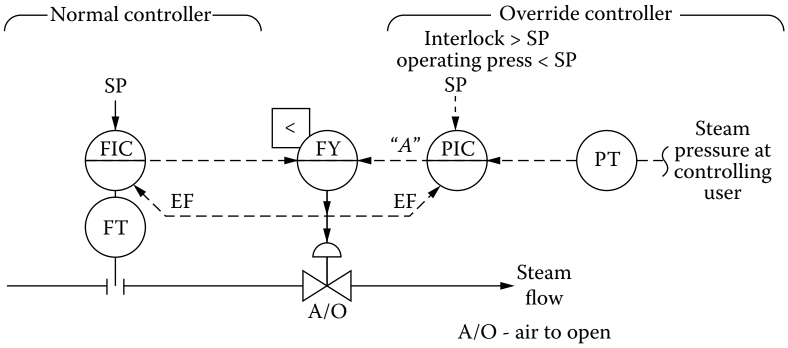


FIG. 2.28b
Flow control with pressure override limits the steam pressure to a safe value. Both controllers are provided with external feedback (EF) to prevent reset windup and thereby guarantee bumpless transfer.

- The development of high steam header pressures, by diverting some of the steam to a low-pressure header or condenser

Figure 2.28b shows an override control loop in which the normal control maintains flow (FIC) while the safety override control is based on pressure (PIC). In this configuration, the outputs of both controllers are fed to a low-signal selector, which selects the lower of the two. The override controller set point is set at the maximum steam pressure that the process can tolerate, but below the safety interlock set point.

As long as the pressure controller set point is not exceeded, the output of the override PIC controller (signal "A") is blocked by the low selector and cannot reach the steam valve. When the pressure of the steam that is being sent to the users exceeds the set point of the override PIC, that controller's output signal decreases, and when it drops below the flow controller's output, it is selected for throttling the steam control valve. This way, the PIC override will prevent the steam pressure from rising above the controller's set point.

Both the FIC and the PIC control algorithms require "external reset feedback" to avoid integrating their errors when they are idle ("reset windup"), because their output is not selected for control. As was discussed in more detail in Section 2.2, the controllers do not receive their own output signal as reset feedback, but both of them receive the signal selected by the low signal selector (FY). This makes the transfer between the controllers bumpless.

Overriding at a Fixed Point

Figure 2.28b illustrates the case when the override variable (steam pressure) had to be limited to a specific value. Another example of this type of application is shown in Figure 2.28c, where the controls involve cooling with river water. In this application the concern is heat exchanger fouling because if the water outlet temperature exceeds 50°C, exchanger fouling becomes rapid.

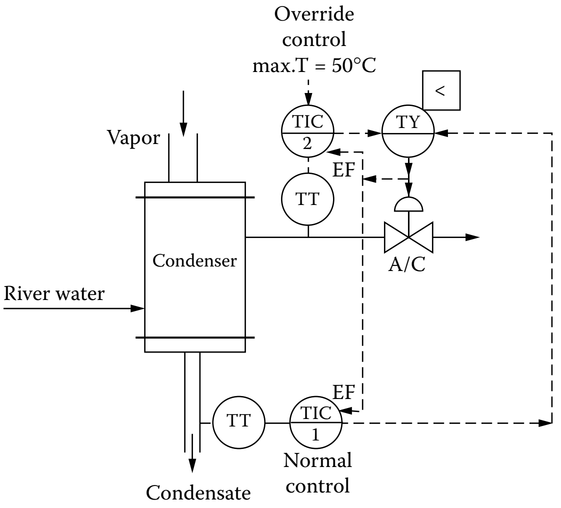


FIG. 2.28c
When the river water outlet temperature rises to 50°C, the override controller (TIC-2) overrides the normal controller (TIC-1) and prevents the outlet temperature from rising beyond 50°C.

Normally, the river water flow is throttled to control the condensate temperature by TIC-1. However, if the river water outlet temperature reaches or exceeds 50°C, the high-temperature override controller (TIC-2) takes over the control of the flow of river water and opens the control valve. This limits the river water outlet temperature to a maximum of 50°C. As a consequence, the condensate temperature will drop below the set point of TIC-1, but the fouling of the heat exchanger will be prevented, and that is the higher priority.

Figure 2.28d describes a control system that protects against overpressurizing a reactor. This protection is provided by overriding the temperature controls and reducing the heat input. In this application a cascade loop throttles the steam flow and controls the reactor temperature if the pressure in the reactor is safe. On the other hand, when the overhead condenser becomes overloaded and its capacity to condense the overhead vapors is insufficient for the rate at which the vapors are generated, the reactor pressure will rise. The pressure controller (PC) provides the high pressure override by taking command of the steam valve (through the low signal selector FY) and throttling down the valve as needed. Both the PC and the FIC are provided with external feedback (EF).

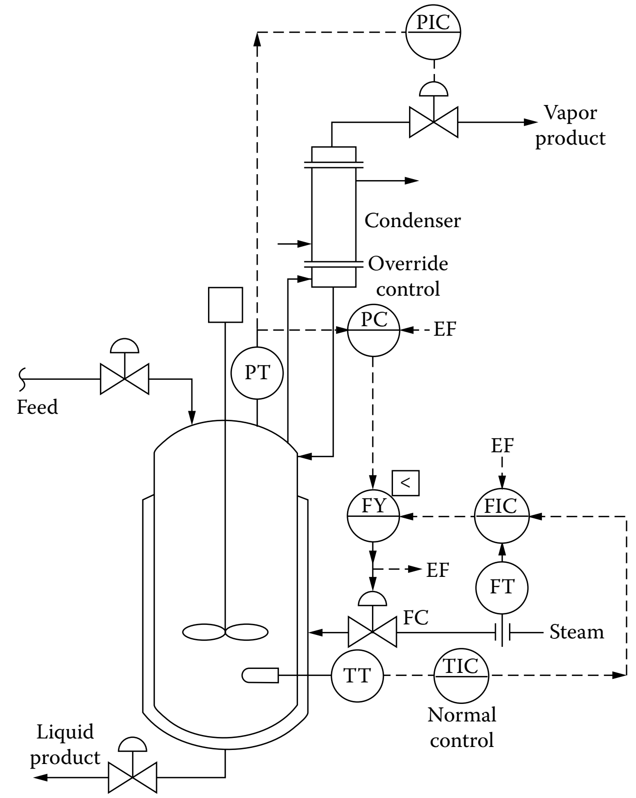


FIG. 2.28d
The temperature controlling cascade loop is overridden in this reactor control configuration when the pressure is high.

Override to Guarantee Valve Closure

Figure 2.28e illustrates another type of override control. Instead of a PID controller, here the high-level override is provided by a signal scaler (LY). The purpose of the high-level override is to make sure that the feed valve of the distillation column is closed when the base level reaches 100%.

The override in this case protects against potential restrictions in the tails (bottoms) flow causing the flooding of the column. A scaler is used because the gain and bias of the LY output must have a preset relationship to the input. This cannot be obtained with a PID controller because the controller output cannot be predicted for a given input.

The scaler output $[f(x)]$ can be predetermined as a gain (K) multiplied by the level transmitter output in percentage ($\%x$) plus a bias (b):

$$f(x) = Kx + b \qquad \textbf{2.28(1)}$$

The scaler can therefore be calibrated to behave in accordance with the relationship shown on the right of Figure 2.28e, where the output is 0% when the level is 100% and the output is 100% when the level is under 75%. This behavior is obtained by setting $K = -4$ and $b = 400$:

$$f(x) = -4x + 400 \qquad \textbf{2.28(2)}$$

As level increases beyond 75%, the output of the scaler will gradually decrease and at some point will drop below

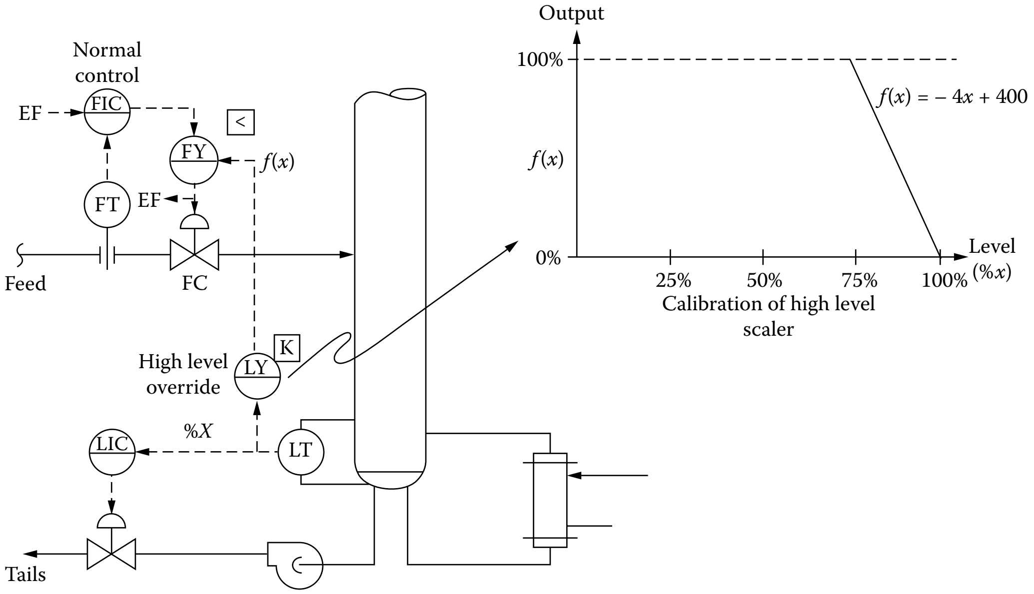


FIG. 2.28e
Instead of a high-level controller, a scaler can also provide a high base level override, which will also prevent the flooding of the column by closing the feed valve.

that of the feed flow controller (FIC), at which point the low signal selector (FY) will select it. At that point LY will override the normal controller (FIC) and will start to throttle the feed valve. The exact point of takeover cannot be predicted because the PID controller output cannot be predicted. However, we definitely know that when the level is 100%, the feed valve will be closed.

The gain of 4.0 was chosen to provide a smooth and stable response. It could be higher or lower depending on how fast or slow the override action should be. The lower the gain, the more likely it is that the override will interfere with flow control when the level is not yet critical. Thus the gain should be set high enough to prevent that while still providing stability.

The scaler does not have an integral mode; it acts like a proportional-only controller. Therefore it will not wind up and does not require external feedback (EF), while the FIC does.

Start-Up and Shut-Down Overrides

Some process startups are too complex to be handled by the operator without the aid of some automatic controls. Automatic start-up controls typically include a ramping signal to open valves and override controls to prevent the violation of constraints. Signal selectors decide which control signal should manipulate each valve. Figure 2.28f illustrates such a strategy for a distillation column.

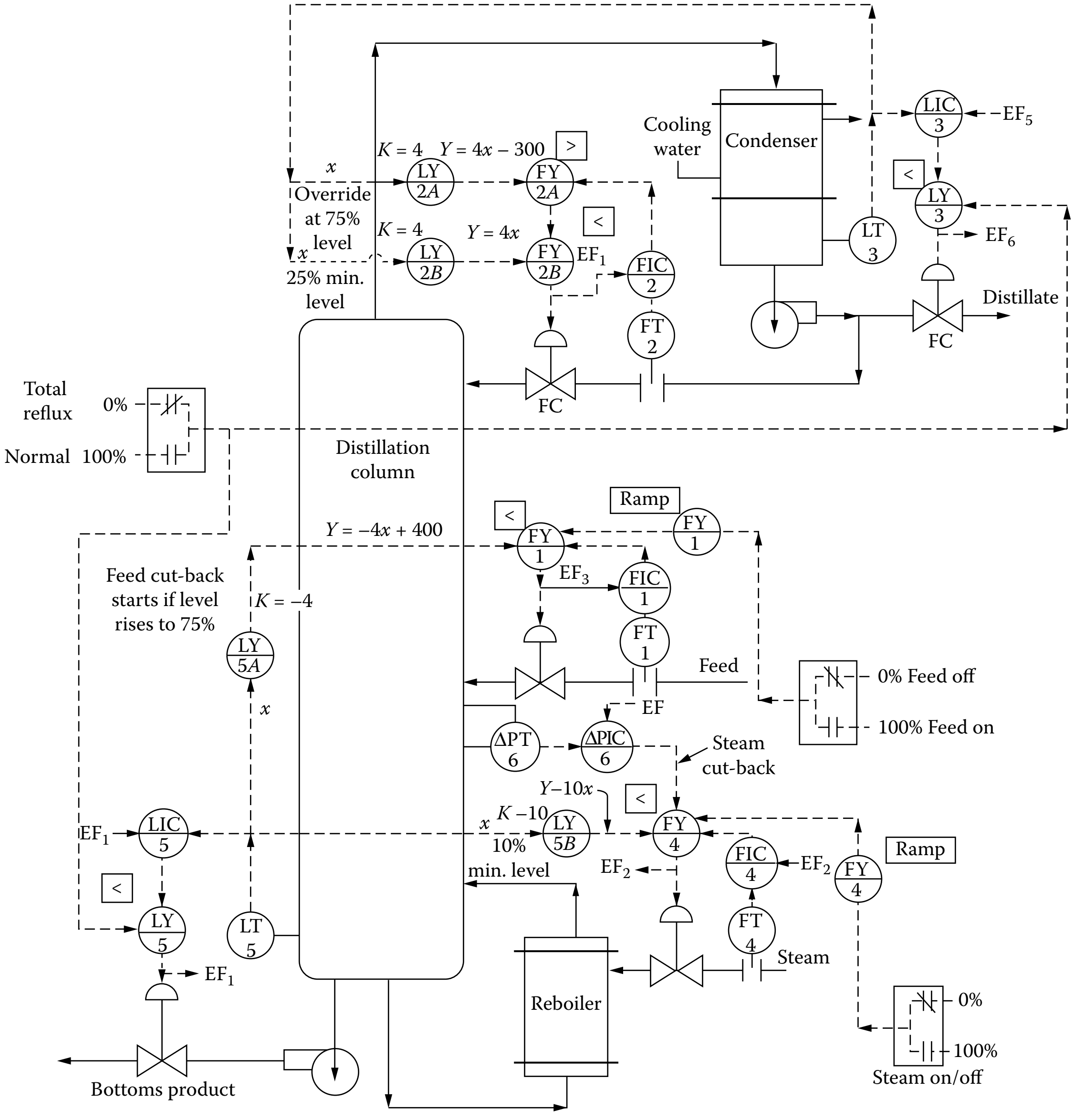


FIG. 2.28f
Overrides required to automatically start up or shut down a distillation tower.

Startup proceeds by opening the cooling water to the condenser first. After that, all the controllers are switched to automatic, maintaining their set points. External reset feedback connections are provided to keep the idle controllers from winding up. For the initial phase of the startup, the total reflux mode is selected, which means that the distillate and tails valves are closed.

The operator turns on the feed by "ramping" the feed valve until the flow set point is reached on FIC-1. As the process fluid accumulates in the base, the operator can turn on the steam. The low-level override (LY-5B) will guarantee that the steam does not come on before a 10% level is established in the base. The steam valve will be ramped open, and the liquid in the reboiler will begin to boil. If the base level gets too high, another override (LY-5A) will throttle back on the feed. This override was described in detail in Figure 2.28e.

As steam comes on, the high delta pressure override (PIC-6) protects the column from "flooding," which can occur when a high pressure differential does not allow the liquid to flow down the column, causing liquid backup on the trays.

As overhead vapors are condensed, the condensate will accumulate in the condenser. As its level rises, the low-level override output (LY-2B) will rise, allowing the reflux flow to come on after the reflux pump was started. The high- and low-level overrides (FY-2A and FY-2B) will return all the condensate to the column (full reflux).

As the temperatures are established in the column, the operator can switch from the total reflux mode of the startup phase to the normal operating mode and thereby the bottoms (tails) and distillate valves will start to open under the direction of their normal controllers (LIC-3 and LIC-5).

The following overrides were used in Figure 2.28f to facilitate automatic startup:

- Low base level (<10%) pinches steam.
- High delta pressure pinches steam.
- High base level (>75%) pinches feed.
- High level in condenser (>75%) increases reflux.
- Low level in condenser (<25%) pinches reflux.

SELECTIVE CONTROL

Limited Availability of Manipulated Variable

In some control applications only one controlled variable exists, but a choice of manipulated variables is present. A typical example is the firing of a process heater with either of two fuels. The choice between the fuels is usually made on the basis of availability.

Fuel *A* may be burned to the limit of its availability and when that limit is reached, it has to be supplemented with fuel *B*. The availability limit can be set manually or by a controller that detects the amount of fuel *A* in storage.

A control system of this type must have the following two features:

1. Capability of manipulating the variable with limited availability while staying below its limit
2. Smooth transition from using one manipulated variable to the other without adversely affecting the controlled variable

Accommodating these requirements necessitates the coordinating of the manipulated variables and the weighing of their effects on the process.

Burning Multiple Fuels Figure 2.28g shows a control in which the temperature of the process is controlled by manipulating two fuels, of which one is of limited availability. Since fuel *A* is less expensive, the objective is to use that fuel to the limit of its availability before starting to use fuel *B*.

In the control system, the set point of the fuel "*A*" flow controller (FIC-A) is limited by the low signal selector (FY-C) to a high limit. This limit could be set manually or could come from the output of a pressure controller on the fuel header.

The computations shown in Figure 2.28g are needed to guarantee the smooth transition between the two modes of operation. The first step in the computation is to convert the difference between the output of the temperature controller

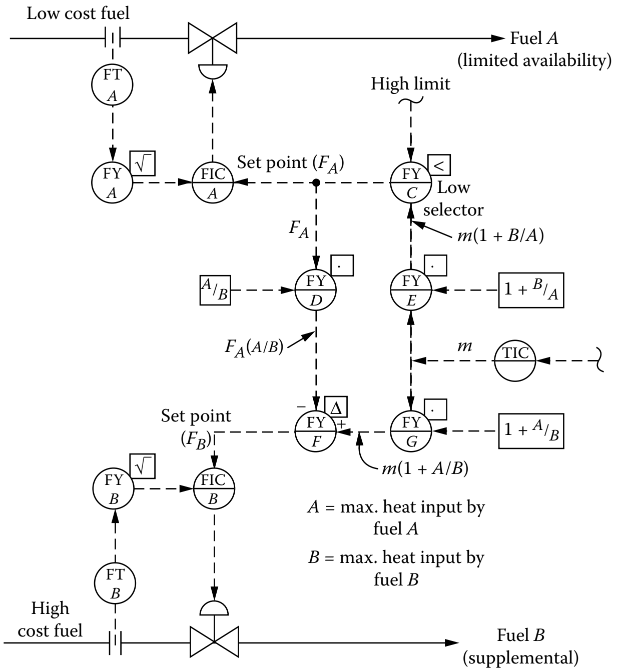


FIG. 2.28g
Temperature is controlled by manipulating the preferred fuel A to the limit of its availability and supplementing it with fuel B as required.

(m) and the equivalent flow of fuel A into a set-point signal for the supplemental fuel flow controller (FIC-B).

The temperature controller output represents the total heat input requirement of the system, which has to be met by the two fuels. This total is converted into the fuel A set point in FY-E, where m is multiplied by a factor of $(1 + B/A)$. Here, B/A is the ratio of the full-scale heat input that can be generated by fuel B to that of fuel A. Through this correction, the controller output and the fuel A set point will be on the same scale. As long as the requirement for fuel A is at or below the high limit set on FY-C, the two inputs to the subtracting relay (FY-F) will cancel one another:

$$m(1 + A/B) - m(1 + B/A)(A/B) = 0 \qquad \textbf{2.28(3)}$$

When the high limit on FY-C is exceeded, the fuel A set point (F_A) becomes that high limit (a constant), and the fuel B set point starts increasing smoothly from zero without affecting the gain of the temperature control loop. The system will also work when the fuel A controller is in manual, as long as the heat load is above the amount that fuel A can supply.

Note that when fuel A is constrained, the temperature controller drives to increase the flow of fuel B. In this control system, external reset feedback is not required in the temperature controller.

Selecting from Multiple Transmitters

Measurement reliability can be improved by selecting a valid transmitter signal from among several signals. Figure 2.28h illustrates the feed controls to a reactor where it is critical to keep the ratio of feed streams within very tight limits. Losing a measurement signal because of a transmitter failure could cause one of the valves to open wide and result in exceeding the ratio limit.

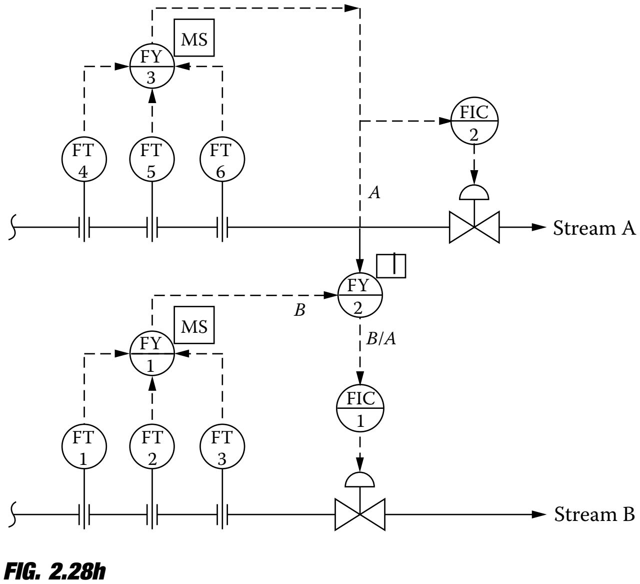


FIG. 2.28h
Selecting the median output of multiple transmitters will keep the control system in operation even if one of the transmitters fails.

If redundant transmitters monitor the flow and a median selector is used to select the signal in the one, that possibility is minimized because two transmitters would have to fail at once for the multiple transmitter configuration to malfunction.

Redundant transmitters or redundant sensors are commonly used in hostile environments (high-temperature, corrosive, dirty, or vibrating surroundings) where transmitter failure rates are high. High or median selectors will keep the control loops operational even if one transmitter fails, avoiding costly shutdowns.

High–Low Selection

Signal selectors can also be used to allow for the selection of the highest or lowest of several measurements for safe control. For example, in order to eliminate the possibility of damage, it is desirable to control fixed-bed chemical reactors on the basis of the hottest temperature, while the location of that temperature can change with both flow rate and time. Thus, temperatures at several locations should be measured and the highest selected as the temperature for control.

ANTIRESET WINDUP OVERRIDES

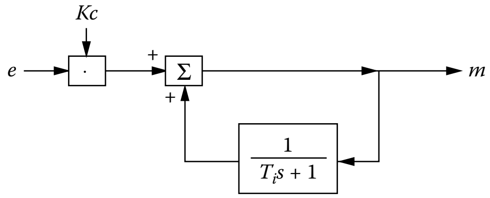
Integral or reset action in a PID controller will cause the controller output to saturate if the controller is off set point because it has no access to the manipulated variable. When the controller experiences such a sustained error, this is called reset windup. It can occur in control loops, where a selector chooses one for active operation from among the outputs of two or more PID controllers while blocking the other. In that case, only one controller at a time commands the valve, while the output(s) of the blocked controller(s) will wind up, unless something is done to limit the reset action.

The PI Controller

The output signal from a PI controller is

$$m(t) = Kc\, e(t) + Kc/Ti \int e(t)\, dt + b \qquad \textbf{2.28(4)}$$

When the PI controller is in automatic but its output is blocked from reaching the valve, and the deviation $e(t)$ is not zero, the error will be integrated until the controller saturates its output either at 100 or 0% depending on the action of the controller.



When $e \neq 0$, positive feedback will cause m to eventually increase to the limit.

FIG. 2.28i
This block diagram illustrates that if the output signal (m) of a controller is blocked from reaching its control valve (the manipulated variable), the resulting sustained error will be integrated until the output (m) reaches a saturated value (100 or 0%).

In the Laplace domain, the PI controller can be described as

$$m(s) = Kc(1 + 1/Ti\ s)\ e(s) \qquad \textbf{2.28(5)}$$

where s is the differential operator d/dt. Equation 2.28(5) can be rearranged to obtain

$$m(s) = Kc\ e(s) + m(s)/(Ti\ s + 1) \qquad \textbf{2.28(6)}$$

The equivalent block diagram of Equation 2.28(6) is given in Figure 2.28i. Here it is shown that the PI controller output signal (m) consists of a "proportional" component ($e(s)$ multiplied by the controller gain Kc) and an integral component that is obtained by feedback of output $m(s)$ through a first-order lag. When $e(s)$ is not zero, $m(s)$ increases (or decreases) to its limit through the positive feedback loop created by the lag function.

Figure 2.28j shows a way to limit this reset windup by breaking the internal feedback loop and providing an external feedback signal (EF) to be fed back through the lag function. For simple loops, the feedback signal can be the signal that is throttling the valve taken from the outlet of the selector.

When the controller is selected for control, the feedback signal is its own output, and when it is not selected, the feedback signal is the output of the selected controller. In the second case, the blocked controller tracks the output of the selected controller through its reset lag.

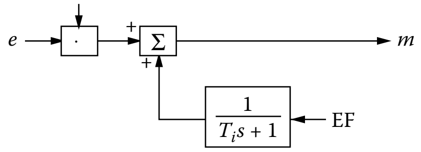


FIG. 2.28j
This block diagram illustrates that in order to eliminate reset windup the blocked controller does not integrate its own output but rather receives an external feedback (EF) from the signal that is operating the control valve (the output of the selector).

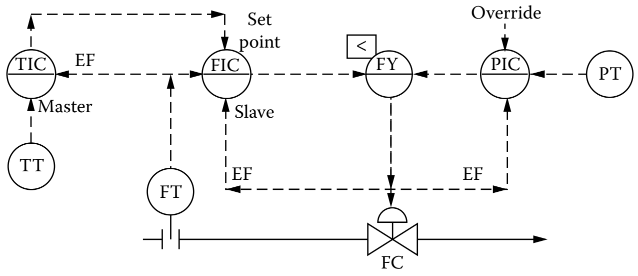


FIG. 2.28k
For TIC, the cascade master, the external feedback signal is the slave measurement. If the slave (FIC) can be overridden by a pressure controller (PIC), the external feedback to both is the manipulated variable going to the valve.

The tracking speed is governed by the integral (reset) time. The longer the integral time, the slower the tracking. Slow integral can be a problem in control configurations where the integral times of the controllers feeding the selector have vastly different values. This is because the slower controller can be wrongly selected due to it not being able to track the selector output fast enough. For this reason, it is sometimes necessary to adapt (automatically change when the controller is blocked) the integral time in the unselected controller to a short time to make sure that the controller will track rapidly.

In the steady state, the unselected controller tracks the selected controller with an offset of $Kc\ e(t)$.

$$m(t) = Kc\ e(t) + F(t) \qquad \textbf{2.28(7)}$$

Cascade and Feedforward Loops Figure 2.28k illustrates the use of external feedback in a cascade loop, and Figure 2.28l illustrates its use in a loop that contains a feedforward signal.

In Figure 2.28l, the external feedback (EF) for the master controller in a cascade loop, which is setting the ratio of flows A and B, is the actual measured flow ratio variable (B/A). The external feedback of this TIC master, whose output is modified by a feedforward signal when it is *multiplied* by A, is the inverse of the feedforward function because the secondary variable B is *divided* by A (Figure 2.28l).

One other type of external reset algorithm is created by modifying the standard PI algorithm by adding a term that integrates the difference between the controller output and the feedback signal.

$$m(t) = Kc\ e(t) + Kc/Ti \int e(t)\ dt + 1/Ti \int [F(t) - m(t)]\ dt \quad \textbf{2.28(8)}$$

When the controller is not selected, the controller output and feedback signal will be different, and integration occurs. The new term cancels out the normal reset action so, in effect, reset action is stopped until that controller resumes control. Equation 2.28(8) behaves exactly like the method used in Figure 2.28j.

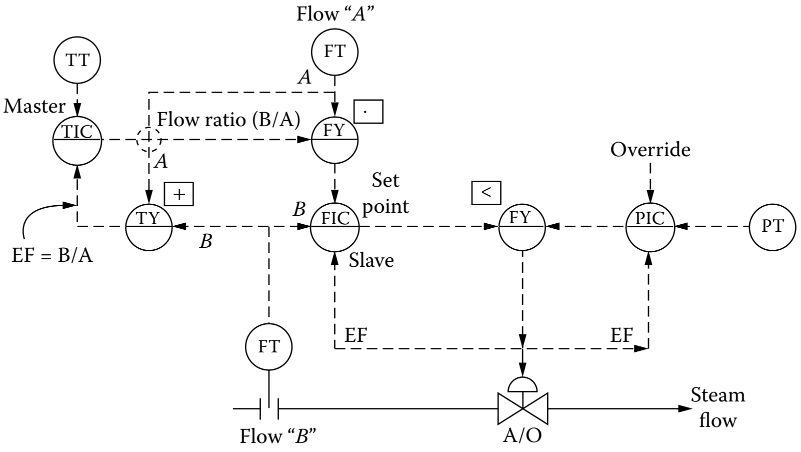


FIG. 2.28l
If the cascade master (TIC) is modulating the ratio of two streams (B/A), the external feedback signal to it will be the actual measured ratio of those flows (B/A). As the slave (FIC) can be overridden by a pressure controller (PIC), the external feedback to both is the manipulated variable going to the valve.

DESIGN PROCEDURE

When designing a control system that includes the feature of selecting the output of one of two controllers, one must consider:

1. The process variables that are to be controlled by the manipulated variable
2. The failure position of the final control element (valve)
3. The safe set points for the controllers
4. The required controller actions
5. The type of selector required (high, low, other)

In the majority of safety applications, a low signal selector is used because control valves, which are installed for reasons of safety, should always fail safe, and one of the possible causes of failure is air failure.

As an application example, consider a unit process consisting of a suction drum, a reciprocating compressor, and a discharge vessel (Figure 2.28m). The purpose of the control system is to prevent the formation of vacuum in the suction drum and the development of excessive pressure in the discharge vessel. Both of these goals can be satisfied by recirculating some of the gas from the discharge vessel to the suction drum. As was discussed in general terms earlier, the design steps include:

Step 1: An open valve failure position must be selected because if the valve is open both the development of vacuum in the suction drum and overpressure in the discharge vessel are prevented.

Step 2: Sensors are placed to detect the variables of interest, namely the vacuum on the suction and the overpressure on the discharge side.

Step 3: Because the development of either high vacuum or high pressure will require the opening of the recirculating valve, the controller actions must be set accordingly.

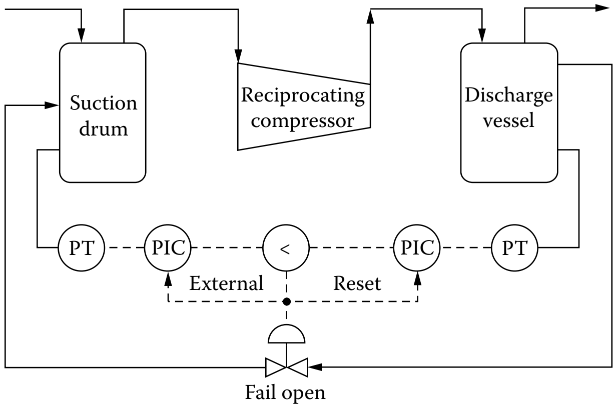


FIG. 2.28m
Safety override controls on a compressor prevent the development of both vacuum in the suction drum and overpressure in the discharge vessel.

Step 4: The action of the suction controller shall be direct (as the measured pressure drops = vacuum rises, the output also drops). The action of the discharge controller shall be reverse (as the measured pressure increases, the output decreases).

Step 5: Because safety is served by opening the recirculating valve, the selector should pick that controller output, which requires a higher valve opening. With a fail open valve, that means a Low Signal Selector.

Step 6: To complete the design, an external reset should be specified for both controllers, which provides them with feedback from the output of the signal selector.

Bibliography

Anderson, L. M., "New Extraction Control System for Steam Turbines," *Hydrocarbon Processing,* February 1976.

Balchen, J. G., and Mummè, K. I., *Process Control Structures and Applications,* New York: Van Nostrand Reinhold, 1988.

Buckley, P. S., "Designing Override and Feedforward Controls," *Control Engineering,* August 1971.

Buckley, P. S., "Protective Controls for a Chemical Reactor," *Chemical Engineering,* April 20, 1970.

Buckley, P. S., *Techniques of Process Control,* New York: John Wiley & Sons, 1964.

Buckley, P. S., and Cox, R. K., "New Development in Overrides for Distillations," *ISA Transactions,* Vol. 10, No. 4, 1971.

Buckley, P. S., et al., *Design of Distillation Column Control Systems,* Research Triangle Park, NC: ISA, 1985.

Buzzard, W. S., "Override Strategies for Analog Control," *Instrumentation Technology,* November 1978.

Eckman, D. P., *Automatic Process Control,* New York: John Wiley & Sons, 1958.

Erickson, K. T., and Hedrick, J. L., *Plant-Wide Process Control,* New York: John Wiley & Sons, 1999.

Gaines, L. D., and Giles, R. F., "Making the Most of Changing Fuel Supplies," *Instruments and Control Systems,* February 1978.

Giles, R. F., and Gaines, L. D., "Integral-Tracking Override is Better than Output-Tracking," *Control Engineering,* February 1978.

Hall, C. J., "Don't Live with Reset Windup," *Instrumentation Technology,* July 1974.

Harriott, P., *Process Control,* New York: McGraw-Hill, 1964.

Kecman, V., *State-Space Models of Lumped and Distributed Systems,* Heidelberg: Springer-Verlag, 1988.

Luyben, W., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* New York: McGraw-Hill, 1990.

Luyben, W. L., and Luyben, M., *Essentials of Process Control,* New York: McGraw-Hill, 1997.

Marlin, T., *Process Control: Designing Processes & Control Systems for Dynamic Performance,* New York: McGraw-Hill, 2000.

McMillan, G., *Tuning and Control Loop Performance,* 2nd ed., Research Triangle Park, NC: ISA, 1990, pp. 45–55.

Murrill, P., *Application Concepts of Process Control,* Research Triangle Park, NC: ISA, 1988.

Murrill, P., *Fundamentals of Process Control,* 2nd ed., Research Triangle Park, North Carolina: ISA, 1991.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Nisenfeld, A. E., and Seemann, R. C., *Distillation Columns,* Research Triangle Park, NC: ISA, 1981.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Rijnsdorp, J. E., MacGregor, J. F., Tyreus, B. D., and Takamatsu, T., *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes,* New York: Pergamon Press, 1990.

Seborg, D., and Mellichamp, D. A., *Process Dynamics and Control,* 2nd ed., New York: McGraw-Hill, 2004.

Shinskey, F. G., *Distillation Control,* 2nd ed., New York: McGraw-Hill, 1984.

Shinskey, F. G., *Energy Conservation through Control,* New York: Academic Press, 1978.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Shunta, J. P., and Cox, R. K., "Sampled Data Controller for Loops with Overrides," paper presented at 74th National AIChE Meeting, New Orleans, March 1973.

Shunta, J. P., and Klein, W. F., "Microcomputer Digital Control — What It Ought to Do," *ISA Transactions,* Vol. 18, No. 1, 1979.

Skogestad, S., Lundstrom, P., and Jacobsen, E. W., "Selecting the Best Distillation Control Configuration," *AIChE Journal,* Vol. 36, May 1990.

Smith, C. A., *Automated Continuous Process Control,* New York: John Wiley & Sons, 2002.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1997.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

Wightman, E. J., *Instrumentation in Process Control,* London: Butterworth, 1972.

2.29 Self-Tuning Controllers

P. E. WELLSTEAD (1995) **B. G. LIPTÁK, S. RENGANATHAN** (2005)

Types of Products on the Market:	A. Stand-alone hardware mostly for temperature control B. Stand-alone hardware for tuning all process loops C. Offered as part of DCS supplier's package D. Self-tuning software package
Partial List of Suppliers (Product Names):	ABB Process Automation, U.S. (MC 5000) (B) (www.abb.com/us/instrumentation) Asea Brown Boveri, Sweden (Novatune) (B, C) (www.abb.com) CAL Controls, U.S. (9900 Autotune) (A) (www.cal-controls.com) Control Techniques, U.K. ("Expert") (B) (www.controltech.co.uk) Eagle Controls, U.K. (Eagle Mini 948) (A) (www.applegate.co.uk) Emerson Process Measurement, U.S. (RS3) (B) (www.emersonprocess.com) Eurotherm, U.S. (815, 8181) (A) (www.eurotherm.com) FGH Controls, U.K. (5900, 556) (A) (www.fgh.co.uk) Fuji Electric, Japan (CC-S) (B) (www.fujielectric.co.jp) Honeywell, U.S. (UDC 6000, E-Max, D-Max) (B) (www.iac.honeywell.com) Invensys, Foxboro, U.S. ("Exact" 761, 760) (B) (www.foxboro.com) Jumo, U.S. (Dicon 5) (A) (www.jumousa.com) Philips, U.K. (KS 4290) (A) (www.philips.co.uk) Self-Tuning Friend, U.K. (D) (www.csc.umist.ac.uk) Siemens, Germany (Sipart dr) (B) (www.sea.siemens.com) Toshiba, Japan (EC300, 2150) (B) (www.tic.toshiba.com) Yokogawa, U.S. (UP25, YEW80) (A) (www.yokogawa.com/us)

INTRODUCTION

In Section 2.18, the methods of process modeling using artificial neural networks (ANN) were described. Some of the concepts used in designing self-tuning controllers use the concepts of the ANN family (Figure 2.29a). Before reading this section, it is advisable to become familiar not only with ANN, but also with the basics of PID controllers, their tuning, and the related subjects of model-based and adaptive controls, which are all covered in this section.

The selection of controller settings that will provide optimum performance is called controller tuning. If the controller is manually tuned and the dynamics of the controlled process change, the tuning has to be done again. If on the other hand, the tuning is done automatically periodically, then it is known as self-tuning. Some DCS and other control system suppliers provide the means to allow the operator or a timer to initiate controller tuning.

Self-tuning controllers are capable of automatically readjusting the controller tuning settings. They are also referred to as auto-tuning controllers and can be stand-alone products, integral parts of distributed computer control systems (DCS), or software packages. The market is dominated by stand-alone products, most of which are able to communicate with other systems or other controllers.

The standard operating principles include self-tuning regulators (STR) and self-tuning temperature controllers using feature-extraction–type tuning methods, which are obtained from step response information taken at startup or when load changes occur. General-purpose stand-alone units can tune by pattern recognition methods (e.g., Foxboro Exact) or by model-based methods (e.g., ABB Novatune).

EVOLUTION

The first self-tuning controllers came from Europe in the early 1970s. The original methods used optimal regulation[1,2] with optimal control,[3] followed by pole-placement[4] somewhat later. The subsequent development involved new algorithms,[5,6] stability theory,[7] and the industrial development of commercial self-tuning controllers. A comprehensive treatment of algorithms

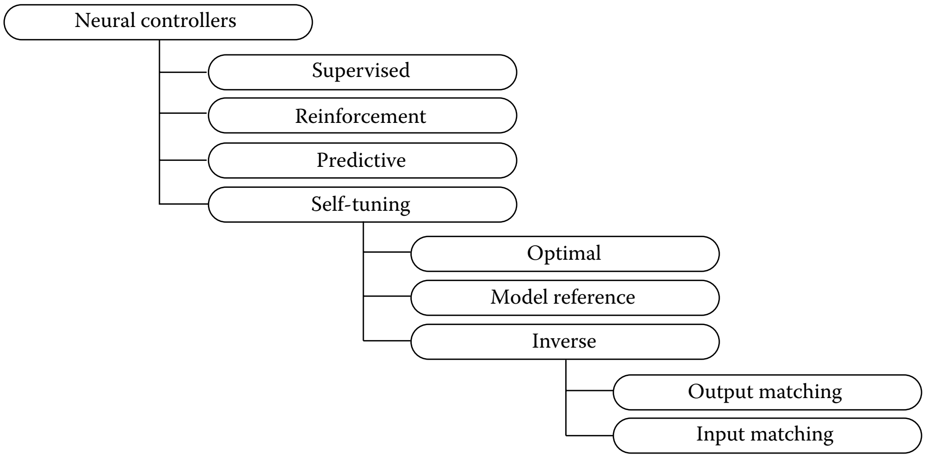


FIG. 2.29a

A hierarchical classification of ANN-based control algorithms. Self-tuning controllers belong to the family of neural controllers, which are distinguished from each other by the way they use the error signal to train the ANN.

for self-tuning control is provided in Reference 8, while nonlinear and adaptive controls are covered in another section of this chapter.

The general principle of operation of the family of self-adaptation is described in Figure 2.29b. This approach is applied when the cause of changes in the control loop response is either unknown or unmeasurable and the adaptation therefore must be based on the response of the loop itself. Several self-adaptive systems are described in the paragraphs that follow, including the self-tuning regulator, the model reference controller, and the pattern-recognizing adaptive controller.

Self-Tuning Regulator (STR)

The main components of a self-tuning regulator are shown in Figure 2.29c. They are as follows:

1. *A system identifier* — This element consists of a process parameter estimation algorithm, which estimates the parameters of the process.
2. *A controller synthesizer* — This synthesizer calculates the new controller parameters as a function of the estimated process parameters specified by the control objective function.
3. *A controller implementation block* — This is the controller whose parameters are updated at periodic intervals by the controller parameter calculator.

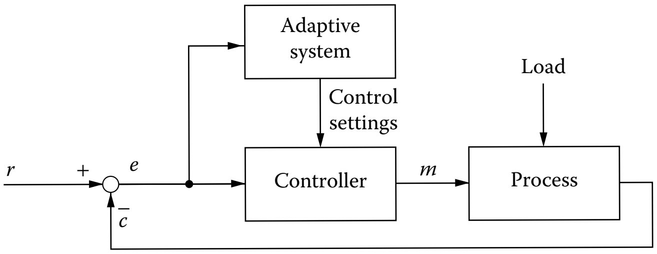


FIG. 2.29b

In self-adaptation, the correction is based on the response of the loop itself. One can view a self-adaptive system as a control loop around a control loop.

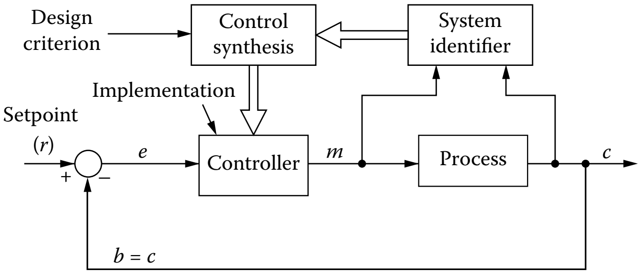


FIG. 2.29c

The main components of a self-tuning regulator.

The STRs are distinguished on the basis of their identifiers and synthesizers. Popular varieties include the minimum-variance, generalized minimum-variance, detuned minimum-variance, dead-beat, and generalized pole-placement controllers.

As shown in Figure 2.29c, the system identifier determines the response of the controlled variable (c) to a change in the manipulated variable (m). Model-based self-tuning algorithms use some method based on recursive estimators. Many commonly used industrial products, however, use pattern recognition or expert/fuzzy logic methods to extract key dynamic response features from a transient excursion in the system dynamics.

The transient excursion can be deliberately introduced by the controller, or preferably, it is the start-up transient or a normally occurring transient of the process. The desired tuning settings for PID, optimal, or other types of control are determined by the control synthesis block, which is also called the controller parameter calculator. The control synthesizer can be simple or sophisticated depending on the rules used.

MODEL-BASED METHODS

There are two classes of model-based self-tuners: optimal and response specification types. Both of these controllers are composed of a reference model, which specifies the desired performance; an adjustable controller, which sets the manipulated so as to bring the process performance as close as possible to that of the reference model; and an adaptation mechanism.

Figure 2.29d schematically shows the organization of the reference model controller.

Optimal Self-Tuning Optimal self-tuning algorithms use optimal regulation theory as their design rule in the controller parameter calculator. The design methods used are minimum variance (MV), generalized minimum variance (GMV), and generalized predictive control (GPC). The MV and GMV methods are, as the names imply, aimed at minimizing the mean square deviation of a process variable from its set point. The minimum variance methods have deficiencies when the process time delay is unknown or variable because they can, if unprotected, become unstable.

The related optimal method GPC was developed to overcome this deficiency by using prediction horizon ideas that occur in DMC (dynamic matrix control).[9] These methods are used in turn-key applications by research and consulting companies. Few industrial controllers use them because of the sophisticated knowhow required, although the ABB Novatune is a widely installed and respected product that uses optimal methods.

Transient Response Self-Tuner Another class of model-based self-tuners uses the desired closed-loop frequency or the transient response characteristics of the loop as the basis for operating the self-tuning algorithm. The accepted technique for this is pole-placement (PP) or pole-assignment (PA). This procedure asks the designer to select the desired closed-loop pole positions, and the self-tuner selects a controller that does this.

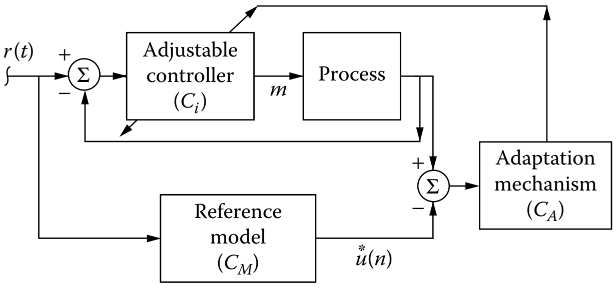


FIG. 2.29d
The main building blocks of a model reference adaptive controller.

The response specification methods can be linked to the optimization methods via various techniques.[8] Pole-placement can handle systems with unknown and variable dead times but requires more computational effort than most optimization-based methods do. Several commercial products offer these methods in one form or another, usually tailored to the requirements of the specific applications.

The model-based methods require that a persistent excitation or "dither" signal be injected into the process. This is normally done in the open loop or at startup, but start-up transients alone are not normally considered sufficient for these self-tuners. The form of the control law is usually a discrete time difference equation that is suitable for digital implementation.

Pattern Recognition Methods

The self-tuning of controllers that utilize pattern recognition procedures is widely used in industrial applications. This method relies upon the introduction of a perturbation of a specified form into the process. This usually is a step or pulse signal. Often the start-up transient is used, which is particularly useful in systems where a special start-up procedure exists, such as the one with temperature controllers.

The process response is then analyzed to extract the key transient performance indicators of the process, which are then used to select the correct controller tuning values. The transient performance indicators can be estimates of rise time, dead time, peak-overshoot, and so on. Some methods use fuzzy or qualitative performance measures, such as the criterion of whether "the time to first overshoot" is very fast, fast, medium, or slow.

The main difference between pattern recognition and the model-based methods is that in the case of pattern recognition no model of the process is constructed. This eliminates problems due to incorrect modeling but limits the usefulness of these methods to processes that can be handled in this way. For example, some of these methods might only work if a dead time and time constant can fully describe the controlled process.

Self-tuners of this type are restricted to three-term (PID) control algorithm applications. In these systems, the PID tuning constants are adjusted according to some tuning rules based upon the measured open-loop transient behavior of the process. The tuning rules are often proprietary variations of such well-known process tuning procedures as the Zeigler Nichols method or others.

A subclass of algorithms is based on testing the stability of the loop. These methods will usually cause the loop to oscillate as in the Zeigler Nichols periodic oscillation (closed-loop) method. Special precautions are applied to avoid runaway oscillation. The period and the controller gain that caused the oscillations are then measured and used to tune the controller.

Eliminating Process Upsets Self-tuning controllers that operate by introducing pulses or other forms of perturbations

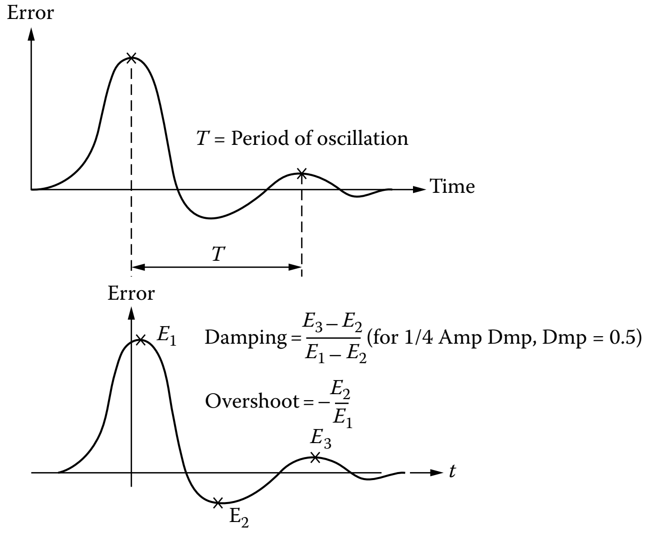


FIG. 2.29e
The control loop tuning is diagnosed by evaluating the period of oscillation (T), the decay or damping ratio, and the amount of overshoot to determine whether the controller is correctly tuned.[10]

are not widely accepted in the process control industry. It is felt that the control loops are unstable enough as it is, and nobody wants additional sources of perturbation.

Therefore, preference is given to self-tuning controllers that do not introduce any upsets but evaluate the controller's response to set-point changes or to other, naturally occurring load variations and other upsets as they take place. In such cases the self-tuning algorithm is usually kept "dormant" (inactive) until an error of some predetermined value (usually at least 1%) develops.

At that point the self-tuning subroutine is activated. After the error has evolved, the self-tuning subroutine checks the response of the controller in terms of the controlled variable's period of oscillation, damping, and overshoot (Figure 2.29e).

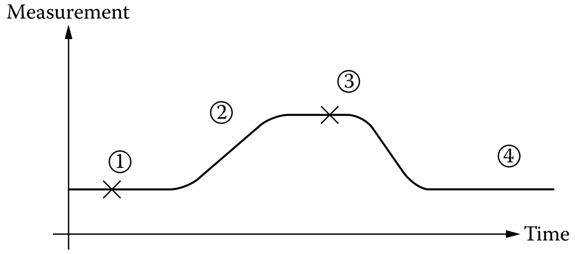


FIG. 2.29g
After evaluating the process response to a step change that increased the process load, another step change is introduced to return the process load to its original value.[10]

Start-Up Sequence As shown in Figure 2.29f, when starting up a control loop for the first time (the pretune phase), the controller is first manually brought to its normal load level of operation. After that, an adjustable-size step change in its output signal to the control value (m) is introduced. This step change in the manipulated variable (m) will cause the controlled variable (c) to draw an S-shaped steady-state reaction curve. From this response curve one can obtain the steady-state process dead time, process sensitivity, and process gain in response to a load increase.

The pretune sequence is continued, as shown in Figure 2.29g, by moving the controller output (m) through another step change back to its original value and determining again the steady-state gain, dead time, and process sensitivity in response to a decrease in load. After these data have been collected and after the "noise-band" of the measurement been identified, these readings are used to automatically determine the start-up PID settings of the controller.

After this pretune or start-up sequence, the self-tuning controller reevaluates its performance after each upset, using the criteria given in Figure 2.29e.

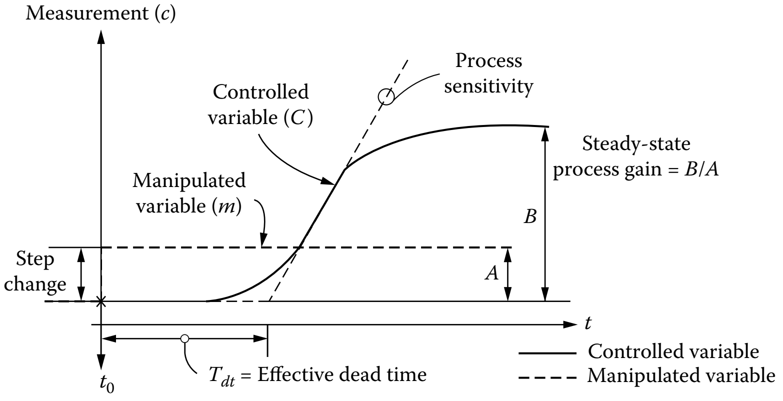


FIG. 2.29f
In the start-up or pretune phase, the self-tuning controller determines the dead time, gain, and sensitivity of the controlled process variable by evaluating its response to a step change.[10]

PERFORMANCE

The performance of such self-tuning controllers has been evaluated on different types of processes. The test results are shown in Figure 2.29h, where the performance ratio between conventionally tuned and self-tuned loops is on the vertical.

The conventionally tuned controllers were tuned for the worst condition, and their performance relative to the self-tuning controllers was evaluated on the basis of the integrated absolute error (IAE), which is the total area under the error curve in Figure 2.29e. In evaluating the test results, one can conclude that the improvement in control quality is impressive for slow processes (low gain) having little or no dead time. These are easy processes to control because they start responding to a correction immediately and can be controlled with high gain (narrow proportional band) controllers.

On the other hand, as the ratio of dead time to time constant rises and as the process becomes faster (process gain increases), the performance of self-tuning controllers deteriorates. When the process gain exceeds two and the dead time exceeds the time constant, the performance of the self-tuning controller actually drops below that of the conventionally tuned loop.

This is not surprising because as the dead time rises, the time it takes for evaluating a response also gets longer, which in turn increases the area under the error curve. For this reason, the use of PID control is not recommended for mostly dead time processes. In such cases a sample-and-hold algorithm (see Section 2.2 in this chapter) is recommended.

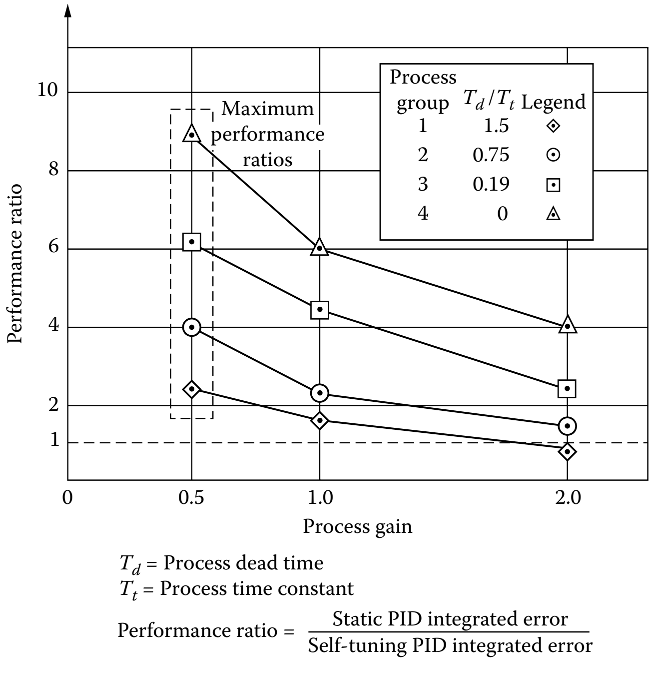


FIG. 2.29h

For slow processes with little dead time, the self-tuning controller outperforms the statically tuned PID controller.[10]

CONCLUSIONS

Process controllers with self-tuning capabilities are mostly stand-alone units that are used for single-loop applications, but self-tuning features are also available in DCS installations or in software packages. The software can be resident in a DCS, or it can be used in general-purpose computers, or perhaps as an external tuning aid.

A special class of self-tuning controllers has evolved for temperature-control applications. These are provided with special tuning procedures (usually based on the response during the warm-up phase or to a set-point change) and include constraints such as minimum overshoot.

Self-tuning controllers used for general-purpose process applications are more sophisticated and are often provided with provisions for multiple loop tuning, feedforward control, gain scheduling, and compensation for actuator and sensor nonlinearity. Most self-tuning controllers will self-tune during startup and will retune both on request and also when upsets occur naturally.

A few self-tuning controllers operate on the basis of continuously introducing excitations into the process. Such excitations can be deterministic set-point steps or doublets (used in pattern recognition methods) or can be persistent set-point variations (often pseudo noise, and usually on the order of 5% of set point).

Self-tuning controllers have gone through substantial development and improvement during the last decades. The better systems no longer depend on pulses or set-point changes for detecting the response of the loop but instead analyze the response to naturally occurring process upsets. These self-tuning controllers give control performance that is superior to that of manually tuned, static PID controllers when the process gain and the process dead time are both low. As the process gain and/or the dead time–to–time constant ratio rises, their performance deteriorates.

References

1. Peterka, V., "Adaptive Digital Regulation of Noisy Systems," Second IFAC Symposium on Identification and Process Parameter Estimation, Paper 6.2, Academica, Prague, 1970.
2. Astrom, K. J., and Wittenmark, B., "On Self-Tuning Regulators," *Automatica*, 9, pp. 185–199, 1973.
3. Clarke, D. W., and Gawthrop, P. G., "Self-Tuning Control," *Proceedings of IEE,* 122, pp. 929–934, 1975.
4. Wellstead, P. E., Edmunds, J. M., Prager, D. L., and Zanker, P. M., "Pole Zero Assignment Self-Tuning Regulator," *International Journal of Control,* 30, pp. 11–26, 1979.
5. Clarke, P. W., Moktadi, C., and Tuffs, P. S., "Generalized Predictive Control," *Automatica*, 23, pp. 137–448, 1987.
6. Lelic, M. A., and Zarrop, M. B., "Generalized Pole-Placement Self-Tuning Controller," *International Journal of Control,* 46, pp. 569–607, 1987.
7. Sastry, S., and Bodson, M., *Adaptive Control: Stability, Convergence and Robustness,* Englewood Cliffs, NJ: Prentice Hall, 1989.
8. Wellstead, P. E., and Zarrop, M. B., *Self-Tuning Systems,* New York: Wiley, 1991.

9. Seborg, D., Edgar, E., and Mellichamp, D., *Process Control,* New York: Wiley, 1990.
10. Kraus, T. W., and Myron T. J., “Self-Tuning PID Controller Uses Pattern Recognition Approach,” *Control Engineering,* June 1984.

Bibliography

Abonyi, J., *Fuzzy Model Identification for Control,* Boston: Birkhauser, 2003.

Astrom, K. J., and Hagglund, T., *Automatic Tuning of PID Controllers*, 2nd ed., Research Triangle Park, NC: ISA Press, 1995.

Astrom, K. J., and Wittenmark, B., *Adaptive Control*, Singapore: Pearson Education, 2001.

Astrom, K. J., and Wittenmark, B., *Adaptive Control*, 2nd ed., Reading, MA: Addison-Wesley, 1995.

Badgwell, T. A., and Qin, S. J., “Review of Nonlinear Model Predictive Control Applications,” *Nonlinear Predictive Control: Theory and Practice,* IEEE Publishing, 2001.

Chindambram, M., *Computer Control of Processes*, New Dehli: Narosa Publishing, 2002.

“Distributed Controllers,” *Measurements and Control,* April 1991.

Ender, D. B., “Troubleshooting Your PID Control Loop,” *InTech,* May 1992.

Fabri, S. G., and Kadirkamanthan, V., *Functional Adaptive Control: An Intelligent Systems Approach,* Heidelberg: Springer-Verlag, 2001.

Howell, J., and Rost, D., “PC Software Tunes Plant Startup,” *InTech,* June 1988.

Hunt, K. J., and Sbarbaro, D., “Neural Networks for Nonlinear Internal Model Control,” *IEE Proceedings/D,* Vol. 138, 1991.

Maciejowski, J. M., *Predictive Control with Constraints,* New York: Addison-Wesley Longman, 2000.

McMillan, G., “A New Era for Model Predictive Control,” *Control,* Vol. 13, No. 5, 2000.

Qin, S. J., and Badgwell, T. A., “An Overview of Industrial Model Predictive Control Technology,” *Chemical Process Control,* 5th International Conference on Chemical Process Control, AIChE and CACHE, 1997.

Rossiter, J. A., *Model-Based Predictive Control,* Boca Raton, FL: CRC Press, 2003.

Seborg, D. E., Edgar, T. F., and Shah, S. L., “Adaptive Control Strategies for Process Control,” *AIChE Journal,* 32, 1986.

Stephanopoulos, G., and Han, C., “Intelligent Systems in Process Engineering: A Review,” *Computers and Chemical Engineering,* Vol. 20, 1996.

Suykens, J. A. K., et al. *Artificial Neural Network for Modelling and Control of Non-Linear Systems,* New York: Kluwer Academic, 1996.

Tao, G., *Adaptive Control Design and Analysis,* New York: John Wiley & Sons, 2003.

VanDored, V., *Techniques for Adaptive Control,* Oxford: Butterworth-Heinemann, 2002.

Wade, H. L., “High-Capability Single-Station Controllers: A Survey,” *InTech,* September 1988.

Wellstead, P. E., and Zarrop, M. B., *Self-Tuning Systems,* New York, Wiley, 1991.

Widrow, B., and Stearns, S. D., *Adaptive Signal Processing,* Englewood Cliffs, NJ: Prentice Hall, 1985.

Wojsznis, K. W., Blevins, L. T., and Nixon, M., “Easy Robust Optimal Predictive Controller,” Advances in Instrumentation and Control, New Orleans, ISA TECH 2000.

Wojsznis, K. W., et al., “Practical Approach to Tuning MPC,” *ISA Transactions,* Vol. 42, January 2003.

2.30 Sliding Mode Control in Process Industry

O. CAMACHO (2005)

Sliding mode control (SMC) is a powerful robust nonlinear control (see Section 2.19) technique that has been intensively developed during the last 35 years.[1,2] Sliding mode control has, for many years, been accredited as one of the key approaches for the systematic design of robust controllers for complex nonlinear dynamic systems operating under conditions of uncertainty. The most important advantage of sliding mode controllers is their innate insensitivity to parameter variations and disturbances once in the sliding mode, thereby eliminating the necessity of exact modeling.[1] Another distinguishing feature is their capability of order reduction, which enables the simplification of design and the decoupling of the system. With these advantages, sliding mode control is a promising area of study for both theoretical and application-oriented problems.

The potential of control resources may be used to the fullest extent within the framework of nonlinear control methods since the actuator limitations and other performance specifications may be included in the design procedure. The scope of SMC studies includes varied problems such as mathematical methods, design principles, and applications.[2–5]

The SMC design is composed of two steps. In the first step, a custom-made surface should be designed. While on the sliding surface, the plant's dynamics is restricted to the equations of the surface and is robust to match plant uncertainties and external disturbances. In the second step, a feedback control law should be designed to provide convergence of a system's trajectory to the sliding surface; thus, the sliding surface should be obtained in a finite time. The system's motion on the sliding surface is called the *sliding mode*.[1,2,5] Perfect tracking can be achieved at the price of control chattering.[7] *Chattering* is a high-frequency oscillation around the desired equilibrium point. It is undesirable in practice because it involves high control activity and can excite high-frequency dynamics ignored in the modeling of the system.[1–7]

The aim of this section is to present the overall concept of sliding mode control. Some references to industrial applications in electromechanical systems are discussed, and a practical approach, using a reduced-order model of chemical processes, for single-input single-output (SISO) systems and multiple-input multiple-output (MIMO) systems is also presented. Furthermore, a proposed SMC implementation of a PID algorithm is also suggested.

DESIGN PROCEDURE

The idea behind SMC is to define a surface along which the process can slide to its desired final value.[1] The structure of the controller is intentionally altered in accordance with a prescribed control law as its state crosses the surface.

First Step

The first step in SMC is to define the sliding surface, $S(t)$, which represents a desired global behavior, such as stability and tracking performance. The $S(t)$ selected in this work, presented by Slotine and Li,[6] is a proportional, integral-differential equation acting on the tracking-error expression:

$$S(t) = \left(\frac{d}{dt} + \lambda\right)^n \int_0^t e(t)dt \qquad \textbf{2.30(1)}$$

where $e(t)$ is the tracking error, that is, the difference between the reference value or set point, $r(t)$, and the output measurement, $x(t)$, namely $e(t) = r(t) - x(t)$. λ is a tuning parameter that helps to define $S(t)$. This term is selected by the designer and determines the performance of the system on the sliding surface. n is the system order.

The control objective is to ensure that the controlled variable is equal to its reference value at all times, meaning that $e(t)$ and its derivatives must be zero. Once the reference value is reached, Equation 2.30(1) indicates that $S(t)$ has reached a constant value, meaning that $e(t)$ is zero at all times; it is desired to set

$$\frac{dS(t)}{dt} = 0 \qquad \textbf{2.30(2)}$$

Equation 2.30(2) is called the *sliding condition.*[1,2]

Second Step

As a second step, once the sliding surface has been selected, attention must be turned to designing the control law that drives the controlled variable to its reference value and satisfies Equation 2.30(2).

The SMC control law, $U(t)$, consists of two additive parts: a continuous part, $U_C(t)$, and a discontinuous part, $U_D(t)$. That is

$$U(t) = U_C(t) + U_D(t) \qquad \textbf{2.30(3)}$$

The continuous part is given by

$$U_C(t) = f(x(t), \quad r(t)) \qquad \textbf{2.30(4)}$$

where $f(x(t), r(t))$ is a function of the controlled variable and the reference value.

The continuous part of the controller is obtained by combining the process model and sliding condition, Equation 2.30(2). The discontinuous part is nonlinear and represents the switching element of the control law.

The discontinuous part of the controller is discontinuous across the sliding surface. Mainly, $U_D(t)$ is designed based on a relay-like function, i.e., $U_D(t) = K_D$ sign $S(t)$, because it allows for changes between the structures with a hypothetical infinitely fast speed. In practice, however, it is impossible to achieve the high switching control because of the presence of finite time delays for control computations or limitations of the physical actuators, thus causing chattering around of the sliding surface.[1–7]

The aggressiveness to reach the sliding surface depends on the control gain (i.e., K_D), but if the controller is too aggressive it can collaborate with the chattering. To reduce the chattering, one approach is to reattach the relay-like function by a saturation or sigma function,[7] which can be written as follows:

$$U_D(t) = K_D \frac{S(t)}{|S(t)| + \delta} \qquad \textbf{2.30(5)}$$

where K_D is the tuning parameter responsible for the reaching mode. δ is a tuning parameter used to reduce the chattering problem.

Figure 2.30a shows the effect of δ variations on the process output. Figure 2.30b shows the effect of K_D variations on the system trajectory on a phase plane, which illustrates how the rate of the trajectory moves to reach its final value from an initial state to a final state. For example, when $K_D = K_3$ is used, the error rate moves faster, reducing the error until the final value is reached, than when $K_D = K_1$ or K_2 are used to tune the controller. Therefore, K_D represents a gain, which provides the aggressivity of the controller. Figure 2.30c confirms the previous conclusion in the time domain.

In summary, the control law usually results in fast motion to bring the state onto the sliding surface, $U_D(t)$, and a slower motion to proceed until a desired state $U_C(t)$, is reached. See Figure 2.30d.

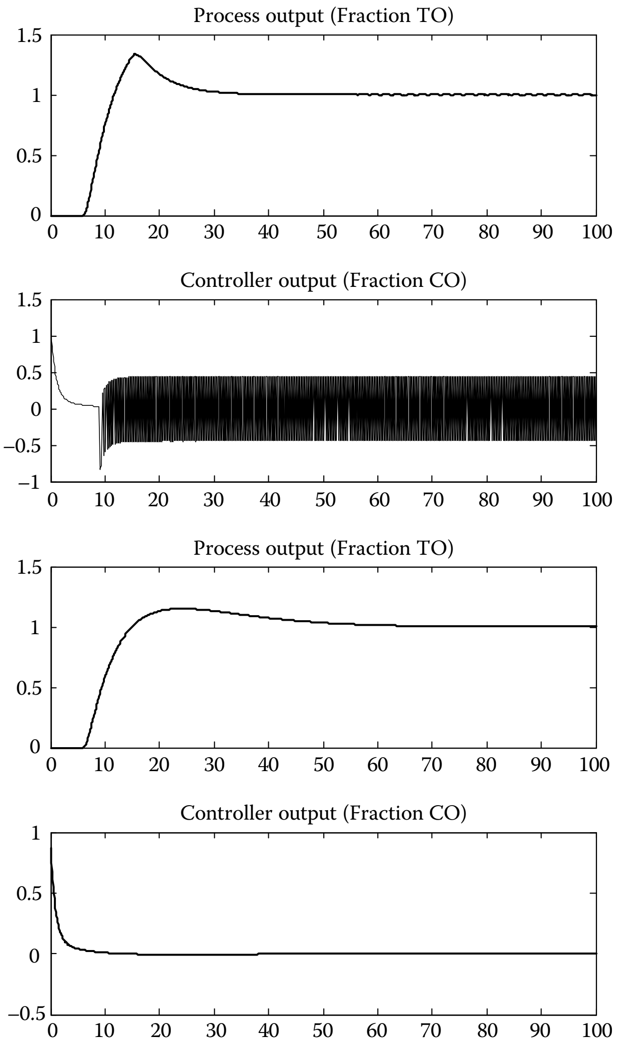


FIG. 2.30a
Chattering reduction using a saturation function. (a: $\delta = 0$, b: $\delta = 1.5$).

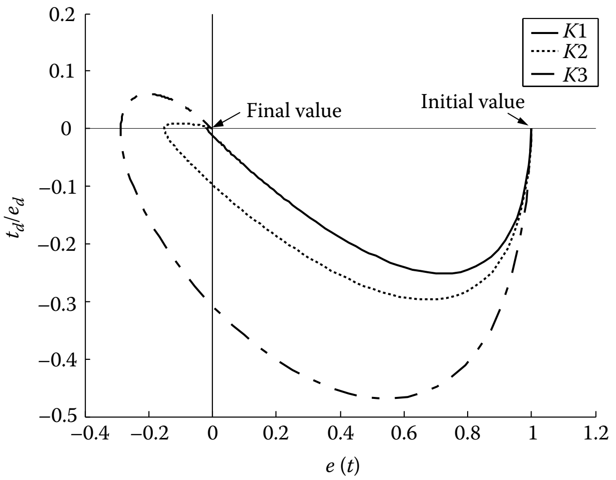


FIG. 2.30b
Graphical interpretation for K_D variations ($K_1 < K_2 < K_3$).

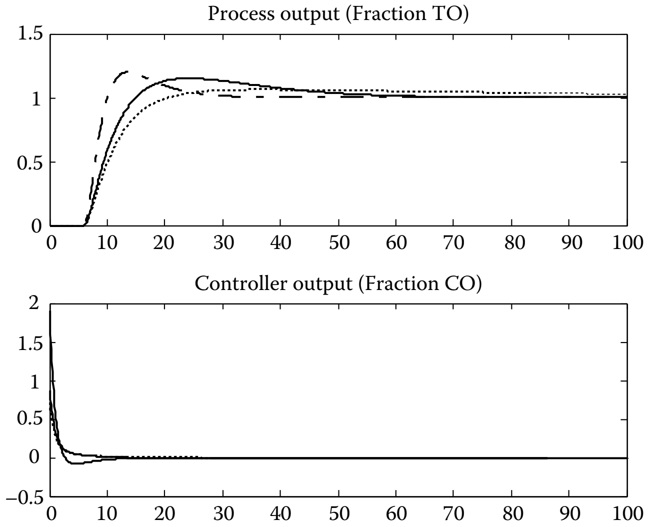


FIG. 2.30c
System responses to a set point change for different K_D values ----: $K_D = 0.25$. —: $K_D = 0.44$ —·—:· $K_D = 3.54$.

APPLICATIONS

Although a large number of reports talk about SMC applications, few process companies have reported their results publicly. Therefore, the information presented here will be limited to the contents of the reports found in the literature. Two types of applications will be described, first to electromechanical systems and second, the applications for chemical processes.

Electromechanical Systems

Sliding mode control has most been often used in the electrical control of machines, including the control of induction motors, speed control, servo systems applications, power systems, power supply switching, robotics, gas turbines, and turbojet engines, as well as in aircraft systems and anti-skid automotive systems. Good references for electromechanical system applications can be found in Utkin, Guldner, and Shi[8] and Edwards and Spurgeon.[9] A special issue of ASME[10] is also recommended for readers interested in applications in this control field.

Chemical Systems

Even though SMC has been widely investigated for a variety of processes, as mentioned earlier, to the best of our knowledge none of the previous works have resulted in a general approach for process industry. Since industrial chemical processes modeled using first principles tend to be of higher order and complexity, the traditional SMC procedures present disadvantages in their application.

An efficient alternative could be the use of empirical modeling methods. Empirical methods use low-order linear models; usually, first-order plus deadtime (FOPDT) models are adequate for process control analysis and design. This kind of model is able to adequately represent the dynamics of many industrial processes, especially chemical processes, over a range of frequencies,[11,12] and is easily obtained from the popular reaction curve (Section 2.22).

Although reduced order models cause uncertainties, which arise from the imperfect knowledge of the model, and nonlinear effects of the process contribute to performance degradation in the controllers, a sliding mode controller (SMCr) could be designed on the basis that the robustness of the controller will compensate for modeling errors that arise from the linearization of the nonlinear model of the process. Thus, if a general reduced-order model of the process can be obtained and chattering can be reduced, this control strategy can

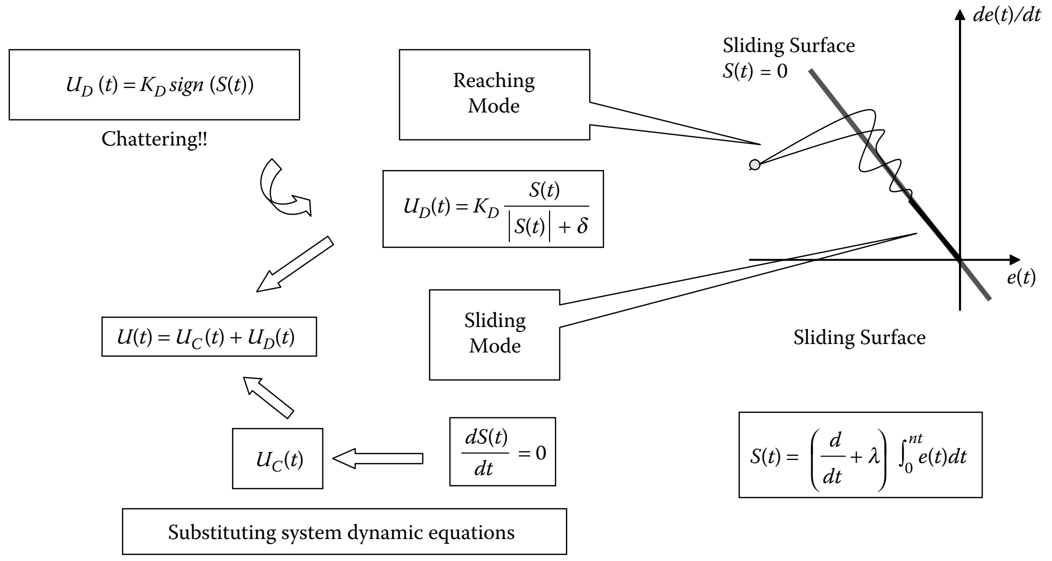


FIG. 2.30d
Sliding mode control components.

become one of the most important discoveries in the field of process control.

SMC for SISO Chemical Process Control

Here a general equation of the SMCr for SISO systems is presented. First, it is discussed in connection with self-regulating processes. After that, the subjects of inverse response, integrating, and a robust dead time compensation are covered.

Self-Regulating Processes SMCr is a type of model-based control (MBC) (Section 2.13), and the order of the model is proportional to the numbers of tuning parameters. The controller equation was derived by using an FOPDT process model, which is the recommended approach[11,12] (see Figure 2.30e). The controller obtained significantly simplifies the application of SMC to industrial processes having this type of dynamics.

Based on the references,[13,14] the complete SMCr is

$$U(t)=\left(\frac{t_0\tau}{K}\right)\left[\frac{X(t)}{t_0\tau}+\lambda_0 e(t)\right]+K_D\frac{S(t)}{|S(t)|+\delta} \qquad \textbf{2.30(6a)}$$

with

$$S(t)=\text{sign}(K)\left[-\frac{dX(t)}{dt}+\lambda_1 e(t)+\lambda_0\int_0^t e(t)dt\right] \qquad \textbf{2.30(6b)}$$

Figure 2.30f shows a schematic of the controller. These equations present advantages from the process control point of view. First, they have a fixed structure depending on the λ parameters and the characteristic parameters of the FOPDT model; second, the action of the controller is considered (in the sliding surface equation), by including the term sign(K). Note that sign(K) only depends on the static gain; therefore, it never switches. For industrial applications, Equation 2.30(6b) represents a PID algorithm.[15]

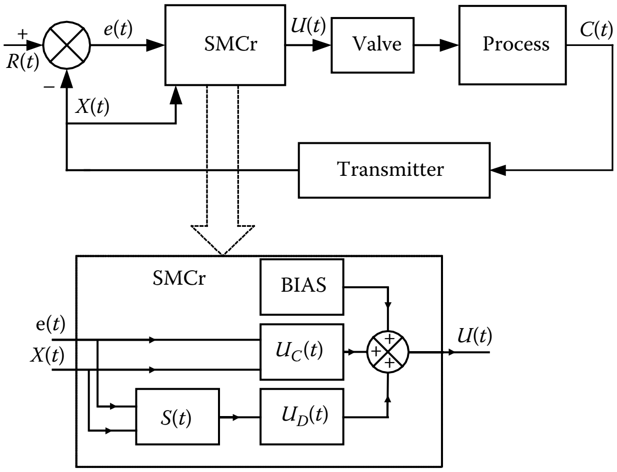


FIG. 2.30e
Schematic of a sliding mode controller.

To complete the controller, the following tuning equations can be considered:[13,14]

- For the continuous part of the controller and the sliding surface,

$$\lambda_1=\frac{t_0+\tau}{t_0\tau}\ [=]\ [\text{time}]^{-1} \qquad \textbf{2.30(7a)}$$

$$\lambda_0\le\frac{\lambda_1^2}{4}\ [=]\ [\text{time}]^{-2} \qquad \textbf{2.30(7b)}$$

- For the discontinuous part of the controller,

$$K_D=\frac{0.51}{|K|}\left[\frac{\tau}{t_0}\right]^{0.76}\ [=]\ [\text{fraction CO}] \qquad \textbf{2.30(7c)}$$

$$\delta=0.68+0.12(|K|\,K_D\lambda_1)\ [=]\ [\text{fraction TO/time}] \qquad \textbf{2.30(7d)}$$

Equations 2.30(7c) and 2.30(7d) are used when the signals from the transmitter and controller are in fractions (0 to 1). Sometimes, the control signals are scaled in percentages, so the signals are in 0-to-100% ranges. In these cases, the values of K_D and δ are multiplied by 100.

Inverse Response Systems When the initial direction of the process response is contrary to the direction of the final steady state, the process is called a nonminimum phase or inverse response process. This characteristic is only related to the input/output behavior of a process. Several chemical processes can have this kind of response, such as drum boilers, reactors, and reboilers.

This kind of system can be described by the following transfer function,

$$G_1(s)=\frac{K(1-\tau_1 s)}{(\tau_2 s+1)\cdot(\tau_3 s+1)\cdots(\tau_n s+1)} \qquad \textbf{2.30(8)}$$

It can be approximated by

$$G_1(s)\cong K\frac{e^{-t_o s}}{\tau s+1} \qquad \textbf{2.30(9)}$$

This means that the step response of the nonminimum phase system (Equation 2.30[8]) can be approximated by an

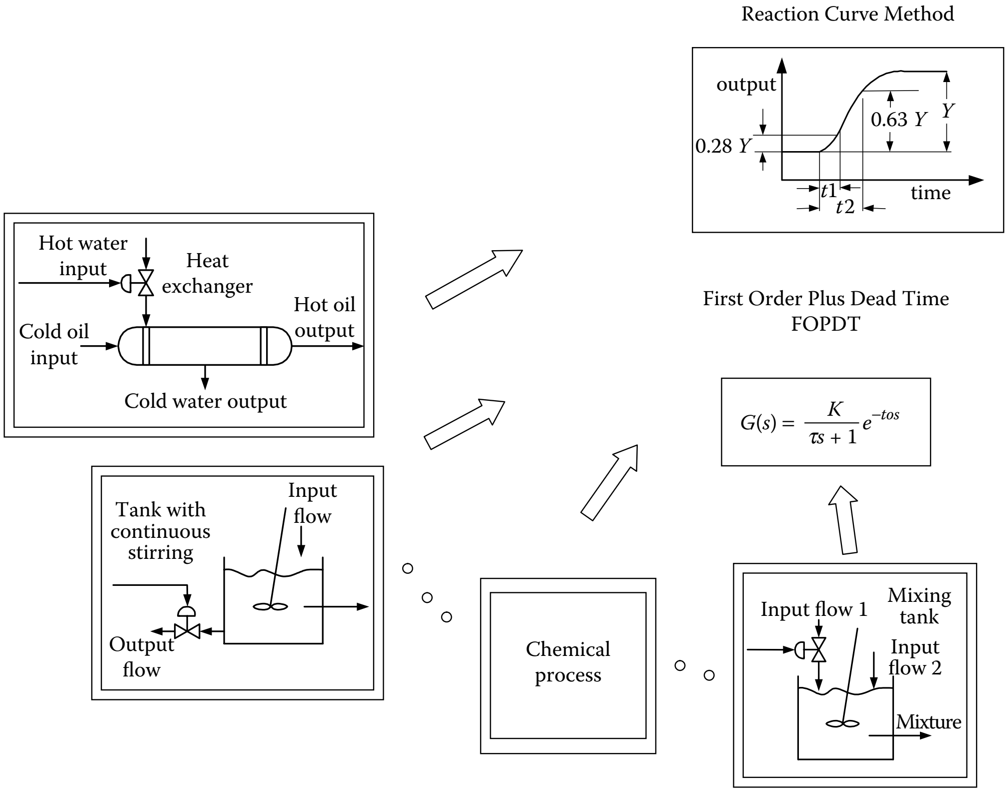


FIG. 2.30f
Empirical model to design Sliding mode controllers for chemical processes.

FOPDT model, such as the one in Equation 2.30(9) (see Section 2.22). Therefore, the design procedure for systems represented using FOPDT models can be applied.[14,16] The resulting SMCr is similar to the ones given by Equation 2.30(6a) and 2.30(6b). The equations for the first tuning estimates of the controller parameters were obtained[16] and are listed in Table 2.30k.

Integrating Systems In general, the following model can approximate integrating systems with dead time:

$$G(s) = K\frac{e^{-t_o' s}}{s} \qquad \textbf{2.30(10)}$$

The identification of the characteristic parameters of this model (K, τ, and t_o) is included in Section 2.22. Using Equation 2.30(10) and the procedure described previously, the controller obtained is[17]

$$U(t) = \frac{\lambda_o t_o'}{K} e(t) + K_D \frac{S(t)}{|S(t)| + \delta} \qquad \textbf{2.30(11)}$$

Equation 2.30(11) does not have the term that includes the deviation of the controlled variable; it suggests a modification, which is summarized in Figure 2.30g. Tuning settings for integrating processes are listed in Table 2.30k.

Robust Dead Time Compensator Dead time is present in many industrial processes, where it creates a serious control problem. The performance of feedback control systems will drop drastically as the ratio of dead time to time constant increases. The Smith predictor (SP—see Section 2.27) or dead time compensator (DTC), as it is also known, has many weak points, including possible instability and poor performance, if the model is inaccurate, and poor response to disturbances.[18] In addition, the original structure of the SP cannot reject constant load disturbance for processes with integration.

Camacho and Smith[14] showed the design of an SMCr from an FOPDT. The design of SMCrs for dead time processes

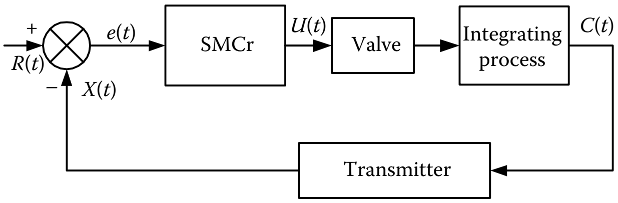


FIG. 2.30g
Modified sliding mode controller for integrating systems.

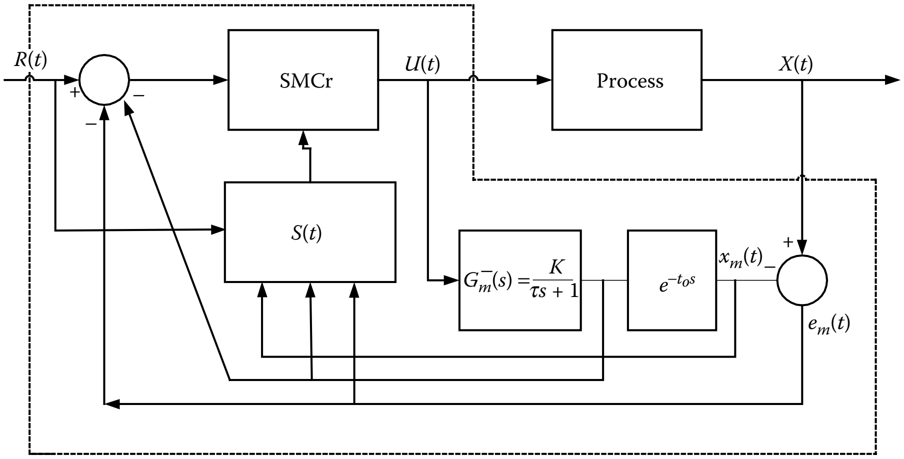


FIG. 2.30h
Robust dead time compensator scheme.

requires some assumptions,[13,14] which can be solved using the internal model approach. It was shown that by merging these two control techniques, a robust one (SMC) and a predictive one (IMC), the resulting controller will provide the benefits of both.[19] Thus, the mixing of SMC and predictive structures can produce better controllers than the original ones.[20]

The robust dead time compensator is designed in two steps: First, an invertible model of the process is chosen, and from it, the sliding surface is designed. Second, the reaching condition is satisfied, and the robustness is guaranteed. Figure 2.30h shows the proposed scheme. The nonlinear process has been modeled as an FOPDT. This model can be separated into two parts:

$$G_m^+ = e^{-t_o s} \qquad \textbf{2.30(12)}$$

$$G_m^- = \frac{K}{\tau s + 1} \qquad \textbf{2.30(13)}$$

Since $G_m^-(s)$ eliminates the dead time term from the model, this simplification facilitates the SMC design.

The controller obtained is as follows:

$$S(t) = \text{sign}(k)\left[e_m^-(t) + \lambda\int_0^t e(t)dt\right] \qquad \textbf{2.30(14)}$$

$$u(t) = \frac{\tau}{K}\left[\frac{x_m^-(t)}{\tau} + \lambda\underbrace{[R(t) - x_m(t) - e_m(t)]}_{e(t)}\right] + K_D\frac{s}{|s| + \delta} \qquad \textbf{2.30(15)}$$

$e_m^-(t)$ is the error between the reference, $R(t)$, and the model output without deadtime, $x_m^-(t)$, $x(t)$ is the controlled variable, and $e_m(t)$ is the error between the process output and the complete model output, also known as modeling error.

It is observed that this sliding surface includes a prediction term, $e_m^-(t)$, which gives a faster response to reach the set point than the original one, when processes present an elevated controllability relationship (t_0/τ). The tuning parameters for this controller are included in Table 2.30k.

SMC for MIMO Chemical Process Control

In this section, a summary is provided of the SMCr modification that should be applied when the chemical processes are multivariable[10] (see Section 2.18). The objective is to extend the applicability of the SMCrs, which were designed for SISO processes, to multivariable systems.

Industrial processes are by nature multivariable. Control of multivariable and nonlinear processes is not an easy task because of the interaction among variables (see Section 2.12) and because of the presence of dead time. The interaction results from the high probability that a manipulated variable can affect several controlled variables. This interaction among the variables degrades control quality. The proposed approach discussed below assumes that the total control system consists of multiple SMCrs acting in addition to a decoupling system (Section 2.12).

The amount of interaction between the variables can be determined by the steady-state gains K_{ij}, using the relative gain array (Section 2.25) or Bristol Array method,[11,12] which allows obtaining a dimensionless interaction index μ_{ij}. In systems with strong interaction, a common practice is to use decoupling of the interacting loops.[11,12]

Once the system is decoupled, it is possible to use a developed SMCr based on an FOPDT model. The original SMCr controller has a set of tuning equations for SISO-type processes. The idea in this part is to consider the effect of the multivariable characteristic parameters and the interaction index in the controller tuning equations.

Using the concepts of decoupling, a multivariable control system can be represented as:

$$\begin{bmatrix} C_1(s) \\ C_2(s) \\ \dots \\ C_n(s) \end{bmatrix} = \begin{bmatrix} G'_{P11}(s) & 0 & \dots & 0 \\ 0 & G'_{P22}(s) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & G'_{Pnn}(s) \end{bmatrix} \begin{bmatrix} M_1(s) \\ M_2(s) \\ \\ M_N(s) \end{bmatrix} \qquad \mathbf{2.30(16)}$$

where $C_i(s)$ represents the i-controlled variable; $M_j(s)$ is the j-manipulated variable; and $G_{pij(s)}$ is an FOPDT transfer function relating the i-controlled and j-manipulated variables, respectively. Once the system is decoupled, it is possible to use SMCr based on an FOPDT model.[13]

The parameters t_{ojj}, τ_{jj}, and K_{jj} needed to calculate the initial tuning settings of the controller are obtained from the open-loop step responses (see Section 2.22), for the *j-j pairing* of the manipulated and controlled variable pairs after decoupling. In comparison with the original SMCr tuning values, only the K_{Dj} tuning parameter has changed to include the interaction index. This provides the necessary aggressiveness to reach the sliding surface in multivariable applications (Table 2.30k).

Discussion of the Results Due to space limitations, detailed results are not included here, but they can be seen for the minimum-phase case in Camacho and Smith,[14] for the inverse response case in Reference 16 and for the integrating systems case in Reference 17.

A new control scheme, mixing the internal model approach with the SMC concept for processes with high dead time–to–time constant ratios was tested in Reference 19. Finally, the multivariable systems case can be seen in the ASME special issue.[10]

Summarizing the results, the same SMCr designed for minimum phase systems can be used with good results for a wide class of nonlinear chemical processes. These SMCrs gave robust performance on processes with dead time and in applications with modeling errors and disturbances.

SMCR IMPLEMENTATION OF PID ALGORITHM

Digital PID algorithms are widely used in distributed control systems (DCS), in programmable logic controllers (PLC), and in remote terminal unit (RTU) applications. A common PID algorithm is as follows:

$$\mathrm{MV}(t) = Kc\left[e(t) + \frac{1}{\tau_I}\int_0^t e(t)dt - \tau_D\frac{d\mathrm{CV}(t)}{dt}\right] \qquad \mathbf{2.30(17)}$$

where MV(t) is the manipulated variable, τ_I is the integral time, τ_D is the derivative term, and CV(t) is the controlled variable. Note that the derivative term is calculated using the measured controlled variable, not the error.

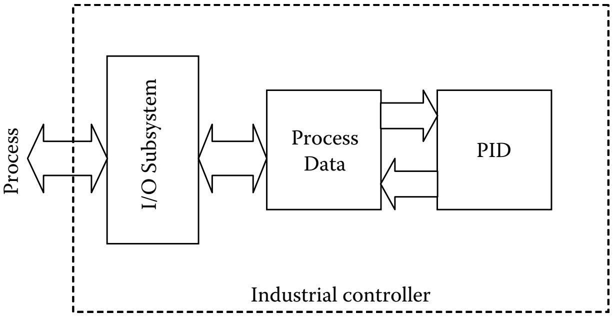


FIG. 2.30i
Implementation of PID algorithm in an industrial controller.

In this respect, Equations 2.30(6b) and 2.30(17) are similar. The discussion here will focus on implementing the SMCr based on the PID algorithms,[15] using the presently existing control devices and plant operator mentality. The PLC-based implementation of a discrete algorithm for a PID is shown in Figure 2.30f.

The term *X(t)* in Equation 2.30(6b) is the same as the term CV*(t)* in 2.30(17). The next step is to tune the PID based on the SMCr tuning equations. From Equation 2.30(6b) and Equation 2.30(17), the tuning parameters of the PID algorithm can be converted into the tuning parameters of the sliding surface $S(t)$, as follows:

$$Kc = \lambda_1 \qquad \mathbf{2.30(18)}$$

$$\tau_I = \lambda_1/\lambda_O \qquad \mathbf{2.30(19)}$$

$$\tau_D = (\lambda_1)^{-1} \qquad \mathbf{2.30(20)}$$

With the above conversions, the PID algorithm can be expressed in terms of the sliding surface term $S(t)$. The SMCr is divided in two parts—the continuous part, which is an algebraic function of the process variables and the set point, and the discontinuous part, which is also an algebraic function of $S(t)$, which is calculated from the PID algorithm that exists on the PLCs or RTUs used in industry.

Figures 2.30i and 2.30j show the implementation of this approach, where the output of the PID, which was an input

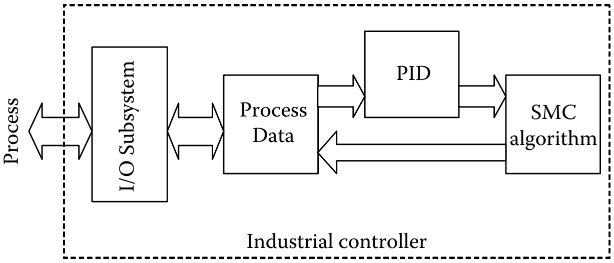


FIG. 2.30j
Implementation of SMCr using an industrial PID controller.

TABLE 2.30k
Summary of the Tuning Equations Used for SMCrs

	Self-Regulating	*Inverse Response*	*Integrating*	*Dead-Time Compensator*	*Multivariable*
λ_1	$\frac{t_0+\tau}{t_0\tau}$	$\frac{t_0+\tau}{t_0\tau}$	$\frac{1}{t_0}$	$\lambda \le \frac{1}{\tau+t_0}$	$\frac{t_{ojj}+\tau_{jj}}{t_{0jj}\ \tau_{jj}}$
λ_0	$\le \frac{\lambda_1^2}{4}$	$\le \frac{\lambda_1^2}{8}$	$\left(\frac{\lambda_1}{4}\right)^2$	—	$\le \frac{\lambda_{ij}^2}{4}$
K_D	$\frac{0.51}{\lvert K\rvert}\left[\frac{\tau_j}{t_{oj}}\right]^{0,76}$	$\frac{0.064}{K}\left[\frac{\tau}{t_0}\right]^{0.76}$	$\frac{0.64}{K}\left[\frac{1}{t_0}\right]^{0.76}$	$K_D \ge \frac{0.8}{\lvert K\rvert}\left(\frac{\tau}{t_0}\right)^{0.76}$	$\mu_{ij}\ \frac{0.51}{\lvert K_{jj}\rvert}\left[\frac{\tau_{jj}}{t_{0jj}}\right]^{0.76}$
δ	$0.68+0.12(KK_D\lambda_1)$	$0.68+0.12\ (KK_D\lambda_1)$	$0.68+0.12(KK_D\lambda_1)$	$0.68+0.12(KK_D\lambda)$	$0.68+0.12(K_{ij}K_{Dij}\lambda_{jj})$

to the process (manipulated variable), became an input to the discontinuous part of the SMCr.

SMCr Implementation Methodology

Three steps are required to implement the total algorithm. In each iteration, the term $S(t)$ is calculated from the PID output. The continuous and discontinuous parts of the controller algorithm are easily programmable algebraic equations. This procedure is based on the PID algorithm and already exists in most industrial controllers.

The sequence of statements summarizing the proposed algorithm is:

```
a1 = t0*tau/K
do forever
  input x
  et = x - ref
  y  = PID(et)
  st =    sign(k)*y
  ut = (x/K + a1*lambda0*et) +
          Kd*(st/(abs(st)+delta)
End
```

Table 2.30k provides a summary of the tuning equations for the SMCrs. The parameters (t_0, τ, and K) that are required to calculate the initial tuning settings of the controller are obtained from the open-loop step response (Section 2.22).

CONCLUSIONS

In the past, SMC has been an important theoretical research field. Today, it is becoming a good source of solutions to real-world problems. Its potential is limited only by the imagination of the people working in process control. Therefore, the prospects for the application for SMC in the processing industries are immense.

This section has described SMC and provided some examples of industrial applications in electromechanical systems and a practical approach for using a reduced-order model in chemical process applications. SISO and MIMO systems were also discussed.

The SMCr can be implemented from a PID algorithm, where the PID represents the sliding surface, and the rest of the controller is built using algebraic blocks. This represents an advantage from a process control engineer's point of view because the controller is of a fixed structure. This controller, having adjustable parameters, can be easily implemented in DCS.

Finally, the approach proposed for chemical industry applications has the simplicity of PID controllers and the robustness of SMCrs, and the procedure used can be easily extended to electrical and mechanical systems if reduced-order models can be obtained for those processes.

References

1. Utkin, V. I., "Variable Structure Systems with Sliding Modes," *Transactions of IEEE on Automatic Control,* AC–22, pp. 212–222, 1977.
2. Hung, J. Y., Gao, W., and Hung, J. C., "Variable Structure Control: A Survey," *IEEE Transactions on Industrial Electronics,* 40(1), pp. 2–21, 1993.
3. Sira-Ramirez, H., and Llanes-Santiago, O., "Dynamical Discontinuous Feedback Strategies in the Regulation of Nonlinear Chemical Processes," *IEEE Transactions on Control Systems Technology,* 2(1), pp. 11–21, 1994.
4. Colantino, M. C., Desages, A. C., Romagnoli, J. A., and Palazoglu, A., "Nonlinear Control of a CSTR: Disturbance Rejection Using Sliding Mode Control," *Industrial & Engineering Chemistry Research,* 34, pp. 2383–2392, 1995.
5. Young, K. D., Utkin V. I., and Özgumer, Ü., "A Control Engineer's Guide to Sliding Mode Control," *IEEE Transactions on Control Systems Technology,* 7(3), pp. 328–342, 1999.
6. Slotine, J. J., and Li, W., *Applied Nonlinear Control,* Englewood Cliffs, NJ: Prentice Hall, 1991.
7. Zinober, A. S. I., *Variable Structure and Liapunov Control,* London: Springer-Verlag, 1994.

8. Utkin, V., Guldner, J., and Shi, J. X., *Sliding Mode Control in Electro-Mechanical Systems,* London: Taylor and Frencis, 1999.
9. Edwards, C., and Spurgeon, S., *Sliding Mode Control: Theory and Applications,* London: Taylor and Francis, 1999.
10. Camacho, O., and Rojas, R., "A General Approach of Sliding Mode Control for Nonlinear Chemical Processes," *Transactions of the ASME Journal of Dynamic Systems, Measurements, and Control,* 122(4), pp. 650–656, 2000.
11. Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1997.
12. Marlin, T. E., *Process Control,* New York: McGraw-Hill, 1995.
13. Camacho, O., "A New Approach to Design and Tune Sliding Mode Controllers for Chemical Processes," Ph.D. Dissertation, University of South Florida, Tampa, 1996.
14. Camacho, O., and Smith, C., "Sliding Mode Control: An Approach to Regulate Nonlinear Chemical Processes," *ISA Transactions,* 39(2), pp. 205–218, 2000.
15. Camacho, O., Smith, C., and Chacón, E., "Toward an Implementation of Sliding Mode Control to Chemical Processes," *Proceedings of ISIE '97,* Guimaraes-Portugal, pp. 1101–1105, 1997.
16. Camacho, O., Rojas, R., and Garcia, W., "Variable Structure Control Applied to Chemical Processes with Inverse Response," *ISA Transactions,* 38, pp. 55–72, 1999.
17. Camacho, O., García, W., and Rojas, R., "Sliding Mode Control: A Robust Approach to Integrating Systems with Deadtime," *Proceedings of IEEE-ICCDCS,* Margarita, Venezuela, 1998.
18. Sunan, H., Kiong, T. K., and Heng, L. T., *Applied Predictive Control,* London: Springer-Verlag, 2002.
19. Camacho, O., Smith, C., and Moreno, W., "Development of an Internal Model Sliding Mode Controller," *Industrial & Engineering Chemistry Research,* 42(3), pp. 568–573, 2003.
20. Perez De La Parte, M., Camacho, O., and Camacho, E. F., "Development of a GPC–Sliding Mode Controller," *ISA Transactions,* 41(1), 2002.

Bibliography

Utkin, V., *Sliding Modes and Their Applications in Variable Structure Systems,* Moscow: Mir Publishing, 1978 (translation of a book published by Nauka, Moscow, 1974 in Russian).

Itkis, U., *Control Systems of Variable Structure,* New York: Wiley, 1976.

Filippov, A., *Differential Equations with Discontinuous Right-Hand Sides,* Dordrecht: Kluwer, 1988.

Zinober, A. S. (Ed.), *Deterministic Non-Linear Control,* U.K.: Peter Peregrinus Limited, 1990.

Young, K.-K. D. (Ed.), *Variable Structure Control for Robotics and Aerospace Application,* Amsterdam: Elsevier, 1993.

Utkin, V., *Sliding Modes in Control and Optimization,* Berlin: Springer-Verlag, 1992.

Young, K. D., and Ozguner, U. (Eds.), *Variable Structure Systems, Sliding Mode and Nonlinear Control,* Heidelberg: Springer-Verlag, 1999.

2.31 Software for Fuzzy Logic Control

D. H. F. LIU (1995) **J. ABONYI** (2005)

In 1965, Professor L. A. Zadeh of the University of California, Berkeley, presented a paper outlining fuzzy theory. In about 1970, fuzzy theory began to produce results in Japan, Europe, and China. Fuzzy control has found applications in cement kinks, papermaking machines, and polymerization reactors. Nowadays, fuzzy controllers are also used to control household appliances, such as dishwashers, washing machines, video cameras, and rice cookers, even though this fact is not always advertised. Among the main characteristics of these processes are their time-varying, nonlinear behavior and the low frequency of their measurements.

Fuzzy logic is a very human concept, potentially applicable to a wide range of processes and tasks that require human intuition and experience. Fuzzy logic control can be applied by means of software, dedicated controllers, or fuzzy microprocessors emdebbed in digital products. Application flexibility combined with inherent simplicity and wide ranges of capabilities give fuzzy logic technology great potential for growth.

INTRODUCTION

Today's manufacturing processes present many challenging control problems; among these are nonlinear dynamic behavior, uncertain and time-varying parameters, and unmeasured disturbances. In the past decade, the control of these systems has received considerable attention in both academia and industry. While advanced control strategies — coupled with cheaper and powerful computer systems — have made nonlinear process control much more practical, there is still a considerable gap between control theory and industrial practice. It is frustrating for the control theory community that elegant and comprehensive frameworks for system analysis and design are seldom applied in the process industry, which still applies the well-known PID controller (90% of control loops) and relies on manual control in complex process situations.

Conventional control theory may fall short if the model of the process is difficult to obtain, (partly) unknown, or highly nonlinear. The design of controllers for seemingly easy everyday tasks such as driving a car or grasping a fragile object continues to be a challenge for robotics, while human beings easily perform these tasks. Similarly, some industrial processes are controlled in one of the following ways:

- The process is controlled manually by experienced operators.
- The process is controlled by an automatic control system that needs manual online "trimming."

Many industrial processes operated by humans cannot be automated using conventional control techniques since the performance of these controllers is often inferior to that of the operators. One of the reasons is that linear controllers, which are commonly used in conventional control, are not appropriate for the control of nonlinear processes.

Another reason is that humans aggregate various kinds of information and combine control strategies, which cannot be integrated into a single analytic control law. The underlying principle of knowledge-based (expert) control is to capture and implement experience and knowledge available from experts (e.g., process operators).

A specific type of knowledge-based control is fuzzy rule-based control, where the control actions corresponding to particular conditions of the system are described in terms of fuzzy if–then rules.

To investigate the current application issues of advanced control techniques, a survey was made in Japan in 1995.[1] Information was obtained about the number and the evaluation of control applications, key factors for successful and failed implementations, etc. The investigated control techniques were classified as follows:

- *Advanced PID:* I-PID and two-degrees-of-freedom PID, decoupling PID, dead time compensation, gain scheduling, and auto tuning control
- *Modern Control Theory:* linear quadratic Gaussian (LQG) regulator, observer, Kalman filter, model predictive control (MPC), adaptive control, H-infinity control, repetitive control, exact linearization control, optimization control
- *FAN (Fuzzy, Artificial intelligence, Neural network):* Fuzzy control, rule-based control, neural network control

The survey showed that advanced PID-type control is widely applied, and about 30% of the 110 respondents have already used this type of control in their factories. MPC and fuzzy control are the most widely used of the modern control and FAN techniques. Slightly less than 40% of the factories have applied these solutions. Compared to a survey in 1989, the applications of decoupling PID, dead time compensation, Kalman filter, model predictive control, rule-based control, fuzzy control, and optimization have increased substantially.

Although modern control theory is rarely applied, except model-predictive control, 60 to 70% of the respondents are satisfied with the results, and the satisfaction with FAN is going up. These tendencies indicate that a huge demand exists in the industry for new fuzzy control solutions.

The objective of this section is to present the principle of fuzzy modeling and control, identify and explain the various design choices, and provide an overview of the implementation tools for engineers.

PRINCIPLE OF FUZZY SYSTEMS

Fuzzy control is a control method based on fuzzy logic. Fuzzy logic is not really "fuzzy." Just as fuzzy logic can be described simply as "computing with words rather than numbers," fuzzy control can be described simply as "control with sentences rather than equations."

For many real-world applications, a great deal of information is provided by human experts, who do not reason in terms of mathematics but instead describe the system verbally through vague or imprecise statements such as

IF The Temperature is **Up THEN** The Pressure is **High.**

Because so much human knowledge and expertise come in terms of such verbal rules, one sound engineering approach is to try to integrate such linguistic information into the modeling process. A convenient and common approach is to use fuzzy logic concepts to cast the verbal knowledge into a conventional mathematical representation.

Fuzzy logic facilitates the computerized representation of this kind of knowledge, which subsequently can be fine-tuned using process experiments (e.g., based on input/output process data). From this basis, the fuzzy system is a computation framework based on the concepts of fuzzy sets, fuzzy if–then rules, and fuzzy reasoning.

This section will introduce the reader to the structure of fuzzy models. It will not attempt to provide a broad survey of the field. For such a survey the reader is referred to *An Introduction to Fuzzy Control,* by Driankov, Helerndoorn, and Reinfrank,[2] *Fuzzy Control,* by Passino and Yurkovic,[3] or *A Course in Fuzzy Systems and Control*, by Wang.[4]

Fuzzy Sets

Conventional set theory is based on the premise that an element either belongs to or does not belong to a given set. Fuzzy set theory takes a less rigid view and allows elements to have degrees of membership of a particular set such that elements are not restricted to either being in or out of a set but are allowed to be "somewhat" in. In many cases this is a more natural approach.

For example, consider the case of a person describing the atmospheric temperature as being "hot." If one was to express this concept in conventional set theory one would be forced to designate a distinct range of temperatures, such as 25°C (77°F) and over, as belonging to the set "hot." That is, *hot* = [25;∞)°C. This seems contrived because any temperature that falls just slightly outside this range would not be a member of the set, even though a human being may not be able to distinguish between it and one that is just inside the set. In fuzzy set theory, a precise representation of imprecise knowledge is not enforced since strict limits of a set are not required to be defined; instead, a *membership function* is defined.

A membership function describes the relationship between a variable and the degree of membership of the fuzzy set that corresponds to particular values of that variable. This degree of membership is usually defined in terms of a number between 0 and 1, inclusive, where 0 implies total absence of membership, 1 implies complete membership, and any value in between implies partial membership of the fuzzy set. This may be written as follows: $A(x) = [0;\, 1]$ for $x \in U$, where $A(x)$ is the membership function and U is the universe of discourse that defines the total range of interest over which the variable x should be defined.

For example, to define membership of the fuzzy set "hot," a function that rises from 0 to 1 over the range 15 to 25°C (59 to 77°F) may be used, i.e.,

$$A(x) = \begin{cases} 0 & x < 15°\text{C} \\ \dfrac{x-15}{10} & 15 \le x \le 25° . \\ 1 & x > 25°\text{C} \end{cases} \qquad \textbf{2.31(1)}$$

This implies that 15°C (59°F) is not hot; 20°C (68°F) is a bit hot; 23°C (73.4°F) is quite hot; and 30°C (86°F) is truly hot. Specific measurable values such as 15 and 20 are often referred to as crisp values or fuzzy singletons, to distinguish them from fuzzy values, such as *hot,* which are defined by a fuzzy set. Fuzzy values are sometimes also called linguistic values. As Figure 2.31a illustrates, this definition is more reflective of human or linguistic interpretations of temperatures and hence better approximates such concepts.

While seeming imprecise to a human being, fuzzy sets are mathematically precise in that they can be fully represented by exact numbers. They can therefore be seen as a method of tying together human and machine knowledge representations. Given that such a natural method of representing information in a computer exists, information processing methods can be applied to it with the use of fuzzy systems.

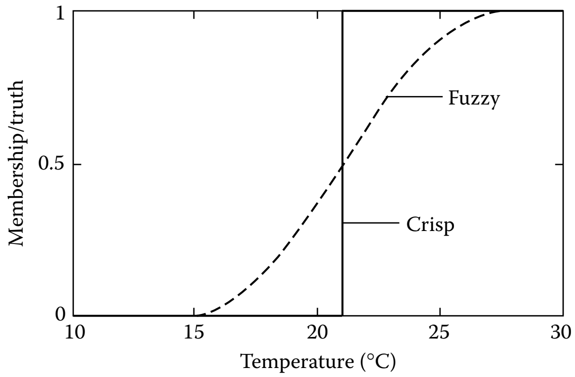


FIG. 2.31a
Representation of high temperature in terms of a "fuzzy set."

Fuzzy Systems

A fuzzy controller has a set of rules that it uses to calculate the final control action. Each rule is a linguistic expression describing the control action to be taken in response to a given set of process conditions. The rules are in the familiar if–then format, and formally the if side is called the *condition* and the then side is called the *conclusion* (more often, perhaps, the pair is called *antecedent–consequent* or *premise–conclusion*). IF (CONDITION) THEN ACTION.

Take for instance a typical fuzzy controller with two rules:

1. **IF** *error* is ***Neg*** **and** *change in error* is ***Neg*** then output is ***NB***
2. **IF** *error* is ***Neg*** **and** *change in error* is ***Zero*** then output is ***NM***

The input value "Neg" is a *linguistic term,* which is short for *Negative*; the output value "NB" stands for *Negative Big* and "NM" for *Negative Medium.* The computer is able to execute the rules and compute a control signal that depends on the measured inputs *error* and *change in error.*

The collection of such rules is called a *rule base.* The basic idea of fuzzy logic control is to formulate sets of rules for automatic operation based on practical experience and knowledge about manual control of the process. In fuzzy logic, the condition of a rule is fulfilled to a certain degree, and each rule will influence the result of the set of rules in accordance with its grade of fulfillment. To illustrate this procedure, the basic configuration of a fuzzy system is shown in Figure 2.31b.

As is depicted in this figure, a fuzzy system involves the following components:[5]

Data preprocessing processes the controller's inputs in order to obtain the inputs of the static fuzzy system. It will typically perform some of the following operations on the input signals:

- *Signal scaling.* The physical values of the input of the fuzzy system may differ significantly in magnitude. By mapping these to proper normalized (but interpretable) domains via scaling, one can instead work with signals of roughly the same magnitude. This is accomplished by normalization gains that scale the input into the normalized domain [−1, 1]. Values that fall outside the normalized domain are mapped onto the appropriate endpoint.

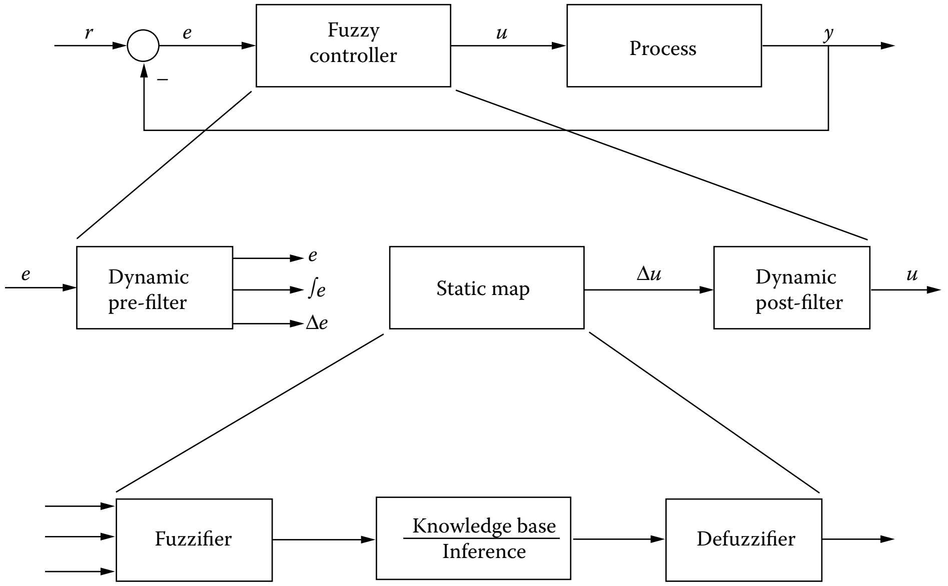


FIG. 2.31b
Fuzzy controller in a closed-loop configuration (top panel) consists of dynamic filters and a static map (middle panel). The static map is formed by the knowledge base, inference mechanism, and fuzzification and defuzzification interfaces.

- *Dynamic filtering.* In a fuzzy PID controller, for instance, linear filters are used to obtain the derivative and the integral of the control error, *e*. Nonlinear filters are found in nonlinear observers and in adaptive fuzzy control where they are used to obtain the fuzzy system parameter estimates.
- *Feature extraction.* Through the extraction of different features, numeric transformations of the controller inputs are performed. These transformations may be Fourier or wavelet transforms, coordinate transformations, or other basic operations performed on the fuzzy controller inputs.

Fuzzification maps the crisp values of the preprocessed input of the model into suitable fuzzy sets represented by membership functions.

The *rule base* is the cornerstone of the fuzzy model. The expert knowledge, which is assumed to be given as a number of if–then rules, is stored in a fuzzy rule base. According to the consequent proposition, there are three distinct classes of fuzzy models:

- *Fuzzy linguistic models* (*Mamdani models*), where both the antecedent and consequent are fuzzy propositions. Hence, a general rule of a linguistic or Mamdani fuzzy model is given by R_j: **If** x_1 is $A_{1,j}$ **and … and** x_n is $A_{n,j}$ **then** y is B_j, where R_j denotes the j-th rule, $j = 1,\ldots, N_r$, and N_r is the number of the rules. The antecedent variables represent the input of the fuzzy system, x_i. $A_{i,j}$ and B_j are fuzzy sets described by membership functions.
- *Fuzzy relational models,* which are based on fuzzy relations and relational equations.[6] These models can be considered as a generalization of the linguistic model, allowing one particular antecedent proposition to be associated with several different consequent propositions via a fuzzy relation.
- *Takagi–Sugeno* (*TS*) *fuzzy models,* where the consequent is a crisp function of the input variables, $f_j(\mathbf{x})$, rather than a fuzzy proposition.[7]

$$R_j\text{: }\mathbf{If}\quad x_1 \text{ is } A_{1,j} \quad \mathbf{and}\ldots\mathbf{and} \quad x_n \text{ is } A_{n,j}$$
$$\mathbf{then} \quad y = f_i(x) \qquad \mathbf{2.31(2)}$$

Mamdani (linguistic) fuzzy models with either fuzzy or singleton consequents are usually used as direct closed-loop controllers, while Takagi–Sugeno (TS) fuzzy systems are typically used as a supervisory controllers of fuzzy models of dynamical processes.

Inference engine. The inference mechanism or inference engine is the computational method that calculates the degree to which each rule fires for a given fuzzified input pattern by considering the rule and label sets. A rule is said to *fire* when the conditions upon which it depends occur. Since these conditions are defined by fuzzy sets which have degrees of membership, a rule will have a degree of firing or *firing strength.* The firing strength is determined by the mechanism that is used to implement the *and* in the expression. There are different methods for implementing each of the logical operators, and the reader is referred to Reference 2 for the details on these.

As can be seen in the example in Figure 2.31c, the logical product (AND connection) of a group of grades is simply the minimum value of all of the grades. Thus, in rule 5, the logical product of 1 for the control error and 0.5 for the change of the control error is 0.5. This procedure is repeated for each of the rules, as shown in Figure 2.31c. The logical sum combines the results of the rules. This is shown in the bottom of the corner of Figure 2.31c.

Defuzzification. A defuzzifier compiles the information provided by each of the rules and makes a decision from this basis. In linguistic fuzzy models the defuzzification converts the resulted fuzzy sets defined by the inference engine to the output of the model to a standard crisp signal. The centroid

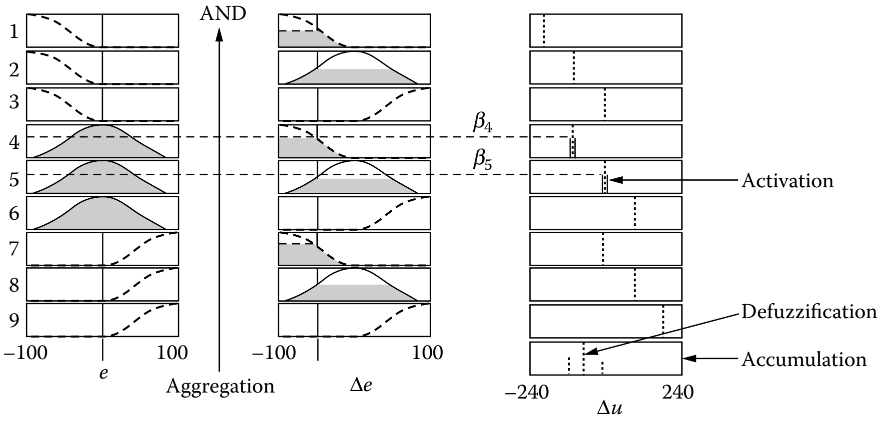


FIG. 2.31c
Example for a fuzzy inference.

method of defuzzification takes a weighted sum of the designated consequences of the rules according to the firing strengths of the rules. In the example shown in Figure 2.31c, which is the special case of Takagi–Sugeno fuzzy models, this type of defuzzification results in the following simple form:

$$y = \sum_{j=1}^{N_r} \beta_j(\mathbf{x}) f_j(\mathbf{x}) \tag{2.31(3)}$$

where $\beta_j(\mathbf{x})$ is the normalized degree of fulfillment (firing strength) of the rules:

$$\beta_j(\mathbf{x}) = \frac{\prod_{j=1}^{n} A_{i,j}(x_i)}{\sum_{i_1=1}^{N_r}\prod_{j=1}^{n} A_{i,j}(x_i)}. \tag{2.31(4)}$$

Postprocessing. The preprocessing step gives the output of the fuzzy system based on the crisp signal obtained after defuzzification. This often means the scaling of the output. Operations this post-filter may perform include:

- *Signal scaling.* A denormalization gain can be used, which scales the output of the fuzzy system to the physical domain of the actuator signal.
- *Dynamic filtering.* In some cases, the output of the fuzzy system is the increment of the control action. The actual control signal is then obtained by integrating the control increments. Of course, other forms of smoothing devices and even nonlinear filters may be considered.

EXAMPLE: FUZZY PI CONTROLLER

The previously presented decomposition of a controller to a static map and dynamic filters can be done for most classical control structures. To see this, consider a proportional–integral (PI) controller,

$$u(t) = K\left(e(t) + \frac{1}{T}\int e(t)dt\right) + \text{bias}$$

which can be expressed in discrete form by a simple difference equation: $\Delta u(k) = q_1 e(k) + q_2 \Delta e(k)$, where $\Delta e(k) = e(k) - e(k-1)$ and q_1, q_2 are the parameters of the controller that define a hyperplane shown in Figure 2.31d.

This linear form can be generalized to a nonlinear function: $\Delta u(k) = f(\mathbf{x})$, where the inputs of this static mapping are $x_1 = e(k)$ and $x_2 = \Delta e(k)$. In a fuzzy controller this nonlinear function f is represented by a fuzzy mapping.

Since the rule base represents a static mapping between the antecedent and the consequent variables, external dynamic filters must be used to obtain the desired dynamic behavior of the controller (Figure 2.31b). Furthermore, as the PI controller is defined in incremental form, signal processing is required both before and after the fuzzy mapping to obtain the control signal: $u(k) = u(k-1) + \Delta u(k)$.

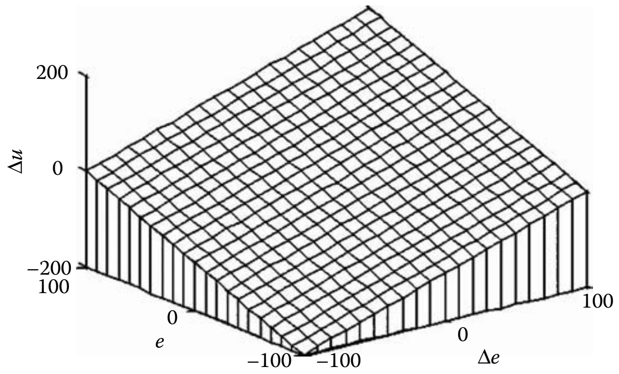


FIG. 2.31d
Control surface of a linear PI controller.

The set of rules of the PI fuzzy controller can be expressed in the following form:

$$\begin{aligned} R_j: &\text{IF} \quad e(k) \text{ is } A_{1,j} \quad \text{and} \quad \Delta e(k) \text{ is } A_{2,j} \\ &\text{THEN} \quad \Delta u_j = q_{1,j} e(k) + q_{2,j} \Delta e(k) \end{aligned} \tag{2.31(5)}$$

where Δu_j denotes the output of the j-th implication, and $q_{p,j}(p = 1, 2)$ are consequent parameters of the sub-PI controllers.

When the rule consequents are constants instead of local PI controllers, the resulted Mamdani fuzzy system can be viewed as defining a piecewise constant function with extensive interpolation.

An example of such a rule base is given in Table 2.31e. Three linguistic terms are used for each variable (N, Negative; ZE, Zero; P, Positive). Each entry of the table defines one rule, e.g., R_{11}: "**If** *e* is N **And** Δ *e* is N **then** *Δu* = –200." When the fuzzy inference shown in Figure 2.31b is used to calculate the output of the fuzzy model, the resulting control surface obtained by plotting the inferred control action for discretized values of e and Δe is shown in Figure 2.31f.

TABLE 2.31e
Rule Base of a Fuzzy PI Controller

e/Δe	*N*	*ZE*	*P*
N	–200	–100	0
ZE	–100	0	100
P	0	100	200

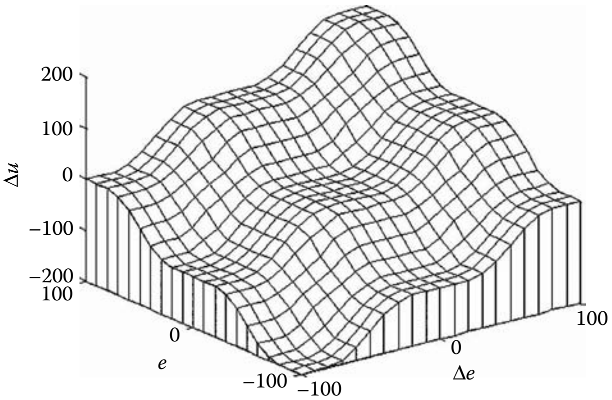


FIG. 2.31f
Nonlinear control surface of a fuzzy PI controller.

Direct and Supervisory Fuzzy Control

Since the first applications of PID-type fuzzy controllers, the application range of fuzzy systems has widened substantially. Although in most cases a fuzzy controller is used for direct feedback control, it can also be used on the supervisory level, e.g., as a self-tuning device in a conventional PID controller. Also, fuzzy control is no longer only used to directly express *a priori* process knowledge. For example, a fuzzy controller can be derived from a fuzzy model obtained through system identification. Therefore, a very general definition of fuzzy control can be given: *A fuzzy controller contains a (nonlinear) mapping that has been defined using fuzzy if–then rules.* There are four main sources for such fuzzy rules:

- *Expert experience and control engineering knowledge.* One classical example is the operator's handbook for a cement kiln.[8] The most common approach to establishing such a collection of rules of thumb is to question experts or operators using a carefully organized questionnaire.
- *Based on the operator's control actions.* Fuzzy *if–then* rules can be deduced from observations of an operator's control actions or a log book. In this case the rules express input/output relationships.
- *Based on a fuzzy model of the process.* A linguistic rule base may be viewed as an inverse model of the controlled process. Thus, the fuzzy control rules might be obtained by inverting a fuzzy model of the process. This method is restricted to relatively low-order systems, but it provides an explicit solution, assuming that fuzzy models of the open- and closed-loop systems are available.[9] Another approach is *fuzzy identification* or fuzzy model-based control (see the following subsection).
- *Based on learning (adaptive fuzzy control).* An adaptive or learning control system is used to cut down the amount of required human intervention, to increase the flexibility of the algorithm, to improve the control performance, and to reduce the initial design time and cost. Learning ability is one of the essential attributes of intelligent controllers. Being intelligent means that the controller can improve its future performance based on the information it has gained in the past.

Because of the recent surge of these techniques, one of the recent research directions is to design fuzzy systems that have learning capabilities. The first adaptive fuzzy controller was proposed by Procyk and Mamdani (1979).[10] This self-organizing algorithm was the basis of the algorithms of Árva and Nemeth (1991),[11] Renders, Saerens, and Hugues (1997),[12] and others. Nowadays, a promising approach is the combination of neural networks and fuzzy logic systems.

Artificial neural networks (ANN) are easy to train and have known convergence properties.[13] Because they are opaque, the model that results is not interpretable, and it is difficult to incorporate linguistic and/or first-principle knowledge into the identification procedure. Conversely, fuzzy models provide transparency through their linguistic interpretability. The combination of these methods resulted in neuro-fuzzy systems.[14,15]

The rule base of these neuro-fuzzy models is initialized by expert knowledge, and the model parameters are determined via similar methods that are used for the training of neural networks. As the neuro-fuzzy model can initialize and learn linguistic rules, the modeling framework can be considered as a direct transfer of knowledge. This is the powerful advantage of adaptive fuzzy systems in comparison with other nonlinear model structures (e.g., neural networks).

Classical Fuzzy Control Algorithms

The early work in fuzzy control was motivated by a desire to mimic the control actions of an experienced human operator (knowledge-based part) and obtain smooth interpolation between discrete outputs that would normally be obtained (fuzzy logic part). Hence, these fuzzy controllers perform just as well as the operators whose knowledge was used for their design. Note, this approach is not only for control applications. If the target system to be emulated is a human physician or a credit analyst, then the resulting fuzzy inference systems become fuzzy expert systems for diagnosis and credit analysis, respectively.

As is illustrated in Figure 2.31g, besides this direct control scheme, fuzzy controllers are being used in various other control schemes:

Scheme a. In this configuration, it is up to the operator to control the process in the existing (conventional) mode or to switch the controls over to a fuzzy control strategy. Often, the conventional loops represent an existing control scheme that has been controlling the

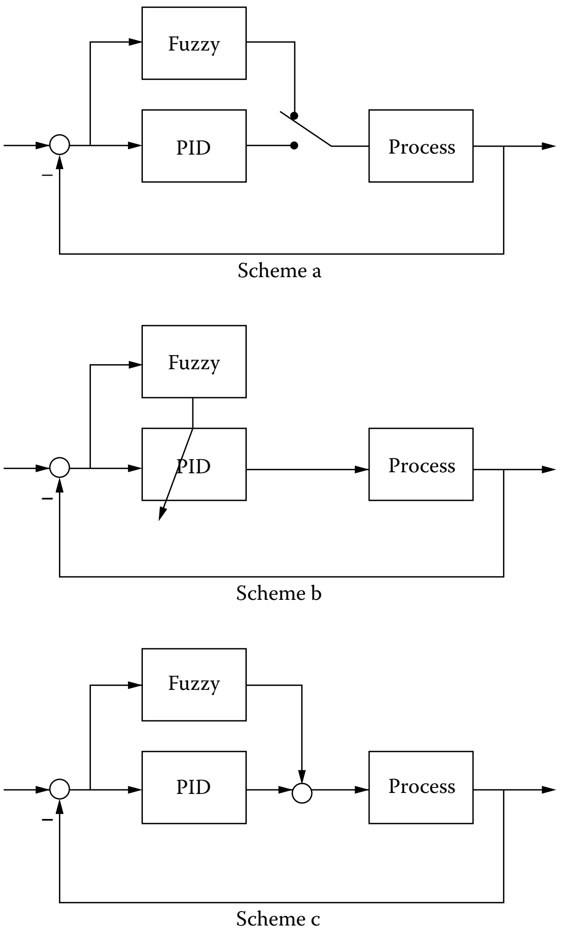


FIG. 2.31g
Fuzzy control configurations: (a) fuzzy replaces PID, (b) fuzzy adjusts PID parameters, and (c) fuzzy adds to PID output.

process before installation of a high-level strategy. The operator has to decide which of the two alternatives is the most likely to produce the best control performance.

Scheme b. In this configuration, the high-level strategy is used to adjust the tuning parameters of the conventional control loops. A common problem with linear PID controllers used for control of highly nonlinear processes is that the set of controller parameters produces satisfactory performance only when the process is within a small operational region. Outside this region, other controller settings are necessary, and these adjustments may be made automatically by a high-level strategy. This is called *gain scheduling* since such algorithms were originally used to change controller gains. A gain scheduling controller contains a linear controller whose parameters are changed as a function of the operating point in a preprogrammed way. It requires a thorough knowledge of the plant, but it is often a good way to compensate for nonlinearities and parameter variations. Additional sensor measurements can be used as *scheduling variables* that govern the change of the controller parameters, often by means of a look-up table.

Scheme c. This scheme can be viewed as a collaboration of linear and nonlinear control actions; the controller may be a linear PID controller, while the fuzzy controller is a supplementary nonlinear controller. Normally, conventional control systems using PID controllers are capable of controlling the process when the operation is steady and close to normal conditions. However, if sudden changes occur or if the process enters abnormal situations, then this configuration may be useful to bring the process back to normal operation as fast as possible. This scheme can also be interpreted as a *feedforward–feedback control,* where a measurable disturbance is being compensated. Generally, this solution requires a good model, but if a mathematical model is difficult or expensive to obtain, a fuzzy model may be useful.

These approaches have resulted in many successful applications, which are listed in Table 2.31h.

POLYMERIZATION REACTOR EXAMPLE

In this example the fuzzy control of the batch production of expanded polystyrene (EPS) is investigated. The typical unit of batch processes in polymer industries is an autoclave

TABLE 2.31h
Application of Conventional Fuzzy Controllers

Start-up of a catalytic reactor by a fuzzy controller	Yamashita, 1988[16]
Fuzzy control of a polymerization reactor	Roffel, 1991[17]
Fuzzy control of an iron ore sintering plant	Arbeithuber, 1995[18]
Self-tuning of a PI controller using fuzzy logic, for a construction unit testing apparatus	Berger, 1996[19]
Adaptive fuzzy control of the molten steel level in a strip-casting process	Lee, 1996[20]
Temperature control of highly exothermic batch polymerization reactors	Ni, 1997[21]
pH fuzzy control of automated industrial electroplating of gold	Jinxiang, 1997[22]
Fuzzy self-organizing pH control of an ammonia scrubber	Ylen, 1997[23]
Fuzzy logic control of an unstable biological reactor	Inamdar, 1997[24]
A study of gain scheduling control applied to an exothermic CSTR	Lagerberg, 1997[25]
Modeling and control of continuous stirred tank reactor for thermal copolymerization	Hwang, 1997[26]
Fuzzy control of a batch polymerization reactor	Abonyi, 1997[27]

with a heating–cooling system, where the reactor temperature is controlled by adjusting the temperature of the water recirculating though the jacket of the reactor. Compared to continuous processes, the control of batch processes is more difficult due to the complex sequential control tasks and the time-varying operating conditions. A further difficulty arises from the fact that vinyl polymerization is an exothermic autocatalytic reaction.

Several researchers have addressed the issue of control problems in reactor systems of this type. Juba and Hamer[28] and Berber[29] provide an excellent overview on the challenges in batch reactor control and suggest several control strategies. Davidson proposes a control algorithm that combines knowledge of the process with the logic used by the skilled operator and compares the result with standard PI controllers.[30] Examples of the application of fuzzy control to chemical reactors can also be found in the literature.[9] Here the benefits of fuzzy control schemes will be presented.

The aim of the development of fuzzy controllers was to obtain control performance better than that of the PID controller. The parameters of the designed controllers were determined based on simulation experiments. The simulator of the process has been built using MATLAB and C based on the first-principle mathematical model of the process.[31–33] During the tuning of the controller the following performance index was minimized by sequential quadratic programming

$$Q = \sum_{k=1}^{N} e(k)^2 + \lambda \sum_{k=1}^{N} \Delta u(k)^2$$

where λ is a weighting factor to maintain the dynamics of the control signal (0.2).

As we have shown in the previous example, conventional control solutions (PID algorithms, gain-scheduling) can be realized by choosing the proper fuzzy structure and can be improved by applying Takagi–Sugeno fuzzy models. The rules of the fuzzy PI controller were formulated by Equation 2.31(1). The performance of the fuzzy controller is illustrated in Figure 2.31j. As Table 2.31i demonstrates, the extension of the classical control schemes with nonlinearities defined by fuzzy systems resulted in 5 to 30% improvement in the performance index of the controllers.

TABLE 2.31i
Comparison of the Performances of Various Controllers

Control Algorithm	*Performance Index*
Optimal PI	1.883
Optimal gain-scheduled PI	1.633
Optimal PI-type fuzzy	1.579
Optimal fuzzy supervised PI controller	1.509

This example showed that fuzzy control is not only useful to handle heuristic control knowledge, but it can be a proper tool where the existing linear control algorithm should be replaced by a better, more effective nonlinear control algorithm.

MODEL-BASED FUZZY CONTROL

In spite of the practical successes, conventional fuzzy control has some drawbacks. First, fuzzy control rules are experience oriented and/or obtained after time-consuming trial-and-error experiments. Second, when a significant change in the system occurs that is outside of the operator experience, re-tuning of the fuzzy controller is necessary. Moreover, it is much easier to obtain information on how a process responds to particular inputs than to record how and why an operator responds to particular situations. These difficulties explain the recent surge of interest in model-based designs of fuzzy controllers, where fuzzy or inverse fuzzy models are used to design the controller.[34–36]

Fuzzy modeling can be effectively used in the identification of process dynamics in order to cope with nonlinear and complex systems.[9,36–38] The applicability of fuzzy models and

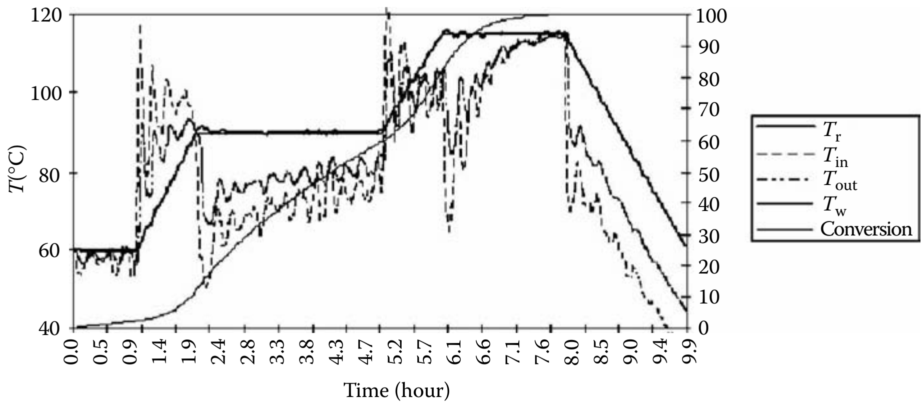


FIG 2.31j
Performance of the fuzzy supervised optimal PI controller.

TABLE 2.31k
Application of Fuzzy Techniques in Model-Based Control

Fuzzy modeling and control of multilayer incinerator	Sugeno, 1986[39]
A model-based controller	Posthlethwaite, 1994[40]
Fuzzy model-based control: stability, robustness, and performance issues	Johansen, 1994[41]
Predictive control of a batch fermentation process	Foss, 1995[42]
Application of fuzzy relational modeling to industrial product quality control	Qian, 1995[43]
Fuzzy-neural-net-based inferential control for a high-purity distillation column	Luo, 1995[44]
Fuzzy models in process optimal control	Keil, 1995[45]
Predictive fuzzy control applied to the sinter strand process	Hu, 1997[46]
Building a model-based fuzzy controller	Posthlethwaite, 1996[47]
Fuzzy relational model-based control, applying stochastic and iterative methods for model identification	Sing, 1996[48]
Long-range predictive control using fuzzy process models	Linkens, 1996[49]
A new indentation algorithm for fuzzy relational models and its applications in model-based control	Posthlethwaite, 1997[50]
Model-based design of fuzzy control systems	Abonyi, 2003[9]

model-based control techniques is illustrated by several laboratory-scale or industrial examples overviewed in Table 2.31k.

Inverse Fuzzy Model-Based Control

The simplest way to control a process using a fuzzy model is to use the inverse process model as the controller. Several methods can be applied to obtain the inverse fuzzy model of a given process, but the following two are the most often used:

- Identification of the inverse model from input/output data
- Inversion of the original model

If an ideal model of the process is available, the control is perfect and the input/output relationship is stable.[51] This situation of perfect control is impossible to achieve because an exact inversion of the process can only be found in special situations, and the model is never identical to the process, resulting in a mismatch between the model and the process.

The internal model controller (IMC) structure consists of a controller, a process model, a filter, and the process itself, as shown in Figure 2.31l.

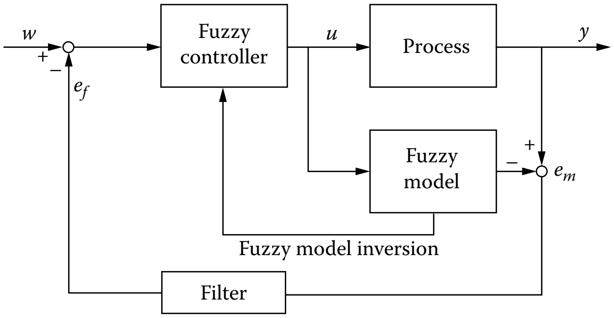


FIG. 2.31l
Fuzzy model in IMC architecture.

In this control arrangement, the process model is placed in parallel with a real process. The difference between the system and its model represents the modeling error and unmodeled process disturbances. This filtered difference is used by the controller to compensate the disturbance and the effects of the modeling error. The IMC structure provides a direct method for the design of feedback controllers. According to the above-mentioned properties of the control scheme, if a good model of the plant is available, the closed-loop system will provide good setpoint control even if unmeasured upsets or disturbances occur.

Reference 52 describes an adaptive fuzzy controller that is based on a fuzzy relational model and is applied to a laboratory-scale liquid level rig. Posthlethwaite has also applied fuzzy relational models for the identification and control of nonlinear, mainly first-order processes.[50] Contrary to these solutions, special types of Takagi–Sugeno fuzzy models can be analytically inverted. Such controllers were applied to control a radial industrial-cooling blast designed for building air conditioning[53] and for pressure control of a fermenter.[36]

When the process and control variables are subjected to level and rate constraints, and/or the system has significant time-delay or nonminimum-phase behavior, the inversion must be done for *p*-step-ahead.[54,55]

Fuzzy Model-Based Predictive Control

The *p*-step-ahead (*p*-inverse) controller can be seen as a special case of model predictive control (MPC). The nonlinear fuzzy models can be incorporated into this MPC scheme, as it is depicted in Figure 2.31m.

Among the several possible solutions of MPC, mainly the linearization-based approach has been utilized, because when a linear model is extracted from a fuzzy model, the controller can use classical linear model-based control algorithms.[56,57] Such local design techniques derived from model predictive control have been applied in References 9 and 58 through 62.

Operating Regime-Based Modeling

The modeling framework that is based on combining a number of local models, where each local model has a predefined operating region in which the model is valid, is called an *operating regime-based model.* The basic idea of operating

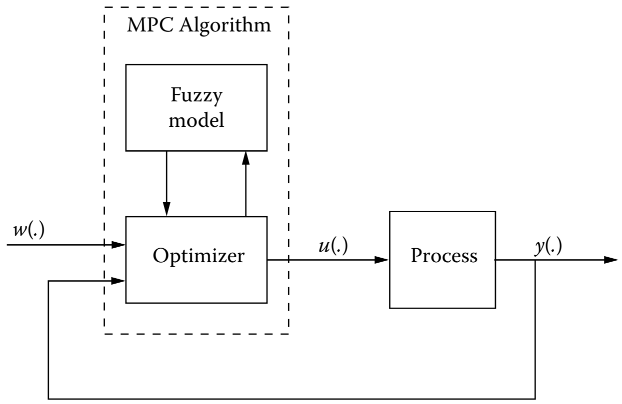


FIG. 2.31m
The role of a fuzzy model in model predictive control architecture.

regime-based modeling is that every model has a limited range of validity. This may be restricted by the modeling assumptions for a mechanistic model, or by the experimental conditions under which the identification data was obtained for an empirical model.

The model with a range of validity that is less than the operating regime of the process is called a *local model,* as opposed to a global model that is valid for the full range of the process operations. Global modeling is a complicated task because of the need to describe the interactions among a large number of phenomena that appear globally. Local modeling, on the other hand, may be considerably simpler, because locally a smaller number of phenomena may be relevant, and their interactions can be simpler. The local models can be combined into a global model using an interpolation technique.

The main advantage of this framework is its transparency. Both the concept of operating regimes and the model structure are easy to understand. This is important since the model structure can be interpreted in terms of operating regimes but also quantitatively in terms of individual local models. Furthermore, the operating regimes can be represented as a fuzzy set.[36,38] This representation is appealing because many systems change behavior smoothly as a function of the operating point, and the soft transition between the regimes introduced by the fuzzy set representation captures this feature in an elegant fashion. An example of such fuzzy models is the Takagi–Sugeno model.

Takagi–Sugeno Fuzzy Models

It is assumed that the MIMO dynamic process can be represented by the following nonlinear vector function: $\mathbf{y}(k+1) = f(\mathbf{y}(k),\ldots,\mathbf{y}(k-n_a+1), \mathbf{u}(k-n_d), \ldots, \mathbf{u}(k-n_d-n_b+1))$, where $\mathbf{y} = [y_1,\ldots y_{Ny}]^T$ is an -dimensional output vector, $\mathbf{u} = [u_1,\ldots\ u_{Nu}]^T$ is an N_u-dimensional input vector, n_a and n_b are maximum lags considered for the outputs and inputs, respectively, n_d is the minimum discrete dead time, and f represents the nonlinear model.

This MIMO Nonlinear Autoregressive with eXogenous Input (NARX) model can be represented by a Takagi–Sugeno fuzzy model,

$$R_j: \textbf{If}\quad z_1 \quad \text{is} \quad A_{1,j} \quad \textbf{and} \ldots \textbf{and} \quad z_n \quad \text{is} \quad A_{n,j} \quad \textbf{then}$$

$$\mathbf{y}^j(k+1) = \sum_{i=1}^{n_a} \mathbf{A}_i^j \mathbf{y}(k-i+1) + \sum_{i=1}^{n_b} \mathbf{B}_i^j \mathbf{u}(k-i-n_d+1) + \mathbf{c}^j \qquad \mathbf{2.31(6)}$$

where $\mathbf{A}_i^j$, $\mathbf{B}_i^j$ matrices and the $\mathbf{c}^j$ vector represent the j-th linear multivariable model, where the operating regime of this model in defined by the antecedent part of the j-th fuzzy rule. In these rule antecedents $A_{i,j}(z_i)$ represents the j-th antecedent fuzzy sets for the i-th input, where $\mathbf{z} = [z_1,\ldots z_n]$ is a “scheduling” vector, which is usually a subset of the previous process inputs and outputs,

$$\mathbf{z} = \{\mathbf{y}_1(k),\ldots,\mathbf{y}_1(k-n_a+1),\ldots,\mathbf{y}_{Ny}(k-n_a+1),\mathbf{u}_1(k-n_d),\ldots,$$
$$\mathbf{u}_1(k-n_d-n_b+1),\ldots,\mathbf{u}_{Nu}(k-n_d-n_b+1)\}$$

The proposed fuzzy model can be seen as a multivariable linear parameter varying (LPV) system model, where at the z operating point, the fuzzy model represents the following linear time-invariant (LTI) model,

$$\mathbf{y}(k+1) = \sum_{i=1}^{n_a} \mathbf{A}_i(\mathbf{z})\mathbf{y}(k-i+1)$$
$$+\sum_{i=1}^{n_b} \mathbf{B}_i(\mathbf{z})\mathbf{u}(k-i-n_d+1) + \mathbf{c}(\mathbf{z}) \qquad \mathbf{2.31(7)}$$

with

$$\mathbf{A}_i(\mathbf{z}) = \sum_{j=1}^{N_r} \beta_j(\mathbf{z})\mathbf{A}_i^j, \qquad \mathbf{B}_i(\mathbf{z}) = \sum_{j=1}^{N_r} \beta_j(\mathbf{z})\mathbf{B}_i^j,$$

$$\mathbf{c}(\mathbf{z}) = \sum_{j=1}^{N_r} \beta_j(\mathbf{z})\mathbf{c}^j,$$

where $0 \le \beta_j(\mathbf{z}) \le 1$ is the normalized truth value of the j-th rule calculated by Equation 2.31(4).

DISTILLATION COLUMN EXAMPLE

The examined process is a first-principle model of a binary distillation column. The column has 39 trays, a reboiler, and a condenser. The simulation model was developed by Skogestad.[63] The simulated system covers the most important effects for the dynamic of a real distillation column.

The studied column operates in *LV* configuration with two manipulated variables (reflux and boilup rate, $u_1 = L,V$, $u_2 = V$) and two controlled variables (top and bottom impurities, $y_1 = 1 - x_D$, $y_2 = x_B$).

The goal of this example was to improve the performance of the process by replacing the PID controller with an advanced control strategy. First, a linear model predictive controller (MPC) was designed, but it did not improve the performance of the control loop significantly. In order to obtain much better performance, the linear model of the process has been replaced by a nonlinear operating regime-based model of the distillation column.

The identified Takagi–Sugeno fuzzy model was used to represent the nonlinear dynamic system that is based on the interpolation between local linear MIMO ARX models. As the process gain varies in direct proportion with the concentrations, the fuzzy sets — the operating regions of the local models — are defined on the domain of the product impurities. This results in the MIMO Takagi–Sugeno (TS) fuzzy model structure, which was given by Equation 2.31(6).

The order of the local models is chosen to be $n_a = n_b = 2$. The process was assumed to have no time delay, $n_d = 0$. This rule base defines a grid-type partition of the operating regime. As Figure 2.31n shows, this fuzzy model consists of 16 multivariable local models.

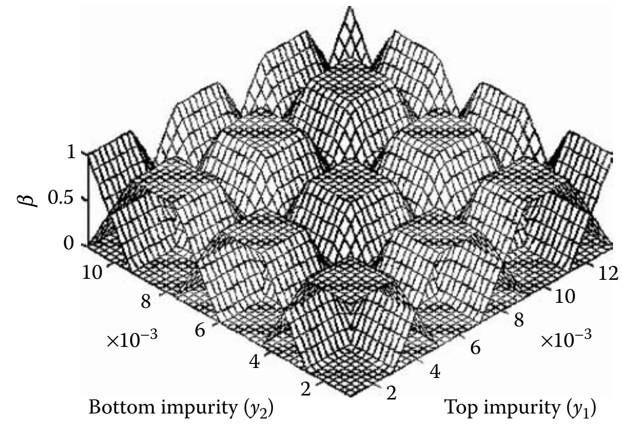


FIG. 2.31n
Operating regions of local linear models defined by fuzzy sets.

Because of the utilized nonlinear model, in the resulting MPC a nonlinear optimization problem had to be solved. To avoid this, a linear model was extracted from the nonlinear fuzzy model at each sampling time based on the linear-parameter-varying model interpretation of the Takagi–Sugeno fuzzy system. The multivariable generalized predictive controller (GPC)[64] is based on a linear ARX model extracted from the fuzzy model at each sampling time.[9]

Hence, based on the $\mathbf{y}(k)$ measured outputs, the previous values of the $\mathbf{u}$ control signal, the future set-points $\mathbf{w}$, and the $\mathbf{A}_j(\mathbf{z})$, $\mathbf{B}_j(\mathbf{z})$, $\mathbf{c}(\mathbf{z})$, parameters of the actual linear approximation of the nonlinear system (see Figure 2.31o), the multivariable general predictive controller (GPC) calculates that future control sequence vector $\{\mathbf{u}(k + j)\}$ at time k that minimizes a quadratic cost function. The results obtained are promising and hold the potential for industrial applications (Figure 2.31p).

SOFTWARE AND HARDWARE TOOLS

The development of fuzzy controllers requires intensive interaction with the user-operator of the process. Therefore, special developer tools have been introduced by various software (SW) and hardware (HW) suppliers such as Omron, Siemens, Aptronix, Inform, and National Semiconductors. Fuzzy control is also gradually becoming a standard option

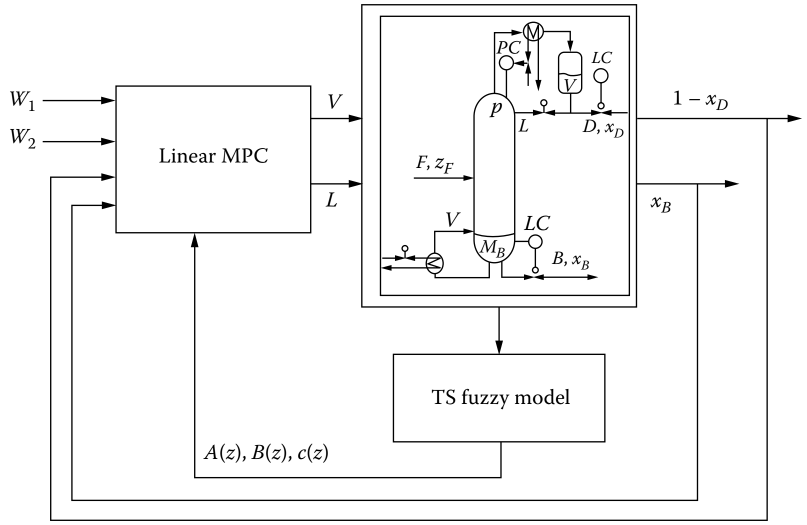


FIG. 2.31o
Fuzzy model-based predictive control of a distillation column, which is controlled in an LV-configuration.

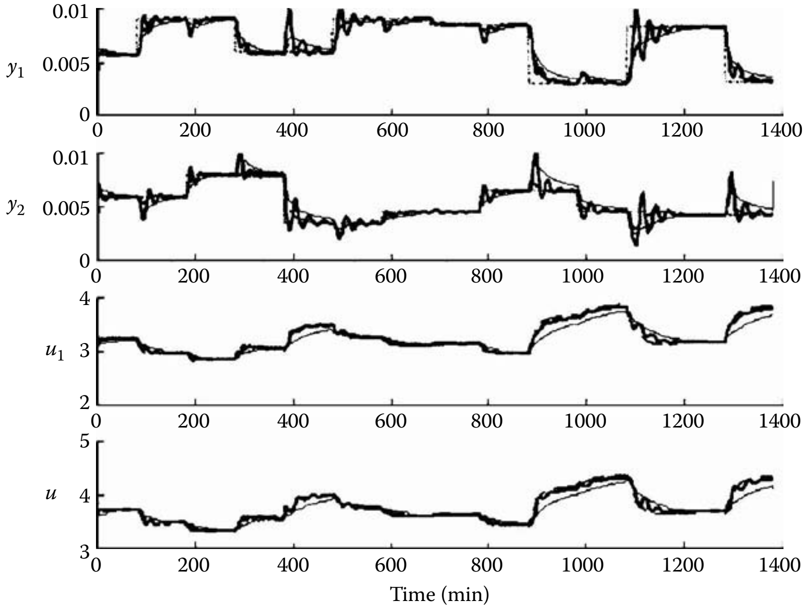


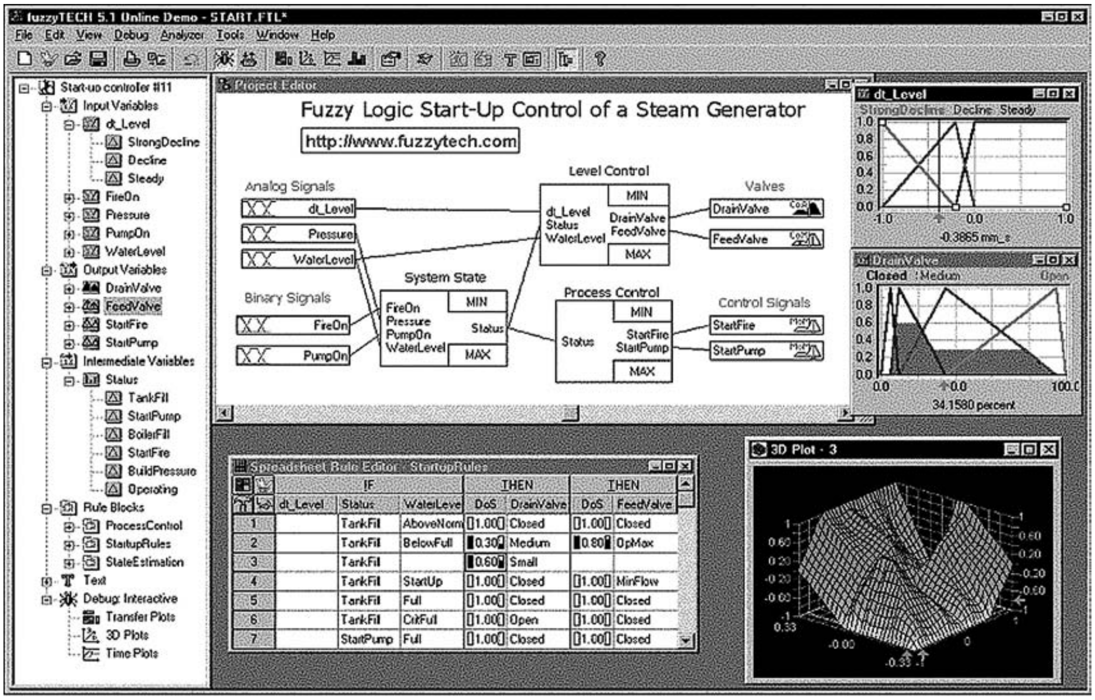
FIG. 2.31p
Performance of the constrained discrete PID (thick line), the constrained linear MPC (thin line) and the constrained nonlinear MPC (thick dashed line) for the distillation column.

in plant-wide control systems, such as the systems from Honeywell.

An extensive list of these HW and SW tools is given at www.fmt.vein.hu/softcomp/fuzzy. Although the user interfaces of these tools have been developed for different tasks, most of them consist of the following blocks.

Project Editor

The heart of the user interface is a graphical *project editor* that allows the user to build a fuzzy control system from basic blocks. An example of such an editor is given in Figure 2.31q.



	IF			THEN		THEN	
	dt_Level	Status	WaterLeve	DoS	DrainValve	DoS	FeedValve
1		TankFill	AboveNorm	1.00	Closed	1.00	Closed
2		TankFill	BelowFull	0.30	Medium	0.80	OpMax
3		TankFill		0.60	Small		
4		TankFill	StartUp	1.00	Closed	1.00	MinFlow
5		TankFill	Full	1.00	Closed	1.00	Closed
6		TankFill	CritFull	1.00	Open	1.00	Closed
7		StartPump	Full	1.00	Closed	1.00	Closed

FIG. 2.31q
Example of a project editor.

Input and output variables can be defined and connected to the fuzzy inference unit either directly or via preprocessing or postprocessing elements such as dynamic filters, integrators, and differentiators. The functions of these blocks are defined by the user, using the C language or its modification. Several fuzzy inference units can be combined to create more complicated (e.g., hierarchical or distributed) fuzzy control schemes.

Rule Base and Membership Functions

The fuzzy logic rules determine how the system reacts to input conditions. A rule-based controller is easy to understand and easy to maintain for a nonspecialist end-user. The rule base and the related fuzzy sets (membership functions) are defined using the rule base and membership function editors. This is where the designer can implement the control and optimization strategies for a solution. To best serve the needs of different applications, most of the tools provide several different rule editors.

The *spreadsheet rule editor* follows the familiar style of a table. The user can define complete rule bases just by pointing and clicking. Many experienced users prefer the *matrix rule editor* for complex rule blocks. Entering rules in FTL (Fuzzy Technology Language) format enables the rule definition in the most flexible way. Either way the rules are defined, the designer can always use any other editor to process or analyze them further.

The *membership functions editor* is a graphical environment for defining the shape and position of the membership functions. Figure 2.31q gives an example of the various interface screens.

Analysis and Simulational Tools

After the rules and membership functions have been designed, the function of the fuzzy controller can be tested using tools for *static analysis* and *dynamic simulation.*

Input values can be entered from the keyboard or read from a file in order to check whether the controller generates expected outputs. The degree of fulfillment of each rule, the adjusted output fuzzy sets, and the results of rule aggregation and defuzzification can be displayed online or logged in a file. For a selected pair of inputs and a selected output, the control surface can be examined in two or three dimensions.

Some packages also provide functions for automatic checking of completeness and redundancy of the rules in the rule base. For example, fuzzyTECH's 2D and 3D Analyzers let the designer analyze the transfer characteristics of a fuzzy logic system or its components in any possible way. Any modification of the system is immediately reflected in all analyzers. This facilitates "interactive" development of a solution.

Dynamic behavior of the closed loop system can be analyzed in simulation. While these tools comprise all simulation features for the fuzzy logic system under design, in some types of software it is also possible to use standard simulation software packages such as Matlab/Simulink™ or MatrixX™.

Code Generation and Communication Links

Once the fuzzy controller is tested using the software analysis tools, it can be used for controlling the plant either directly from the environment (via computer ports or analog inputs/outputs), or through generating a run-time code. Most of the programs generate a standard C-code and also a machine code for a specific hardware, such as microcontrollers or programmable logic controllers (PLCs). In this way, existing hardware can be also used for fuzzy control. Besides that, specialized fuzzy hardware is marketed, such as fuzzy control chips or fuzzy coprocessors for PLCs.

An example of a fuzzy processor is Motorola 68HC12 MCU. This 16-bit microcontroller family includes four fuzzy logic instructions in addition to the memory and on-chip peripheral functions. A fuzzy inference kernel on the HC12 takes one-fifth as much code space and executes more than 10 times faster compared to an HC11 general purpose MCU.

It should be noted that after the design of the fuzzy system an equivalent controller could be implemented using conventional techniques; in fact, any fuzzy controller could be represented by a classical look-up table.

CONCLUSIONS

From the control engineering perspective, a fuzzy controller is a nonlinear controller. The linguistic nature of fuzzy control makes it possible to handle expert knowledge concerning how the process should be controlled or how the process behaves. The interpolation aspect of fuzzy control has led to the viewpoint where fuzzy systems are seen as smooth function approximation schemes.

According to Michael Athans, fuzzy control methods are "parasitic"; they simply implement interpolations of control strategies obtained by other means.[65] This is partly true. Most of the fuzzy control algorithms can be interpreted as gain-scheduled controllers. Therefore, some people attack the fuzzy control community by stating that the final control and/or modeling algorithm just boils down to a nonlinear gain schedule that could be obtained by other interpolation methods, too.

Fuzzy control is a new technique that should be seen as an extension to existing control methods and not their replacement. It provides an extra set of tools which the control engineer has to learn to use where it makes sense. Nonlinear and partially known systems that pose problems to conventional control techniques can be tackled using fuzzy control. Fuzzy techniques provide a man–machine interface, which facilitates the acceptance, validation, and transparency of the process model or controller very much.[66] In this way, the control engineering is a step closer to achieving a higher level of automation in places where it has not been possible before.

References

1. Takatsu, H., and Itoh, T., "Future Need for Control Theory in Industry—Report of the Control Technology Survey in Japanese Industry," *IEEE Transactions on Control Systems Technology,* 7, 298–305, 1989.
2. Driankov, D., Helerndoorn, H., and Reinfrank, M., *An Introduction to Fuzzy Control,* Berlin: Springer-Verlag, 1993.
3. Passino, K. M., and Yurkovic S., *Fuzzy Control,* New York: Addison-Wesley, 1998.
4. Wang, L. X., *A Course in Fuzzy Systems and Control,* New York: Prentice Hall, 1997.
5. Yager, R. R., and Filev, D. P., *Essentials of Fuzzy Modeling and Control,* New York: John Wiley & Sons, 1994.
6. Yi, S. Y., and Chung, M. J., "Identification of Fuzzy Relational Model and Its Applications to Control," *Fuzzy Sets and Systems,* 59, 1, 25–33, 1993.
7. Takagi, T., and Sugeno, M., "Fuzzy Identification of Systems and Its Applications to Modeling and Control," *IEEE Transactions on Systems, Man, and Cybernetics,* 15, 1, 116–132, 1985.
8. Holmblad, L. P., and Østergaard, J.-J., "Control of a Cement Kiln by Fuzzy Logic," in M. M. Gupta and E. Sanchez (Eds.), *Fuzzy Information and Decision Processes,* Amsterdam: North-Holland, 1982.
9. Abonyi, J., *Fuzzy Model Identification,* Boston: Birkhauser, 2003.
10. Procyk, T. J., and Mamdani, E. H., "A Linguistic Self-Organizing Process Controller," *Automatica,* 15, 15–30, 1979.
11. Arva, P., and Nemeth, S., "Learning Algorithm in Fuzzy Control," *Annales Univ. Sci. Budapest., Sect. Comp.,* 12, 25–32, 1991.
12. Renders, J. M., Saerens, M., and Hugues, "Fuzzy Adaptive Control of a Certain Class of SISO Discrete-Time Processes," *Fuzzy Sets and Systems,* 85, 49–61, 1997.
13. Narendra, K. S., and Parthasarathy, K., "Identification and Control of Dynamical Systems," *IEEE Transactions on Neural Networks,* 1, 4–27, 1990.
14. Jang J.-S. R., and Sun, C.-T., "Neuro-Fuzzy Modeling and Control," *Proceedings of the IEEE,* 83, 378–406, 1995.
15. Jang J.-S. R., "ANFIS: Adaptive-Network-Based Fuzzy Inference Systems," *IEEE Transactions on Systems, Man and Cybernetics,* 23, 3, 665–685, 1993.
16. Yamashita, Y., Matshumoto, S., and Suzuki, M., "Start-Up of a Catalytic Reactor by Fuzzy Controller," *Journal of Chemical Engineering of Japan,* 21, 277–281, 1988.
17. Roffel, B., and Chin, P. A., "Fuzzy Control of a Polymerization Reactor," *Hydrocarbon Processing,* June 1991, pp. 47–49.
18. Arbeithuber, C., Jorgl, H. P., and Raml, H., "Fuzzy Control of an Iron Ore Sintering Plant," *Control Engineering Practice,* 3, 12, 1669–1674, 1995.
19. Berger, M., "Self-Tuning of a PI Controller Using Fuzzy Logic, for a Construction Unit Testing Apparatus," *Control Engineering Practice,* 4, 6, 785–790, 1996.
20. Lee, D., and Kang, T., "Adaptive Fuzzy Control of the Molten Steel Level in a Strip-Casting Process," *Control Engineering Practice,* 4, 11, 1511–1520, 1996.
21. Ni, H., Debelak, K., and Hunkeler, D., "Temperature Control of Highly Exothermic Batch Polymerisation Reactors," *International Journal of Applied Polymerisation Science,* 63, 1997, 761–772, 1997.
22. Jinxiang, Z., Zaiman, Z., Yanbin, Q., Linru, Y,. Suqin, W., and Lianyu, L., "pH Fuzzy Control of Automated Industrial Electroplating of Gold," *Chemical Engineering Technology,* 20, 576–580, 1997.
23. Ylen, J. P., and Jutila, P., "Fuzzy Self-Organising pH Control of an Ammonia Scrubber," *Control Engineering Practice,* 5, 9, 1233–1244, 1997.
24. Inamdar, S. R., and Chiu, M. S., "Fuzzy Logic Control of an Unstable Biological Reactor," *Chemical Engineering Technology,* 20, 414–418, 1997.
25. Lagerberg, A., and Breitholtz, C., "A Study of Gain Scheduling Control Applied to an Exothermic CSTR," *Chemical Engineering Technology,* 20, 435–444, 1997.
26. Hwang, W. H., Chey, J. I., and Rhee, H. K., "Modelling and Control of Continuous Stirred Tank Reactor for Thermal Copolymerisation," *International Journal of Applied Polymerisation Science,* 67, 921–931, 1998.
27. Abonyi, J., Nagy, L., and Szeifert, F., "Fuzzy Control of Batch Polymerization Reactors," *IEEE International Conference on Intelligent Systems*, INES'97, 1997, pp. 251–255.
28. Juba, M. R., and Hamer, J. W., "Progress and Challenges in Batch Process Control," in *Proceedings of the CPC III, CHACE,* Amsterdam, 1986.
29. Berber, R., "Control of Batch Reactors: Review," *Trans IChemE,* 74, part A, 3–20, 1996.
30. Davidson, R. S., "Using Process Information to Control a Multipurpose Batch Chemical Reactor," *Computers Chem. Engng.,* 13, 83–85, 1989.
31. Cawthon, G. D., and Knaebel, K. S., "Optimization of Semibatch Polymerization Reactors," *Computers Chem. Engng.,* 13, 63–72, 1989.
32. Nemeth, S., and Thyrion, F. C., "Study of the Runaway Characteristics of Suspension Polymerization of Styrene," *Chemical Engineering Technology,* 18, 315–323, 1985.
33. Choi, K. Y., and Lei, G. D., "Modeling of Free-Radical Polymerization of Styrene by Bifunctional Iniciators," *AIChE Journal,* 33, 2067–2076, 1987.
34. Babuška, R., te Braake, H. A. B., Krijgsman, A. J., and Verbruggen, H. B., "Comparison of Intelligent Control Schemes for Real-Time Pressure Control," *Control Engineering Practice,* 4, 11, 1585–1592, 1996.
35. Palm, R., Driankov, D., and Hellendoorn, H., *Model Based Fuzzy Control,* Heidelberg: Springer-Verlag, 1996.
36. Babuška, R., *Fuzzy Modelling for Control,* Boston: Kluwer Academic Publishers, 1998.
37. Babuška, R., and Verbruggen, H. B., "An Overwiew of Fuzzy Modeling for Control," *Control Engineering Practice,* 4, 11, 1996, 1593–1606, 1996.
38. Hellendoorn, H., and Driankov, D. (Eds.), *Fuzzy Model Identification: Selected Approaches,* Berlin: Springer-Verlag, 1997.
39. Sugeno, M., and Kang, G. T., "Fuzzy Modeling and Control of Multilayer Incinerator," *Fuzzy Sets and Systems,* 18, 329–346, 1986.
40. Posthlethwaite, B., "A Model-Based Fuzzy Controller," *Transactions of IChemE,* 72, A1, 38–46, 1994.
41. Johansen, T. A., "Operating Regime-Based Process Modelling and Identification," Ph.D. thesis, University of Trondheim, The Norwegian Institute of Technology, 1994.
42. Foss, B. A., Johansen, T. A., and Sorensen, A. V., "Nonlinear Predictive Control Using Local Models—Applied to a Batch Fermentation Process," *Control Engineering Practice,* 3, 389–396, 1995.
43. Quian, Y., and Tessier, J. C., "Application of Fuzzy Relational Modelling to Industrial Product Quality Control," *Chemical Engineering Technology,* 18, 330–336, 1995.
44. Luo, R. F., and Zhang, Z. J., "Fuzzy-Neural-Net–Based Inferential Control for a High-Purity Distillation Column," *Control Engineering Practice,* 3, 31–40, 1995.
45. Keil, F., Stoyanov, S., and Shunova, E., "Fuzzy Models in Process Optimal Control," *Hungarian Journal of Industrial Chemistry,* 23, 201–206, 1995.
46. Hu, J. Q., and Rose, E., "Predictive Fuzzy Control Applied to the Sinter Strand Process," *Control Engineering Practice,* 5, 247–252, 1997.
47. Posthlethwaite, B., "Buliding a Model-Based Fuzzy Controller," *Fuzzy Sets and Systems,* 79, 3–13, 1996.
48. Sing, C. H., and Postlethwaite, B., "Fuzzy Relational Model-Based Control Applying Stochastic and Iterative Methods for Model Identification," *Transactions of IChemE,* 74, A1, 70–76, 1996.

49. Linkens, D. A., and Kandiah, S., "Long-Range Predictive Control Using Fuzzy Process Models," *Chemical Engineering Research and Design,* (*Transactions of IChemE*), 74, Part A, 77–88, 1996.
50. Posthlethwaite, B., Brown, M., and Sing, C. H., "A New Identification Algorithm for Fuzzy Relational Models and its Applications in Model-Based Control," *Transactions of IChemE,* 75, A1, 453–458, 1997.
51. Garcia, C. E., and Morari, M., "Internal Model Control. 1. A Unifying Review and Some New Results," *Ind. Eng. Chem. Process Des. Dev.,* 21, 308–323, 1982.
52. Graham, P., and Newel, R. B., "Fuzzy Adaptive Control of a First-Order Process," *Fuzzy Sets and Systems,* 31, 47–65, 1989.
53. Fischer, M., and Nelles, M., "Predictive Control Based on Local Linear Fuzzy Models," *International Journal of Systems Science,* 29, 679–697, 1998.
54. Hernandez, E., and Arkun, Y., "Study of Control-Relevant Properties of Backpropagation Neural Network Models of Nonlinear Dynamical Systems," *Computers and Chemical Engineering,* 16, 4, 227–240, 1992.
55. Sistu, P. B., and Bequette, B. W., "Nonlinear Predictive Control of Uncertain Processes: Application to a CSTR," *AICHE Journal,* 37, 11, 1711–1722, 1991.
56. Driankov, D., and Palm, R. (Eds.), *Advances in Fuzzy Control,* Heidelberg: Springer-Verlag, 1998.
57. Palm, R., and Rehfuess, U., "Fuzzy Controllers as Gain Scheduling Approximations," *Fuzzy Sets and Systems,* 85, 233–246, 1997.
58. Abonyi, J., Nagy, L., and Szeifert, F., "Fuzzy Model-Based Predictive Control by Instantaneous Linearization," *Fuzzy Sets and Systems,* 120, 1, 109–122, 2001.
59. Biassizo, K. K., Skrjanc, I., and Matko, D., "Fuzzy Predictive Control of Highly Nonlinear pH Process," *Comp. Chem. Eng.,* 21/S, 613–618, 1997.
60. Espinosa, J. J., and Vandewalle, J., "Predictive Control Using Fuzzy Models," in *Advances in Soft Computing — Engineering Design and Manufacturing,* Roy, R., Ed., London: Springer-Verlag, 1998.
61. Fisher, M., and Isermann, R., "Inverse Fuzzy Process Models for Robust Hybrid Control," in *Advances in Fuzzy Control,* Driankov, D., and Palm, R., Eds., Heidelberg: Physica-Verlag, 1998, pp. 103–128.
62. Linkens, D. A., and Kandiah, S., "Long-Range Predictive Control Using Fuzzy Process Models," *Chemical Engineering Research and Design, (Transactions of IChemE*), 74, Part A, 77–88, 1996.
63. Skogestad, S., "Dynamics and Control of Distillation Columns," *Chem. Eng. Res. Des. (Trans IChemE),* 75, 539–562, 1997.
64. Clarke, D. W., and Mohtadi, C., "Properties of Generalized Predictive Control," *Automatica,* 25, 6, 859–875, 1989.
65. Athans, M., Debate with Prof. Lofti A. Zadeh, 7th European Conference on Intelligent Techniques and Soft Computing (EUFIT '99), Aachen, Germany, September 1999.
66. Verbruggen, H. B., and Brujin, P. M., "Fuzzy Control and Conventional Control: What Is (and Can Be) the Real Contribution of Fuzzy Systems?" *Fuzzy Sets and Systems,* 90, 151–160, 1997.

2.32 Stability Analysis, Transfer Functions

C. F. MOORE (1970) **P. D. SCHNELLE, JR.** (1985) **B.G. LIPTÁK** (1995)
R. J. SMITH II (2005)

Control theory, which is the topic of this chapter, is approached differently by industry and academia. At some universities, process control courses have been treated as yet another course in mathematics, dealing with differential equations and analysis in the frequency domain. In industry, process control is discussed in the "time domain" (the domain that we understand better because we live in it) and less emphasis is given to the highly mathematical and abstract "frequency domain analysis," which makes mathematical manipulations easier to perform.

Because of this difference, even the terminology and language used in industry are different from those used in academia and in testing laboratories. This can sometimes be confusing because what the process industry calls a set point becomes a reference signal in academia, what is a controlled variable becomes process output, what is a manipulated variable becomes process input, and what is dead time becomes a time lag.

Although the frequency domain is friendly to partial differential equations, it is not so friendly to common-sense understanding and Murphy's Law-type reality. For these reasons many of the sections dealing with control loop stability in this handbook work in the time domain and use industrial terminology.

Because Chapter 2, of which this section is part, also contains a substantial amount of material on the subjects of simulation and modeling, we have used both sets of terms in the various sections. In this section, which provides an in-depth treatment to frequency domain transform theories and transfer functions used for linearization and stability analysis, it is appropriate to stay in the frequency domain.

INTRODUCTION

Most techniques used in the analysis of control problems require descriptive mathematical equations. This section describes some of the mathematical tools needed for the analysis of control problems.

One of the main topics of this section is the subject of transforms, which allow one to handle difficult problems in a simple manner. As the use of logarithms simplifies multiplication to addition, so do Laplace and z transforms perform a similar function in the solution of differential and difference equations, respectively.

State-space representation is another useful method for expressing multiple-input multiple-output (MIMO) difference and differential equations. This permits large systems to be handled as a first-order matrix system using linear algebra techniques.

Some of the other control system analysis tools discussed here are the block diagram, for signal flow representation, and linearization, which is a technique to convert nonlinear equations into linear form. Several graphical techniques are also extremely useful in the analysis of dynamic systems. These include the Bode diagram, Nyquist plot, and Nyquist array methods, which are all covered in this section.

One of the main purposes of control system analysis is to evaluate and guarantee that particular systems will be stable in their operation. Stability criteria that are discussed in this section include the Descartes, Routh's, and Nyquist criteria.

LAPLACE TRANSFORMS

One very useful tool in the analysis of differential equations is the principle of Laplace transforms. The Laplace transform concept is widely used in process control and was used to provide the basic framework upon which most automatic control theory was based.

The goal of any transform operation is to transform a difficult problem into a form that is more convenient to handle. Once the desired manipulations or results have been obtained from the transformed problem, an inverse transformation can be made to obtain the solution of the original problem.

For example, logarithms are a transform operation by which problems of multiplication and division can be transformed to problems of addition and subtraction. Laplace transforms perform a similar function in the solution of differential equations. The Laplace transform of a linear ordinary differential equation results in a linear algebraic equation. The algebraic problem is usually much simpler to solve than the corresponding differential equation would be.

Once the Laplace domain solution has been found, the corresponding time domain solution can be determined by an inverse transformation. There also exist very powerful control analysis and design techniques available for *s*-domain and frequency domain models. These include Bode and Nyquist analyses, to be discussed later.

The Laplace transform of a time domain function $f(t)$ will be noted by the symbol $F(s)$, and will be defined as follows:

$$F(s) = L[f(t)] = \int_0^\infty f(t)e^{-st}\,dt \qquad \mathbf{2.32(1)}$$

where $L[f(t)]$ is the symbol for indicating the Laplace transformation of the function in brackets. The variable s is a complex variable ($s = a + jb$) introduced by the transformation. All time-dependent functions in the time domain become functions of s in the Laplace domain (*s*-domain).

For the mathematical concept of the Laplace transformation to be meaningful, certain restrictions are placed on the function $F(s)$. However, in most practical work no such difficulties are encountered; therefore, this issue will not be considered in this treatment, and only concepts normally used in process control will be covered.

Theorems

A number of theorems exist that facilitate the use of Laplace transform techniques. The following are some of the most useful ones for causal signals.

Linearity Theorem (Superposition) This theorem holds when the result is the same if two or more "signals" are manipulated together or separately:

$$L[K\,f(t)] = KL[f(t)] = K\,F(s) \quad (K = \text{constant}) \qquad \mathbf{2.32(2)}$$

$$L[f_1(t) \pm f_2(t)] = F_1(s) \pm F_2(s) \qquad \mathbf{2.32(3)}$$

Real Differentiations Theorem First-order differential

$$L\left[\frac{d}{dt}f(t)\right] = sF(s) - f(0) \qquad \mathbf{2.32(4)}$$

General *n*th-order differential

$$L\left[\frac{d^n}{dt^n}f(t)\right] = s^nF(s) - s^{n-1}f(0) - s^{n-2}\frac{d}{dt}f(0) - s\frac{d^{n-2}}{dt^{n-2}}f(0) - \frac{d^{n-1}}{dt^{n-1}}f(0) \qquad \mathbf{2.32(5)}$$

Real Integration Theorem

$$L\left[\int f(t)dt\right] = \frac{F(s)}{s} \qquad \mathbf{2.32(6)}$$

In general,

$$L\left[\int^1\!\!\int^2 \cdots \int^n f(t)dt^n\right] = \frac{1}{s^n}F(s) \qquad \mathbf{2.32(7)}$$

Initial Value Theorem

$$f(0) = \lim_{t\to\infty} f(t) = \lim_{s\to\infty} sF(s) \qquad \mathbf{2.32(8)}$$

Final Value Theorem

$$f(\infty) = \lim_{t\to\infty} f(t) = \lim_{s\to 0} sF(s) \qquad \mathbf{2.32(9)}$$

The direct and inverse Laplace transformation of a particular function can be obtained from direct integration of Equation 2.32(1) and/or from the applications of the above theorems. Extensive tabulations of specific transform pairs are also available in most mathematical tables. A few of the more common transform pairs encountered in control analysis are tabulated in Table 2.32a.

First-Order Lag

The following two examples illustrate the solution of differential equations using the underlying principle of the Laplace transform technique. First consider a simple first-order lag described earlier:

$$\tau\frac{d}{dt}c(t) + c(t) = K\,r(t) \qquad \mathbf{2.32(10)}$$

where

$$c(0) = 0.0$$

and

$$r(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases} \qquad \mathbf{2.32(11)}$$

The general procedure in the solution of the above equation is to:

1. Transform the differential equation to the Laplace domain.
2. Solve the resulting algebraic equations for the system output, $C(s)$.
3. Take inverse Laplace transformation of the expression describing $C(s)$ to determine the corresponding time domain solution.

Because of the Linearity Theorem, each term in the differential equation 2.32(10) can be transformed individually:

$$L\left[\tau\frac{d}{dt}c(t)\right] + L[c(t)] = L[K\,r(t)] \qquad \mathbf{2.32(12)}$$

TABLE 2.32a
Laplace Transform Pairs

Transform, F(s)	*Function, f(t)*
1	$\delta(t)$: Unit impulse
$\frac{1}{s}$	$u(t)$: Unit Step
$\frac{1}{s^2}$	t: Unit ramp
$\frac{n!}{s^{n+1}}\ (n = 1, 2, \ldots)$	t^n: nth order ramp
$\frac{1}{s \pm a}$	$e^{\mp at}$: Exponential
$\frac{1}{s(s \pm a)}$	$\frac{1}{\pm a}(1 - e^{\mp at})$
$\frac{s}{s^2 + a^2}$	$\cos at$
$\frac{a}{s^2 + a^2}$	$\sin at$
$\frac{s}{s^2 - a^2}$	$\cosh at$
$\frac{a}{s^2 - a^2}$	$\sinh at$
$\frac{b}{(s + a)^2 + b^2}$	$e^{-at} \sin bt$: Damped sine
$\frac{s + a}{(s + a)^2 + b^2}$	$e^{-at} \cos bt$: Damped cosine
$\frac{1}{(s + a)^n}$	$\frac{1}{(n-1)!} t^{n-1} e^{-at}$
$\frac{e^{-as}}{s}$	$u(t - a)$
$\frac{e^{-as}}{s^2}$	$\begin{cases} 0 & (0 < t < a) \\ t - a & (t > a) \end{cases}$
$\frac{1 - e^{-as}}{s}$	$\begin{cases} 1 & (0 < t < a) \\ 0 & (t > a) \end{cases}$
$\log \frac{s - a}{s - b}$	$\frac{1}{t}(e^{bt} - e^{at})$
$\tan^{-1} \frac{a}{s}$	$\frac{1}{t} \sin at$

The first term can be determined by the use of the Linearity and the Real Differentiations Theorems:

$$L\left[\tau \frac{d}{dt} c(t)\right] = \tau L\left[\frac{d}{dt} c(t)\right] = \tau(sC(s) = c(\varnothing^{\circ})) \qquad \textbf{2.32(13)}$$

The second term by definition is

$$L[c(t)] = C(s) \qquad \textbf{2.32(14)}$$

The third term can be determined from the table of transform pairs in Table 2.32a.

$$L[K\, r(t)] = \frac{K}{s} \qquad \textbf{2.32(15)}$$

The entire Laplace domain equation therefore will read

$$\tau s\, C(s) + C(s) = \frac{K}{s} \qquad \textbf{2.32(16)}$$

This can be solved for $C(s)$:

$$C(s) = \frac{K}{s(1 + \tau s)} \qquad \textbf{2.32(17)}$$

For the inverse transformation, Table 2.32a provides the following transform pair:

$$L\left[\frac{1}{\pm a}(1 - e^{\pm at})\right] = \frac{1}{s(s \pm a)} \qquad \textbf{2.32(18)}$$

Therefore,

$$c(t) = K(1 - e^{-t/\tau}) \qquad t > 0 \qquad \textbf{2.32(19)}$$

Partial Fraction Expansion

Frequently the necessary inverse transformation may not be directly available in the Laplace transform tables at hand. In such cases the function must be expanded in terms of the roots of the denominator of the Laplace expression, namely,

$$F(s) = \frac{A(s)}{B(s)} = \frac{C_1}{(s + r_1)} + \frac{C_2}{(s + r_2)} + \frac{C_3}{(s + r_3)} + \cdots + \frac{C_0}{(s + r_0)} \qquad \textbf{2.32(20)}$$

where

$$B(s) = (s - r_1)(s - r_2)(s - r_3) \cdots (s - r_n)$$

$$r_1, r_2, r_3, \ldots r_n = \text{roots of } B(s) \qquad \textbf{2.32(21)}$$

$$C_1, C_2, C_3, \ldots C_n = \text{constants in partial fraction}$$

The procedure for evaluating the constants in the expansion depends on the nature of the roots, which can be:

1. Real and distinct
2. Real and repeated
3. Complex conjugates

When the roots are real and distinct, the expansion of $F(s)$ is

$$F(s) = \frac{A(s)}{(s-r_1)(s-r_2)\cdots(s-r_n)} \qquad \textbf{2.32(22)}$$

$$F(s) = \frac{C_1}{(s-r_1)} + \frac{C_2}{(s-r_2)} + \cdots + \frac{C_n}{(s-r_n)} \qquad \textbf{2.32(23)}$$

and the inverse transformation is

$$f(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t} + \cdots C_n e^{r_n t} \qquad \textbf{2.32(24)}$$

where

$$\left.\begin{aligned} C_1 &= \underset{s\to r_1}{\text{limit}}\,[(s-r_1)\,F(s)] \\ C_2 &= \underset{s\to r_2}{\text{limit}}\,[(s-r_2)\,F(s)] \\ &\;\;\vdots \\ C_n &= \underset{s\to r_n}{\text{limit}}\,[(s-r_n)\,F(s)] \end{aligned}\right\} \qquad \textbf{2.32(25)}$$

For the case when the roots are real and repeated,

$$F(s) = \frac{A(s)}{(s-r_1)\cdots(s-r_j)^q\cdots(s-r_n)} \qquad \textbf{2.32(26)}$$

(The jth root is repeated q times.)

$$F(s) = \frac{C_1}{(s-r_1)} + \cdots + \left[\frac{C'_q}{(s-r_j)^q} + \frac{C'_{q-1}}{(s-r_j)^{q-1}} + \cdots + \frac{C'_1}{(s-r_j)}\right] + \cdots + \frac{C_n}{(s-r_n)} \qquad \textbf{2.32(27)}$$

The inverse transform is

$$f(t) = C_1 e^{r_1 t} + \cdots + [C'_q t^{q-1} + C'_{q-1} t^{q-2} + \cdots + C_1{}']e^{r_j t} + \cdots + C_1 e^{r_n t} \qquad \textbf{2.32(28)}$$

where

$$C'_q = \underset{s\to r_j}{\text{limit}}\,\{(s-r_j)^q\,F(s)\} \qquad \textbf{2.32(29)}$$

$$C'_{q-1} = \underset{s\to r_j}{\text{limit}}\left\{\frac{1}{1!}\frac{d}{ds}[(s-r_j)^q\,F(s)]\right\} \qquad \textbf{2.32(30)}$$

$$C_{q-k} = \underset{s\to r_j}{\text{limit}}\left\{\frac{1}{k!}\frac{d^k}{ds^k}[(s-r_j)^q\,F(s)]\right\} \qquad \textbf{2.32(31)}$$

where

$$k! = 1\cdot 2\cdot 3\cdot 4\cdot 5\cdots k.$$

For the case when the roots are complex,

$$F(s) = \frac{A(s)}{(s-r_1)\cdots(s-a-jb)\,(s-a+jb)\cdots(s-r_n)} \qquad \textbf{2.32(32)}$$

$$F(s) = \frac{C_1}{(s-r_1)} + \cdots + \frac{1}{2jb}|\pi_j|\left(\frac{e^{i\alpha}}{s-a-jb} - \frac{e^{-j\alpha}}{s-a+jb}\right) + \cdots + \frac{C_n}{(s-r_n)} \qquad \textbf{2.32(33)}$$

The inverse transformation results in

$$f(t) = C_1 e^{r_1 t} + \cdots + \frac{1}{b}|\pi_j|\,e^{at}\sin(bt+\alpha) + \cdots + C_n e^{r_n t} \qquad \textbf{2.32(34)}$$

$$\pi_j = \underset{s\to a+jb}{\text{limit}}\,[(s-a-jb)\,(s-a+jb)\,F(s)] \qquad \textbf{2.32(35)}$$

This results is a complex polynomial that can be expressed in terms of a magnitude, π_j, and an angle, α. For example,

$$\begin{aligned} F(s) &= \frac{10}{s(s^2+0.5s+0.4)} \\ &= \frac{10}{s(s+0.25-j0.58)\,(s+0.25+j0.58)} \end{aligned} \qquad \textbf{2.32(36)}$$

$$F(s) = \frac{C_1}{s} + \frac{1}{2jb}|\tau_j|\left(\frac{e^{i\alpha}}{s-a-jb} - \frac{e^{-j\alpha}}{s-a+jb}\right) \qquad \textbf{2.32(37)}$$

where $a = -0.25$; $b = 0.58$.

$$C_1 = \underset{s\to 0}{\text{limit}}\left[s\left(\frac{10}{s(s^2+0.5s+0.4}\right)\right] = 25 \qquad \textbf{2.32(38)}$$

$$\pi_j = \underset{s\to 0.25-j\,0.58}{\text{limit}} \left[(s^2+0.5s+0.4)\left(-\frac{10}{s(s^2+0.5s+0.4}\right)\right] = \frac{10}{-0.25+j0.58} \qquad \textbf{2.32(39)}$$

$$|\pi_j| = \frac{10}{\sqrt{(0.25)^2+(0.58)^2}} = \frac{10}{0.63} = 15.9 \qquad \textbf{2.32(40)}$$

$$\alpha = \text{angle of } \pi_j = 0 - \tan^{-1}\frac{0.58}{-0.25} = -113.5 \qquad \textbf{2.32(41)}$$

Therefore,

$$c(t) = 25 + \left(\frac{15.9}{0.58}\right) e^{-0.25t} \sin(0.58t - 113.5°) \quad \textbf{2.32(42)}$$

Z TRANSFORMS

Laplace transformation is a powerful tool for the solution of linear *differential* equations and is particularly useful when these equations represent the dynamic behavior of continuous systems. For certain kinds of discontinuous systems, whose dynamic behavior can be defined by linear *difference* equations, there exists another kind of transformation calculus, the *z*-transformation. It is particularly applicable to the study and design of sampled data control.

Let us suppose that we have a continuous function, $f(t)$, such as that shown at the top in Figure 2.32b by curve (1) and that this signal is to be fed into a sampling device such as that shown in Figure 2.32b(2).

At fixed intervals of time of period T, the switch closes briefly, allowing the signal $f(t)$ to be transmitted. If the length of time the switch is closed is infinitesimally short, the switch or sampler may be treated as an impulse modulator. That is to say, the sampler puts out a train of impulses $I(t)$, spaced from one another by the time interval T (Figure 2.32b (3)). The intermittent function at the switch output is designated as $f^*(t)$.

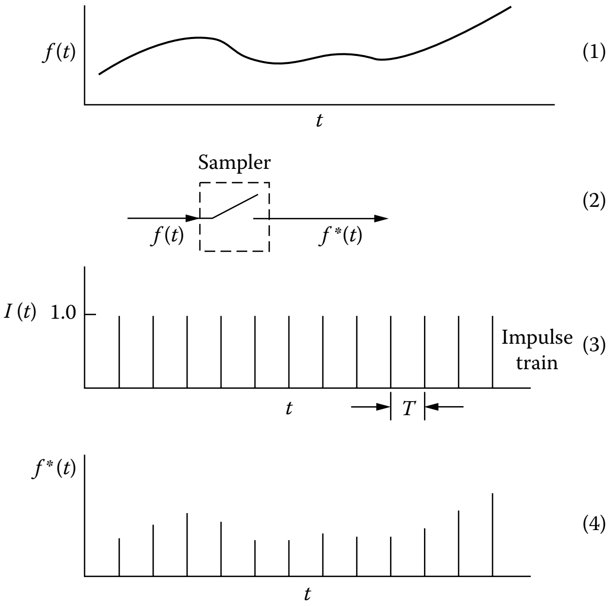


FIG. 2.32b
Sampling of continuous signals.

The sampling is represented mathematically as

$$\begin{aligned} I(t) &= U'(t-T) + U'(t-2T) + U'(t-3T) + \cdots \\ &\quad + U'(t-kT) + \cdots \end{aligned} \quad \textbf{2.32(43)}$$

$$= \sum_{n=0}^{+\infty} U'(t-nT) \quad \textbf{2.32(44)}$$

where U' is the unit impulse. We may now define $f^*(t)$ as the product of the input signal and the switching function:

$$f^*(t) = f(t)I(t) \quad \textbf{2.32(45)}$$

Substituting Equation 2.32(44) into Equation 2.32(45) we obtain

$$f^*t = \sum_{n=0}^{\infty} f(nT)U'(t-nT) \quad \textbf{2.32(46)}$$

The Laplace transformation of both sides leads to

$$f^*(s) = \sum_{n=0}^{\infty} f(nT)e^{-nTs} \quad \textbf{2.32(47)}$$

The form of this equation is not particularly convenient for the study of sampled data control systems. Let us therefore introduce a new variable z, which is defined as

$$z = e^{Ts} \quad \textbf{2.32(48)}$$

Then by substitution into Equation 2.32(47) we obtain the z transform:

$$f^*(s) = \sum_{n=0}^{\infty} f(nT)z^{-n} \quad \textbf{2.32(49)}$$

$$= F(z) \quad \textbf{2.32(50)}$$

$F(z)$ is seen to be an infinite series in powers of z^{-1}.

The following three theorems are useful in z transforms:

1. The shifting theorem: If $z[f(t)] = F(z)$ then

$$z[f(t+T)] = z[F(Z) - f(0)] \quad \textbf{2.32(51)}$$

2. The initial value theorem:

$$\lim_{t\to 0} f^*(t) = \lim_{z\to\infty} f(z) \quad \textbf{2.32(52)}$$

3. The final value theorem:

$$\lim_{t\to 0} f^*(t) = \lim_{z\to 1}\left(\frac{z-1}{z}\right)F(z) \quad \textbf{2.32(53)}$$

If the sampling frequency is at least twice as high as the highest frequency of interest in the continuous input signal, the information contained in the *z* transform of the sampler output is equivalent to that contained in the Laplace transform of the sampler output.

To put it another way, if *T* is the sampling period, then a complete reconstruction of the input signal can be obtained from the sampler output signal for input sinusoidal frequencies up to $\omega = (2\pi/T)1/2 = \pi/T$ radians/unit time.

As was done with the Laplace transform pairs in Table 2.32a, a partial list of the *z* transforms and inverse *z* transform pairs is tabulated in Table 2.32c.

First-Order Lag

Solve the following first-order linear difference equation, given $w(0) = -1$:

$$w^*(t+T) + 2w^*(t) = t^* \qquad \textbf{2.32(54)}$$

Taking the *z* transformation of Equation 2.32(54) one obtains

$$z[W(z)+1] + 2W(z) = \frac{Tz}{(z-1)^2}$$

or

$$(z+2)\,W(z) = \frac{Tz}{(z-1)^2} - z$$

from which $W(z)$ can be solved:

$$W(z) = \frac{Tz}{(z+2)(z-1)^2} - \frac{z}{z+2} \qquad \textbf{2.32(55)}$$

To find the inverse transform of the first term, it is expanded (without the constants and the *z* in the numerator) in partial fractions:

$$\frac{1}{(z+2)\,(z-1)^2} = \frac{K_1}{z+2} + \frac{K_{22}}{(z-1)^2} + \frac{K_{21}}{z-1} \qquad \textbf{2.32(56)}$$

Using the previously applied method one finds

$$K_1 = \left.\frac{1}{(z-1)^2}\right|_{z=-2} = \frac{1}{9},$$

$$K_{22} = \left.\frac{1}{z+2}\right|_{z=1} = \frac{1}{3},$$

$$K_{21} = \left[\frac{d}{dz}\left(\frac{1}{z+2}\right)\right]_{z=1} = -\frac{1}{9}$$

Substituting these constants in Equation 2.32(56), one can now write Equation 2.32(55) in its partial-fraction form:

$$W(z) = \frac{T}{9}\left[\frac{z}{z+2} + \frac{3z}{(z-1)^2} - \frac{z}{z-1}\right] - \frac{z}{z+2} \qquad \textbf{2.32(57)}$$

Finally, one obtains, using transform Table 2.32c:

$$w^*(t) = \frac{T}{9}\left[\left(1-\frac{9}{T}\right)(-2)^{t/T} + 3\frac{t}{T} - 1\right]^*$$

$$= \frac{T}{9}\sum_{n=0}^{\infty}\left[\left(1-\frac{9}{T}\right)(-2)^n + 3n - 1\right]\delta(t-nT) \qquad \textbf{2.32(58)}$$

The solution in Equation 2.32(58) is given in the form of a train of impulses of varying strength.

TABLE 2.32c
z Transform Pairs

Transform F(z)	*Transform f(t)*
1	$\delta(t)$: Unit impulse
$\frac{z}{z-1}$	$u(t)$: Unit Step
$\frac{Tz}{(z-1)^2}$	t: Unit Ramp
$\frac{T^2 z(z+1)}{(z-1)^3}$	t^2
$\frac{z}{z-e^{aT}}$	e^{at} : Exponential
$F(e^{aT}z)$	$e^{-at}f(t)$
$\frac{z \sin bT}{z^2 - 2z\cos bT + 1}$	$\sin bt$
$\frac{z(z-\cos bT)}{z^2 - 2z\cos bT + 1}$	$\cos bt$
$\frac{z \sinh bT}{z^2 - 2z\cosh bT + 1}$	$\sinh bt$
$\frac{z(z-\cosh bT)}{z^2 - 2z\cosh bT + 1}$	$\cosh bt$
$z[F(z) - f(0)]$	$f(t+T)$
$z^2[F(z) - f(0)] - zf(T)$	$f(t+2T)$
$z^{-n}F(z)$	$f(t-nT)U(t-nT)$

Inverse *z* Transform

There are two basic ways of performing the inverse *z* transformation. The first of these is analogous to the inverse Laplace transformation (i.e., partial fractions). The *z* transform, $F(z)$, is expressed in partial-fraction terms, each sufficiently simple that it may be found in transform tables.

The other method involves expressing $F(z)$ as a sequence of z^{-1}, by dividing the denominator of $F(z)$ into the numerator. Then

$$F(z) = a_0 + \frac{a_1}{z} + \frac{a_2}{z^2} + \frac{a_3}{z^3} + \frac{a_4}{z^4} + \cdots \frac{a_n}{z^n} + \cdots \qquad \textbf{2.32(59)}$$

$$\frac{1}{z^k} = e^{-kTs} \qquad \textbf{2.32(60)}$$

$$L^{-1}\left(\frac{1}{z^k}\right) = U'(t - kT) \qquad \textbf{2.32(61)}$$

$$\begin{aligned} f^*(t) &= a_0 U'(0) + a_1 U'(t - T) \\ &\quad + a_2 U'(t - 2T) + \cdots \end{aligned} \qquad \textbf{2.32(62)}$$

This method therefore leads to an intermittent function, a series of impulses at the sampling instants. $f^*(t)$ may therefore be written $f^*(nT)$. Note that the sequence $f^*(t)$ or $f^*(nT)$ may or may not be convergent and that the coefficients a_1, a_2, … a_n are the amplitudes of $f(t)$ at $t = 0$, t, $2T$, …, nT.

As an example, find

$$Z^{-1}\left[\frac{Tz}{(z-1)^2}\right]$$

Before we perform the long division, it is advantageous to arrange both the numerator and the denominator in ascending powers of z^{-1}:

$$\begin{aligned} \frac{Tz}{(z-1)^2} &= T\frac{z^{-1}}{(1 - z^{-1})^2} = T\frac{z^{-1}}{1 - 2z^{-1} + z^{-2}} \\ &= T(z^{-1} + 2z^{-2} + 3z^{-3} + \cdots + kz^{-k} + \cdots) = \sum_{n=0}^{\infty}(nT)z^{-n} \end{aligned} \qquad \textbf{2.32(63)}$$

Hence

$$f^*(t) = Z^{-1}\left[\frac{Tz}{(z-1)^2}\right] = \sum_{n=0}^{\infty}(nT)\,\delta(t - nT) \qquad \textbf{2.32(64)}$$

STATE SPACE REPRESENTATION

Application and design of control systems are predicated on the description of the "process" by means of difference and/or differential equations. Here, state space notation (discussed further in Section 2.33) is introduced as a conventional formulation for representing these mathematical systems.

State space notation allows nth-order difference/differential equations and coupled sets of difference/differential equations to be expressed as vector–matrix equations. This then enables these dynamic systems to be manipulated, transformed, and studied by means of simple linear algebraic procedures. State space representations for simple second-order systems and for 2 × 2 (two inputs, two outputs) multivariable systems will also be developed and demonstrated in the following paragraphs.

Vector and Matrix Operations

The reader with no background in this area is referred to Reference 1 because the material presented here requires an understanding of definitions and operations pertinent to the application of vector and matrix methods in modern control. In the equations below, the vector quantities are underscored and matrices are designated by capital letters.

Second-Order Model

A general continuous state space model is described by a system of linear differential equations of the first order.

$$\text{state equation } \underline{\dot{x}}(t) = A\underline{x}(t) + B\underline{u}(t) \qquad \textbf{2.32(65)}$$

$$\text{output equation } \underline{y}(t) = C\underline{x}(t) \qquad \textbf{2.32(66)}$$

$\underline{x}(t) - n \times l$ state variables

$\underline{u}(t) - m \times l$ manipulated variables

$\underline{y}(t) - j \times 1$ output variables

$A - n \times n$ state parameter matrix

$B - m \times m$ input parameter matrix

$C - j \times j$ output parameter matrix

The discrete state space formulation is completely analogous.

$$\text{state equation } \underline{x}(n+1) = A\underline{x}(n) + B\underline{u}(n) \qquad \textbf{2.32(67)}$$

$$\text{output equation } \underline{y}(n) = C\underline{x}(n) \qquad \textbf{2.32(68)}$$

These are very simple state space models intended for explanation. Various segmented forms are of more practical use in general application.[2]

The $\underline{u}(t)$ vector is the manipulated or disturbance input. It can be thought of as a list of variables that are used to control or disturb the dynamic process. The $\underline{y}(t)$ vector is the output or measured variable vector. It can be thought of as a list of things one can see or measure about the process.

The $\underline{x}(t)$ vector is the state vector. The state vector is of a dimension consistent with the order of the dynamic process. It can be thought of as the output of the integrators or delay elements in a dynamic block diagram or signal flow diagram. The conversion of a simple second-order difference equation into a discrete state space representation will help illustrate this concept. Consider the following second-order linear difference equation:

$$w(n+2) + b_1 w(n+1) + b_0 w(n) = e(n) \qquad \textbf{2.32(69)}$$

Rearranging Equation 2.32(69):

$$w(n+2) = -b_1 w(n+1) - b_0 w(n) + e(n) \qquad \mathbf{2.32(70)}$$

It is also true from the definition of the time delay that

$$w(n+1) = w(n) \qquad \mathbf{2.32(71)}$$

If one defines the state vector $\underline{x}(n)$ as

$$\underline{x}(n) = \begin{bmatrix} w(n) \\ w(n+1) \end{bmatrix} \qquad \mathbf{2.32(72)}$$

From Equation 2.32(72) it is apparent that

$$\underline{x}(n+1) = \begin{bmatrix} w(n+1) \\ w(n+2) \end{bmatrix} \qquad \mathbf{2.32(73)}$$

If one defines the input vector $\underline{u}(n)$ as

$$\underline{u}(n) = \begin{bmatrix} 0 \\ e(n) \end{bmatrix} \qquad \mathbf{2.32(74)}$$

and defines the output vector $\underline{y}(n)$ as

$$\underline{y}(n) = \begin{bmatrix} w(r+2) \\ 0 \end{bmatrix} \qquad \mathbf{2.32(75)}$$

The state system can now be represented as

$$\underline{x}(n+1) = A\,\underline{x}(n) + \underline{b}\,\underline{u}(n) \qquad \mathbf{2.32(76)}$$

$$\underline{y}(n) = \underline{c}x(n) \qquad \mathbf{2.32(77)}$$

or combining Equations 2.32(70) to 2.32(73):

$$x_1(n+1) = w(\underline{n}+1) = w(\cancelto{x_1(n)}{n}) \qquad \mathbf{2.32(78)}$$

$$x_2(n+1) = w(n+2) = -b_1 w(n \cancelto{x_2(n)}{+} 1) - b_0 w(\cancel{n}) + e(n) \qquad \mathbf{2.32(79)}$$

$$y_1(n) = x_2(n) \qquad \mathbf{2.32(80)}$$

Rewriting yields

$$x_1(n+1) = x_1(n) + 0\,x_2(n) + 0\,e(n) \qquad \mathbf{2.32(81)}$$

$$x_2(n+1) = -b_0 x_1(n) - b_1 x_2(n) + 1\,e(n) \qquad \mathbf{2.32(82)}$$

$$y_1(n) = 0\,x_1(n) + 1\,x_2(n) + 0\,e(n) \qquad \mathbf{2.32(83)}$$

Putting in vector–matrix notation gives the following expression:

$$\underline{x}(n+1) = \underbrace{\begin{bmatrix} 1 & 0 \\ -b_0 & -b_1 \end{bmatrix}}_{A} \underline{x}(n) + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\underline{b}} \underline{e}(n) \qquad \mathbf{2.32(84)}$$

$$\underline{y}(n) = \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\underline{c}} x(n) \qquad \mathbf{2.32(85)}$$

The second-order difference Equation 2.32(79) is now a first-order matrix difference equation of dimension two. This state space formulation is convenient for use in many control analysis methods.

Multiple-Input Multiple-Output

State space notation is particularly useful in representing multivariable dynamic systems. Multivariable implies that the system under consideration has more than one input (disturbance and/or manipulated) and more than one output (measured and/or control). Multiple-input multiple-output systems are typically described by sets of coupled difference or differential equations. Consider the following set of equations:

$$\frac{dL_1}{dt} = f_1(w_1) - f_2(w_2, L_1, L_2) \qquad \mathbf{2.32(86)}$$

$$\frac{dL_2}{dt} = f_3(w_1) - f_4(L_2) \qquad \mathbf{2.32(87)}$$

These equations are coupled because L_2 affects L_1 through f_2 of Equation 2.32(86) and L_1 affects both L_1 and L_2 through f_1 and f_3.

Assume that this coupled set of equations can be linearized (see discussion later on linearization). The linearized versions of Equations 2.32(86) and 2.32(87) are given below.

$$\frac{dL_1}{dt} = a_1 L_1 + a_2 L_2 + b_1 w_1 + b_2 w_2 \qquad \mathbf{2.32(88)}$$

$$\frac{dL_2}{dt} = a_3 L_1 + a_4 L_2 + b_3 w_1 + b_4 w_2 \qquad \mathbf{2.32(89)}$$

Define

$$\underline{x} = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix}, \quad \underline{u} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \quad \dot{\underline{x}} = \begin{bmatrix} \frac{dL}{dt}1 \\ \frac{dL}{dt}2 \end{bmatrix} \qquad \mathbf{2.32(90)}$$

Assume one can measure L_1 and L_2 so that

$$\underline{y} = \begin{bmatrix} L_1 \\ L_2 \end{bmatrix} \qquad \mathbf{2.32(91)}$$

The multiple-input multiple-output state space vector–matrix formulation of Equations 2.32(88) and 2.32(89) is

$$\dot{\underline{x}} = \underbrace{\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}}_{A} \underline{x} + \underbrace{\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}}_{B} \underline{u} \qquad \textbf{2.32(92)}$$

$$\underline{y} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{C=I} \underline{x} \qquad \textbf{2.32(93)}$$

TRANSFER FUNCTIONS

A notation used to describe the dynamics of a particular process or system is the transfer function. For a system with an input $r(t)$ and an output $c(t)$, the transfer function $G(s)$ is defined as the ratio of the Laplace transform of the output of the system $C(s)$ divided by the Laplace transform of the input to the system $R(s)$.

$$G(s) = \frac{C(s)}{R(s)} \qquad \textbf{2.32(94)}$$

An inherent assumption in applying the transfer function is that the process is initially at steady state, meaning that

$$\frac{d}{dt}c(0) = \frac{d^2}{dt^2}c(0) = \cdots = \frac{d^n}{dt^n}c(0) = 0.0 \qquad \textbf{2.32(95)}$$

$$\frac{d}{dt}r(0) = \frac{d^2}{dt^2}r(0) = \cdots = \frac{d^n}{dt^n}r(0) = 0.0 \qquad \textbf{2.32(96)}$$

and

$$c(0) = K\,r(0) \qquad \textbf{2.32(97)}$$

where K is the steady-state gain of the process. Under such conditions the Laplace transform of a particular differential equation can be obtained by substituting

$$s \leftrightarrow d/dt(\ ) \qquad \textbf{2.32(98)}$$

$$s^2 \leftrightarrow d^2/dt^2(\ ) \qquad \textbf{2.32(99)}$$

$$s^n \leftrightarrow d^n/dt^n(\ ) \qquad \textbf{2.32(100)}$$

$$1/s \leftrightarrow \int (\ )\,dt \qquad \textbf{2.32(101)}$$

$$1/s^2 \leftrightarrow \int\int (\ )\,dt^2 \qquad \textbf{2.32(102)}$$

$$1/s^n \leftrightarrow \int\int\cdots\int^n (\ )\,dt^n \qquad \textbf{2.32(103)}$$

$$C(s) \leftrightarrow c(t) \qquad \textbf{2.32(104)}$$

$$R(s) \leftrightarrow r(t) \qquad \textbf{2.32(105)}$$

The transfer function can then be determined by solving for $C(s)/R(s)$. Consider the three examples described below.

Second-Order Lag

Consider the following second-order equation:

$$\frac{d^2}{dt^2}c(t) + 2\xi\omega_n\frac{d}{dt}c(t) + \omega_n^2 c(t) = K\,\omega_n^2\,r(t) \qquad \textbf{2.32(106)}$$

After the substitutions:

$$s^2C(s) + 2\xi\omega_n sC(s) + \omega_n^2 C(s) = K\omega_n^2 R(s) \qquad \textbf{2.32(107)}$$

Therefore,

$$\frac{C(s)}{R(s)} = \frac{K\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \qquad \textbf{2.32(108)}$$

PID Controllers

A typical proportional, integral, and derivative (PID) controller is given by

$$m(t) = K_c\left(e(t) + \frac{1}{T_i}\int e(t)\,dt + T_d\frac{d}{dt}e(t)\right) \qquad \textbf{2.32(109)}$$

After the substitutions:

$$M(s) = K_c\left(E(s) + \frac{1}{T_i s}E(s) + T_d s\,E(s)\right) \qquad \textbf{2.32(110)}$$

Therefore,

$$\frac{M(s)}{E(s)} = K_c\left(1 + \frac{1}{T_i s} + T_d s\right) \qquad \textbf{2.32(111)}$$

When using a particlar manufacturer's PID algorithm, it is advisable to review the specification on how the algorithm is implemented. (See Section 2.21.)

Multiple-Input Multiple-Output

Multivariable state space representations can also be provided by transfer functions. They are referred to as *transfer function matrices.* Consider the following state space system:

$$\dot{\underline{x}} = A\underline{x} + B\underline{u} \qquad \textbf{2.32(112)}$$
$$\underline{y} = C\underline{x}$$

Taking the Laplace transform of both equations yields

$$s\underline{x}(s) - \underline{x}(0) = A\underline{x}(s) + B\underline{u}(s) \qquad \textbf{2.32(113)}$$

$$\underline{y}(s) = C\underline{x}(s) \qquad \textbf{2.32(114)}$$

Rearranging gives

$$\underline{x}(s) = (sI - A)^{-1}\,\underline{x}\,(0) + (sI - A)^{-1}\,B\,\underline{u}(s) \qquad \textbf{2.32(115)}$$

and

$$\underline{y}(s) = C(sI - A)^{-1}\,\underline{x}\,(0) + C(sI - A)^{-1}\,B\,\underline{u}(s) \qquad \textbf{2.32(116)}$$

The transfer function matrix is defined as

$$G(s) = C(sI - A)^{-1}\,B \qquad \textbf{2.32(117)}$$

Consider the 2 × 2 state space system in Equations 2.32(92) and 2.32(93).

$$(sI - A) = \begin{bmatrix} s & o \\ o & s \end{bmatrix} - \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} s - a_1 & -a_2 \\ -a_3 & s - a_4 \end{bmatrix} \qquad \textbf{2.32(118)}$$

$$(sI - A)^{-1} = \begin{bmatrix} s - a_4 & a_2 \\ a_3 & s - a_1 \end{bmatrix} \Big/ \text{C.E.} \qquad \textbf{2.32(119)}$$

$$\underbrace{S^2 - (a_1 + a_4)s + (a_1 a_4 - a_2 a_3)}_{\text{Characteristic Equation (C.E.)}} \qquad \textbf{2.32(120)}$$

if $C = I$ then

$$G(s) = C(sI - A)^{-1}\,B =$$

$$= \begin{bmatrix} \dfrac{(s - a_4)b_1 - a_2 b_3}{\text{C.E.}} & \dfrac{(s - a_4)b_2 - a_2 b_4}{\text{C.E.}} \\[2ex] \dfrac{(s - a_1)b_3 - a_3 b_1}{\text{C.E.}} & \dfrac{(s - a_1)b_4 - a_3 b_2}{\text{C.E.}} \end{bmatrix} \qquad \textbf{2.32(121)}$$

Consider element 1,1 of $G(s)$:

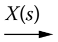
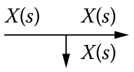
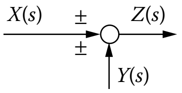
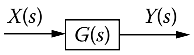
$$G_{11}(s) = \frac{b_1(s - (a_4 b_1 + a_2 b_3))}{s^2 - (a_1 + a_4)s + (a_1 a_4 - a_2 a_3)}\;\frac{x_1(s)}{u_1(s)} \qquad \textbf{2.32(122)}$$

This is the transfer function that relates $x_1(s)$ and $u_1(s)$ in the single-input single-output case. It has a second-order denominator (or two poles) and a first-order numerator (or one zero), as discussed later in this section in connection with stability.

BLOCK DIAGRAMS

In control system analysis, it is not convenient to use the various mathematical equations, while a block diagram format of representation can give further insight. Such diagrams not only provide an organized description of the flow of information and energy, but they also, in the framework of the transfer function notation, facilitate the simultaneous solution of the differential equations that describe the system.

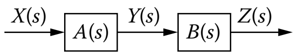
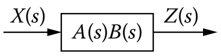
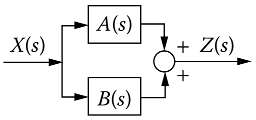
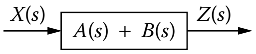
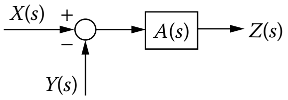
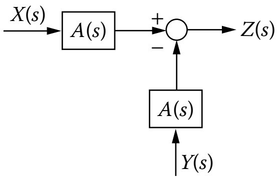
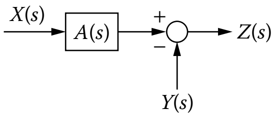
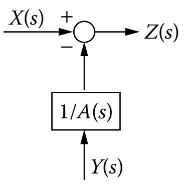
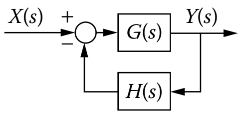
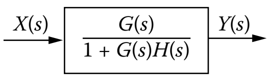
TABLE 2.32d
Symbols Used in Block Diagrams

Symbol	*Interpretation*
	Input or output signal; arrow gives direction.
	Branch point. Division of a signal to give two or more paths without modification.
	Summing point. $Z(s) = \pm X(s) \pm Y(s)$
	System element. $Y(s) = G(s)X(s)$

The standard symbols used in the construction of block diagrams are listed in Table 2.32d. The two main symbols are a circle, which indicates summation of two signals, and a rectangle, which indicates multiplication of a signal by a constant K or by a transfer function $G(s)$. It is important to note that both symbols indicate linear operations.

In using block diagrams to analyze control problems it is important to be able to convert from one form to another. Such manipulations are termed block diagram algebra. Table 2.32e lists some examples of conversions.

TABLE 2.32e
Conversions Used in Block Diagrams

Original Form	*Converted Form*
	
	
	
	
	

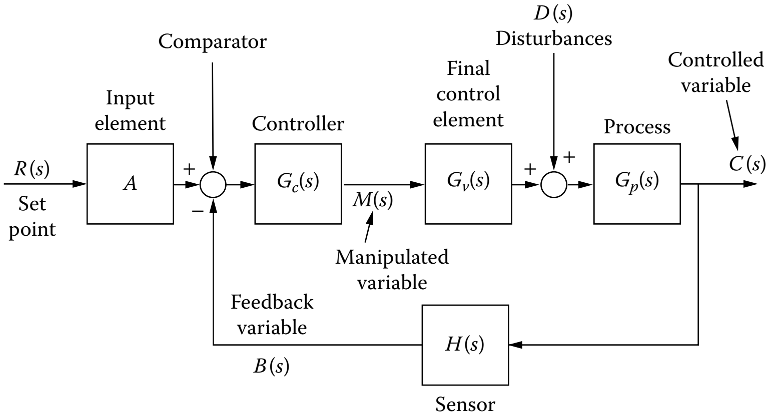


FIG. 2.32f
The block diagram of a typical control loop.

In control analysis the most frequent block diagram manipulations involve feedback loops of some sort. The rules for reducing such systems into a single transfer function can be summarized as follows.

$$\text{Output} = \frac{(\text{Input})\begin{pmatrix}\text{Product of blocks in the forward}\\ \text{path between input signal}\\ \text{and output signal}\end{pmatrix}}{1+\begin{pmatrix}\text{Product of blocks in the}\\ \text{control loop}\end{pmatrix}} \qquad \textbf{2.32(123)}$$

For systems with more than one input, Equation 2.32(123) becomes

$$\text{Output} = \sum_{i=1}^{N} \frac{(\text{Input-}i)\begin{pmatrix}\text{Product of blocks in the forward}\\ \text{path between input-}i\\ \text{and output signal}\end{pmatrix}}{1+\begin{pmatrix}\text{Product of blocks in the}\\ \text{control loop}\end{pmatrix}} \qquad \textbf{2.32(124)}$$

For example, consider the feedback-type control system in Figure 2.32f, where

$$G_c(s) = K_c \text{ (proportional control)} \qquad \textbf{2.32(125)}$$

$$G_p(s) = \frac{K}{1+\tau s} \text{ (first-order feedback)} \qquad \textbf{2.32(126)}$$

$$H(s) = \frac{K_H}{1+\tau_H s} \text{ (first-order feedback)} \qquad \textbf{2.32(127)}$$

Substituting,

$$c(s) = \frac{\dfrac{AK_cK}{1+\tau s}}{1+K_c\left(\dfrac{K}{1+\tau s}\right)\left(\dfrac{K_H}{1+\tau_H s}\right)}R(s) + \frac{B\left(\dfrac{K}{1+\tau s}\right)}{1+K_c\left(\dfrac{K}{1+\tau s}\right)\left(\dfrac{K_H}{1+\tau_H s}\right)}D(s) \qquad \textbf{2.32(128)}$$

or

$$c(s) = \frac{AK_cK(1+\tau_H s)}{(1+\tau s)(1+\tau_H s)+K_cKK_H}R(s) + \frac{BK(1+\tau_H s)}{(1+\tau s)(1+\tau_H s)+K_cKK_H}D(s) \qquad \textbf{2.32(129)}$$

LINEARIZATION

The great majority of techniques used in the analysis of control problems, including block diagrams, are dependent on the existence of mathematical equations in a linear form. By their very nature, most systems are nonlinear to one degree or another and therefore must be approximated by some linear equation. The technique used to linearize a nonlinear function can best be illustrated by considering a nonlinear equation such as

$$Y = \phi(X_1, X_2, X_3, \ldots) \qquad \textbf{2.32(130)}$$

where Y = dependent variable; X = independent variables; and ϕ = nonlinear function.

The equation can be expanded about a point $(X_{1i}, X_{2i}, X_{3i}, \ldots)$ using a Taylor's series expansion in which the higher-order terms are ignored.

$$Y = Y_i + \left.\frac{\delta Y}{\delta X_1}\right|_i (X_1 - X_{1i}) + \left.\frac{\delta Y}{\delta X_2}\right|_i (X_2 - X_{2i}) + \left.\frac{\delta Y}{\delta X_3}\right|_i (X_3 - X_{3i}) + \cdots \tag{2.32(131)}$$

or

$$y = K_1 x_1 + K_2 x_2 + K_3 x_3 + \cdots \tag{2.32(132)}$$

where

$$y = Y - Y_i \tag{2.32(133)}$$

$$x = X - X_i \tag{2.32(134)}$$

$$K_1 = \left.\frac{\delta Y}{\delta X_1}\right|_i = \text{constant} \tag{2.32(135)}$$

$$K_2 = \left.\frac{\delta Y}{\delta X_2}\right|_i = \text{constant} \tag{2.32(136)}$$

$$K_3 = \left.\frac{\delta Y}{\delta X_3}\right|_i = \text{constant} \tag{2.32(137)}$$

For example, consider the development of the transfer function representation for the water tank shown in Figure 2.32g. A material balance on the tank yields

$$\rho Q_{\text{in}} - \rho Q_{\text{out}} = \frac{d}{dt}(\rho H K_s) \tag{2.32(138)}$$

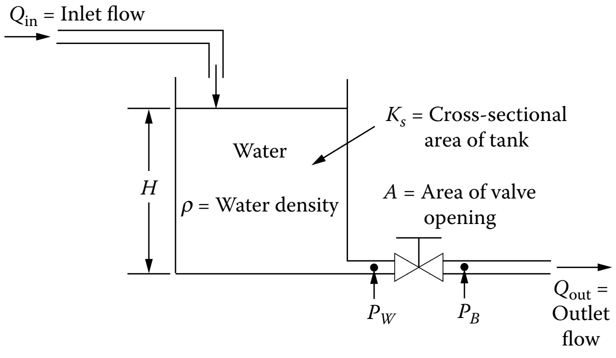


FIG. 2.32g
The operation of a water tank illustrates a nonlinear process.

Flow out of the tank is a function of the area of the valve opening and of the pressure drop across the valve, which can be determined by the following equation:

$$Q_{\text{out}} = C_d A \sqrt{(2g_c/\rho)(P_w - P_B)} \tag{2.32(139)}$$

where ρ = water density; C_d = orifice discharge coefficient (constant); P_B = backpressure (constant); P_w = water pressure at valve; and A = area of valve opening.

The pressure at the valve is caused by the height of water column in the tank:

$$P_w = \rho \frac{g}{g_c} H \tag{2.32(140)}$$

Therefore the valve equation is

$$Q_{\text{out}} = C_d A \sqrt{\frac{2g_c}{\rho}\left(\rho \frac{g}{g_c} H - P_B\right)} \tag{2.32(141)}$$

Substituting into the material balance yields

$$Q_{\text{in}} - C_d A \sqrt{\frac{2g_c}{\rho}\left(\rho \frac{g}{g_c} H - P_B\right)} = \rho K_3 \frac{d}{dt}(H) \tag{2.32(142)}$$

which equation is nonlinear and must be linearized before a Laplace transform can be determined. Using the linearization technique described earlier, the nonlinear equation can be expressed as a linear function about a normal operating point $(Q_{\text{out}i}, Q_{\text{in}i}, A_i, H_i)$ as follows:

$$q_{\text{out}} = K_1 a + K_2 h \tag{2.32(143)}$$

where

$$q_{\text{out}} = (Q_{\text{out}} - Q_{\text{out}i}) \tag{2.32(144)}$$

$$a = (A - A_i) \tag{2.32(145)}$$

$$h = (H - H_i) \tag{2.32(146)}$$

$$K_1 = \left.\frac{\delta Q_{\text{out}}}{\delta A}\right|_i = C_d \sqrt{\frac{2g_c}{\rho}\left(H_i \rho \frac{g}{g_c} - P_B\right)} \tag{2.32(147)}$$

$$K_2 = \left.\frac{\delta Q_{\text{out}}}{\delta H}\right|_i = C_d A_i \sqrt{\frac{2g_c/\rho}{H_i \rho \frac{g}{g_c} - P_B}\left(\rho \frac{g}{g_c}\right)} \tag{2.32(148)}$$

Expressing Q_{in} as a deviation from a reference operating condition, $Q_{\text{in}i}$, the linearized process differential equation is

$$q_{\text{in}} - K_1 a - K_2 h = K_3 \frac{dh}{dt} \tag{2.32(149)}$$

The Laplace transformation of this equation yields:

$$Q_{\text{in}}(s) - K_1 A(s) - K_2 H(s) = K_3 s H(s) \qquad \textbf{2.32(150)}$$

or

$$H(s) = \frac{Q_{\text{in}}(s)}{K_2 + K_3 s} - \frac{K_1 A(s)}{K_2 + K_3 s} \qquad \textbf{2.32(151)}$$

Therefore the two transfer functions that describe the water tank process are

$$\frac{H(s)}{Q_{\text{in}}(s)} = \frac{1}{K_2 + K_3 s} \qquad \textbf{2.32(152)}$$

$$\frac{H(s)}{Q_{\text{in}}(s)} = \frac{K_1}{K_2 + K_3 s} \qquad \textbf{2.32(153)}$$

GRAPHIC REPRESENTATIONS

Analytical solutions to dynamic control problems are typically very tedious. In practice these systems of equations are analyzed using computer-aided graphical techniques. Two of these graphic methods will be discussed here—Bode plots and Nyquist plots. In the subsequent paragraph where stability is discussed, these plotting techniques will prove valuable.

Bode Plots

Bode plots are useful graphical techniques for the analysis of dynamic systems. The Bode plot correlates certain parameters as functions of frequency. Consider the closed-loop equation based on Figure 2.32h.

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)\,H(s)} \qquad \textbf{2.32(154)}$$

If the $r(t)$ input to this system is a sine wave, the steady-state output $c(t)$ will be a sine wave of different amplitude and shifted in phase. If this change in amplitude and phase shift were recorded versus various different frequency $r(t)$ inputs and plotted on rectangular coordinates, these would be the magnitude and phase Bode plots, respectively.

Equation 2.32(154) can be examined in the frequency domain by substituting $s = j\omega$.

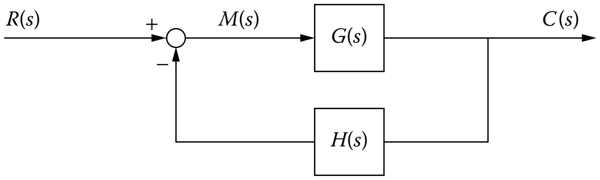


FIG. 2.32h
Feedback control loop.

The frequency response, magnitude versus frequency, for the control ratio

$$\frac{C(j\omega)}{R(j\omega)} = \frac{G(j\omega)}{1 + G(j\omega)H(j\omega)} \qquad \textbf{2.32(155)}$$

can be determined for any given frequency. For each value of frequency, Equation 2.32(155) yields a phasor quantity (see "Nyquist Plots," below) whose magnitude is $C(j\omega)/R(j\omega)$ and whose phase angle α is the angle between $C(j\omega)$ and $R(j\omega)$.

For a given sinusoidal input signal, the input and steady-state output are of the following forms:

$$r(t) = R \sin \omega t \qquad \textbf{2.32(156)}$$

$$c(t) = C \sin(\omega t + \alpha) \qquad \textbf{2.32(157)}$$

An ideal system may be one in which

$$\alpha = 0^\circ$$

$$R = C \qquad \textbf{2.32(158)}$$

for $0 < \omega < \infty$. Curves 1 in Figure 2.32i represent the ideal system. If the above equations are analyzed, it is found that an instant transfer of energy must occur from the input to the output in zero time. This is a prerequisite for faithful reproduction of a step input signal. In reality, this generally cannot be achieved, since in any physical system there is energy dissipation and there are energy-storage elements. Curves 2 and 3 in Figure 2.32i represent actual systems.

The Bode plot provides significant information concerning the time response of the system. Two features of the frequency plot are the maximum value M_m and resonant frequency ω_m. The time response is qualitatively related to these values M_m and ω_m, which can be determined from the frequency response plots.

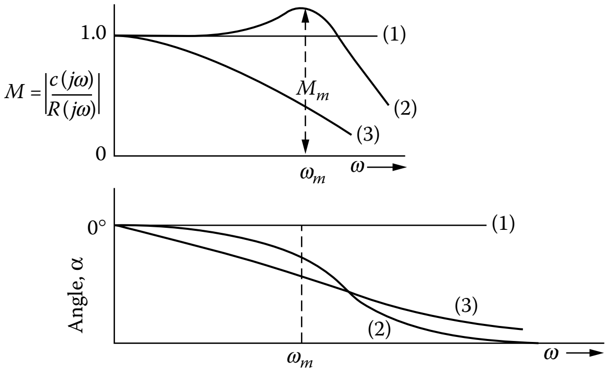


FIG. 2.32i
Bode plots [frequency response characteristic of C(jω)/R(jω) . . in rectangular coordinates].

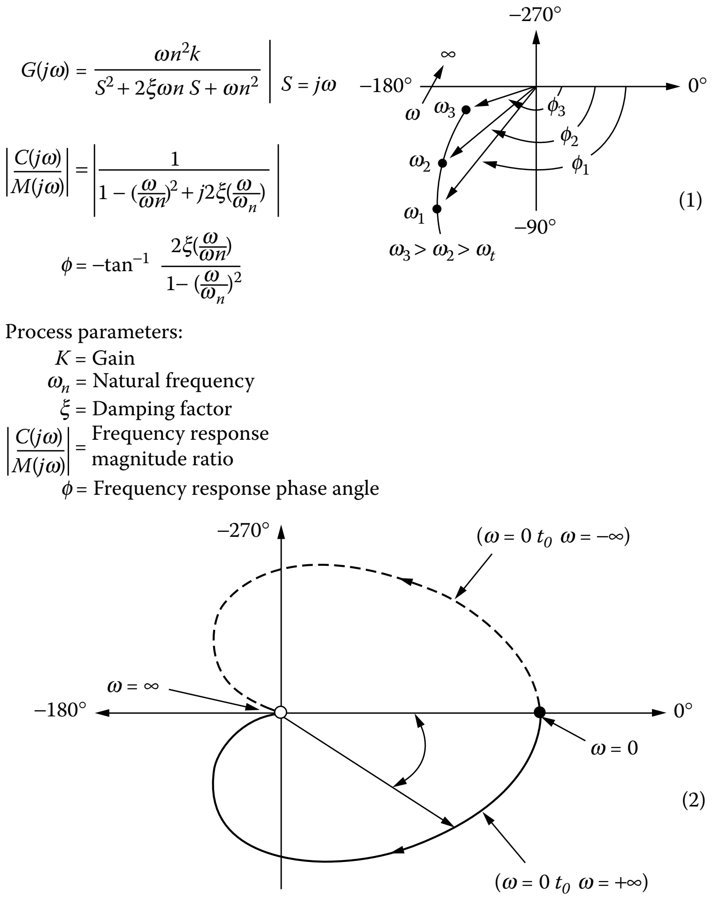


FIG. 2.32j
(1) Phasor-locus development in the complex plane. (2) The direct Nyquist plot.

The magnitude and phase versus frequency information is generated from the phasor quantity. A transfer function can be shown to equal the complex quantity.

$$G(j\omega) = \alpha + j\omega = |G(j\omega)|\, e^{i\theta(\omega)} \qquad \textbf{2.32(159)}$$

$$|G(j\omega)| = \sqrt{\alpha^2 + \omega^2}, \qquad \text{magnitude} \qquad \textbf{2.32(160)}$$

$$\theta(\omega) = \tan^{-1}(\omega/\alpha), \qquad \text{phase} \qquad \textbf{2.32(161)}$$

The quantity typically plotted on the magnitude Bode plot is the decibel, defined as

$$20 \log |G(j\omega)| \qquad \textbf{2.32(162)}$$

Techniques for generating Bode plots vary from quick asymptotic approximation to relatively simple computer packages.

Nyquist Plots

Another method of representing the open-loop steady-state sinusoidal response is the polar plot or Nyquist plot. From this plot, the stability and the frequency response of the closed-loop system can be obtained (see discussion later on stability), and the closed-loop time response of the system can be predicted.

Consider the forward transfer function:

$$|G(j\omega)| = \frac{C(j\omega)}{M(j\omega)} \qquad \textbf{2.32(163)}$$

For a given frequency, $G(j\omega)$ is a phasor quantity whose magnitude is $|C(j\omega)/M(j\omega)|$ and whose phase angle is $\phi(\omega)$. Representing this quantity in polar coordinates can be done by a phasor whose length corresponds to the magnitude vector $|C(j\omega)/R(j\omega)|$ and whose angle with respect to the positive real axis corresponds to $\phi(\omega)$, as shown in Figure 2.32j(1).

Thus a set of values of $G(j\omega)$ for frequencies between 0 and ∞ yields a set of phasors whose tips are connected by a smooth curve, as shown in Figure 2.32j(2). The polar plot in Figure 2.32j(2) contains all the necessary information and is of this form.

Transfer Function Matrix The reader has now been introduced to several forms of graphical representations employed in synthesis work. These techniques, whether computer generated or done by hand, can be very useful in analyzing dynamic systems.

An extension of the Nyquist plot is also very useful in analyzing interactions in multivariable systems. As defined earlier, the transfer function matrix is a matrix of single-input single-output transfer functions representing a multiple-input multiple-output system.

It is possible to take this matrix of transfer functions and plot a matrix of Nyquist plots. Figure 2.32k(1) shows such a matrix for a possible dynamic system described by Equation 2.32(121). This plot alone allows the designer to study the stability and frequency response characteristics of the individual single-input single-output transfer functions.

Another very useful property that can be calculated and displayed graphically is the row dominance property. The dominant transfer function of a given row of a transfer function matrix is the element whose time behavior would dominate (i.e., faster response within a given frequency range). Diagonal row dominance has the following definition for $Q(s)$ [$Q(s) = G(s)^{-1}$ or the inverse transfer function matrix].

$$|q_{ii}(s)| \sum_{\substack{j=1 \\ j \neq 1}}^{m} |q_{ij}(s)| > 0 \qquad \textbf{2.32(164)}$$

This implies that for any s, the modulus of each element on the diagonal of $Q(s)$ exceeds the sum of the moduli of the off-diagonal elements. The row dominance can be graphically illustrated. For each Nyquist plot, at specified frequencies, the modulus of all other same-row elements is calculated. This number is used as the radius for a circle drawn on the Nyquist contours with circle center on the Nyquist contours at the frequency specified [refer to Figure 2.32k(2)]. These circles are called Gershgorin circles.

When this is done for a range of frequencies, the circles sweep out a band called a Gershgorin band. If the origin does not lie within the Gershgorin band, the element is said to be dominant on that row. For example, in Figure 2.32k(2), the g_{11} and g_{21} plots do not have Gershgorin bands that circle the origin; therefore, they are dominant on their respective rows. This matrix is then column 1 dominant, not diagonally dominant as we would like from a decoupling standpoint. Later, compensation will be designed to achieve this dominance.

Multiple-input multiple-output Nyquist arrays bigger than 2 × 2 are very cumbersome to calculate by hand. Computer methods are typically necessary for these systems (see Section 2.38).

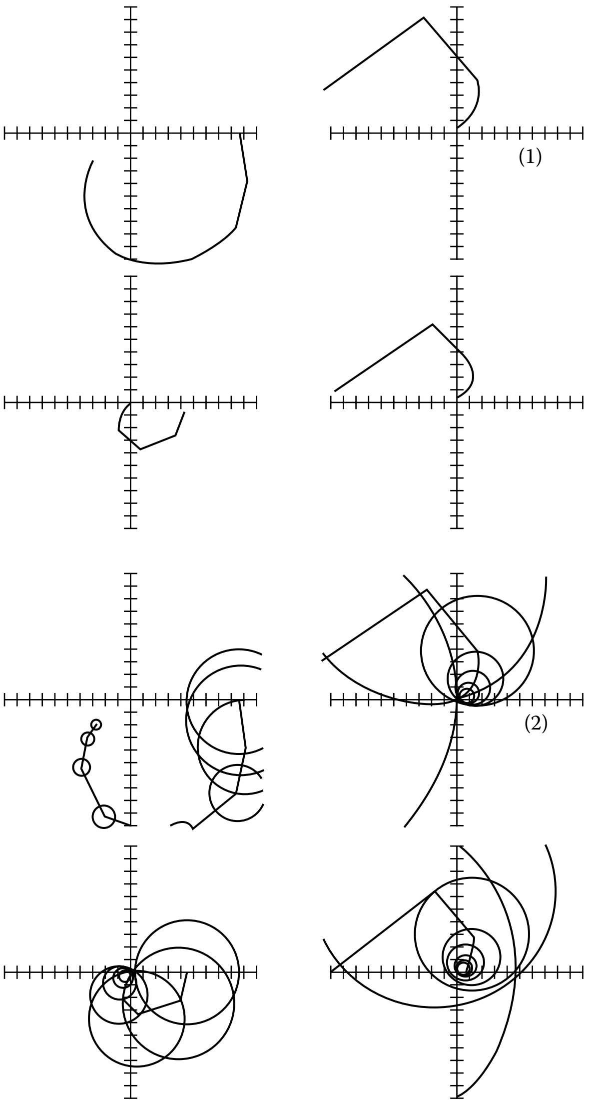


FIG. 2.32k
(1) Nyquist array plot. (2) Nyquist array plot with Gershgorin bands.

STABILITY

A stable linear time-invariant system or element is one in which the system response is always bounded. Any bounded system input is called a BIBO (bounded-input, bounded-output).

While most processes encountered in the process industries (with the exception of a few chemical reactors) are inherently stable, a feedback system employed to control the process can lead to a potentially unstable system. Mathematically, the stability of a linear time-invariant system can be determined by an analysis of the roots of the *characteristic equation* from the differential equation describing the process (which corresponds to the roots of the denominator of the transfer function).

The characteristic equation can be written as follows:

$$\sum_{i=0}^{n} a_i p^i = a_n(s-p_1)(s-p_2)\ldots(s-p_n) = 0 \qquad \textbf{2.32(165)}$$

Consider Figure 2.32l(1). Here the roots of the characteristic equation (for a given K) are on the imaginary axis, and the system is oscillating. In Figure 2.32l(2,3), the roots are to the right of the imaginary axis (i.e., the real part of the root is positive). Both systems are unstable. Their time responses are growing in amplitude. However, in many situations it may not be necessary to actually obtain the exact position of the roots in order to learn the nature of these roots.

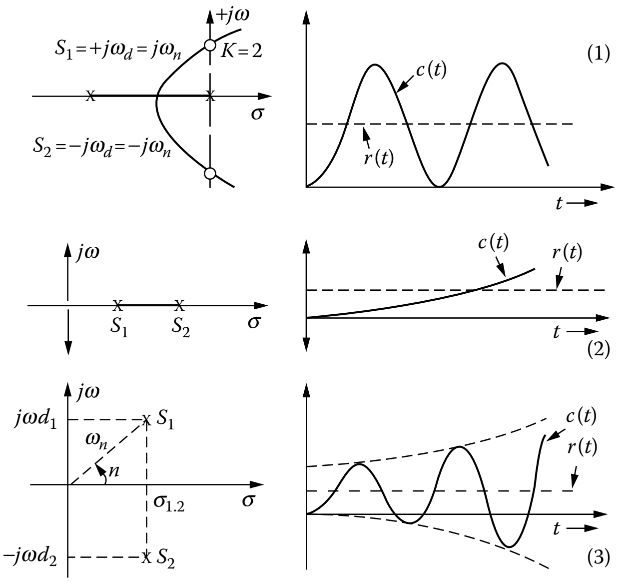


FIG. 2.32l
Roots plotted and associated time response showing (1) continuous oscillations, (2) real root unstable condition, (3) underdamped unstable pole placement.

Descartes' Rule of Signs

A simple rule-of-thumb analysis that can provide much insight into the nature of roots is Descartes' rule of signs:

$$\text{Let } f(s) = a_n s^n + a_{n-1}s^{n-1} + \cdots + a_2 s^2 + a_1 s^1 + a_0 \qquad \textbf{2.32(166)}$$

be a polynomial with real coefficients and $a_0 \neq 0$:

1. The number of positive real zeros of f is either to the number of variations in sign of $f(s)$ or is less than that number by an even integer.
2. The number of negative real zeros of f is either to the number of variations in sign of $f(-s)$ or is less than that number by an even integer.

If the terms in the denominator are arranged in descending powers of s, Descartes' rule of signs states that the number of positive *real* roots cannot exceed the number of variations in sign from term to term.

Therefore a simple *necessary* but *not sufficient* condition of stability that can be immediately assessed for any transfer function is that all the coefficients of each term in the denominator must be the same sign. If any variations in sign exist, the system is unstable. If no variations in sign exist, the system is potentially stable; however, instabilities could result from complex roots in the right-hand plane that Descartes' rule is unable to predict.

For example, for the transfer function

$$F(s) = \frac{3}{(s+3)(s+4)} \qquad \textbf{2.32(167)}$$

the characteristic equation is given by $Q(s) = (s + 3)(s + 4) = 0$, and the roots of the equation, −3, and −4, are in the left half plane. This transfer function is stable. For instance, if one of the roots would have been positive then the system would have been unstable. For complex systems that are stable, the characteristic equation roots would have to be in the imaginary axis and the left half plane.

Routh's Criterion

One absolute method of determining whether complex or real roots lie in the right-hand plane is by the use of Routh's criterion. The method entails systematically generating a column of numbers that are then analyzed for sign variations. The characteristic equation of the polynomial is set to zero. If a transport lag exists, Routh's Criterion cannot be used.

The first step is to arrange the denominator of the transfer function into descending powers of s. All terms including those that are zero should be included.

$$A(s) = a_n s^n + a_{n-1}s^{n-1} + a_{n-2}s^{n-2} + \cdots + a_0 \qquad \textbf{2.32(168)}$$

Next, arrange the coefficients of s according to the following schedule:

$$\begin{array}{ccccc} a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots \\ a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots \end{array} \qquad \textbf{2.32(169)}$$

Now, expand according to the following manner:

$$\begin{array}{c|ccccc} s^n & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdots \\ s^{n-1} & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdots \\ s^{n-2} & b_1 & b_2 & b_3 & b_4 & \cdots \\ s^{n-3} & c_1 & c_2 & c_3 & \cdots & \\ \vdots & : & & & & \\ s^2 & g_1 & g_2 & & & \\ s^1 & h_1 & & & & \\ s^0 & i_1 & & & & \end{array} \qquad \textbf{2.32(170)}$$

The power of s (left of bar) is included to help keep track of were one is located in the array. The equations for the coefficients in the additional rows of above array 2.32(170) are as follows:

$$b_1 = -\frac{1}{a_{n-1}}\begin{vmatrix} a_n & a_{n-2} \\ a_{n-1} & a_{n-3} \end{vmatrix} \qquad b_2 = -\frac{1}{a_{n-1}}\begin{vmatrix} a_n & a_{n-4} \\ a_{n-1} & a_{n-5} \end{vmatrix}, \ldots$$

$$c_1 = -\frac{1}{b_1}\begin{vmatrix} a_{n-1} & a_{n-1} \\ b_1 & b_2 \end{vmatrix} \qquad c_2 = -\frac{1}{b_1}\begin{vmatrix} a_{n-1} & a_{n-5} \\ b_1 & b_3 \end{vmatrix}, \ldots$$

2.32(171)

This process is continued until all new terms are zero. Routh's criterion says that the number of roots (real or complex) of the denominator of the transfer function that lie in the right-hand plane is equal to the number of changes of sign in the leftmost column of the array shown in Equation 2.32(170). If any zeros appear in the first element of any of the nth rows, the transfer function is marginal or instable.

As an example, consider a system described by the following transfer function:

$$g(s) = \frac{s+1}{s^3 + 2s^2 + 5s + 24} \qquad \textbf{2.32(172)}$$

An initial analysis made using Descartes' rule of thumb indicates that, since all the signs of the denominator are the same, the system appears to be stable; however, Routh's criterion must be applied in order to determine this conclusively. The Routh array is

$$\begin{array}{c|ccc} s^3 & 1 & 5 & 0 \\ s^2 & 2 & 24 & 0 \\ s^1 & -7 & 0 & \\ s^0 & 24 & 0 & \end{array} \qquad \textbf{2.32(173)}$$

An inspection of the leftmost column indicates two sign changes (one from +2 to –7 and the other from –7 to +24); therefore, there are two roots in the right-half plane, resulting in an unstable system.

Nyquist Criterion

In the process industries, the Routh criterion has the limitation of not being applicable to systems containing dead time. Another more general stability criterion is the Nyquist criterion. This is based on the "frequency response" of a process and can be easily applied to dead time or to other distributed parameter effects.

To demonstrate the use of the Nyquist criterion, consider the Nyquist plot diagram of Figure 2.32j(2). From this plot the stability can be determined by an investigation of the $s = -1$ point in the complex plane. The Nyquist stability criterion: Let $G(s)H(s)$ be the ratio of two polynomials in s. If $G(s)H(s)$ is within and on C, and if there are no poles or zeros on C, then

$$N = z - P \qquad \textbf{2.32(174)}$$

where N = net number of encirclements of point $s = -1$ in a clockwise direction; z = number of zeros of $G(s)H(s)$ that lie in the right-hand plane; and P = number of poles of $G(s)H(s)$ that lie in the right-hand plane.

The system is unstable if the contour either encircles or passes through the (–1, 0) point in a clockwise direction.

CONCLUSIONS

While the analysis of transfer functions seems to be a mathematics course for process control systems, the possession of the knowledge to deal with systems in both the time and frequency domain allows one to understand what is happening in those systems. Using the stability analysis techniques described in this section will allow the reader to design stable control systems.

References

1. Hoffman, K., and Kunze, R., *Linear Algebra,* Englewood Cliffs, NJ: Prentice Hall, 1961.
2. Geld, A., *Applied Optimal Estimation,* Cambridge, MA: MIT Press, 1974.
3. Zill, D., and Cullen, M., *Differential Equations with Boundary-Value Problems,* 3rd ed., Boston: PWS-Kent, 1993.

Bibliography

Buckley, P. S., "Dynamics of Pneumatic Control by Pattern Recognition," *Instruments and Control Systems,* March 1970.

Buckley, P. S., "A Modern Perspective on Controller Tuning," Texas A&M 30th Annual Symposium on Instrumentation for the Process Industries, January 1973.

Buckley, P. S., *Techniques of Process Control,* New York: John Wiley & Sons, 1964.

Campbell, D. P., *Process Dynamics,* New York: John Wiley & Sons, 1958.

Chen, C. T., *Introduction to Linear System Theory,* New York: Holt, Rinehart and Winston, 1970.

Dutton, K., Thompson, S., and Barraclough, B., *The Art of Control Engineering,* Essex/London: Addison Wesley Longman, 1997.

Harriott, P., *Process Control,* New York: McGraw-Hill, 1964.

ISA-S26 Standard, "Dynamic Response Testing of Process Control Instrumentation," 1968.

Joseph, B., et al., "Knowledge Acquisition and Knowledge Representation in an Expert System for Control of a Batch Process," Paper Presented at AIChE National Meeting, Houston, 1991.

Kecman, V., *State-Space Models of Lumped and Distributed Systems,* Berlin: Springer-Verlag, 1988.

Koening, D. M., *Control and Analysis of Noisy Processes,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Looney, R., "A Thermal Sine Wave Generator for Speed of Response Studies," ASME Paper 54-5A-38, 1954.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

McMillan, G., *Tuning and Control Loop Performance,* 2nd ed., Research Triangle Park, NC: ISA, 1990.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1991.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* 3rd ed., Englewood Cliffs, NJ: Prentice Hall, 1996.

Randhawd, R. S., “Understanding the ‘Lazy’ Integral,” *Instruments and Control Systems,* January 1983.

Rijnsdorp, J. E., MacGregor, J. F., Tyreus, B. D., and Takamatsu, T., *Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes,* New York: Pergamon Press, 1990.

Roffel, B., and Chin, P., *Computer Control in the Process Industries,* Chelsea, MI: Lewis Publishers, 1989.

Rosenbrock, H. H., *Computer-Aided Control System Design,* London: Academic Press, 1974.

Seborg, D., et al., *Process Dynamics and Control,* New York: Wiley, 1989.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

2.33 State Space Control

I. VAJK (2005)

Please note that in this section the language of control system is used, which differs somewhat from the terminology used by engineers working in the process industries. Therefore it might be of value to note that when the term *system* appears in this text, it refers to the process; when *system output* is used, that term refers to the controlled variable $y(t)$; when *system input* is referred to, that term is equivalent to the manipulated variable $u(t)$; and when the term *reference signal* is used, it, in the everyday PID terminology, means set point.

INTRODUCTION

Using state space control, the poles of the system to be controlled can be placed to arbitrary positions. To achieve a satisfactory dynamic system behavior, the roots of the closed-loop characteristic equations are to be assigned properly. The predefined system behavior can be achieved by applying proportional feedback for all the state variables of the system.

The controller given by state variable feedback is realizable if and only if the system is controllable. The control strategy requires access to all the state variables of the system. In practice, however, this condition is rarely satisfied by most systems. This implies that the unmeasured state variables are supposed to be estimated. This is achievable if and only if the system is observable.

Consequently, the complete control strategy will turn out to be a combination of a state observer and a state feedback, where the unmeasured state variables will be substituted by their estimated values. The state space control techniques using both state variable estimation and feedback can be converted to traditional feedback control schemes realized by transfer functions.

STATE SPACE DESCRIPTION

Linear dynamic systems can be described in a number of mathematical forms. One of these is the *impulse response* description. Assuming that the impulse response is available, the system output [in process control, the controlled variable $y(t)$] at an arbitrary time can be determined, if up to that time the history of the system input [in process control, the manipulated variable $u(t)$] is known. [The controlled variable $y(t)$—system output—can be calculated by evaluating a convolution integral.]

Another way to describe a linear dynamic process is to introduce the use of state variables. In this case there is no need to know how the system developed to arrive to a certain state at a given time instant; the system output can be calculated by using the instantaneous value of the state variables and that of the system input. When selecting state variables in a system, it is reasonable to choose variables that are unable to exhibit abrupt changes.

Consider the state space model (or in short, *state model*) of a system given by

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)\\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t)\end{aligned} \qquad \mathbf{2.33(1)}$$

where $\mathbf{x}(t)$ is the state vector with n entries, while $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are the system input and output signals, respectively.

In the case of single-input single-output (SISO) systems both $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are scalar quantities. The $\mathbf{A}$(n by n) and $\mathbf{B}$(n by 1) matrices in Equations 2.33(1) will determine how the state variables develop from an initial state if a given input is applied to the system. The matrices $\mathbf{C}$(1 by n) and $\mathbf{D}$(1 by 1) tell how to find the system output once the state variables and the input signal are available for the calculations.

The evaluation of Equations 2.33(1) suggests that the system dynamics is described by n first-order differential equations. Note that Equations 2.33(1) represent an appropriate way to uniformly describe linear systems with multiple inputs and multiple outputs (MIMO systems). Finally observe that the $\mathbf{D}$ matrix allows a direct contribution from the input $\mathbf{u}(t)$ to the output $\mathbf{y}(t)$. In all real processes it takes some time for a change in the manipulated variable to have an effect on the controlled variable. This delay between the input and the output results in $\mathbf{D} = \mathbf{0}$.

A state model is not unique in the sense that several sets of $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ exist to ensure input/output equivalence. Equivalent state models can be linked together by using a simple linear transformation, called similarity transformation. Assume that $\mathbf{T}$ (n by n) is an arbitrary nonsingular matrix, then $\mathbf{x}(t)$ can be transformed to a new state representation $\mathbf{z}(t)$ by

$$\mathbf{z}(t) = \mathbf{T}\mathbf{x}(t). \qquad \mathbf{2.33(2)}$$

The transformed state model can simply be derived as

$$\begin{aligned}\dot{\mathbf{z}}(t) &= \mathbf{TAT}^{-1}\mathbf{z}(t) + \mathbf{TBu}(t)\\ \mathbf{y}(t) &= \mathbf{CT}^{-1}\mathbf{z}(t) + \mathbf{Du}(t)\end{aligned} \qquad \mathbf{2.33(3)}$$

Among the state models several canonical forms exist (e.g., controllability, observability, diagonal forms). System controllability and observability are important features attached to state models:

- A system is controllable if there exists a finite control input to govern the system from an arbitrary initial state to an arbitrary final state within a finite time horizon. The ability to control (controllability) can be checked by considering the rank of the controllability matrix

  $$\mathbf{Q} = [\mathbf{B} \quad \mathbf{AB} \quad \mathbf{A}^2\mathbf{B} \quad \ldots \quad \mathbf{A}^{n-1}\mathbf{B}] \qquad \mathbf{2.33(4)}$$

  The system is controllable if and only if the controllability matrix is of full rank (rank is n). In case of MIMO systems the controllability matrix is not a square matrix, so one way to check if the rank of **Q** is n is to consider $\mathbf{QQ}^T$, which is an $n \times n$ matrix. If $\mathbf{QQ}^T$ is not singular it indicates that **Q** is of full rank and the system is controllable.
- A system is observable if from knowing the input and output records of a system for a finite time period the initial value of the state variables can be calculated. A system is observable if and only if the observability matrix

  $$\mathbf{M} = \begin{bmatrix}\mathbf{C}\\ \mathbf{CA}\\ \ldots\\ \mathbf{CA}^{n-1}\end{bmatrix} \qquad \mathbf{2.33(5)}$$

  is of full rank (n). Again, for MIMO systems the singularity of $\mathbf{M}^T\mathbf{M}$ can alternatively be checked.

Equations 2.33(1) exhibit the fundamental description of SISO time-invariant linear dynamic systems. Using the Laplace transformation we obtain:

$$\begin{aligned}s\mathbf{x}(s) - \mathbf{x}(t_0) &= \mathbf{Ax}(s) + \mathbf{Bu}(s)\\ \mathbf{y}(s) &= \mathbf{Cx}(s) + \mathbf{Du}(s)\end{aligned} \qquad \mathbf{2.33(6)}$$

where $\mathbf{x}(t_0)$ is the initial value of the state vector and s stands for the Laplace operator. Rearranging Equations 2.33(6), the Laplace transform of the output signal can be expressed as follows:

$$\mathbf{y}(s) = (\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D})\mathbf{u}(s) + \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(t_0) \qquad \mathbf{2.33(7)}$$

Observe the term

$$H(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} \qquad \mathbf{2.33(8)}$$

establishing the relation between the input and output signals. $H(s)$ is the well-known *transfer function* of the system. Some algebra leads to

$$H(s) = \frac{\mathbf{C}\,\mathrm{adj}(s\mathbf{I} - \mathbf{A})\mathbf{B}}{\det(s\mathbf{I} - \mathbf{A})} + \mathbf{D} = \frac{B(s)}{A(s)} \qquad \mathbf{2.33(9)}$$

where $B(s)$ and $A(s)$ are both polynomials. It can be shown that for systems that are controllable *and* observable, $B(s)$ and $A(s)$ are relatively prime (often called coprime) polynomials, i.e., polynomials having no common roots. Note that if $\mathbf{D} = \mathbf{0}$ holds, then the order of $B(s)$ is less than the order of $A(s)$. Systems with $\mathbf{D} = \mathbf{0}$ are also called strictly proper systems. The rest of this section is devoted to strictly proper systems.

CONTROL LAW DESIGN

As it was indicated earlier, state variable feedback can ensure arbitrary pole locations for the closed-loop system. In other words, using state variable feedback, the poles of the original system (i.e., the poles of the system with no state variable feedback) can be moved to arbitrarily assigned positions in the complex plane. The question here is how to find the gains along the feedback path to achieve the required pole locations. Assume first that all the state variables are directly available (measured) for feedback. Later on, this condition will be relaxed.

State Variable Feedback

Assume that the feedback is applied around the state model

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{Ax}(t) + \mathbf{Bu}(t)\\ \mathbf{y}(t) &= \mathbf{Cx}(t)\end{aligned} \qquad \mathbf{2.33(10)}$$

by using the linear combination of the state variables according to

$$\mathbf{u}(t) = -\mathbf{K}\,\mathbf{x}(t) + \mathbf{u}_r(t) \qquad \mathbf{2.33(11)}$$

where $\mathbf{u}_r(t)$ is the set point (scalar value in case of SISO systems) of the system and **K** is the gain matrix (essentially a row vector in case of SISO systems). The block diagram of the closed-loop system is shown in Figure 2.33a. Designing a controller is after all a procedure to find appropriate values for the gain matrix. Substituting Equation 2.33(11) back to Equation 2.33(10) results in the following differential equation for the closed-loop system:

$$\dot{\mathbf{x}}(t) = (\mathbf{A} - \mathbf{B}\mathbf{K})\mathbf{x}(t) + \mathbf{B}\,\mathbf{u}_r(t) \qquad \mathbf{2.33(12)}$$

Using the Laplace transforms, we obtain:

$$y(s) = \mathbf{C}(s\mathbf{I}_n - \mathbf{A} + \mathbf{BK})^{-1}\mathbf{B}u_r(s) + \mathbf{C}(s\mathbf{I}_n - \mathbf{A} + \mathbf{BK})^{-1}\mathbf{x}(t_0) \qquad \mathbf{2.33(13)}$$

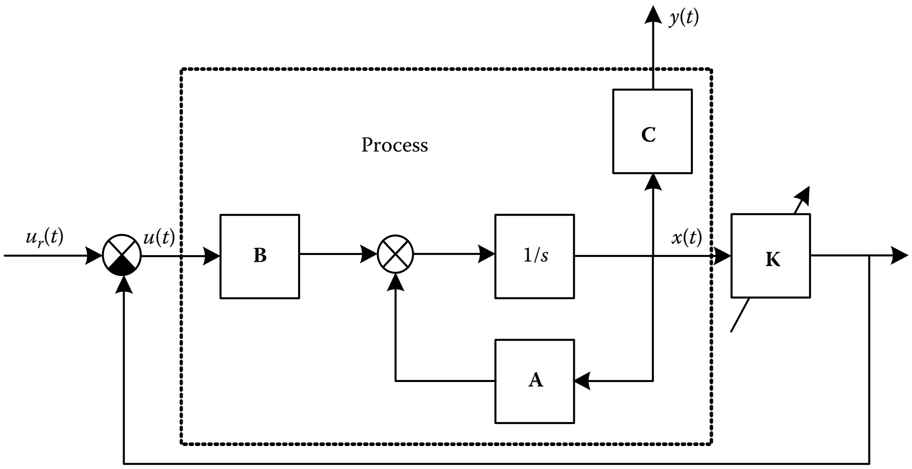


FIG. 2.33a
Controlled process with state variable feedback.

or

$$y(s)=\frac{\mathbf{C}\,\mathrm{adj}(s\mathbf{I}_n-\mathbf{A}+\mathbf{BK})}{\det(s\mathbf{I}_n-\mathbf{A}+\mathbf{BK})}(\mathbf{B}u_r(s)+\mathbf{x}(t_0)) \qquad \mathbf{2.33(14)}$$

The above relation clearly shows that the closed-loop behavior is determined by the closed-loop characteristic polynomial:

$$\alpha(s)=\det(s\mathbf{I}_n-\mathbf{A}+\mathbf{BK}) \qquad \mathbf{2.33(15)}$$

Also, the overall relation between the set point and the output can be expressed by

$$\frac{y(s)}{u_r(s)}=\mathbf{C}(s\mathbf{I}_n-\mathbf{A}+\mathbf{BK})^{-1}\mathbf{B}=\frac{B(s)}{\alpha(s)} \qquad \mathbf{2.33(16)}$$

Ignoring the effect of the nonzero initial value $\mathbf{x}(t_0)$, Equation 2.33(16) shows that the state variable feedback just applied is equivalent to applying a series term $A(s)/\alpha(s)$ directly for the system given by the transfer function $B(s)/A(s)$.

Next, the elements of the **K** gain matrix will be determined such that the roots of the closed-loop characteristic equation are $\{-s_1, -s_2, \ldots -s_n\}$. Using Equation 2.33(15) results in

$$\det(s\mathbf{I}_n-\mathbf{A}+\mathbf{BK})=(s+s_1)(s+s_2)\ldots(s+s_n) \qquad \mathbf{2.33(17)}$$

Then the entries of **K** can be determined by matching the coefficients of the identical powers of s in the polynomials of the two sides in Equation 2.33(17).

Ackermann derived a nice compact algorithm to find **K** directly. Introduce the polynomial

$$\alpha(s)=(s+s_1)(s+s_2)\ldots(s+s_n)=s^n+\alpha_{n-1}s^{n-1}+\cdots+\alpha_0 \qquad \mathbf{2.33(18)}$$

and the matrix polynomial

$$\alpha(\mathbf{A})=\mathbf{A}^n+\alpha_{n-1}\mathbf{A}^{n-1}+\cdots+\alpha_0\mathbf{I} \qquad \mathbf{2.33(19)}$$

The Ackermann formula presents the gain matrix by

$$\mathbf{K}=[0 \quad 0 \quad \ldots \quad 1]\mathbf{Q}^{-1}\alpha(\mathbf{A}) \qquad \mathbf{2.33(20)}$$

where **Q** denotes the controllability matrix. The Ackermann formula shows why the inverse of the controllability matrix should be nonsingular to get a gain matrix. If **Q** is nonsingular the gain matrix obtained by the Ackermann formula is unique.

An even simpler way to determine the coefficients of the gain matrix can be obtained if the state model is assumed to be available in controllability form. Denoting the gain vector associated with the controllability form by $\mathbf{K}_c$, consider a polynomial $K(s)$ with coefficients delivered by $\mathbf{K}_c$. Then, the characteristic polynomial of the closed-loop system simply becomes

$$\alpha(s)=A(s)+K(s) \qquad \mathbf{2.33(21)}$$

Equation 2.33(21) offers a direct way to determine the feedback gain matrix; namely, $K(s)$ can be calculated knowing the goal of the design, i.e., the location of the closed-loop poles.

The presented state variable feedback can also be considered as a generalization of the traditional PD compensation. The generalization is accomplished in the sense that all the closed-loop poles are moved to desired positions. Consequently, all the components of the system response transients are under control. At the same time, according to the fundamental property of the PD compensation, the steady-state error will not go to zero if the system is disturbed by additive constant noise. This becomes only achievable by inserting integrator(s) into the loop.

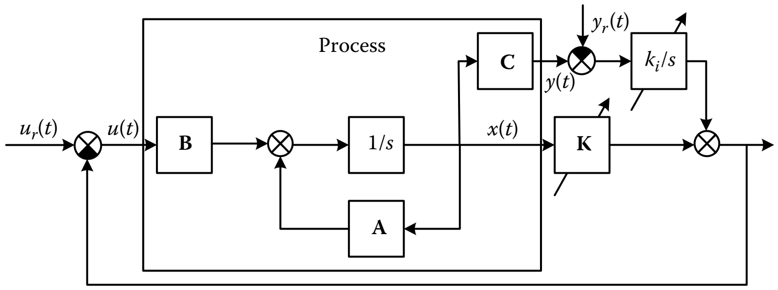


FIG. 2.33b
State feedback with PI action.

State Feedback with PI Control

The state variable feedback concept outlined in the previous section represents a certain generalization of the PD compensation technique. To ensure zero steady-state error, an additional integrator should be inserted into the loop. This can easily be done by augmenting the size of the state vector by one more entry. Denoting the new state variable by $z(t)$ we have

$$\dot{z}(t) = y(t) - y_r(t) \tag{2.33(22)}$$

where $y_r(t)$ denotes the reference signal. The description of the augmented system becomes:

$$\begin{aligned} \dot{\bar{\mathbf{x}}}(t) &= \bar{\mathbf{A}}\bar{\mathbf{x}}(t) + \bar{\mathbf{B}}u(t) + \bar{\mathbf{B}}_r y_r(t) \\ y(t) &= \bar{\mathbf{C}}\bar{\mathbf{x}}(t) \end{aligned} \tag{2.33(23)}$$

where

$$\begin{aligned} &\bar{\mathbf{x}}(t) = \begin{bmatrix} \mathbf{x}(t) \\ z(t) \end{bmatrix} \quad \bar{\mathbf{A}} = \begin{bmatrix} \mathbf{A} & 0 \\ \mathbf{C} & 0 \end{bmatrix} \quad \bar{\mathbf{B}} = \begin{bmatrix} \mathbf{B} \\ 0 \end{bmatrix} \\ &\bar{\mathbf{B}}_r = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \quad \bar{\mathbf{C}} = [\mathbf{C} \quad 0] \end{aligned} \tag{2.33(24)}$$

The state variable feedback will be realized by using the augmented state vector for feedback, according to

$$u(t) = -\bar{\mathbf{K}}\bar{\mathbf{x}}(t) = -\mathbf{K}\mathbf{x}(t) - k_i z(t) = -\begin{bmatrix} \mathbf{K} & k_i \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ z(t) \end{bmatrix} \tag{2.33(25)}$$

Applying the above feedback, the characteristic polynomial of the closed-loop system can be written as

$$\begin{aligned} \bar{\alpha}(s) &= \det(s\mathbf{I}_{n+1} - \bar{\mathbf{A}} + \bar{\mathbf{B}}\bar{\mathbf{K}}) = \det\begin{bmatrix} s\mathbf{I}_n - \mathbf{A} + \mathbf{BK} & \mathbf{B}k_i \\ \mathbf{C} & s \end{bmatrix} \\ &= s\det(s\mathbf{I}_n - \mathbf{A} + \mathbf{BK}) + \mathbf{C}\operatorname{adj}(s\mathbf{I}_n - \mathbf{A} + \mathbf{BK})\mathbf{B}k_i \end{aligned} \tag{2.33(26)}$$

Using the transfer functions introduced earlier, Equation 2.33(26) leads to

$$\bar{\alpha}(s) = s(A(s) + K(s)) + B(s)k_i \tag{2.33(27)}$$

Relation to Equation 2.23(21) is clearly seen from the above form. Given $\bar{\alpha}(s)$, $A(s)$, and $B(s)$, the feedback polynomial $K(s)$ and the k_i coefficient can both be determined.

The block scheme of the state variable feedback extended by the integrator is shown in Figure 2.33b. The augmented system is capable of rejecting nonzero steady-state error. Note that the closed-loop system exhibits identical operational performance for following the reference signal (servo property), as well as for rejecting additive disturbances acting on the system output.

A final remark here relates to the size of the feedback gain matrix. For SISO systems **K** is a vector with n entries, while for MIMO systems **K** is an $m \times n$ matrix, where m denotes the number of inputs. The entries of the gain matrix are adjusted to locate the closed-loop poles; however, there are only n poles to be located, resulting in n equations to determine $m \times n$ coefficients in the gain matrix. Consequently, there exist several gain matrices resulting in identical closed-loop characteristic equations.

OBSERVER DESIGN

So far a state variable feedback technique has been studied assuming that all the state variables are available for the feedback. This condition is rarely met in practice. In the following, how to modify the control strategy derived so far to be able to incorporate the unavailable state variables into the feedback structure will be discussed. The procedure developed to estimate the state variables based on the input/output records of the system is called state estimation or state reconstruction, and the related functional unit is called an *observer.*

Full Order Observer

The state variables are to be estimated by the following linear model

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{Q}\hat{\mathbf{x}}(t) + \mathbf{R}\mathbf{u}(t) + \mathbf{S}\mathbf{y}(t) \qquad \mathbf{2.33(28)}$$

where $\hat{\mathbf{x}}(t)$ denotes the estimated state vector. Formally, designing an observer is equivalent to determining the **Q**, **R**, and **S** matrices. One classical solution of this problem is to use a feedback in the observer itself driven by the deviation between the measured and estimated system outputs:

$$\begin{aligned}\dot{\hat{\mathbf{x}}}(t) &= \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{L}(\mathbf{y}(t) - \hat{\mathbf{y}}(t)) \\ \hat{\mathbf{y}}(t) &= \mathbf{C}\hat{\mathbf{x}}(t)\end{aligned} \qquad \mathbf{2.33(29)}$$

The above expression clearly shows how the estimated system output is derived from the estimated state vector. The related block scheme, also called the *asymptotic state observer,* is shown in Figure 2.33c. The error in the state estimation can be expressed by

$$\tilde{\mathbf{x}}(t) = \hat{\mathbf{x}}(t) - \mathbf{x}(t), \qquad \mathbf{2.33(30)}$$

whose development is described by the following differential equation:

$$\dot{\tilde{\mathbf{x}}}(t) = (\mathbf{A} - \mathbf{L}\mathbf{C})\tilde{\mathbf{x}}(t) + \mathbf{L}\mathbf{y}(t) \qquad \mathbf{2.33(31)}$$

The error in the state estimation develops according to the characteristic polynomial

$$\beta(s) = \det(s\mathbf{I}_n - \mathbf{A} + \mathbf{L}\mathbf{C}) \qquad \mathbf{2.33(32)}$$

Various choices for the gain matrix **L** allow locating the poles of the above characteristic polynomial.

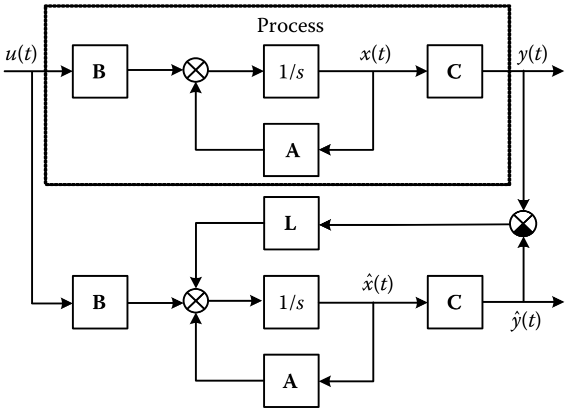


FIG. 2.33c
Block scheme of the full order observer.

Assume working with single output systems. To design the way as the error in the state estimation decays, the poles of the characteristic Equation 2.33(32) should be located to predefined $-s_1, -s_2, \ldots -s_n$ values. To determine the entries of the vector **L**, polynomial equation

$$\det(s\mathbf{I}_n - \mathbf{A} + \mathbf{L}\mathbf{C}) = (s + s_1)(s + s_2)\ldots(s + s_n) \qquad \mathbf{2.33(33)}$$

should be solved. Again, this can be done by matching the coefficients of the left-side and right-side polynomials in a very similar way as the feedback gain vector was determined by Equation 2.33(17) earlier. Introducing the polynomial

$$\beta(s) = (s + s_1)(s + s_2)\ldots(s + s_n) = s^n + \beta_{n-1}s^{n-1} + \cdots + \beta_0 \qquad \mathbf{2.33(34)}$$

and the matrix polynomial

$$\beta(\mathbf{A}) = \mathbf{A}^n + \beta_{n-1}\mathbf{A}^{n-1} + \cdots + \beta_0\mathbf{I} \qquad \mathbf{2.33(35)}$$

the related Ackermann formula takes the following form:

$$\mathbf{L} = \beta(\mathbf{A})\mathbf{M}^{-1}[0 \quad 0 \quad \ldots \quad 1]^T \qquad \mathbf{2.33(36)}$$

where **M** still denotes the observability matrix. Equation 2.33(36) shows that the observer feedback gain vector **L** exists if and only if **M** is invertible. In other words, to be able to apply the Ackermann formula and to have a unique solution for the observer gain vector **L**, the system should be observable.

The structure of the observer shown in Figure 2.33c can be transferred to an equivalent form shown in Figure 2.33d. According to this structure, the observer itself can be interpreted as a servo controller with a reference signal represented by the system output and a controlled signal represented by the estimated output signal. Further on the system parameters are stored in **A**, **B**, and **C**, while the system state variables are the estimated states.

Comparing Figures 2.33a and 2.33d, it is seen that the structure of the state variable control and that of the full order

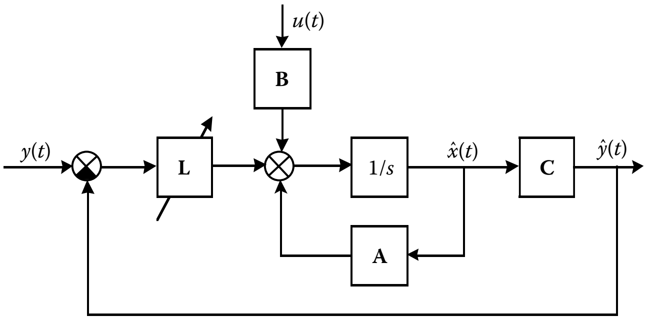


FIG. 2.33d
Full order observer as a servo problem.

observer resemble each other to a large extent. The output signal, as well as the **L** and **C** matrices, play an identical role in the observer as the control signal, as well as the **B** and **K** matrices do in the state variable control. Parameters in matrix **L** and parameters in matrix **K** are to be freely adjusted for the observer and for the state variable control, respectively. In a sense, calculating the controller and observer feedback gain matrices represents dual problems. In this case, duality means that any one of the structures shown in Figures 2.33a and 2.33d can be turned to its dual form by reversing the direction of the signal propagation, interchanging the input and output signals ($u \leftrightarrow y$), and transforming the summation points to signal nodes and vice versa.

The function realized by the observer can also be expressed in the form of a transfer function. Assuming zero initial state variable conditions, the solution of Equation 2.33(29) results in

$$\hat{y}(s) = \mathbf{C}(s\mathbf{I}_n - \mathbf{A} + \mathbf{LC})^{-1}(\mathbf{B}u(s) + \mathbf{L}y(s)) \qquad \mathbf{2.33(37)}$$

Elementary matrix operations used for Equation 2.33(32) lead to a polynomial separation:

$$\beta(s) = \det(s\mathbf{I}_n - \mathbf{A} + \mathbf{LC}) = A(s) + L(s) \qquad \mathbf{2.33(38)}$$

where

$$A(s) = \det(s\mathbf{I}_n - \mathbf{A}) \qquad \mathbf{2.33(39)}$$

and

$$L(s) = \mathbf{C}\mathrm{adj}(s\mathbf{I}_n - \mathbf{A})\mathbf{L} \qquad \mathbf{2.33(40)}$$

Using the observability form for the state model studied, coefficients of the $L(s)$ polynomial are identical to the entries of matrix **L**. Then using the above polynomials, Equation 2.33(37) turns to

$$\beta(s)\hat{y}(s) = (A(s) + L(s))\hat{y}(s) = B(s)u(s) + L(s)y(s) \qquad \mathbf{2.33(41)}$$

For MIMO systems the **L** matrix consists of $n \times p$ entries, where p denotes the number of the system outputs. However, the required location of the observer poles defines only n equations, thus the problem—just as when designing a MIMO state variable feedback control—is underdetermined.

Full Order Observer for Constant Error

It is a well-known fact from classical control theory that the application of proportional controllers ensures finite, nonzero control error when constant set points or constant disturbances are applied. Not surprisingly at all, if the output signal the observer intends to follow contains a constant, nondecaying, nonzero component (typically derived from constant disturbance acting on the input), the state estimation will be biased. The concept of duality suggests that this bias can be eliminated by adding an integrator to the observer circuit.

Consider a system together with an unknown constant additive noise (augmented system):

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}(u(t) + w(t)) \\ \dot{w}(t) &= 0 \\ y(t) &= \mathbf{C}\mathbf{x}(t) \end{aligned} \qquad \mathbf{2.33(42)}$$

Following the technique applied earlier, the state variables of the augmented system can be estimated along the following equations:

$$\begin{aligned} \dot{\hat{\mathbf{x}}}(t) &= \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}(u(t) + \hat{w}(t)) + \mathbf{L}(y(t) - \hat{y}(t)) \\ \dot{\hat{w}}(t) &= l_i(y(t) - \hat{y}(t)) \\ \hat{y}(t) &= \mathbf{C}\hat{x}(t) \end{aligned} \qquad \mathbf{2.33(43)}$$

The characteristic polynomial of the augmented system is

$$\bar{\beta}(s) = \det\begin{bmatrix} s\mathbf{I}_n - \mathbf{A} + \mathbf{LC} & -\mathbf{B} \\ l_i\mathbf{C} & s \end{bmatrix} \qquad \mathbf{2.33(44)}$$

The roots of the characteristic equation can be affected via the feedback elements in **L** and l_i. Using equivalent transformations for the above determinant, Equation 2.33(44) can be separated by

$$\bar{\beta}(s) = s(A(s) + L(s)) + B(s)l_i \qquad \mathbf{2.33(45)}$$

Then in the knowledge of $\bar{\beta}(s)$, $A(s)$, and $B(s)$, the coefficients in $L(s)$, and l_i itself, can be calculated.

Reduced Order Observer

If no significant disturbance is acting on the system output, there is no need to construct an observer providing an error in the estimated output signal converging to zero only asymptotically, rather than having an error in the estimated output signal being constantly zero. It is seen in Figure 2.33d that the feedback is utilizing the error in the estimated output signal and the state estimation is influenced by this error via the linear gain **L**. The error in the estimated output signal is forced to be constantly zero if the state estimation provides

$$\mathbf{y}(t) = \mathbf{C}\hat{\mathbf{x}}(t) \qquad \mathbf{2.33(46)}$$

If the output signals of the system are linearly independent from each other, they can be considered as state variables. Introduce the state vector such that all the output variables show up in the state vector. Then the state vector can be separated according to

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}_r(t) \end{bmatrix} \qquad \mathbf{2.33(47)}$$

where $\mathbf{z}_r(t)$ contains the state variables not measured directly. The transformation leading to the state vector by Equation 2.33(47) can be found by

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}_r(t) \end{bmatrix} = T\mathbf{x}(t) = \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix} \mathbf{x}(t) = \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{Tx}(t) \end{bmatrix} \qquad \mathbf{2.33(48)}$$

The goal now is to find an estimation for the inaccessible $\mathbf{z}_r(t)$ component of the state vector as the state vector of the following linear state model:

$$\dot{\hat{\mathbf{z}}}_r(t) = \mathbf{Q}\hat{\mathbf{z}}_r(t) + \mathbf{Ru}(t) + \mathbf{Sy}(t) \qquad \mathbf{2.33(49)}$$

Find the matrices **Q**, **R**, and **S** in Equation 2.33(49) such that the estimation error

$$\tilde{\mathbf{z}}_r(t) = \hat{\mathbf{z}}_r(t) - \mathbf{z}_r(t) \qquad \mathbf{2.33(50)}$$

asymptotically converges to zero.

The state model using the new (transformed) state vector takes the following form:

$$\begin{aligned} \begin{bmatrix} \dot{\mathbf{y}}(t) \\ \dot{\mathbf{z}}_r(t) \end{bmatrix} &= \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix} \mathbf{A} \begin{bmatrix} \mathbf{H} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}_r(t) \end{bmatrix} + \begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix} \mathbf{Bu}(t) \\ \mathbf{y}(t) &= \mathbf{C}\begin{bmatrix} \mathbf{H} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{y}(t) \\ \mathbf{z}_r(t) \end{bmatrix} \end{aligned} \qquad \mathbf{2.33(51)}$$

where the notation by

$$\begin{bmatrix} \mathbf{C} \\ \mathbf{T} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{H} & \mathbf{F} \end{bmatrix} \qquad \mathbf{2.33(52)}$$

has been applied. Performing the above operations, Equation 2.33(51) becomes

$$\begin{aligned} \dot{\mathbf{z}}_r(t) &= \mathbf{TAFz}_r(t) + \mathbf{TAHy}(t) + \mathbf{TBu}(t) \\ \dot{\mathbf{y}}(t) &= \mathbf{CAFz}_r(t) + \mathbf{CAHy}(t) + \mathbf{CBu}(t) \\ \mathbf{y}(t) &= \mathbf{CFz}_r(t) + \mathbf{CHy}(t) = \mathbf{y}(t) \end{aligned} \qquad \mathbf{2.33(53)}$$

Then, using Equation 2.33(53), the following relation is derived

$$\begin{aligned} \dot{\hat{\mathbf{z}}}_r(t) &= \mathbf{TAF}\hat{\mathbf{z}}_r(t) + \mathbf{TAHy}(t) + \mathbf{TBu}(t) \\ \hat{\mathbf{x}}(t) &= \mathbf{F}\hat{\mathbf{z}}_r(t) + \mathbf{Hy}(t) \end{aligned} \qquad \mathbf{2.33(54)}$$

to estimate the state variables. The functional unit realizing the above state estimation algorithm is called reduced order observer; see Figure 2.33e for the block diagram. The error related to the estimation of the inaccessible state variables can be expressed by

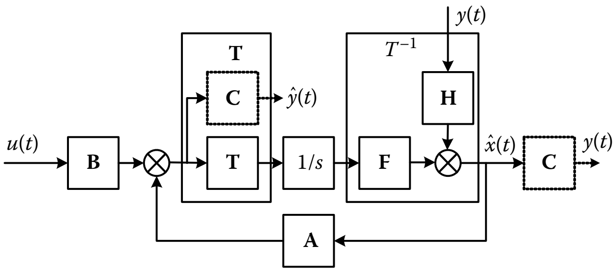


FIG. 2.33e
Block scheme of the reduced order observer.

$$\dot{\tilde{\mathbf{z}}}_r(t) = \mathbf{TAF}\dot{\tilde{\mathbf{z}}}_r(t), \qquad \mathbf{2.33(55)}$$

while the error related to the original state vector can be written as

$$\dot{\tilde{\mathbf{x}}}(t) = \mathbf{FTA}\tilde{\mathbf{x}}(t) = (\mathbf{I}_n - \mathbf{HC})\tilde{\mathbf{x}}(t) \qquad \mathbf{2.33(56)}$$

The initial condition of the original state vector can be calculated by $\tilde{\mathbf{x}}(t_0) = \mathbf{F}\tilde{\mathbf{z}}_r(t_0)$. Selecting appropriate values in **H**—keeping in mind that $\mathbf{CH} = \mathbf{I}_p$ holds—the error by Equation 2.33(56) can arbitrarily be reduced. Note that here *p* denotes the number of the system outputs.

Using the constraint by

$$\mathbf{HC} + \mathbf{FT} = \mathbf{I}_n \qquad \mathbf{2.33(57)}$$

(see Equation 2.33[52]), the reduced order observer shown in Figure 2.33e can equivalently be transformed to the form shown in Figure 2.33f. Note that the observer is using one single matrix (**H**) to reconstruct the state vector. Selecting the elements in matrix **H**, the constraint by $\mathbf{CH} = \mathbf{I}_p$ should be taken into account. The number of the state variables involved in the observer shown in Figure 2.33f is *not* minimal; however, it is operating as a reduced order observer.

Assuming one single output signal and a state model where this single output is one of the state variables, we have

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix}, \qquad \mathbf{2.33(58)}$$

as well as a matrix **H**

$$\mathbf{H} = [1 \quad h_1 \quad \dots \quad h_{n-1}]^T = [1 \quad \tilde{\mathbf{H}}^T]^T \qquad \mathbf{2.33(59)}$$

of special structure. Separating the states as before leads to

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{bmatrix} \qquad \mathbf{2.33(60)}$$

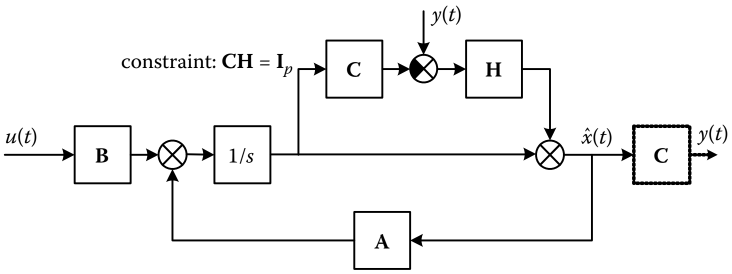


FIG. 2.33f
Reduced order observer with extra state variables.

and the form of the reduced order observer is

$$\dot{\hat{\mathbf{z}}}_r(t) = (\mathbf{A}_{22} - \tilde{\mathbf{H}}\mathbf{A}_{12})(\hat{\mathbf{z}}_r(t) + \tilde{\mathbf{H}}y(t)) + (\mathbf{A}_{21} - \tilde{\mathbf{H}}\mathbf{A}_{11})y(t) + (\mathbf{B}_2 - \tilde{\mathbf{H}}\mathbf{B}_1)\mathbf{u}(t)$$

$$\hat{\mathbf{x}}(t) = \begin{bmatrix} y(t) \\ \hat{\mathbf{z}}_r(t) + \tilde{\mathbf{H}}y(t) \end{bmatrix} \qquad \mathbf{2.33(61)}$$

Further on, the error in the state estimation is expressed by

$$\dot{\tilde{\mathbf{z}}}(t) = (\mathbf{A}_{22} - \tilde{\mathbf{H}}\mathbf{A}_{12})\tilde{\mathbf{z}}(t) \qquad \mathbf{2.33(62)}$$

Based on Equation 2.33(62), the disturbance rejection property of the closed-loop system equipped with the reduced order observer will be determined by the roots of the

$$\tilde{\beta}(s) = \det(s\mathbf{I}_{n-1} - \mathbf{A}_{22} + \tilde{\mathbf{H}}\mathbf{A}_{12}) \qquad \mathbf{2.33(63)}$$

characteristic equation. Note that the coefficients in $\tilde{\mathbf{H}}$ represent the freedom to locate the roots of the closed-loop characteristic equation. To be able to complete the design in that way, the subsystem belonging to the states in the reduced observer must be observable.

In case of constant disturbances, even reduced order observers need special treatment. Assume a constant additive disturbance signal acting on the input:

$$\dot{w} = 0 \qquad \mathbf{2.33(64)}$$

Augmenting the state equation by the above relation, the number of the state variables will be increased, and an additional integrator will be part of the observer. The augmented reduced order observer is capable of rejecting disturbances of unknown constant values.

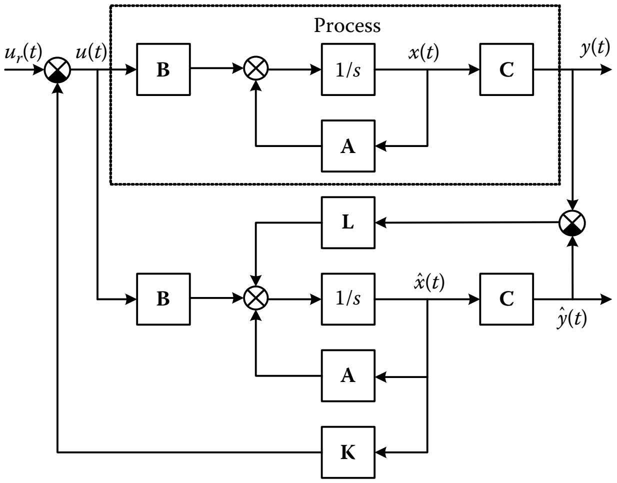


FIG. 2.33g
Block scheme of the combined observer-controller.

COMBINED OBSERVER–CONTROLLER

The state variable feedback and observer strategies discussed in the previous sections can be applied *simultaneously,* as well. The state variable feedback driven by the reconstructed states can be considered as a complete controller (see Figure 2.33g). State reconstruction is realized using the measured variables. In the above part of this chapter, two state variable feedback strategies (a proportional and an extended one by an integrator), as well as four state reconstruction strategies (full order and reduced order observers with or without handling constant input disturbance) were presented. Any pair combined by a selected state variable feedback strategy and a selected observer strategy is applicable.

Combined Observer-Controller Behavior

First consider a closed-loop system with state variable feedback using a full order observer to reconstruct the unmeasured state variables. Again, the system is described by

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)\\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t)\end{aligned} \qquad \mathbf{2.33(65)}$$

while the full order observer is given by

$$\dot{\hat{\mathbf{x}}}(t) = \mathbf{A}\hat{\mathbf{x}}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{L}(\mathbf{y}(t) - \mathbf{C}\hat{\mathbf{x}}(t)) \qquad \mathbf{2.33(66)}$$

Applying the control strategy of

$$\mathbf{u}(t) = -\mathbf{K}\hat{\mathbf{x}}(t) + \mathbf{u}_r(t) \qquad \textbf{2.33(67)}$$

Equations 2.33(65) through 2.33(67) allow deriving the following state equation for the closed-loop system:

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\hat{\mathbf{x}}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{BK} \\ \mathbf{HC} & \mathbf{A}-\mathbf{BK}-\mathbf{HC} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \hat{\mathbf{x}}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{B} \end{bmatrix} \mathbf{u}_\mathbf{r}(t) \qquad \textbf{2.33(68)}$$

where the system state vector and the observer state vector together form the complete state vector of the closed-loop system.

Selecting the state estimation error $\breve{\mathbf{x}}(t) = \hat{\mathbf{x}}(t) - \mathbf{x}(t)$ to replace the estimated state vector in the complete state vector, the following form can be obtained:

$$\begin{bmatrix} \dot{\mathbf{x}}(t) \\ \dot{\breve{\mathbf{x}}}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}-\mathbf{BK} & -\mathbf{BK} \\ \mathbf{0} & \mathbf{A}-\mathbf{HC} \end{bmatrix} \begin{bmatrix} \mathbf{x}(t) \\ \breve{\mathbf{x}}(t) \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \end{bmatrix} \mathbf{u}_\mathbf{r}(t) \qquad \textbf{2.33(69)}$$

The above form allows the derivation of the characteristic polynomial of the closed-loop system in an easy way:

$$\det\left(\begin{bmatrix} s\mathbf{I}_n - \mathbf{A} + \mathbf{BK} & \mathbf{BK} \\ \mathbf{0} & s\mathbf{I}_n - \mathbf{A} + \mathbf{LC} \end{bmatrix}\right)$$
$$= \det(s\mathbf{I}_n - \mathbf{A} + \mathbf{BK})\det(s\mathbf{I}_n - \mathbf{A} + \mathbf{LC}) \qquad \textbf{2.33(70)}$$

Equation 2.33(70) shows that the roots of the closed-loop characteristic equation contain roots determined by the state variable feedback law and roots introduced by the observer. It is also seen that the state variable feedback and the observer can be designed independently from each other. In other words, the setting of the coefficients in the feedback gain vector does not depend on whether the feedback is using the state vector itself or its reconstructed value. In the literature, this property is referred to as the *separation property* or *separation theory.* As far as the concept of setting the dynamics of the state variable feedback and that of the observer, it is reasonable to design fast enough transient dynamics for the observer to force the error in the state vector reconstruction to converge quickly to zero. "Fast enough transient" means here that the observer dynamics are to be designed faster than the dynamics of the system being observed. Assigning too-short time constants for the observer may lead to unwanted closed-loop behavior; specifically, the state estimation may become extremely sensitive, especially in cases of noisy measurements.

Considering SISO systems, find the transfer functions between the reference signal and the control input, as well as between the system output and the control input, respectively. Based on Equations 2.33(66) and 2.33(67), the control input can be written by

$$u(s) = -\mathbf{K}(s\mathbf{I}_n - \mathbf{A} + \mathbf{LC} + \mathbf{BK})^{-1}\mathbf{L}(y(s) - y_r(s)) \qquad \textbf{2.33(71)}$$

For the sake of simplicity, assume zero initial conditions for the state vectors involved. Also, selecting a special reference signal, a closed-loop system exhibiting identical servo and disturbance rejection properties will be designed. Equation 2.33(71) can be expressed in the form of a rational transfer function:

$$u(s) = -\frac{F(s)}{G(s)}(y(s) - y_r(s)) \qquad \textbf{2.33(72)}$$

where the nth order denominator polynomial above can be expressed by

$$G(s) = \det(s\mathbf{I}_n - \mathbf{A} + \mathbf{LC} + \mathbf{BK}), \qquad \textbf{2.33(73)}$$

while the numerator polynomial $F(s)$ is given by

$$F(s) = \mathbf{K}\operatorname{adj}(s\mathbf{I}_n - \mathbf{A} + \mathbf{LC} + \mathbf{BK})\mathbf{L}. \qquad \textbf{2.33(74)}$$

Using the above polynomials the characteristic equation of the closed-loop system is

$$P(s) = G(s)A(s) + F(s)B(s) \qquad \textbf{2.33(75)}$$

In harmony with Equation 2.33(70), it is seen that the closed-loop characteristic equation can be derived by multiplying the characteristic equations represented by the state variable feedback and the observer, respectively:

$$P(s) = \alpha(s)\beta(s) = G(s)A(s) + F(s)B(s) \qquad \textbf{2.33(76)}$$

Studying the behavior of the closed-loop system, no distinction can be made between the poles of the state variable feedback and the observer. Theoretically, all the poles of the closed-loop system can be assigned for the design procedure. Given a system with $A(s)$ and $B(s)$ and a design goal specified by $P(s)$, the polynomials $F(s)$ and $G(s)$ are to be determined to complete the closed-loop control. Assuming that controllability and observability both hold, as well as $ord(P(s)) = 2n$, solving the Diophantine Equation 2.33(76), a unique solution is obtained for $F(s)$ and $G(s)$. The controller derived is strictly proper.

For stable systems the presented control structure can also be transformed into an IMC (internal model control) form (see Figure 2.33h). In this structure a parallel process model is applied, preceded by a serial compensator. The overall transfer function becomes

$$\frac{y(s)}{y_r(s)} = \frac{F(s)B(s)}{\alpha(s)\beta(s)} \qquad \textbf{2.33(77)}$$

Note that the number of state variables realizing the IMC structure is *3n,* while the characteristic equation of the system remains as $\alpha(s)\beta(s)$ after eliminating the common terms.

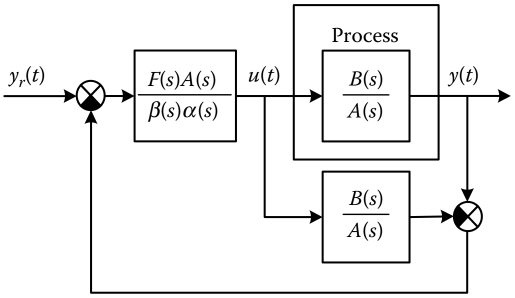


FIG. 2.33h
Observer-controller as internal model controller.

If the measured control input is available for realizing the control strategy, it is reasonable to derive another form of the controller utilizing the control signal. Assume that $\mathbf{u}(t)$ is the control input directly driving the process and $\mathbf{u}_0(t)$ is the control signal released by the controller (see Figure 2.33i). The actual feedback and observer relations can be written as

$$\begin{aligned}\dot{\hat{\mathbf{x}}}(t) &= (\mathbf{A}-\mathbf{LC})\hat{\mathbf{x}}(t)+\mathbf{Bu}(t)+\mathbf{L}(\mathbf{y}(t)-\mathbf{y}_r(t))\\ \mathbf{u}_0(t) &= -\mathbf{K}\hat{\mathbf{x}}(t)\end{aligned} \qquad \textbf{2.33(78)}$$

or for SISO systems in the form of transfer functions:

$$u_0(s) = -\mathbf{K}(s\mathbf{I}_n - \mathbf{A}+\mathbf{LC})^{-1}(\mathbf{B}u(s)+\mathbf{L}(y(s)-y_r(s))) \qquad \textbf{2.33(79)}$$

With some straightforward manipulations:

$$\beta(s)u_0(s) = -F(s)(y(s)-y_r(s))-(G(s)-\beta(s))u(s). \qquad \textbf{2.33(80)}$$

In the above expression

$$\frac{F(s)}{\beta(s)}$$

and

$$\frac{G(s)-\beta(s)}{\beta(s)}$$

are strictly proper transfer functions. The transfer function form of the combined observer–controller is shown in Figure 2.33i.

The presented versions require various amounts of computing power but result in identical control actions for a stable, linear, disturbance-free process. At the same time, for practical applications with slightly nonlinear behavior or input saturation, the exact calculations may lead to different closed-loop behavior.

Note that the polynomial approach shown above can accomplish even more complex control strategies. The state variable feedback technique followed so far required the polynomial $P(s)$ to be separable according to $P(s) = \alpha(s)\beta(s)$. A possible generalization may assume $P(s)$ to be freely selectable. The control problem then is set up as follows: given $A(s)$, $B(s)$, $P(s)$, and $\beta(s)$ find the polynomials $F(s)$ and $G(s)$ such that

$$\frac{F(s)}{\beta(s)}$$

and

$$\frac{G(s)-\beta(s)}{\beta(s)}$$

are both strictly proper. The control strategy derived in that way is more general than the state variable feedback control supported by an observer.

Due to the limited available space, not all the eight versions (two feedback solutions and four observers) will be discussed here. In general, each additional integrator or constant input disturbance estimation will increase the order of the controller by one. Applying a reduced order observer will reduce the order of the controller. The resulting $F(s)/\beta(s)$ and $F(s)/G(s)$ blocks will be just proper, not strictly proper transfer functions.

Transfer-Function Interpretation

A general controller is driven by the reference signal, the output signal, and the control signal. Assuming that the controller is linear and the number of the states used by the

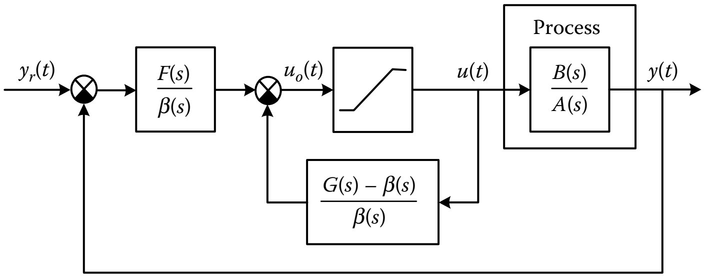


FIG. 2.33i
Transfer function form of the combined observer-controller.

TABLE 2.33j
Orders of the Transfer-Function Polynomials Equivalent to the Combined Observer–Controller

Feedback	*Observer*	*ord* $\alpha(s)$	*ord* $\beta(s)$	*No. States*	*No. Integr.*	*ord* $G(s)$	*ord* $F(s)$
P	O	n	n	n	0	n	$n-1$
PI	O	$n+1$	n	$n+1$	1	n	n
P	EO	n	$n+1$	$n+1$	1	n	n
PI	EO	$n+1$	$n+1$	$n+2$	2	n	$n+1$
P	R	n	$n-1$	$n-1$	0	$n-1$	$n-1$
PI	R	$n+1$	$n-1$	n	1	$n-1$	n
P	ER	n	n	n	1	$n-1$	n
PI	ER	$n+1$	n	$n+1$	2	$n-1$	$n+1$

controller is q, the most complicated possible controller can be described by

$$p_f(s)u(s) = p_{\mathrm{yr}}(s)y_r(s) - p_y(s)y(s) - p_u(s)u(s) \qquad \mathbf{2.33(81)}$$

The above controller is realizable if and only if ord($p_f(s)$) ≥ ord($p_{\mathrm{yr}}(s)$), ord($p_f(s)$) ≥ ord($p_y(s)$) and ord($p_f(s)$) > ord($p_u(s)$) are held for the polynomials involved. Designing identical servo and disturbance rejection properties we have $p_{yr}(s) = p_y(s)$. Then the control law by Equation 2.33(81) simplifies to

$$p_f(s)u(s) = p_y(s)(y_r(s) - y(s)) - p_u(s)u(s) \qquad \mathbf{2.33(82)}$$

resulting in

$$u(s) = \frac{p_y(s)}{p_f(s) + p_u(s)}(y_r(s) - y(s)) \qquad \mathbf{2.33(83)}$$

On the other hand, the general form of the controller design using the state variable feedback technique is based on the following form (i is the number of pure integrators in the controller to be designed):

$$P(s) = \alpha(s)\beta(s) = G(s)A(s)s^i + F(s)B(s) \qquad \mathbf{2.33(84)}$$

Solution of the above Diophantine equation results in the controller polynomials. The control law itself turns out to be

$$u(s) = \frac{F(s)}{s^i G(s)}(y_r(s) - y(s)). \qquad \mathbf{2.33(85)}$$

Comparing Equations 2.33(83) and 2.33(85), it is seen that the state variable feedback control is equivalent to a polynomial-based controller with $p_y(s) = F(s)$ and $p_f(s) + p_u(s) = s^i G(s)$.

If the control input is available and used for the control law calculations, the integrators included in the controller should be separated. In the case of state variable feedback control, the control input can be calculated by

$$\beta(s)u(s) = \frac{F(s)}{s^i}(y_r(s) - y(s)) - (G(s)s^c - \beta(s))u(s) \qquad \mathbf{2.33(86)}$$

In other words, the control input calculations are always filtered according to the characteristic polynomial of the observer.

Table 2.33j summarizes the state variable feedback and observer methods discussed in this section. It contains the order of the polynomials in the Diophantine equations and the number of the state variables involved for various control strategies. The following notations are used:

P	State variable feedback (proportional feedback from the state variables)
PI	State variable feedback plus extra integral feedback from the output
O	Basic full order asymptotic observer
EO	Extended asymptotic observer (observer with constant disturbance)
R	Reduced order observer
ER	Extended reduced order observer (reduced order observer with constant disturbance)
No. states	Number of the state variables in the controller
No. integr.	Number of the integrators in the controller

The table shows that the state variable feedback extended by an integrator and the proportional state variable feedback with an observer extended by constant disturbance estimation result in identical control actions. Although these control configurations are associated with identical $F(s)$ and $G(s)$ polynomials, under practical conditions (e.g., control signal saturation, handling slight nonlinearities) they will exhibit different closed-loop behavior because the detailed realizations of the algorithms are different.

CONCLUSIONS

This section overviewed the state variable feedback control of continuous-time linear dynamic processes. Using static gain coefficients for the state variables, the location of the closed-loop poles can be assigned in an arbitrary way. In order to improve both the servo and disturbance rejection properties of the closed-loop system, it is reasonable to extend this basic feedback structure by a term integrating the output signal.

State variables that are not directly accessible but are needed to obtain the feedback signal can be replaced by their estimates. The state estimate can be obtained by full order or reduced order observers.

It has been shown how the state variable control together with an observer can be interpreted in the form of transfer functions. Though only continuous-time problems have been discussed, all the control methods and control algorithms that were discussed can also be transformed to handle discrete-time systems.

Instead of describing the design criterion by the desired location of the roots of the closed-loop characteristic equation, the feedback gain matrix can be found as a result of optimizing a loss function (LQ problem). In that case the feedback gain matrix will be calculated as the solution of a Riccati equation.

The estimated states can be found not only by assigning the poles of the observer characteristic polynomial, but also by a well-elaborated solution based on stochastic considerations and assumptions related to the disturbances (Kalman filter technique). The two approaches are based on different assumptions; however, they lead to identical control actions for SISO systems.

Bibliography

Aplevich, J. D., *The Essentials of Linear State-Space Systems,* New York: John Wiley & Sons, 2000.

Astrom, K. J., and Wittenmark, B., *Computer-Controlled Systems: Theory and Design,* Englewood Cliffs, NJ: Prentice Hall, 1997.

Bertsekas, D. P., *Dynamic Programming and Optimal Control,* Vols. I and II, Belmont, MA: Athena Scientific, 2001.

Bryson, A. E., *Dynamic Optimization,* Reading, MA: Addison-Wesley, 1999.

Bryson, A. E., and Ho, Y., *Applied Optimal Control,* New York: Hemisphere/Wiley, 1975.

Burl, J. B., *Linear Optimal Control: H2 and H-Infinity Methods,* Reading, MA: Addison-Wesley, 1999.

Control system toolbox for use with MATLAB. User's Guide. The Math Works, 1998.

Decarlo, R. A., *Linear Systems. A State Variable Approach with Numerical Implementation,* Englewood Cliffs, NJ: Prentice Hall, 1989.

D'Souza, A. F., *Design of Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1988.

Goodwin, G. C., Graebe, S. F., and Salgado, M. E., *Control System Design,* Englewood Cliffs, NJ: Prentice Hall, 2001.

Iserman, R., *Digital Control Systems. Volume I. Fundamentals, Deterministic Control,* Heidelberg: Springer-Verlag, 1989.

Iserman, R., *Digital Control Systems. Volume II. Stochastic Control, Multivariable Control, Adaptive Control, Applications,* Heidelberg: Springer-Verlag, 1991.

Kailath, T., *Linear Systems,* Englewood Cliffs, NJ: Prentice Hall, 1980.

Kwakernaak, H., and Sivan, R., *Linear Optimal Control Systems,* New York: John Wiley & Sons, 1972.

Levine, W. S. (Ed.), *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.

Lewis, F. L., and Syrmos, V. L., *Optimal Control,* New York: John Wiley & Sons, 1995.

Lyashko, S. I., *Generalized Optimal Control of Linear Systems with Distributed Parameters,* Dordrecht: Kluwer, 2002.

Maciejowski, J. M., *Multivariable Feedback Design,* Reading, MA: Addison-Wesley, 1989.

Middleton, R. H., and Goodwin, G. C., *Digital Control and Estimation. A Unified Approach,* Englewood Cliffs, NJ: Prentice Hall, 1990.

Naidu, D. S., *Optimal Control Systems,* Boca Raton, FL: CRC Press, 2003.

Sinha, P. K., *Multivariable Control. An Introduction,* New York: Marcel Dekker, 1984.

2.34 Statistical Process Control

D. H. F. LIU (1995) **J. F. TATERA** (2005)

Software Package Suppliers:	
	ABB Inc. (www.abb.com) Aspen Technology Inc. (www.aspentech.com) CyboSoft (www.cybosoft.com) Foxboro Co. (www.foxboro.com) Honeywell (www.acs.honeywell.com and www.honeywell.com/imc) JMP Software (www.jmp.com) Lifestar (www.qimacros.com) Minitab Inc. (www.minitab.com) Northwest Analytical (www.nwasoft.com) Oil Systems Inc.—PI (www.osisoft.com) Pavillion Technologies Inc. (www.pavtech.com) Rosemount Inc./Emerson Process Management (www.rosemount.com and www.emersonprocess.com) Siemens (www.sea.siemens.com/ia) Software Toolbox (www.softwaretoolbox.com) StatPoint LLC. (www.statpoint.com) StatSoft (www.statsoft.com) Stochos Inc. (www.stochos.com) Wonderware (www.wonderware.com) Yokogawa (www.yokogawa.com/us) Zontec Inc. (www.zontec-spc.com)

INTRODUCTION

The goal of statistical process control (SPC) is to detect whether a process has undergone statistical abnormality — that is, a shift from its normal statistical behavior. The use of statistical techniques to detect variations in product quality and consistency dates back to Walter Shewhart's work at Bell Laboratories in the early 1900s. His work resulted in the development of statistical quality charts (Shewhart charts). These charts are still used for analyzing patterns in product variability. In the 1940s and 1950s, W. Edwards Deming's work in statistical quality control (SQC) methodology evolved into a 14-point management program for quality improvement. His approach emphasized the application of statistical principles to control the production process.

As with most other applications of statistics, different statistical methods operate under certain assumptions. These methods are subject to significant abuse and misuse when the assumptions are not understood and the methods are applied to data sets that they should not be used on. Most of these assumptions have to do with the nature of the data set (homogeneous, randomness, distribution shape, data collection techniques, etc.). A number of different charting techniques are available. This allows the user to select the one(s) that are most appropriate for their data set. Before going into the specifics, some general comments are appropriate.

SPC and Process Control

Just as a proportional-integral-derivative (PID) controller needs to be tuned to properly control individual loops, so an SPC process can be thought of as a means of tuning the overall process. SPC might not be considered by process control engineers as part of process control. It is a powerful statistical tool to help verify whether a process is performing properly and to help troubleshoot the process if a correction or improvement is desired.

We usually think of control in terms of a direct correction applied to the process in response to a change in the control variable. This is the case with both the open-loop control configuration (where a human manually initiates the correction) and the closed-loop (where an automatic controller initiates it).

One might think of SPC as an extreme form of open-loop control. SPC uses statistical process monitoring techniques to determine the status of the process, a combination of statistical and troubleshooting techniques to identify the cause(s) of upsets and the process area(s) requiring improvement/adjustment, and verification techniques to ensure that the adjustment(s) have resulted in the desired change.

Because of the open-loop decision-making nature of SPC, many believe that SPC charting should be done manually. This view is based at least partially on the assumption that a human observer will pay closer attention during periods of upsets and therefore will be better prepared for the troubleshooting and problem-solving phases of the process. Those sharing this view prefer to keep the observer actively involved in the SPC process.

In the name of productivity and to make better use of our powerful computing tools, many organizations have implemented forms of automated SPC charting and monitoring. When automated, the SPC calculations and charting are usually implemented in a Plant Information Data Base and may be displayed either in that system and/or a process control computer. Most of those who have automated the first phase are still performing the nonmonitoring portions of the SPC process manually. A few of the more aggressive automation types are attempting to implement diagnostic and process adjustment activities with artificial intelligence or other computer models.

Continuous Processes

It needs to be emphasized that all statistical techniques are not valid for all data sets. SPC tools that were developed for discrete manufacturing need to be adapted and/or interpreted differently when applied to a continuous process environment. Many statistical techniques involve assumptions that are not valid on all data sets. These may include issues such as the randomness of the data, the shape of the distribution, and data collection technique issues. It should also be noted that choosing an appropriate chart(s) for the data set and/or problem can greatly simplify the SPC process. Different charting techniques can offer a different view of the same data set.

SPC techniques are the heart and soul of today's Six Sigma quality programs. Six Sigma is a very popular and powerful process improvement tool, which is implemented by a team. The team is usually led by operations people and includes process control engineers and operations, quality, or middle-management personnel. It is normally targeted at measuring the Process Capability of both manufacturing and nonmanufacturing processes and then focuses on improving those processes.

What Is Statistical Control?

All processes exhibit variation. There is no cause for adjusting the total process unless the variation exceeds statistically acceptable limits. If the process is within acceptable statistical limits, changes would only introduce instability. Just as a PID controller needs to be tuned to properly control a loop, so the SPC process can be thought of as a procedure to help determine the tuning parameters for the total process. In other words with SPC, the control is usually on a process scale and not for individual loops.

What does it mean for a process to be in statistical control? A process is considered in control if it is experiencing only normal upsets and variation that have often occurred in the past. A process is out of statistical control when the change is sustained and it brings the controlled variable outside the limits of its normal range of variation. In that case one should try to identify the cause of the change in order to decide how to respond to it.

Understanding the types of variation allows one to predict the performance of a stable process. If an "in-control" process is making large quantities of out-of-specification product, the control limits need to be changed. Process control limits are not design specifications but are values that describe past performance of the process. Therefore, in order to change the SPC limits, one needs to change the process.

SPC TOOLS AND TECHNIQUES

SPC problem-solving techniques include:

- Analysis of processes for stability and the effects of process modifications (control charts and capability indices)
- Defining problems and setting priorities (Pareto charts)
- Identifying causes for good and bad performance (cause and effect diagrams—fishbone charts/diagrams)
- Quantifying relationships between process or product variables and other variables (scatter plots or other correlation tools)

Control Charts

The most widely used SPC tools are control charts. On these charts, the measurements of product samples are plotted to show their centering (XBar chart) and dispersion values (R chart). The centering value is the average of the samples. The dispersion value is the frequency distribution from that average. These charts are used to distinguish random, uncontrollable variations in measured parameters from variations that are controllable and should be corrected. To control product quality, the aim is to keep the product variation within the random pattern.

Figure 2.34a shows a centering and a dispersion chart. Both charts have similar components: centerline, upper limit, lower limit, and sample size. The centerline, also called the average, is shown with a solid line. The average of all the samples should be at this value. The upper and lower limits are shown by dotted lines. If the average is achieved, the averages of n samples should be less than the upper limit and greater than the lower limit. Variations within the limits are

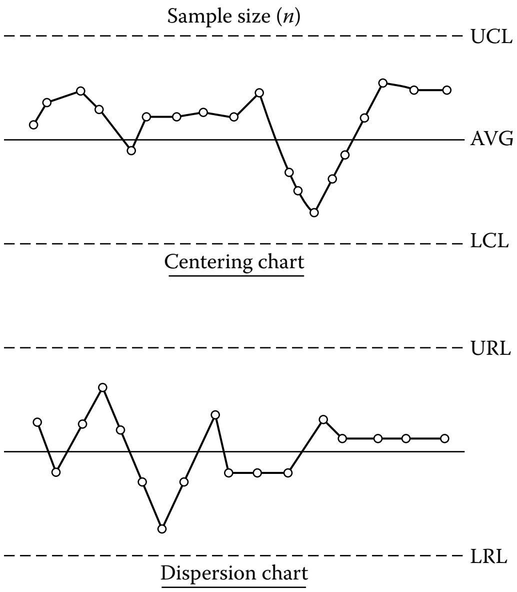


FIG. 2.34a
Shewhart chart for a subgroup sampling method. The centering (XBar) chart helps determine whether the centering of the process is stable. The Dispersion (R) chart helps determine whether the spread of the process is stable.

considered to be normal. Variations outside the limits are considered to be abnormal.

Charts and Tables

Tables 2.34b, 2.34c, and 2.34d show various control strategies for both the single point sampling method and for the subgroup sampling method. Also described are points to be stored for various types of centering charts. Tables 2.34e, 2.34f, and 2.34g give the equations for single point centering chart limits and

TABLE 2.34b
Control Chart Types for a Single Point Sampling Method

Chart Family	*Chart Type*	*Description*
Centering Chart	Individual	Uses individual point values to provide a trend.
	Geometric Mean	Uses the geometric mean to provide exponential smoothing.
	Moving Average	Uses the moving average to remove noise and indicate an overall trend.
Dispersion Chart	Range (R chart)	Commonly called an R chart. Plots the range which is the difference between the highest and lowest sample in the range. The number of samples for calculating the range is configured by the user.

Courtesy of Rosemount Inc.

TABLE 2.34c
Single Point Centering Chart Point Values

Centering Chart Type	*Point to Be Stored*
Individual	New value
Geometric Mean	(last point value) $(1 - W)$ + (new value) (W), where W is the specified weight of a new point
Moving Average	Average of the new individual and the last n individuals, where n is the number of samples for range

Courtesy of Rosemount Inc.

TABLE 2.34d
Control Chart Types for a Subgroup Sampling Method

Chart Family	*Chart Type*	*Description*
Centering Chart	Average (XBar)	Commonly called an XBar chart. Plots the average of data in the specified range. All subgroup samples are saved until the specified size is met. The average is then calculated, and the value written to disk.
	Median	Plots the median value in the same manner as the average chart plots the average value.
Dispersion Chart	Range (R chart)	Commonly called an R chart. Plots the range which is the difference between the highest and lowest sample in the subgroup. Samples are collected until the subgroup is complete, then sorted in order of size. The average or median is determined for the centering chart, and then the range is calculated.
	Range with Sigma	Plots the R values in the same manner as in the range chart, but the limits are calculated using a derivation value (sigma) that was entered during configuration.
	Root Mean Squared	Used for subgroups of 12 to 25 points. Uses the root mean square (RMS) to calculate the limits and centerline.
	Root Mean Squared with Sigma	Plots points in the same manner as the RMS chart, but the limits and the centerline are calculated using a deviation value (sigma) that was entered during configuration.
	Standard Deviation	Plots the standard deviation of the subgroup points.

Courtesy of Rosemount Inc.

TABLE 2.34e
Equations for Single Point Centering Chart Limits

Centering Chart Type	*Control Limit*	*Parameters*
Individual— "Limit Calculation Type" Range	$\text{UCL} = \text{AVG}\,X + \frac{3(\text{AVG}\ R)}{1.128}$ $\text{LCL} = \text{AVG}\,X - \frac{3(\text{AVG}\ R)}{1.128}$	R = range value X = subgroup average
Individual— "Limit Calculation Type" Standard Deviation	$\text{UCL} = \text{AVG}\,X + \frac{3\sqrt{\sum_{i=1}^{n}(X_i - \text{AVG}\,X)^2}}{n-1}$ $\text{LCL} = \text{AVG}\,X - \frac{3\sqrt{\sum_{i=1}^{n}(X_i - \text{AVG}\,X)^2}}{n-1}$	$n = 2$ X_i = sample value X = subgroup average
Geometric Mean	$\text{UCL} = \text{AVG}\,X + \frac{3\sqrt{\frac{W}{2-W}(\text{AVG}\,R)}}{1.128}$ $\text{LCL} = \text{AVG}\,X - \frac{3\sqrt{\frac{W}{2-W}(\text{AVG}\,R)}}{1.128}$	R = range value W = weight of the point X = subgroup average
Moving Average	$\text{UCL} = \text{AVG}\ X + (1.88)\ (R)$ $\text{LCL} = \text{AVG}\ X - (1.88)\ (R)$	R = range value X = subgroup average

Courtesy of Rosemount Inc.

single point dispersion chart values. Tables 2.34h, 2.34i, 2.34j, and 2.34k give the equations for subgroup centering chart values and chart limits.

Interpretation of Charts

The XBar chart (Table 2.34d) or Shewhart chart is used in combination with rule-based checks. It can detect whether a process output has undergone statistical abnormality—that is, a shift from its normal statistical behavior. On noisy processes, a simple time trend does not make this kind of deviation visually obvious, but simple rule checks can.

Centering charts generated from single point measurements (Table 2.34b) indicate that a process is not in statistical control if any one of the following is true:

1. One or more points fall outside the control limits.
2. Seven or more consecutive points fall on the same side of the centerline.
3. Ten of 11 consecutive points fall on the same side of the centerline.
4. Three or more consecutive points fall on the same side of the centerline and all are located closer to the control limit than to the centerline.

Centering charts generated from the subgroup sampling method indicate abnormal fluctuations when:

1. Any single subgroup value is more than three standard deviations away from the centerline (or set point).
2. Two consecutive subgroup values are more than two standard deviations away from the centerline, on the same side of the centerline.
3. Three consecutive subgroup values are more than one standard deviation away from the centerline, on the same side of the centerline.

TABLE 2.34f
Single Point Dispersion Chart Values

Dispersion Chart Type	*Range Value Equation*	*Parameters*
Range	R (for only two samples) = $\lvert R_{\max} - R_n \rvert$ R (for more than two samples) = $R_{\max} - R_n$	R_n = minimum range value $R_{\max}$ = maximum range value

Courtesy of Rosemount Inc.

TABLE 2.34g
Equations for Single Point Dispersion Chart Limits

Dispersion Chart Type	*Control Limit Equation*	*Parameters*
Range	UCL = (3.27) (AVG R) LCL = (0) (AVG R) CL = AVG R	R = average range value

Courtesy of Rosemount Inc.

TABLE 2.34h
Equations for Subgroup Centering Chart Values

Centering Chart Type	*Equation*	*Parameters*
Average, $\bar{X}$	$\frac{x_1 + x_2 + x_3 \cdots + x_n}{n}$	$X_1 \ldots X_n$ = individual sample values $\bar{X}$ = subgroup average n = number of values in the subgroup

Courtesy of Rosemount Inc.

TABLE 2.34i
Equations for Subgroup Centering Chart Limits

Centering Chart Type	*Control Limit Equation*	*Parameters*
Average	$\text{UCL} = \text{AVG } \bar{X} + (A_2)(\text{AVG } R)$ $\text{LCL} = \text{AVG } \bar{X} - (A_2)(\text{AVG } R)$	n = number of values in the subgroup R = range of the subgroup A_2 = value from Table 1.20j

Courtesy of Rosemount Inc.

TABLE 2.34j
Constants for Subgroup Centering Chart Limits

n	A_2	*n*	A_2
—	—	11	0.29
2	1.88	12	0.27
3	1.02	13	0.25
4	0.73	14	0.24
5	0.58	15	0.22
6	0.48	16	0.21
7	0.42	17	0.20
8	0.37	18	0.19
9	0.34	19	0.19
10	0.31	20	0.18

Courtesy of Rosemount Inc.

TABLE 2.34k
Equation for Subgroup Dispersion Chart Value

Dispersion Chart Type	*Equation*	*Parameters*
Range	$R_{\max} - R_{\min}$	R = range of the subgroup (values sorted highest to lowest)

Courtesy of Rosemount Inc.

4. Two out of three consecutive subgroup values are more than two standard deviations away from the centerline, with all three values on the same side of the centerline.
5. Five consecutive subgroup values are on the same side of the centerline.

Another process violation detection chart, called the moving range chart, differs from an Xbar chart in that each datum is an actual scanned data point rather than a subgroup average of several samples.

The Purpose of the Charts

When utilizing SPC charts one needs to remember their purpose and choose them accordingly. The purpose of a good chart is to give information about the data that is not self-evident.

Graphs can help make recognition of data patterns and features much easier than most tables or lists. Shewhart's first principle for understanding data is that no data have meaning apart from their context. Shewart also provided two rules for the presentation of data:

1. Data should always be presented in a way that preserves the evidence in the data for all the predictions that might be made from them.
2. Whenever average, range, or histogram charts are used to summarize data, the summary should not mislead the user into taking any action that the user would not take if the data were presented in a time series.

With all of the charting techniques that are available, one needs to be careful of which to use and when. These are powerful techniques that can be helpful but if misapplied can be misleading.

Figures 2.34l and 2.34m depict histograms of a relatively large (500-data point) data set that is in statistical control. Histograms tend to obscure (at a minimum) the sequential information in the data.

Using Charts

Utilizing an average verses an X-plane histogram could lead one to differing conclusions regarding a process/data set.

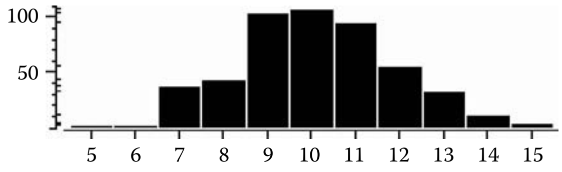


FIG. 2.34l
Histogram of 500 observations in the "X-Plane." (Courtesy of SPC Press Inc.)

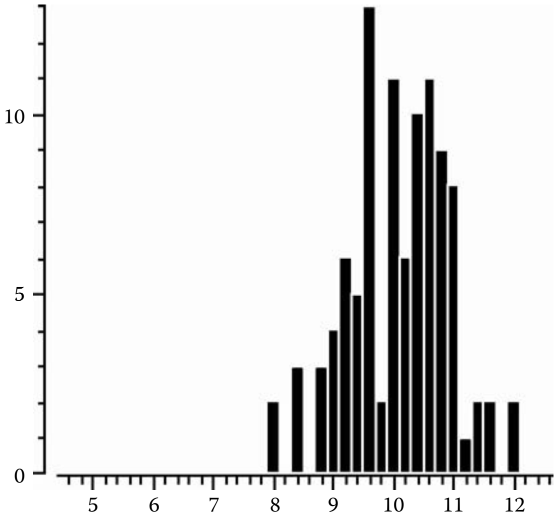


FIG. 2.34m
Histogram of the data provided in Figure 2.34l. Data set in the "Average Plane." (Courtesy of SPC Press Inc.)

Averaged charts tend to hide the frequency and magnitude of outliers. If only a simple glance is made at these charts (which represent the same process, which is in control), it could lead one to think that the first depicted a process in tighter control (it appears to have a very smooth normal distribution shape) and the second appears to be only generally normal in shape.

A quick glance could also lead one to think that the first indicated a process with a wider control limit (data range from 7 to 15) than the second (data range of 8 to 12). Also, the smooth-appearing normal distribution of the first as compared to the more jagged-appearing normal distribution of the second could mislead one to assume that the first represents a process that is in better control.

This comparison of two very similar charts of the same data is meant to encourage the review of data in many different forms, using many types of charts. This can help one to see different features in the data and better understand one's process. Patterns within the data may contribute more to understanding the process than a macro statistical overview of the data would. This is especially true for distributions that exhibit highly nonnormal tendencies (single tails, bimodal, etc.).

PROCESS CAPABILITY

An unstable, out-of-statistical-control process cannot be evaluated for its capability. Process Capability is the heart of the famed Six Sigma program. The Six Sigma program is intended to answer the following question: "Is the process capable of producing a product within the specifications to the statistical Six Sigma (Standard Deviations)?" If it is not, the Six Sigma program is to assist the user to identify the weak points that are keeping it from achieving this level of performance.

Upon achieving a state of statistical control, a process can be evaluated to determine whether it is capable of meeting the desired specification(s) or requirement(s). Usually a minimum of 20 to 25 subgroups is desirable to get a representative picture of the process. In order to characterize the capability of the process, its natural deviation is compared to the width of the specifications. This includes the evaluation of both the upper specification limit (USL) and the lower specification limit (LSL) independently to account for the location of the average process reading relative to the specifications. If there is only one limiting specification (LSL or USL), then the capability only needs to be calculated on that side.

The standard deviation ($\hat{\sigma}$) is used to evaluate the common level of process variation and is calculated by

$$\hat{\sigma} = \frac{\bar{R}}{d_2} \qquad 2.34(1)$$

where $\hat{\sigma}$ indicates that the value is an estimate, d_2 is a scaling factor, which is based on n samples in the subgroup, and $\bar{R}$ is the average range. The values of d_2 for 5, 10, 15, 20, and 25 samples (n) are, respectively, 2.326, 3.078, 3.472, 3.735, and 3.931.

A common technique for reporting process capability is the use of a ratio called the C_p index. The C_p index is the ratio that is obtained by dividing the distance to specification by the distance to the common cause variation. The C_p index is calculated as follows:

$$C_p \text{ for the upper side: } C_{pu} = \frac{\text{USL} - \bar{X}}{3\hat{\sigma}} \qquad 2.34(2)$$

where $\bar{X}$ is the subgroup average.

$$C_p \text{ for the lower side: } C_{pl} = \frac{\bar{X} - \text{LSL}}{3\hat{\sigma}} \qquad 2.34(3)$$

The minimum of these two values will be the worst of the two cases and is reported as the C_p or C_{pk} index.

A C_p value over 1.33 would indicate a capable (acceptable, saying that 99.73% of individual readings are within specification) process, while a C_p index of less than 1.0 would indicate a "noncapable" process. A C_p index between 1.0 and 1.33 would indicate a marginal process. C_{pu} values for four different process states are shown in Figure 2.34n. For the latter cases, the next step is to improve the process.

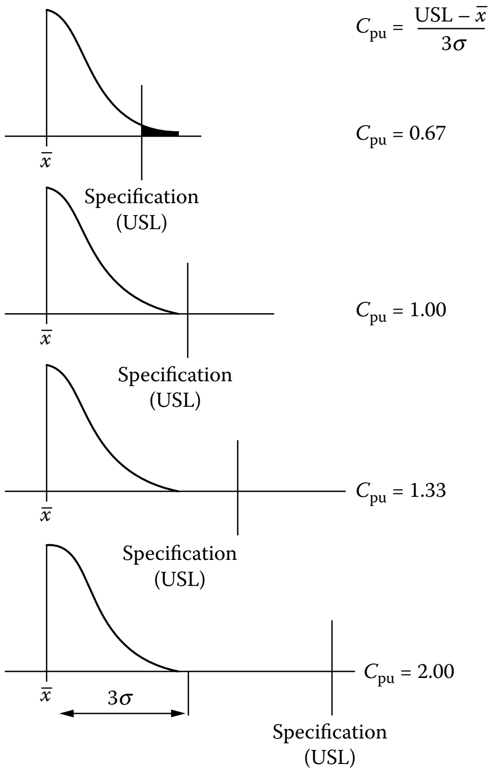


FIG. 2.34n
C_{pu} values for four different process states: C_{pu} less than 1.0, a noncapable process; C_{pu} between 1.0 and 1.33, a marginal process; C_{pu} equal to 1.33 or greater, a capable process.

There are three general ways to increase the values of the C_p index:

1. Move the specifications or set points (if they have not been properly established).
2. Move the process average, if the specification has only one limit (unilateral), or center the process average if the process has two limits (bilateral).
3. Reduce the common cause variation.

Identifying Causes

The goal of all SPC techniques is to first determine that the process behavior is not normal and next to identify/diagnose the cause(s) of unusual (good or bad but not normal) performance. One of the cause-and-effect tools used in SPC is the Ishikawa cause-and-effect diagram (also called the fishbone diagram).

Figure 2.34o depicts an Ishikawa fishbone diagram. This diagram is actually a thought process flow diagram (sort of a cause-and-effect checklist/map), which helps the user to systematically determine the root causes of problems. It begins with the major potential causes and assists/guides the user to work backwards through the listed causes to determine a root cause.

The user must first identify all the potential causes that can affect the process/product problem being studied. This diagnostics process can be lengthy and involve the use of designed experiments and many other techniques to study the process. The goal of this diagnostics process is to obtain a process that is in control. If the abnormal performance was bad, the goal is to eliminate its cause. If the abnormal performance was favorable, the goal is to incorporate it into the new normal process (process improvement).

Implementing SPC Concepts

The implementation of real-time SPC increases the demands on instrumentation, communication networks, and computer technology. Developments in microchip technology offer improved accuracy and stability for primary sensors. The use of digital transmitters eliminates the error contribution of

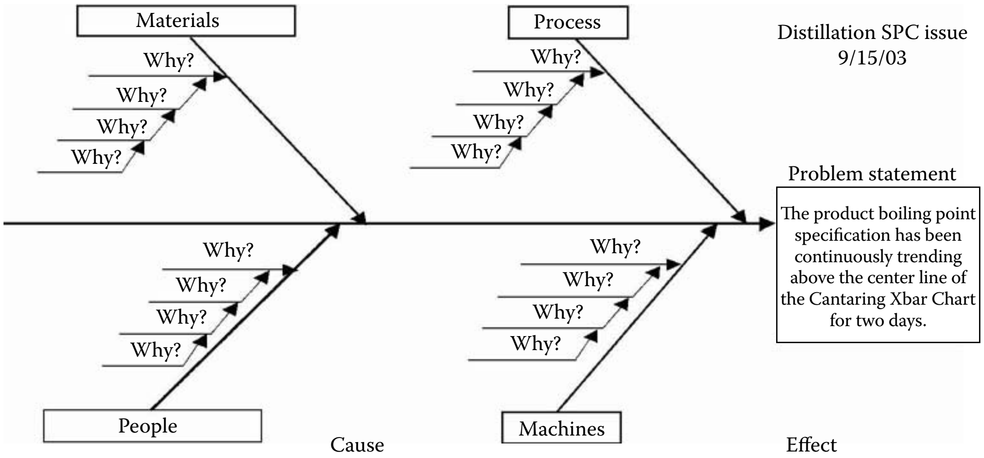


Fig. 2.34o
A fishbone diagram is a powerful diagnostic tool to help identify and define problems. (Created with QI Macros for Excel SPC Software.)

analog-to-digital conversion. Microprocessor-based “smart” transmitters contribute to better accuracy and higher reliability by their improved rangeability. Automatic pressure and temperature compensation, remote calibration, and self-diagnostics also contribute to better data quality.

The range of process variables that can be measured has also increased. Advances in online analytical instrumentation have made it possible to measure a variety of physical properties and chemical compositions that could not be directly detected in the past. (For more information, refer to Volume 1 of this handbook, titled *Process Measurement and Analysis*.)

Good quality data input (measurement) is useful for online SPC only if the control system network can transfer the process information rapidly enough for real-time data manipulation. The use of MAP (Manufacturing Automation Protocol) for integration of different vendors’ products into a common communications network facilitates streamlined data transmission. The MAP OSI (Open System Interconnection) model defines communication network functions in seven layers, allowing software programs to be utilized over different networks.

Although most of these techniques predate the computer, the ability to interface a personal computer with the control system also increases the feasibility of adding real-time SPC to existing plants. Intelligence at the I/O interface level enables routine operating functions to be handled locally, while conserving higher-level processing capability for SPC-type applications.

It is not necessary to implement any or all phases of an SPC program in a real-time mode. One only needs to automate those portions of an SPC program that one feels comfortable with, and the nonautomated SPC techniques can still be performed on a non-real-time basis. Many of those who have automated the data collection and monitoring phases are still performing the nonmonitoring portions of the SPC process manually. Others are implementing diagnostic and process adjustment activities with artificial intelligence or other computer models.

DATA STORAGE, COMPUTATION, AND DISPLAY

The ability to store statistical records of individual process data points is essential for generating and analyzing SPC control charts. Historical information is particularly useful for statistical analysis of cause-and-effect relationships using fishbone diagrams. Figure 2.34p shows the Pareto chart that enables diagnosis of the most likely causes of off-spec production runs at the component level by searching the historical database.

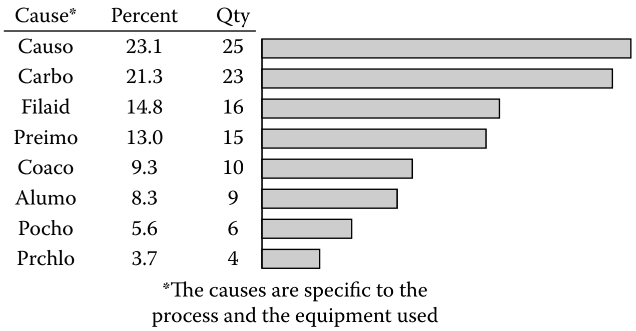


FIG. 2.34p
A historical data search, presented in the form of a Pareto chart, can help identify the most likely causes of off-spec production runs.

The desirability of performing identical mathematical studies as well as providing real-time data points warrants a dedicated SPC computer for many chemical process industry applications. Figure 2.34q is a scatter plot showing the use of a linear regression algorithm to explore the relationship between a selected product property (particle size) and an input variable (mixing time). Knowing this relationship, the operator can adjust the problem variable (mixing time) to restore the product (particle size) to meet specifications.

Statistical calculations translate into practical control results through the interpretation of statistical control charts. Configurable CRT displays are prerequisites to maximizing the use of online SPC charting capabilities. Automatic SPC alarm notifications and summary displays can isolate the problem input variables.

A computerized technique called online diagnostics is used to determine and anticipate the causes of plant operating problems before they actually occur. This diagnostic technique is basically a mix of SPC and fault-diagnostic principles that are programmed into an expert system that operates online. Thereby, process operators can anticipate and react to potential problems before they occur.

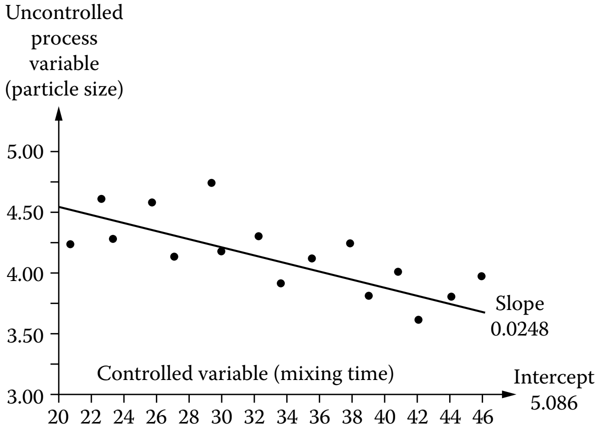


FIG 2.34q
Scatter plots can be used to investigate relationships between process variables and can indicate how much a controlled variable (mixing time) should be adjusted to bring an uncontrolled variable (particle size) within specifications.

Bibliography

Braverman, J. D., *Fundamentals of Statistical Quality Control,* Reston, VA: Reston Publishing, 1981.

Deming, W. E., *Out of Crisis,* Cambridge, MA: MIT Center of Advanced Engineering Study, 1986.

Doherty, J., "Detecting Problems with SPC," *Control*, November 1990, pp. 70–73.

Grant, E. L., and Leavenworth, R. S., *Statistical Quality Control,* 6th ed., New York: McGraw-Hill, 1988.

Holmes, D. S., "Time Series Analysis Overcomes SPC Problems," *Control,* February 1991, pp. 36–38.

Scherkenback, W. W., *The Deming Route to Quality and Productivity, Roadmaps and Roadblocks,* Rockville, MD: Mercury Press, 1986.

Shewhart, W. A., *Economic Control of Quality Manufactured Product,* New York: Van Nostrand Company, 1980.

Wheeler, D. J., *Advanced Topics In Statistical Process Control,* Knoxville, TN: SPC Press, 1995.

Wheeler, D. J., *Understanding Variation — The Key To Managing Chaos,* Knoxville, TN: SPC Press, 2000.

Wheeler, D. J., and Chambers, D. S., *Understanding Statistical Process Control,* Knoxville, TN: Keith Press, 1986.

Wolske, B. K., "Implementing SPC Concepts on a Real-Time Basis," *Chemical Processing,* March 1988, pp. 52–56.

2.35 Tuning PID Controllers

P. W. MURRILL (1970) **P. D. SCHNELLE, JR.** (1985) **B. G. LIPTÁK** (1995)
J. GERRY, M. RUEL, F. G. SHINSKEY (2005)

In order for the reader to fully understand the content and concepts of this section, it is advisable to first become familiar with some basic topics. These include gains, time lags and reaction curves (Section 2.22), the PID control modes (Section 2.3), feedback and feedforward control (Section 2.9), and relative gain calculations (Section 2.25).

Controllers are designed to eliminate the need for continuous operator attention when controlling a process. In the automatic mode, the goal is to keep the controlled variable (or process variable) on set point. The controller tuning parameters determine how well the controller achieves this goal when in automatic mode.

DISTURBANCES

The purpose of a controller is to keep the controlled variable as close as possible to its set point at all times. How well it achieves this objective depends on the responsiveness of the process, its control modes and their tuning, and the size of the disturbances and their frequency distribution.

Sources

Disturbances arise from three different sources: set point, load, and noise. Noise is defined as a random disturbance whose frequency distribution exceeds the bandwidth of the control loop. As such, the controller has no impact on it, other than possibly amplifying it and passing it on to the final actuator, which can cause excessive wear and ultimate failure.

Set point and load changes affect the behavior of the control loop quite differently, owing to the dynamics in their path. A controller tuned to follow set point changes tends to respond sluggishly to load variations, and conversely a controller tuned to correct disturbances tends to overshoot when its set point is changed.

Set Point The set point is the desired value of the controlled variable and is subject to adjustment by the operator. In a continuous process plant, most of the control loops operate as regulators, having a set point that remains unchanged for days and even months at a time.

Examples of variables held at constant set points are drum-level and steam temperature of a boiler, most pressure and level variables, pH of process and effluent streams, most product-quality variables, and most temperature loops. Set-point response is of no importance to these loops, but they must contend with load upsets minute by minute. In fact, the only loops in a continuous plant that must follow set-point changes are flow loops.

Batch plants have frequent transitions between steady states, some of which require rapid response to set-point changes with minimal overshoot. However, some of these changes are large enough to saturate the controller, particularly at startup. This can cause integral windup, which requires special means of prevention to overcome.

The Load Only pure-batch processes — where no flow into or out of the process takes place — operate at constant load, and that load is zero. All other processes can expect to encounter variations in load, which are principally changing flow rates entering and leaving vessels. A liquid-level controller, for example, manipulates the flow of one liquid stream, while other streams represent the load. Feedwater flow to a boiler is manipulated to control drum level and must balance the combined flows of steam and blowdown leaving to keep level at set point. The load changes frequently and often unpredictably, but the set point may never change.

In a typical temperature control loop, the load is the flow of heat required to keep temperature constant. Liquid entering a heat exchanger will require a certain flow of steam to reach a controlled exit temperature. Variations in liquid flow and inlet temperature will change the demand for steam flow manipulated to keep exit temperature at set point.

Dynamics

The term *process dynamics* can refer to capacitance, inertia, resistance, time constant, dead time or their combinations. There is no dynamics involved with changing the set point, unless intentionally placed there for the purpose of filtering the set point.

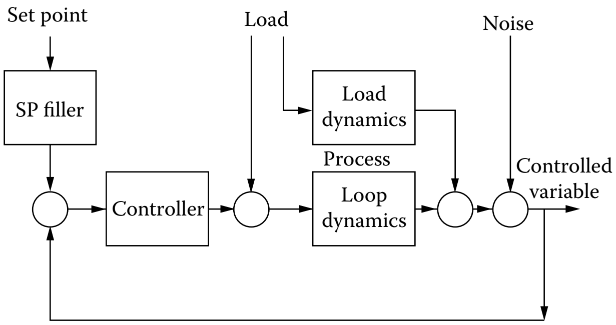


FIG. 2.35a
Load variables always pass through the dominant dynamic elements.

However, there is *always* dynamics in the load path. Load variables are principally the flow rates of streams similar to those manipulated by the controller. Therefore the dynamics in their path to the controlled variable are similar — and in most cases identical — to the dynamics in the loop itself.

Figure 2.35a presents all the essential elements of a control loop, showing its disturbance sources and dynamics. Most frequently, the dynamics are common to both the load disturbance and the controller output, meaning that the load and manipulated streams enter the process at the same point. An example would be the control of composition of a liquid at the exit of a blender, where both the manipulated and load streams making up that blend are introduced at a common entry point.

Less often encountered is the process where the dynamics in these two paths differ. An example of this is a shell-and-tube heat exchanger, wherein the temperature of a liquid leaving the tube bundle is controlled by manipulating the flow of steam to the shell. The shell may have more heat capacity than the tubes, causing the temperature to respond more slowly to a change in steam flow than to a change in liquid flow. Nonetheless, these two dominant lags will typically not differ greatly.

Step Responses

Step testing is recommended for all control loops where the frequency content of the disturbance variables is not specified. There are cases of periodic disturbances, and they can pose special problems for control loops that themselves are capable of resonating at a particular period. They are found principally in cascade loops and in process interactions where controllers manipulate valves in series or in parallel. These are considered in other sections of this work. Another period disturbance is the cyclic operation of such cleaning devices as soot-blowers.

For the general case, the step disturbance is the most difficult test for the controller in that it contains all frequencies, including zero. In fact, the frequency content of the step is identical to that of integrated white noise — therefore, it is an excellent test for loops subject to random disturbances.

Steps are also quite common in industry, representing conditions caused by sudden startup and shutdown of equipment; starting and stopping of multiple burners, pumps and compressors; and capacity changes of reciprocating compressors. If a control loop can respond adequately to a step disturbance, then a ramp or exponential disturbance will have less of an impact on it.

The step is also the easiest test to apply, requiring only a size estimate, and can be administered manually. Pulses require duration estimates, and doublet pulses require balancing. Step changes in set point are the usual disturbance applied to test or tune a loop, even for loops that operate at constant set point. Figure 2.35b illustrates a step response with 1/4 decay ratio.

The usual result of tuning a controller for set point response is to reduce its performance to variations in load. Therefore, the effectiveness of a controller and its tuning as a load regulator need to be determined by simulating a step load change.

Simulating a Load Change

Some controllers have an adjustable output bias. An acceptable simulation of load change, when the controller is in automatic and at steady state, is a step change in the value of this bias. The value of the controller output prior to the step is an indication of the current plant load because the loop was in a steady state. The step in bias in that case moves the controller output to another value, which disturbs the controlled variable and causes the controller to integrate back to its previous steady-state output.

Alternatively, controllers that can be transferred "bumplessly" between manual and automatic modes (most do — all

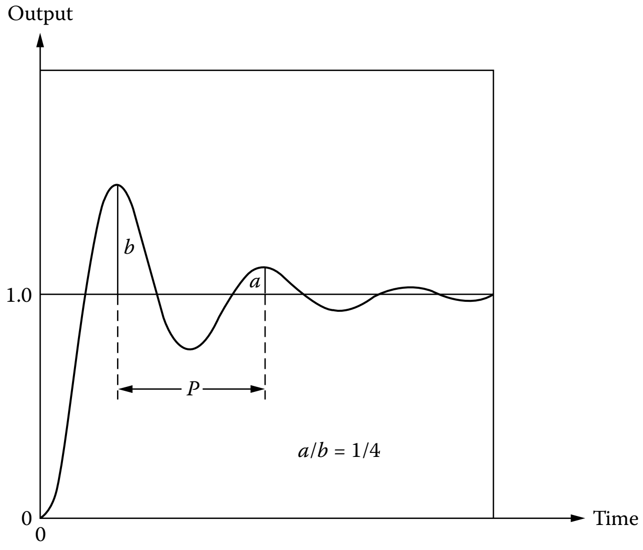


FIG. 2.35b
Step response curve of a control loop tuned for 1/4 decay ratio.

should) allow simulating a load change by using that feature. This is done by waiting until the loop is at steady state and on set point (zero deviation). At that point switching to the *manual* mode and stepping the output by the desired amount in the desired direction, and immediately (before a deviation develops) transferring back to the *automatic* mode.

This procedure can be followed for all but the fastest loops, such as flow loops. For them, a step in set point is acceptable, both because flow loops must follow set-point changes, and because for them, set-point tuning gives acceptable load response.

Comparing Set-Point and Load Responses

The steady-state process gain of a flow loop is typically between 1 and 2, as indicated by the controller output being between 50 and 100% when the flow measurement is at full scale. The proportional gain of a typical flow controller is in the range of 0.3 to 1.0, with the higher number associated with the process that has the lower steady-state gain. Therefore, the proportional loop gain for a typical flow loop is in range of 0.6 to 1.0. As a result, a step change in set point will move the controller output approximately the correct amount to produce the same change in flow, by proportional action alone, that gives excellent set-point response.

This is not the case for other loops. Level has the opposite behavior. To maintain a constant level, the controller must match the vessel's inflow and outflow precisely. Changing the set point will cause the controller to change the manipulated flow, but only temporarily — when the level reaches the new set point, the manipulated flow must return to its original steady-state value.

Therefore, *no* steady-state change in output is required for a level controller to respond to a set-point change. The Integrated Error (IE) sustained by a controller following a disturbance varies directly with the change in output between its initial and final steady states. In response to a set-point change, the level controller has the same initial and final steady-state output values and hence sustains zero integrated error.

As a result, the error that is integrated while the level is approaching the new set point will be matched exactly by an equal area of overshoot. In other words, set-point overshoot is unavoidable in a level loop unless set-point filtering is provided.

Most other processes, such as temperature, pressure, and composition, have steady-state gains higher than those of a flow process. But more importantly, they are also dominated by lags, which allows the use of a higher controller proportional gain for tight load regulation. When this high proportional gain is multiplied by the process steady-state gain, the resulting loop gain can be as high as 5 to 10 or more.

A set-point step then moves the controller output far more than required to drive the controlled variable to the new set point, producing a large overshoot. To minimize set-point overshoot, the controller must be detuned, with lower gain and longer integral time than is optimum for load regulation, or a filter must be applied to the set point.

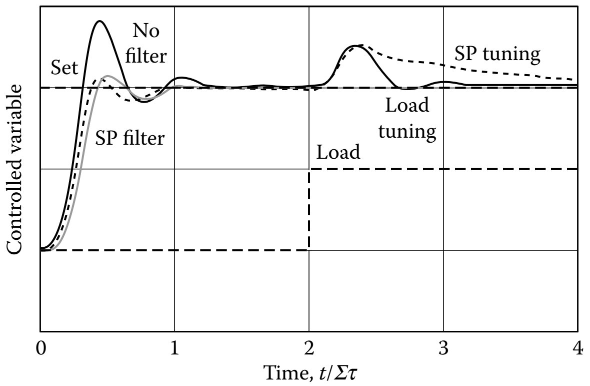


FIG. 2.35c
Set-point tuning slows load recovery for lag-dominant processes.

Figure 2.35c compares responses to steps in set point and load for a process with distributed lag such as a dashed exchanger, distillation column, or stirred tank. The time scale is normalized to $\Sigma\tau$, which is the time required for the distributed lag to reach 63.2% of the full response to a step input in the open loop. It is also the residence time of liquid in a stirred tank.

If the PID settings are adjusted to minimize the Integrated Absolute Error (IAE) to the set point change, the dashed response curve is produced (SP tuning). Note that following a step change in load, the return to set point is sluggish. This is commonly observed with lag-dominant processes. The PID settings that produce the minimum-IAE load response, shown in black (no filter), result in a large set-point overshoot, however.

Set-Point Filtering

If optimum load rejection is desired, without the large set-point overshoot that it produces on lag-dominant processes, the proportional response to set-point changes needs to be reduced.

Some PID controllers have the option to eliminate proportional action on the set point altogether. This tends to produce a set-point undershoot, which can significantly dashed the controller response, and it should never be used in the secondary controller of a cascade system. (Incidentally, derivative action should never be applied to the set point, as this always produces overshoot.)

Some controllers can reduce the controller's proportional gain when it acts on set point changes, either through the use of a lead-lag filter, or by the use of a specially structured algorithm. This adjustment allows separate optimization of set-point response, after the PID settings have been tuned to optimize the controller's load response. The filter used for the loop whose response is shown in Figure 2.35c has applied only half the controller's proportional gain to the set point.

OPEN-LOOP TUNING

Applying a step to the process is simple and can be used to tune the loop and to obtain a simple model for the process. Two methods are widely used. The first is the “process reaction curve,” which is not used to calculate the process model but is used to obtain the tuning parameters for rejecting the upsets caused by load changes.

The second method uses the “process model” by obtaining a simple process model; the tuning parameters are calculated from this model based on either a load rejection or a set point change criterion.

Process Reaction Curve

When a process is at steady state and it is upset by a step change, it usually starts to react after a period of time called the dead time (Figure 2.35d). After the dead time, most processes will reach a maximum speed (reaction rate), then the speed will drop (self-regulating process) or the speed will remain constant (integrating process).

When tuning a loop to remove disturbances caused by load changes, the controller must react at its maximum rate of reaction, and the strength of the reaction will correspond to the maximum speed. Hence, to tune the loop, it is not necessary to know the process model. It is sufficient to know the dead time and the maximum speed to calculate the tuning parameters.

Figure 2.35e illustrates the response of a temperature loop after a step change in the controller output. This example will be used throughout this section to illustrate the differences between the recommended settings arrived at by using the various tuning techniques.

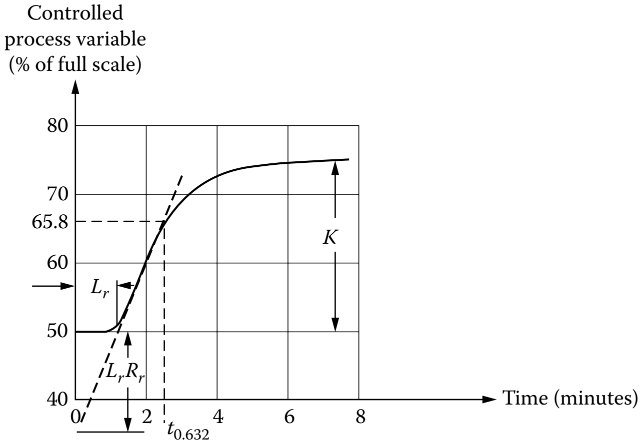


FIG. 2.35d
Reaction curve of a self-regulating process, caused by a step change of one unit in the controller output. $L_r = t_d$ is dead time, R_r is reaction rate, and K is process gain.

From this curve, it is not possible to determine the process model since the curve is too short to tell whether the reaction rate remains constant (integrating process) or goes down (self-regulating process). From such a test, the model cannot be found but the tuning parameters for load rejection can be estimated.

As can be seen from Figure 2.35e, a 10% change in CO was applied at 10 s and the temperature started to increase

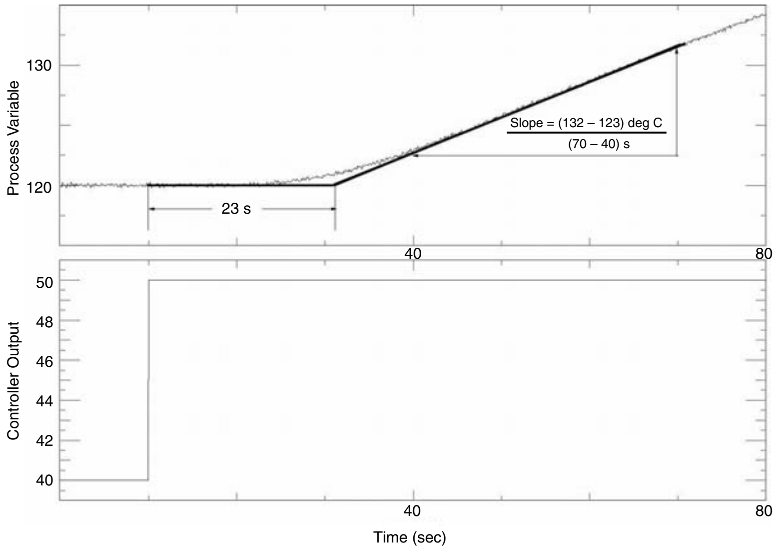


FIG. 2.35e
Process reaction curve in response of a change in controller output (CO). The process variable (PV) range is 0 to 300 degC and the CO range is 0 to 100%.

TABLE 2.35f
Equations for Calculating the Ziegler–Nichols Tuning Parameters for an Interacting Controller

Type of Controller	*P (gain)*	*I (minutes/repeat)*	*D (minutes)*
P	$\frac{\Delta \mathrm{CO}}{R_r * \mathrm{td}}$	—	—
PI	$0.9 * \frac{\Delta \mathrm{CO}}{R_r * td}$	3.33 t_d	—
PID	$1.2 \frac{\Delta \mathrm{CO}}{R_r * td}$	2 t_d	0.5 t_d

23 s seconds later at 33 s. Hence the dead time (t_d) is 23 s and the reaction rate (speed) is

$$R_r = \frac{\Delta \mathrm{PV}}{\Delta t} \qquad \textbf{2.35(1)}$$

The slope is:

$$\frac{\Delta \mathrm{PV}}{\Delta t} = \frac{132\,\mathrm{deg\,C} - 123\,\mathrm{deg\,C}}{70\ \mathrm{s} - 40\ \mathrm{s}} = \frac{9\,\mathrm{deg\,C}}{30\ \mathrm{s}} = \frac{\frac{9\,\mathrm{deg\,C}}{300\,\mathrm{deg\,C}} * 100}{30\ \mathrm{s}}$$
$$= \frac{3\%}{30\ \mathrm{s}} = 0.1\%/\mathrm{s} = 6.0\%/\mathrm{min}$$

After the dead time (t_d) and the reaction rate (R_r) have been determined, the controller settings are calculated by using the equations in Table 2.35f.

If a PI controller is to be used for the process that was tested in Figure 2.35e, the values are:

P = 0.9*10%/(0.1%/s X 23 s) = 3.9

I = 3.33 * 23 s = 76.6 s = 1.28 minutes.

Ziegler and Nichols recommend using the ratio of the controller output divided by the product of the slope and the dead time to calculate the proportional gain. The ideal process has a small dead time and a small slope, so that the controller can aggressively manipulate the controller output to bring the process back to set point.

The integral and derivative are calculated using the dead time. The proportional is calculated using the slope and the dead time. If the slope is high, then the controller gain must be small because the process is sensitive; it reacts quickly. If the dead time is long, the controller gain must be small because the process response is delayed and therefore the controller cannot be aggressive. If one can reduce the slope and the dead time of any process, it will be easier to control.

One of the advantages of open-loop tuning over the closed-loop tuning technique is its speed because one does not need to wait for several periods of oscillation during several trial-and-error attempts. The other advantage is that one does not introduce oscillations into the process with unpredictable amplitudes. In open-loop tuning, the user selects the upset that is introduced, and it can be small. Yet another advantage is that this test can be performed prior to the installation of the control system.

The disadvantages are also multiple. The open-loop test is not as accurate as the closed-loop one because it disregards the dynamics of the controller. Another disadvantage is that the S-shaped reaction curve and its inflection point are difficult to identify when the measurement is noisy and/or if a small step change was used.

Because of the above considerations, a good approach is to use the open-loop method of tuning in order to obtain the first set of initial tuning constants for a loop during startup. Then, refine these settings once the system is operating by retuning the loop using the closed-loop method.

Process Model

There are many ways to use and interpret the dead time (t_d) and reaction rate (R_r) values obtained from the open-loop tuning method. Most open-loop methods are based on approximating the process reaction curve by a simpler system. Several techniques are available to obtain a model.

The most common approximation by far is a pure time delay (dead time) plus a first-order lag. One reason for the popularity of this approximation is that a real-time delay of any duration can only be represented by a pure time delay because there is yet no other simple and adequate approximation. Theoretically, it is possible to use higher-than-first-order lags plus dead time, but accurate approximations are difficult to obtain. Thus the real process lag is usually approximated by a pure time delay plus a first-order lag. This approximation is easy to obtain, and it is sufficiently accurate for most purposes.

The process's dead time is the time period following an upset during which the controlled variable is not yet responding. The time constant is a period between the time when a response is first detected and the time when the response has reached 63.2% of its final (new steady-state) value. The time constant is also the time it would take for the controlled variable to reach its final value if the initial speed were maintained.

Bump Tests Figure 2.35g shows the Ziegler–Nichols procedure to approximate that process reaction curve with a first-order lag plus a time delay. The first step is to draw a straight-line tangent to the process reaction curve at its point of maximum rate of ascent (point of infection).

Although this is easy to visualize, it is quite difficult to do in practice. This is one of the main difficulties in this procedure, and a considerable number of errors can be introduced at this point. The slope of this line is termed the reaction rate R_r. The time between the instant when the bump was applied and the time at which this line intersects the initial value of the controlled variable prior to the test is the dead time, or transport time delay t_d.

Figure 2.35g illustrates the determination of these values for a one-unit step change (ΔCO) in the controller output

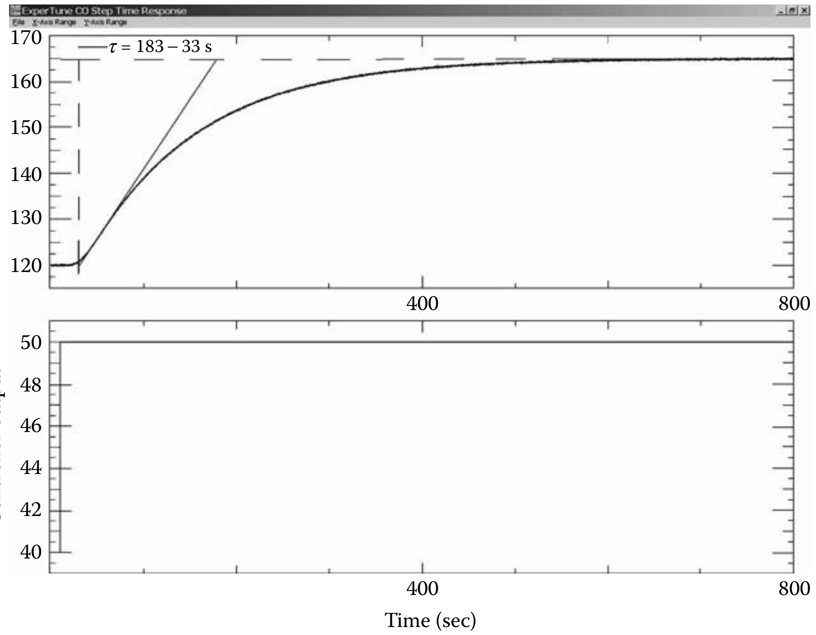


FIG. 2.35g
Maximum slope curve, Fit 1.

(manipulated variable) to a process. If a different-size step change in controller output was applied, the value of τ_d would not change significantly.

However, the value of R_r is essentially directly proportional to the magnitude of the change in controller output. Therefore, if a two-unit change in output was used instead of a one-unit change, the value of R_r would be approximately twice as large. For this reason the value of R_r used in Equation 2.35(2) or any other must be the value that would be obtained for a one-unit change in controller output.

In addition to the dead time and reaction rate, the value of the process gain K must also be determined as follows:

$$K = \frac{\text{final steady-state change in controlled variable (\%)}}{\text{change in controller ouput (control unit)}} \qquad \mathbf{2.35(2)}$$

There is a second method to determine the pure time delay plus the first-order lag approximation. In order to distinguish between these two methods, they will be called Fit 1 (described in Figure 2.35g) and Fit 2 (described in Figure 2.35h). The only difference between these two is in how the first-order time constant is obtained. In case of Fit 2, the time constant of the process is determined as the difference between the time when the dead time ends and the time when the controlled variable has covered 63.2% of the distance between the pre-test steady state and the new one. The dead time determination by both fits is the same and was already described.

Another method to determine the dead time is to measure the time when the PV moves by 2% of the total change.

The first-order lag time constants are given by:

$$\text{Fit1: } \tau_{F1} = K/R_r = 1.5/0.1\%/\text{s} = 150 \text{ s}$$
$$\text{(same slope as previous section)} \qquad \mathbf{2.35(3)}$$

$$\text{Fit 2: } \tau_{F1} = t_{63.2\%} - t_0 = 155 \text{ s} - 33 \text{ s} = 122 \text{ s} \qquad \mathbf{2.35(4)}$$

In Equation 2.35(4), $t_{63.2\%}$ is the time necessary to reach 63.2% of the final value, and t_0 is the time elapsed between the CO change and the beginning of the PV change. Note that the parameters for Fit 1 are based on a single point on the response curve, which is the point of maximum rate of ascent. However, the parameters obtained with Fit 2 are based on two separate points.

Studies[3] indicate that the open-loop response based on Fit 2 always provides an approximation that is as good or better than the Fit 1 approximation. A typical curve resulting from the above procedure is shown in Figure 2.35i.

The response shown in Figure 2.35i resulted from a ten-unit change in controller output. For different step changes, K and R_r must be adjusted accordingly. From a curve such as in Figure 2.35i, a number of parameters can be determined. The controller settings are calculated from the equations in Table 2.23j:

Table 2.35k compares the results obtained in terms of process gain (K), time constant ($\tau(s)$), dead time ($t_d(s)$), and the resulting controller gain (K_c) and integral time setting (T_i(min)) of a PI controller.

In the following discussion, the process model determined by the second fit will be used to compare against a variety of tuning criteria.

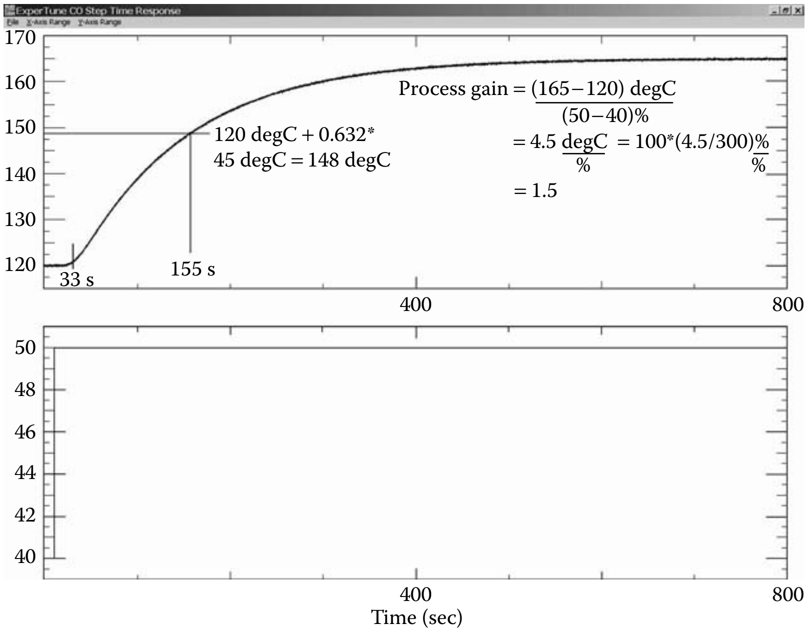


FIG. 2.35h
Bump test, Fit 2 curve.

Comparing the Tuning Methods One of the earliest methods for using the process reaction curve was proposed by Ziegler and Nichols. When using their process reaction curve method, which was described in connection with Figure 2.35e, only R_r and t_d or t_0 must be determined. Using these parameters, the empirical equations to be used to predict controller settings to obtain a decay ratio of 1/4 are given in Table 2.35j in terms of K, t_d, and τ.

In developing their equations, Ziegler and Nichols considered processes that were not "self-regulating." To illustrate,

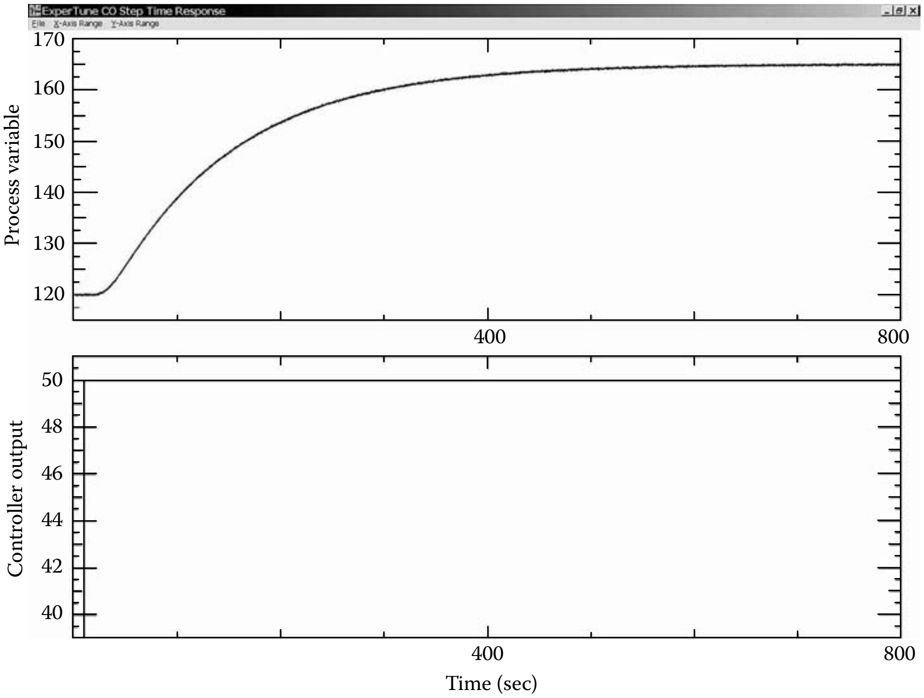


FIG. 2.35i
A typical reaction curve using the dead time and time constant obtained by a bump test.

TABLE 2.35j

Ziegler–Nichols' Recommendations to Obtain the Tuning Parameters for an Interacting Controller Based on the Readings Calculated from a Model

Type of Controller	*P (gain)*	*I (minutes/repeat)*	*D (minutes)*
P	$\frac{\tau}{K*t_d}$	—	—
PI	$0.9*\frac{\tau}{K*t_d}$	3.33 t_d	—
PID	$1.2\frac{\tau}{K*t_d}$	2 t_d	0.5 t_d

consider the level control of a tank with a constant rate of liquid outflow. Assume that the tank is initially operating at constant level. If a step change is made in the inlet liquid flow, the level in the tank will rise until it overflows. This process is not "self-regulating."

On the other hand, if the outlet valve opening and outlet backpressure are constant, the rate of liquid removal increases as the liquid level increases. Hence, in this case, the level in the tank will rise to some new position but will not increase indefinitely, and therefore system will be self-regulating. To account for self-regulation, Cohen and Coon[4] introduced an index of self-regulation μ defined as:

$$\mu = R_r L_r / K \qquad \textbf{2.35(5)}$$

TABLE 2.35k

The Tuning Setting Recommendations for a PI Controller Resulting from the Three Methods of Testing Described

Testing Method Used	*K*	τ *(s)*	t_d *(s)*	K_c	T_i *(min)*
Reaction curve				3.91	1.28
Process model Fit1, ZN	1.5	150	23	3.91	1.28
Process model Fit 2, ZN	1.5	122	23	3.18	1.28

Note that this term can also be determined from the process reaction curve. For processes originally considered by Ziegler and Nichols, μ equals zero and therefore there is no self-regulation. To account for variations in μ, Cohen and Coon suggested the equations given in Table 2.35l in terms of t_d and τ.

In case of a proportional control, the requirement that the decay ratio be 1/4 is sufficient to ensure a unique solution, but for the case of proportional-plus-reset control, this restraint is not sufficient.

Another constraint in addition to the 1/4 decay ratio can be placed on the response to determine the unique values of K_c and T_i. This second constraint can be to require that the control area of the response be at its minimum, meaning that the area between the response curve and the set point be the smallest. This area is called the error integral or the integral of the error with respect to time.

With the proportional-plus-reset-plus-rate controller (PID), the same problem of not having a unique solution exists even when the 1/4 decay ratio and the minimum error integral constraints are applied. Therefore, a third constraint must be chosen to obtain a unique solution. Based on the work of Cohen and Coon, it has been suggested that this new constraint could have a value of 0.5 for the dimensionless group $K_c K t_d / \tau$. The tuning relations that will result from applying these three constraints are given in Table 2.35l. This method has been referred to as the 3C method.[5–7]

Greg Shinskey suggested a variation to the above, where the proportional gain and the integral time are increased.

TABLE 2.35l

Comparison of Equations Recommended by Ziegler–Nichols, Shinskey, Cohen–Coon, and 3C for the Determination of the Tuning Settings for PID Controllers

	Ziegler–Nichols	*Shinskey*	*Cohen–Coon*	*3C*
P	$KK_c = (t_d/\tau)^{-1.0}$	$(t_d/\tau)^{-1.0}$	$(t_d/\tau)^{-1.0} + 0.333$	$1.208(t_d/\tau)^{-0.956}$
P	$KK_c = 0.9(t_d/\tau)^{-1.0}$	$0.95(t_d/\tau)^{-1.0}$	$0.9(t_d/\tau)^{-1.0} + 0.082$	$0.928(t_d/\tau)^{-0.946}$
I	$\frac{T_i}{\tau} = 3.33(t_d/\tau)$	$4.0(t_d/\tau)$	$\frac{3.33(t_d/\tau)[1+(t_d/\tau)/11]}{1+2.2+(t_d/\tau)}$	$0.928(t_d/\tau)^{-0.583}$
P	$KK_c = 1.2(t_d/\tau)^{-1.0}$	$0.855(t_d/\tau)^{-1.0}$	$1.35(t_d/\tau)^{-1.0}+0.270$	$1.370(t_d/\tau)^{-0.950}$
I	$\frac{T_i}{\tau} = 2.0(t_d/\tau)$	$1.6(t_d/\tau)$	$\frac{2.50(t_d/\tau)[1+(t_d/\tau)/5]}{1+0.6(t_d/\tau)}$	$0.740(t_d/\tau)^{0.738}$
D	$\frac{Td}{\tau} = 0.5(t_d/\tau)$	$0.6(t_d/\tau)$	$\frac{0.37(t_d/\tau)}{1+0.2(t_d/\tau)}$	$0.365(t_d/\tau)^{0.950}$

Integral Criteria Tuning[8]

Table 2.35m provides the controller settings that minimize the respective integral criteria to the ratio t_d/τ. The settings differ if tuning is based on load (disturbance) changes as opposed to set point changes. Settings based on load changes will generally be much tighter than those based on set point changes. When loops tuned to load changes are subjected to a set point change, a more oscillatory response is observed.

Which Disturbance to Tune for

With tuning parameters calculated for load rejection, the integral time (T_i) and derivative time (T_d) will depend mostly on the dead time (t_d) of the process.

In contrast, if the tuning parameters are calculated for a set-point change, the integral time will be longer and the derivative time will be shorter, and they will depend mostly on the time constant of the process.

The relationship between the controller settings based on integral criteria and the ratio t_0/τ is expressed by the tuning relationship given in Equation 2.35(6).

$$Y = A\left(\frac{t_0}{\tau}\right)^B \qquad \textbf{2.35(6)}$$

where $Y = KK_c$ for proportional mode, τ/T_i for reset mode, T_d/τ for rate mode; A, B = constant for given controller and mode; t_0, τ = pure delay time and first-order lag time constant.

Hence, using these equations,

$$K_c = \frac{A}{K}\left(\frac{t_0}{\tau}\right)^B \qquad \textbf{2.35(7)}$$

$$\frac{1}{T_i} = \frac{A}{\tau}\left(\frac{t_0}{\tau}\right)^B \qquad \textbf{2.35(8)}$$

$$T_d = \tau * A\left(\frac{t_0}{\tau}\right)^B \qquad \textbf{2.35(9)}$$

Lambda Tuning

Lambda tuning originated from Dahlin in 1968; it is based on the same IMC theory as MPC,[4,5] is model-based, and uses a model inverse and pole-zero cancellation to achieve the desired closed-loop performance.

Lambda tuning is a method to tune loops based on pole placement. This method ensures a defined response after a set-point change but is generally too sluggish to properly reject disturbances. Promoters for this method often claim that all loops should be tuned on the basis of Lambda tuning. Doing so, the controllers are almost in "idling mode" and

TABLE 2.35m

Tuning Settings for Load and Set Point Disturbances

		Load Change		*Set Point Change*	
		A	*B*	*A*	*B*
IAE	P	0.902	–0.985		
	P	0.984	–0.986	0.758	–0.861
	I	0.608	–0.707	1.020	–0.323
	P	1.435	–0.921	1.086	–0.869
	I	0.878	–0.749	0.740	–0.130
	D	0.482	1.137	0.348	0.914
ITAE	P	0.490	–1.084		
	P	0.859	–0.977	0.586	–0.916
	I	0.674	–0.680	1.030	–0.165
	P	1.357	–0.947	0.965	–0.855
	I	0.842	–0.738	0.796	–0.147
	D	0.381	0.995	0.308	0.929
ISE	P	1.411	–0.917		
	P	1.305	–0.959		
	I	0.492	–0.739		
	P	1.495	–0.945		
	I	1.101	–0.771		
	D	0.560	1.006		
ZN	P	1.000	–1.000		
	P	0.900	–1.000		
	I	0.333	–1.000		
	P	1.200	–1.000		
	I	0.500	–1.000		
	D	0.500	1.000		
CCC	P	1.208	–0.956		
	P	0.928	–0.946		
	I	1.078	–0.583		
	P	1.370	–0.950		
	I	1.351	–0.738		
	D	0.365	0.950		
Shinskey	P	1.000	–1.000		
	P	0.952	–1.000		
	I	0.250	–1.000		
	P	0.855	–1.000		
	I	0.625	–1.000		
	D	0.600	1.000		
4 to 1 decay	P	1.235	–0.924		
Critical damping (no overshoot, maximum speed)	P	0.300	–1.000	0.300	–1.000
	P	0.600	–1.000	0.350	–1.000
	I	0.250	–1.000	T_i=1.16τ	
	P	0.950	–1.000	0.600	–1.000
	I	0.420	–1.000	$T_i = \tau$	
	D	0.420	1.000	0.500	1.000

when the process load changes or other disturbance occurs, the time to eliminate this disturbance is quite long for most processes, because the integral time selected equals the process time constant.

The promoters also suggest the use of a closed-loop time constant, which is three times the process time constant. Doing so, the response time in the automatic mode will be three times longer than in manual. Therefore, the response time will be slower in automatic. This is adequate if no disturbance occurs but if no disturbance occurs, the control loop is not needed.

"Lambda tuning" refers to all tuning methods where the control loop speed of response is a selectable tuning parameter. The closed loop time constant is referred to as "Lambda" (λ). Therefore, following a set-point change, the PV will reach set point as a first-order system (same type of response as in the manual mode when the CO is changed).

Lambda tuning has been widely used in the pulp and paper industry, but control specialists are starting to realize that it is often too sluggish to handle disturbances. For a first-order plus dead time model

$$T_i = \tau \tag{2.35(10)}$$

$$K_c = \frac{1}{K} * \frac{T_i}{\lambda + t_d} \tag{2.35(11)}$$

where λ = closed loop time constant; it is recommended to use $\lambda = 3\tau$.

TABLE 2.35n

The Tuning Setting Recommendations for a PI Controller Resulting from the Criteria Listed

Criteria	*Tuning*	K_c	T_i *(min)*
Load change criteria	Ziegler–Nichols	3.18	1.28
	CCC	3.00	0.71
	Shinskey	3.37	1.53
	IAE	3.40	1.03
SP change criteria	IAE	2.13	1.16
	Lambda	0.21	2.03

The performance of lambda tuning is unacceptable for correcting upsets caused by load changes if the process time constant is larger than the dead time. This is the case with pressure, level, and temperature control applications. With flow loops, the results are similar to other methods since the time constant is in the same order of magnitude as the dead time.

In Table 2.35n, Fit 2 (Figure 2.35h) will be used as the reference to compare the process models found using the different tuning criteria. For load and set-point responses of the different tuning techniques, see Figures 2.35o, p, and q.

Adjusting Robustness To remove oscillations in a control loop, hence to increase the robustness, it is necessary to give

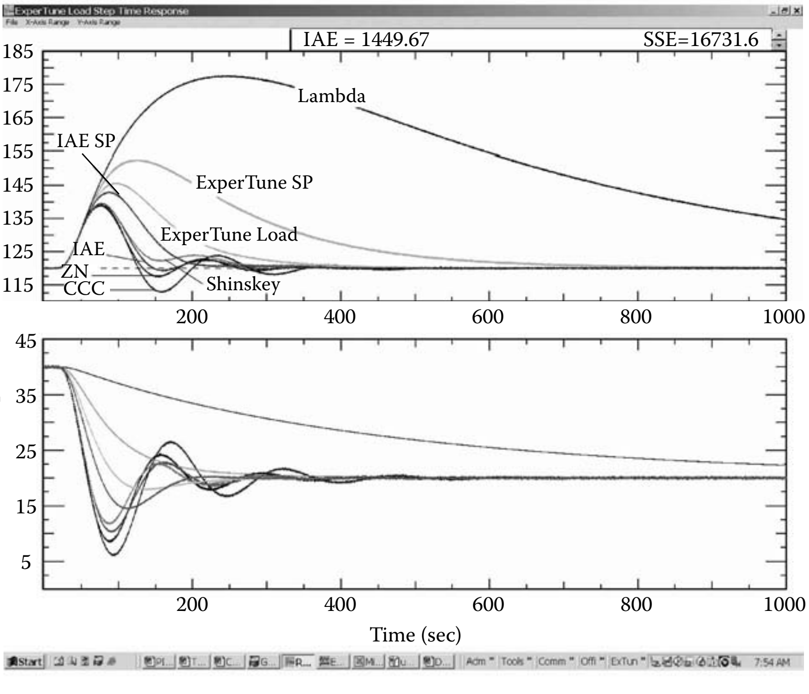


FIG. 2.35o

Load responses of the different tuning techniques (example).

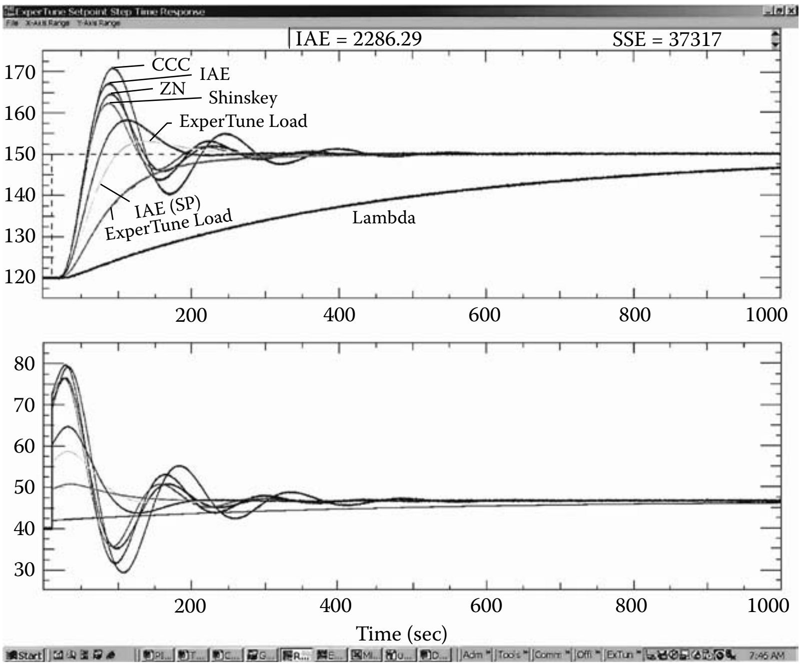


FIG. 2.35p
Set-point responses of the different tuning techniques (example).

up some performance. By reducing the proportional gain in a control loop, the robustness will be increased and the oscillations will be reduced or removed.

As a rule of thumb, dividing the proportional gain by a factor of two will eliminate the oscillations; dividing again the proportional gain by a factor of two will remove most of the overshoot. (For more on robustness, see Section 2.26.)

Digital Control Loops

Digital control loops differ from continuous control loops by having the continuous controller replaced by a sampler, a discrete control algorithm calculated by the computer, and a hold device (usually a zero-order hold). In such cases Moore et al. have shown that the open-loop tuning methods presented

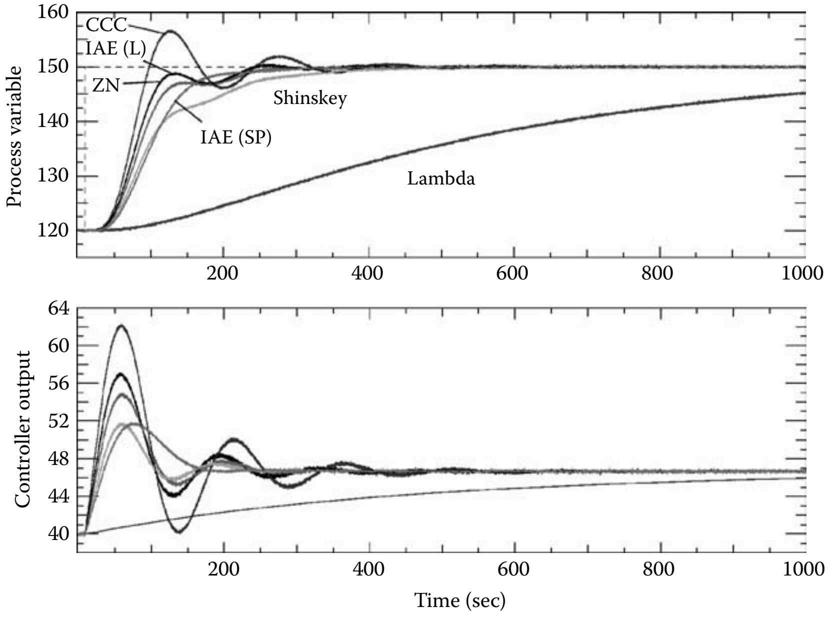


FIG. 2.35q
Set-point responses of the different tuning techniques (example) with a controller where the P is applied only on PV changes.

previously may be used, considering that the dead time used is the sum of the true process dead time and one-half of the sampling time, as expressed by Equation 2.35(12):

$$t_0' = t_0 + T/2 \qquad \textbf{2.35(12)}$$

where T is the sampling time. t_0' is used in the tuning relationships instead of t_0. (Section 2.38 deals with the subject of controller tuning by computer.)

CLOSED-LOOP RESPONSE METHODS

As has been discussed previously, in the open-loop method of tuning, the controller does not even have to be installed in order for the controller settings to be determined. When the closed-loop method is used, the controller is in automatic. Described below are the two most common closed-loop methods of tuning, the ultimate method and the damped oscillation method.

Ultimate Method

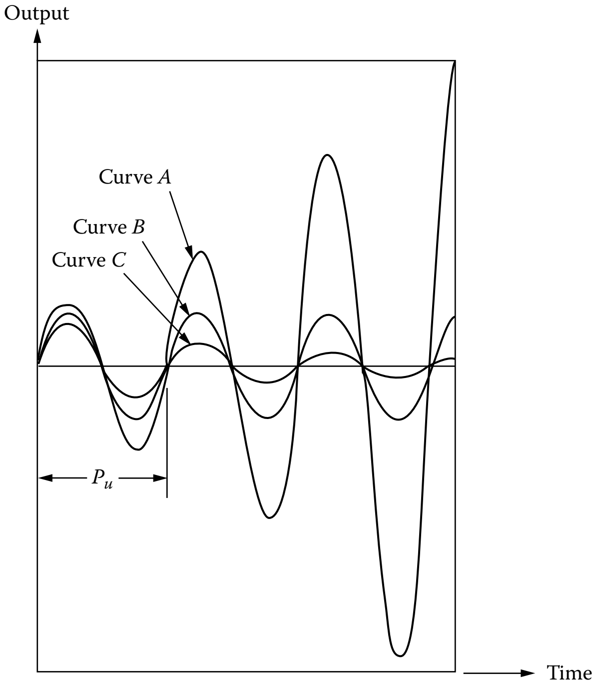
One of the first methods proposed for tuning controllers was the ultimate method, reported by Ziegler and Nichols[1] in 1942. This method is called "the ultimate method" because its use requires the determination of the ultimate gain (sensitivity) and the ultimate period. The ultimate gain K_u is the maximum allowable value of gain (for a controller with only a proportional mode) for which the system is stable. The ultimate period is the response's period with the gain set at its ultimate value (Figure 2.35r).

In order for a closed loop to display a quarter of the amplitude damping (Figure 2.35b), its loop gain must be at 0.5. This means that the *product* of the gains of all the components in the loop — composed by the process gain (G_p), the sensor gain (K_s), the transmitter gain (K_t), the controller gain (K_c = 100/PB), and the control valve gain (K_v) — must be at 0.5. When the loop is in sustained, undampened oscillation, the gain product of the loop is 1.0 and the amplitude of cycling is constant (Curve B in Figure 2.35r).

The period when the closed loops oscillate depends mostly on the amount of dead time in the loop. The period of oscillation in flow loops is 1 to 3 seconds; for level loops, it is 3 to 30 seconds (sometimes minutes); for pressure loops, 5 to 100 seconds, for temperature loops; 0.5 to 20 minutes; and for analytical loops, from 2 minutes to several hours.

When controlled by analog controllers (no dead time added by sampling), plain proportional loops oscillate at periods ranging from two to five dead times, PI loops oscillate at periods of three to five dead times, and PID loops at around three dead time periods.

The settings determined by this method will be based on load disturbance rejection and will not be suitable for set-point changes. The tuning parameters are in fact calculated on the basis of the distance where the loop will operate from instability. Settings for load rejection are not too far from instability, but tuning parameters for set-point change are different, and the process model is needed for their determination.



Curve A: unstable, runaway oscillation
Curve B: continuous cycling, marginal stability
Curve C: stable, damped oscillation

FIG. 2.35r
Ultimate gain is the gain that causes continuous cycling (Curve B) and ultimate period (Pu) is the period of that cycling.

The optimum integral and derivative settings of controllers vary with the number of modes in the controller (P = 1, PI = 2 mode, PID = 3 mode) and with the amount of dead times in the loop. For noninteracting PI controllers with no noticeable dead time, one would set the integral (T_i) for about 75% of the period of oscillation.

As the dead time-to-time constant ratio rises, the integral setting becomes a smaller percentage of the oscillation period — around 60% when the dead time equals 20% of the time constant, about 50% when their ratio is at 50%, about 33% when they are equal, and about 25% when the dead time exceeds the time constant.

For noninteracting PID loops with no dead time, one would set the integral minutes/repeat (I) to a value equal to 50% of the period of oscillation and the derivative time (D) to about 18% of the period. As dead time rises to 20% of the time constant, (I) drops to 45% and (D) to 17% of the period. At 50% dead time, (I) = 40% and (D) = 16%. When the dead time equals the time constant, (I) = 33% and (D) = 13%; finally, if dead time is twice the time constant, (I) = 25% and (D) = 12%.

Tuning Example The same example as was used earlier will be used to illustrate the "ultimate" closed-loop tuning method. The aim of this tuning process is to determine the controller gain or proportional band that would cause sustained, undampened oscillation (K_u) and to measure the corresponding period of oscillation, called the ultimate period (P_u). The steps in this tuning sequence are as follows:

1. Set all controller dynamics to zero. In other words, set the integral to infinite (or maximum) minutes per repeat or zero (or minimum) repeats per minute and set derivative to zero (or minimum) minutes.
2. Set the gain or proportional band to some arbitrary value near the expected setting (if known) or at $K_c = 1$ (PB = 100%) if no better information is available.
3. Let the process stabilize. Once the PV is stable, introduce an upset. The simplest way to do that is to move the set point up or down by a safe amount (for example, move it by 2% for half a minute) and then return it to its original value.

The result will be an upset in the PV resembling the characteristics of curve A, B, or C in Figure 2.35r. If the response is undampened (curve A), the gain (or proportional) setting is too high (proportional narrow); inversely, if the response is damped (curve C), the gain (or proportional) setting is too low (proportional wide). Therefore, if the response resembles curve A, the controller gain is increased; if it resembles curve C, the gain is reduced, and the test is repeated until curve B is obtained.

After one or more trials, the state of sustained, undampened oscillation will be obtained (curve B), and at that point the test is finished. (Make sure that the oscillation is a sinusoidal and not a limit cycle.) Next, read the proportional gain that caused the sustained oscillation. This is called the "ultimate gain" (K_u), and the corresponding period is the ultimate period of oscillation (P_u).

Once the values of K_u and P_u are known, one might use the recommendations of Ziegler–Nichols (Table 2.35s) or the recommendations that were described earlier, which also consider the dead time-to-time constant ratio. No one tuning is perfect, and experienced process control engineers do come up with their own "fudge factors" based on experience.

TABLE 2.35s

Tuning Parameters Based on the Measurement of K_u and P_u Recommended by Ziegler–Nichols for a Noninteracting Controller

Type of Controller	*P (gain)*	*I (minutes/repeat)*	*D (minutes)*
P	0.5 K_u	—	—
PI	0.45 K_u	$P_u/1.2$	—
PID	0.6 K_u	$P_u/2$	$P_u/8$

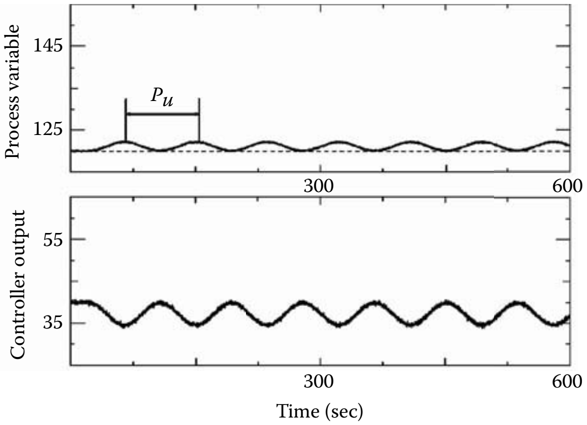


FIG. 2.35t

Ultimate cycling response of the same process that was tested in Figure 2.35e.

In order to use the ultimate gain and the ultimate period to obtain the controller settings for proportional controllers, Ziegler and Nichols correlated the decay ratio vs. gain expressed as a fraction of the ultimate gain for several systems. From the results they concluded that if the controller gain is set to equal one-half of the ultimate gain, it will often give a decay ratio of 1/4, i.e.,

$$K_c = 0.5K_u \text{ (PB} = 2\text{PB}u) \qquad \textbf{2.35(13)}$$

By analogous reasoning and testing, the equations in Table 2.35s were found to also give reasonably good settings for noninteracting two- and three-mode controllers. Again it should be noted that these equations are empirical and exceptions abound. For the same example as before, Figure 2.35t illustrates the ultimate cycling response.

The ultimate gain and period obtained from Figure 2.35t are: $K_u = 7.75$ and $P_u = 87$ s.

Hence the recommended tuning settings for the process that was used in the example are: $K_p = 3.49$ and $T_i = 1.21$ minutes.

There are a few exceptions to the tuning procedure described here because in some cases, decreasing the gain makes the process more unstable. In these cases, the "ultimate" method will not give good settings. Usually in cases of this type, the system is stable at high and low values of gain but unstable at intermediate values. Thus, the ultimate gain for systems of this type has a different meaning. To use the ultimate method for these cases, the lower value of the ultimate gain is sought.

Advantages and Disadvantages The main advantage of the closed-loop tuning method is that it considers the dynamics of all system components and therefore gives accurate results at the load where the test is performed. Another advantage is that the readings of K_u and P_u are easy to read and the period of oscillation can be accurately read even if the measurement is noisy.

TABLE 2.35u
Harriott Tuning Parameters for a Noninteracting Controller Calculated from $K_{c1/4}$ Obtained to Reach a Quarter-of-Amplitude Decay

Type of Controller	*P (gain)*	*I (minutes/repeat)*	*D (minutes)*
P	$K_{c1/4}$	—	—
PID	adjusted	$P/1.5$	$P/6$

The disadvantages of the closed-loop tuning method are that when tuning unknown processes, the amplitudes of undampened oscillations can become excessive (unsafe) and the test can take a long time to perform. One can see that when tuning a slow process (period of oscillation of over an hour), it can take a long time before a state of sustained, undampened oscillation is achieved through this trial-and-error technique. For these reasons, other tuning techniques have also been developed and some of them are described below.

Damped Oscillation Method

Harriott has proposed a slight modification of the previous procedure. For some processes, it is not feasible to allow sustained oscillations and therefore, the ultimate method cannot be used. In this modification of the ultimate method, the gain (proportional control only) is adjusted, using steps analogous to those used in the ultimate method, until a response curve with 1/4 of the decay ratio is obtained. However, with this tuning method, it is necessary to note only the period P of the response.

Again it should be noted that the equations in Table 2.35u are empirical and exceptions abound.

Example for Damped Cycling From Figure 2.35v the gain and period are found to be $K_{c1/4} = 7.75$ and $P = 87$ s. Hence the recommended tuning settings for the PI controller are $K_p = 3.49$ and $T_i = 1.21$ minutes.

After these modes are set, the sensitivity is again adjusted until a response curve with 1/4 of a decay ratio is obtained.

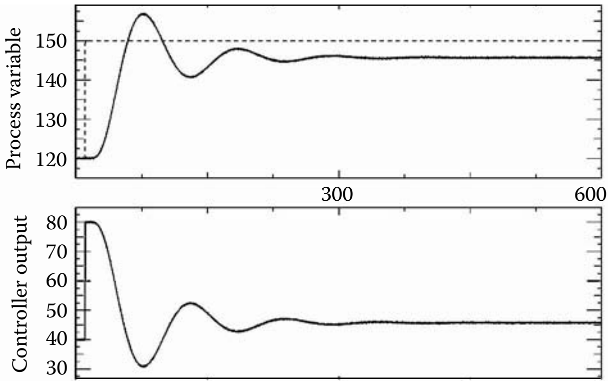


FIG. 2.35v
Damped cycling for the example.

This method usually requires about the same amount of work as the ultimate method since it is often necessary to experimentally adjust the value of the gain to obtain a decay ratio of 1/4. It is also possible to use this method to use a different decay ratio criterion.

Advantages and Disadvantages In general, there are two major disadvantages to the ultimate and damped oscillation methods. First, both are essentially trial-and-error methods, since several values of gain must be tested before the ultimate gain or the gain to give a 1/4 decay ratio are to be determined. To make one test, especially at values near the desired gain, it is often necessary to wait for the completion of several oscillations before it can be determined whether the trial value of gain is the desired one.

Second, while one loop is being tested in this manner, its output may affect several other loops, thus possibly upsetting an entire unit. While all tuning methods require that some changes be made in the control loop, other techniques require only one and not several tests, unlike the closed-loop methods.

Also, if the tuning parameters are too aggressive, the expected response can be obtained by increasing the proportional band (or decreasing the proportional gain). The integral and derivative settings probably need to be modified. The proportional gain has to be reduced to 3.5 to have a quarter-of-amplitude decay.

COMPARISON OF CLOSED AND OPEN LOOP

Table 2.35w provides a comparison of open-loop and closed-loop results for the process example used in Figure 2.35e.

FREQUENCY RESPONSE METHODS

Frequency response methods for tuning controllers involve first determining the frequency response of the process, which is a process characteristic. From this, tuning can be developed. Frequency response methods (FRM) may have several advantages over other methods: These are:

1. FRM require only one process bump to identify the process. The bump can be a change in automatic or manual,

TABLE 2.35w
Comparison of Closed and Open Loop Test Results

	Type of Tuning Test	K_c	T_i *(min)*
Open loop	Reaction curve	3.91	1.28
	Model Fit 2	3.18	1.28
Closed loop	Closed loop, ultimate cycling	3.49	1.21
	Closed loop, damped cycling	3.5	1.20

and be either a pulse, step, or other type of bump. A setpoint change provides excellent data from FRM.

2. FRM do not require any prior knowledge of the process dead time or time constant. With the other time response methods, one often needs a dead time estimate and a time constant estimate.
3. FRM do not require any prior knowledge of the process structure. Time response methods often require the user to have such model structure knowledge, i.e., whether it is first or second order or whether it is an integrator. For FRM-based tuning none of this is required; only the process data are needed.

Obtaining the Frequency Response

The process frequency response is a graph of amplitude ratio and phase vs. oscillation or sine wave frequency. If one injects a sine wave into a linear process at the controller output, the PV will also display a sine wave. The output (PV) sine wave will probably be of smaller height relative to the input and will be shifted in time.

The ratio of the heights is the amplitude ratio at the frequency of the input sine wave. The shift in time is the phase shift or phase lag. A time shift resulting in the trough of the output when aligned with the crest of the input is generally thought to be 180 degrees out of phase or 180 degrees of phase lag.

By applying a variety of sine wave inputs to a process, one can obtain a table of amplitude ratios and phase lags dependent on the sine wave frequency. If one plots these, the result will be the frequency response of the process.

Using Fourier analysis computer software programs one can calculate the process frequency response from a bump, pulse, or any other signal that applies sufficient excitation to the controller output. In the evaluation both the CO and PV trends are used. The data provided for these programs should start from a settled state, experience a quick change, and end settled.

Any one of the responses in Figures 2.35e, g, h, i, o, p, q, and v would provide adequate data for frequency response–based testing. Figure 2.35x shows the typical frequency response arrived at by the use of computer software.

PID Tuning Based on Frequency Response

In most processes, both the amplitude ratio and the phase angle will decrease with increasing frequencies. Assuming

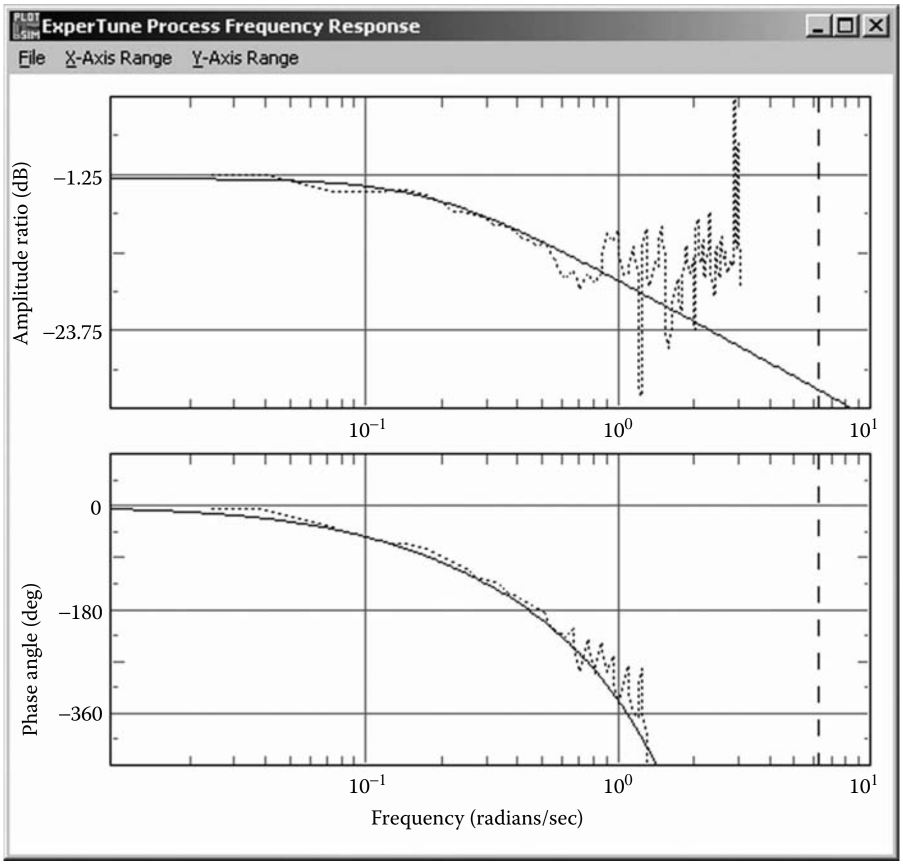


FIG. 2.35x
Typical model (solid line) and actual (dash line) process frequency response.

that the combined phase and amplitude ratio decreases with frequency when the process and the controller frequency responses are combined, the following general stability rule applies: A control system will be unstable if the open-loop frequency response has an amplitude ratio that is larger than one when the phase lag is 180 degrees.

To provide proper tuning, a margin of safety in the gain and phase is desired. Tuning constants are therefore adjusted to result in the highest gain at all frequencies and yet achieve a certain margin of safety or stability. This is best accomplished using computer software.

FINE TUNING

The performance of a controller can be tested by simulating a load change in the closed loop as described at the beginning of this section. This can be especially useful if the initial tuning is not satisfactory or if the process characteristics have changed. It is simply a trial-and-error method of recognizing and approaching an optimum — or otherwise desirable — load response.

Optimum Load Response

As described earlier, the optimum load response is generally considered to be that which has a minimum IAE, in that it combines minimum peak deviation with low integrated error and short settling time. The second curve from the bottom in each of Figures 2.35y and z represents a minimum-IAE load response for a distributed lag under PI control.

This second curve from the bottom has a symmetrical first peak, low overshoot, and effective damping. The time scale of these curves is normalized to the 63.2% open-loop step response of the distributed lag, identified as $\Sigma\tau$. On either side of the optimum curve in both figures are other response curves, which were produced by changing one of the controller settings.

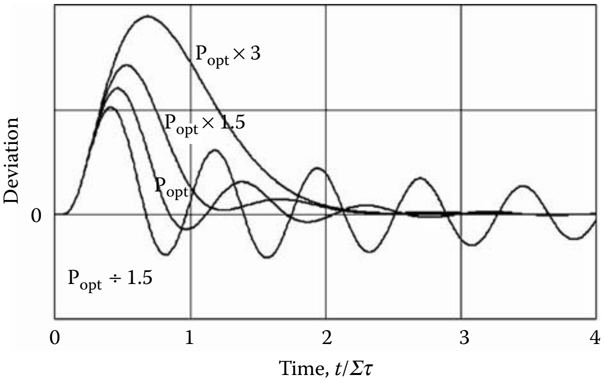


FIG. 2.35y
The proportional setting affects the height of the first peak and its symmetry.

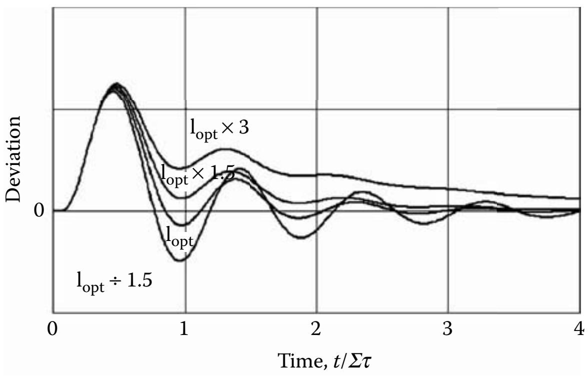


FIG. 2.35z
The integral setting primarily affects the location of the second peak.

In Figure 2.35y, the proportional band of the controller is the parameter being adjusted. Note that increasing the proportional band increases the height of the first peak and also its damping; but along with the increased damping comes a loss of symmetry. In Figure 2.35z, it is the integral time of the controller that is being adjusted. Note that it has no effect at all on the first peak but determines the location of the second, i.e., the overshoot or undershoot of the process variable's deviation.

The integrated error IE of a standard PID controller varies directly with the product of its proportional band and integral time. While increasing either of these settings improves damping, it also increases IE in direct proportion, and therefore it costs performance. Increasing both settings above their optimum compounds this effect.

Effect of Load Dynamics

All of the tuning rules described in this section to optimize load regulation will apply only to loops in which the dynamics in the load path are identical to those in the path of the controller output. Yet Figure 2.35a shows the possibility that the dynamics are different in the two paths, for example in the case of heat exchangers.

Any dead time in the load path, or variation thereof, will have no effect on the load response curve, simply delaying it by more or less time. The dominant lag in the load path is what determines the shape of the leading edge of the response curve.

As a demonstration, a distributed lag in a PI control loop was simulated to assess the effects of variations in load dynamics. Three different load-response curves appear in Figure. 2.35aa representing ratios of $\Sigma\tau_q/\Sigma\tau_m$ varying from 0.5 to 2, where subscripts q and m identify the load and manipulated-variable paths, respectively.

In all three cases, the PI controller has been tuned to minimize IAE following a simulated load change, simulated by stepping the controller output, which produces the center curve represented by $\Sigma\tau_q/\Sigma\tau_m = 1$. However, when a true load step

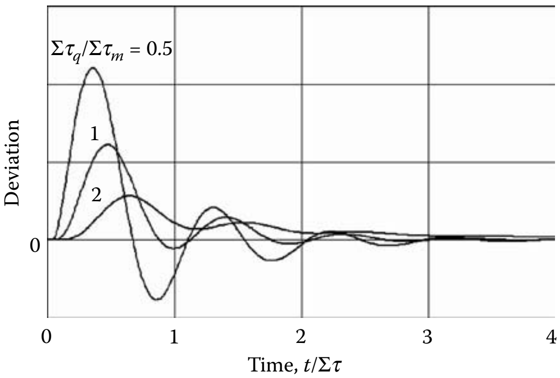


FIG. 2.35aa
Load dynamics affect the shape of the response curve.

passes through different dynamics than in the controller-output path, the resulting response curve is no longer minimum-IAE.

As might be expected, faster dynamics in the load path produce a faster rising deviation, which therefore peaks at a higher value before the controller can overcome it. The larger peak is then followed by a large overshoot. By contrast, slower dynamics in the load path reduce the peak deviation, followed by an undershoot. The IAE for the curve where $\Sigma\tau_q/\Sigma\tau_m = 0.5$ is actually only 6% higher than the minimum, but for the case where $\Sigma\tau_q/\Sigma\tau_m = 2$, the IAE is actually 53% above the minimum, indicating substantial need for correction. The appearance of the exaggerated overshoot and the undershoot indicate that those responses are no longer minimum-IAE, and that the integral time needs adjusting. Retuning the controllers for minimum-IAE response in the presence of different load dynamics resulted in the development of the correction factors plotted in Figure 2.3bb.

To use these factors, one should first tune the controller to minimize IAE following a load change simulated by stepping the controller output. After that the appropriate correction factors should be found for the expected ratio of load-to-manipulated lag $\Sigma\tau_q/\Sigma\tau_m$.

The integral time will need correcting regardless of the controller used—it must be multiplied by a factor represented by the top curve for a PID controller and the middle curve for a PI controller. If the controller is a noninteracting (ideal, parallel) PID, then its proportional band should be multiplied by the correction factor, which is indicated by the dashed line. PI and interacting (series) PID controllers do not require a correction to the proportional setting, and no correction to the derivative setting is required for any controller.

Symbols, Abbreviations

CO	Controller output
FRM	Frequency response method
G_p	Process gain
IAE	Integral of absolute error
IE	Integral error
ISE	Integral of squared error
ITAE	Integral of absolute error × time
ITSE	Integral of squared error × time
K_p	Proportional gain for a PID controller
K_u	Ultimate controller gain
PV	Process variable or measurement
R_r	Reaction rate, slope
RRT	Relative response time, the time to remove most of a disturbance
SP	Set point
τ (Tau)	Process time constant (seconds)

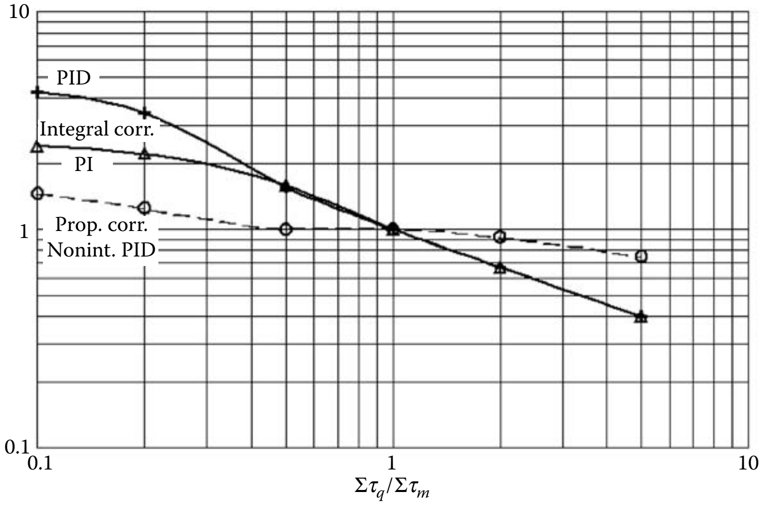


FIG. 2.35bb
To obtain minimum-IAE response, integral time and possibly the proportional band may require correction for load dynamics.

$t_0(t_d)$ Process dead time (seconds)

T_i Integral time (s) -for a PID controller

T_d Derivative time (s) for a PID controller

τ_F PV filter time constant

t_u Ultimate period

Bibliography

Astrom, K.J., and Hagglund, T., Automatic Tuning of PID Controllers, Research Triangle Park, NC: ISA Press, 1988.

Bjornsen, B.G., *A Fluid Amplifier Pneumatic Controller,* Control Engineering, June 1965.

Box, G.E.P., and Jenkins, G.M., Time Series Forecasting and Control, San Francisco: Holden-Day, 1970.

Caldwell, W.I., Coon, G.A., and Zoss, L.M., Frequency Response for Process Control, New York: McGraw-Hill, 1959.

Campbell, D.P., Process Dynamics, New York: John Wiley & Sons, 1958.

Chen, C.T., Introduction to Linear System Theory, New York: Holt, Rinehart and Winston, 1970.

Corripio, A.B., Tuning of Industrial Control Systems, Research Triangle Park, NC: ISA Press, 1990.

Gerry, J.P., *How to Control Processes with Large Dead Times,* Control Engineering, March 1998.

Gerry, J.P., *PID settings sometimes fail to make the leap,* Intech, November 1999.

Gerry, J.P., *Tune Loops for Load Upsets VS Setpoint Changes,* CONTROL Magazine, September 1991.

Higham, J.D., *Single-Term Control of First- and Second-Order Processes with Dead Time,* Control, February 1968.

Hougen, J.O., *Experiences and Experiments with Process Dynamics,* Chemical Engineering Process Monograph Series, Vol. 60, No. 4, 1964.

Howell, J., Rost, D., and Gerry, J.P., *PC Software Tunes Plant Startup,* InTech, June 1988.

Kraus, T.W., and Myron, T.J., *Self-Tuning PID Controller Uses Pattern Recognition Approach,* Control Engineering, June 1984.

Larson, R.E., et al., *State Estimation in Power Systems, Part I and Part II,* IEEE Trans. on Power Apparatus and Systems, Vol. PAS 89, No. 3, March 1970.

Lelic, M.A., and Zarrop, M.B., *Generalized Pole-Placement Self-Tuning Controller,* Int. J. Control, 46, 1987, pp. 569–607.

Lipták, B.G., *How to Set Process Controllers,* Chemical Engineering, November 1964.

Lopez, A.M., et al., *Controller Tuning Relationships Based on Instrumentation Technology,* Control Engineering, November 1967.

Luyben, W.L., Process Modeling, Simulation and Control for Chemical Engineers, 2nd edition, New York: McGraw-Hill, 1990.

Mehra, R.K., *Identification of Stochastic Linear Dynamic Systems Using Kalman Filter Representation,* AIAA Journal, January 1971.

Mehra, R.K., *On-Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering,* IEEE Trans. on Automatic Control, Vol. AC 16, No. 1, February 1971.

Miller, J.A., et al., *A Comparison of Open-Loop Techniques for Tuning Controllers,* Control Engineering, December 1967.

Murrill, P.W., Automatic Control of Processes, Scranton, PA: International Textbook Co., 1967.

Raven, F., Automatic Control Engineering, New York: McGraw-Hill, 1961.

Rovira, A.A., Ph.D. dissertation, Louisiana State University, Baton Rouge, LA, 1969.

Rovira, A.A., Murrill, P.W., and Smith, C.L., *Tuning Controllers for Set-Point Changes,* Instruments and Control Systems, December 1969.

Ruel, M., *Loop Optimization: before You Tune,* Control Magazine, Vol. 12, No. 03 (March 1999), pp. 63–67.

Ruel, M., *Loop Optimization: Troubleshooting,* Control Magazine, Vol. 12, No. 04, (April 1999), pp. 64–69.

Ruel, M., *Loop Optimization: How to Tune a Loop,* Control Magazine, Vol. 12, No. 05 (May 1999), pp. 83–86.

Ruel, M., Instrument Engineers' Handbook, chap. 5.9, Plantwide Control Loop Optimization, 3rd ed., Bela G. Liptak, Ed., Boca Raton, FL: CRC Press, 2002, 40 pp.

Shinskey, F.G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Shinskey, F.G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Shinskey, F.G., *Interaction between Control Loops,* Instruments and Control Systems, May and June 1976.

Salmon, D.M., and Kokotovic, P.V., *Design of Feedback Controllers for Nonlinear Plants,* IEEE Trans. on Automatic Control, Vol. AC 14, No. 3, 1969,

Sastry, S., and Bodson, M., Adaptive Control: Stability, Convergence and Robustness, Englewood Cliffs, NJ: Prentice Hall, 1989.

Seborg, D., Edgar, E., and Mellichamp, D., Process Control, New York: Wiley, 1990.

Seborg, D.E., Edgar, T.F., and Shah, S.L., *Adaptive Control Strategies for Process Control,* AI Chem. E. Journal, 32, 1986, pp. 881–913.

Smith, C.A., and Corripio, A.B., Principles and Practice of Automatic Process Control, New York: John Wiley & Sons, 1985.

Smith, C.L., *Controller Tuning That Works,* Instruments and Control Systems, September 1976.

Smith, C.L., and Murrill, P.W., *A More Precise Method for the Tuning of Controllers,* ISA Journal, May 1996.

Smith, C.L., *Controllers—Set Them Right,* Hydrocarbon Processing and Petroleum Refiner, February 1966.

Smith, O.J.M., *Close Centrol of Loops with Dead Time,* Chemical Engineering Progress, May 1957.

Sood, M., *Tuning Proportional-Integral Controllers for Random Load Changes,* M.S. thesis, Department of Chemical Engineering, University of Mississippi, University, MS, 1972.

Sood, M., and Huddleston, H.T., *Optimal Control Settings for Random Disturbances,* Instrumentation Technology, March 1973.

Stephanopoulos, G., Chemical Process Control: An Introduction to Theory and Practice, Englewood Cliffs, NJ: Prentice Hall, 1984.

Suchanti, N., *Tuning Controllers for Interacting Processes,* Instruments and Control Systems, April 1973.

Wade, H.L., *High-Capability Single-Station Controllers: A Survey,* InTech, September 1988.

Wells, C.H., *Application of Modern Estimation and Identification Techniques to Chemical Processes,* AIChE Journal, 1971.

Wells, C.H., and Mehra, R.K., *Dynamic Modeling and Estimation of Carbon in a Basic Oxygen Furnace,* paper presented at 3rd Int. IFAC/IFIPS Conf. on the Use of Digital Computers in Process Control, Helsinki, Finland, June 1971.

Wellstead, P.E., and Zarrop, M.B., Self-Tuning Systems, New York: Wiley, 1991.

Widrow, B., and Stearns, S.D., Adaptive Signal Processing, Englewood Cliffs, NJ: Prentice Hall, 1985.

Ziegler, J.G., and Nichols, N.B., *Optimal Settings for Automatic Controllers,* ASME Trans., November 1942.

Ziegler, J.G., and Nichols, N.B., *Optimal Settings for Controllers,* ISA Journal, August 1964.

2.36 Tuning Level Control Loops

H. L. WADE (2005)

INTRODUCTION

Liquid level control loops, while among the most common control loops, have some unique and very distinctive characteristics:

- Liquid level is usually not a self-regulating process, but an integrating one.
- Intuitive general rules of thumb used for tuning, such as "If it's cycling, reduce the gain." do not apply and will often produce the opposite effect.
- Liquid level control loops, if once properly tuned, do not usually need retuning, although they may *appear* to go out of tune due to the onset of valve sticking.

This section first presents an approach to level control tuning, which is based on an idealized process model. Many actual level control loops can be approximated by this idealized process model. Later in the section, some of the characteristics of nonidealized level control systems are considered.

Most processes require extensive testing to obtain even their approximate process models. Most liquid level control loops, however, readily yield to an analytical approach because their process models can be formulated, desired performance parameters established, and from this their controller tuning parameters can be calculated. Once this is done, other attributes of the control loop such as the period of oscillation can be predicted.

The counterintuitive nature of level controllers makes their tuning by trial-and-error techniques difficult. On the other hand, the determination of their tuning parameters by analytical means is a natural choice and for that reason liquid level control loops should be engineered, not tuned.

In the following paragraphs, the discussion will start with an ideal model. After that, some of the nonideal characteristics of real installations and worst-case conditions (even if they are very unlikely to occur) will be discussed.

IDEALIZED MODEL

An idealized liquid level control system is shown in Figure 2.36a. Attributes of this idealized system are:

- The tank has constant cross-sectional area.
- The level controller is cascaded to a flow controller.
- A valve positioner is installed on the flow control valve.
- All inflow goes to outflow; the tank is merely a buffer storage tank.
- The maximum outflow is the same as the maximum inflow.
- The tank is of significant size relative to the flow rates.
- There is no thermal effect such as in the case of boiler drum level control.
- The level controller set point is constant.

The consequences of the above attributes are:

- The level process is linear.
- Up- or downstream pressure, line loss, or pump curve have no effect on loop behavior.

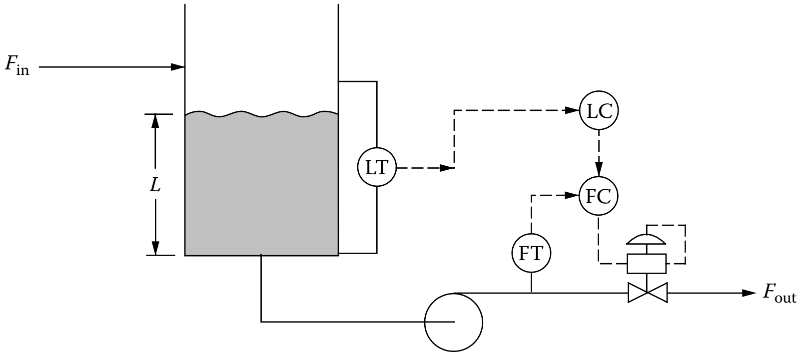


FIG. 2.36a
Liquid level control of an ideal process.

- The control loop performance is not affected by valve size.
- There is no dead time in the loop.
- The speed of response of the flow loop is significantly faster than that of the level in the tank; therefore, the dynamics can be ignored.
- Response to set-point change need not be considered because set-point changes are rarely made. The critical consideration is the response of the loop to load disturbances.

Figure 2.36a describes a common installation where the level controller manipulates the outflow from the tank in response to changes in inflow. The level controller can also manipulate the inflow in response to varying demands for the outflow. The discussion in the following paragraphs is applicable to both cases.

Time Constant of the Tank

A key parameter for the analysis of this process is the tank holdup time, also called the *tank time constant.* If the tank geometry (diameter and distance between the level taps) and maximum outflow rate (flow rate corresponding to 100% of the level measurement span) are known, then the tank time constant can be calculated as

$$T_L = \frac{Q}{F} \qquad \mathbf{2.36(1)}$$

where T_L = tank time constant; Q = tank hold up quantity, between upper- and lower-level sensor taps; and F = maximum flow rate. Q and F should be in compatible units, such as "gallons" and "gallons per minute," in which case the time constant will be in minutes.

A block diagram of the control loop with a PI controller is shown in Figure 2.36b. If the loop operates at constant set point, then the response to a load disturbance (i.e., to a change in the inflow F_{in}) is of more interest, and the set-point response can be neglected. However, in addition to the response of the level to a change in F_{in}, there is also interest in the response of the outflow, F_{out}, to a change in. F_{in}. The transfer functions of these two responses can be derived from Figure 2.36b:

$$\frac{L(s)}{F_{\text{in}}(s)} = \frac{\frac{s}{T_L}}{s^2 + \frac{K_C}{T_L}s + \frac{K_C}{T_I T_L}} \qquad \mathbf{2.36(2)}$$

$$\frac{F_{\text{out}}(s)}{F_{\text{in}}(s)} = \frac{\frac{K_C}{T_L}s + \frac{K_C}{T_I T_L}}{s^2 + \frac{K_C}{T_L}s + \frac{K_C}{T_I T_L}} \qquad \mathbf{2.36(3)}$$

where K_C is the controller gain; T_I is the integral time of the controller in minutes per repeat; and T_L is the time constant of the tank in minutes.

According to these equations, the loop acts as a second-order system. These transfer functions can also be written using the traditional servo-mechanism terminology as

$$\frac{L(s)}{F_{\text{in}}(s)} = \frac{\frac{s}{T_L}}{s^2 + 2\zeta\omega_n s + \omega_n^2} \qquad \mathbf{2.36(4)}$$

$$\frac{F_{\text{out}}(s)}{F_{\text{in}}(s)} = \frac{2\zeta\omega_n s + \omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \qquad \mathbf{2.36(5)}$$

where ζ, the damping factor, and ω_n, the natural frequency, are given by

$$\zeta = \frac{1}{2}\sqrt{\frac{K_C T_I}{T_L}} \qquad \mathbf{2.36(6)}$$

$$\omega_n = \sqrt{\frac{K_C}{T_I T_L}} \qquad \mathbf{2.36(7)}$$

The damping factor is a dimensionless number; the natural frequency is in radians per minute, if T_I is in minutes per repeat and T_L is in minutes.

Many practicing engineers may be more familiar with the term decay ratio (DR), rather than the damping

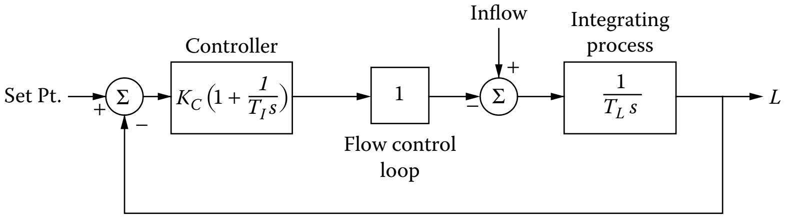


FIG. 2.36b
Block diagram of a liquid level control loop for an ideal process.

TABLE 2.36c
Equations for Calculating the Tuning Parameters of a PI Level Controller for an Ideal Level Process

Tuning Parameter	Underdamped ζ < 1		*Critically Damped* ζ = 1
	Rigorous	*Simplified*	
K_C	$2\zeta e^{-\zeta f(\zeta)}\left(\frac{\Delta F_{\text{in}}}{\Delta L_{\max}}\right)$	$2\zeta e^{-\zeta f(\zeta)}\left(\frac{\Delta F_{\text{in}}}{\Delta L_{\max}}\right)$	$\frac{2}{e}\left(\frac{\Delta F_{\text{in}}}{\Delta L_{\max}}\right)$ $(e = 2.71828\cdots)$
T_I	$2\zeta e^{\zeta f(\zeta)}\left(\frac{T_L\,\Delta L_{\max}}{\Delta F_{\text{in}}}\right)$	$4\zeta^2\left(\frac{T_L}{K_C}\right)$	$4\left(\frac{T_L}{K_C}\right)$

factor, ζ. The damping factor and decay ratio are related as follows:

$$\zeta = \frac{-\ln(\text{DR})}{\sqrt{4\pi^2 + (\ln(\text{DR}))^2}} \qquad \textbf{2.36(8)}$$

$$DR = \exp\left(\frac{-2\pi\zeta}{\sqrt{1-\zeta^2}}\right) \qquad \textbf{2.36(9)}$$

For example, the familiar quarter-amplitude decay ratio in terms of damping factor is $\zeta = 0.215$.

Determining Tuning Parameters

In addition to knowing the tank hold-up time, T_L (see Equation 2.36[1]), the analytical determination of tuning parameters requires choosing values for three design parameters:

1. ΔF_{in}—The maximum anticipated step change in disturbance (inflow) that can be expected, in percent of full scale measurement of the inflow
2. $\Delta L_{\max}$—The maximum allowable deviation from set point, in percent of full scale level measurement, resulting from a step disturbance of size ΔF_{in}
3. DR—the desired decay ratio after such a step disturbance

Once the above three values have been determined, a value for T_L can be obtained and Equation 2.36(8) can be used to convert decay ratio to damping factor. This will then permit one to derive the equations tabulated in Tables 2.36c, 2.36d, and 2.36e.

Table 2.36c provides equations to calculate tuning parameters, while Tables 2.36d and 2.36e give equations for the calculation of various characteristics of the level in the tank and of the response to a step change in inflow. In Tables 2.36c, 2.36d, and 2.36e:

$$f(\zeta) = \frac{1}{\sqrt{1-\zeta^2}}\tan^{-1}\frac{\sqrt{1-\zeta^2}}{\zeta} \qquad \textbf{2.36(10)}$$

TABLE 2.36d
Equations for Calculating Some of the Predicted Behavior Attributes of Control on an Ideal Level Process

Behavior Attribute	Underdamped ζ < 1		*Critically Damped* ζ = 1
	Rigorous	*Simplified*	
Arrest time—T_{aL}	$f(\zeta)e^{\zeta f(\zeta)}\left(\frac{T_L\,\Delta L_{\max}}{\Delta F_{\text{in}}}\right)$	$\frac{f(\zeta)}{2\zeta}T_I$	$\frac{T_I}{2}$
Period—P	$\frac{2\pi}{\sqrt{1-\zeta^2}}e^{\zeta f(\zeta)}\left(\frac{T_L\,\Delta L_{\max}}{\Delta F_{\text{in}}}\right)$	$\frac{\pi}{\zeta\sqrt{1-\zeta^2}}T_I$	N/A
IAE	$\left(\frac{1+e^{\frac{-\zeta\pi}{\sqrt{1-\zeta^2}}}}{1-e^{\frac{-\zeta\pi}{\sqrt{1-\zeta^2}}}}\right)e^{2\zeta f(\zeta)}\left(\frac{T_L(\Delta L_{\max})^2}{\Delta F_{\text{in}}}\right)$	Same as ←	$e^2\left(\frac{T_L(\Delta L_{\max})^2}{\Delta F_{\text{in}}}\right)$

TABLE 2.36e

Equations for Calculating the Predicted Behavior Attributes of Level Controls

Behavior Attribute	*Underdamped* $\zeta < 1$ *Rigorous*	*Underdamped* $\zeta < 1$ *Simplified*	*Critically Damped* $\zeta = 1$
Maximum change in outflow $\Delta F_{\text{out-max}}$	$(1+e^{-2\zeta f(\zeta)})\Delta F_{\text{in}}$	$(1+e^{-2\zeta f(\zeta)})\Delta F_{\text{in}}$	$(1+e^{-2})\Delta F_{\text{in}}$
Arrest time—T_{aF}	$2f(\zeta)e^{\zeta f(\zeta)}\left(\dfrac{T_L\,\Delta L_{\max}}{\Delta F_{\text{in}}}\right)$	$2T_{\text{aL}}$	$2T_{\text{aL}}$
Max rate of change of outflow $\left(\dfrac{dF_{\text{out}}}{dt}\right)_{\max}$	$\zeta \le \frac{1}{2}$ $\left[\exp\left(-\zeta f(\zeta)-\dfrac{1}{\sqrt{1-\zeta^2}}\tan^{-1}\left(\dfrac{(1-4\zeta^2)\sqrt{1-\zeta^2}}{\zeta(3-4\zeta^2)}\right)\right)\right]\left(\dfrac{(\Delta F_{\text{in}})^2}{T_L\,\Delta L_{\max}}\right)$	$\zeta \le \frac{1}{2}$ Same as ←	
	$\frac{1}{2} < \zeta < 1$ $2\zeta\, e^{-\zeta f(\zeta)}\left(\dfrac{(\Delta F_{\text{in}})^2}{T_L\,\Delta L_{\max}}\right)$	$\frac{1}{2} < \zeta < 1$ Same as ←	$\dfrac{2}{e}\left(\dfrac{(\Delta F_{\text{in}})^2}{T_L\,\Delta L_{\max}}\right)$

In the three tables, the columns labeled "rigorous" provide equations that are entirely a function of the following four fixed or chosen parameters—T_L, ζ (as determined from the chosen decay ratio), ΔF_{in}, and $\Delta L_{\max}$. The column labeled "simplified" produces the same results, calculated in terms of a previously calculated quantity.

Once the tuning parameters have been calculated, the predicted behavior of the level and the outflow can be calculated from Tables 2.36d and 2.36e. In Table 2.36d, the level arrest time, T_{aL}, is the time period beginning at the disturbance and ending when the maximum deviation from set point is reached.

In Table 2.36e, the outflow arrest time, T_{aF}, is the time period that begins with the disturbance and ends when the maximum change in outflow is reached. The maximum rate of change of outflow is provided because it is this quantity, rather than the size of the outflow change itself, that can be the maximum disturbance to a downstream process unit.

The equations given in Tables 2.36c, 2.36d, and 2.36e describe this tuning technique, but they are not very useful due to the large amount of computation required. For three specific decay ratios, Tables 2.36g and 2.36h provide equations for calculating tuning settings, level and outflow parameters. The three decay ratios chosen are: 1) critically damped, 2) quarter-decay ratio and 3) 1/20 decay ratio.

The critically damped decay ratio is chosen because it is a recognized basis for tuning. Quarter amplitude damping is chosen because of its familiarity. The third response, although less familiar, is chosen because it provides both the minimum IAE and the lowest maximum rate of change in outflow. Figure 2.36f depicts the level responses when the level controllers are tuned on the basis of these three forms of responses with equal values of maximum deviation.

Tables 2.36g and 2.36h also provide correction factors to account for two real-world phenomena:

1. The use of non-cascade control
2. The presence of dead time in the control loop

In the equations, θ is the ratio,

$$\frac{\text{Dead time}}{T_L}$$

and K_V is the valve gain. The correction factors were determined as a "best fit" to simulation results, for the values of θ between 0.0 and 0.5.

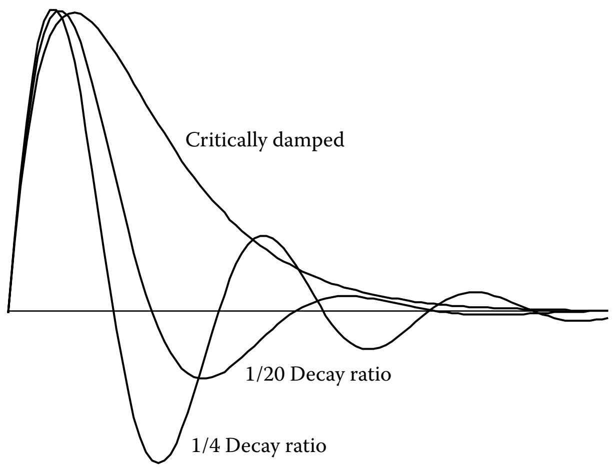


FIG. 2.36f

Response of the levels to equal maximum deviation when the controllers are tuned for the noted three decay ratios.

TABLE 2.36g

Working Equations for Calculating Tuning Parameters for a PI Level Controller for a Process with Dead Time and Non-Cascade Control

ΔF_{in} = max. step change in disturbance

$L_{\max}$ = max. allowable deviation of level from set point

T_L = hold up time, minutes

$\theta = \dfrac{\text{Date Time}}{T_L} \quad (0 \le \theta \le 0.5)$

K_V = Valve Gain, if non-cascade = 1.0 if level is cascaded to flow

Decay Ratio DR	*Damping Factor* ζ	*Gain* K_C	*Integral Time* T_I
Critically damped	1.0	$\dfrac{0.74\,\Delta F_{\text{in}}}{(1-\theta)^{0.5}\,K_V\,\Delta L_{\max}}$	$\dfrac{4.0\,T_L}{(1-\theta)^{0.5}\,K_V K_C}$
0.05	0.430	$\dfrac{0.5\,\Delta F_{\text{in}}}{(1-\theta)\,K_V\,\Delta L_{\max}}$	$\dfrac{0.74\,T_L}{(1-\theta)^{1.25}\,K_V K_C}$
0.25	0.215	$\dfrac{0.32\,\Delta F_{\text{in}}}{(1-\theta)^{1.5}\,K_V\,\Delta L_{\max}}$	$\dfrac{0.19\,T_L}{(1-\theta)^{2.4}\,K_V K_C}$

Example

Assume that a tank has the following specifications:

Tank diameter 5.0 feet
Distance between level transmitter taps 8.0 feet
Maximum outflow (upper end of outflow transmitter span) 250 gpm

Assume also that the level controller is the cascade master of a flow controller (valve gain $K_V = 1.0$), and that there is no dead time in the loop (the dead-time/T_L ratio = $\theta = 0$). The tank holdup time is calculated by:

$$\text{Surge volume} = \frac{\pi}{4} \times 5^2 \times 8 = 157.3 \text{ ft}^3$$

$$\text{Surge quantity} = Q = 157.3 \text{ ft}^3 \times 7.48 \frac{\text{gal}}{\text{ft}^3} = 1176.6 \text{ gal}$$

$$\text{Hold up time} = T_L = \frac{1176.6}{250} = 4.7 \text{ min}$$

Also, assume that the worst-case disturbance is anticipated to be a step change in inflow of 10%. In the event of this disturbance, the maximum level deviation should not exceed 5% (about 5 inches) and the system should settle out rapidly, so a decay ratio of 0.05 is chosen.

$$\Delta F_{\text{in}} = 10\%$$

$$\Delta L_{\max} = 5\%$$

With these data, Table 2.36g can be used to calculate tuning parameters:

$$K_C = \frac{0.50 \times 10}{5} = 1.0$$

$$T_I = \frac{0.74 \times 4.7}{1.0} = 3.45 \text{ min/repeat}$$

Tables 2.36h and 2.36i can be used to predict the level and outflow response:

TABLE 2.36h

Working Equations for Calculating Predicted Behavior Attributes of Level Response to a Step Change in Inflow, Level Control Loop with Dead Time and Non-Cascade Control

Decay Ratio DR	*Level Arrest Time* T_{aL}	*Period* P	*IAE*
Critically Damped	$0.5(1-\theta)^{0.96}T_I$	Not Applicable	$7.39(1-\theta)^{0.52} \times T_L \dfrac{\Delta L^2_{\max}}{\Delta F_{\text{in}}}$
0.05	$1.45(1-\theta)^{0.70}T_I$	$8.09(1-\theta)^{1.13}T_I$	$4.61(1-\theta) \times T_L \dfrac{\Delta L^2_{\max}}{\Delta F_{\text{in}}}$
0.25	$3.22(1-\theta)^{0.47}T_I$	$14.93(1-\theta)^{1.56}T_I$	$5.45(1-\theta)^{0.54} \times T_L \dfrac{\Delta L^2_{\max}}{\Delta F_{\text{in}}}$

TABLE 2.36i

Working Equations for Calculating Predicted Behavior Attributes of Outflow Response to a Step Change in Inflow; Level Control Loop with Dead Time and Non-Cascade Control

Decay Ratio *DR*	*Outflow* $\Delta F_{out\text{-}max}$	*Outflow Arrest Time* T_{aF}	*Outflow* *Max Rate of Change*
Critically Damped	$\dfrac{1.14\Delta F_{\text{in}}}{(1-\theta)^{0.1}}$	$1.0\ (1-\theta)^{1.26}T_I$	$\dfrac{0.74}{(1-\theta)^{0.70}}\times\dfrac{\Delta F_{\text{in}}^2}{T_L\Delta L_{\max}}$
0.05	$\dfrac{1.34\Delta F_{\text{in}}}{(1-\theta)^{0.22}}$	$2.50(1-\theta)^{1.36}T_I$	$\dfrac{0.52}{(1-\theta)^{1.37}}\times\dfrac{\Delta F_{\text{in}}^2}{T_L\Delta L_{\max}}$
0.25	$\dfrac{1.55\Delta F_{\text{in}}}{(1-\theta)^{0.25}}$	$6.43(1-\theta)^{1.75}T_I$	$\dfrac{0.61}{(1-\theta)^{1.16}}\times\dfrac{\Delta F_{\text{in}}^2}{T_L\Delta L_{\max}}$

Level arrest time (from time of disturbance to time when maximum deviation is reached):

$$T_{\text{aL}} = 1.45 \times 3.45 = 5.0 \text{ minutes}$$

$$\text{Period } P = 8.09 \times 3.45 = 27.9 \text{ minutes}$$

The period may seem to be excessive; however, since a fast settling behavior was selected (decay ratio of 0.05), the maximum deviation during the second half-cycle will be about 1.1 inches, during the third half-cycle about 0.25 inches, and so on.

NONIDEAL PROCESSES

The discussion so far has been based on an idealized process model. Many real applications will deviate from the idealized model. In the following paragraphs some commonly encountered situations will be described, along with suggested procedures for coping with them.

Irregular Vessel Shapes

For irregularly shaped vessels, such as horizontal or spherical tanks, the direct use of the level measurement signal as the input to the level controller may result in highly nonlinear process characteristics and undesirable control loop behavior. In this case, the loop can be linearized by converting the level measurement into the volumetric holdup in the vessel, which can be computed from the vessel geometry and the actual level. In this case, the converted measurement signal should be scaled in 0 to 100% of maximum volumetric holdup.

No Cascade Loop

If the slave flow control loop shown in Figure 2.36a is not provided, then the holdup time cannot be calculated from Equation 2.36(1), because the maximum outflow cannot be related to the maximum setting of a secondary controller.

To attempt to determine the maximum outflow rate when the outlet valve is wide open would probably be futile because of the unknown variables such as line loss, pump curve effects, and hydrostatic head effects in the tank. In addition, because the process response is nonlinear, the maximum outflow rate with a wide-open valve will vary with the level in the tank.

In this case, what needs to be determined is the apparent holdup time at the nominal operating point and the valve gain. The corresponding block diagram is shown in Figure 2.36j, and the required test is illustrated in Figure 2.36k.

To determine the apparent holdup time and the valve gain at the actual operating point, the operation must be stable, the inflow and the level at the normal set point must be constant, and the controller must be in the automatic mode. In addition, the controller output should be within its extreme limits. It is necessary for inflow to remain constant during the test. The control valve modulating the outflow should have a positioner or at least be as "sticktion free" as possible.

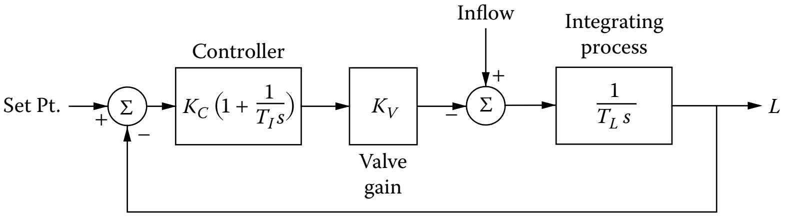


FIG. 2.36j

Block diagram of a conventional liquid level control loop, where there is no cascade slave controller and the level controller directly throttles the valve of an ideal process.

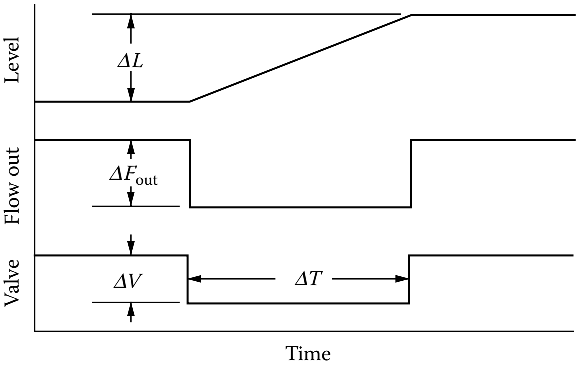


FIG. 2.36k
Illustration of a test to estimate the apparent holdup time and valve gain for noncascade level control.

The testing is done in the following sequence: First, switch the controller to manual and change the previously constant output signal by a small amount, say, $\Delta V\%$ (see Figure 2.36k). In response, the outflow will change by an amount ΔF and the level will start changing. After a certain period of time, say Δt, move the controller output back to where it was before the test. As a consequence, the level should stop changing. Once the level has stabilized, determine the change in level, ΔL, that occurred during the test.

Convert the readings for ΔL, ΔF_{out}, and ΔV into percent of full scale. The equation for estimating the apparent holdup time is

$$T_L = \frac{\Delta F_{\text{out}}\,\Delta t}{\Delta L} \qquad \textbf{2.36(11)}$$

and the valve gain is

$$K_V = \frac{\Delta F_{\text{out}}}{\Delta V} \qquad \textbf{2.36(12)}$$

Dead Time

The ideal level process has neither dead time nor lag time, but real processes can have either or both. For example, in case of the level-flow cascade loop in Figure 2.36a, the flow control loop may have a finite response time. If there is a dead time in the loop, θ, the ratio of dead time to holdup time should be calculated and used in the equations listed in Tables 2.36g, 2.36h, and 2.36i. Similarly, the actual valve gain (K_V) should be used in Table 2.36g (for cascade loops, $K_V = 1.0$).

Unequal In- and Outflows

When determining step changes in the inflow to a vessel, it should always be the actual flow that enters the tank. Hence the change in feed rate, ΔF_{in}, can be due to any cause, such as actual change in vessel feed rate, change in the liquid/vapor ration of the feed, change in reboiler heat, or change in liquid load on the trays in the distillation tower.

Flashing Liquids

There are cases where flashing liquids result in a false indication of level. An example is the “shrink and swell” effect in a boiler drum. The shrink and swell effect can be approximated as dead time; hence, the dead time correction factors in Tables 2.36g, 2.36h, and 2.36i may be used.

Sinusoidal Disturbance

If a sinusoidal variation in inflow is anticipated (for instance, due to cycling of a process controller at an upstream process unit), then the maximum variation in level and the amplitude of variation of outflow should be investigated. If the frequency of the sinusoidal variation is not known in advance, then a worst case condition could be assumed for the investigation, where the frequency of inflow disturbance in the same as the natural frequency of the level control loop.

An oscillating input will cause both the level and the outflow to oscillate with the same period. If the process can be approximated by the ideal model, which was defined at the beginning of this section, and if the inflow oscillates with a known amplitude and period (P_{in}), then Figures 2.36l and 2.36m can be used to determine the amplitude of both the level and of the outflow oscillations.

The undamped natural frequency, ω_n, is calculated from the damping factor, ζ, which is obtained from Table 2.36g, based on the chosen response (decay ratio) to a step disturbance and from the tuning parameters. If the process fits the ideal process model, Equation 2.36(7) can be used to calculate ω_n. Otherwise, the period, P, given by Table 2.36h, should be used to calculate ω_n from the equation below:

$$\begin{aligned}\omega_n &= \frac{2\pi}{P\sqrt{1-\zeta^2}} && \text{if DR} = 1/20 \text{ or } 1/4 \\ &= \frac{2}{T_I} && \text{if Critically Damped}\end{aligned} \qquad \textbf{2.36(13)}$$

From the period of oscillation of the input, P_{in}, the frequency in radians per minute can be calculated from:

$$\omega = \frac{2\pi}{P_{\text{in}}} \qquad \textbf{2.36(14)}$$

The frequency ratio $\frac{\omega}{\omega_n}$ in Figure 2.36l should be used to calculate $K_C \frac{L(\omega)}{F_{\text{in}}(\omega)}$. (If the frequency of the disturbance is unknown, then for a worst case analysis, use a frequency ratio of 1.0.) From this and from the amplitude of the disturbance oscillation, $F_{\text{in}}(\omega)$, one can calculate the predicted amplitude of oscillation of level about the set point, $L(\omega)$.

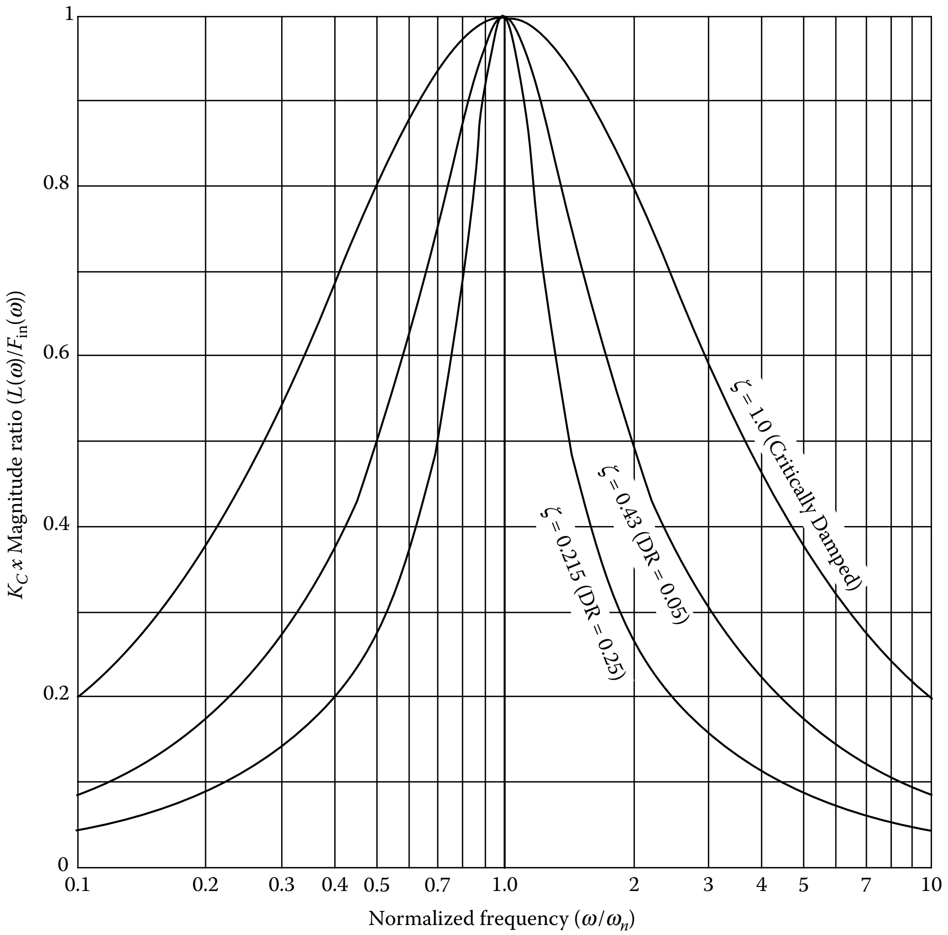


FIG. 2.36l

Magnitude ratio of changes in level to sinusoidal changes to inflow.

If the peak deviation from set point (one half the amplitude of oscillation) exceeds the allowable maximum deviation, then K_C should be increased by the relation:

$$K_{C\text{-new}} = \frac{0.5\,L(\omega)}{\Delta L_{\max}} \qquad \mathbf{2.36(15)}$$

After this, one can return to Tables 2.36g, 2.36h, and 2.36i and calculate T_I and the predicted attributes of the response.

If the magnitude and period of oscillation of the input are assumed or known, use Figure 2.36m to calculate the magnitude ratio of oscillations of outflow to inflow,

$$\frac{F_{\text{out}}(\omega)}{F_{\text{in}}(\omega)}$$

then calculate the amplitude of oscillation of the outflow.

As a final check, if there are level control loops in series, such as in case of a train of distillation towers, it is necessary to check the natural frequency of each tower. Ideally, the natural frequency of each tower should be no more than half the natural frequency of the preceding tower. Since the size of the towers (holdup time) cannot be changed, the best "handle" for adjusting the ratio of the natural frequencies is to increase $\Delta L_{\max}$ for the downstream tower or to decrease $\Delta L_{\max}$ for the tower upstream.

OTHER APPROACHES TO TUNING

Averaging Level Control

Many liquid level loops are not critical and one can tolerate fluctuation, even offset in levels, if it smooths out the flow to a downstream process unit. This can be accomplished by using a proportional-only controller. This technique is called "averaging liquid level control."[1] Assuming that the allowable excursions above and below set point are equal, then the controller gain should be set according to the equation

$$K_C = \frac{100}{2\,\Delta L_{\max}} \qquad \mathbf{2.36(16)}$$

or the proportional band should equal $2 \times \Delta L_{\max}$ and the bias (manual reset) should be set to 50%. With this arrangement, if the disturbance is such that a controller output of 50% is

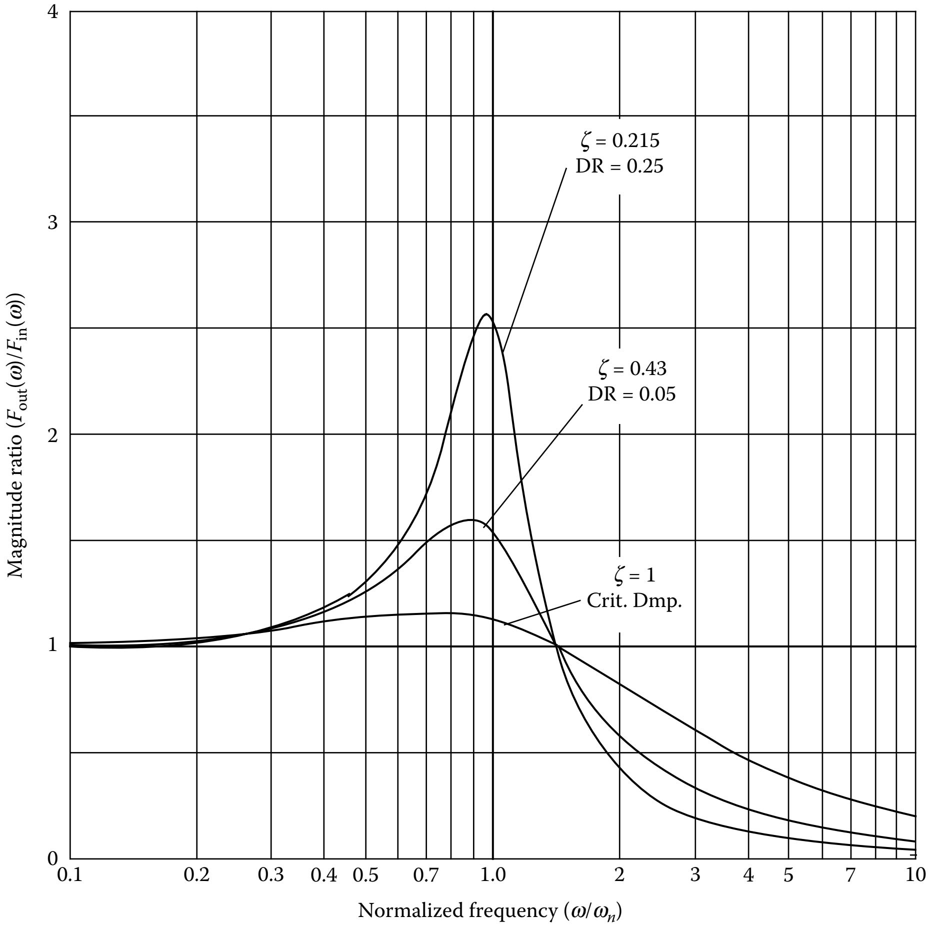


FIG. 2.36m
Magnitude ratio of changes in outflow to sinusoidal changes in inflow.

required, the level will be on set point. Otherwise, there will be a steady-state offset between set point and level measurement. When a step change in inflow occurs, the level will respond as a first-order lag with a time constant equal to the holdup time divided by the controller gain.

This technique ensures that the level never exceeds the limits because at the limits, the outflow is either at 0 or 100%. This technique also provides the lowest rate of change of the outflow, hence the minimum disturbance to a downstream processing unit. The disadvantage of this method is the fact that the level is rarely at set point. This is probably more of a disadvantage from the operator's acceptance point of view than any other.

It might appear that one could achieve the advantages of averaging level control and still maintain the level at the set point by using Equation 2.36(16) to determine the controller gain and then by using a slow reset action. Equation 2.36(6) shows, however, that a large value for T_I will produce a larger damping factor or even an overdamped response. If the response is underdamped, the loop will have a very long period; if overdamped, the return to set point will be excessive. Furthermore, with the reset action causing an effective shift of the proportional band, the positive protection of knowing exactly where the level will be when the controller output is at maximum or minimum is lost.

Controller Gain and Resonance

If it is desired to stay closer to the set point, the controller gain can be increased. Although there is no theoretical upper limit for the gain of a proportional-only controller used on an integrating process, in practice the gain will be limited by resonance that may occur within the loop.

If the level sensor is of the external cage type, then there may be a manometer effect between the liquid in the vessel and the liquid within the level sensor cage. This will appear as an oscillation within the control loop even when the total mass holdup is unchanging. If there is a large surface area of the liquid, this may result in a resonant sloshing, with a period proportional to the cross-sectional areas.

For a probe or a point-source sensor, this will also show up as an oscillation within the loop. Furthermore, splashing, such as from upper trays in a distillation tower, may result in the appearance of noise on the level measurement. Thus there will be a practical limit to the controller gain. With a high gain, any measurement noise present will cause excessive valve action. Therefore, the gain may be reduced, and some integral action is added within the controller.

Nonlinear Gain

Some manufacturers provide a nonlinear control algorithm that has the effect of increasing the controller gain as the measurement gets further away from set point. An example is the "error-squared algorithm," in which a modified error, ê, is calculated as:

$$\hat{e} = e|e|. \qquad \textbf{2.36(17)}$$

When the level is on set point, this algorithm gives a very low gain, which increases as the measurement gets further away from set point. Other manufacturers accomplish a similar operation by linear characterization of the error signal. Sometimes the nonlinear behavior is applied to only one of the controller modes, such as to the proportional mode, with the other controller mode seeing the normal error signal.

Reference

1. Shinskey, F. G., *Feedback Controllers for the Process Industries,* New York: McGraw-Hill, 1994.

Bibliography

Cheung, T. F., and Luyben, L., "Liquid-Level Control in Single Tanks and Cascades of Tanks with Proportional-Only and Proportional-Integral Feedback Controllers," *IEC Fund.,* 18, 1, 15–21, 1979.

Cheung, T. F., and Luyben, L., "Nonlinear and Non-Conventional Liquid Level Controllers," *IEC Fund.,* 19, 1, 93–98, 1980.

Cho, C., *Measurement and Control of Liquid Level,* Research Triangle Park, NC: ISA, 1982.

Marlin, T. E., *Process Control—Designing Process and Control Systems for Dynamic Performance,* 2nd ed., New York: McGraw-Hill, 2000.

MacDonald, K., McAvoy, T., and Tits, A., "Optimal Averaging Level Control," *AIChE J.,* 32, 75–86, 1986.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

Shunta, J., and Fehervari, W., "Non-Linear Control of Liquid Level," *Instr. Techn.,* 43–48. January 1976.

2.37 Tuning Interacting Loops, Synchronizing Loops

M. RUEL (2005)

INTRODUCTION

The reader is advised to also read Sections 2.6, 2.12, and 2.22. This is desirable to gain a full understanding of the phenomenon of interaction and decoupling; the methods discussed in this section cannot be used in all cases, and a full understanding of the other options for overcoming interactions between control loops can be helpful.

MULTILOOP SYSTEMS

When a unit process is controlled by several control loops, there is no magic formula that can tell whether one loop will affect another. This information will only come through an in-depth understanding of the process. If one loop directly interacts with another, oscillation in the first loop will cause oscillation in the second and possibly in other downstream loops. If the same pump feeds two flow loops, oscillation in one loop can cause oscillation in the second.

When a loop is cycling, it is essential to determine whether the process is causing cycling, or whether the cycling is attributable to other loops or possibly to the loop itself. To check the cause, one can switch the particular loop to manual mode; if the cycling continues, it is probable that the cycling is being caused by an external source.

There is also hidden cycling, which occurs if a cycle is present but is hidden by noise. To uncover a hidden cycle, the readings should be collected in the manual mode, and power spectral density analysis should be used on the data collected. In that case, the hidden cycles will show up as peaks.

In some installations the problem is not that the loops interact, but it is imperative that they respond with the same speed. In either case, one should be knowledgeable about the tools that are available to determine loop health and performance:

Control Loop Analysis

The following criteria should be met for a control system to perform in an optimal manner:

- The power spectral density should be flat, no cycling present.
- Cumulative power spectral density should be continuous.
- Statistical analysis: Data distribution should be a bell curve, variability should be small, valve movements should be minimized.
- Development of a process model is desirable to validate the process and to find tuning parameters.
- Robustness analysis is recommended to validate tuning parameters.
- The process should be analyzed to check for hysteresis and backlash, stiction, noise, process model inaccuracy, hidden cycling, and nonlinearities.
- Oscillation should be evaluated by determining the area under the curve, which is a good indicator of control quality.
- Cross-correlation and multivariate techniques can be used to measure interaction between signals and loops They can also help to determine whether multivariable control should be considered.
- Performance indexes, such as variability, IAE, and Harris index, should all be monitored.

INTERACTING LOOPS

An example of a control system with potentially interacting multiple loops is illustrated in Figure 2.37a. Here two liquids

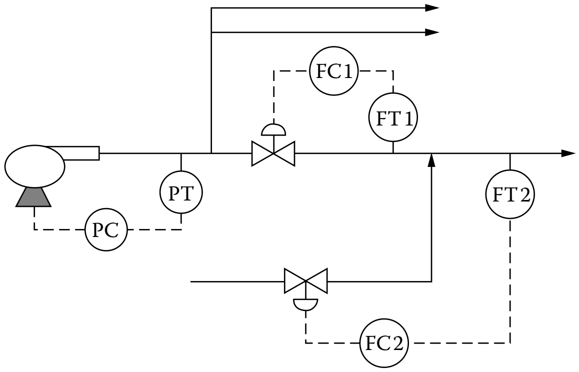


FIG. 2.37a
Illustration of potentially interacting control loops.

are being mixed. The expensive process fluid is controlled by FC1 at a flow rate of 100 GPM. The second flow controller (FC2) adds water to dilute the process fluid by maintaining the total flow between 200 and 400 GPM.

All three of these loops have the potential to be fast. A response time of less than 30 seconds is attainable on all three loops, but which loop should be the fastest? What is the logic behind this decision and why should one be faster than the others in the first place?

The reason why the speeds should be different is because if they are not, the loops can oscillate whenever an upset occurs because the correction generated by one loop upsets the others and this generates cycling. As to which loop should be the fastest, one should evaluate the process to determine which controlled variable needs to be constant in order for the other loop(s) to operate properly.

In Figure 2.37a, by observing the process we would conclude that in order for the flow loop (FC1) to function properly, the upstream pressure to its control valve has to be constant. Because that upstream pressure has to be constant regardless what the flow is, therefore, the pressure loop must be faster than flow loop #1.

If this was not the case, if the pressure loop and the #1 flow loop were tuned to have the same speed of response, they may work for a while, but eventually, when a disturbance occurs, it will cause the two loops to oscillate. For example the following sequence of events could cause oscillation in this control configuration:

1. Another user valve is suddenly closed and this disturbance causes the line pressure to increase.
2. If the pressure control loop is not faster than the flow loop, the flow through FC1 will increase.
3. To correct for the flow increase, FC1 will close down its valve, which in turn will cause the pressure to rise.
4. Eventually the pressure loop will slow down the pump, which will cause the flow to decrease.
5. As the flow drops, FC1 will open its valve to compensate, which will cause the pressure to decrease.
6. In response to the drop in pressure, the PC will speed up the pump, causing the flow to increase again.

In this configuration, if PC is not faster than FC1, steps 3, 4, 5, and 6 will repeat continuously and the two loops will oscillate and potentially resonate.

Likewise, since FC1 controls the process fluid feed to the flow loop controlled by FC2, flow loop 1 must be faster than 2. If this is not the case, a disturbance in flow 1 could cause both flow controllers to react, and oscillation would result.

Tuning to Eliminate the Interaction

When loops interact, it is necessary to make sure that their response speeds are not the same and not even similar because speeds that differ but are close also have the potential to oscillate. To be on the safe side, one should select response speeds that differ by a factor of three to five. If speeds are closer than 3:1, loops may one day start to oscillate. In case of loops that are highly interactive, a speed ratio of up to 10:1 may be required to fully decouple them.

For the control system described in Figure 2.37a, this means that the response time of the pressure loop will determine the response time of flow loop #1, and the response time of flow loop #1 will determine the response time of flow loop #2. In tuning interacting loops, one would do that by placing the downstream loop in manual while tuning the upstream loop; once the upstream loop's speed of response is determined, use a multiple of that to set the downstream controller.

So, for the control system in Figure 2.37a, one would place FC1 and FC2 in manual, while aggressively tuning the pressure controller to provide a high speed of response. The response time of the pressure control loop will determine the system response time. Once the PC is tuned, one would place the pressure loop in automatic so that to the rest of the control system, it would seem as if it were part of the process.

Flow controller FC1 is tuned next, while FC2 is still in manual. FC1 must be tuned for a response time that is at least three times slower than that of the pressure loop response time. For ideal separation it should be 5 to 10 times slower. Once FC1 is tuned, both the PC and FC1 are left in automatic, while FC2 is being tuned. Again, FC2 should be tuned for a response time which is at least 3 times (ideally 5 to 10 times) slower than flow loop FC1.

Therefore, one can sum up the tuning of the three interacting loops into the following three steps:

Step 1. Tune PC for quick response, while other loops are in manual mode. For the purposes of an example, assume that the settling time of this fastest loop turns out to be 30 seconds (Figure 2.37b).

Step 2. Tune FC1 for moderate response, while PC remains in automatic and FC2 in manual mode. Tune FC1 for a settling time of at least three times that of the PC, or at least 90 seconds (Figure 2.37c).

Step 3. Tune FC2 for a slow response while PC and FC1 both remain in the automatic mode. The settling time of FC2 should be at least three times that of the FC1, or at least 270 seconds (Figure 2.37d) .

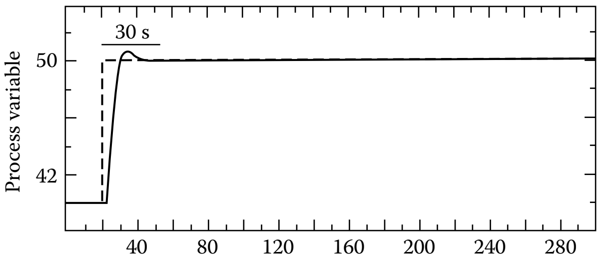


FIG. 2.37b
After the fastest loop is tuned, measuring its response time (settling time), which in this case is 30 seconds.

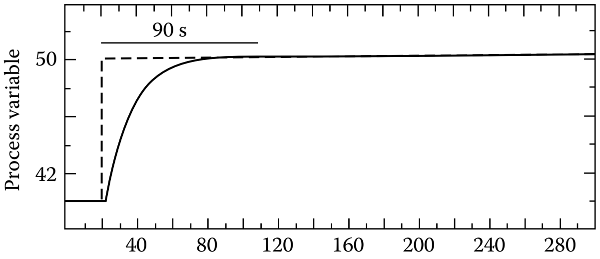


FIG. 2.37c
Tuning the less fast loop for a response time (settling time) that is three times that of the fastest loop or in this case is 90 seconds.

For a summary of the steps required in tuning any number of interacting loops, refer to Table 2.37e.

The method of removing interaction by reducing the speed of response has disadvantages because this can cause the control loops to become sluggish and unable to effectively correct upsets and disturbances. If this is the case the use of more sophisticated techniques of decoupling is recommended, as discussed in Section 2.12.

Cascade loops is another case of interacting loops. Hopefully, in a cascade system, the inner loop has to be faster than the outer loop. (See Section 2.6.)

SYNCHRONIZING LOOPS

In some control systems, the loops do not interact but they do work together. Such configurations are called synchronized loops, and it is desirable for such loops to have the same response time. It is important to note that synchronized loops should be so designed that there is no physical link between them that could cause interaction.

Batch mixing is one example of a control system that should be synchronized. Figure 2.37f illustrates a control system for mixing three ingredients in a mix tank. The goal of such a control system is to maintain the required ratio of the ingredients even during startup or shutdown or when the rate of production changes.

Because the control valves and pipe volumes associated with the three flow control loops are substantially different, it is probable that their process gains, dead times, and time constants also differ. In such cases, if all three loops were tuned for 10% overshoot (or any other criterion), the response times of the loops would not be the same. Therefore, when the rate of production rises and the level controller calls for increased flows, the recipe flow ratios will be out of balance until all three flows reach their new set points and regain stability.

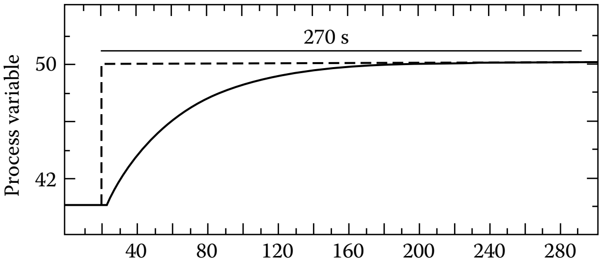


FIG. 2.37d
Tuning the least fast loop for a response time (settling time), which is three times that of the less fast loop or in this case is 270 seconds.

TABLE 2.37e
The Sequence of Steps to be Used in Tuning Any Number of Interacting Control Loops

	Fast →				**Slow**
Steps	**Loop$_1$**	**Loop$_2$**	**Loop$_3$**	...	**Loop$_n$**
1	**Tune**	Manual	Manual	Manual	Manual
2	Automatic	**Tune**	Manual	Manual	Manual
3	Automatic	Automatic	**Tune**	Manual	Manual
	...	...	...	...	...
n	Automatic	Automatic	Automatic	Automatic	**Tune**
n + 1	Automatic	Automatic	Automatic	Automatic	Automatic

To ensure that all three loops move at the same speed, one should determine the response time of the slowest loop and match the response times of the others to it. Normally, the slowest loop is also the one with the largest dead time. The steps involved in tuning synchronizing loops are:

1. Apply an upset (bump test) to each loop. This can be a temporary change of set point.

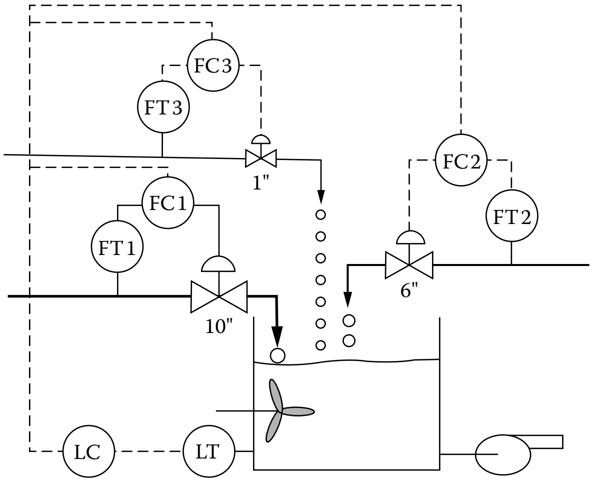


FIG. 2.37f
In order to keep the ratio of ingredients constant during load changes, the loops have to be synchronized (their speed of response has to be the same).

TABLE 2.37g
Summary of Steps Required in Tuning to Synchronize Control Loops

Steps	**Loop$_1$**	**Loop$_2$**	**Loop$_3$**	...	**Loop$_n$**
1	**Test**	**Test**	**Test**	**Test**	**Test**
2	Slowest →				
3					**Tune at maximum speed**
4	Tune at same speed				
5	Automatic	Automatic	Automatic	Automatic	Automatic

2. From the responses of the loops, determine which is slowest.
3. Tune the slowest loop for maximum speed of response and measure the response time that results.
4. Adjust the tuning parameters of the other loops so that they will also have approximately the same response time.

When tuning loops that need to work in harmony, select tuning parameters that give similar response times. If this is done using software, the expected speed can be specified. If done by hand, techniques such as pole placement, Internal Model Control, or Lambda tuning should be used.

Bibliography

Astrom, K. J., *PID Controllers: Theory, Design, and Tuning,* 2nd ed., Research Triangle Park, NC: Instrument Society of America, 1995.

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: Instrument Society of America, 1990.

Gerry, J. P., "Tune Loops for Load Upsets vs. Setpoint Changes," *Control Magazine,* September 1991.

Levine, W., *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.

Lipták, B. G. (Ed.), *Instrument Engineers' Handbook: Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002.

McMillan, G. K., *Tuning and Control Loop Performance,* 3rd ed., Research Triangle Park, NC: ISA, 1994.

Ruel, M., "Loop Optimization: Before You Tune," *Control Magazine,* Vol. 12, No. 3 (March 1999), pp. 63–67.

Ruel, M., "Loop Optimization: Troubleshooting," *Control Magazine,* Vol. 12, No. 4 (April 1999), pp. 64–69.

Ruel, M., "Loop Optimization: How to Tune a Loop," *Control Magazine,* Vol. 12, No. 5 (May 1999), pp. 83–86.

Ruel, M., "Plantwide Control Loop Optimization," Chapter 5.9 in B. G. Lipták, Ed., *Instrument Engineers' Handbook*, 3rd ed., Boca Raton, FL: CRC Press, 2002.

Shinskey, F. G, *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996.

2.38 Tuning by Computer

R. M. BAKKE (1970) **P. D. SCHNELLE** (1985) **B. G. LIPTÁK** (1995) **M. RUEL** (2005)

Tuning, Modeling and Diagnostics Software Package Suppliers:	ABB Inc. (www.abb.com) Adersa (www.adersa.asso.fr) Aspen Technology Inc. (www.aspentech.com) CyboSoft (www.cybosoft.com) DOT Products (www.dot-products.com) Emerson Process Management (www.easydeltav.com) ExperTune (www.expertune.com) Foxboro Co. (www.foxboro.com) Honeywell (www.acs.honeywell.com and www.honeywell.com/imc) Hyprotech, Ltd. (www.hyprotech.com) Invensys (www.invensys.com) JMP Software (www.jmp.com) Lifestar (www.qimacros.com) Minitab Inc. (www.minitab.com) MYNAH Technologies division of Experitec, Inc. (www.mynah.com) Northwest Analytical (www.nwasoft.com) Oil Systems Inc. — PI (www.osisoft.com) Omega Simulation Co, Ltd. (www.omegasim.co.jp) Pavillion Technologies Inc. (www.pavtech.com) Process Systems Enterprise Ltd. (www.psenterprise.com) ProTuner, Techmation (www.protuner.com) Rosemount Inc./Emerson Process Management (www.rosemount.com and www.emersonprocess.com) Shell Global Solutions (www.shellglobalsolutions.com) Siemens (www.sea.siemens.com/ia) SIMSCI-ESSCO division of Invensys (www.simsci.com) Software Toolbox (www.softwaretoolbox.com) Stochos Inc. (www.stochos.com) Trident Computer Resources Inc. (www.tridentusa.com) Wonderware (www.wonderware.com) Xanalog Corporation (www.xanalog.com) Yokogawa (www.yokogawa.com/us) Zontec Inc. (www.zontec-spc.com)

The tuning steps that were described in Section 2.35 can all be performed automatically under computer control. The pattern recognition type self-tuning controllers do this by applying the "open-loop" method of tuning during their "pretune phase" (step change in manual mode) and by using the "closed-loop" method to evaluate the loop performance during upsets.

Computers can also perform "model-based" tuning. In that case, the performance of the proposed tuning constants is first checked on a mathematical model of the process, and only if they perform well on the model are they used on the real process. Because of the overlap between some of the earlier sections in this chapter and this one, it is suggested that the reader also read those to gain a complete understanding of the tuning capabilities of computers and of digital systems.

INTRODUCTION

Computer programs that were developed to aid in the *design* of process control systems are gaining general popularity and acceptance. Several good commercially available packages exist. These computer programs include nonlinear simulation tools, linearizing routines, predictor and estimation techniques, linear quadratic Gaussian methods, time series, and frequency response techniques. This section will discuss the approach to a design problem using such methods.

The *maintenance* of control systems is equally important as is a good workable design. Many good control schemes fall into disuse because of a lack of proper fine-tuning, failure to adapt to changing process requirements, or poor maintenance. As time progresses and the operating processes change in the plant, some controllers might have to be detuned to prevent cycling or switched to a manual mode of operation. Good maintenance cannot be substituted by tuning. If equipment fails, the loop should not be re-tuned; the piece of equipment should be repaired.

Computer-aided control analysis tools can be thought of in two categories. First there are the techniques that are useful in control system design. These are typically used in the design of new plants or new control systems. The nature of such tools is most often offline because that is the safest mode for investigating new control methods.

Maintenance-related software is the second category of the computer-aided tools of control analysis required in maintenance of existing systems. The nature of these techniques is frequently online. They involve trying to identify the process dynamics while the plant is operating. They use step testing techniques and/or recursive estimation methods for adaptive tuning. After the process model is updated, these techniques can be used to calculate new control system parameters.

In addition to keeping the controllers tuned, other methods are available to improve the quality and reliability of process measurements. Overall process balance calculations and the use of predictor/estimator filters (i.e., Kálmán filters) can help to improve the quality of measurements. These better-quality measurements are contributing to better control performance, which will be discussed in more detail in the following paragraphs.

PROCESS MODELING

A good control system can only be designed, if the process is well understood, so that it can be mathematically modeled (Sections 2.13 to 2.17). Some control systems are less sensitive to model errors than others. However, as the requirements for improved control performance increase, so does the need for model accuracy.

Process control engineers use many types of process models. Each type of model is preferred for some group of applications. Table 2.38a provides a list of some of these computer models and of their typical applications.

Time and Frequency Domains

Software packages that are available to tune loops and to optimize a process can use either time series or frequency analysis techniques. To obtain the best PID tuning settings and the best process model, frequency response methods have several advantages over other methods. These are:

- Frequency response methods require only one bump to identify the dynamics of the process.
- The bump can be made in manual mode or can be a set-point change in the automatic mode.
- The change can be a pulse, a step, or bumps of other characteristics.
- Frequency response methods do not require prior knowledge of the dead time or time constant of the process.
- In contrast, when using time response methods, one often needs an estimate of both the dead time and the time constant.
- In addition, when using time response methods, the type of the process model has to be known.

TABLE 2.38a
Types of Computer Models Used in Various Control Applications

Computer Model Type	*Typical Application*
Large-scale nonlinear dynamic simulation	Used for offline investigation of process, process dynamics, and control application; large effort required to generate models
Frequency response models, single-input single-output (SISO)	Used to investigate open-loop behavior and specify closed-loop feedback control
Frequency response models, multiple-input multiple-output (MIMO)	MIMO models are used to study interaction; model typically generated from state space models or plant testing
State space models, MIMO Continuous and discrete linear models	Used for control system design with methods such as linear quadratic optimal and pole placement control; used for estimate/predictor models; models typically generated from linearization of simulation models or plant tests
Time series models, SISO Difference equations	Used for adaptive control and state parameter estimation
Simple low order models SISO continuous time	Used for quick time domain model fits to plant tests for control-loop tuning

Currently the main tool employed for control system design is dynamic process simulation using basic physical laws such as time-dependent balance relationships of mass, heat, momentum, etc. This type of simulation effort is very time consuming and requires an in-depth understanding of the process and the availability of large dynamic modeling software packages, which are available commercially. Many of these packages have several common features.

First, the simulation models require the input of many differential equations representing the process. The packages usually handle the numerical integration and output of the desired time variables with very little effort on the part of the user.

Once the process has been adequately modeled, the control system is tuned by dynamic simulation. The time behavior of the closed-loop system is studied and is changed in order to try new intuitively devised schemes or control designs that are recommended by other computer design methods such as the linear quadratic optimal or the frequency response methods.

Single-input single-output transfer function models allow control systems to be analyzed in the frequency domain. Computer-generated graphics are used by many control analysis packages to draw root locus, Bode plots, Nichols charts, and Nyquist plots. The individual transfer function models are typically generated from linearized differential equations, plant pulse or step testing, or transformed state space models (see Section 2.33). All of these procedures have been computerized, which makes the analysis more convenient, and are very helpful in control system design and specification in the offline mode.

State Space Models

State space models are nothing more than linearized matrix representations of multiple-input multiple-output or high-order differential or difference equations. The convenient matrix formulation of the state space models makes them great candidates for manipulation and study by computer. These linear models can be used for time simulation (over the range where the linearization is valid). They can be transformed to transfer function matrix representations. They can be used to formulate linear quadratic optimal controllers. They can also be used to build predictor/estimators (i.e., Kálmán filters).

State space control models are very flexible and are a common starting point for many computer-aided control analysis packages. The state space model is useful for offline design and online analysis and recursive study.

The above techniques are basically design tools used offline. One popular online use for state space models is the estimator/predictor configuration. Here the computer evaluates the state space model of the process alongside with the real process. The measurable information about the process is fed into the computer model. The model predicts or estimates those states of the process that cannot be measured. This information is then used for the control of the process.

The methods used for estimating unmeasured states vary. One popular method is Kálmán filtering. In this method the estimated states are updated with information from the process model and measure their inputs on the basis of the statistical properties of the measurement noise. Details about Kálmán filtering and other estimation/prediction techniques can be found in References 1 and 2.

These filtering techniques improve measurement quality and utility, thus improving the entire control system. Certain modeling applications can be used to signal instrument problems by monitoring the deviation between the output that is predicted by the model and the output of the actual process. If this deviation becomes too large, it is possible that sensor maintenance is required.

Time Series Models

Time series models are also useful in adaptive control applications. Discrete time models are used to represent single-input single-output systems processes. The parameters of the time series models are estimated using recursive techniques. The updated model resets the online controller to keep the loop adequately tuned. This again is an online computer application that is primarily aimed at improving control quality.

The low-order single-input single-output continuous model consists of a series of time lags associated with loop elements (valve, process, transmitter, etc.). Usually, the controlled processes also have transport delay (pure dead time) resulting, for example, from transport lag in pipes. Unlike nuclear reactors and modern missiles, these processes are usually stable without feedback controllers, just operating in the open loop.

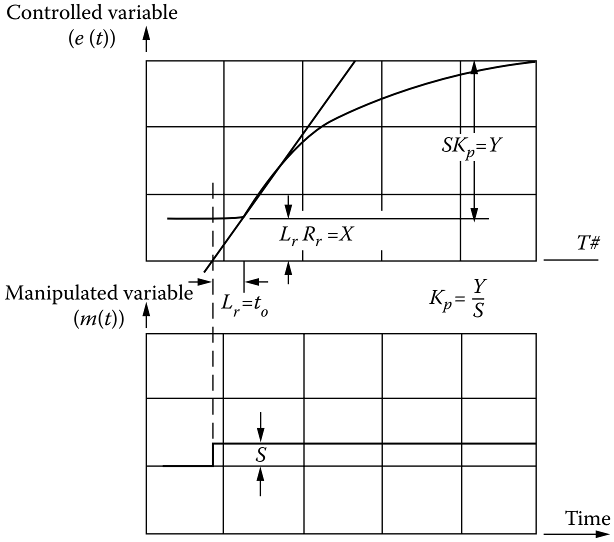
A general mathematical structure for these simple processes is composed of a first-order lag and dead time, as described in Equation 2.38(1). This model has three parameters: gain, time constant, and dead time.

$$\frac{dc(t)}{dt} = \frac{1}{\tau}[Km(t - t_0) - c(t)] \qquad \textbf{2.38(1)}$$

where t = variable time; $\frac{dc(t)}{dt}$ = derivative of the controlled variable with respect to time; $c(t)$ = controlled process variable, the output of the process; m = manipulated process variable; K = process gain; τ = process time constant; and $t_0 = t_d = L_r$ apparent process dead time.

This mathematical structure has the advantage of constrained dimensionality. Dead time in the model conveniently lumps the higher-order process time delays and the actual processes' dead time into a single parameter.

This model is used to tune a simple feedback loop by fitting the model to the process. This fitting is done by control loop tunings to achieve the desired response.



Processes that drift during the conventional step response test will invalidate the measured parameters X, Y and L_r.

FIG. 2.38b

First-order response to a step change in manipulated variable, while the controller is in manual, can determine the process gain (Kp) and dead time (Lr).

FITTING SIMPLE MODELS

Step Response

The conventional step response test (Figure 2.38b), which is performed while the controller is in manual, is probably the most popular method for fitting the three-term process model. This method works well for process loops that, after an upset, will achieve a new steady state. In addition, the common characteristics of these processes is that their time constant τ is small and that they do not appreciably drift during the test.

However, if the process gain ($K_p = Y/S$) is unknown, which is often the case, this technique has the disadvantage that the step size (S) must be selected by trial and error. This is because too small a step might provide a reaction curve that is hard to read, and too large a step might result is an unacceptable upset.

To fit this three-parameter model to the process while using this test, the following measurements have to be made manually or automatically:

1. Step size of the manipulated variable (S)
2. The apparent dead time ($L_r = t_0$)
3. The total change in the controlled variable ($X = L_r R_r$) over the period of the apparent dead time ($L_r = t_0$), if the controlled variable is changing at its maximum rate of change (R_r)
4. The final value controlled variable's response (Y)

After these four values have been obtained, the following parameters can be determined:

$$R_r = \frac{X}{SL_r} \tag{2.38(2)}$$

$$K_p = \frac{Y}{S} \tag{2.38(3)}$$

$$\tau = \frac{K_p}{R_r} \tag{2.38(4)}$$

Modified Step Response

The conventional step response test is not practical in loops with large time constants. Furthermore, it does not provide a check on the accuracy in performing the test itself unless repeated. Even if the test is repeated, errors can still occur in unit conversion.

A combination of a modified step response and an ultimate sensitivity test overcomes the above limitations (Figure 2.38c). When performing the modified step response, the operator retains the step just long enough to be certain that the process

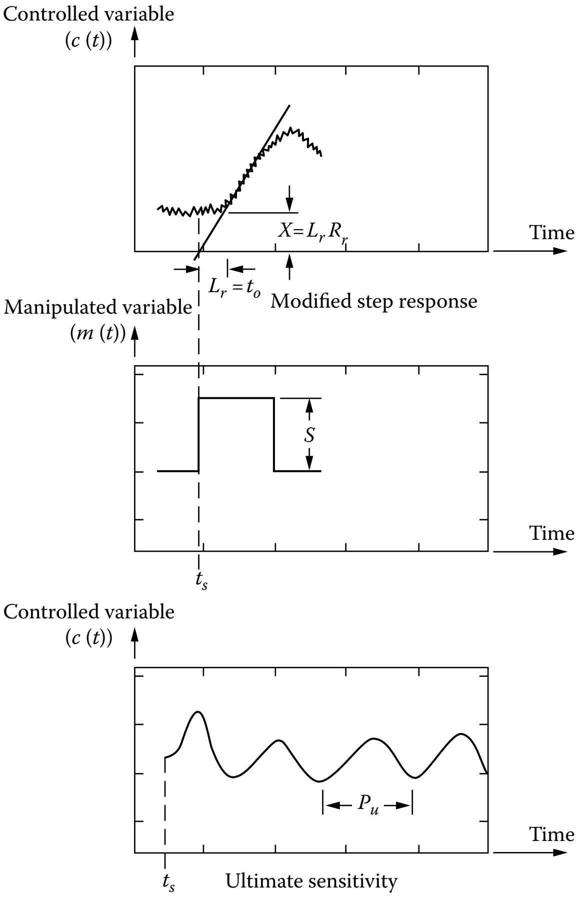


FIG. 2.38c

The combination of the modified step response test and the ultimate sensitivity test is better suited for processes with dead time.

response reaches the maximum rate of change and then returns the manipulated variable to its initial value. Because in this test the controlled variable does not have to achieve a new steady state, the selection of step size *S* is not critical. The maximum rate of rise in this case too is obtained by Equation 2.38(2).

This technique should not be confused with pulse testing, which requires analysis of the total pulse response curve.

Ultimate Gain Test

The ultimate gain (K_u) test calls for increasing the gain of a proportional controller until the process cycles at constant amplitude, without convergence or divergence. The following equations describe the behavior of the system:

$$\frac{K_u K_p}{\sqrt{\left(\frac{2\pi\tau}{P_u}\right)^2 + 1}} = 1 \qquad \textbf{2.38(5)}$$

$$\tan^{-2}\left(\frac{2\pi\tau}{P_u}\right) + \frac{2\pi L_r}{P_u} = \pi \qquad \textbf{2.38(6)}$$

where K_u is the ultimate gain of the controller and P_u is the ultimate cycle period. These equations hold true if proportional-only control is continuously applied or if sampling is used at such a rate that it is apparently continuous with respect to either the dead time (L_r) or the time constant (τ). Equations 2.38(5) and 2.38(6) can be reduced to determine the process time constant and the gain:

$$\tau = \frac{P_u}{2\pi}\left[\tan\left(\pi - \frac{2\pi L_r}{P_u}\right)\right] \qquad \textbf{2.38(7)}$$

$$K_p = \frac{1}{K_u}\sqrt{\left(\frac{2\pi\tau}{P_u}\right)^2 + 1} \qquad \textbf{2.38(8)}$$

With τ calculated from Equation 2.38(7) and the gain from Equation 2.38(8), Equation 2.38(4) can be used to give a cross-check on the accuracy of the test and on the validity of the model structure.

Discrete Cycling

The process gain can be determined by another equation by discrete cycling of the loop while using a sample rate, which equals the dead time ($L_r = t_0$). If the gain required to reach ultimate sensitivity is K_{ud}, then the conditions of stability are defined by:

$$K_{\mathrm{ud}} K_p [1 - e^{t_0/\tau}] = 1 \qquad \textbf{2.38(9)}$$

Process gain can be computed from this test by

$$K_p = \frac{1}{K_{\mathrm{ud}}} \frac{e^{t_0/\tau}}{e^{t_0/\tau} - 1} \qquad \textbf{2.38(10)}$$

The amplitude of cycling obtained with the discrete test will be different from that of the continuous test, which provides an additional means of checking whether the gain is linear. The degree to which this test can be automated is strictly a function of available hardware and software.

If a valid model cannot be obtained, the instrument engineer should tune a controller by trial and error, or, if a more scientific approach is desired, then more general model structures and identification approaches should be tried. If the process gain is nonlinear, frequent adjustment or online adaptation should be considered.

The reader should also refer to the tables provided in Section 2.35, which provide equations to calculate the tuning parameters for a variety of tuning objectives. The tests described can be automated by the use of software packages, which not only perform these tests but also can analyze some of the results.

TUNING CRITERIA AND FORMULAS

For those processes that fit the first-order plus dead time model, a variety of approaches are described in the literature for computing the tuning settings. With conventional control applications, the techniques that have been developed were found to work well. A set of equations are provided below to give guidance in the tuning of controllers. These transfer function equations allow the calculation of the manipulated variables $M(s)$ (where s is the Laplace operator), for the control modes P, PI, and PID, as a function of the process error $E(s)$:

Control Mode: P $\quad M(s) = K_c E(s) \qquad$ **2.38(11)**

Control Modes: PI $\quad M(s) = K_c(1 + 1/T_i s)E(s) \qquad$ **2.38(12)**

Control Modes: PID $\quad M(s) = K_c(1 + 1/T_i s + T_d S)E(s) \qquad$ **2.38(13)**

In the following, a number of tables are provided that contain some traditional tuning criteria and the corresponding tuning equations for P, PI, and PID controllers, when they are tuned to correct for load disturbances in the process. Table 2.38d lists a variety of tuning criteria that have been discussed in earlier sections. Tables 2.38e, 2.38f, and 2.38g list the equations to be used in determining the proportional, integral, and derivative settings for the various criteria listed in Table 2.38d. Depending on which performance criterion is selected from Table 2.38d and on the number of control modes, the formulas can be selected manually or automatically (per-programming) from the next three tables.

TABLE 2.38d

Controller Tuning Criteria Selection Table[1]

$$\text{ISE — 1} = \int_0^\infty [c(t) \times c(\infty)]^2 \, dt$$

$$\text{ISE — 2} = \int_{\theta_0}^\infty [c(t) - c(\infty)]^2 \, dt$$

$$\text{ISE — 3} = \int_0^\infty \left[\frac{c(t) - c(\infty)}{c(\infty)} \right]^2 dt$$

$$\text{IAE — 1} = \int_0^\infty \left| c(t) - c(\infty) \right| dt$$

$$\text{IAE — 2} = \int_{\theta_0}^\infty \left| c(t) - c(\infty) \right| dt$$

$$\text{IAE — 3} = \int_0^\infty \left| \frac{c(t) - c(\infty)}{c(\infty)} \right| dt$$

$$\text{ITAE — 1} = \int_0^\infty \left| c(t) - c(\infty) \right| t \, dt$$

$$\text{ITAE — 2} = \int_{\theta_0}^\infty \left| c(t) - c(\infty) \right| t \, dt$$

$$\text{ITAE — 3} = \int_0^\infty \left| \frac{c(t) - c(\infty)}{c(t)} \right| t \, dt$$

TABLE 2.38e

Tuning Equations for Proportional-Only Controller Tuned for Correcting for a Load Disturbance

$$K_c = \frac{A}{K_u} \left(\frac{t_0}{\tau} \right)^B$$

Constants

Criterion	*A*	*B*
Ultimate	2.133	−0.877
¼ Decay	1.235	−0.924
ISE—1	1.411	−0.917
ISE—2	0.9889	−0.993
ISE—3	0.6659	−1.027
IAE—1	0.9023	−0.985
IAE—2	0.6191	−1.067
IAE—3	0.4373	−1.098
ITAE—1	0.4897	−1.085
ITAE—2	0.4420	−1.108
ITAE—3	0.3620	−1.119

TABLE 2.38f

Tuning Equations for PI Controller Tuned for Correcting for a Load Disturbance

$$K_c = \frac{A}{K_u} \left(\frac{t_0}{\tau} \right)^B$$

$$\frac{I}{T_i} = \frac{A}{\tau} \left(\frac{t_0}{\tau} \right)^B$$

Constants

Criterion	*Controller Mode*	*A*	*B*
ISE	Proportional	1.305	−0.960
	Reset	0.492	−0.739
IAE	Proportional	0.984	−0.986
	Reset	0.608	−0.707
ITAE	Proportional	0.859	−0.977
	Reset	0.674	−0.680

TABLE 2.38g

Tuning Equations for PID Controller Tuned for Correcting for a Load Disturbance

$$K_c = \frac{A}{K_u} \left(\frac{t_u}{\tau} \right)^u$$

$$\frac{I}{T_i} = \frac{A}{\tau} \left(\frac{t_0}{\tau} \right)^B$$

$$T_d = \tau A \left(\frac{t_0}{\tau} \right)^B$$

Constants

Criterion	*Controller Mode*	*A*	*B*
ISE	Proportional	1.495	−0.945
	Reset	1.101	−0.771
	Rate	0.560	1.006
IAE	Proportional	1.435	−0.921
	Reset	0.878	−0.749
	Rate	0.482	1.137
ITAE	Proportional	1.357	−0.947
	Reset	0.842	−0.738
	Rate	0.381	0.995

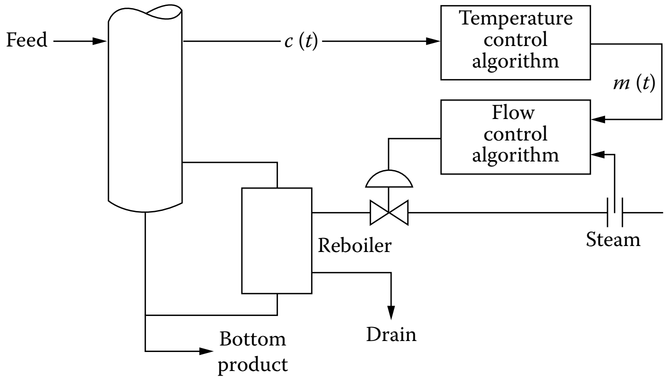


FIG. 2.38h
The tuning example is based on the temperature cascade loop shown in this figure.

Example

In this example the previously outlined tuning method will be applied to a pilot plant distillation unit under the direct digital control of a time-shared computer. Figure 2.38h illustrates the unit process on which the process operator in cooperation with a computer tunes the cascade loop, which controls the temperature of this unit.

Figure 2.38i shows the modified step response method of testing, where the upper two plots show how the operator applies a step in the open loop and obtains a response of the controlled variable, which allows for the determination of the ultimate sensitivity. This is then followed by the closed-loop portion of the test, which is shown in the lower two plots. Here the ultimate gain of the process is obtained.

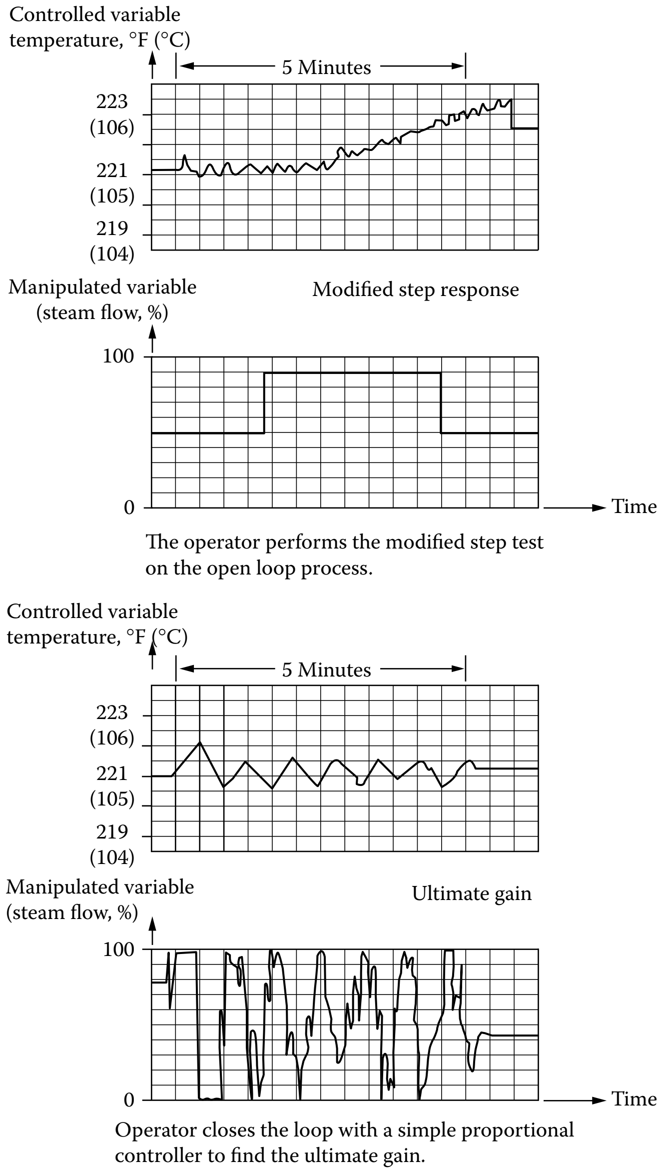
Once the data are collected, Equation 2.38(8) is used to calculate the process gain (K_p = 74%) This can than be verified by using Equation 2.38(4) to also calculate K_p = (time constant – T)(maximum rate of rise – R_r), which is considered to be adequate verification.

Tuning and Diagnostic Software

Some software packages include a variety of automatic means for obtaining better process models and improved tuning settings and also provide data for better maintenance.[2] The information below is taken from Section 5.6 of Volume 3 of this handbook, where the topic is treated in more detail.

Some of these software packages can perform a variety of analyses (auto-correlation, cross-correlation, power spectral density, statistical, linearity, etc.) They can also identify a variety of process models (integrator, dead time, first-order and higher orders, etc.); measure signal and valve noise; evaluate valve hysteresis, stiction, and nonlinearity; and evaluate the time and frequency response of the loop. In addition, they can evaluate and recommend control algorithm variations (error, reset, filtering, and sampling).

These advanced software packages can also provide performance analysis (IAE, ISE, response time, robustness, variability, statistics) and can suggest multivariable control strategies (cascade, feed forward, Smith predictor).



The operator performs the modified step test on the open loop process.

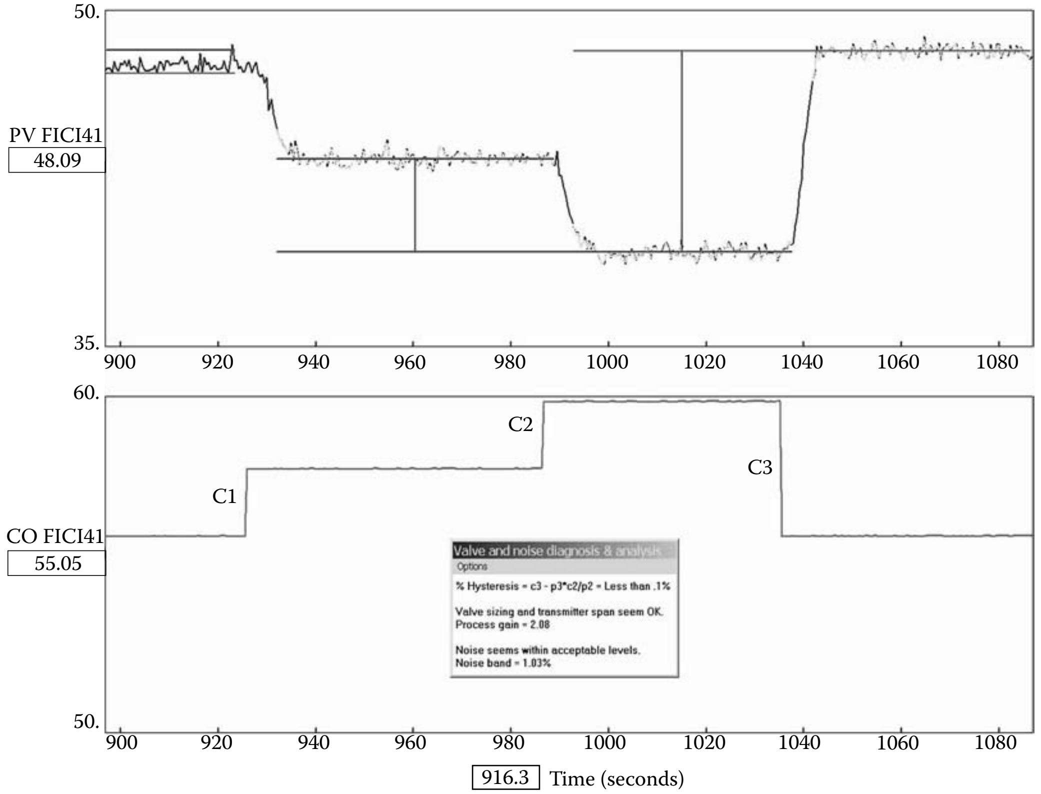
Operator closes the loop with a simple proportional controller to find the ultimate gain.

FIG. 2.38i
The modified test response (top two plots) is obtained in the open loops as a step response, while the ultimate gain is determined (bottom two plots) in the closed-loop mode.

Example

A flow loop in a refinery was analyzed and tuned using such a software package. Figure 2.38j illustrates the test for valve hysteresis, which was completed in the manual mode in less than 3 minutes.

Software tools were used to analyze the data collected. It was found that the gain of this flow process varied from 1.3 to 2.4, which gives a linearity (maximum gain over



% Hysteresis = c3 - p3*c2/p2 = Less than .1%

Valve sizing and transmitter span seem OK.
Process gain = 2.08

Noise seems within acceptable levels.
Noise band = 1.03%

FIG. 2.38j
Valve hysteresis test results, which were obtained in less than 3 minutes of manual testing.

minimum gain) of 1.8; the response time of the loop exceeded 1 minute; hysteresis was under 0.1%; and stiction (the sticking of the valve) was 1.1%. Based on the data collected, it was found that the process gain is 2.3, the dead time of the process is 3.6 seconds, and the time constant is 2.4 seconds.

As a consequence of this evaluation, the sticking valve was repaired, and then the loop was re-tuned.[3] Figure 2.38k provides a tabulation of the performance statistics of this flow loop before (left) and after (right) of the tuning. Figure 2.38l provides a visual indication of the valve sticking (left) before maintenance was performed

PV FIC141 Statistics		
Sample (raw)	0.33	
Num of Points	1819	
Time Min	0	
Time Max	801	
PV Min	46.47	
PV Max	49.56	
Range	3.088	
Mean (μ)	47.99	
Standard deviation (δ)	0.544	
μ ± δ	**47.446–48.534**	
μ ± 2δ	**46.902–49.078**	= 2.176
Variance	0.2959	
Variability	2.27%	
Variability	39	

PV FIC141 Statistics		
Sample (raw)	0.33	
Num of Points	1817	
Time Min	0	
Time Max	800	
PV Min	46.82	
PV Max	48.97	
Range	2.156	
Mean (μ)	47.65	
Standard deviation (δ)	0.258	
μ ± δ	47.392–47.908	
μ ± 2δ	47.134–48.166	= 1.032
Variance	0.06656	
Variability	1.08%	
Variability	0.475	

CO FIC141 Statistics		
Travel	8.13	878/day
Reversals	14	1510/day

CO FIC141 Statistics		
Travel	0	0/day
Reversals	1	108/day

FIG. 2.38k
Performance data of the flow loop before (left) and after (right) the correcting of the sticking valve and re-tuning of the controller.

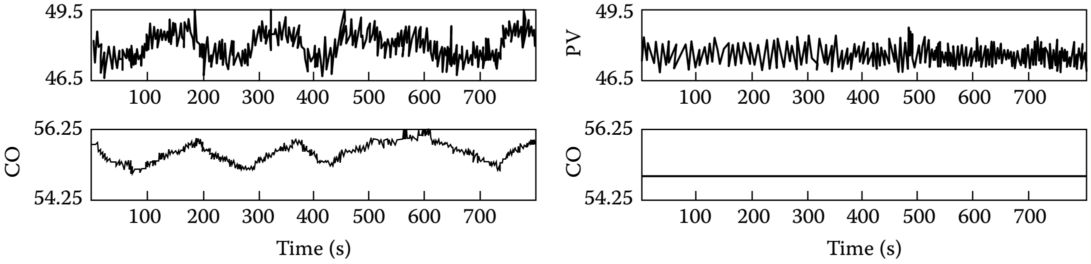


FIG. 2.38l
Visual description of the performance of the flow loop before (left) and after (right) the correcting of the sticking valve and re-tuning of the controller.

> on it. As can be seen, the controller output (*CO*) had to change over 1% (from about 55 to 56%) before the valve would respond. The improvement in loop performance can be seen by comparing the fluctuation of flow between 46.5 and 49.5% (left) to about half as much (right), achieved by the fixing of the valve and the re-tuning of the loop.

CONCLUSIONS

With the various techniques described in this section, the operator can check the controller tuning and the process model accuracy periodically or when operating conditions change. Up-to-date modeling and tuning should greatly improve process performance and safety because most of these techniques maintain process safety by allowing the operators to remain in the loop during the critical testing phase.

In addition, cost reductions are also realized when reducing the amount of trial-and-error tuning because the operator need not wait to bring the process to a final value in the step response test. This saves time when dealing with slow processes. Test signal size is not critical, reducing trial-and-error selection.

Time-shared systems with DCS or supervisory control have greater scaling and sensitivity capabilities than conventional industrial instruments. Also, it can perform unit conversion and can be easily programmed to abort the tests automatically and set off alarms if necessary.

In general, the computer can be a very valuable tool for control system analysis and design in both online and offline testing. Offline techniques range from full-blown computer simulations of process and the associated control system to simple linear first-order plus dead time type models generated from step testing. Online techniques can involve state space optimal control (Section 2.33) and Kálmán filtering techniques,[4] or they can be as simple as some of the automated tuning procedures that were discussed in this section.

The computer has become a very important tool, and software is widely available for the process control and instrument engineer to use to implement a variety of offline techniques of testing. Online techniques are also becoming more popular as the process control computer is becoming part of most control system installations.

Testing, modeling, and tuning improve process performance and result in cost reductions by reducing the amount of trial-and-error tuning. Time-shared systems with DCS, PLC, or supervisory control offer a convenient vehicle to assist operators in tuning, while data acquisition can be done automatically. In minutes a loop can be analyzed and re-tuned.

References

1. Lopez, A. M., Miller, J. A., Smith, C. I., and Murrill, P. W., *Instrument Technology,* Research Triangle Park, NC: ISA, November 1967, p. 57.
2. Lipták, B. G. (Ed.), *Instrument Engineers' Handbook: Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002 (Section 5.6).
3. Ruel, M., "Stiction: The Hidden Menace," *Control Magazine,* Vol. 13, No 11, November 2000, pp. 69–76.
4. Blevins, T. L., Brown, M. McMillan, G., and Wojsznis, W. K., *Advanced Control Unleashed,* Research Triangle Park, NC: ISA, 2003.

Bibliography

Aplevich, J. D., *The Essentials of Linear State-Space Systems,* New York: John Wiley & Sons, 2000.

Astrom, K. J., and Wittenmark, B., *Computer-Controlled Systems: Theory and Design,* Englewood Cliffs, NJ: Prentice Hall, 1997.

Badgwell, T. A., and Qin, S. J., "Review of Nonlinear Model Predictive Control Applications," in B. Kouvaritakis and M. Cannon, Eds., *Nonlinear Predictive Control: Theory and Practice,* IEE Publishing, 2001, pp. 3–32.

Bertsekas, D. P., *Dynamic Programming and Optimal Control,* Vols. I and II, Athena Scientific, 2001.

Blevins, T. L., McMillan, G. K., Wojsznis, W. K., and Brown, M. W., *Advanced Control Unleashed,* Research Triangle Park, NC: ISA, 2002.

Blevins, T. L., Brown, M., McMillan, G., and Wojsznis, W. K., *Advanced Control Unleashed,* Research Triangle Park, NC: ISA, 2003.

Bryson, A. E., *Dynamic Optimization,* Reading, MA: Addison-Wesley, 1999.

Chan, P. C., *Process Control, A First Course with Matlab,* New York: Cambridge University Press, 2002, pp. 21–24.

Cheng, G. S., and Huo, L. Q., "Leading Edge Plant Control System — Model-Free Adaptive Control Puts Ling-Yuan Iron and Steel in League of Its Own," *InTech,* October 2003.

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: Instrument Society of America, 1990.

Dahlin, E. B., "Designing and Tuning Digital Controllers," *Instruments and Control Systems,* June 1968.

Goodwin, G. C., Graebe, S. F., and Salgado, M. E., *Control System Design,* Englewood Cliffs, NJ: Prentice Hall, 2001.

Haykin, S. S., *Neural Networks: A Comprehensive Foundation,* 2nd ed., Englewood Cliffs, NJ: Prentice Hall, 1998.

Higham, J. D., "Single-Term Control of First- and Second-Order Processes with Dead Time," *Control,* February 1968.

Howell, J., and Rost, D., "PC Software Tunes Plant Startup," *InTech,* June 1988.

Kraus, T. W., and Myron, T. J., "Self-Tuning PID Controller Uses Pattern Recognition Approach," *Control Engineering,* June 1984.

Lelic, M. A., and Zarrop, M. B., "Generalized Pole-Placement Self-Tuning Controller," *International Journal of Control,* 46, 1987, pp. 569–607.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill, 1990.

Lyashko, S. I., *Generalized Optimal Control of Linear Systems with Distributed Parameters,* Dordrecht: Kluwer, 2002.

Maciejowski, J. M., *Predictive Control with Constraints,* London: Pearson Education Limited, 2002.

Mansy, M. M., McMillan, G, K., and Sowell, M. S., "Step into the Virtual Plant," *Chemical Engineering Progress,* February 2002, pp. 56–61.

Marlin, T. E., *Process Control, Designing Process and Control Systems for Dynamic Performance*, 2nd ed., New York: McGraw-Hill Higher Education, 2000.

McMillan, G. K., *Tuning and Control Loop Performance,* 3rd ed., Research Triangle Park, NC: ISA, 1994

McMillan, G. K., *Process Industrial Instruments and Control Handbook,* New York: McGraw-Hill, 1999.

McMillan, G., "A New Era for Model Predictive Control," *Control,* Vol. 13, No. 5, pp. 45–50, 2000.

McMillan, G. K., and Cameron, R. A., *Models Unleashed — Virtual Plant and Model Predictive Control Applications,* Research Triangle Park, NC: ISA, 2003.

Naidu, D. S., *Optimal Control Systems,* Boca Raton, FL: CRC Press, 2003.

Nelles, O., *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models,* London: Springer-Verlag, 2000.

Omatu, S., Khalid, M., and Yusof, R., *Neuro-Control and its Applications,* London: Springer-Verlag, 1996.

Ou, J., and Rhinehart, R. R., "Grouped Neural Network Model Predictive Control," *Control Engineering Practice,* Vol. 11, No. 7, 2003, pp. 723–732.

Qin, S. J., and Badgwell, T. A., "An Overview of Industrial Model Predictive Control Technology," in Chemical Process Control – V, pages 232–256, Fifth International Conference on Chemical Process Control, AIChE and CACHE, 1997.

Rossiter, J. A., *Model-Based Predictive Control,* Boca Raton, FL: CRC Press, 2003.

Rothstein, M. B., "Tuning a Nearly Optimal Control by Digital Simulation," *Instrumentation Technology,* February 1972.

Rovira, A. A., Murrill, P. W., and Smith, C. L., "Tuning Controllers for Set-Point Changes," *Instruments and Control Systems,* December 1969.

Ruel, M., "Loop Optimization: How to Tune a Loop," *Control Magazine,* Vol. 12, No. 5, May 1999, pp. 83–86.

Sastry, S., and Bodson, M., *Adaptive Control: Stability, Convergence and Robustness,* Englewood Cliffs, NJ: Prentice Hall, 1989.

Seborg, D. E., Edgar, T. F., and Shah, S. L., "Adaptive Control Strategies for Process Control," *Chemical Engineering Journal,* 32, 1986, pp. 881–913.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* 2nd ed., New York: John Wiley & Sons, 2003.

Shinskey, F. G., *Feedback Controllers for the Process Industries,* New York: McGraw Hill, 1994.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw Hill, 1996.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

Smith, O. J. M., "Close Control of Loops with Dead Time," *Chemical Engineering Progress,* May 1957.

Sood, M., "Tuning Proportional-Integral Controllers for Random Load Changes," M.S. Thesis, Dept. of Chemical Engineering, University of Mississippi, 1972.

Sood, M., and Huddleston, H. T., "Optimal Control Settings for Random Disturbances," *Instrumentation of Technology,* March 1973.

Suykens, J. A. K., Vandewalle, J. P. L., and De Moor, B. L. R., *Artificial Neural Networks for Modelling and Control of Non-Linear Systems,* New York: Kluwer Academic, 1996.

Thayer, D., "Tuning Direct Digital Control," *Instruments and Control Systems,* October 1967.

VanDoren, V., *Techniques of Adaptive Control,* Burlington, MA: Butterworth-Heinemann, 2003.

Wellstead, P. E., and Zarrop, M. B., *Self-Tuning Systems,* New York: Wiley, 1991.

Wojsznis, W., Gudaz, J., Mehta, A., and Blevins, T., "Practical Approach to Tuning MPC," *ISA Transactions,* Vol. 42, No. 1, January 2003, pp. 149–162.

Wood, R. K., "Computerized Design and Analysis of Control Loops," *Instrumentation Technology,* 1971.

Transmitters and Local Controllers

3

3.1
CONTROLLERS—PNEUMATIC 460

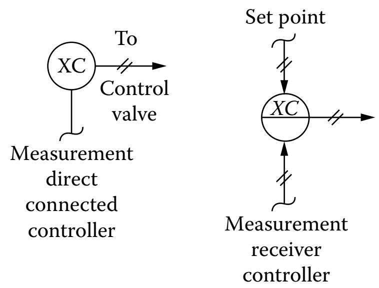
3.2
CONTROLLERS—ELECTRONIC 478

3.3
CONVERTERS AND DAMPENERS 488

3.4
RELAYS FOR COMPUTING AND PROGRAMMERS 493

3.1 Controllers—Pneumatic

C. L. MAMZIC (1970, 1985, 1995) **P. M. B. SILVA GIRÃO** (2005)



Flow sheet symbols

Types:	Receiver controllers: Indicating, recording, miniature, high-density miniature, and large case
Direct-connected controllers:	Blind, indicating, or recording, in large or medium case for field or panel mounting.
Application:	Receiver controllers: Control of any variable that can be measured and translated into an air pressure by a pneumatic transmitter; includes automatic and manual control and set point adjustment.
Direct-connected controllers:	Have own measuring element in contact with the process; measuring elements available include pressure, differential pressure, temperature, level, pH, thermocouple, radiation pyrometer, and humidity.
Typical Front Panel Size:	Miniature: 6 in. × 6 in. (150 mm × 150 mm)
High-density miniature:	3 in. × 6 in. (75 mm × 150 mm)
Large case:	15 in. × 20 in. (375 mm × 500 mm)
Minimum Response Level:	Less than 0.01% of full scale
Input Range (measurement):	3 to 15 PSIG (0.2 to 1.0 bar) or direct-connected
Output Range (signal to valve):	3 to 15 PSIG (0.2 to 1.0 bar)
Repeatability:	± 0.2% of span
Inaccuracy:	± 0.25 to 2% of span
Displays:	Set point, process variable (measurement), controlled variable (output to control valve), deviation, and balance
Maximum Frequency Response:	Flat to 30 Hz
Maximum Zero Frequency Gain:	750
Control Modes:	Manual, on-off, proportional (P), integral (I-reset), derivative (D-rate), floating, differential gap, follow-up, cascade
Costs:	Direct-connected, indicating, field-mounted controller costs range from $500 to $2000; miniature indicating controller costs $1000 to $2500; miniature high-density indicating controller indicating $1500 to $2500; miniature recording controller costs

$2000 to $3000; large-case indicating controller costs $1500 to $2500; and large-case recording controller costs $2000 to $3000

Partial List of Suppliers:	ABB Group (www.abb.com)
	Ametek (www.ametek.com)
	Barton (www.barton-instruments.com/index2.php)
	Bristol Babcock (www.bristolbabcock.com)
	Emerson Process Management (www.emersonprocess.com)
	Foxboro Co. (www.foxboro.com/m&i/specifications/controllers/)
	Honeywell Automation and Control (hbctradeline.honeywell.com/Catalog/Pages/default.asp)
	Leslie Controls, Inc (www.lesliecontrols.com)
	Powers Process Controls (www.powerscontrols.com)
	Samson AG (www.samson.de/pdf_en/_ek16_re.htm)
	Spence Engineering Company, Inc. (www.spenceengineering.com/Handbook/index.htm)
	Trautomation (www.trautomation.com/automation/categorie.nsf/)

HISTORY AND DEVELOPMENT

Pneumatic controllers were first introduced at the turn of the twentieth century. They logically followed the development of diaphragm-actuated valves in the 1890s. Early types were all direct connected, local mounting, indicating, or blind types. Large-case indicating and circular chart recording controllers appeared around 1915. All early models incorporated two-position, on/off action or proportional action. It was not until 1929 that reset action was introduced. Rate action followed in 1935.

Until the late 1930s, all controllers were direct connected and therefore had to be located close to the process. Pneumatic transmitters were not introduced until the later 1930s. To make them compatible, the large-case pressure recording and indicating controllers were easily converted into receiver controllers. This made remote mounting practicable, and centralized control rooms became a reality. Because of the inherent advantages, the combination of pneumatic transmitters and receiver controllers quickly became popular. Since the recording and indicating receiver controllers were quite large, control rooms and panel boards were likewise spacious. Additionally, all control boards had a monotonous look and usually came in one color—black.

A revolution in design occurred in 1948 with the introduction of miniature instruments. Here the concept of the controller evolved into a combination of a small, approximately 6 in. × 6 in. (150 mm × 150 mm), panel front indicating and recording control station and a blind receiver controller. The station permitted the operator to monitor the measured variable (process variable), set point, and valve output; it allowed the operator to switch between, and operate in, either the automatic or manual control modes.

Miniature controllers ushered in the era of the graphic panel, in which the instruments are inserted into graphic symbols representing the attendant process apparatus. Control rooms became more compact, control boards more meaningful and colorful, and because operators quickly developed a "feel" for the process, training time was considerably reduced.

Nevertheless, graphic panels were also wasteful of space and presented major modification problems each time the process was changed. This led to the evolution of the semi-graphic panel, in which a graphic symbol diagram of the process appeared above the miniature instruments mounted in neatly spaced rows and columns.

In 1965, miniature, high-density mounting style stations appeared. The new lines brought with them the most advanced ideas in displays, operating safety and simplicity, packaging, installation simplicity, and servicing facility. Along with some of the standard miniature controllers, they offered computer compatibility along with some unique control capabilities that had previously been impractical.

The early 1980s saw some important new entries in the pneumatic controller market. These included pneumatic controllers using RTD and thermocouple type sensors and pneumatic controllers with microprocessor-based serial communication modules for tie-in to distributed control systems.

As first analog (Section 3.2), and later digital electronic (Section 4.4) and DCS-based software controllers (Chapter 4) became available, they gradually took over the controller marketplace, but pneumatic units are still used in many existing plants and in locations where intrinsic safety is essential.

OPERATING PRINCIPLES

There are both force and motion balance pneumatic controllers. A receiver-type pneumatic controller based on the force balance principle (moment-balance) is shown schematically in Figure 3.1a. A process transmitter (lower left) senses the measured variable (i.e., process variable, e.g., pressure, temperature, flow) and transmits a proportional air pressure to the measured variable (MV) bellows of the controller.

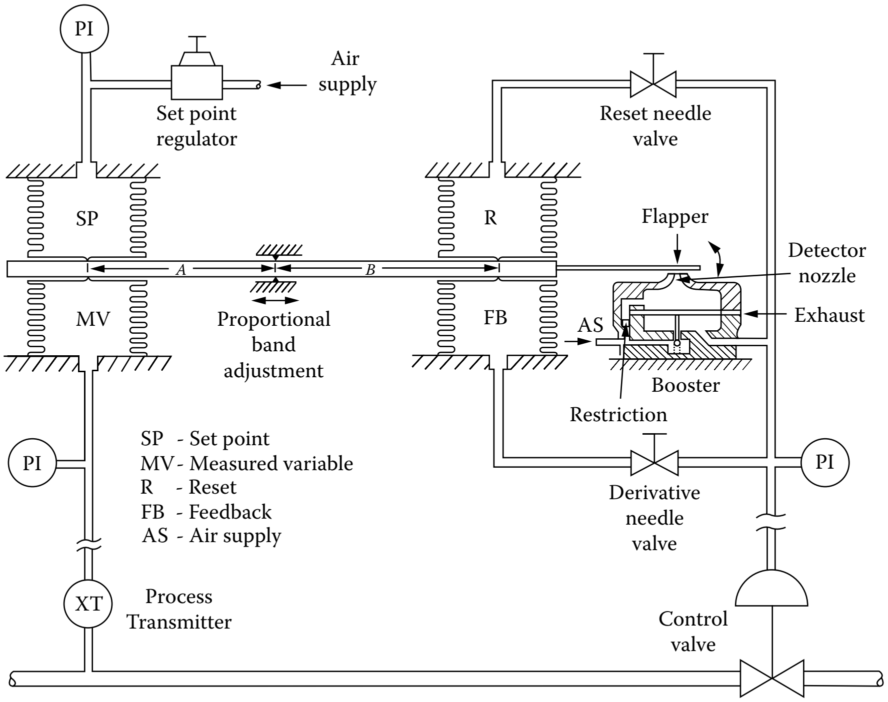


FIG. 3.1a

Moment-balance controller.

The controller compares the measured variable against the set point (SP) and sends a corrective air signal (error or deviation) to manipulate the control valve, thereby completing the feedback control loop.

Figure 3.1a depicts a so-called direct acting controller; the output of the controller increases when the measured variable increases. The controller action is selected as a function of the failure position of the control valve and the relationship between the controlled and manipulated variables. To illustrate this relation with an example, visualize a process cooler. The purpose of the control loop on that cooler is to maintain the temperature at the outlet of the cooler. If the cooling water valve fails closed, a *rise* in the detected temperature will require further opening of the coolant valve, which means that the air signal to the valve should *drop* and therefore the controller has to be reverse acting.

In Figure 3.1a, the controller consists of two sets of opposed bellows of equal area, acting at opposite ends of a force beam that rotates about a movable pivot. Extending from the right end of the beam is a flapper that baffles the detector nozzle of the booster relay.

Booster Circuit

Supply air connects to the pilot valve of the booster and flows through a fixed restriction into the top housing and out the detector nozzle. The flapper is effective in changing the backpressure on the nozzle as long as the clearance is within one-fourth of the nozzle diameter. The restriction size is selected on the basis that the continuous air consumption will be reasonable and that it will be large enough not to clog with typical instrument air. The nozzle, on the other hand, must be large enough that when the flapper has a clearance of one-fourth nozzle diameter, nozzle backpressure drops practically to atmospheric. It must not be so large, however, that the seating of the flapper becomes too critical.

A typical size of restriction is 0.012 in. ID (0.3 mm) while the nozzle would be 0.050 in. ID (1.3 mm). The nozzle backpressure is a function of flapper position. (For a more detailed discussion of circuits, refer to Section 1.4, “Electronic versus Pneumatic Instruments.”)

The exhaust diaphragm senses nozzle backpressure and acts on the pilot valve. If the backpressure increases, it pushes down on the valve, opening the supply port to build up the underside pressure on the diaphragm until it balances the nozzle backpressure. If the backpressure decreases, the diaphragm assembly moves upward, allowing the valve to close off the supply seat while opening an exhaust seat in the center of the diaphragm. This allows the underside pressure to exhaust through the center mesh material of the diaphragm assembly until the pressures balance.

Proportional Response

Since a pressure range of 3 to 15 PSIG (0.2 to 1.0 bar) is an almost universal standard for representing 0 to 100% of the range of a measured variable, a set point, and the output of receiver controllers, this description will assume the operation to be in this range.

To understand the proportional response, first assume that the derivative needle valve is wide open and that the reset needle valve is closed with 9 PSIG (0.6 bar) mid-scale pressure, trapped in the reset bellows R (Figure 3.1a). If the set point is adjusted to 9 PSIG (0.6 bar), when the measured variable equals set point, the flapper will automatically be positioned so that the booster output, acting on the feedback bellows (FB), will be equal to the reset pressure, namely 9 PSIG (0.6 bar). The reason for this is that the force beam will only come to equilibrium when all of the moments cause a rotation of the beam with attendant repositioning of the flapper and change in feedback pressure until the moment balance is restored.

If the pivot is positioned centrally, where moment arm A equals B, then for every 1 PSI (0.067 bar) difference between set point and measured variable there will be a 1 PSI (0.067 bar) difference between reset and controller output, or feedback. This represents a 100% proportional band setting, or a gain of one. Percent proportional band is defined as the input change divided by the output change times 100. Proportional bands are typically adjustable from 5 to 500%. Gain equals the ratio of output change to input change. For a description of both the proportional band and of controller action, refer to Figure 3.1b.

$$\text{PB}\,\% = \frac{\text{change in input}}{\text{change in output}} \times 100 \qquad \textbf{3.1(1)}$$

$$\text{gain} = \frac{\text{output change}}{\text{input change}} = 100/\text{PB} \qquad \textbf{3.1(2)}$$

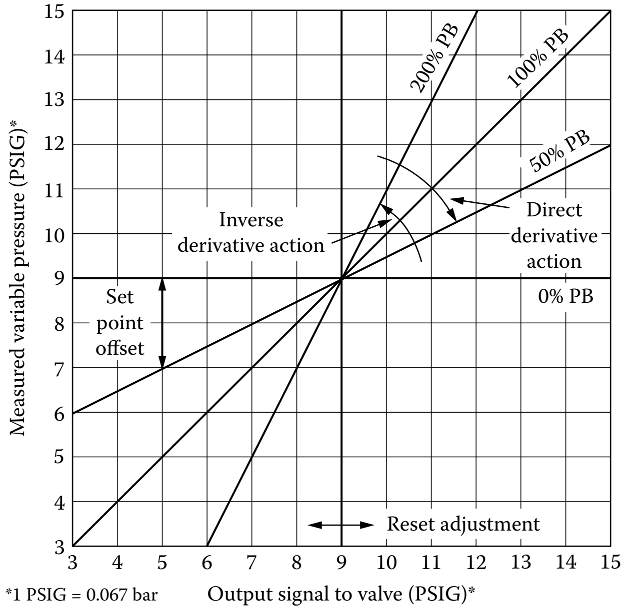


FIG. 3.1b
Graphic representation of control functions.

If the pivot in Figure 3.1a is shifted to the right, to the point where moment arm A is four times greater than moment arm B, then every 1-PSI (0.067-bar) change in the measured (also called controlled) variable results in a 4-PSI (0.27-bar) change in the output (also called manipulated) variable. This gives a proportional band of 25%, or a gain of four.

If the pivot is moved to the left, the ratio is reversed and the proportional band would be 400%, which corresponds to a gain of 0.25. If the pivot could be moved to the right to coincide with the center of the R and FB bellows, the most sensitive setting would be achieved, i.e., approaching 0% band or infinite gain. In this state, the slightest difference between MV and SP would rotate the flapper to change the output to 0 PSI (0 bar) or full supply pressure, depending upon the action of the error (on/off control).

Reset or Integral Response

If it were practical to use a 1 or 2% proportional band on all processes, proportional action alone would be sufficient for most processes. However, most process loops become unstable at such or even wider bands. Flow control loops, for example, usually require more than a 200% band for stability. Assuming that the process cannot tolerate a band narrower than 50%, in that case (Figure 3.1b), the controller can maintain the measured variable on set point only when valve pressure is 9 PSIG (0.6 bar).

If the valve pressure has to be 5 PSIG (0.3 bar), it can only occur when the measured variable deviates from set point by 2 PSI (0.14 bar) (a 16.7% of full scale error). In most cases this amount of error, more correctly termed offset, is intolerable. Nevertheless, in real control systems the valve pressure must change as the load changes.

If the load is such as to require a 5-PSIG (35-kPa) valve pressure, one way of eliminating the offset would be to *manually* change the reset pressure R to 5 PSIG. The output or valve pressure would then be 5 PSIG when the error is zero. Initially, when proportional-only controllers were used, such manual reset was used, but that is totally unacceptable today.

It is a simple matter to make this reset action automatic. It only requires that the reset bellows be able to communicate with the controller output pressure through some adjustable restriction such as a needle valve. The reset action must be tuned to the process in such a way as to allow the process sufficient time to respond. Too fast a reset speed, in effect, makes the controller "impatient" and results in instability. Too slow a speed results in stable operation, but the offset is permitted to persist for a longer period than necessary.

To describe the automatic reset action, assume that set point is at 9 PSIG (0.6 bar) and the reset is trapped at 9 PSIG, while valve pressure feedback, because of load change, must be at 5 PSIG (0.3 bar). According to Figure 3.1b, with the proportional band at 50%, the measured variable will be controlled at 7 PSIG (0.46 bar).

If the reset needle valve is then opened slightly, the pressure in the reset bellows will gradually decrease. As it drops from 9 to 7 PSIG (0.6 to 0.46 bar), the measured variable pressure will rise from 7 to 8 PSIG (0.46 to 0.5 bar). The reset pressure will continue to drop until it exactly equals output, i.e., 5 PSIG (0.3 bar). At this point, the measured variable will exactly equal set point, 9 PSIG (0.6 bar). With this configuration, regardless of where the valve pressure must be, the controller will ultimately provide the correct valve pressure with no offset between set point and measured variable.

Reset action can also be understood by looking at it from the controller design perspective. We can see that the controller cannot come to equilibrium as long as there is any difference in pressure between reset and feedback. This is because if there is, the open communication through the reset needle valve will cause the reset to continue to change, which in turn directly reinforces the feedback pressure.

Reset and feedback, in turn, cannot be equal unless the set point and the measured variable are equal to each other. This fact alone ensures that the controller will maintain corrective action until it makes the measured variable exactly equal to set point regardless of where feedback must be. Referring to Figure 3.1b, reset in effect shifts the proportional band lines along a horizontal axis at set-point level. It makes the middle of the band coincide with the required valve pressure.

Reset time is the time required for the reset action to produce the same change in output that the proportional action would if the error remained constant (the available integral time settings range from 0.01 to 60 minutes per repeat). For example, if the response to an error of a plain proportional controller is a 1-PSI (0.067-bar) change in its output and the error is sustained, the integral mode with a 1 minute/repeat reset setting will require 1 minute to eliminate that offset. The integral setting can have the units of either time/repeat or repeats/unit of time.

Derivative Response

If a needle valve is inserted between the booster output and the feedback bellows as in Figure 3.1a, it delays the rebalancing action of the feedback bellows and causes the controller to give an exaggerated response to changes in the measured variable. The degree of exaggeration is in proportion to the speed or rate at which the measured variable is changing. (The term derivative action refers to the slope or the rate of change.)

Derivative or rate action is particularly effective in the control of slow processes, such as in temperature control loops. It compensates for lag or inertia. For a sudden change of even small magnitude, it provides an extra "kick" to the control valve. This is because the derivative mode assumes that upsets will continue at the rate they are occurring and corrects right away for the error that would evolve one derivative time later. Conversely, when the error is dropping as the controlled process variable is returning to set point, the rate action anticipates that the inertia of the process will carry it below the set point and begins cutting back the valve response accordingly.

The method shown in Figure 3.1a for adding derivative action to the has some serious limitations. If the derivative needle valve is located as shown, the derivative action will interact with both the proportional and the integral responses and, in fact, will follow the proportional response. This design therefore is useless in preventing overshoot on startup or when large upsets occur, and this derivative also interacts with set point changes. Therefore an independent derivative unit is needed, as shown in Figure 3.1e and described later.

Derivative action can be viewed as if it temporarily changed the proportional band and therefore temporarily changed the slope of the lines in Figure 3.1b. If the error is rising, the change is clockwise; if it is dropping, the change is counterclockwise; if the error is constant, no change occurs. Derivative time is the amount of time (usually in minutes) by which the rate action anticipates into the future and leads the feedback pressure during a steady ramp input change. (Derivative settings can usually be adjusted from 0.01 to 60 minutes.)

MINIATURE CONTROLLER DESIGNS

Two designs of force-balance receiver controllers are shown in Figures 3.1c and 3.1d. The controller in Figure 3.1c closely resembles that of Figure 3.1a, except that the bellows all act from one side against a pivoted "wobble plate." The wobble plate acts as the nozzle's baffle. Rotating the pivot axis

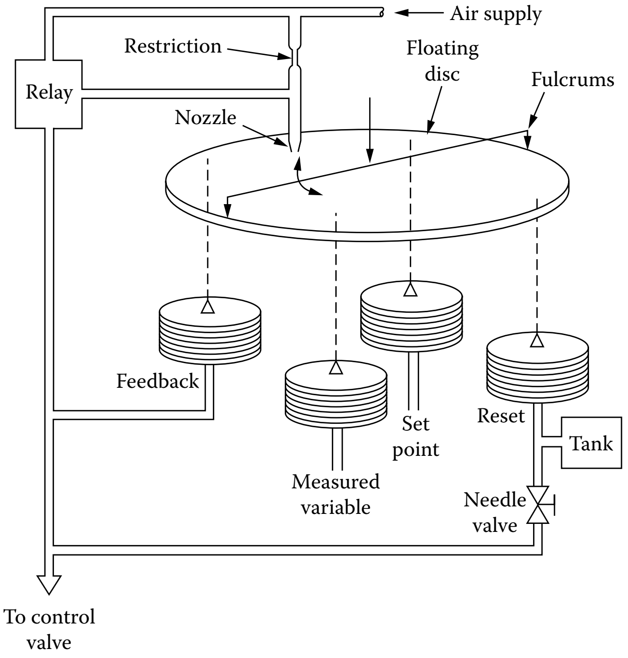


FIG. 3.1c
Typical moment-balance controller.

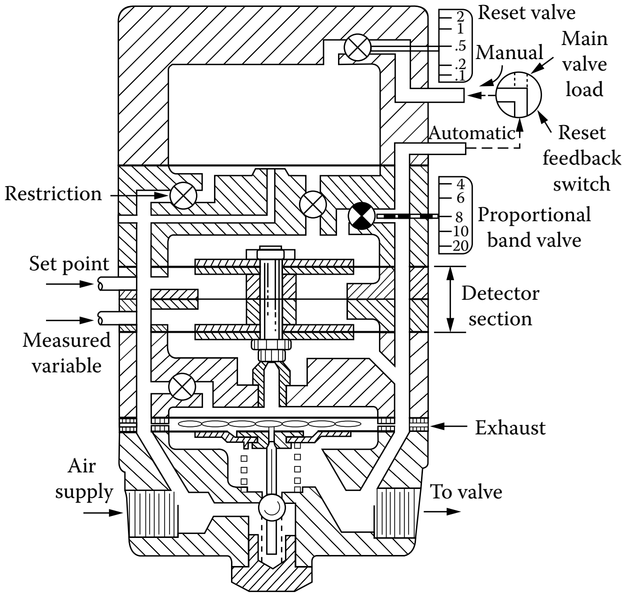


FIG. 3.1d
Force-balance controller.

changes the proportional band. When the pivot axis coincides with the reset and feedback bellows, 0% proportional band results; when it coincides with the measured variable and set point bellows, infinite band results. When the axis bisects the two bellows' axis, 100% proportional band results. Otherwise, the operation is the same as described for Figure 3.1a.

The controller in Figure 3.1d is constructed of machined aluminum rings, which are separated by rubber diaphragms with bolts holding the assembly together. The lower portion of the controller forms the booster section. This is quite similar to the booster in Figure 3.1a.

The detector section consists of three diaphragms. The upper and lower diaphragms have equal areas, while the center diaphragm has half the effective area of the other two. Reset pressure acts in the top chamber. Assume this pressure is at mid-scale and that the reset needle valve is closed. The reset pressure acts on the top diaphragm, which is part of a 1:1 repeating relay. Supply air passes through a restriction and out the exhaust nozzle. The diaphragm baffles the nozzle to make the backpressure equal to the reset pressure.

Assume further that the proportional band needle valve is closed. The pressure then acting on top of the detector section will be the reset pressure as reproduced by the 1:1 relay since the two chambers are connected via a restriction. If the measured variable signal equals the set point, the detector diaphragm assembly, with its integral nozzle seat, will baffle off the nozzle so as to make the controller output, or valve pressure, equal to the reset pressure, thus balancing all of the forces acting on the detector.

If the measured variable is then increased by 1 PSI (0.067 bar), the increase in pressure acts downward on the lower diaphragm as well as upward on the center diaphragm. The net effect is the same as having the pressure act downward on a diaphragm of half the area of the lower diaphragm. Since the output pressure acts upward on the full area of the lower diaphragm, it needs to increase only 1/2 PSI (0.03 bar) to bring the forces to balance. Since a 1-PSI (0.067-bar) change in input resulted in a 1/2-PSI change in output, the proportional band is said to be at 200% or the gain is said to be 0.5.

If the proportional band needle valve were wide open, so that it provided negligible resistance, then if the measured variable pressure increased ever so slightly above set point, the full effect of the resultant change in output would be felt on top of the detector stack. This would cause the output to increase further, which in turn would feed upon itself, and the action would continue to regenerate until the output reached its maximum limit. Therefore, with the proportional band needle valve wide open, the narrowest proportional band is obtained.

If the proportional band needle valve is set to where its resistance equals that of the restriction separating the reproducing relay from the top of the detector section, then the following action results. If the measured variable deviates from set point by 1 PSI (0.067 bar) in an increasing direction, instantly the output will rise 1/2 PSI (0.03 bar) because of the construction of the detector section. The difference in pressure between the controller output and the reproducing relay will cause a flow through the reset needle valve and the intermediate restriction.

Since the resistance of the two is equal, the pressure drop will divide equally, causing a 1/4-PSI (0.017-bar) increase on the top of the detector section. This 1/4-PSI increase directly causes the output to increase 1/4 PSI, which further causes the pressure on top of the detector to increase by 1/8 PSI (0.008 bar). The action continues until equilibrium is obtained, with the output having changed a total of 1 PSI (0.067 bar) and with the pressure on top of the detector section having increased 1/2 PSI (0.03 bar). Since a 1-PSI change in variable resulted in a 1-PSI change in output, this needle valve opening provides a 100% proportional band.

The reset action in this controller is similar to that in Figure 3.1a in that any change in reset pressure propagates down through the unit to directly affect the output. The controller will not come to equilibrium until all forces are balanced, i.e., the measured variable will have to equal set point and the reset pressure will have to equal controller output.

Derivative Relay

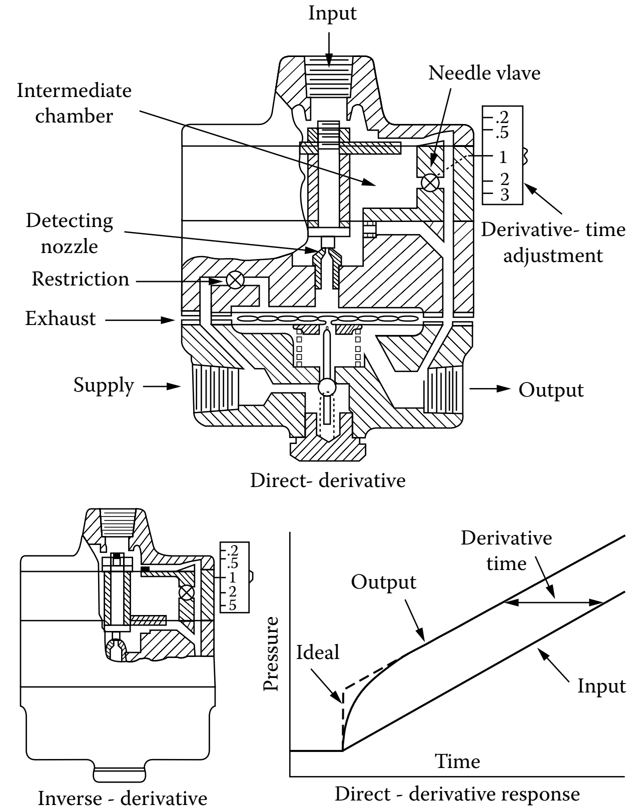
It was noted earlier that the type of derivative circuit used in Figure 3.1a interacted with the effects of the other control modes. Derivative action is more effective if it is noninteracting and if it can be applied *ahead* of the contributions of the proportional and reset actions of the controller. Such a derivative unit can be built into the controller, or it can be a separate relay as shown in Figure 3.1e. This relay employs diaphragms, but the design can also use bellows.

In this design the signal from some process transmitter is connected to the input port at the top. Because of the difference between input and output diaphragm areas, the result of the force balance design is that a step change in input pressure will produce an amplified step change in output pressure.

In a steady-state condition (with no change in input), the output pressure acts on both sides of the output diaphragm — and output pressure rebalances the input pressure directly. The output pressure is connected to the intermediate chamber through the needle valve. There is therefore a lag between a change in output and a change in intermediate pressure.

If the input changes with a continuous ramp, the intermediate pressure will lag the output by a constant amount, proportional to the rate of change in output. Thus, the intermediate pressure will partially rebalance the input, reducing the effective gain. The result is that the output will continuously lead the input by a definite amount, which is proportional to the rate of change in input. The time by which the output leads the input is the "derivative time." The graduated needle valve is used to set the derivative time.

An inverse derivative relay is also shown in Figure 3.1e. It works in the opposite manner and thereby attenuates high-frequency signals. It can therefore serve as a noise filter and stabilizing relay on "noisy" processes.



Direct- derivative

Inverse - derivative

Direct - derivative response

FIG. 3.1e
Direct and inverse derivative relays.

Miniature Control Stations

Miniature control stations with a panel face of nominally 6 in. × 6 in. (150 mm × 150 mm) and inserted into individual cutouts having approximately 10 in. (250 mm) center-to-center distances are one of the common types found on central control room panels in the various process industries. Figure 3.1f shows a typical cross section of some of the types of units available.

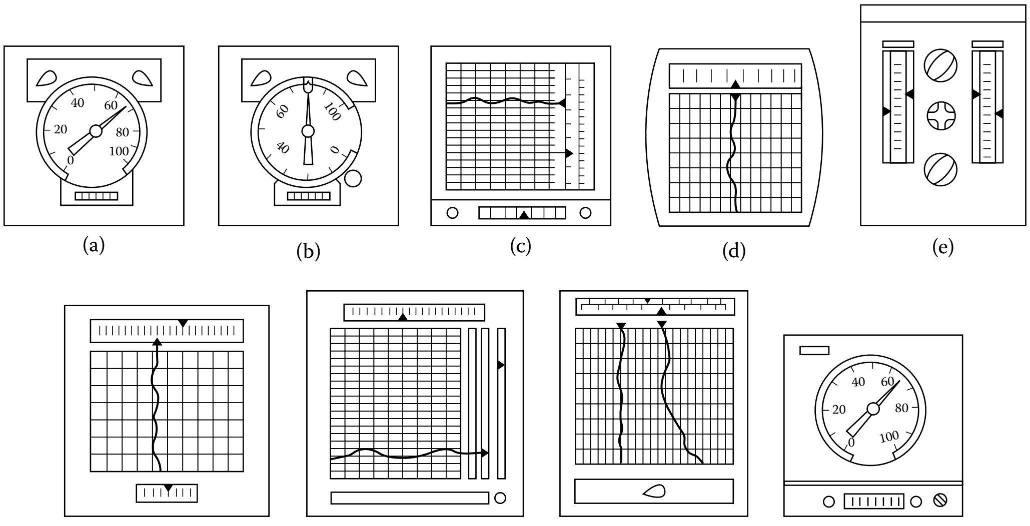


FIG. 3.1f
Types of miniature pneumatic controllers. (a) Typical indicating control station. (b) Indicating control station with 12 o'clock scanning features. (c) Recording control station with 30-day strip-chart and vertical moving pen. (d) Recording control station with horizontal moving pen and daily chart tear-off feature. (e) Indicating control station with two duplex vertical scale indicators (f) Recording control station with no "seal" position. (g) Recording control station with servo-operated pen. (h) Recording control station with procedureless switching. (i) Indication control station with instant procedureless switching.

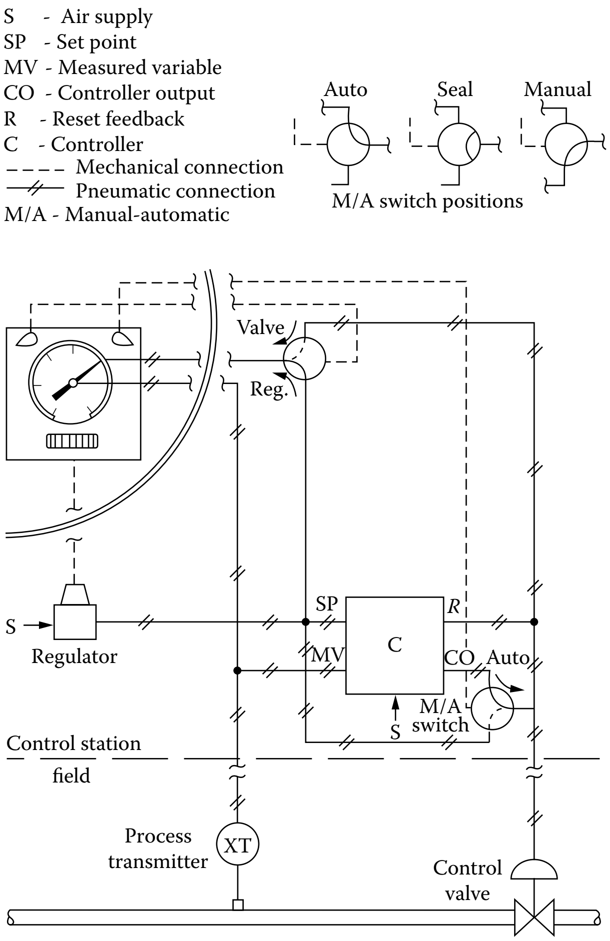


FIG. 3.1g
Miniature control station with integral-mounted controller.

The interconnections of an indicating miniature control station are shown in Figure 3.1g. The measured variable is indicated on the center pointer, and the set point is indicated on the peripheral pointer of a duplex gauge. On automatic, the operator changes the set point by adjusting the set point regulator and noting the set point on the gauge.

The controller is connected to the control valve via the manual–automatic switch. The controller compares the measured variable with the set point and manipulates the control valve to bring the variable on set point. If the operator wishes to switch to manual, he or she notes the valve pressure by operating the upper left-hand switch on the station. Next, the operator turns the right-hand switch to "seal," which isolates the controller from the control valve. After that, the operator adjusts the regulator to match the noted valve pressure and then turns the right-hand switch to the manual position, which connects the regulator directly to the valve. In manual, the operator directly adjusts the valve, while the controller reset follows the changes that are made to the valve.

In switching back to automatic, the operator goes to "seal" position and adjusts the set point regulator to match the measured variable. If the set point and measured variable are equal at switchover, and if the reset is equal to the valve pressure, the controller output should then equal the valve pressure, and the switchover is effected without a "bump."

Four-Pipe System Since a lag exists in the transmission of pneumatic signals, the dynamic capability of a control loop can be affected by increasing the distance between the controller and the process. Transmission lag will interfere with the performance of fast control loops such as liquid flow control but will not be significant if the distance is up to 100 ft (30 m).

For control loops where the transmission lag cannot be tolerated, the controller can be mounted locally, near the measuring transmitter and control valve, as shown in Figure 3.1h.

In this design, four air signal tubes are run between the control station and the field-mounted equipment. These carry the measured variable, set point, valve pressure, and relay operating pressure lines. Hence, the name "four-pipe system," whereas the configuration in Figure 3.1g is referred to as a "two-pipe system." Since the lines going back to the station amount to dead-ended parallel connections, the dynamics of

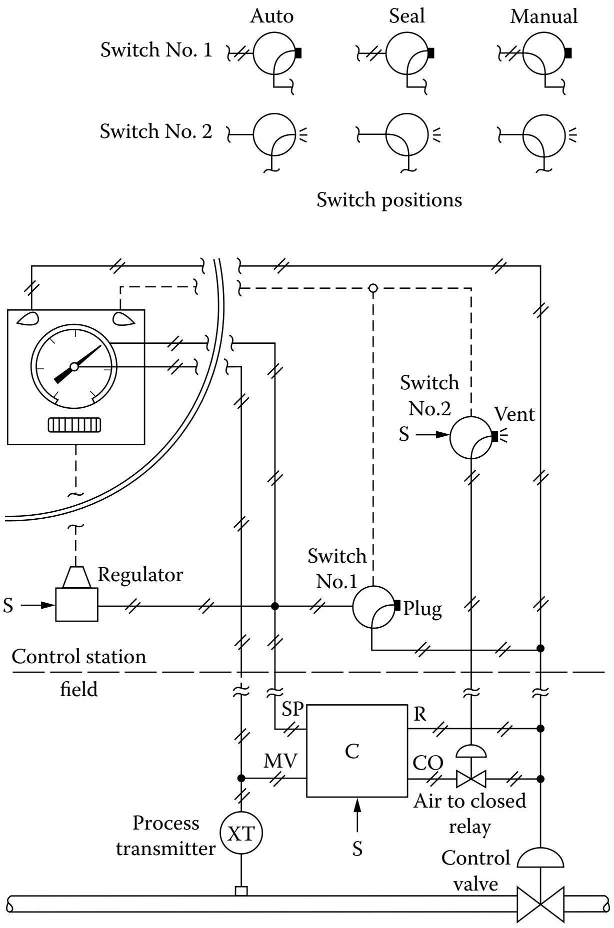


FIG. 3.1h
Miniature control station with field-mounted controller (four-pipe system).

the control loop are the same as they would be with any closed-loop system.

In switching from automatic to manual, the operator switches the loop to the “seal” position, which actuates the cutoff relay, to isolate the controller from the valve and permits the operator to change the set point regulator to match the noted valve pressure. This pressure is then connected to the valve when the operator turns the switch to manual. Returning to automatic involves the same procedure as was described for Figure 3.1g.

The disadvantages of this circuit are that:

1. It is more costly to run four transmission tubes between the control station and the field-mounted equipment.
2. The controller settings cannot be adjusted from the control panel.

Effect of Transmission Distance Since there is a transportation lag caused by pneumatic transmission, control is affected as the distance between the process and controller increases. Figure 3.1i shows the effect of increasing transmission distance between the process and the controller. As can be seen, a 10% upset in liquid flow will result in longer and longer recovery times as the transmission distance increases. Since liquid flow control is one of the fastest processes, this amounts to a worst-case example; the effect on slower processes will be proportionately less.

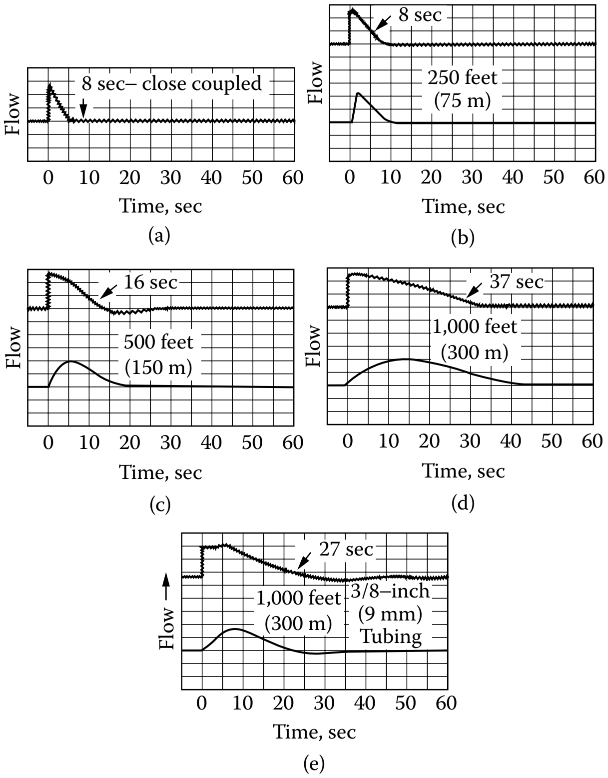


FIG. 3.1i
Effect of transmission distance on control of a liquid flow control process (worst-case example) with 10% step upset. Upper noisy curves show flows recorded locally; lower smooth curves show flows recorded remotely at controller. (J.D. Warnock, “How Pneumatic Tubing Size Influences Controllability,” Instrumentation Technology, *February 1967.)*

From the charts it can be seen that with all instruments close-coupled it took 8 seconds for the system to recover. At a transmission distance of 250 ft (75 m), the recovery time was still approximately 8 seconds. At 500 ft (150 m), it was 16 seconds, and at 1000 ft (300 m), 37 seconds. These results were obtained using 1/4 in. (6 mm)-OD tubing, which is conventional.

When 3/8 in. (9 mm) tubing was used, there was a significant improvement. At 1000 ft (300 m), the recovery time was reduced from 37 to 27 seconds. An equivalent electronic control loop had an 8-second recovery time. The upper, noisy record on the charts (b), (c), (d), and (e) is a plot of the flow as it was recorded at the transmitter. The lower, smoother records are the flows as they appeared on the remotely located recorder-controller.

If this lag is objectionable, the four-pipe system shown in Figure 3.1h can be used, and if it was, the dynamic performance would be equivalent to that of the closed-loop system. Other options include the installation of booster relays in the transmission lines or the use of larger-diameter transmission tubing.

Remote-Set Stations If a modification as shown in Figure 3.1j is added to the basic station in Figure 3.1g, the control station can accommodate a remote set point signal from sources such as remote ratio or proportioning relay, primary cascade controller, or computer.

Computer-Set Station The addition of a stepping motor to the set point regulator or to the set point motion transmitter as in Figure 3.1k allows the control station to be set from a digital computer. The station provides the option of switching the loop from computer control of the set point to manual, as shown in the Figure. The stepper motor, which operates the regulator, can be driven by time-duration signals or by individual up-down pulses. A resolution of at least 1000 pulses for full scale is provided.

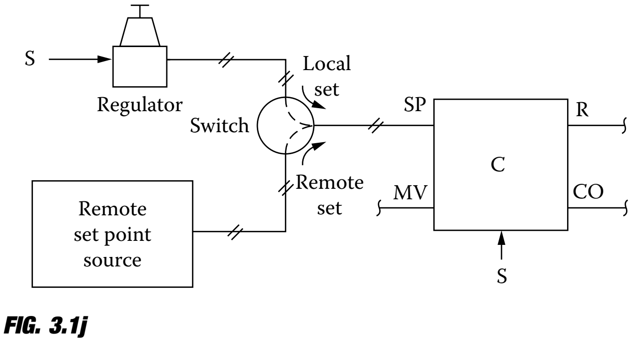


FIG. 3.1j
Control station circuit for remote set point adjustment.

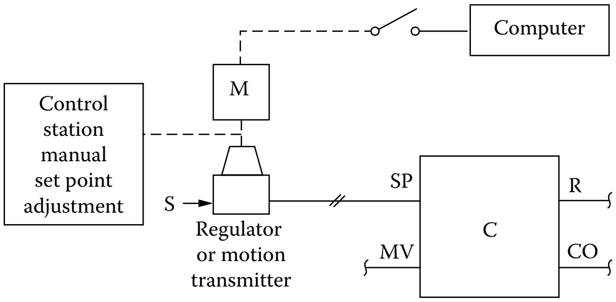


FIG. 3.1k
Control station circuit for computer-adjusted set point control.

Single-Station Cascade Cascade control can be implemented either with two controller stations, with the master controller (such as Figure 3.1g) generating the set point for the slave (or secondary), which is connected as in Figure 3.1j. The other option is to use a single station, such as shown in Figure 3.1l. The latter scheme not only eliminates one of the two stations but also offers operating safety and convenience.

A common problem with having separate master and slave stations is that when the operator switches the slave to manual, he or she often forgets to also switch the master station to manual as well. In such situation, the primary controller has no influence on the manipulated variable and will not be balanced when the loop is returned to cascade. This is particularly undesirable because cascade circuits are usually used on the most critical loops.

In Figure 3.1l, the regulator has three functions:

1. Set point to primary controller in cascade control
2. Set point to secondary controller for independent secondary control
3. Manual valve setting

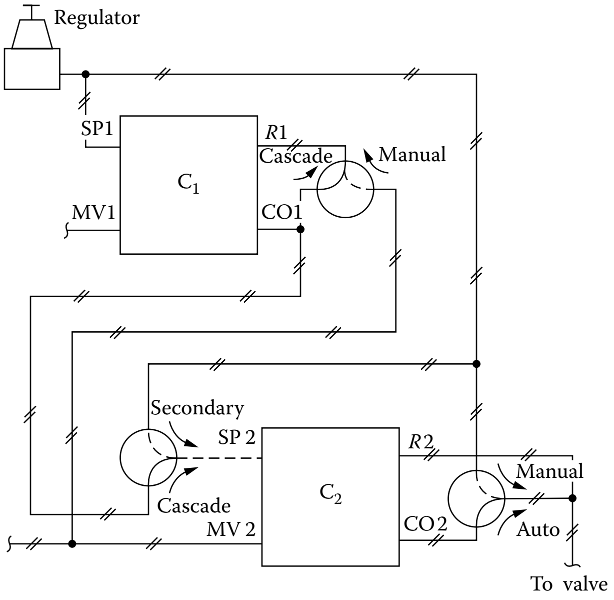


FIG. 3.1l
Single-station cascade controller.

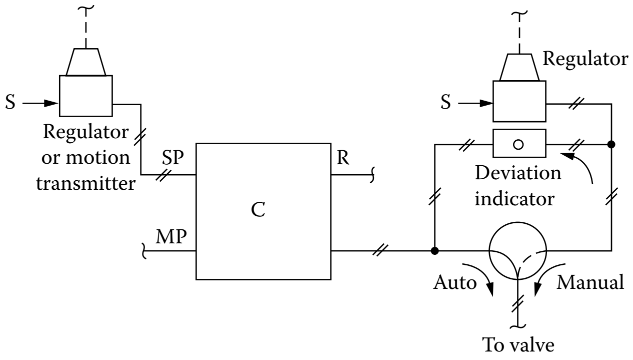


FIG. 3.1m
The use of two-regulator stations eliminates the need for a "seal" position.

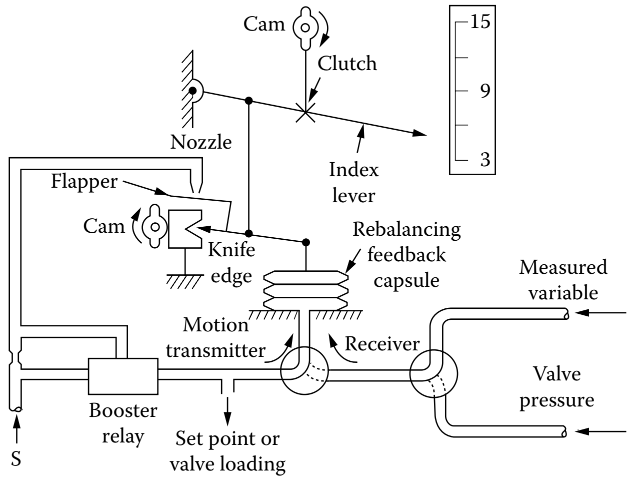
There is a seal position between each step while the regulator is set for its upcoming function. The key to this station is the concept used in making the secondary measured variable (MV2) the reset feedback of the primary controller while on manual or secondary control. Versions are also available that allow cascade, independent automatic control on the primary and manual modes of operation.

"No Seal" Station By using two regulators or two motion transmitters in a station (Figure 3.1m), the need for a "seal" position can be eliminated. In this configuration, when the operator wishes to switch to manual, he or she adjusts the manual regulator to match the controller output while viewing a deviation indicator. When the deviation is zero, the operator transfers control.

Procedureless Switching Stations With procedureless switching, the operator simply turns a switch, and the station automatically takes care of the pressure-balancing requirements. Two designs are shown in Figure 3.1n. The mechanism on the top (a) is a motion transmitter/receiver combination. This motion transmitter provides the set point pressure when the controller is in automatic and the valve pressure when in manual—similarly to the regulator in Figure 3.1g.

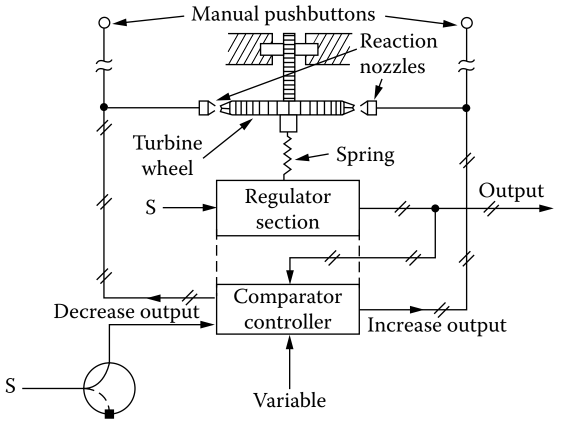
As a motion transmitter, a friction clutch holds the index lever at whatever position the operator has set it. A restriction-nozzle circuit senses the position and converts it to a proportional pneumatic output that is fed back to the rebalancing bellows. A 0 to 100% movement of the index gives a 3- to 15-PSIG (0.2- to 1.0-bar) output pressure. When acting as a receiver, supply pressure is cut off from the restriction nozzle circuit and the pressure to be sensed is admitted to the rebalancing bellows.

The friction clutch is disconnected so that the index lever can be moved by the rebalancing bellows. The unit is so designed and calibrated that a 3- to 15-PSIG (0.2- to 1.0-bar) sensed pressure produces a 0 to 100% index movement. Therefore, as the operator moves the switch from automatic



Alternating motion transmitter/receiver

(a)



Self-synchronizing regulator

(b)

FIG. 3.1n
Two approaches to self-balancing and to procedureless switching in control stations.

to manual, the following actions take place in sequence and automatically:

The index is declutched; supply is cut off from the restriction nozzle circuit; controller output (valve) pressure is admitted to the bellows causing the index level to take a position proportional to the pressure; controller output is disconnected from the valve line; the clutch is engaged; supply pressure is readmitted to the restriction nozzle circuit; and the unit again acts as a motion transmitter, now providing valve pressure.

Switching back to automatic involves the same sequence, in reverse order. This is the same sequence as carried out in Figure 3.1g, except that here it is automatic.

A second system for procedureless switching involves the use of self-synchronizing regulators or synchros (see design b in Figure 3.1n). One regulator provides set point, while the other is used for manual valve loading. As the name implies, this regulator can synchronize itself to some varying pneumatic pressure and thereby provide automatic balancing.

The regulator employs a reaction nozzle circuit that results in very low spring force (approximately 1 oz or 0.28 N) to develop a 3- to 15-PSIG (0.2- to 1-bar) output. The setting spring is adjusted by rotation of a turbine wheel with an integral lead screw (part b in Figure 3.1n). If supply is connected to the comparator controller section, air is transmitted to the increase–decrease nozzles to make the regulator section output match the variable input pressure. If supply is cut off from the comparator controller, the regulator section output remains locked in, with the memory being a function of lead screw position. The unit can then be driven manually by the operator.

When this control station is in automatic, the set point synchro is manually adjusted by the operator, while the valve-operating synchro keeps itself matched to the controller output to allow instant transfer to manual. In manual, the operator adjusts the valve-loading synchro while the set point synchro tracks the measured variable.

In addition to procedureless switching, these stations can also be switched from a remote source, manually or automatically. They can also be gang-switched, and they can be operated in parallel. For example, it is possible to have one station in the central control room and the other out in the field. The operator can use either station and whichever is not in active service keeps itself fully synchronized and ready to take active duty at any instant.

High-Density Stations

Miniature high-density stations have a typical panel size of 3 in. × 6 in. (7.5 mm × 150 mm), mount adjacent to each other, and allow very compact and efficient panel arrangements (Figure 3.1o). The units incorporate packaging features that simplify panel construction and design and facilitate servicing. Much of the design is aimed at making the job of the operator simpler, faster, and safer, in line with the present trend to consolidate control rooms, minimize the number of operators, and handle increasingly fast, complex, and critical processes.

Mid-Scale Scanning Station Figures 3.1p, 3.1q, and 3.1r show three types of high-density control stations featuring a mid-scale deviation scanning pointer. The pointer is driven either by a differential detector or differential servo that compares the measured variable against set point. If the two are equal, the red deviation pointer is positioned at mid-scale, where it is screened off by a green scan band. If there is a deviation, the red pointer stands out prominently.

In Figure 3.1p, a fixed, nominal 4 in. (100 mm) vertical scale is employed, and there are separate pointers to indicate set point and measured variable. The station uses the two-regulator approach to achieve “no-seal” switching as in Figure 3.1m. The operator does have to balance pressures before switching, however.

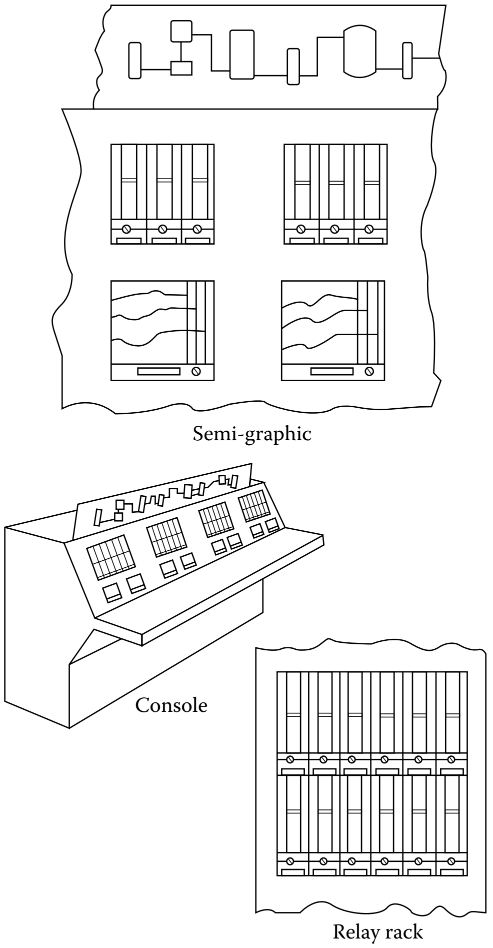


FIG. 3.1o
Typical mounting arrangements of high-density control stations.

In Figure 3.1q, the station employs an expanded scale, which provides greater readability. The only indication on the scale is deviation, however, and this requires that the set point transmitter scale and deviation servo stay in calibration relative to each other in order to provide an accurate reading of the variable. While the expanded scale gives greater readability, it does have to be moved to bring the reading on scale when the variable makes any excursions.

This station and the one in Figure 3.1r use the two-regulator approach to eliminate the need for a seal position. In both cases the operator balances pressures prior to switchover. In Figure 3.1q, the operator notes deviation on a ball-in-tube indicator. In Figure 3.1r, the valve switch has a detent action while the indicator switch operates at the mid-throw position, so that the operator moves the integral switch lever back and forth across center while manually matching pressures before switching. The controller in Figure 3.1r is a deviation type actuated by displacement of a deviation link in the indicator circuit. The controller acts to hold the link at its "zero" position.

While thousands of these control stations have been used in control rooms, mid-scale scanning stations have evolved

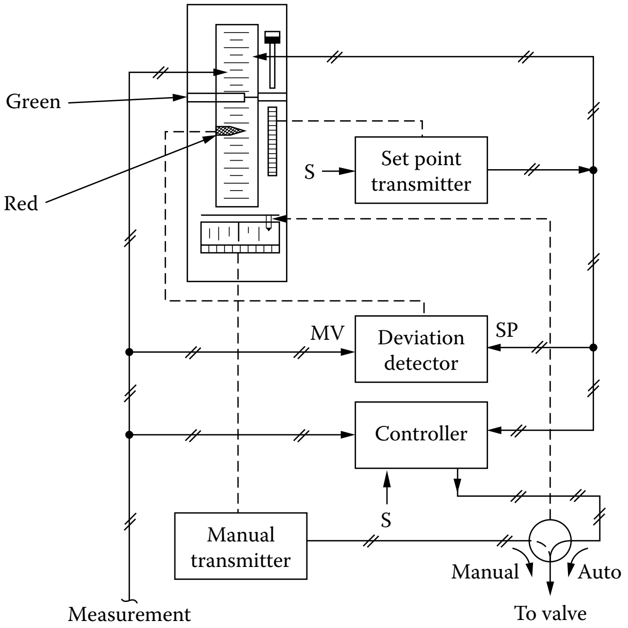


FIG. 3.1p
Functional diagram of high-density station with mid-scale scanning and individual indication of set point and measured variable.

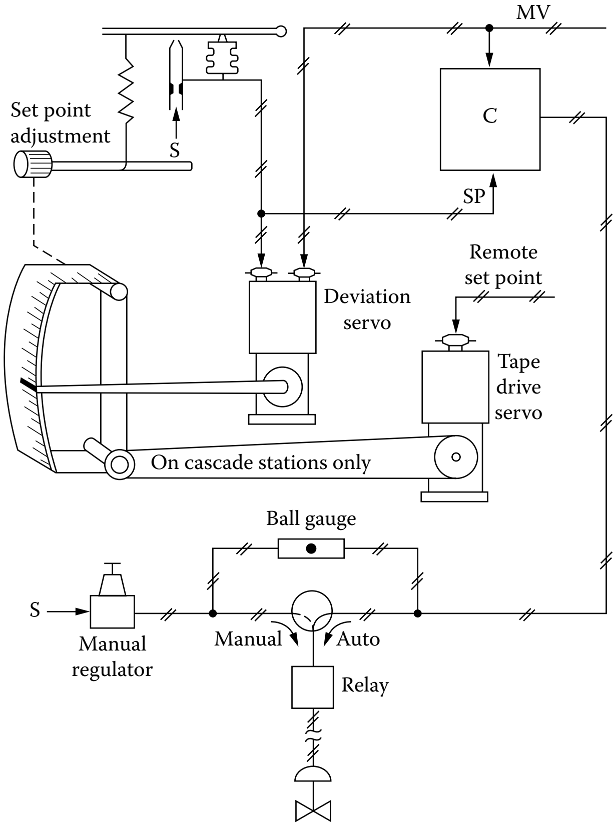


FIG. 3.1q
Mid-scale scanning station with expanded scale.

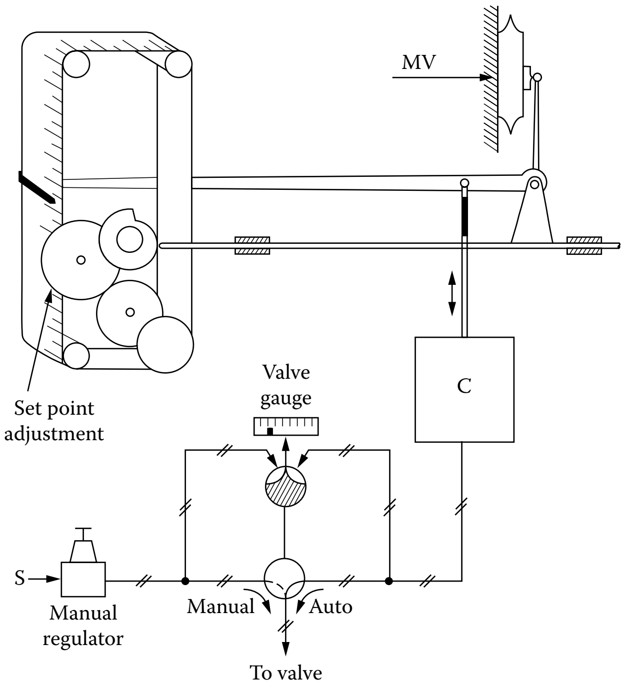


FIG. 3.1r
Mid-scale scanning station, expanded scale, and deviation controller.

into a type combining mid-scale scanning with procedureless switching, using principles described in the next paragraph in connection with Figure 3.1t

Procedureless Switching Station Stations shown in Figures 3.1s and 3.1t offer procedureless switching for the operator. Both have a fixed 4 in. (100 mm) scale, separate indication of set point and measured variable, and a scanning technique that allows the set point indicator to overlap the measured variable indicator when the controlled variable is on set point.

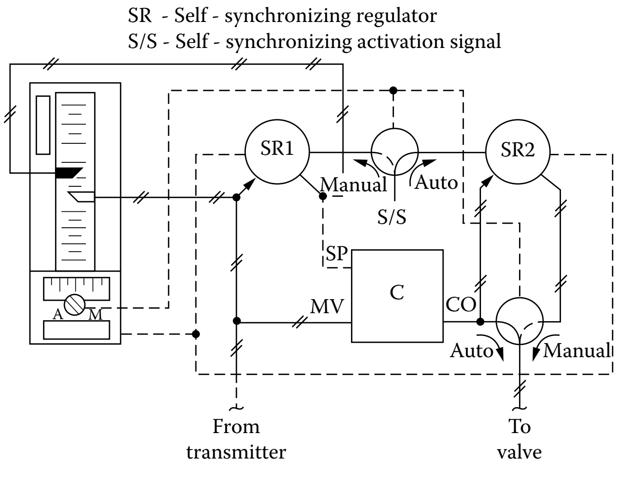


FIG. 3.1s
Self-synchronizing control station with procedureless switching.

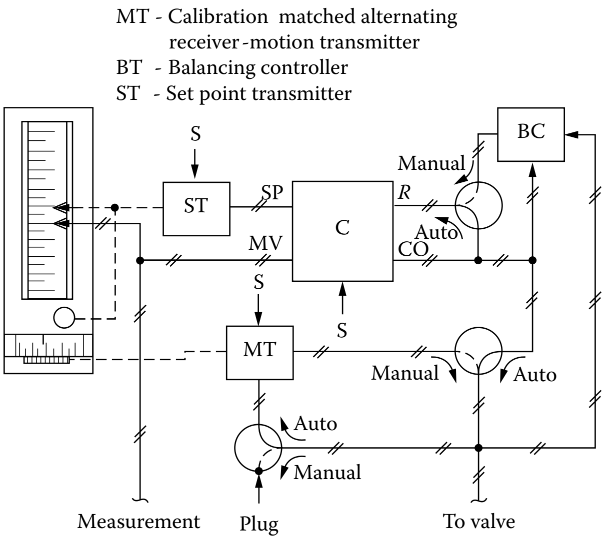


FIG. 3.1t
Self-balancing control station.

The station in Figure 3.1s employs two self-synchronizing regulators or synchros (see Figure 3.1n), one for set point and the other for manual valve loading. When the controller is in automatic, the operator manually adjusts the set point synchro, while the valve-loading synchro automatically tracks controller output. In manual, the operator adjusts the valve-loading synchro while the set point synchro tracks the measured variable. Thus, the station is always balanced allowing for instant transfer of the control mode.

Like its 6 in. × 6 in. (150 mm × 150 mm) counterpart, these stations can also be gang-switched, switched remotely or automatically, and operated in parallel from different locations while maintaining themselves in synchronism, and they are also available in single-station cascade arrangements.

Motion Transmitter/Receiver The station in Figure 3.1t employs a dual function motion transmitter/receiver for manual valve loading and valve pressure indication. This unit is similar to that described in Figure 3.1n. In automatic, the index lever is declutched and the feedback capsule, connected to the valve pressure line, moves the index accordingly. In manual, the clutch engages, and the mechanism reverts to a motion transmitter that provides manual valve loading. This allows procedureless switching to manual.

Switching to automatic is also procedureless, assuming that the set point of the process does not change. If the operator wishes the controller to operate at some new value, compared to where the process had been set on manual, the operator must change the set point to that value prior to switching to automatic. However, to facilitate switching to automatic on loops where the set point remains fixed, the station incorporates a separate balancing controller that operates while the station is in manual.

The balancing controller manipulates controller reset pressure to keep the controller output equal to the manual valve loading even though the measured variable may be off set point. This allows the operator to switch to automatic while off the intended set point and yet have the pressures balanced at switchover and therefore the system to return to set point without overshoot.

This feature of the circuit is somewhat limited on narrow proportional band applications, which is usually the case with slow processes. This is because it takes little deviation from set point before the reset would have to be at either a vacuum or considerably above the air supply pressure to obtain the balance between valve loading and controller output.

LARGE-CASE DESIGNS

Receiver Controllers

Because of their size, large-case receiver controllers are even less frequently used as miniature designs. They nevertheless still find use in some plants and industries where the 24-hour circular chart is traditional and preferred.

Most large-case controllers operate on a displacement balance principle. The set point is a mechanical index setting. The measured variable acts on a pressure spring such as a bellows, spiral, or helix that moves the recorder pen or indicator pointer. A differential linkage detects any deviation between the index and pen position and actuates the flapper-nozzle system in an effort to bring the deviation down to zero.

One example of a large-case recording controller is shown in Figure 3.1u. In this design, if the pen moves clockwise, the differential link moves upward, and the bell-crank moves the flapper toward the nozzle. The resultant increase in nozzle backpressure is reproduced by the relay, whose output is connected to the control valve and the housing of the proportioning bellows. The pressure increase is transmitted to the small inner bellows, causing the two connected inner bellows to move to the right. The spring in the left inner bellows compresses while the spring in the right distends.

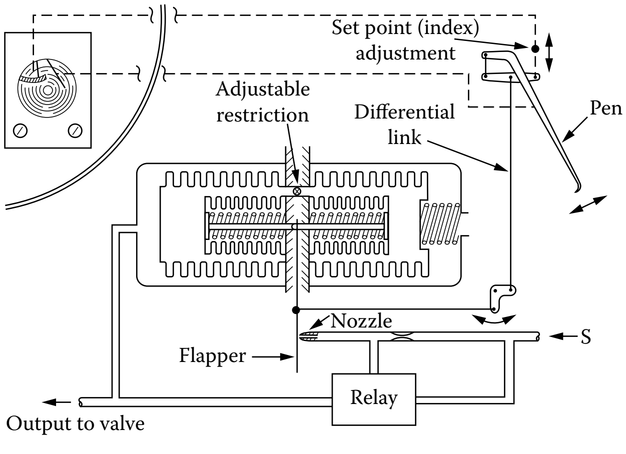


FIG. 3.1u
Large-case recording controller.

The motion also causes the large right-hand bellows to move to the right against the housed spring. As the center rod joining the inner bellows moves to the right, the flapper is moved back away from the nozzle. This negative feedback results in proportioning action. The greater this negative feedback, the greater the change that will be required in the measured variable to obtain a given change in the valve. Adjusting the linkage to change the amount of this negative feedback changes the proportional band.

Opening the adjustable restriction between the two large bellows allows flow from the bellows at higher pressure to the one at lower pressure. In this example, it would flow from the left to the right bellows, causing the inner bellows to move left, moving the flapper toward the nozzle and increasing output further. This action would continue to regenerate until the pen finally returned to the index, where full balance would be achieved with the pressure equal in the two large bellows and with the two inner bellows centered.

Large-case controllers are also available with remote air-operated set point adjustment as required in cascade and ratio control.

Direct-Connected Controllers

Direct-connected controllers have their own measuring elements, which, as the term implies, are directly connected to the process. They therefore eliminate the need for a transmitter. However, the fact that direct-connected controllers must be located in the vicinity of the process limits their use to mounting on local control panels. With this design, the process connections must be run to the local control panel, which is costly, troublesome, and hazardous. For these reasons, these units are seldom used, even on smaller installations and on local panels.

Direct-connected large-case controllers of the indicating and recording type predated the receiver type units by some 20 years. In the first designs, the receiver controllers were direct-connected pressure controllers with a 3- to 15-PSIG range.

The operation of direct-connected large-case controllers is the same as that of large-case receiver controllers. The pen or pointer arm, instead of being actuated by pressure from a process transmitter, is actuated by its own built-in process pressure detector. The cross sections of some of the measuring elements that are available with large-case controllers are shown in Figure 3.1v.

These include sensors for the detection of pressure, absolute pressure, draft, vacuum, differential pressure, liquid level, and filled systems for temperature measurement. Basic electrical measurements involved in thermocouples, resistance bulbs, radiation pyrometers, and pH probes are accommodated in the potentiometer versions of large-case pneumatic controllers.

As with the large-case receiver controller, the direct-connected ones are also available with remote set point adjustment for cascade and ratio control applications.

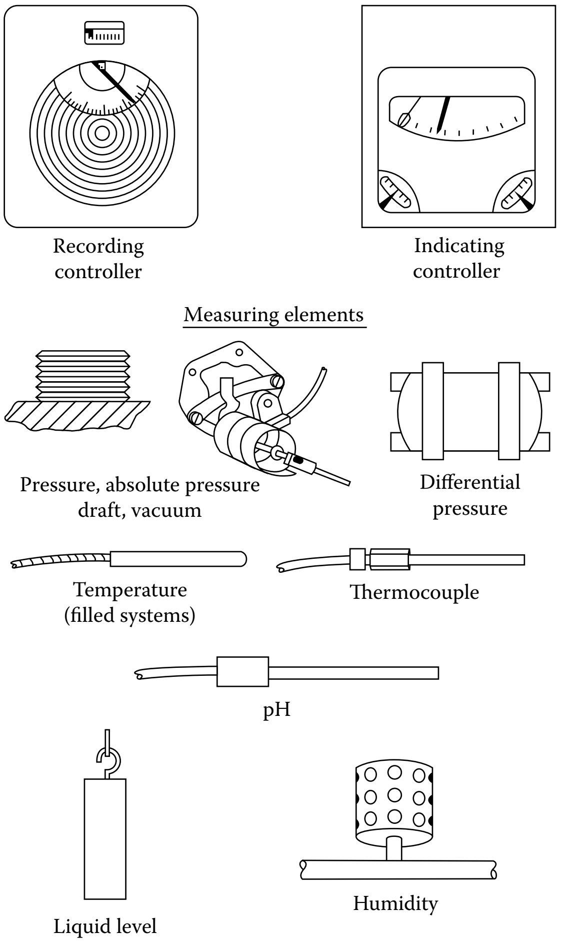


FIG. 3.1v
Direct-connected large-case recording and indicating controllers can be provided with a variety of direct sensors, including the shown measuring elements.

FIELD-MOUNTED CONTROLLERS

Direct-Connected

Direct-connected controllers mounted in the field are smaller than the large-case instruments. They usually have weatherproof cases and are available in both indicating and blind designs. They can be pipe-mounted, mounted directly on a valve or surface, or flush-mounted on a local panel. They include their own measuring elements. Figure 3.1w shows some typical types and mounting arrangements.

These units are the least expensive pneumatic controllers, and they are expected to be less accurate than the previously discussed designs are. They find use in small local installations and as local field loops in larger plants. Every plant has some noncritical loops that do not require auto–manual switching or the displaying of their controlled variable or access to their set point from the control room. These units are used as local pressure, temperature, and level regulators.

These field-mounted controllers can be had with on/off, differential gap, proportional, reset, and derivative modes of control. Their principles of operation are similar to that discussed in connection with Figures 3.1a and 3.1w.

Receiver-Type

Some field-mounted local controllers receive a signal from a transmitter rather than having their own measuring element. Such controllers are used if the measurement must be transmitted to the control board for recording, alarm, or indication, but the controller is local. Field-mounted receiver controllers are shown in Figures 3.1c and 3.1d. Remote adjustment of set point is also an option with these controllers.

Pneumatic with Electronic Detectors

The temperature controller illustrated in Figure 3.1x, receives a temperature measurement signal from either a thermocouple or a resistance temperature detector (RTD). These indicating pneumatic controllers provide the accuracy, convenience, and range of electrical temperature sensing but without the need for external electrical power. A built-in generator accepts a conventional pneumatic pressure source and produces electrical power for the controller.

The controller compares the process temperature signal with an operator-adjustable set point and delivers a pneumatic control signal to a final control element, which then moves the process temperature towards the set point. Process temperature

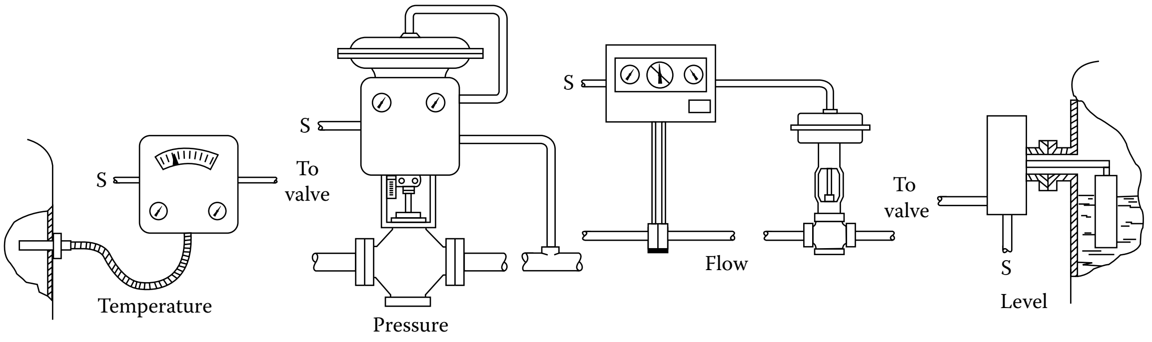


FIG. 3.1w
Typical installations of direct-connected, locally mounted controllers.

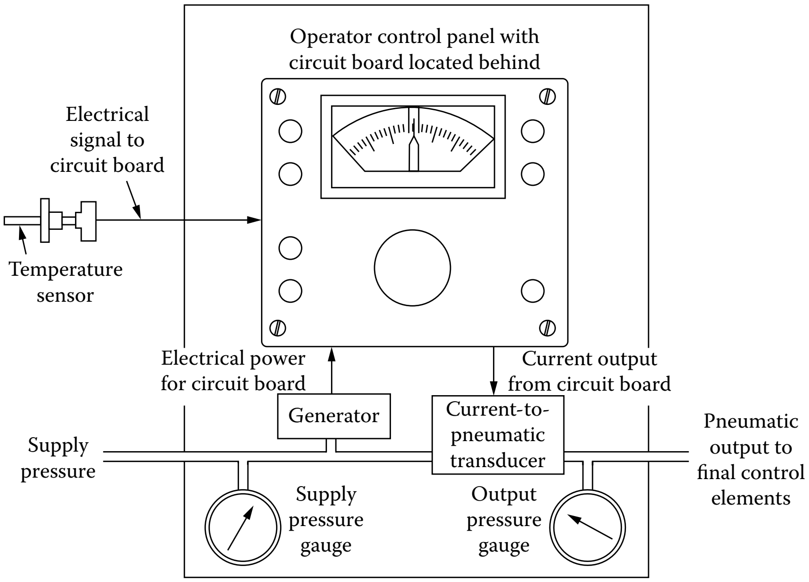


FIG. 3.1x
Pneumatic controller with electronic temperature sensor.

and set point are both indicated on this controller. These controllers are available with PI and PID control modes and with or without bumpless and balanceless auto/manual switching.

In some designs, the air supply is also used to generate the required electric power for the unit and therefore, the maximum "instantaneous" air supply requirement could reach 80 SCFH (2.3 normal m^3/hr), while the maximum "steady state" air consumption is only 35 SCFH (1.0 normal m^3/hr).

In this controller, a thermocouple or RTD temperature sensor signal is applied to the circuit board, which electrically compares the signal with the set point value and acts on the error with its proportional, reset, and rate control modes to restore the process temperature to the set point value.

SPECIAL PNEUMATIC CONTROLLERS

Connections for Digital Highways

Figure 3.1y illustrates a pneumatic controller, which is provided with microprocessor-based digital highway communication modules for integration into distributed control systems (Figure 3.1y). By thus making it possible to communicate over a data highway, pneumatic controls were made compatible with distributed systems.

The serial communications module reports, upon command, the current value of the set point, process variable, and valve output and the operating mode that the station is in. The communications module can also receive and execute commands, which can change set points or outputs or operating modes. Miniature transducers convert the pneumatic signals to electric, from which the signals are then converted to digital. The serial link operates at 19.2 kilobaud.

The first data highways were introduced in the late 1970s. Fieldbus is an architecture that provides a communication link between all instruments and all computers using a standard interface. Several manufacturers have entered the fieldbus technology market. However, the Internet seriously threatens the fieldbus market, because Web-based communication between instruments and computers may eventually lead to the replacement of some fieldbus components.

Special Controls

Feedforward, ratio, cascade, and other loop configurations can be easily implemented with pneumatic hardware. Two of the popular control configurations involve selective and batch

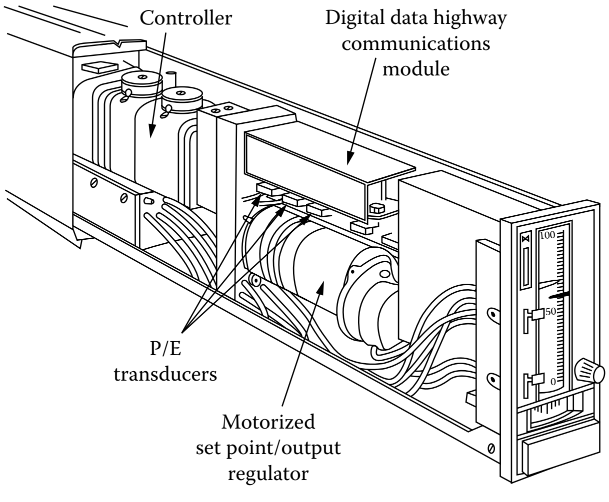


FIG. 3.1y
Pneumatic controller with digital highway communication module for use in microprocessor-based distributed control systems.

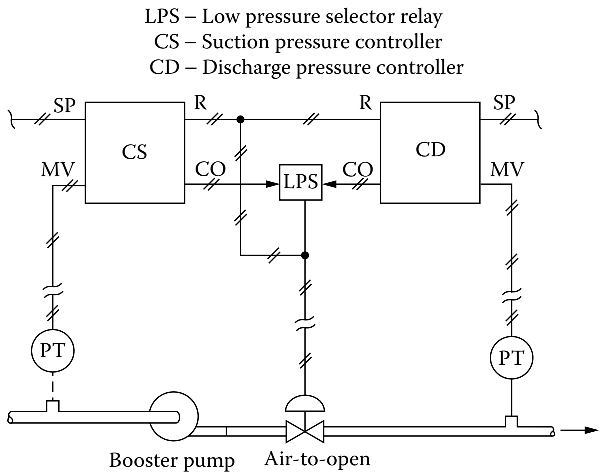


FIG. 3.1z
Automatic override or selector control circuit on pipeline booster pump.

control. Controllers are sometimes packaged specially for these applications.

Selective Control With automatic selector control, also called override or limit control, two or more control loops are connected to a common valve. In this configuration, when the conditions are normal, the normal controller has command of the valve. However, if some abnormal condition arises, one of the other loops can automatically take over control to keep the plant operation within safe limits.

Unlike safety shutdown systems, here the plant is kept in operation, although its production rate might be cut back as much as necessary to stay within safe limits. When the abnormal condition abates, the normal loop resumes control. Figure 3.1z shows a control system that can be used on a booster pump station that is serving a transcontinental pipeline.

Under normal conditions the discharge pressure of the pump is being controlled. If the suction pressure gets too low, however, as would be the case if the booster pump upstream failed or if a line rupture occurred, the discharge controller would open the valve wide, which would lower the suction pressure beyond safe limits, causing cavitation, which could seriously damage the pump.

To protect against this, a suction pressure controller is added, which is set to the low safe limit, and a low-pressure selector is installed on the two controller outlets. Since the control valve is air-to-open, the low selector chooses the output of that controller, which is asking for the valve to be less open. Under normal conditions, when the suction pressure is adequate, the output of the discharge pressure controller will be the lower and hence, it will throttle the valve. If suction pressure drops to the set point of the suction pressure controller, it automatically takes over control.

A key requirement for correct implementation of this system is to provide both controllers with a live reset feedback from the valve pressure. In this way the controlling unit has its reset acting normally while the standby controller is prevented from saturating or winding up. This way, when either controller takes over, its reset will exactly match the pressure of the signal reaching the valve at the instant of switchover.

Batch Control Figure 3.1aa shows a control system for batch pressure control. Batch control is special because when a batch is completed and the manual valve is closed while the controller is in automatic, the reset mode of the controller will keep integrating the error until the controller saturates (winds up).

At the next startup, if the controller is PI only, or if it is a PID controller but its derivative mode is interacting with the proportional and reset actions (Figures 3.1a and 3.1c), the controller output will stay saturated until the error changes sign. Therefore, the controller will stay inactive until the process measurement crosses over set point. This will result in a considerable overshoot.

One solution is to add the antireset windup relay shown in Figure 3.1aa. This is simply a throttling relay set to operate at 15 PSIG (1 bar), which corresponds to the wide-open position of the control valve. As long as the controller output is below

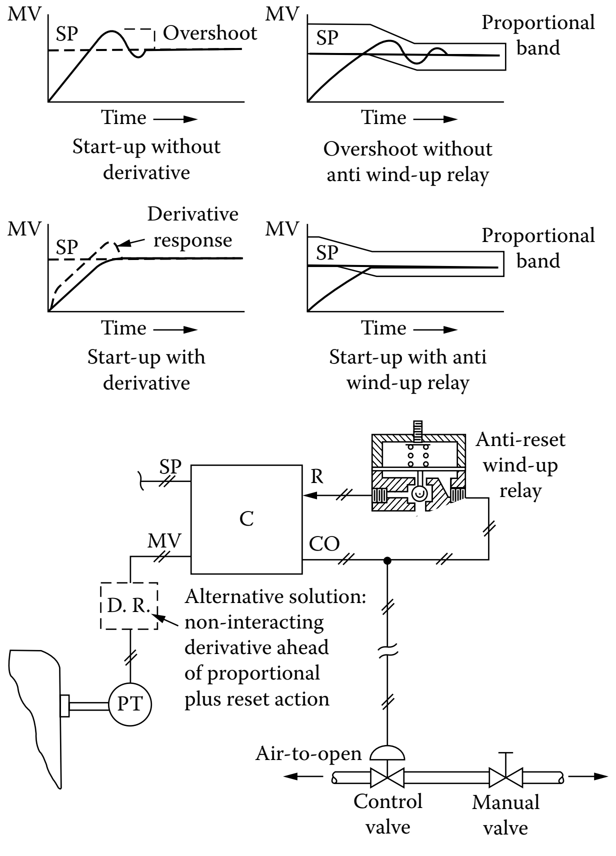


FIG. 3.1aa
Alternative circuits for eliminating over-peaking during startup of batch processes.

15 PSIG, the relay transmits the output to the reset feedback connection of the controller and the reset acts normally.

If the output goes above 15 PSIG, the relay begins exhausting the reset feedback line until the output pressure drops to 15 PSIG. Thus it does not affect control except when the system is shut down. At that time it lowers the reset pressure to whatever value it has to be in order to limit its output to 15 PSIG. This allows the proportional action to be active during startup so that it can prevent overshoot. How effective the antireset windup protection is depends upon the quality of the tuning of the controller.

An alternative solution to reset windup on batch applications is to use either a separate derivative unit or a controller with a built-in derivative unit ahead of the proportional-plus-integral sections. This way the derivative unit's output crosses over the set point sooner than the variable itself does, and this starts the reset to unwind, which is in time to prevent overshoot.

The effectiveness of this circuit also depends upon proper tuning. Too little derivative allows some overshoot; too much causes initial undershoot.

Bibliography

"Analog Controllers," *Measurements and Control,* February 1992.

"Analogue Pneumatic Signal for Process Control Systems," IEC 60382, 1991.

Buckley, P. S., "Dynamic Design of Pneumatic Control Loops: Parts I and II," *InTech,* April and June 1975.

Buckley, P. S., and Luyben, W. L., "Designing Longline Pneumatic Control Systems," *Instrumentation Technology,* April 1969.

Connel, B., *Process Instrumentation Applications Manual,* New York, McGraw-Hill, 1996.

"Controllers with Analogue Signals for Use in Industrial-Process Control Systems. Part 1: Methods of Evaluating the Performance," IEC 60546-1, 1987.

Goldfeder, L. B., "Analog Control Architecture: Integral vs. Split," *InTech,* April 1977.

Jones, B.E., *Instrumentation, Measurement and Feedback,* New York: McGraw-Hill Book Co., 1977.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Parr, E. A., *Hydraulics and Pneumatics: A Technicians and Engineers Guide,* Oxford: Butterworth-Heinemann, 1999.

Patrick, D. R., and Fardo, S. W., *Industrial Process Control Systems,* Albany, NY: Delmar Learning, 1997.

Pessen, D. W., *Industrial Automation: Circuit Design and Components,* New York: John Wiley & Sons, 1989.

"Quality Standard for Instrument Air," ANSI/ISA-7.0.01-1996.

"Reset Feedback in Pneumatic Controllers," Moore Products Co., Spring House, PA.

Seborg, D., et al., *Process Dynamics and Control,* New York: Wiley, 1989.

Senbon, T., Hanabuchi, F. (Eds.), *Instrumentation Systems — Fundamentals and Applications,* Berlin: Springer-Verlag, 1987.

Shinskey, F. G., "Control Topics," Publ. No. 413–1 to 413–8, Foxboro Co.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1997.

Stenerson, J., *Industrial Automation and Processs Control,* Upper Saddle River, NJ: Prentice Hall, 2002.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall Inc., 1984.

3.2 Controllers—Electronic

P. M. B. SILVA GIRÃO (2005)

Types:	Analog, digital.
Applications:	Single-loop or multi-loop; single or multi-variable.
Control Modes:	Analog: Manual, two-step (on/off), multi-step, proportional (P, throttling), integral (I, reset), derivative (D, rate), cascade, selective, feed forward, limiting, ratio and adaptive. Digital: all types listed for analog, plus multi-variable interactive, model based, neural and fuzzy logic.
Input Ranges:	4 to 20 mA (most common), 0 to 20 mA, 0 to 10 V, 2 to 10 V.
Typical Output Range:	4 to 20 mA.
Displays:	Set point, controlled (process) variable, manipulated (output) variable, deviation, alarms, limits.
Typical Front Panel Size:	1/32 DIN (1.89 in. × 0.95 in. [48 mm × 24 mm]), 1/16 DIN (1.89 in. × 1.89 in. [48 mm × 48 mm]), 1/8 DIN (1.89 in. × 3.78 in. [48 mm × 96 mm]), 1/4 DIN (3.78 in. × 3.78 in. [96 mm × 96 mm]).
Supply:	AC: 24, 120, 230 V. DC: 12 V.
Inaccuracy:	Analog: 0.5 to 1% of span; Digital: from under 0.1 to 0.5% of span (zero < 0.2%; gain < 0.1%).
Costs:	*On/off controllers (switches):* The simplest versions of both analog and digital units can be obtained for $100. HVAC quality direct-connected controllers range from $200 to $500. Process control quality analog PID controllers are in the range of $1000 to $1500, and microprocessor-based local digital controllers, with more options than their analog counterparts are in the $1000 to $2000 range.
Direct Signal Processor (DSP) Controller Capabilities:	Can be programmed for almost anything, including linearization, compensation for valve and actuator characteristics, a variety of control modes (including batch control), and a variety of algorithms, both linear and nonlinear, self-tuning and self-optimization.
Partial List of Suppliers:	ABB Group (www.abb.com) Action Instruments (www.actionio.com) Analogic Corp. (www.analogic.com) Athena Controls Inc. (www.athenacontrols.com) Barber-Colman Industrial Instruments (www.barber-colman.com) Barton (www.barton-instruments.com/index2.php) Bristol Babcock (www.bristolbabcock.com) Dwyer Instruments Inc. (www.dwyer-inst.com) Emerson Process Management (www.emersonprocess.com) Fisher Controls International, Inc. (www.fisher.com) Foxboro-Invensys (www.foxboro.com) Honeywell Automation and Control (www.honeywell.com/acs/index.jsp)

ICS Triplex (www.icstriplex.com)
ISE, Inc. (www.instserv.com/prod01.htm)
Jumo Process Control. Inc. (www.jumousa.com)
Love Controls Corp. (www.love-controls.com)
Matricon Inc. (www.matricon.com)
Omega Engineering Inc. (www.omega.com)
Parker Hannifin Corp. (www.pneutronics.com)
Partlow Process Instruments (www.partlow.com)
Powers Process Controls (http://www.powerscontrols.com/)
Robertshaw IDP, an Invensys Co. (www.robertshawindustrial.com)
Samson AG (www.samson.de/pdf_en/_ek16_re.htm)
Siemens Energy & Automation (www.sea.siemens.com)
Smar International (www.smar.com/products/technology.asp)
Spence Engineering Company, Inc. (www.spenceengineering.com/Handbook/index.htm)
Thermo Electric Co., Inc. (www.thermo-electric-direct.com)
Thermosystems (www.thermosystems.it)
Toshiba International Corp. (www.tic.toshiba.com)
Triplett Corp. (www.triplett.com)
United Electric Controls (www.ueonline.com)
Watlow (www.watlow.com)
Westinghouse Process Control (www.westinghousepc.com)
Wilkerson Instrument Co. (www.wici.com)
Yokogawa Corp. of America (www.yca.com)

INTRODUCTION

In this handbook, electronic controllers are discussed in three chapters, and there is some overlap between them. Chapter 2 is devoted to control theory; it discusses a number of software packages that include control algorithms. This chapter concentrates on local controllers. Chapter 4 covers control room equipment, including a variety of electronic controllers and DCS systems. Because electronic controllers can be mounted in either the control room or locally out in the field (properly protected), they are discussed in both Chapters 3 and 4.

Electronic controllers entered the market in the late 1950s using vacuum tube technology. In the beginning, the process variable and the controller output signals were both voltage signals, but they were not immune to noise. In addition, if the signal were to drive several series loads, it would degrade. Because current signals provide better accuracy, the 4- to 20-mA signal was adopted as standard. In the 1960s and 1970s, the performance of electronic analog controllers increased due to the use of solid-state electronics.

Digital control systems using mainframe computers were considered since the mid-1950s and early 1960s. However, it was the introduction of the microprocessor in the 1980s that boosted the use of digital process and today one can say that digital electronic controllers dominate the process controller market. The microprocessor allowed the implementation of more sophisticated and better-performing control algorithms and led to the production of controllers whose configuration and tuning can be easily changed or adapted to different applications.

The microprocessor has also contributed decisively to the development of new control approaches, such as DCS, which is discussed in Chapter 4; PLCs, which are covered in Chapter 5; and digital networks, which are the topic of Volume 3 of this handbook.

An electronic controller can be an ac- or dc-powered device whose inputs (set point and measurement [controlled process variable]) and output (manipulated variable) are electric signals. When the controller is manipulating the opening of an air-operated valve, its electronic output signal is converted to an air signal by an I/P or V/P (current-to-air or voltage-to-air) converter. A typical electronic control loop configuration is shown in Figure 3.2a.

Electronic controllers can be analog or digital. Analog controllers are real-time devices that operate on continuous analog signals. The operation of digital controllers is software-based and implements control algorithms, which are very fast but still are intermittent devices in comparison to their analog counterparts. Analog electronic controllers are cheaper than their digital counterparts when the control requirements are small (e.g., single control loop and simple on/off or plain proportional control algorithms).

Digital controllers are more complex because they need (1) to convert analog input signals to digital ones before they can be processed and then to convert the digital output signals back to analog and (2) because digital controls and displays are usually more expensive. In spite of the relative costs favoring analog designs when the control requirements are minimal, the trend is clearly in favor of the digital design, not only because of its flexibility, superior performance, and communications

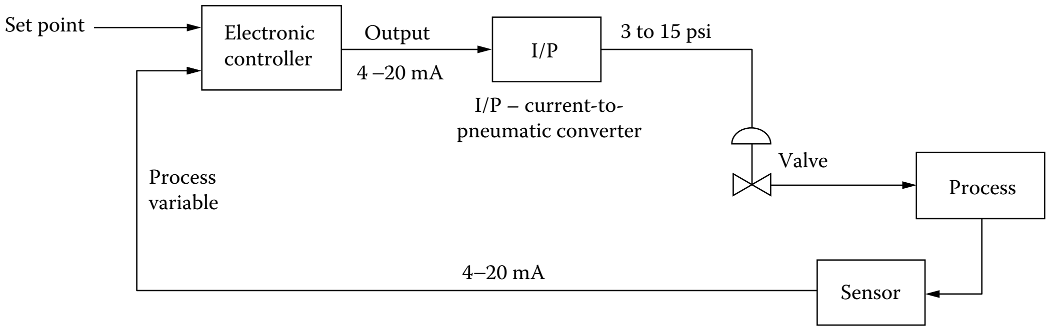


FIG. 3.2a
The main components of an electronic process control loop with a pneumatically operated control valve, which necessitates the use of a current-to-air (I/P) converter.

abilities, but also because of its relative economy in case of large control systems.

ANALOG ELECTRONIC CONTROLLERS

Today, analog electronic controllers are solid-state electronic instruments using operational amplifiers. The amplifiers have high gain, high input impedance, very low offset, and high common-mode rejection. When mounted locally, they must be protected against the environment, and they require special care in selecting the controller's components and packaging.

The majority of analog electronic controllers are used to implement the less sophisticated control algorithms (on/off, P, PI, PID, cascade, ratio, etc.) usually in single control loops that are measuring only one process variable. Sometimes, they are also used as backup (for computer failure) of critical loops in systems that are controlled by multi-loop digital controllers.

As shown in Figure 3.2a, an analog electronic controller receives the controlled process variable from a 4- to 20-mA electronic transmitter, compares that signal to its set point and generates a 4- to 20-mA output based on its control algorithm. This signal is converted into a 3- to 15-PSIG pneumatic signal if the manipulated element is an air- operated throttling valve.

Analog On/Off Switch

Figure 3.2b shows the block diagram of an analog based on/off electronic switch. (For a discussion of regular field mounted on/off switches refer to Volume 1 of this handbook and for panel- or console-mounted switches, see Chapter 4 in this volume.)

In Figure 3.2b, the 4- to 20-mA to voltage conversion is achieved by inserting a resistor into the current loop. The isolation amplifier serves to provide the separation between the transmission loop and the on/off controller electronics. The voltage switch provides the means for switching from direct to reverse action. In the direct mode, the in- and outputs move in the same direction, while in the reverse mode an increase in the controlled variable (input) results in a drop in the manipulated variable (output).

In Figure 3.2b, the hysteresis of the voltage comparator (Schmitt trigger type) provides an adjustable dead band to reduce the duty or rate cycle. Voltages at the switch inputs and comparator's output can be used to monitor set point, controlled variable, and manipulated variable values. The set point may be provided either (1) locally using a dc power supply and a voltage divider, or (2) by the output of another controller (cascade control) or (3) by a remote source.

For the manual control of the loop, the input of the voltage to a 4- to 20-mA converter must be disconnected from the comparator output and connected to the remote and adjustable dc supply (e.g., a power supply and a adjustable voltage divider).

Analog PID Controller

PID control algorithms are used on processes with multiple capacities and large and/or fast-changing loads. The output of the PID controller (E_o), which is also called the manipulated variable (m), is a function of the difference between the controlled variable and the set point or error (e). The classical

I/V – current (4–20 mA)-to-voltage converter
V/I – voltage-to-current (4–20 mA) converter

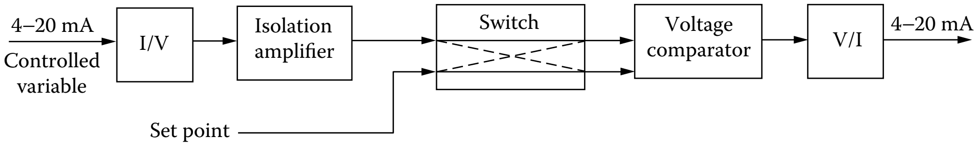


FIG. 3.2b
ON/OFF analog electronic controller.

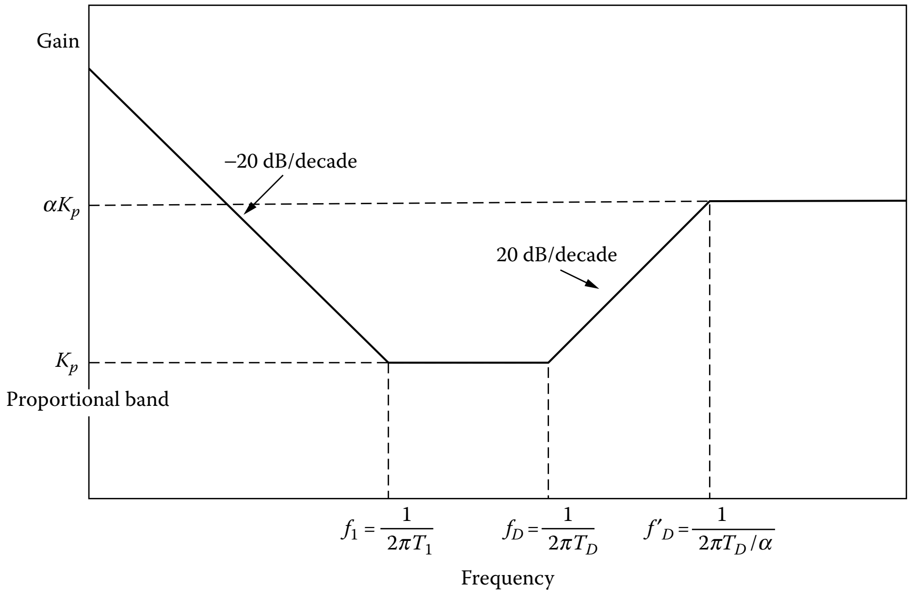


FIG. 3.2c
Bode representation of the PID controller algorithm corresponding to Equation 3.2(3).

differential equation relating the two is:

$$m = E_o = K_P e + K_D \frac{de}{dt} + K_I \int e\,dt \qquad 3.2(1)$$

where K_P is the proportional gain, K_D the derivative time, and $1/K_I$ the integral time. The transfer function $G(s)$, which relates E_o with e in the frequency domain, is obtained by Laplace transformation of Equation 3.2(1) yielding:

$$G(s) = K_P\left(1 + T_D s + \frac{1}{T_I s}\right) \qquad 3.2(2)$$

where $T_D = K_D/K_P$ and $T_I = K_P/K_I$. Most analog electronic PID algorithms are not this noninteracting variety in the time domain but rather in the frequency domain:

$$G(s) = K_P\left(\frac{1 + T_I s}{T_I s}\right)\left(\frac{1 + T_D s}{1 + (T_D/\alpha)s}\right) \qquad 3.2(3)$$

where $K_P = 100/\mathrm{PB}$, (PB is proportional band), $T_D = K_D/K_P$ is a rate (derivative) time, $T_I = K_P/K_I$ is a reset (integral) time, and α is the derivative gain, usually between 5 and 20, that ensures that the output does not change significantly in response to electrical noise (incomplete derivative controller). The Bode frequency-response of this equation is presented in Figure 3.2c.

PID Components Figure 3.2d shows one possible implementation of a PID analog electronic controller. In this figure, IA is a very-high-input impedance amplifier that amplifies the difference between the controlled (process) variable and the set point. If the controlled variable and the set point are equal, the output of the IA (the error, e) is equal to the reference voltage. The same objective can be obtained with an operational amplifier in a difference configuration, but because the voltage that is to be amplified is floating, the

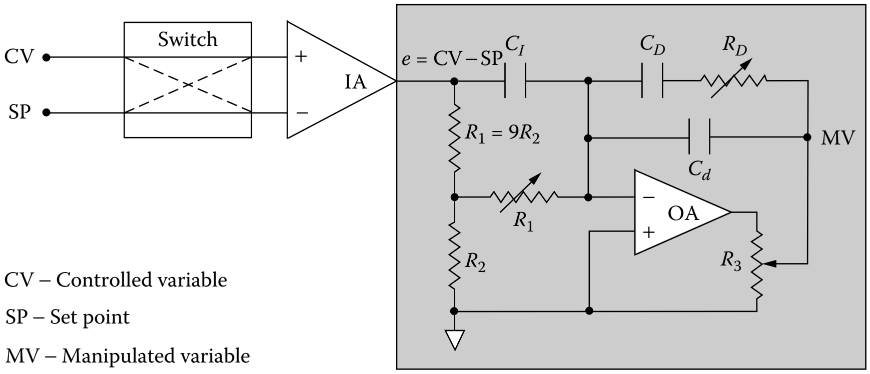


FIG. 3.2d
PID analog electronic controller circuitry including an error amplification stage.

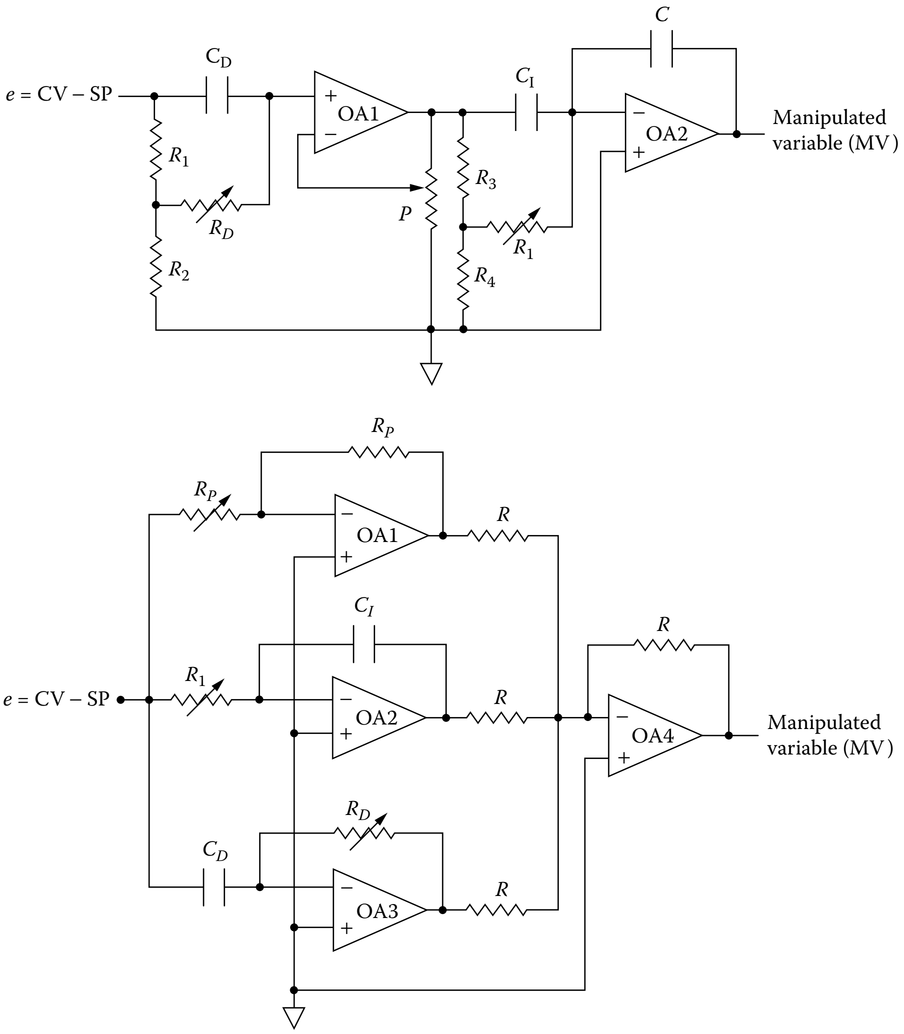


FIG. 3.2e

Operational blocks of two PID controller algorithms corresponding to Equation 3.2(3) (top) and Equation 3.2(1) (bottom).

higher common-mode rejection ratio of instrumentation amplifiers suits this application better.

The switch is in front of the instrumentation amplifier and serves to implement the direct or reverse action of the controller. As shown in Figure 3.2b but not in Figure 3.2d, a current-to-voltage converter and an isolation amplifier would complete the path of the controlled variable to the switch input.

OA is a very high-impedance operational amplifier. The transfer function of this amplifier configuration conforms to Equation 3.2(3). $R_I C_I$ define the reset time constant that determines f_I on the Bode plot of Figure 3.2c. R_1 and R_2 divide the error voltage by 10 so that R_1 is 1/10 of the value required to obtain the same time constant if this voltage divider were not used. The rate time constant of the controller is set by the combination of R_D, C_D, and C_d.

These components determine the lead or lag at f_D and f_d on the Bode diagram of Figure 3.2c. The voltage divider R_3 adjusts the feedback voltage (proportional-band adjustment), the minimum feedback voltage corresponding to the highest gain. Voltage-to-4- to 20-mA conversion, not shown in Figure 3.2d, is usually implemented by feeding the output voltage to a current amplifier.

Figure 3.2e shows two circuits that can also be used instead of the one inside the shaded box of Figure 3.2d for PID analog electronic controller implementation. The top one also implements Equation 3.2(3) while the one on the bottom implements Equation 3.2(1).

Both controllers, which are described in Figures 3.2d and 3.2e, implement PID algorithms that act on the deviation or error. As it is discussed in more detail in Chapter 2, there are several other options as a function of the controlled process. Figure 3.2f shows an algorithm implementation where the derivative mode acts on the measurement (derivative-in-front PID), which tends to speed up the return to set point in some processes.

Displays and Communications In an analog electronic controller the values of set point, controlled variable, deviation, and manipulated variable are represented by the amplitudes of slowly varying electric voltages. Indicating dc analog voltmeters

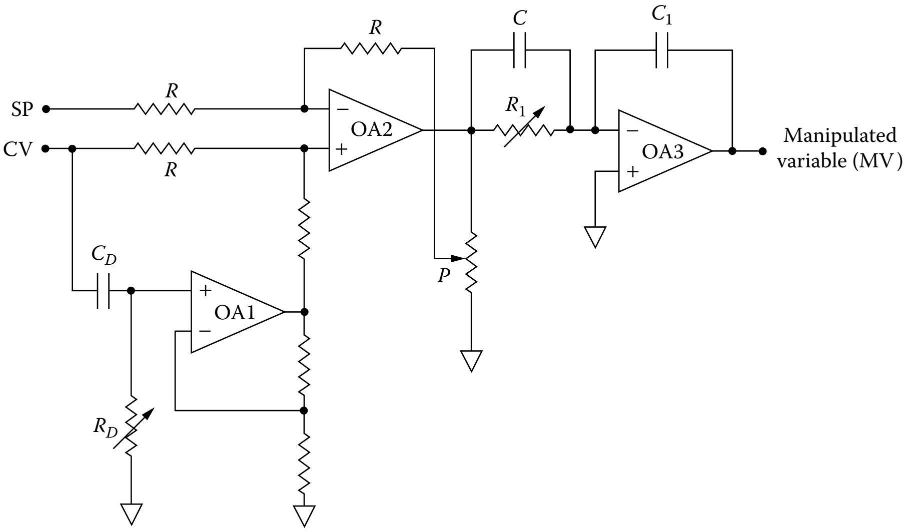


FIG. 3.2f

Operational block of a PID analog electronic controller where the derivative mode acts on the measurement instead of the error.

(e.g., D'Arsonval type) with scales graduated either in units of the process variable or more often in percentage of the span are used for monitoring these values. Abnormal conditions are signaled by voltage- or current-operated sound alarms. Even in analog controllers, the low price of numeric, alphanumeric, and bar graph displays and of their respective circuit drivers favors their use.

One feature that is common in today's analog electronic controllers is the capability for digital interfacing. Interfacing circuits and boards for serial communication, such as RS485, are quite inexpensive, available from different vendors, and fairly easy to incorporate in the controller. As already mentioned, this feature is decisive when the controller is to be integrated into a distributed or supervised control system.

Reset Windup Any controller that has an integral mode will saturate when a deviation between the controlled variable and the set point persists. This is because the reset capacitor in the controller will continue to charge so long as the error exists. This phenomenon is called integral or reset windup and is important in process phases during which the controller is switched to manual. This is the case, for example, after startup, where the controller can saturate its output before it is switched into automatic. Problems resulting from reset windup can be minimized by including circuitry that limits the output voltage of the reset amplifier.

Automatic-to-Manual Switching The manipulated variable (controller output) should stay constant while switching from the automatic to the manual mode of operation and vice versa, because any sudden change in this signal will upset (bump) the process.

One possible way to obtain bumpless transfer is to use a switch that is controlled by voltage transition and is enabled by pressing the front panel button, which commutates only if a comparator circuit detects that the output voltage of the manual operation block crosses the controller output voltage. In this design, if the output of the manual operation block changes very quickly, a bump may still occur.

To avoid this problem, slewing buttons (step-up, step-down) are used for manual adjustment. Controllers provided with automatic commutation means are sometimes called procedureless controllers.

Stability In analog electronic controllers the feedback amplifiers and feedback networks must be carefully designed to ensure that the controller is stable under all operating conditions. Drift due to component leakage is also a problem to overcome. Currents even as low as some nanoamperes can still produce significant output drift due to the extremely large time constants and very high input impedance of the amplifiers. Thus, it is important to carefully select not only the components of the operational amplifier but also the design of the circuit's layout.

DIGITAL ELECTRONIC CONTROLLERS

Computers of medium and large size are used in process control, while local digital controllers are mainly microprocessor based. Initially, microprocessor-based controllers were developed to control more than one loop, which reduced the per-loop cost. Nowadays, however, the price of microprocessors has dropped to a level that allows their usage in both multi-loop and single-loop controllers.

Other processing structures available to support the implementation of digital controllers are programmable logic devices (PLDs), namely, field-programmable gate arrays (FPGAs). Control algorithms such as PID can be implemented in FPGAs.[1] Nevertheless, FPGA-based digital process controllers are still rare because in many aspects they are outperformed by microprocessor-based controllers.

Performance

Digital controllers outperform analog controllers in several aspects. The accuracy of a digital controller is mainly dependent on the performance of the analog-to-digital (A/D) and the digital-to-analog (D/A) converters it incorporates. Thermal, drift, nonlinearity, and hysteresis errors are minimized in digital controllers, which result in errors of 0.1% or less of the span.

The sophistication and number of available control algorithms and number and topology of control loops and special functions (alarm, totalizing, display) that can be implemented with digital controllers, and the low cost or minimal additional hardware required to do so, are factors in favor of this type of controller. Digital controllers with such options are usually less expensive than the equivalent analog controllers.

Communication Capabilities

Digital techniques allow more information to be transmitted faster and with fewer errors than does analog communication. In addition, a digital controller can easily communicate with a digital computer.

The expected lifetime of local electronic controllers is typically 10 years. Integrated circuits (ICs) are the basis of electronic controllers and, due to the fast rate of development in this domain, some ICs become obsolete in a few years or even less. This is particularly true with digital components that support digital controllers. On the other hand, local controllers operate sometimes in harsh environments.

Microprocessor-Based Controllers

For many years, the name microprocessor was associated with a digital system composed of a central processor unit (CPU), memory, and input/output. The term microprocessor was introduced in 1972 by Intel to designate a large-scale integrated (LSI) circuit with the central processor unit functions of a computer implemented by Intel.

Following Intel's terminology, the microprocessor is viewed as central processor unit and the expression *microprocessor-based controller* refers to any digital electronic controller that is implemented around a CPU. This terminology applies to all digital process controllers, programmable logic controllers (PLCs), micro-controllers, and microcomputers. DCS systems are discussed in Chapter 4 and PLCs in Chapter 5 of this handbook. Here we concentrate on microcontroller—and microcomputer-based electronic controllers that can be mounted locally in the plant.

Operation

In contrast to analog controllers, which operate continuously, digital controllers implement their control algorithms in a batch mode. The CPU, which is the heart of the digital controller, operates sequentially in cycles. The cycles include the steps of process data collection and preprocessing (e.g., linearization), if necessary, calculation of the new value of the manipulated variable, and checking for alarm functions and limits before the new value of the manipulated variable is returned to the analog domain.

The displays are also updated during each cycle. This requires the calculation of engineering values of the controlled variable and the set point, the calculation of other process parameters, and the actualization of the display. This cycle of procedures repeats at rates from 0.05 to 0.5 s, typically 0.1 s. In other words, the process is evaluated and the output of the controller is updated 10 times per second.

This means that a digital controller with a single CPU is not in touch with the process except in the instants of sampling. Thus, the controller is not aware of all events that occur during the cycle. Therefore, all events that need to be continuously monitored must be latched by dedicated hardware for later processing. The communications between the controller and other instruments are handled either by using the time when the CPU is idle between the end of one cycle and the beginning of the next or on an interrupt basis.

Hardware

Microprocessor-based process controllers are data acquisition and processing systems with analog output capabilities. Figure 3.2g shows a block diagram of such a controller. It consists of an analog-to-digital conversion block, a digital-to-analog conversion block, a micro-processing block, a display and controls section, and a communications port. The micro-processing block includes not only the processing element (CPU) but also the memory. Memory used for program storage is of the read-only type (ROM); memory for temporary data storage is of the random-access type (RAM). In addition, the controller has two output ports, one of the serial type for communications and another of the parallel type with individually addressable lines.

The conditioning and analog-to-digital conversion block of the controller provides the processor with the set point and controlled variable values for the control algorithm. The conditioning part of the block provides the interfacing between the controller's input signals and the analog-to-digital conversion section and thus its design depends on both.

The remaining components include an analog multiplexer, a programmable-gain amplifier, a sample-and-hold and an analog-to-digital converter (ADC). The CPU controls these components. The ADC must be chosen taking into account the accuracy and resolutions required for the controller and the conversion rate used.

For ADCs, two good choices are the dual-slope and the successive approximation designs. The first is of the averaging type and thus is less influenced by noise while the second can be implemented using the controller's CPU and few additional components.

Sensors with frequency output (e.g., measurand range converted into a 10- to 50-kHz frequency range) are becoming more and more popular, and some microprocessor-based

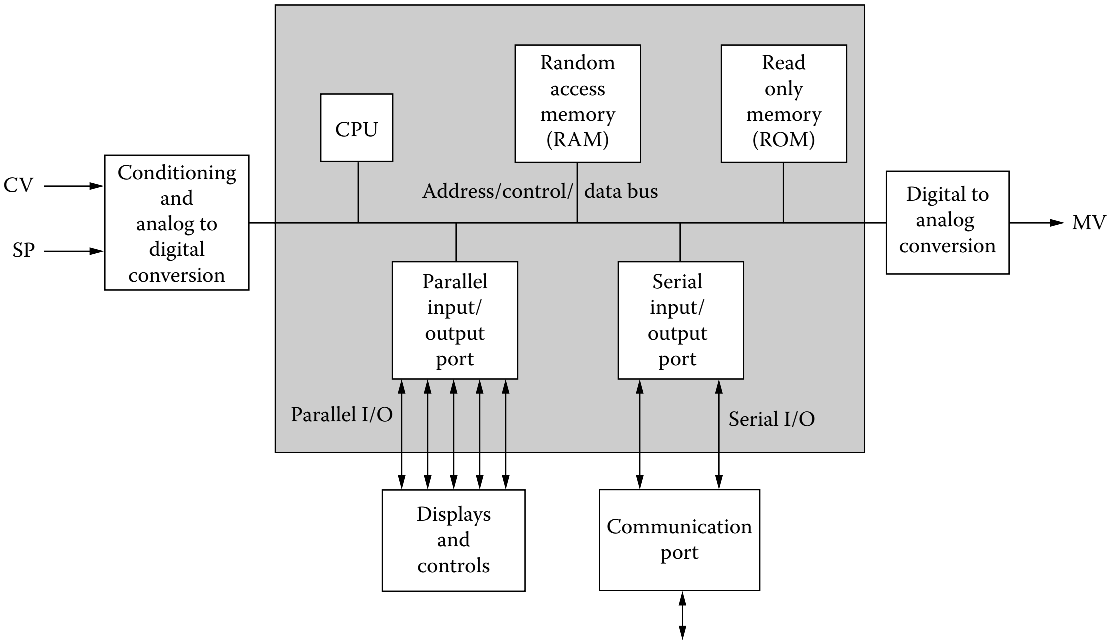


FIG. 3.2g
Typical hardware block diagram of a microprocessor-based process controller.

controllers are able to directly interface with them. In this case, the analog-to-digital conversion is easier because it depends on pulse counting, and thus the analog-to-digital conversion block is not required.

The function of the digital-to-analog conversion (DAC) block is to convert the calculated value of the manipulated variable into a dc voltage or current. DACs use an R-2R resistor ladder network whose natural output is a current. Some manufacturers also offer digital controllers with air pressure output.[2]

Displays and Communications Digital displays with either a bar graph and a digital readout that provide the variable values in engineering units are both good choices since they allow fast reading of values and also fast detection of abnormal conditions.

For adjusting digital controllers, switches, pushbuttons, keyboards, and joysticks are the natural replacements for potentiometers because they are more accurate to input data and easier to read. Figure 3.2h shows the face plates of three microprocessor-based controllers, which all have these types of displays. As a rule, displays and controls are connected to the parallel port of the processing block.

Communication between the controller and other systems is in serial formats to minimize wiring. The serial port of the processing block interfaces the processor buses, but the remaining of the transmission and communication channel can and is implemented in several different ways, some standardized others not, which makes interconnection of equipment from different manufacturers difficult if not impossible.

It is true that several manufacturers offer controllers that are compatible with standard hardware interfaces. Unfortunately, most suppliers provide only the minimum information a user needs to establish communication with other equipment. What is even worse is that in many cases the user has to configure or even has to program the equipment (e.g., computer) to code and decode the commands and data communicated by the controller.

The diversity of solutions is undesirable for users, but to date the definition and general acceptance of a single standard has been resisted by manufacturers. The existence of several fieldbuses is another example of lack of standardization. Nevertheless, some attempts have been made to standardize communications in control systems. The IEEE 1451 series of standards[3] is a good example, but it has had little success to date.

Micro-Controllers and Backup Power In the micro-processing block in Figure 3.2g, the CPU and the other components are separately shown. This corresponds to the architecture of a microcomputer. When the CPU, memory, and I/O are integrated into one circuit chip, the processor unit is called a micro-controller.

Microcomputers and micro-controllers differ also in other aspects. Microcomputers run at higher clock rates, have larger memory blocks, and are easier to program. For these reasons, microcomputers are presently better suited for process control applications. Nevertheless micro-controllers, which were initially designed to operate as embedded systems, are growing in complexity and functionality.

Some micro-controllers today can not only perform as a microcomputer but can also include, in the single chip, a variety of additional functions. These include analog-to-digital and digital-to-analog conversion, memory arrays, both of RAM

FIG. 3.2h
Front panel view of microprocessor-based controllers: temperature (upper left), general-purpose single-loop or cascade (bottom) and multi-loop (Courtesy of Foxboro Co. and Smar International.)

and of electrical alterable or erasable programmable read only memory (EAPROM, EEPROM), of reasonable sizes and advanced input/output capabilities (e.g., "high" power signal handling). Micro-controllers are inexpensive (less than $10) and thus potentially interesting as hardware support for digital process controllers.

Some processes require control modes and algorithms involving a heavy arithmetic workload and thus fast memory access or heavy data transfer. In those cases, the processing block must use special processors called digital signal processors (DSPs).

The operating parameters of a digital controller are entered by the operator and must be stored in the memory. It is good practice to provide backup power, so that the stored data are not lost and do not need to be reloaded after a power interruption (nonvolatile memory).

Possible solutions are: (1) incorporation of a long-life or rechargeable battery in the controller to maintain power to the RAM during loss of primary power, (2) use of EAPROMs or EEPROMs for parameter storage. Since the number of write/erase cycles in EAPROMs and EEPROMs is limited, the lifetime of this last solution may prove to be incompatible with the expected lifetime of the controller.

Software

The standard control algorithm for process control still is the PID algorithm. Its velocity algorithm is:

$$u_{n\Delta t} = u_{(n-1)\Delta t} + K_P \left(\begin{array}{l} (e_{n\Delta t} - e_{(n-1)\Delta t}) + \dfrac{\Delta t}{T_I} e_{n\Delta t} \\ + \dfrac{T_D}{\Delta t}(e_{n\Delta t} - 2e_{(n-1)\Delta t} + e_{(n-2)\Delta t}) \end{array} \right) \tag{3.2(4)}$$

where $u_{n\Delta t}$ is the controller's output in sampling instant $n\Delta t$ (current value); $u_{(n-1)\Delta t}$ is the controller's output value in sampling instant $(n-1)\Delta t$ (previous value); $e_{n\Delta t}$, $e_{(n-1)\Delta t}$, and $e_{(n-2)\Delta t}$ are deviations at sampling instants $n\Delta t$, $(n-1)\Delta t$ and $(n-2)\Delta t$, respectively; K_P, T_I, and T_D are the proportional gain, integral, and derivative times, respectively; and Δt is the time interval between samples.

In the velocity algorithm, integral is replaced by addition, and the simplification leads to a requirement of calculating only the change that has occurred in the error since the last sampling; the change in error is weighted by the appropriate constants.

The performance of a microprocessor-based controller naturally depends on its hardware. Nevertheless, it is the software component that ultimately dictates the controller's capabilities. The controller's cost reflects this. The same hardware may support controllers with different specifications and different prices.

General-purpose digital controllers are designed to meet the expectations of a broad spectrum of users rather than fully exploiting the capabilities of the hardware. Direct signal processor (DSP)-based controllers, for instance, can be programmed for almost anything, including linearization, compensation for valve and actuator characteristics, a variety of control modes (including batch control), and a variety of algorithms, both linear and nonlinear, self-tuning and self-optimization. Many of these features do not require the computation power of a DSP and can be easily implemented with basic processing hardware.

The suitability of a microprocessor-based controller for a specific application depends not only on the algorithms it contains, but also on their implementation and parameter adjustment capabilities. Since most controllers have no or just some extremely restricted reprogramming capabilities, the user should study their capabilities before purchasing. This is particularly the case with batch process control, where the control algorithm is usually dependent on several variables.

CONCLUSIONS

Local electronic controllers were described in this section. Some of these controller designs are similar to the ones that are located in central control rooms. Because the control room equipment is covered in Chapter 4, it is not discussed here in detail. Therefore, the reader is encouraged to also review the sections titled: Annunciators and Alarms, CRT Displays, Digital Readouts, Indicators, Analog Displays, Switches, Pushbuttons, Keyboards, and Touch Screen Display in Chapter 4.

References

1. Masmoudi, N., Samet, L., Kharaat, M. W., and Kamoun, L., "Implementation of a PID Controller on FPGA," Proceedings of ICESC 1997, Vol. 2, 550–554, Cario, Egypt, December 1997.
2. Alamo Controls, Inc., http://www.alamocontrols.com/mline_pneumatic.html.
3. "Standard for a Smart Transducer Interface for Sensors and Actuators," IEEE 1451, New York: Institute of Electrical and Electronics Engineers, 1999.

Bibliography

"Analogue Signals for Process Control Systems. Part 1: Direct Current Signals," IEC 603881-1, 1982.

"Analogue Signals for Process Control Systems. Part 2: Direct Voltage Signals," IEC 60381-2, 1978.

"Application of Safety Instrumented Systems for the Process Industries," ANSI/ISA-84.01-1996.

Astrom, K. J., and Hagglund, T., *PID Controllers: Theory, Design, and Tuning,* 2nd ed., Research Triangle Park, NC: Instrument Society of America, 1995.

Ball, S., *Analog Interfacing to Embedded Microprocessors: Real World Design,* Newnes, 2001.

Boller, A., "The Fieldbus from Production to Process Automation," INTERKAMA, 1999.

"Compatibility of Analog Signals for Electronic Industrial Process Instruments," ANSI/ISA-50.00.01-1975 (R2002).

Connel, B., *Process Instrumentation Applications Manual,* New York: McGraw-Hill, 1996.

"Controllers with Analogue Signals for Use in Industrial-Process Control Systems. Part 1: Methods of Evaluating the Performance," IEC 60546-1, 1987.

"Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems," IEC-61508 series, Parts 1 to 7.

Galovic, Z., "Advantages of Using Embedded Controllers on the Factory Floor," ETCON 2001.

Harris, T. J., Seppala, C. T., and Desborough, L. D., "A Review of Performance Monitoring and Assessment Techniques for Univariate and Multivariate Control Systems," *Journal of Process Control,* 9, 1–17, 1999.

"Function Blocks for Industrial-Process Measurement and Control Systems—Part 1: Architecture," IEC/PAS 61499-1, 2000.

"Function Blocks for Industrial-Process Measurement and Control Systems—Part 2: Software Tools Requirements," IEC/PAS 61499-2, 2001.

"Function Blocks for Industrial-Process Measurement and Control Systems—Part 4: Rules for Compliance Profiles," IEC 614999-4, 2002.

Hintz, K. J., and Tabak, D., *Architecture, Implementation, and Programming,* New York: McGraw-Hill, 1992.

"Industrial-Process Measurement and Control—Documentation of Application Software," IEC 61506, 1997.

Newman, H. M., *Direct Digital Control of Building Systems: Theory and Practice*, New York: Wiley Interscience, 1994.

Pal, J. K., and Tzafestas, S. G., *Real Time Microcomputer Control of Industrial Processes,* Boston: Kluwer Academic Publishers, 2002.

Pessen, D. W., *Industrial Automation: Circuit Design and Components,* New York: John Wiley & Sons, 1989.

Predko, M. M., *Handbook of Microcontrollers,* New York: McGraw-Hill, 1999.

Sadhukhan, D., and Dever, T. M., "The Application of Dedicated Controllers to Parallel/Serial Exhauster Trains," ISA EXPO 1998.

Senbon, T., and Hanabuchi, F. (Ed.), *Instrumentation Systems—Fundamentals and Applications,* Berlin: Springer-Verlag, 1987.

Shinskey, F. G., *Process Control Systems: Application, Design, and Tuning,* New York: McGraw-Hill, 1996.

St. Clair, D. W., "The PID Algorithm," members.aol.com/pidcontrol/pid_algorithm.html.

Tocci, R. J., and Ambrosio, F. J., *Microprocessors and Microcomputers: Hardware and Software,* Upper Saddle River, NJ: Prentice Hall, 2003.

3.3 Converters and Dampeners

A. F. MARKS (1970, 1985) **B. G. LIPTÁK** (1995, 2005)

Signal Ranges:	Pneumatic: 3 to 15, 3 to 27, 6 to 30 PSIG (0.2 to 1.0, 0.2 to 1.8, 0.4 to 2.0 bar) Voltage: 0 to 1 millivolts DC to 1 to 5 volts DC or 0 to 5 volts AC Current: 1 to 5, 4 to 20, 10 to 50 milliamperes DC
Inaccuracy:	Generally 1/10 to 1/2% of full scale
Cost:	Most standard units at around $300 with special designs up to $2000
Partial List of Suppliers:	ABB Automation Inc. (www.abb.com/processautomation) Acromag Inc. (www.acromag.com) Action Instruments (www.actionio.com) Air Monitor Corp. (www.airmonitor.com) Ametek U.S. Gauge Div. (www.ametekusg.com) Devar Inc. (www.devarinc.com) Dresser Instrument (www.ashcroft.com) Emerson Process Management (www.easydeltav.com) Endress + Hauser (www.systems.endress.com) Fairchild Industrial Prod Co. (www.fairchildproducts.com) Flowserve – Valtek Control Products (www.flowserve.com) Honeywell Sensing and Control (www.honeywell.com/sensing) Invensys Process Systems (www.invensysips.com) ITT Conoflow (www.ittconoflow.com) Moore Industries International Inc. (www.miinet.com) Omega Engineering Inc. (www.omega.com) Robertshaw (www.robertshawindustrial.com) Rosemount Inc. (www.rosemount.com) Scanivalva Corp. (www.scanivalve.com) Siemens Energy & Automation Inc. (www.sea.siemens.com/ia) Thermo Electric Inc. (www.thermo-electric-direct.com) Triad Controls Inc. (www.triadcontrols.com) United Electric Controls (www.ueonline.com)

INTRODUCTION

A wide variety of converters are used in process control. This section provides only a partial overview of these devices. For more details on the individual categories, it is recommended to review the specific sections that deal with them.

In Volume 1 (*Measurement*) of this handbook, the TC and RTD converters are discussed in Chapter 4 and the various pressure repeaters and pressure scanners are covered in Chapter 5. In this volume (*Process Control*), the current-to-air (I/P) converters are also discussed under control valves in Chapter 6, and the DCS-related analog-to-digital (A/D) or digital-to-analog (D/A) converters are covered in Chapter 4.

When interfacing different digital devices, they too require conversion, as the digital information can be coded differently and the communications can follow different formats. These topics are also addressed in Chapter 4 of this volume and also in Volume 3, which is devoted to software.

PNEUMATIC CONVERTERS AND DAMPENERS

In the sense that they convert a process pressure into a standard output air signal, all pneumatic pressure transmitters

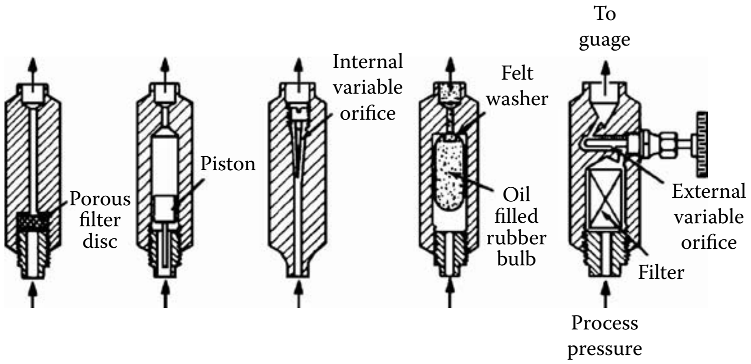


FIG. 3.3a
Pulsation dampener design variations.

(covered in Chapter 5 of Volume 1) are converters. Similarly, the various pneumatic pressure boosters and multiplying and reversing relays are also converters and are discussed in Chapter 6 of this volume.

Pulsation Dampeners

Pressure systems, both pneumatic and liquid-filled ones, transmit process noise very rapidly. A good example is the output of a reciprocating compressor or positive displacement pump. In order to obtain precise measurements when the process pressure is pulsating, it is necessary to dampen the pressure pulses by restricting the flow passages. Some of the various types of snubbers and pulsation dampeners are illustrated in Figure 3.3a.

One design consists of a fitting with a corrosion-resistant porous metal filter disc. By the use of such a device, the equilibrium reading on the outlet is delayed by about 10 seconds. Another snubber design depends for its damping action on a small piston in the inlet fitting, which rises and falls with pressure impulses and thereby absorbs shock and surge. Still another snubber design uses the adjustable restriction created by a micro-valve in the inlet fitting to damp pulsation. This differs from a needle valve in two important ways: the filler consists of stainless steel wool to prevent plugging, and it does not quite shut off. This is to prevent “shutting-in” a false reading.

Other forms of pulsation or noise dampening can be obtained by using dashpots in motion balance devices or by restricting the flow between bellows in fluid-filled devices. For further information on snubbers and pulsation dampeners refer to Chapter 5 in Volume 1 (*Measurement*) of this handbook.

DIGITAL CONVERTERS

Analog-to-digital and digital-to-analog converters are covered in Chapter 4 of this volume, where I/O devices are also discussed. Analog–to–on/off and digital–to–on/off logic devices, such as monitor switches, relays, and counters, are also discussed in Chapter 5.

ANALOG CONVERTERS

Pneumatic-to-Electronic Converters

The pneumatic-to-electronic transducer is used wherever pneumatic signals must be converted to electronic signals for any of the following reasons:

1. Connecting existing pneumatic plants to computers
2. Transmission over large distances
3. Input to an electronic logger
4. Input to telemetering equipment
5. Instrument air not available at the receiver controller

In principle, any of the electronic pressure transmitters could be used, but in practice, special, improved-accuracy devices are used. The air signals are at low pressure levels (3 to 15 PSIG or 0.2 to 1.0 bar), and many of the standard pressure detectors are not sensitive or not linear enough at these pressures.

A P/I transducer should be at least 1/2% of full-scale accurate and preferably 1/4% to preserve the integrity of the initial signal. Since the total error is the square root of the mean squares of the individual component errors, the greater the precision of the P/I transducer the better the signal.

Because of this need for higher accuracy, most P/I transducers use a bellows input and a motion balance sensor. A typical high-quality P/E converter is shown in Figure 3.3b.

Pneumatic-to-electronic converters are also discussed in Chapter 4 of this volume and in Chapter 5 of Volume 1 (*Measurement*).

Current-to-Air Converters

Electropneumatic transmitters are also referred to as converters and transducers. They are important because they form

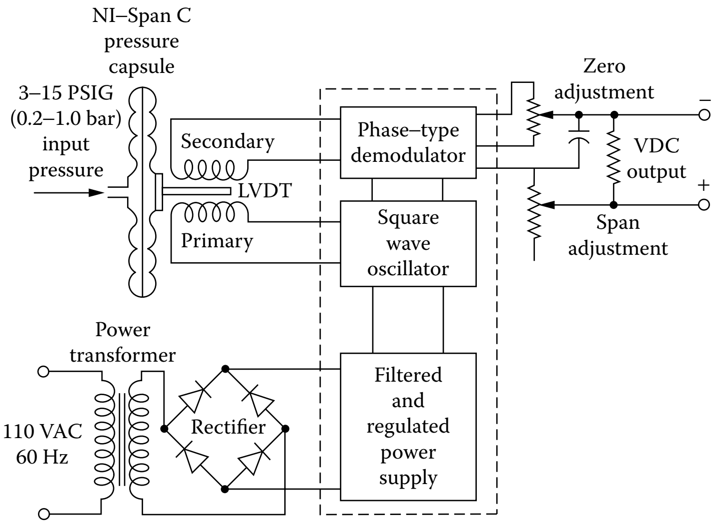


FIG. 3.3b
Pneumatic-to-electronic converter.

the link between electrical measurements in a pneumatic control system. They also convert electronic controller outputs into air pressure signals for the operation of pneumatic control valves. (These devices are also discussed in Chapter 6.)

Figure 3.3c illustrates one of these converter designs and also lists the various electric devices with which it can be combined. The input is usually a DC current in the range of 1 to 5, 4 to 20, or 10 to 50 milliamperes. An Alnico permanent magnet creates a field that passes through the steel body of the transmitter and across a small air gap to the pole piece. A multi-turn, flexure-mounted voice coil is suspended in the air gap. The input current flows through the coil creating an electromagnetic force that tends to repel the coil and thus converts the current signal into a mechanical force.

Since the total force obtainable in a typical voice coil motor with such small current inputs is only on the order of some ounces, a different approach, namely, the use of a reaction nozzle, is employed to convert the force into a pneumatic output

Types of inputs
1. Thermocouple amplifier
2. Resistance bulb amplifier
3. Speed detector amplifier
4. Radiation detector amplifier
5. pH amplifier
6. AC current or power converter
7. Magnetic flowmeter amplifier
8. Electronic controller output

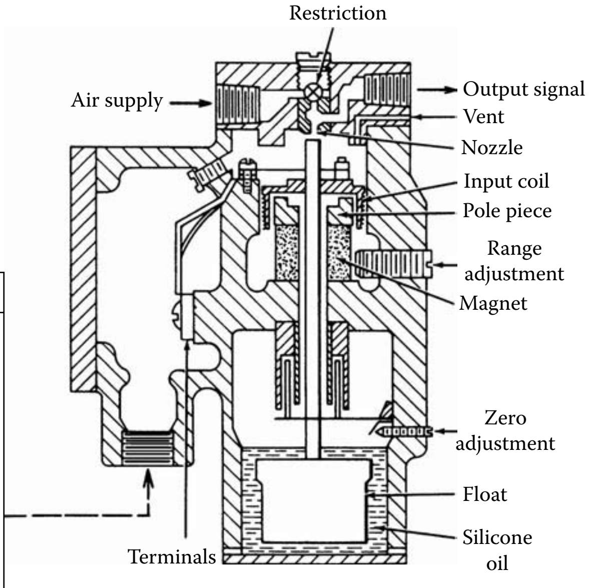


FIG. 3.3c
Electric-to-pneumatic transducers or transmitters can have a variety of input sources.

pressure. In this circuit, the supply air flows through a restriction and out the detector nozzle. The reaction to the air jet as it impinges against the nozzle seat supplies the counterbalancing force to the voice coil motor. The nozzle backpressure is the transmitted output pressure.

To make the transmitter insensitive to vibration, the voice coil is integrally mounted to a float submerged in silicone oil. The float is sized so that its buoyant force equals the weight of the assembly, reducing the net force to zero.

The zero of the transmitter is adjusted by changing the force on a leaf-spring. Span is adjusted by turning the range-adjusting screw, which changes the gap between the screw and the magnet, thus shunting some of the magnetic field away from the pole piece.

Millivolt-to-Current Converters

Millivolt-to-current converters are widely used in the measurement of temperature, using thermocouples or other millivolt-generating sensing elements. They are also utilized in converting the output signals of analyzers into higher-level transmission signals. A typical millivolt-to-current converter is illustrated in block diagram form in Figure 3.3d.

When these devices are used to convert thermocouple signals, they are also referred to as temperature transmitters (See Chapter 4 in Volume 1). In the mid-1960s several companies developed miniaturized converters for local mounting in the thermocouple head, which is mounted directly on the thermowell (Figure 3.3e). These devices have been discussed in more detail in Section 4.13 of the Measurement volume of this handbook.

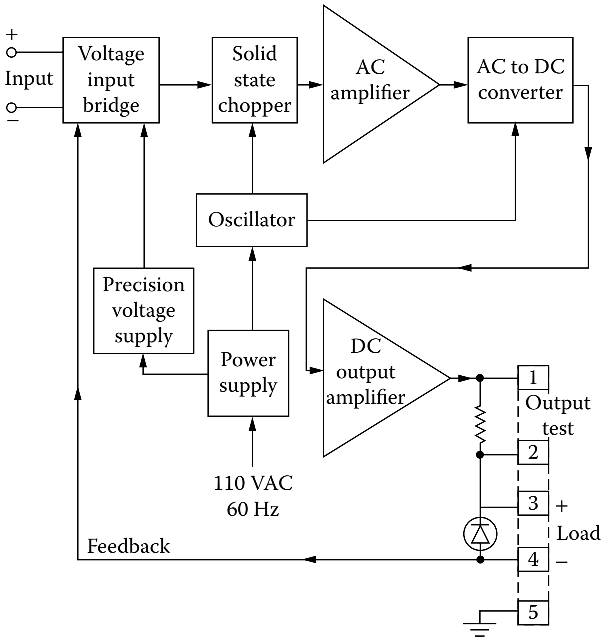


FIG. 3.3d
The block diagram of a millivolt-to-current converter.

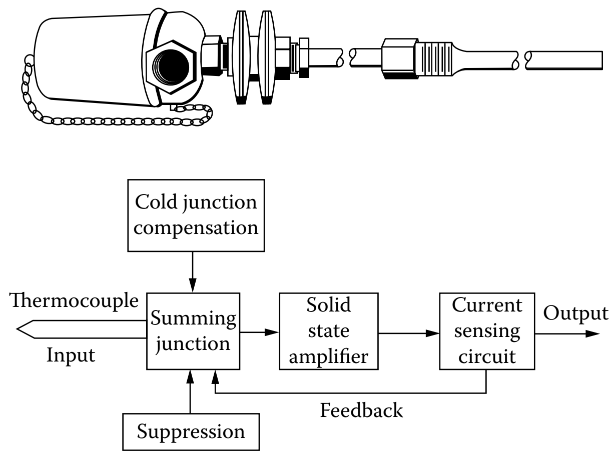


FIG. 3.3e
Thermocouple-to-current converter, which is mounted directly on the thermowell.

Voltage-to-Current Converters

Converters are also available for the conversion of higher than millivolt voltages into transmission signals (Figure 3.3f). These usually consist of voltage dividers (and rectifiers if necessary) to reduce voltages to a level compatible with the receivers.

Current-to-Current Converters

Current-to-current transducers are available to convert AC signals to DC (or DC to AC). They also serve to amplify or reduce the current levels as necessary. They are sometimes used in the power industry, but not too widely because other methods are capable of receiving and handling higher current levels.

Levels of alternating current can be changed, when necessary, by current transformers. A common signal level in the power industry is 0 to 5 amperes. Direct currents are re-ranged by putting a series resistor (low ohms) in the circuit and reading the voltage drop across it.

A milliampere converter for AC to DC is shown in Figure 3.3g. This Figure shows three separate devices: a current transformer, an AC-to-DC milliampere converter, and a current-to-air converter. The function of the transformer is to scale down the current to the range normally used for direct AC metering. The AC/DC converter makes this signal compatible with the

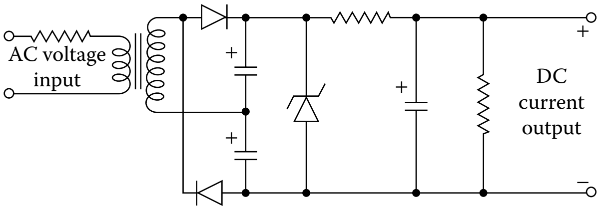


FIG. 3.3f
Solid-state voltage-to-current converter.

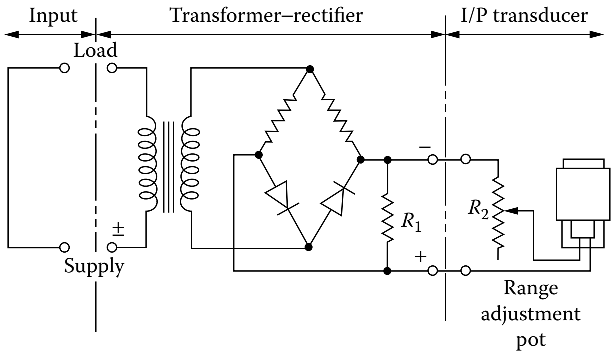


FIG. 3.3g
AC-to-DC milliampere converter with integral I/P transducer.

usual DC milliampere transmission systems. DC to DC "converters" are sometimes used for isolation of electrical circuits, such as in intrinsically safe systems. In such installations they are usually called current repeaters or barrier repeaters.

Resistance-to-Current Converters

Resistance measurements are common in temperature measurements and also in resistance or strain gauge sensors. The circuits used are similar to those of millivolt-to-current converters except that the front end is a resistance bridge instead of a voltage bridge (Figure 3.3h).

The strain gauge bridge, a special form of resistance element, is described in Sections 5.7 and 7.25 of the *Process Measurement* volume of this handbook. The strain gauge elements may take the place of two of the resistors in the resistance bridge shown in Figure 3.3h.

ELECTRONIC NOISE REJECTION

It is a general consensus that the analog electronic signal leads should be twisted pairs, and that they should be run in metallic conduit or be shielded. Less generally agreed upon is a recommendation that these conduits or shielded bundles of wiring be isolated from high-power wiring to motors, but this too is followed in most cases.

DC signals of 4 to 20 or 10 to 50 milliamperes can be run in the same conduit with telephone wires, low-voltage DC signals, and/or thermocouple leads. They should not be run in the same conduit or in the same shielded bundle with alarm signals or power wiring.

AC signals (which are much less common) should be run in twisted pairs, with each pair shielded, and then in conduit or shielded bundles. Some common-mode noise rejection is built into the instruments, but this is not sufficient if computers are being used or if higher-level spikes from motor loads are experienced.

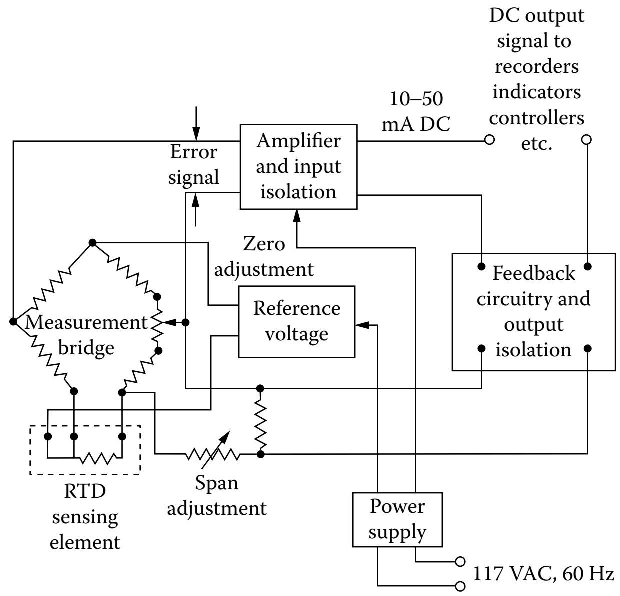


FIG. 3.3h
Resistance-to-current converter.

The various methods of conditioning digital signals, and the subject of computer-compatible wiring practices in general, are covered in Section 1.8 of this volume.

Bibliography

Dyer, S. A., *Wiley Survey of Instrumentation and Measurement,* New York: Wiley, 2001.

Farmer, E., "Pneumatics in a Digital World," *Instruments and Control Systems,* March 1979.

Holben, E. F., "A Digital-to-Pneumatic Transducer," *Instruments and Control Systems*, January 1970.

ISA, "Industrial Measurement Series: Pressure" (video), 2002.

Korn, G., *Interactive System Simulation,* New York: McGraw-Hill, 1989.

Luyben, W. L., et. al., *Plantwide Process Control,* New York, McGraw-Hill, 1998.

McNeil, E. B., "A User's View of Instrumentation," *Chemical Processing*, October 1982.

Nachtigal, C. N., *Instrumentation and Control*, New York: Wiley, 1990.

NUREG/CR-6633, "Advanced Information System Design," USNRC, March 2000.

Oda, S., et al., "Communications between the DCS and Sub-Systems," 1992 ISA Conference, Houston, October 1992.

Pinkowitz, D. C., "A Designer's Guide to Deglitching DACs," *Instruments and Control Systems,* April 1983.

Sheridan, T. B., "Interface Needs for Coming Industrial Controls," *Instrumentation Technology,* February 1979.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

3.4 Relays for Computing and Programmers

C. L. MAMZIC (1970, 1985) **R. GILBERT** (1995) **B. G. LIPTÁK** (2005)

Large-Case, Cam-Actuated Time Function Generators (Programmers)

Cost:	$1200 to $2500
Partial List of Suppliers:	ABB Automation Inc. (www.abb.com/processautomation) Bristol Babcock Inc. (www.controlwave.com) Emerson Process Management (www.easyddeltav.com) Honeywell Sensing and Control (www.honeywell.com/sensing) Invensys Process Systems (www.invensysips.com) Siemens Energy & Automation (www.sea.siemens.com/ia)

Large-Case, Adjustable Range and Hold Programmers

Cost:	$2200 to $3500
Partial List of Suppliers:	ABB Automation Inc. (www.abb.com/processautomation) Invensys Process Systems (www.invensysips.com) Texas Analytical Controls Inc. (www.tac-controls.com)

Miniature and Large-Case Pneumatic Profile Tracers (Programmers)

Inaccuracy:	± 0.25% of full scale
Cost:	$1800 to $4000
Partial List of Suppliers:	Aro Corp. (http://www.hydraulic-supply.com/html/productline/mfgprod/aro-corp.htm) Gaston County Dyeing Machine Co. (http://www.gaston-county.com/) Partlow Corp. (http://partlow.ttistore.com/) Pneucon, Inc. (http://www.pneucon.net/) Siemens Energy & Automation Inc. (www.sea.siemens.com)

Electric Line and Edge Follower Programmers

Inaccuracy:	± 0.25% of full scale
Cost:	$1800 to $4000
Partial List of Suppliers:	Emerson Process Management (www.easydeltav.com) Honeywell Sensing and Control (www.honeywell.com/sensing) Jumo Process Control Inc. (www.jumousa.com) Moeller Electric Corp. (http://www.moellerusa.net/) MK Juchheim GmbH. (http://www.jumo.de/)

Timing and Programming Equipment

Cost: $500 to $5000

Partial List of Suppliers:
ABB Automation Inc. (www.abb.com/processautomation)
Aromat Corp. (www.aromat.com)
Automatic Timing & Controls Co. (www.automatictiming.com)
Crouzet Corp. (www.crouzet.com)
Danaher Controls (www.dancon.com)
Honeywell Sensing and Control (www.honeywell.com/sensing)
Invensys Process Systems (www.invensysips.com)
Johnson Controls Inc. (www.johnsoncontrols.com)
Jumo Process Control Inc. (www.jumousa.com)
Kessler-Ellis Products Co. (www.kep.com)
Love Controls Corp. (www.lovecontrols.com)
Newport Electronics Inc. (www.newportus.com)
Omron Electronics Inc. (www.omron.com/oei)
Red Lion Controls (www.redlion-controls.com)
Square D/Schneider Electric Co (www.squared.com)

Drum Programmers for On-Off Event Sequencing

Cost: $2500 to $8000

Partial List of Suppliers:
Barber-Coleman Co. (http://www.barber-colman.com/)
Honeywell Sensing and Control (www.honeywell.com/sensing)
Moore Controls SA Pty Ltd. (www.moore.co.za)
Omron Electronics Inc. (www.omron.com/oei)
Siemens Energy & Automation Inc. (www.sea.siemens.com)
Texas Instruments Far East Inc. (www.ti.com)

Computing Relays

Inaccuracy: ±1/2% for all types except the differentiating and integrating relays, which are ±15 % if uncompensated and ±1% if compensated in specially built units

Cost: Prices shown are for pneumatic (and electronic) devices:
High and low selectors $80 to $180 ($125 to $300)
Adding and subtracting relays $100 to $200 ($250 to $500)
Square root extractors and function generators $400 to $800 ($250 to $500)
Scaling and proportioning relays $500 to $1000 ($250 to $500)
Multiplying and dividing relays $600 to $1200 ($400 to $600)

Partial List of Suppliers:
ABB Automation Inc. (www.abb.com/processautomation)
Acromag Inc. (www.acromag.com)
Devar Inc. (www.devarinc.com)
Dymec Inc. (www.dymec.com)
Emerson Process Management (www.easydeltav.com)
Fairchild Industrial Products Co. (www.fairchildproducts.com)
Invensys Process Systems (www.invensysips.com)
Johnson Controls Inc. (www.johnsoncontrols.com)
Moore Industries International Inc. (www.miinet.com)
MTL Inc. (www.mtl-inst.com)
Robertshaw (www.robertshawindustrial.com)
Rochester Power Instruments (www.rochester.com)
Ronan Engineering Co. (www.ronan.com)
Transmation Products Group (www.transmation.com)
United Electric Controls (www.ueonline.com)
Wilkerson Instrument Co. Inc. (www.wici.com)

INTRODUCTION

In our digital age, most computing and timing functions are done in software. Volume 3 of this handbook is devoted to the description of such and many other software packages. The topics discussed in this section are from the analog age when mathematical and time function generators were still dedicated hardware components. Computing functions were performed by pneumatic and analog electronic relays, while timing and sequencing was controlled by programmers. Some of the common relay functions and their symbols are listed in Table 3.4a.

From the controller's viewpoint automation is, in effect, a set of computing and timing elements that carry out the mathematical operations required to implement a control scheme. Summation, subtraction, multiplication, division, raising to a power, extracting the root, integration, and differentiation are important examples of such computing operations. Popular timing functions include lag elements, time function generators, ramp and hold programmers, event sequencers, and profile programmers. Both timing and computing functions are still available as stand-alone products or in combinations that add up to a complete controller.

MATHEMATICAL FUNCTIONS

Most control systems require that the controller perform some sort of mathematical calculation before a control action is initiated. These calculations involve algebra as well as integral and differential calculus. The popular computing elements perform summations, subtractions, multiplication, division, raising to a power, extracting roots, and specific types of differentiation and integration. In the discussion that follows, these analog computing relays are grouped into pneumatic and electronic categories.

PNEUMATIC RELAYS

In some existing plants, pneumatic computing relays are still used because of their simplicity, reliability, and safety advantages when used in processes with high fire and explosion potentials.

Multiplying and Dividing Relays

In the force bridge multiplier-divider shown in Figure 3.4b, input pressures act on bellows in chambers *A*, *B*, and *D*. The output is a feedback pressure in chamber *C*. The bridge consists of two weigh-beams that pivot on a common movable fulcrum, with each beam operating a separate feedback loop. Any unbalance in moments on the left-hand beam causes a movement of the fulcrum position until a moment-balance is restored. An unbalance in moments on the right-hand beam results in a change in output pressure until balance is restored.

TABLE 3.4a
*Relay Functions and Their Symbols**

Symbol	*Function*	
1-0 or ON-OFF	Automatically connect, disconnect, or transfer one or more circuits, provided that this is not the first such device in a loop. (See Note 13 in Table 1.4b)	
Σ or ADD	Add or totalize (add and subtract), with two or more inputs.	
Δ or DIFF	Subtract (with two or more inputs)	
±	Bias (single input)	
AVG.	Average	
% or 1 : 3 or 2 : 1 (typical)	Gain or attenuate (input:output), with single input	
	Multiple (two or more inputs)	
÷	Divide (two or more inputs)	
$\sqrt{\ }$ or SQ. RT.	Extract square root	
x^n or $x^{1/n}$	Raise to power	
f(*x*)	Characterize	
1 : 1	Boost	
> or HIGHEST (Measured Variable)	High-select. Select highest (higher) measured variable (not signal, unless so noted).	
< or LOWEST (Measured Variable)	Low-select. Select lowest (lower) measured variable (not signal, unless so noted).	
REV.	Reverse	
	Convert	
a. E/P or P/I (typical)	For input/output sequences of the following:	
	Designation	Signal
	E	Voltage
	H	Hydraulic
	I	Current (electrical)
	O	Electromagnetic or sonic
	P	Pneumatic
	R	Resistance (electrical)
b. A/D or D/A	For input/output sequences of the following:	
	A	Analog
	D	Digital
∫	Integrate (time integral)	
D or *d*/*dt*	Derivative or rate	
I/D	Inverse derivative	

Note: The use of a box enclosing a symbol is optional. The box is intended to avoid confusion by setting off the symbol from other markings on a diagram.

*Permission by ISA to abstract from its standard ANSI/ISA S5.1 is gratefully acknowledged.

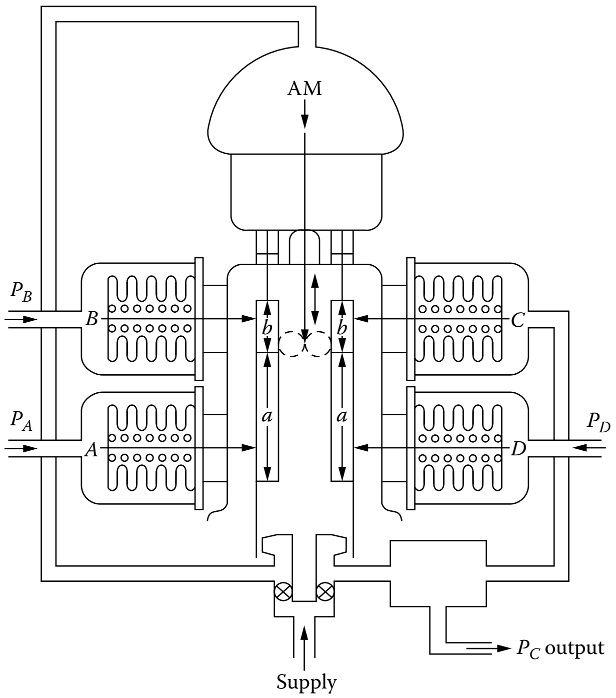


FIG. 3.4b
Multiplying and dividing relay, utilizing a pneumatic force bridge.

The equations that characterize the force bridge operation with respect to each subunit diaphragm in the unit shown in Figure 3.4b can be conveniently reduced to:

$$C = \frac{B \times D}{A} \qquad \textbf{3.4(1)}$$

Multiplication results when the two input variables are connected to chambers *B* and *D*. Division results when the dividend is connected to either chambers *B* or *D*, with the divisor connected to *A*. Simultaneous multiplication and division results when *B*, *D*, and *A* chambers are used.

The significant advantage of a cam-actuated multiplying and dividing relay is that it can operate with practically any type of nonlinear function that can be cut on a cam. These can include logarithmic functions, as in pH measurement, and computation in narrow, suppressed ranges of measurement, which results in good resolution.

The pure multiplier-divider in Figure 3.4c, when used for temperature and pressure compensation, for example, uses input signals proportional to the total absolute temperature and pressure range, starting with zero. Since the usable temperature and pressure range might be a small percentage of the total measurement range, the results might lack precision.

In Figure 3.4c, input pressure P_1 and output P_0 act on double-diaphragm capsules, and the net resultant force in each is in the direction of the larger area diaphragm. Input P_1 creates force *Y*, which pulls the baffle, pivoted at *A*, away from the nozzle. Output pressure P_0 creates force *X*, which moves the baffle closer to the nozzle. The θ input/output relationship is a function of the angle of the nozzle beam.

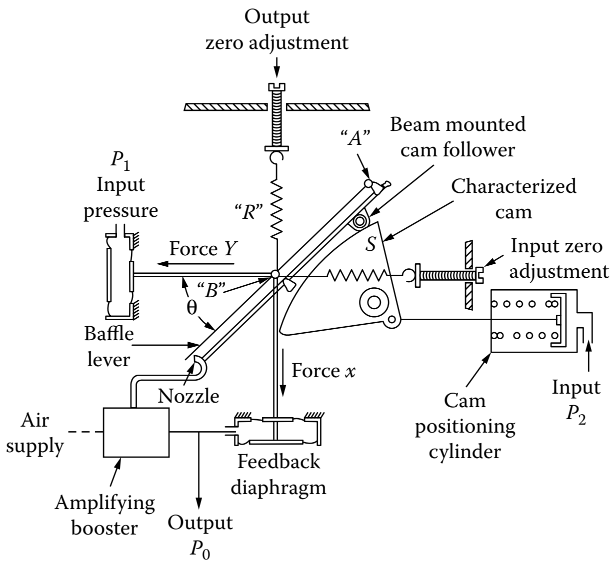


FIG. 3.4c
Multiplying and dividing relay characterized by a pneumatic cam.

When angle θ is 45 degrees, the relationship is 1 : 1. This can also be considered the multiplication factor or gain, *K*. At larger angles, *K* is greater than 1, and at smaller angles, smaller than 1. The multiplicand P_2 acts on the cam-positioning cylinder and thereby changes the nozzle beam angle in accordance with the cam characteristic. The zero adjusting springs subtract the 3-PSI (0.2-bar) zero from P_1 and set a 3-PSI (0.2-bar) zero on the output, respectively.

Adding, Subtracting, and Inverting Relays

In the force balance arithmetic computing relay, Figure 3.4d, a signal pressure in chamber *A* acts downward on a diaphragm with unit effective area. A signal in chamber *B* also acts downward on an annular diaphragm configuration, likewise having an effective area of unity. Signal pressures in chambers *C* and *D* similarly act upward on unit effective diaphragm areas. Any unbalance in forces moves the diaphragm assembly with its integral nozzle seat. The change in nozzle seat clearance changes the nozzle backpressure and hence, changes the output pressure, which is fed back into chamber *D* until force balance is restored. The following equation describes the operation of the relay:

$$T = A + B - C \pm K \qquad \textbf{3.4(2)}$$

K is the spring constant. It is adjustable to give an equivalent bias of ±15 PSI (1.24 bar).

The relay in Figure 3.4e is a modification of Figure 3.4d in that it incorporates additional input chambers and output feedback chambers. It can be used to add and/or average up to nine inputs. Figure 3.4e is an averaging relay for five inputs. The averaging feature keeps all signals in the same

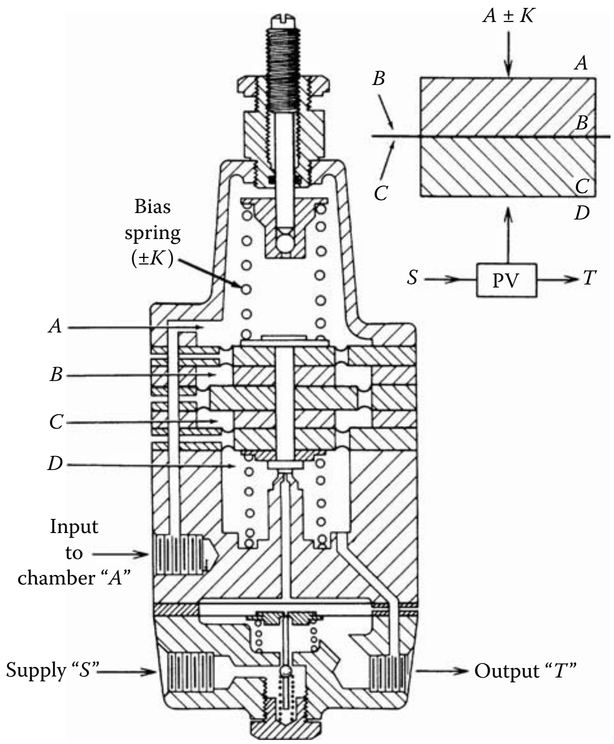


FIG. 3.4d
Pneumatic, adding, subtracting, inverting, and biasing relay.

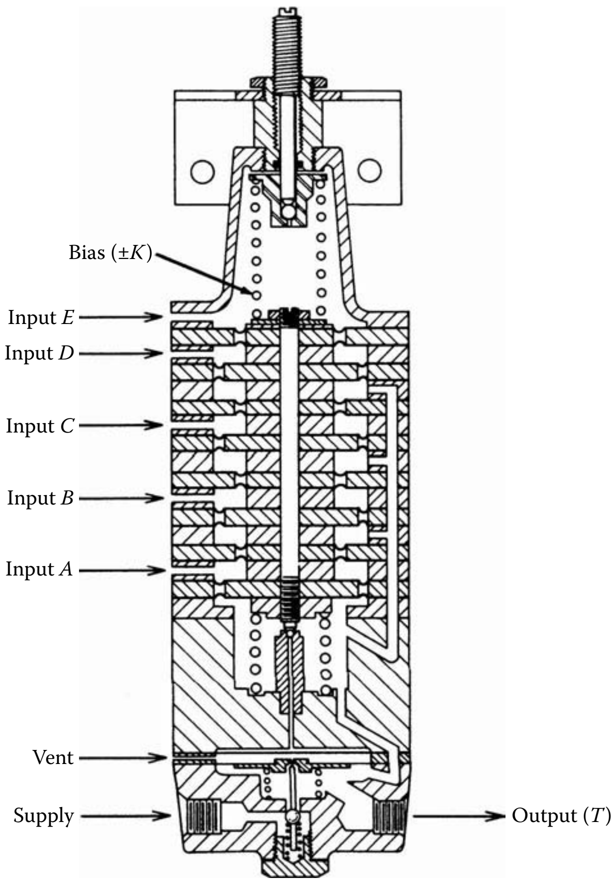


FIG. 3.4e
Multi-input averaging relay.

standard 3- to 15-PSIG (0.2- to 1.0-bar) range. Its characteristic equation is:

$$T = \frac{A + B + C + D + E}{5} \pm K \qquad \textbf{3.4(3)}$$

The relay shown in Figure 3.4d provides inverting or reversing action by setting the loading of the bias spring to a maximum and by connecting the input to subtracting chamber C. If the bias is set at +18 PSIG (1.2 bar), then a 3- to 15-PSIG (0.2- to 1.0-bar) signal in chamber C results in a 15- to 3-PSIG (1.0- to 0.2-bar) output. The equation describing this operation is:

$$T = K - C \qquad \textbf{3.4(4)}$$

Differentiating Relay

A differentiating relay produces an output proportional to the rate of change of the input. Figure 3.4f shows an ideal pneumatic differentiating relay. The relay is basically similar in construction to the relay shown in Figure 3.4d, except that the annular effective diaphragm area between chambers *B* and *C* is more than ten times the effective area of the small diaphragms between chambers *A* and *B*—giving a gain of greater than ten. The input signal is transmitted unrestricted to chamber *B* and passes to chamber *C* through an adjustable restriction. When the input is steady, the forces resulting from pressures in *B* and *C* chambers cancel each other, so that the output equals the zero-spring setting (usually mid-scale, 9 PSIG [0.6 bar] if both positive and negative rates are to be measured).

If the input pressure changes, a differential develops across the restriction. The relay transmits an output proportional to this differential. For accurate results, this differential must be directly proportional to the rate of change of input. Using a needle valve that produces laminar flow provides a linearly proportional volumetric flow, but the differential developed across the needle valve is a function of the mass flow, which varies with static pressure because of compressibility. This compressibility error is approximately ±15%. The effect can be fully compensated, however, by the addition of a variable volume to chamber *C*, the restricted chamber. As the static pressure increases, tending to make the differential smaller because of higher mass flow rate, the volume increases proportionately to maintain a constant differential. The needle valve setting determines the rate time constant.

The compensated relay is not a standard piece of hardware. In most cases, a noncompensated differentiating relay is satisfactory.

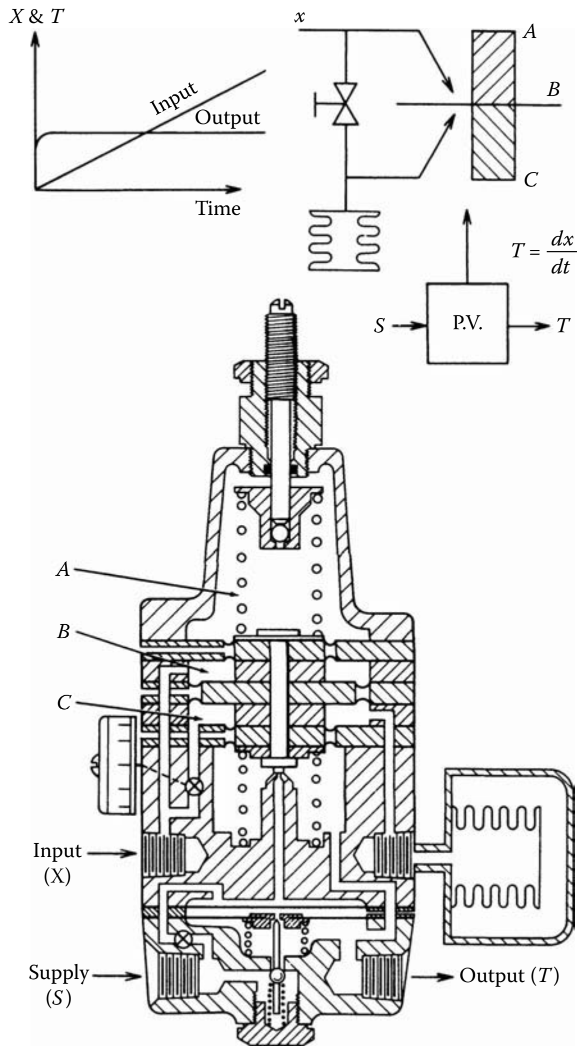


FIG. 3.4f
Differentiating relay.

Scaling and Proportioning Relays

Scaling, or proportioning, involves multiplication by a constant. Several approaches are available:

1. Special fixed-ratio relays
2. Pressure transmitters
3. Proportional controllers
4. Adjustable ratio relays

The fixed-ratio scaler is the simplest if the correct ratio is available and if adjustability and exact ratio are unnecessary. Figure 3.4g shows such a relay. The input pressure is connected to the top chamber and acts on the upper diaphragm. Output acts upward on the small bottom diaphragm. The gain is a function of the relative effective areas of the large and small diaphragms as determined by the dimensions of the diaphragm ring. The bottom spring applies a negative bias to the input and the adjustable top spring allows exact zero setting. The operating equation is:

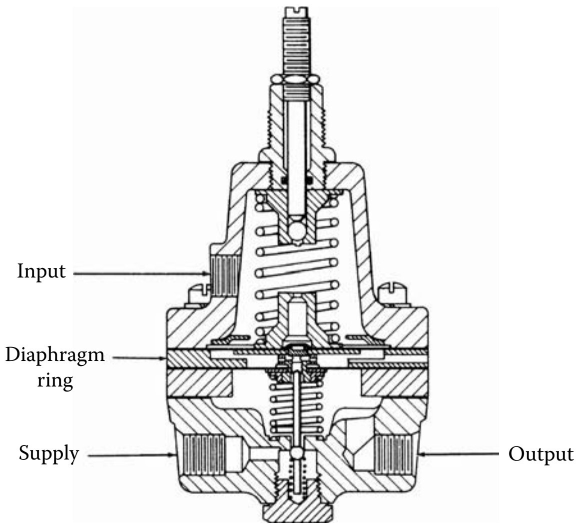


FIG. 3.4g
Fixed-ratio amplifying relay.

$$T = AP_1 + K \qquad \textbf{3.4(5)}$$

where A is the gain constant and K is the spring setting.

Where the scaling must be exact and does not have to be adjusted periodically, pressure transmitters are an economical, reliable, and accurate choice. Where the scaling factor must be modified occasionally, conventional ratio relays, which often consist of the proportioning section of a controller, are commonly used.

Integrating Relay

Integration, the reverse of differentiation, essentially involves measurement of accumulated pressure resulting from a flow that is proportional to the offset (from some chosen reference) of the input variable. Figure 3.4h shows an ideal integration relay. The input signal loads chamber *B*. The output, it should be noted, is the accumulated pressure in chamber A, *not* the booster pilot output.

The input signal determines the pressure differential across the needle valve. As in the case of the differentiation relay, with laminar flow across the needle valve, the volumetric flow is directly related to the differential. The mass flow, however, which determines the accumulated pressure, still varies with the static pressure because of compressibility. This effect is also compensated by connecting a variable volume to chamber *A*. The needle valve sets the proportionality constant of the integrator.

Neither the compensated differentiating relay nor the compensated integrator is available as standard hardware. Usually, the noncompensated relay (actually a proportional-speed floating controller) is satisfactory.

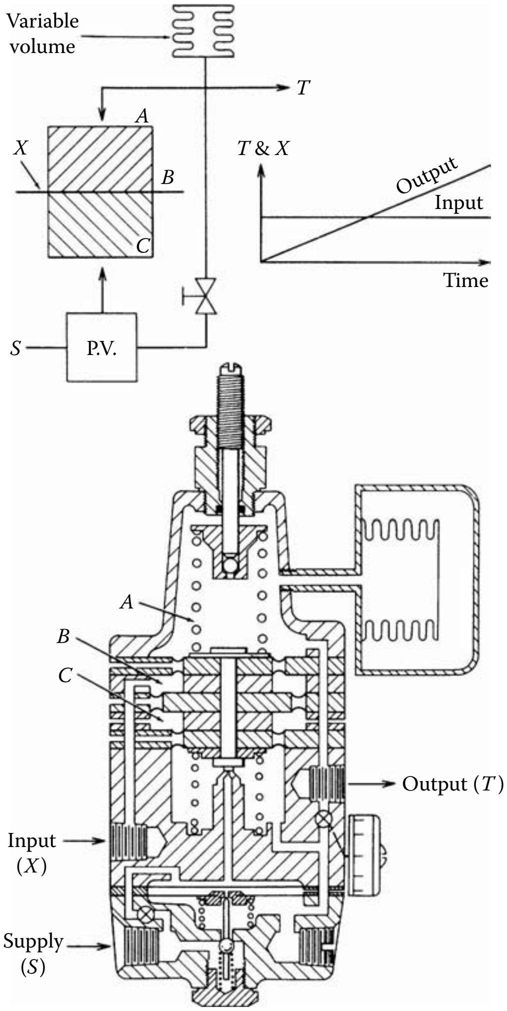


FIG. 3.4h
Integrating relay.

Square Root Extractor Relay

Square root extraction is commonly required to linearize signals from differential pressure-type flow transmitters. The force bridge, Figure 3.4b, provides square root extraction when the output is connected in common to the *A* and *C* chambers, giving the equation:

$$C^2 = B \times D \qquad \textbf{3.4(6)}$$

Other solutions are based on:

1. Use of a cam-characterized function generator
2. A geometric relationship, namely, change in cosine compared with the change in included angle, for smaller angular displacements (Figure 3.4i)

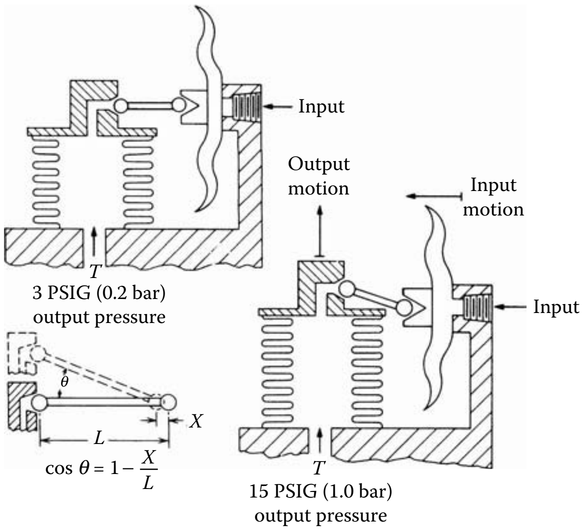


FIG. 3.4i
Square root extractor.

Starting with the input and output at 3 PSIG (0.2 bar), an increase in input causes the floating pilot link to restrict the pilot nozzle. This increases the output pressure and moves the output feedback bellows upward until balance is restored. Since the length of the floating link is fixed, the angular displacement produced by movement of the output bellows follows the relationship:

$$\cos\theta = 1 - \frac{X}{L} \qquad \textbf{3.4(7)}$$

A plot of the angle θ (output displacement) versus X (input displacement) in this equation shows the relationship to be virtually an exact square root for small angular motion.

High- and Low-Pressure Selector and Limiter

Selector relays are used in override systems. The high- pressure selector relay compares two pressures and transmits the higher of the two in its full value. In Figure 3.4j, the two input pressures act against a free-floating flapper disk. The differential pressure across the flapper always results in closure of the low-pressure port.

In the low-pressure selector, Figure 3.4k, if input *A* is less than input *B*, the diaphragm assembly throttles the pilot

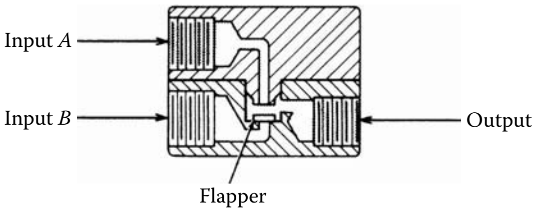


FIG. 3.4j
High-pressure selector relay.

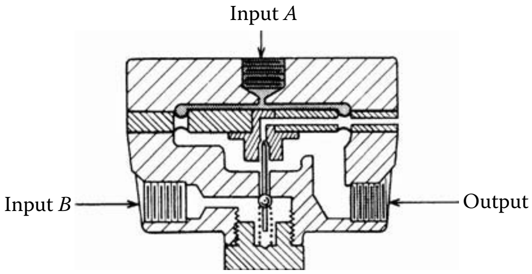


FIG. 3.4k
Low-pressure selector relay.

plunger to make the output equal to input A (the conventional action of a 1 : 1 booster relay). If input B is less than A, the supply seat of the pilot plunger is wide open so that pressure B is transmitted in its full value.

By connecting a set reference pressure into one of the ports of the high-pressure selector, a low-limit relay results. Conversely, by connecting a set reference pressure into the low-pressure selector, a high-limit relay results. Limit relays are available with the reference-setting regulator built into the relay.

ELECTRONIC COMPUTING ELEMENTS

Although electronic instrumentation is based on large-scale integrated circuits, computing functions based on discrete component operational amplifier, i.e., op amp, are still available. In addition, op amp concepts are at the heart of all the current and future electronic computing elements. Therefore, it is appropriate to review the electronic computing functions within their op amp context.

Multiplying and Dividing

In Figure 3.4l, inputs e_1 and e_2 are multiplied in the diode bridge. Conduction of the diodes in the bridge is dependent upon the relative magnitude of the inputs with respect to the constant slope of the sawtooth input. The output of the diode bridge is a trapezoid, the area of which can be calculated by the following equation:

$$\text{area} = e_1 e_2 \tan\theta \qquad \mathbf{3.4(8)}$$

The angle θ is established by the constant slope of the sawtooth, and thus

$$\text{area} = K e_1 e_2 \qquad \mathbf{3.4(9)}$$

The output voltage, e_0, is amplified and filtered to a DC signal, and its voltage level will therefore be proportional to the area and, consequently, to the product of e_1 and e_2.

Adding another diode bridge to the multiplier circuit produces a multiplier/divider (Figure 3.4m). The input to the amplifier is the output difference from the two bridge networks.

$$e_0 = A(Ke_1e_2 - Ke_3e_0) \qquad \mathbf{3.4(10)}$$

where A = gain of the amplifier. Rearranging the equation yields:

$$e_1e_2 = \frac{e_0}{AK} + e_3e_0 \qquad \mathbf{3.4(11)}$$

The term e_0/AK is very small if the amplifier gain is high, and thus

$$e_0 = \frac{e_1e_2}{e_3} \qquad \mathbf{3.4(12)}$$

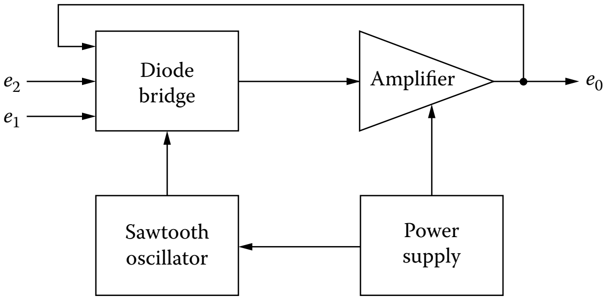


FIG. 3.4l
Multiplier.

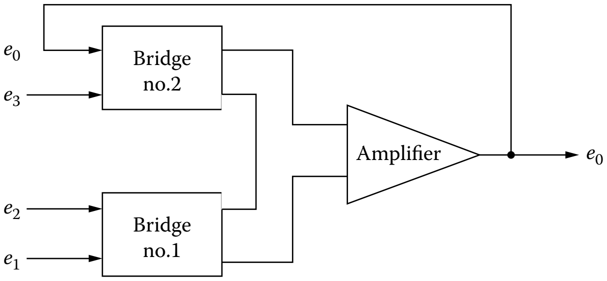


FIG. 3.4m
Multiplier-divider.

Adding, Subtracting, and Inverting

In Figure 3.4n, the two input potentials e_1 and e_2 are compared in the multiple comparator, which produces a proportional output to the amplifier. The current paths of the two inputs can be the same or opposite, resulting in either an adding or subtracting circuit, respectively.

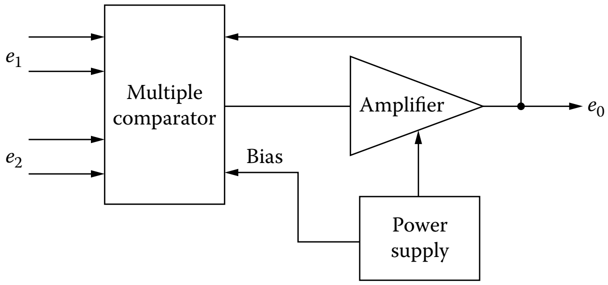


FIG. 3.4n
Adder, subtracter, and inverter.

Inverting is accomplished by biasing the comparator to produce maximum output with no input. Applying a reverse input (i.e., a reverse current input with respect to bias current) causes the output to decrease with increasing input. The feedback signal is such that the amplifier acts as a unity gain network.

Differentiating

The input amplifier in Figure 3.4o is a capacitor-coupled so that only the rate of change of the input signal is seen by the amplifier. Two diodes in the feedback of the amplifier allow its output to go positive or negative (depending on the direction of the rate of change) by an amount equal to the forward drop across the diodes (only a few tenths of a volt). The output amplifier inverts and amplifies this signal by its open-loop gain.

A small positive feedback is applied to the last amplifier to prevent output from "chattering" at the diodes' switching point.

Integrating

The first amplifier in Figure 3.4p, a simple inverting type, performs the integration function as the charge accumulates across the capacitor of the RC (resistor-capacitor) network. The second amplifier is an inverting, general-purpose type, which relates the output directly to the input.

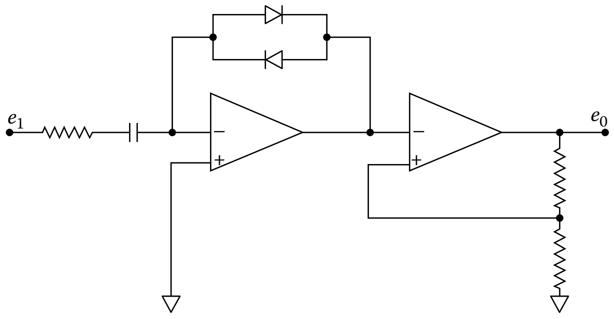


FIG. 3.4o
Differentiator.

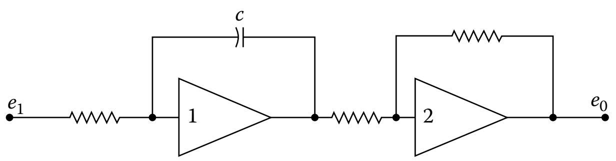


FIG. 3.4p
Integrator.

Scaling and Proportioning

Simple electronic scaling or proportioning involves combining a voltage divider circuit with an amplifier. The voltage divider circuit is connected to either the input or output side, depending upon whether the gain is to be greater or less than unity.

In Figure 3.4q, the amplifier comes to balance when Δe_i equals zero. Since the voltage divider is on the output, only a portion of the amplifier output is fed back to counterbalance the input voltage. Therefore, the output will rise above e_1, resulting in gains greater than one. The operation can be expressed as:

$$e_0 = e_1 \frac{(R_1 + R_2)}{R_1} \tag{3.4(13)}$$

An electronic scaler with a gain of less than one is constructed by a simple alteration of the circuit shown in Figure 3.4q. The voltage divider is moved from the amplifier output and placed across the e_1 input. If the signal from the middle of the voltage divider is fed to the amplifier's "+" terminal, then the resulting gain will be less than one. This can be expressed as:

$$e_0 = e_1 \frac{R_2}{R_1 + R_2} \tag{3.4(14)}$$

Square Root Extracting

The square root converter combines a DC amplifier with a negative feedback diode network. As current into the amplifier increases, the amplifier gain decreases with decreased feedback resistance in the diode network. The gain typically varies according to seven straight line segments that approximate a square root function. This is accomplished by having seven diode-resistance

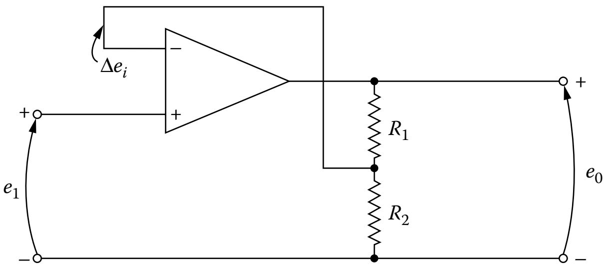


FIG. 3.4q
Scaler with gain greater than one.

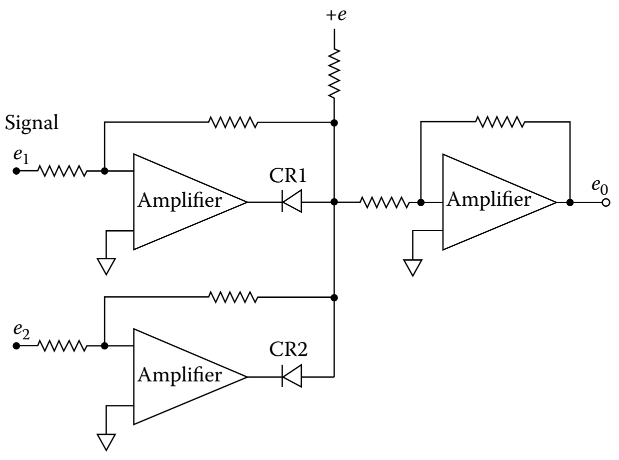


FIG. 3.4r
High-voltage selector.

paths in the feedback network automatically parallel each other with increasing input. The output stabilizes when the diode network modified feedback counterbalances the input.

High- and Low-Voltage Selector and Limiter

The higher of the two positive inputs in Figure 3.4r causes a higher negative potential at the cathode of one of the diodes (CR1 or CR2). The forward bias of this diode passes the higher input and reverse-biases the other diode to isolate the lower input. Thus if signal e_1 drops below signal e_2, CR2 is forward-biased to pass signal e_2 and CR1 is reverse-biased to isolate signal e_1.

All the amplifiers are unity gain inverter types. Substituting a fixed input for one of the variables produces a low-limit relay. To obtain a low-voltage selector, the diodes shown in Figure 3.4r are inverted and a negative supply is used. Thus, the least positive input forward-biases one of the diodes (by the least negative potential applied to the anodes of the diodes). This automatically reverse-biases the other diode and isolates the higher input from the output.

TIMING ELEMENTS

Common technologies for time management devices include discrete analog and digital circuits, microprocessor-based time functions, and pneumatic timing devices. Each of these choices may offer an advantage over the other, depending on the type of timing operation and process application of interest. These timing options are discussed below and also in Chapter 5, where the emphasis is on digital software-based systems.

Here the timing elements are grouped for convenience into analog, digital, and microprocessor categories. In most process applications digital and microprocessor-based timing functions provide the same accuracy as their analog equivalents. Analog timing elements are still used in some cases because of historic, maintenance, and process environment reasons, that is, people have used them in the past and are comfortable with them, technical support personnel needed to repair them are available in the plant, and they function well under various harsh plant operating conditions. All of the analog time functions discussed here are also available as digital or microprocessor-based units.

Analog Timing Operations

A simple but highly useful time function generator is the first-order lag function. In many industrial control schemes, motors and other loads must be energized in a time sequence provided by a lag function. First-order time functions are also used in controllers that provide reset and rate actions. Although a first-order time function is certainly available as a digital circuit or as a microprocessor-based unit, it can also be implemented in its analog or pneumatic versions.

Electronic Time Delay Analog first-order lag circuits are common and are often used to generate delay times on applications involving the starting of motors. Such circuits are simple variable resistance and capacitance circuits that provide output voltages as a function of time as set by the values of the two variable circuit components. The integral equation that describes this type of circuit is as given in Equation 3.4(15):

$$V(t) = \mathrm{IR}(1 - e^{-\tau/\mathrm{RC}}) \qquad \textbf{3.4(15)}$$

Voltages from circuits that follow Equation 3.4(15) vary exponentially with time. Thus, a motor starter that is monitoring $V(t)$ will not respond until the voltage reaches a level it recognizes. The time delay that it takes for the voltage input signal to the motor to go from zero volts to the starter's trigger voltage is controlled by the actual values of R and C, which are selected for the specific motor application.

Pneumatic Time Delay The equivalent RC circuit in its pneumatic form is usually described by Equation 3.4(16):

$$\tau dP_i/dt + P_i = P_0 \qquad \textbf{3.4(16)}$$

The solution of this differential equation with respect to $P_i(t)$, which is the input pressure as a function of time, gives a pressure–time profile similar to the shape of the voltage–time profile obtained from the solution of Equation 3.4(15).

The construction of a pneumatic system that follows Equation 3.4(16) is not difficult. The input air signal is fed to a laminar flow throttling device and a small-capacity single-port storage tank. The time delay in the output pressure signal is a function of the storage tank size and the restriction in the throttling device.

As with its analog circuit counterpart, a pneumatic time lag device must be able to vary the time constant, τ, over a range from 5 seconds to about 20 minutes. On commercially available units, the resistance and capacity parameters are designed such that τ has an adjustable range from 5 to 300 seconds for each capacity. Therefore, a four-equal-capacity parallel system would allow an adjustable range from 20 seconds to 20 minutes.

Digital Timing Operations

This subject is discussed in more detail in Volume 3 and in Chapter 5 of this volume.

The main distinction between the analog and digital worlds is the concept of a discrete event. For timing elements, the disadvantages of a discrete, i.e., digital- or microprocessor-based, timing function becomes significant in a control application only when the time between output signal updates is long; if the circuit does not quickly update the output signal to the process heater, then the ramp profile will have a lot of fluctuation in it.

All of the timing functions are available as digital- or microprocessor-based products. Their selection depends on several constraints. First, the selected timing circuit must update its control signals at least twice as fast as the process can change the value of the process variable that is being controlled. Second, the selected timing element should be able to survive the chemical and mechanical environment it is to be subjected to. Finally, the product must function properly despite electrical noise.

Microprocessor Timing Functions When compared to the older technologies, one major difference is the convenience with which additional control functions can be included with the microprocessor-based system. On the other hand it is desirable to make sure that the unit meets government and industry safety standards for the expected application. These include fire and explosion standards in hazardous environments. Second, will it meet the mechanical vibration and electrical requirements for the environment in which it will be installed? Third, do the input signal conditions required and the output signal characteristics delivered by the unit match the needs of the application?

PROGRAMMERS

Step Programmers

A step programmer is a process sequencer that does not have a practical analog circuit equivalent.

Step programmers are used for on/off event sequencing, a subject that is discussed in more detail in Chapter 5 under the topics of PLCs and logic elements. These step programmers do not provide an analog output. A typical design consists of a perforated drum (Figure 3.4s). Each perforation represents a step in one channel. Drums are available with from 30 to 100 steps and from 16 to 93 channels. Inserting a nylon plug into a hole results in a switch actuation on the corresponding step and channel. The stepping can be initiated by a remote sensor switch, counter, timer, or pushbutton.

These units are easily programmed and can replace complex logic and interlock circuits. They can not only replace such systems but can eliminate the need for their custom design and construction as well.

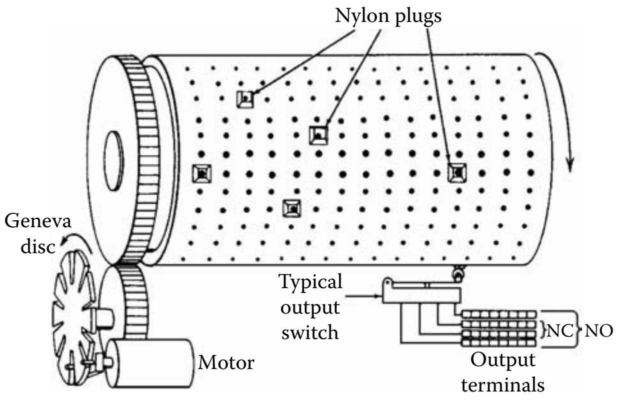


FIG. 3.4s
Drum programmer for on/off event sequencing.

Related to the step programmer is the multi-channel cam timer in which a number of individually adjustable cams are mounted on a single drive shaft and provide event-sequencing control.

Profile Programmers

There are three types of process profile programmers that use the same basic time function control mechanisms. Although the cam-type, the belt-type, and the electrostatic-line-type programmers have different mechanical operating principles, they fundamentally perform the same process control function. Each uses a tracing element that follows the movement of a rotating cam, belt, or electrostatic line. Each has a motion detection subsystem that converts the tracing element movements into a predetermined time-dependent control signal. The time dependence of the control signal depends on the shape, cut, or position of the programmer's cam, belt, or electrostatic line.

Cam Programmer Figure 3.4t shows the general layout for a cam-driven profile programmer. The cams of the cam-type programmers can be driven by pneumatic or electric motor drives that move the set-point index. This motion in the tracing element in turn generates a time-dependent standard, i.e., 3 to 15 PSIG, 4 to 20 mA DC, etc., output signal. The cams can be

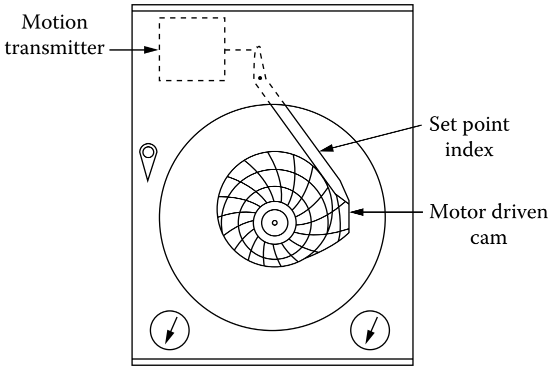


FIG. 3.4t
Cam-type programmer.

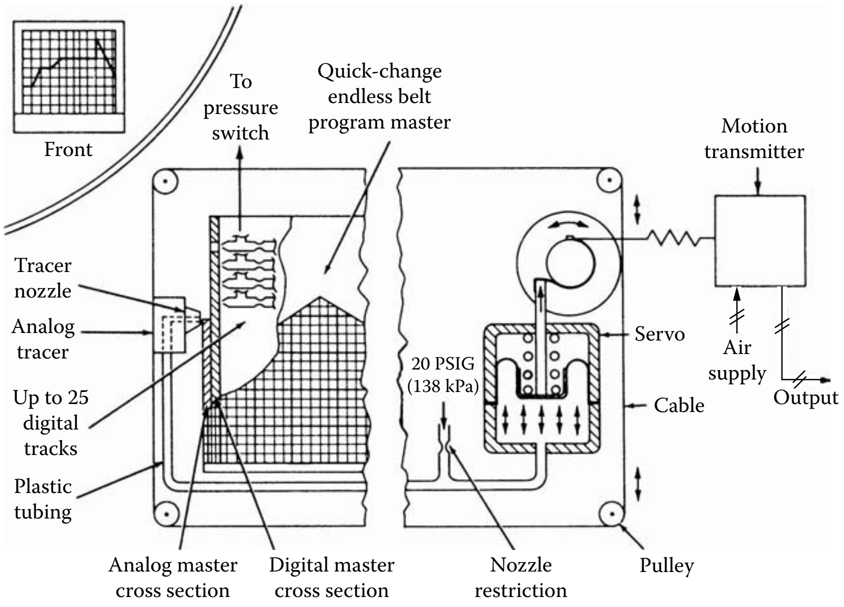


FIG. 3.4u
Pneumatic profile tracer programmer with synchronized on/off sequence control switches.

made of plastic or metallic materials. Traditionally, cam programmers have incorporated an integral controller, a sensor element, and a circular chart recorder in the same package.

Cam programmers are used only on batch processes for the following reasons.

- The making of the new cams and changing them is not a simple matter.
- The cam rise is limited for mechanical reasons to about a 50-degree cam rotation for full-scale movement of the tracing element.
- Although graphic lines are scribed onto the cam surface to facilitate cutting the cam to the desired shape, these coordinates are curvilinear. It is more difficult to lay out the cam design for cutting with such coordinates than it is to program a device that uses rectilinear coordinates, such as a belt programmer.

Belt Programmers Figure 3.4u illustrates an early belt-type profile programmer with pneumatic output. This specific unit was housed in a 6 in. by 6 in. recorder case. The time profile program was stored on a laminated, endless-belt plastic master. It combines an analog set-point program with up to 25 synchronizes digital tracks for operation of logic circuits, auxiliary equipment, solenoid valves, lights, etc. There is no limit to the slope the programmer can follow—even slopes of 90 degrees can be accommodated.

Since the master program can be quickly changed, these programmers were often used where the program does require periodic change and where accurate reproduction of the program is essential, as in textile dyeing processes. The complete program is stored on the master, thus eliminating the need for having an operator make various program settings for each change and therefore eliminating the chance for human error in setting the program. These programmers are accurate to within 1/4% of full scale, which makes them applicable when accuracy is one of the critical requirements of the operation.

The endless belt master can be made up by plotting the desired analog program on the rectilinear chart and cutting the top portion away with scissors. The second layer serves as a backing and is also used to program the synchronized digital tracks. At any point of the program where a switch action is desired, a hole can be punched in with a conductor's punch. The back of the analog program has a pressure-sensitive adhesive that joins the two sections. A splice finishes the makeup of the master.

In operation, a motor drives the master program. A cable-mounted tracer nozzle senses the step on the analog program profile. The backpressure of the nozzle actuates a servo which, through the cable drive, keeps the tracer following the profile. Operating from the same servo drive is an accurate force-balance-type motion detector. Sensing the back side of the digital master are a series of vertically aligned nozzles. Normally, their backpressure is high, since the master baffles the nozzles. However, if a punched hole presents itself, the backpressure of that particular nozzle drops to zero, actuating the connected pressure switch.

Line Follower Programmers Electric line and edge follower programmers will perform with less than ±1/4% error. In the electrostatic line follower design, Figure 3.4v, the desired program curve is etched into a chart with a conductive surface, dividing it into two electrically isolated surfaces. These surfaces are energized by oppositely phased AC voltages, which establish a gradient across the gap. A noncontacting probe is used to sense the electrostatic fields developed by the surfaces. It energizes a

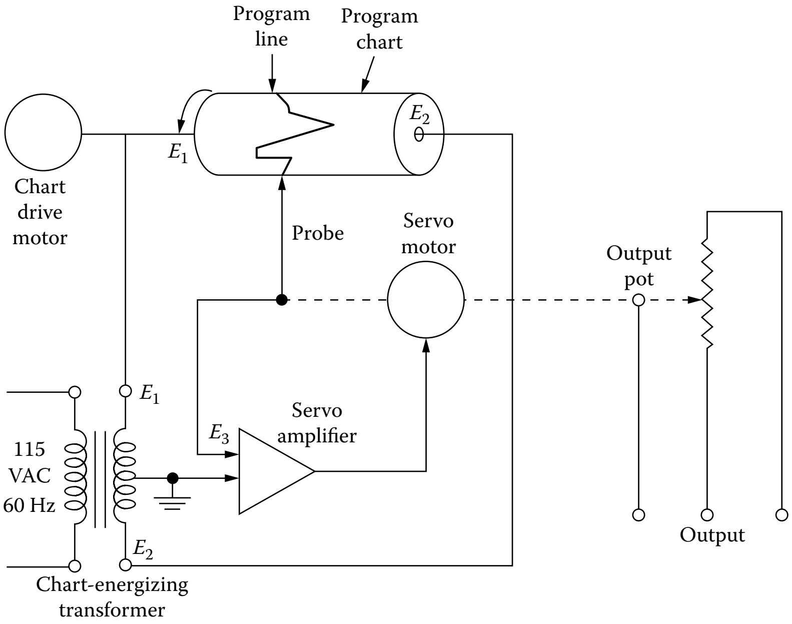


FIG. 3.4v
Electrostatic line follower programmer.

servo amplifier, which keeps the probe at zero potential as it tracks the line. Attached to the servo drive is the wiper of a potentiometer with its output proportional to line position.

The photoelectric line follower type functions to keep the line centered between two slightly overlapping pickup heads. The detector must be manually set over the line at startup, and slope rate is limited by the speed of the follower mechanism. The photoelectric edge programmer is a common variation of the electrostatic line programmer. This type of programmer has a higher slope rate limit than does the line programmer. The edge detection system is basically a photocell detector attached to the tracing element. Line and edge programmers include digital control tracks as common additional features.

Adjustable Ramp-and-Hold Programmers For such batch processes in which the controlled variable must rise at a controlled rate, then hold at some preset value and, possibly, fall at a controlled rate, programmers such as that in Figure 3.4w are often preferable to cam types. This is particularly the case if the program must be changed periodically.

These too are usually packaged as large-case circular chart recorders. In this type of programmer the set-point index is driven by a constant-speed motor. The rate of rise is set by adjustment of an interrupter timer, which makes contact for a set percentage of the basic timer cycle time. The movement of the index is therefore actually in steps, but the steps are so small that the operation is, for all practical purposes, continuous. The set-point rises until it coincides with the hold point index, at which point the hold timer is energized while the interrupter timer is deenergized. Controlled cooling rate requires driving the set-point index in reverse. This type of programmer can come complete with a controller and a direct sensing element.

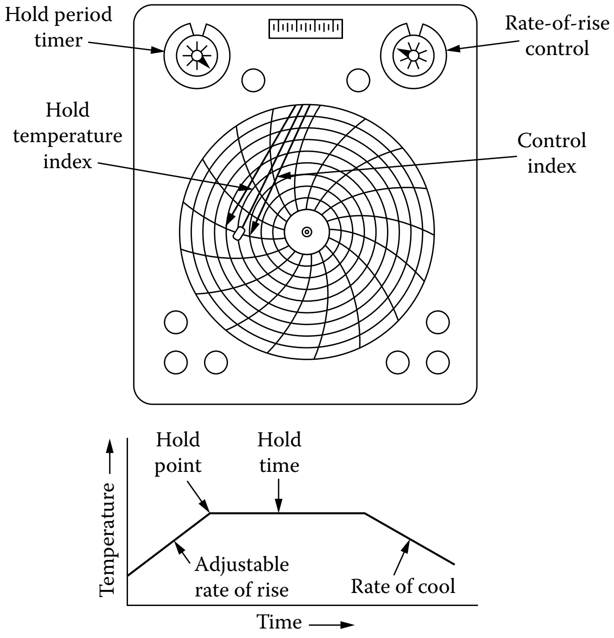


FIG. 3.4w
Adjustable ramp-and-hold programmer.

Bibliography

ARC Report, “Batch Process Automation Strategies,” ARC Advisory Group, October 1999.

Buckley, P. S., "Designing Long-Line Pneumatic Control Systems," *Instrumentation Technology,* April 1969.

Dyer, S. A., *Wiley Survey of Instrumentation and Measurement,* New York: Wiley, 2001.

Eckman, D. P., *Automatic Process Control,* New York: John Wiley & Sons, 1958.

Farmer, E., "Pneumatics in a Digital World," *Instruments and Control Systems*, March 1979.

Gassett, L. D., "Pneumatic or Electronic?" *Chemical Engineering,* June 2, 1969.

Johnson, C. D., *Process Control Instrumentation Technology,* Prentice Hall, 7th ed., 2002.

Leonessa, A., et al., *Nonlinear Switching Control Design,* Heidelberg: Springer-Verlag, August 2000.

Luyben, W. L., et al., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Mamzic, C. L., "Pneumatic Controls Interface with Computers," *Instruments and Control Systems,* October 1975.

Nachtigal, C.N., *Instrumentation and Control,* New York: Wiley, 1990.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, 1985.

"Timing and Programming Equipment," *Measurements and Control,* September 1993.

3.5 Telemetering Systems

C. H. HOEPPNER (1970) **J. VENCZEL** (1985) **B. G. LIPTÁK** (1995) **H. EREN** (2005)

Types:	Wire telemetry and wireless telemetry
Sensors:	Common sensors for both types; additionally, IC and intelligent sensors for digital instruments
Components:	Collection of many components: computers, microprocessor and micro-controller systems, intelligent and common sensors and transducers, instruments, controllers, wire and wireless communication devices and components, communication protocols and software, repeaters, satellite systems, GPS systems
Displays:	Computer peripherals, numeric (LED or LCD) displays
Communications:	Wired base-band and multiplexed, RF, microwave, optical, sonic, satellite communications
Networking:	Unlimited network capabilities
Vendors (partial list):	Acroamatics Telemetry Systems (www.acroam.com) Athena Controls, Inc. (www.athenacontrols.com) Automated Control Systems (www.testdevices.com) Barton Instrument Systems Ltd (www.barton-instruments.com) Bristol Babcock (www.bristolbabcock.com) Crompton Instruments Ltd (www.crompton-instruments.com) Data Flow Systems, Inc. (www.dataflowsys.com) Datatel Telemetry Electronik GmbH (www.datatel-telemetry.com) D&D Security Products, Inc. (www.ddsp.com) Devar, Inc. (www.devarinc.com) Dranetz Technologies, Inc. (www.dranetz.com) Dwyer Instruments, Inc. (www.dwyer-inst.com) Dynalco Controls (www.dynalco.com) EMI Technologies, Inc. (www.emitechnologies.com) Garmin, Ltd. (www.garmin.com) Geomation Measurement & Control Systems (www.controls-ez.com) Harris Corp., Government Aerospace Systems Div. (www.harris.com) Industrial Control Links, Inc. (www.iclinks.com) Industrial Instruments & Supplies, Inc. (www.iisusa.com) Internet Telemetry Corp.(www.itelemetry.net) IR Telemetrics Inc. (www.irtelemetrics.com) Kahn Instruments, Inc. (www.kahn.com) Keithley Instruments, Inc. (www.keithley.com) Koehler Inst. Co., Inc. (www.koehlerinstrument.com) Lockheed Martin Corp. (www.lockheedmartin.com) Microwave Data Systems, Inc. (www.microwavedata.com) Myers Engineering International, Inc. (www.myerseng.com) Nihon Kohden Corporation (www.nihonkohden.com) Ono Sokki Technology Inc. (www.onosokki.net) Philip Medical Systems (www.medical.philips.com) Precision Devices, Inc. (www.predev.com)

Pribusin Inc. (www.pribusin.com)
REMEC, Inc. (www.remec.com)
Satellink, Inc. (www.satellink.com)
SunTech Medical Instruments (www.suntechmed.com)
Tateyama Electr. Corp (www5.ocn.ne.jp/~tateyama/telemetrysys/en/)
Telescada, Inc. (www.telescada.com)
Teletronics Technology Corporation (www.ttcdas.com)
Texas Instruments (www.ti.com)
Turner BioSystems (www.turnerbiosystems.com)
Warren-Knight Instrument Co. (www.warrenknight.com)
Weston Aerospace (www.westonaero.com)
Wireless Data Corp. (www.ce-mag.com)
Yokogawa Corp. (www.yokogawa.com)

INTRODUCTION

Telemetry is a process of gathering information from remote locations. The data obtained from field instruments is transmitted to a convenient location for processing and recording purposes. Telemetry can be performed by different methods: electromagnetic, optical, electrical, sonic, Internet, etc. Recently, radio frequency (RF) and microwave telemetry methods have been used extensively. This section largely concentrates on wireless telemetry systems. However, the use of optical fiber systems allows the measurement of broad bandwidth and provides high immunity to noise and interference. Here, telemetry and telemetry systems that are based on electrical and electromagnetic principles are covered.

Electrical telemetry methods can be divided into groups depending on the transmission methods that they use. There are two basic types: wire telemetry and wireless telemetry. Wire telemetry uses wire transmission utilizing coaxial cables, twisted wire pairs, telephone systems, common communication buses, or similar communication means. It offers simple and inexpensive solutions for data flow and networking of devices. It is extensively used where the basic infrastructure and wiring systems already exist, as in the case of electric power lines that can be used as wire telemetry carriers.

Wireless telemetry is somewhat more complex as it requires radio frequency transmitters and receivers. Despite the complexity, it is widely used because it can transmit information over longer distances without wires from normally inaccessible areas. It can also operate at high speeds and can have the capacity to transmit several parallel channels of information at the same time. The applications of other wireless telemetry methods such as ultrasound and infrared radiation are limited to certain environments, such as marine telemetry.

Telemetry systems can operate repeatedly without any adjustments and calibrations under widely ranging environmental conditions, such as high-temperature and high-pressure situations. Telemetry is extensively used in space exploration and military applications for telemeasurements of distant variables or telecommandment of actuators and controllers.

Some land-mobile vehicles, such as passenger and cargo trains, also use telemetry systems, either wireless or by using some of the existing power distribution networks to transmit data to a central station. In medical applications, the use of telemetry increases the quality of life of patients, giving them mobility while still being monitored. Several medical applications are based on implanting a sensor in a patient and transmitting the data to be further analyzed and processed either by radio or by adapted telephone lines. In industrial applications, fiber optic communications are used in hazardous and electrically noisy environments.

The design and methodology required for a fully functional telemetry system depends on the application requirements and the environmental characteristics. From the applications point of view, telemetry can be grouped into four main categories: telemetry in medical and life sciences, industrial telemetry, space telemetry, and others. Application examples are provided at the end of this section.

BASIC TELEMETRY CONCEPTS

Classical Configuration

A telemetry system consists of many components, as illustrated in Figure 3.5a. The main components are:

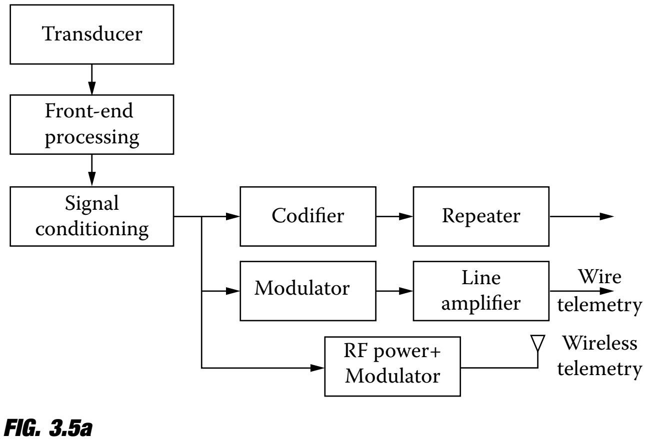


FIG. 3.5a
Basic components of a telemetry system.

1. Sensors and transducers, which convert physical variables into electrical signals for processing
2. Preconditioning circuits, which serve to amplify the low-level signals, limit their bandwidth, and realize impedance matching of inputs and outputs
3. A main signal processing unit for further signal conditioning
4. A sub-carrier oscillator, whose signal is modulated by the output of the transducer
5. A coding circuit, which can be a digital encoder or an analog or digital modulator, and which serves to prepare the signal to match the characteristics of transmission channel
6. A radio transmitter for wireless telemetry, modulated by the composite signal
7. An impedance line adapter in case of wire transmission, to match the input impedance of the line to the output impedance of the circuits
8. For wireless communication, a transmitting antenna
9. Dedicated or general computer and/or microprocessor systems with supporting peripherals

The receiving end of the telemetry system consists of similar modules. In the case of wireless telemetry, these modules are:

1. A receiving antenna suitably designed to operate in the radio frequency band of interest
2. A radio receiver with a suitable demodulation circuitry
3. Demodulation circuits for each of the transmitted channels in case of multiple transmissions
4. Computers and/or microprocessor systems with supporting peripherals

Both in wire or wireless telemetry, the transmitted signals can be analog or digital. The digital systems send data digitally as a finite set of symbols, each representing one of the possible values of the signals at the time of sampling. In case of wire telemetry, the antenna and the radio receiver are substituted by a generic front-end circuit to amplify and process the signals.

The effective communication distance in a wireless system is limited by the power radiated by the transmitting antenna, the receiver's sensitivity, and the bandwidth of the radio frequency signal. As the bandwidth increases, the contribution of noise to the total signal also increases, and consequently more transmitted power is needed to maintain the same signal-to-noise ratio. This is one of the main limitations of wireless telemetry systems.

In some applications, the transmission is done on base band, after the conditioning circuits. The advantage of base-band telemetry systems is their simplicity, although because of the base-band transmission, they are normally limited to only one channel at low speeds. The base-band concept will be explained in detail in the following sections.

Recent Trends

Nowadays, many instruments use intelligent sensors that include not only the sensor but also the signal processors and telemetric communication capabilities in a single chip. These units are appearing in the marketplace as pressure sensors and accelerometers, bio-sensors, chemical sensors, optical sensors, magnetic sensors, environmental sensors, and so on. Some of these sensors are manufactured with RF transceivers, neural network, neural processors, vision systems, and intelligent parallel processors.

Recently, due to advances in wireless communication technology and intelligent sensors, a new class of devices, called wireless sensors, is finding extensive applications. These small sensor devices integrate micro-sensing and actuation with on-board processing and wireless communication capabilities. In many applications, due to low cost and small size, a large number of sensors can be deployed. These sensors are capable of organizing themselves into multi-hop wireless networks in environmental monitoring, surveillance, and scientific data gathering applications.

In parallel with the rapid developments in the hardware aspects of modern sensors, some new standardized techniques are emerging for intelligent sensors such as the support software and robust communications and networking components. This is making a revolutionary contribution in the development and applications of telemetry. For instance, the IEEE-1451 is a set of standards that define interfaces, network-based data acquisition, and control of sensors. This subject will be revisited later in this section.

Apart from the intelligent and sophisticated sensors, the communications capabilities of many modern electronic instruments are enhancing the widening of their application possibilities. Modern instruments have two levels of communications. The first is communication at the sensor level, which is mainly realized by intelligent sensors. The second is at the device level. At the device level, the communication takes place in wire, wireless, infrared, and sonic forms.

In the implementation of wireless communications of today's instruments and sensors, the "Bluetooth" technology and Near-Field Communication protocols are gaining wider acceptance. Particularly, Bluetooth is considered to be a low-cost and short-range wireless technology to provide communication functionality, ranging from short length wire replacements to wide area networks. As the number of products incorporating technology such as the Bluetooth increases, the development of various types of instruments for many new applications gains momentum.

OPERATING PRINCIPLES AND TYPES

Telemetry can be analog or digital. Nowadays, mostly digital telemetry is used because of ease in data transmission and handling. As far as operating principles are concerned, telemetry can be divided into two main categories: base-band

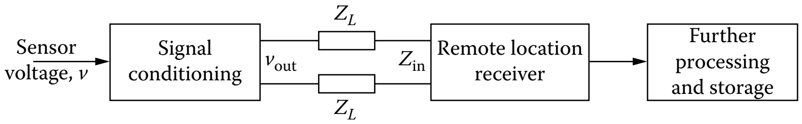


FIG. 3.5b
A voltage-based telemetry system.

telemetry and multiple-channel telemetry. These methods will be explained next, and appropriate comments will be made for their analog and digital applications.

Base-Band Telemetry

Base-band telemetry uses a wire line to communicate the signals obtained from the sensors and transducer to some remote location. There are two basic types: amplitude-based telemetry and frequency-based telemetry.

Amplitude-Based Telemetry Amplitude-based telemetry is extensively used in industry for short-distance signal transmission. In this system, the signal from the transducer is amplified, normally to a voltage level between 1 and 15 V, and is sent to the receiver through a line consisting of two wires. Figure 3.5b shows a simple voltage-based telemetry system. The permissible distance for transmission depends on the resistance of the line, the input resistance for the receiver, and the electrical noise of the environment.

The limitation on the transmission distance due to noise and impedance of the line can be partially overcome by using current signals instead of voltages, as is shown in Figure 3.5c. The basic theory is that unlike voltage, the current is not affected by noise. In addition, the drop in voltage caused by the wire resistance does not affect the transmitted signal.

However, this system requires an additional step of voltage-to-current conversion. At the receiver end, the signal is converted to voltage by using a suitable resistor. The most common method used in industry is the 4- to 20-mA current transmission. In this configuration, zero voltage corresponds to and is transmitted as a 4 mA current value, while the highest voltage is transmitted as 20 mA. Other standard current ranges are 0 to 5, 0 to 20, 1 to 50, 1 to 5, and 2 to 10 mA.

Frequency-Based Telemetry Similarly to amplitude-based telemetry, frequency-based telemetry is also often used in industry for short-distance signal transmission. Frequency-based telemetry, shown in Figure 3.3d, is known to provide higher immunity to noise. In this method, the signals obtained from the measurements are converted to frequencies by voltage-to-frequency converters. A frequency-to-voltage converter puts signals back to the voltage form at the receiving end. A special form of frequency-based telemetry is pulse telemetry in which the modulating signals are in the form of a train of pulses, which will be explained in detail in the forthcoming paragraphs.

Multiple-Channel Telemetry

Most processes require the measurement of multiple physical variables, thus requiring special methods for handling the signals. In multiple-channel applications, the use of base-band telemetry is not economical, as it would require building a different system for each channel. Instead, multiple-channel telemetry can be configured by sharing a common transmission channel (Figure 3.5e) and by using suitable multiplexing techniques.

The multiplexing can be realized by frequency division multiplexing (FDM) or time division multiplexing (TDM). In frequency division multiplexing, different channels are assigned to different spectrum bands. In time division multiplexing (TDM), the information of different channels is transmitted sequentially.

Frequency Division Multiplexing Frequency division multiplexing (Figure 3.5f) makes use of modulations of signals from each channel onto sinusoidal signals of different frequencies. The sinusoidal signals are called sub-carriers. Each modulated signal is then sent through a low-pass filter to ensure that the bandwidth limits are observed.

After the filtering stage, all the modulated signals are fed into a common summing block, producing what is known as the base-band signal. The base-band signal is then modulated by a carrier signal whose frequency and amplitude depend on the characteristics of the transmission. The carrier signal is then fed into a transmission wire or into an antenna in the case of wireless systems.

At the receiver end, the transmitted signal is detected and demodulated to separate each sub-carrier signal. This is done by feeding the signal into a bank of parallel pass-band filters.

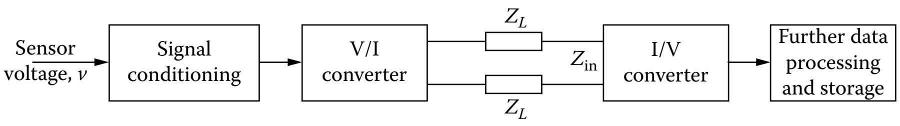


FIG. 3.5c
A current-based telemetry system.

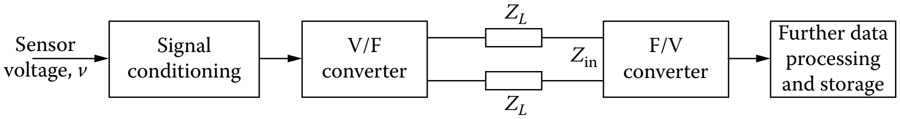


FIG. 3.5d
A frequency-based telemetry system.

Each channel is further demodulated to recover the original measurement signals. One problem with FDM systems is the cross-talk between channels. Cross-talk can be minimized by having a guard band between the spectrums of two contiguous channels. The guard band can be increased to decrease the cross-talk, but then the effective bandwidth also increases. The effective bandwidth is the sum of the bandwidth of all channels, plus the sum of all the guard bands.

The frequency and energy level of the carrier cannot be arbitrary; hence, international agreements on the use of the electromagnetic spectrum must be complied with. There are national organizations as well as international consortiums (e.g., International Consortium on Telemetry Spectrum, ICTS) to regulate the use of the electromagnetic frequency spectrum. The problem of allocation of the telemetry spectrum is common to many nations, whether telemetry is used in support of national defense, the commercial aerospace industry, or space applications.

In the United States, the Federal Communications Commission (FCC) is the body that regulates the allocation of frequencies for various communication services. For informational purposes, the most commonly used telemetry frequency bands and their intended use are shown in Table 3.5g. The allocation of bands is a process subject to change from time to time.

There are three basic techniques of modulation for FDM: amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). Combinations of these modulation techniques are also used—FM/FM, FM/PM, or AM/FM.

Amplitude Modulation Amplitude modulation makes use of the amplitude of a sub-carrier signal that changes in accordance with the value of the measured channel. The resulting signal can be expressed by:

$$v(t) = A_c\,[1 + m(t)]\cos(\omega_c t) \qquad \textbf{3.5(1)}$$

where A_c is the amplitude of the carrier, $m(t)$ the modulating signal, and ω_c the frequency of the carrier.

AM has limitations, particularly due to efficiency. It is possible to overcome some of the limitations by using modulation techniques, such as Double Side Band (DSB), Single Side Band (SSB), and Compatible Single Side Band (CSSB) techniques.

Frequency Modulation Frequency modulation and phase modulation techniques are by far the most common modulation scheme used in FDM telemetry systems. These modulations are inherently nonlinear. For example, the angle modulation can be expressed as:

$$v(t) = A\cos[\omega_c t + \phi(t)] \qquad \textbf{3.5(2)}$$

where $\phi(t)$ carries the information on the signals from the transducers.

The value of the instantaneous frequency can be expressed as:

$$f = \frac{1}{2\pi}\frac{d}{dt}[\omega_c t + \phi(t)] = \frac{\omega_c}{2\pi} + \frac{\mathrm{d}}{dt}\phi(t) \qquad \textbf{3.5(3)}$$

From these equations, it can be seen that the signal $v(t)$ is modulated in frequency. This expression can further be modified in terms of frequency deviation (f_m) and modulation index (β). The frequency deviation is defined as the maximum departure of the instantaneous frequency from the carrier frequency, and the modulation index is defined as the maximum phase deviation. Now this equation can be expressed as:

$$f = \frac{\omega_c}{2\pi} + \frac{\beta_m}{2\pi}\cos(\omega_m t) = f_c + \beta f_m \cos(\omega_m t) \qquad \textbf{3.5(4)}$$

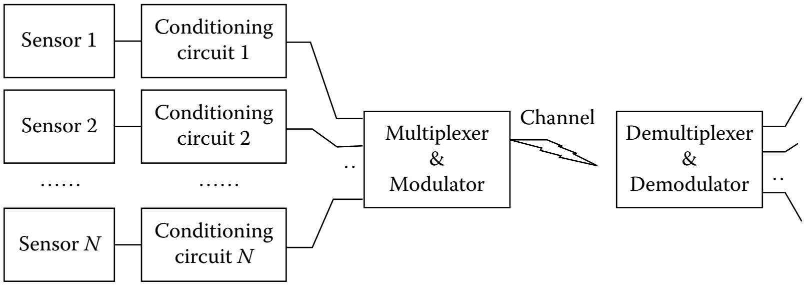


FIG. 3.5e
Block diagram of a multiple-channel telemetry system.

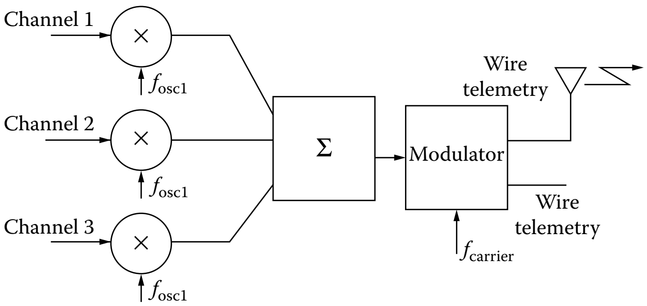


FIG. 3.5f
Frequency division multiplexing (FDM) block diagram.

The maximum frequency deviation is defined as Δf and is given by

$$\Delta f = \beta f_m \qquad \textbf{3.5(5)}$$

Therefore, we can express the equation for the frequency modulated signal as:

$$v(t) = A\cos\left[\omega_c t + \frac{\Delta f}{f_m}\sin(\omega_m t)\right] \qquad \textbf{3.5(6)}$$

The previous equation shows that the instantaneous frequency, f, lies in the range $f_c \pm \Delta f$. A practical rule states that the bandwidth of an FM modulated signal is twice the sum of the maximum frequency deviation and the modulating frequency.

TABLE 3.5g
Frequency Band Allocation for Telemetry in the United States

Frequency Band	*Uses*
72 to 76 MHz	Biotelemetry
88 to 108 MHz	Educational
154 to 174 MHz	Industry
174 to 216 MHz	Biotelemetry
216 to 222 MHz	Multiple purpose
450 to 470 MHz	General
458 to 468 MHz	Biotelemetry
512 to 566 MHz	Biotelemetry
1427 to 1435 MHz	Fixed (land mobiles services, telemetering, and telecommand)
1435 to 1535 MHz	Aeronautical (L-Band)
2200 to 2290 MHz	Mobile (S-Band)
2310 to 2360 MHz	Terrestrial sound broadcasting
2360 to 2390 MHz	Telemetry flight testing

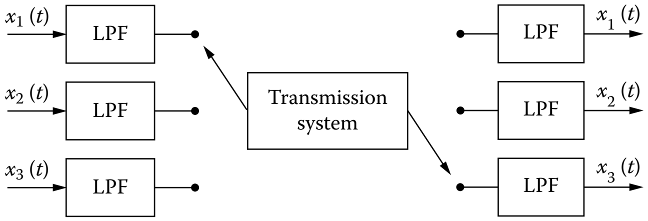


FIG. 3.5h
Block diagram of a time division multiplexing (TDM) system.

Time Division Multiplexing Time division multiplexing (TDM) is a transmission technique that divides the time in different slots and assigns one slot to each measurement channel. The entire transmission bandwidth is assigned to the assigned channel during transmission. The measurement channels are sequentially sampled by a digital switch for a period of time (T). For M measurement channels the period between two consecutive pulses is $T_s/M = 1/M f_s$, where T_s is the sampling period.

At the receiver end, the original signals are recovered by separating the digital signals into different channels by synchronized demultiplexers and by low-pass filtering. Figure 3.5h shows a basic block diagram for an FDM system.

TDM systems have a number of advantages over the FDM. In FDM, sub-carrier modulators and demodulators are required for each channel, while TDM systems need only one set of multiplexer at the transmitter and one set of demultiplexer at the receiver. The TDM signals are resistant to cross-talk. However, in TDM systems, the separation between channels depends on the sampling method adopted. Since, in practice, it is impossible to produce perfectly square pulses without rise and fall times, it is then necessary to provide time guards between pulses, similar to the band guards in FDM.

Time Division Multiplexing can be achieved by either digital or analog methods, as explained below.

Pulse Code Modulation Pulse code modulation (PCM) is a digital technique in which the measured signal is represented by a group of codified digital pulses. Figure 3.5i shows the basic elements of a PCM telemetry system. Two variations of PCM that are commonly used are Delta Modulation (DM) and Differential Pulse Code Modulation (DPCM).

PCM systems transmit data as a serial stream of digital words. An encoder converts the input data into a serial data

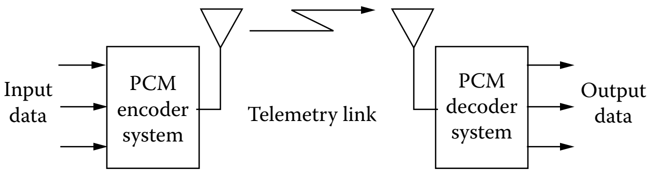


FIG. 3.5i
Block diagram of a pulse code modulation (PCM) telemetry link.

format suitable for transmission. At the receiving end, a decoder converts the serial data back into individual output data signals. The PCM encoder samples the input data and inserts the data words into a PCM frame. Words are assigned specific locations in the PCM frame so the decoder can recover the data samples corresponding to each input signal.

The simplest PCM frame consists of a frame synchronization word followed by a string of data words. The frame repeats continually to provide new data samples as the input data changes.

Pulse Amplitude Modulation *Pulse Amplitude Modulation* (*PAM*) is an analog method along with two other methods, *Pulse Duration Modulation* (*PDM*) and *Pulse Position Modulation* (*PPM*). In PAM, the waveforms are realized from unipolar, nonrectangular pulses whose amplitudes are proportional to values of samples.

Similarly, PDM is made from unipolar, rectangular pulses whose duration or widths depend on the values of samples. PPM is closely related to PDM, in which the information to be transmitted resides on the time location of the pulses rather than in the pulses themselves. The common characteristics of these modulation techniques are:

1. The modulated signal spectrum has strong low-frequency content, particularly near the sampling frequency.
2. There is a need to avoid overlaying of consecutive pulses.
3. The original signals can be reconstructed by using suitable filters after demultiplexing.

TELEMETRY PROTOCOLS, STANDARDS, AND NETWORKS

Telemetry is basically a form of communication of information obtained from measurement devices, which can be termed nodes. Information flow between nodes, individual instruments, and computers is regulated by protocols, as will be explained here.

Protocols and Standards

According to the IEEE (Institute of Electrical and Electronics Engineers), a *protocol* is defined as "a set of conventions or rules that must be adhered to by both communicating parties to ensure that information being exchanged between two parties is received and interpreted correctly." In general, communication protocols are defined by the following:

- *ISO reference model layers:* physical, datalink, network, transport, session, presentation and application layers
- *Network topology:* star, ring, or bus (tree) and so on
- *Data communication modes:* simplex, half-duplex, or duplex
- *Data transmission modes:* synchronous or asynchronous
- *Data rate supported:* several bps (bits per second) to several Gbps, depending on both oscillator frequency and transmission medium
- *Transmission medium supported:* twisted pair, coaxial cable, optical, RF, or microwave, etc.
- *Data format:* mainly based on data transmission modes and individual protocol specifications
- *Error detection methods:* parity, block sum check, or CRC (cyclic redundancy check).
- *Error control methods:* echo checking, automatic repeat request (ARQ), sequence number I, etc.

Generally, telemetry systems are subgrouped together to operate in nodes by the help of communication protocols, micro-controllers, and microprocessors. At the grassroots of this networking lie the fieldbus for industrial systems and other communication protocols developed by vendors to enable reliable and fast data communications among nodes, computers, and individual sensors. Volume 3 of this handbook is devoted to this subject.

Manufacturers offer a bewildering range of buses and protocols, but all protocols comply with the ISO reference model. Some of the typical protocols used in telemetry are: IEC 60870-5 series protocols, ISO 9506, IEC 62056, Bluetooth, and IEEE 802.11.

ISO Reference Model ISO reference model layout for the protocols is shown in Table 3.5j. The reference model has seven layers, of which each is an independent functional unit. Each layer uses functions from the layer below and provides functions for the layer above. The lowest three layers are network-dependent layers; the highest three layers are network- independent layers (application-oriented); and the middle layer (transport layer) is the interface between the two.

All digital devices need to obey the ISO reference model when communicating with each other. In some applications, some of the seven layers of the ISO reference model may be omitted, as in the cases of many fieldbuses.

Bluetooth *Bluetooth* is a protocol based on the ISO reference model; it is used in digital telemetry systems. Initially, Bluetooth was mainly aimed at bringing short-distance wireless interfaces to consumer products on a large scale. As the number of products incorporating Bluetooth wireless technology increased, the development of various types of instruments for a diverse range of applications became widespread.

However, Bluetooth did not escape competition. ExtremeTech reports that Sony and Philips have agreed to jointly work on "Near-Field Communication," a potential competitor to Bluetooth in the short-range personal area network (PAN) market. Also, the IEEE just recently adopted and approved

TABLE 3.5j
ISO Reference Model Layer Layout for the Protocols

No.	*Layer*	*Application*	*Protocols*
7	Application	Common application service elements (CASE); manufacturing message services (MMS); file transfer and management (FTAM); network management.	ISO 8650/2 (DP), RS-511, ISO 8571 (DP), IEEE 802.1.
6	Presentation	Transformation of information such as file transfer. Data interpretation, format and code transformation.	Null/MAP transfer. ISO 8823 (DP).
5	Session	Communication and transaction management, synchronization, administration of control sessions between two or more entities.	ISO Session Kernel. ISO 8237 (IS).
4	Transport	Transparent data transfer, mapping, multiplexing, end-to-end control, movement of data among network elements.	ISO Transport, Class 4. ISO8073 (IS).
3	Network	Routing, switching, segmenting, blocking, error recovery, flow control. Wide area addressing and relaying.	ISO DIS 8473, Network services, ISO DAD 8073 (IS).
2	Link	Transmission of data in local network. Establish, maintain and release data links, error and flow.	IEEE 802.4 Token Bus. IEEE 802.2 Type 1 Connection services.
1	Physical	Electrical, mechanical and packaging specifications. Functional control of data circuits.	ISO/IEEE802.4, phase Coherent Carrier, Broadband 10 Mbs, etc.

the Bluetooth protocol under its WPAN standards, the IEEE-802 series of protocols.

Many Bluetooth products are appearing on the marketplace, from the chip level to sophisticated devices. On the chip level, a typical example of Bluetooth products is an under-$5 chipset from Texas Instruments. It provides up to 1 Mbps for fast data transmission.

On the device level, Sony has introduced in Europe and Japan the ultra-compact fixed lens DSC-FX77 camera with a built-in Bluetooth communication protocol. Many chip-level products target Bluetooth specifications. Point-to-multipoint applications consisting of a base-band controller with flash memory, a reference crystal, and an RFCMOS (radio frequency complementary metal-oxide semiconductor) transceiver.

In some cases a two-chip approach is chosen to ensure a good level of performance and reliability in RF-intensive environments. The architecture is based on independent silicon optimization with digital circuits in standard CMOS and analog parts in BiCMOS (bipolar metal oxide semiconductor) or RFCMOS, enabling cost and size reduction.

IEEE 802.11 The *IEEE 802.11* is a high-bandwidth standard for transfer of large amounts of data; it is extensively used in telemetry systems. It handles spread-spectrum and high-data bursts easily. The IEEE 802.11b standard is supported by three chips compared to two chips or the single-chip solution of Bluetooth. 802.11b is designed as a communication channel to host processors running TCP/IP (transmission central protocol/internet protocol). The encryption length is 64 bits.

IEEE-1451 The *IEEE-1451* standard aims to make it easy to create solutions using existing networking technologies, standardized connections, and common software architectures. The standard allows application software, field network, and transducer decisions to be made independently. It offers flexibility to choose the products and vendors that are most appropriate for a particular application.

As an example, Bluetooth finds wide applications in wireless smart sensor and transducer networks. Interfacing of IEEE-1451 Smart Transducer nodes to a Bluetooth network is gaining momentum. This involves a detailed study of the network communication models specified for smart transducer communication as well as the Bluetooth protocol stack.

CCSDS There are many software packages supporting telemetry applications. Many of these packages are application-specific, tailored for particular applications. An example of such packages is *CCSDS* (Consultative Committee for

Space Data Systems). CCSDS is a telemetry and telecom software library that provides a reference implementation of the international protocol standard for the transmission and reception of data in radio communications with spacecraft.

This library supports the full set of uplink and downlink virtual channels. It has a frame-acceptance and -reporting mechanism (FARM) that supports a sliding window specified within the standard. The library has been written in the C programming language for execution on various computers running the SunOS4, SunOS5, AIX, and VxWorks operating systems.

MLS-STD-1553 The *MLS-STD-1553* standard addresses the acquisition of all the traffic flowing on MIL-STD-1553–type data buses for telemetry and recording purposes. Up to eight data buses within a single system can be used. Constraints such as RF bandwidth and recording times are defined.

As the application of telemetry widens, many forms of supporting hardware and software appear in the marketplace offered by different vendors and organizations. Although most of these products are supported by the existing standards, in some cases new standards appear in lieu of new developments.

Networks

The network is an essential element of telemetry systems. A network includes many instruments, controllers, microprocessors, and computers. The hardware architectures of a network can be centralized, decentralized, hierarchical, or distributed. Many different network topologies exist, as illustrated in Figure 3.5k.

The group arrangement of telemetry systems results in many advantages. For example in message sharing, because all the nodes are connected together by one single transmission medium, nodes can demand information from another node to implement strategies, for setting and resetting, for alarms, and so on. Also, due to the simple data transmission medium and the node configuration, any problem from sensors or transmissions can be isolated, thus helping maintenance without affecting the performance of the network.

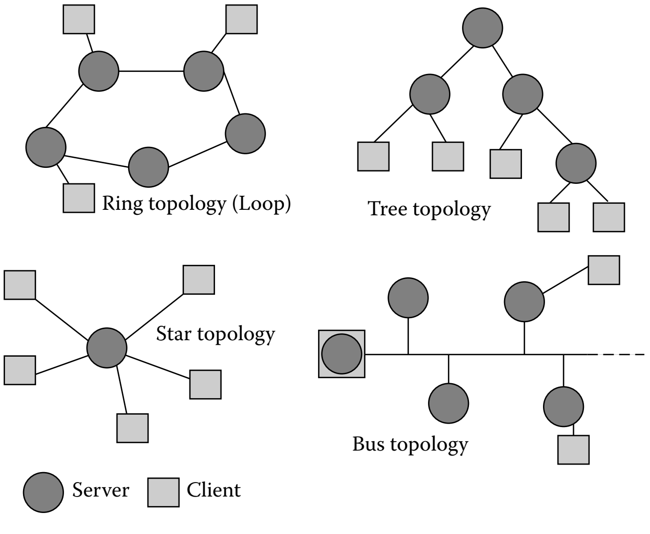


FIG. 3.5k
Common network configurations.

Telemetry Equipment and Hardware

As stated previously, a telemetry system consists of a collection of components such as computers, micro-controllers, wire and wireless transmitters and receivers, transceivers, repeaters, receivers/decoders, sensors, actuators, measurement instruments, controllers, signal converters, plug-and-play components, and protocol converters. Most of these components can be bought off the shelf or can be designed from their basic components. A list of the manufacturers and suppliers of telemetry systems and components is provided on the first page of this section.

If a system is to be designed from the basic principles, there are many single-chip RF transceivers available. A typical example is the LMX3162 from National Instruments. This transceiver is a 48-pin 7 mm × 7 mm × 1.4 mm monolithic device operating at 2.45 GHz wireless systems. It contains phase locked loop (PLL) transmit and receive functions. The 1.3 GHz PLL is shared between transmit and receive sections. The transmitter includes frequency doublers and a high frequency buffer. The receiver consists of a 2.5-GHz low noise mixer, an intermediate frequency (IF) amplifier, a high-gain limiting amplifier, a frequency discriminator, a received signal strength indicator (RSSI), and an analog DC compensation loop. The circuit has on-chip voltage regulation, which allows the use of supply voltages ranging from 3.0 to 5.5 V. Such chips are extensively used in personal wireless communication systems/networks (PCS/PCN), and wireless local area networks (WLANs).

If the data need to be transmitted for long distances in which transmitter and receivers are out of range, wireless bridges and repeaters can be employed. Wireless bridges and repeaters provide long-range point-to-point or point-to-multipoint links. Some of these devices use direct sequence spread spectrum (DSSS) radio technology operating at typical frequencies (i.e., 2.4 GHz). Data may be transmitted at speeds up to 11 Mbps. These bridges comply with standards such as IEEE-802.11b, connecting one or more remote sites to a central server or Internet connection.

On the other hand, industrial plug-and-play or long-range wireless communication devices are offered by many companies such as Microwave Data Systems, Inc. Typical products provide point-to-point and point-to-multipoint communication of data with frequencies from 200 MHz to 2.4 GHz and speeds up to 8 Mbps. Such devices are used for health monitoring and other civil, military, and industrial applications. Networks of low-cost wireless sensors enable monitoring of large civil structures with a large number of sensing nodes.

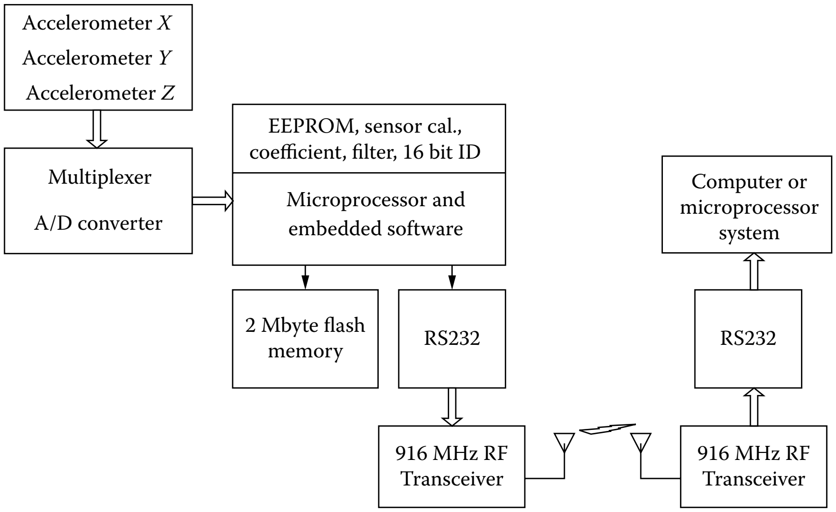


FIG. 3.5l
An intelligent accelerometer sensor equipped with telemetric features.

A typical single-unit data logging transceiver is a wireless data communications device that can serve multichannel sensor arrays to a remote data acquisition system hosted by a computer or a microprocessor. The frequency of transmission is 916 MHz, narrow-band. The RF communication link operates with 19,200 baud, and the device is capable of triggering a sample to be logged (typically from 30 m), or requested data to be transmitted.

A typical application of wireless data logging on a triaxial MEM accelerometer is illustrated in Figure 3.5l. These accelerometers have flash memory (typically 2 to 8 Mbytes). The nodes in the wireless networks may be assigned with 16-bit address lengths, hence are able to connect with thousands of multi-channel sensor clusters. They are powered by 3.1- to 9-V lithium ion AA-size batteries. They draw about 10 mA and contain 10- or 12-bit A/D converters. The dimensions of the total package are 25 mm × 40 mm × 7 mm.

Sensor Networks

Nowadays, geographically distributed sensors can be networked as distributed sensor networks (DSN). The networking requires intelligent sensors that may obey some form of a hierarchical structure.

In recent years, an IEEE-1451 compatible interface between the Internet/Ethernet serial port websensors has been developed. These sensors have direct Internet address. The interface is realized in IEEE-451 Network Capable Application Processors (NCAP). The NCAP connects the Internet through Ethernet. NCAP is a communication board capable of receiving and sending information using standard TCP/IP format. The sensor data are formatted to and from the serial port by one of the followings, RS232, RS485, TII, Microlan/1-wire, Esbus, or I^2C.

The future of sensors will probably be shaped by how they interface into the network to share information. Nevertheless, the importance of sensor networks has been understood by many vendors; hence, the vendors promote proprietary solutions to connect smart sensors into TCP/IP-based networks. As was explained earlier, developments are taking place at both low-level interfaces and at high-level interfaces.

TELEMETRY APPLICATIONS

Telemetry finds a wide range of applications in medicine and life sciences, in industry, in the military, in space exploration, and in other fields.

Medical and Life Science Applications

In recent years, medicine and life sciences have benefited considerably from the use of telemetry techniques. Telemetry has proven to be particularly important in those situations where it is desirable to leave the subjects in a relatively normal physiological and psychological state with minimum interference with their normal pattern of activities. It finds extensive applications in remote locations, where health and safety are the prime concerns. For example, for humans working in hazardous environments, the telemetry of their vital signs is a way to ensure their safety.

Medical Telemetry Medical telemetry is used to detect the regular and/or irregular functioning of the human body. A patient's vital signs, such as electrocardiogram (ECG) waveforms, are monitored continuously and the information is transmitted to a central station for observation and recording. Often wireless telemetry is selected to free patients from being bedridden, thus giving them freedom to move about.

Over the past several years, manufacturers have made tremendous strides in improving the design and functionality of

(1) Bed side monitor (2) handheld transmitters (3) base station

FIG. 3.5m
A telemetric patient monitoring system. (Courtesy of Nihon Kohden Corporation, www.nihonkohden.com)

telemetry systems, allowing more acute patients to become ambulatory during recovery. These systems are now capable of monitoring several different parameters, from ECG waveforms, arrhythmias, and heart rate to pulse oximetry and other vital signs.

Many medical telemetry systems consist of a small digital transmitter worn by the patient, a set of sensors for gathering patient vital signs, a distributed transmitter, and receiving systems. Displays, controls, and recordings are often controlled from a central computer, which is located at the nursing station, where telemetry units dedicated to the diagnosis and treatment of life-threatening illnesses are also located.

A telemetric nurse-based biomedical instrumentation system is illustrated in Figure 3.5m, where (1) shows the telemetry bedside monitor of a patient. It has a 7-inch CRT screen with two waveforms. Vital signs can be graphically or tabularly displayed. The vital signs of patients, such as electrocardiogram (ECG) waveforms, and data can be transmitted, normally wireless, to a remote station for observation and recording.

In Figure 3.5m (2), a patient transmitter for a wireless bedside monitor is illustrated. The base station, in Figure 3.5m (3), can monitor waveforms, data, trends and alarms of up to eight telemetry patients. It has dual-antenna diversity reception, 100-arrhythmia event memory, and a vital signs list.

The use of wireless medical telemetry systems is on the rise worldwide despite a growing problem with electromagnetic overcrowding on the airwaves. Different countries approach congestion of the airwaves in different ways. Some countries assign less crowded regions of the electromagnetic spectrum to telemetry services while others concentrate on the development of modulation and compression techniques to limit bandwidths.

In the United States, the Federal Communications Commission (FCC) is the agency responsible for assigning portions of the electromagnetic spectrum to users other than the federal government. The FCC has issued new rules for medical telemetry in which it assigns portions of the spectrum to telemetry services to operate on an interference-protected basis. In this approach, hospitals will have reasonable assurance that the next generation of medical telemetry devices will operate in this expanded spectrum with minimal interference.

One application of telemetry is the cardiac pacemaker. A bidirectional telemetry process is initiated by the cardiologist when a special signal is sent to the pacemaker. Each pacemaker that is carried by a patient has a different identification code set by the cardiologist before implantation. The typical data from the pacemaker include voltage, impedance, and current consumption of the battery; battery charge level; voltage and current levels; and width, rate, and energy of pulses to record the electrical activity of the heart. All this information can be read and interrogated.

In many cases, especially for patients who are located in rural or isolated areas, monitoring of physiological parameters by wireless transmission is not possible. In these cases, public telephone networks are used for telemetry purposes; the data obtained from the patient is transmitted over telephone lines. However, this approach needs careful attention since the transmission of multiple physiological channels can be problematic in terms of recovering useful information due to frequency band limitations of telephone lines. In this telemetry, speech and data are transmitted simultaneously. This requires an increase in bandwidth, which is achieved by suitable processing devices.

Life Sciences and Bio-Telemetry Telemetry is extensively used in diverse range of life science applications. Telemetry systems equipped with radio transmitters implanted or attached to marine and land animals allow the study of their biological functions and behavior without disturbing their normal way of living. Implanted transmitters have been used to study animals. The transmitter surgically implanted in the body of the animal supplies valuable information on animals' survival patterns.

In life science applications, the range of the transmitter of the telemeter is an important parameter. The range can be from a few centimeters in laboratory experiments to several miles for tracking wildlife. Another critical parameter is the selection of the sensors and transducers that can accurately sense the desired range and at the same time can be protected from corrosion and harmful body fluids. The operating temperature of the sensor can vary from sub-freezing temperatures in the North to the temperature of the deserts. Temperature is important because the life of the batteries powering the telemetry system is largely dependent on the operating temperature.

Figure 3.5n illustrates an implantable sensor/telemetry transmitter to measure the temperature, gross motor activity,

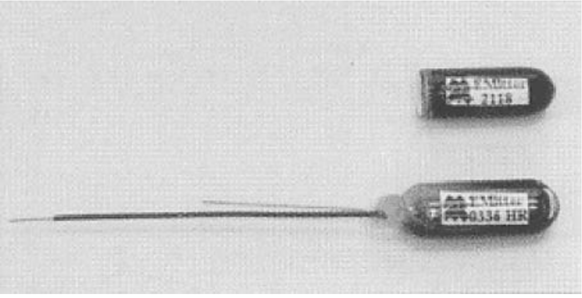


FIG. 3.5n
An implantable sensor/telemetry transmitter. Weight: 1.6 grams, Size: 22 × 8 mm. No batteries required. (Courtesy of Mini Mitter, www.minimitter.com)

and heart rate of an animal. This sensor/transmitter arrangement does not require a battery.

Industrial Telemetry

Telemetry is extensively used in industrial applications. Industrial telemetry enables the automatic and remote monitoring and recording of process variables. By using telemetry, operational efficiencies and safety can be improved. Some applications include oil and gas operations, medical equipment, utilities production and consumption, and intelligent transportation systems.

There are many different types of industrial telemetry systems, since they are offered by a diverse range of vendors competing for the same market. SCADA, which is discussed below, is an example of a well-established telemetry system.

Modern industrial systems use computers, remote terminal units (RTUs), programmable telemetry controllers (PTCs), and so on. PTCs consist of modular intelligent units that can handle multiple inputs (typically 128 I/O) for networking purposes. Remote terminal units can handle virtually thousands of digital and analog inputs and outputs. Some RTUs support local, public switched telephone networks (PSTN), private wire (PW) systems, leased lines (LL), and radio communications.

SCADA Systems SCADA stands for *System Control and Data Acquisition.* SCADA consists of a collection of computers, sensors, and other equipment interfaced by telemetry to monitor and control processes. The uses of SCADA systems are endless as they are only limited by the designer's imagination. Some typical applications include high- and low-voltage power distributions, broadcasting stations, and environmental measurements.

An example of SCADA applications is in water management. Water conservation is critical not only in drought-stricken areas but everywhere, as water is becoming a limited resource due to increases in consumption. Application of SCADA systems enables the remote and effective monitoring and controlling of water levels.

Another application of SCADA in industry is digital pager alarming systems. If an unusual situation occurs in a process, the SCADA system can alert key personnel by sending messages to their pagers. SCADA systems also archive information and generate reports and graphs that are critical to processes.

An important advantage of SCADA systems is that the existing sensors in an application can be incorporated into the overall SCADA system. In doing so, the SCADA system simply adds a new level of intelligence and provides additional capabilities to the existing controls in the plant. In addition to lowering the cost of implementing new technological solutions in existing plants, it can also help to centralize the controls to a central commanding post. SCADA systems can be equipped with alarms so that if one of the subsystems fails, corrective actions can be taken without delay.

FIG. 3.5o
A wireless bridge for SCADA. (Courtesy of Industrial Control Links, Inc., www.iclinks.com)

Figure 3.5o illustrates a ScadaBridge Wireless Remote Terminal Unit (RTU). Such units combine the traditional remote I/O functions of an RTU with a built-in Spread Spectrum radio, thus lowering cost and simplifying installation and support. This particular device serves as a "standard" Modbus Slave RTU in a radio-based SCADA system, or as a wireless I/O bridge that "mirrors" analog and digital signals between two remote locations, providing radio capabilities to non-radio enabled process equipment. The ScadaBridge is "plug-and-play" compatible with SCADA/DCS software packages such as WonderWare, FIX, FactoryLink and Labview — and with PLCs and process controllers.

Space Telemetry

The need to control and gather information from objects put into space is one of the earliest applications of telemetry systems. In the U.S., the current industrial and commercial telemetry standards are largely derived from the telemetry of missiles and other unmanned space objects. Today, telemetry finds extensive applications in space exploration and military use.

In manned space flights, telemetry is used to detect the potential onset of distress and take the necessary precautions. The astronaut's EKG, blood pressure, respiratory rhythm, and other life functions are monitored continuously and telemetered to earth. During the first manned landing on the moon, the voices of the astronauts were combined with 900 other signals, some of them physiological parameters from the astronauts. These complex signals transmitted to Earth, at about 2.2 GHz, were picked up by the ground station in California and relayed to the space flight center in Houston. The different signals were separated and human physiological parameters were extracted as the command ship orbited around the moon.

There are numerous satellite systems around the Earth relying on telemetry systems for spacecraft management. Onboard, they have a Telemetry, Tracking and Command (TT&C) system that supports the daily functioning of the spacecraft. The functions of the TT&C system are numerous. It monitors the performance of satellite subsystems, transmits the data to satellite control centers, supports the determination of orbital parameters, and performs control functions issued by the command center.

The information received from the satellites is used for operational and failure diagnostic purposes. The commonly monitored parameters include voltages, currents, and temperatures of

FIG. 3.5p
Typical space telemetry equipment. (Courtesy of Acroamatics Telemetry Systems, www.acroam.com)

all the major subsystems; the switch status of communication transponders; the pressure in propulsion tanks; and the outputs from altitude sensors.

In modern satellite systems, distributed telemetry systems in modular forms are favored. In this configuration, digital encoders are located in each satellite subsystem and the data from each encoder are sent to a central encoder via a common, time-shared bus. This method reduces the number of wire connections and improves the reliability of the whole telemetry system. Modular design also permits the easy expansion of the initial design and facilitates easy testing during the satellite assembly phase.

Typical space telemetry data processing equipment is shown in Figure 3.5p. This data processor is a virtual machine environment (VME) based system using open architecture and operating systems. It is used in applications where extensive real-time processing of single to multiple streams in complex formats is required. It is capable of real-time data display and recordings.

Other Applications

Apart from medical, life sciences, industrial, and space applications, telemeters are used in many novel applications. For example, vibrations and additional stresses of power lines due to wind and snow can be monitored to prevent fatigue and failure. In this case, the telemetry systems are used as a watchdog before a problem occurs. Some of the telemetry applications are listed below, but such applications are limited only by the imagination of the designer:

- Determination of levels and flow rates of rivers and lakes
- Computer and other equipment fault information systems
- Vending machines connected to mobile phones
- Electric meters interrogated for billing purposes
- Water and gas flow meters monitored for billing and fault diagnosis
- Personal security systems, etc.

Bibliography

Baker, S., “Building Reliability into Your Digital Telemetry System,” *Health Management Technology,* Vol. 23, No. 6, pp. 100–103, 105, 2002.

Boquete, L., Bravo, I., Barea, R., and Garcia, M. A., “Telemetry and Control System with GSM Communications,” *Microprocessors & Microsystems,* Vol. 27, No. 1, pp. 1–8, 2003.

Brooks, T., “Wireless Technology for Industrial Sensor and Control Networks,” SIcon/01, Sensors for Industry Conference, Proc. of the First, ISA/IEEE, Sensors for Industry Conference, 2001, pp. 73–77.

Dunbar, M., “Plug-and-Play Sensors in Wireless Networks,” *IEEE Instrumentation and Measurement Magazine,* Vol. 4, No. 1, March 2001, pp. 19–23.

Eren, H., “Instruments,” In S. A. Dyer (Ed.), *Survey of Instrumentation and Measurement,* New York: Wiley, 2001, pp. 1–14.

Ewing, R. L., and Zohdy, A. A., “Design Approach For Biomedical Smart Sensors,” Proc. of the 44th Symp. on Circuits and Systems, Vol. 1, 2001, pp. 474–477.

Ferrari, P., Flaminni, A., Marioli, D., and Taroni, A., “A Low-Cost Smart Sensor with Java Interface,” IEEE SI/Con, Sensors for Industry Conf. Proc, Houston, TX, 2002, pp. 161–167.

Filipov, A., Srour, N., and Falco, M., “Networked Microsensor Research at ARL and the ASCTA,” IEEE SI/Con, Sensors for Industry Conf. Proc., Houston, TX, 2002, pp. 212–218.

Frenzel, L. E., “After a Slow Start, Bluetooth Shows Its Colors,” *Electronic Design,* Vol. 50, No. 13, 2002, pp. 68–74.

Fukushima R., Kakeda, K., and Fujiyoshi, M., “Intelligent Remote Monitoring and Control System,” *Hitachi Zosen Technical Review,* Vol. 63, No. 3, 2003, pp. 20–23.

IRIG Standards, Telemetry Group. “Test Methods for Telemetry Systems and Subsystems,” Document 118-97. Range Commanders Council, U.S. Army White Sands Missile Range, 1997.

Lee K., “IEEE 1451: A Standard in Support of Smart Transducer Networking,” Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference, Vol. 2, 2000, pp. 525–528.

Lee, K. B., and Schneeman, R. D., “Distributed Measurement and Control Based on the IEEE 1451 Smart Transducer Interface Standards,” *IEEE Transactions on Instrumentation and Measurement*, Vol. 49, No. 3, 2000, pp. 621–627.

Rice, M., and Davis, A., “A Multipath Channel Model for Wideband Aeronautical Telemetry,” IEEE Military Communications Conference Proceedings, Vol. 1, 2002, pp. 622–626.

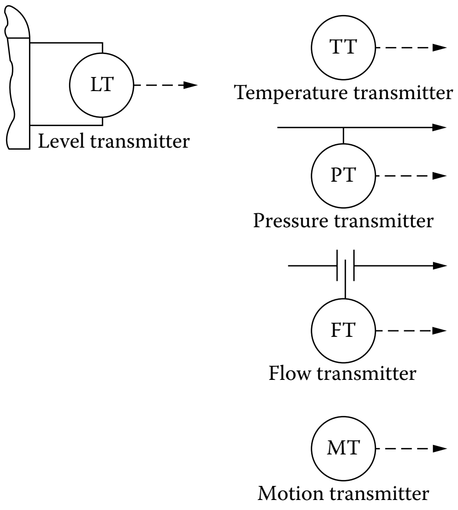
Schildknecht, T., and Reiff, E. C., “Telemetry Station Has Radio and Internet Compatibility,” *Elektrizitaetswirtschaft,* Vol. 102, No.1–2, 2002, pp. 50–53.

Shults, M. C., Rhodes, R. K., Updike, S. J., Gilligan, B. J., and Reining, W. N., “A Telemetry–Instrumentation System for Monitoring Multiple Subcutaneously Implanted Glucose Sensors,” *IEEE Transactions on Biomedical Engineering,* Vol. 41, 1994, pp. 937–942.

Webster, J. G. (Ed.), *The Measurement, Instrumentation, and Sensors Handbook,* Boca Raton, FL: CRC Press, 1999.

3.6 Transmitters—Electronic

A. F. MARKS (1970)
P. B. BINDER, **L. D. DINAPOLI** (1985)
C. G. WALTERS (1995) **D. S. NYCE** (2005)



Flow sheet symbols

Range:

Temperature: Thermocouples can measure as low as –518°F (–270°C) with a K, E, or T-type thermocouple, and as high as 4200°F (2300°C) with a Nanmac model HTP tungsten/rhenium (W/Re) thermocouple. Platinum RTDs can measure from –328 to 1112°F (–200 to 600°C) with Tegam model 8693 probe.

Pressure: Absolute, gauge, and differential pressures. Fullscale ranges from 0.1 in. H_2O (25 Pa) (with Druck model LPM 1000), to 200,000 PSI (1,380 MPa) with Omegadyne model PX91.

Flow: Virtually unlimited. Orifice size can be varied for use with DP cell, electromagnetic induction, paddle, thermal conductivity, etc.; can vary pipe size.

Level: Maximum level is unlimited with bubbler or differential pressure. Radar up to 114 ft (35 m) with Solartron/Mobrey model MRL 700, magnetostrictive up to 100 ft (30.6 m) with MTS Temposonics flexible Level Plus™.

Motion: Full scale ranges from 0.020 in. (0.5 mm) with Schaevitz LVDT model LBB, to over 20 ft (6 m) with MTS Temposonics magnetostrictive models.

Inaccuracy: Fully compensated transmitters are available with error as little as 0.1% of full range over their operating temperature range. Lower cost devices are normally 0.25 to 1%. Transmitters with especially high or low full scale ranges may have errors of 2% or more.

Costs: Process control pressure, flow, and level transmitters in common full-scale ranges cost $500 to $2000. Temperature and motion sensors are typically less than $500. Replacement RTDs and thermocouples are $25 to $100.

Partial list of Suppliers:

ABB (abb.com/instrumentation)
Acromag (www.acromag.com)
Action Instruments (www.actionio.com)
AGM Electronics (www.agmelectronics.com)
Air Monitor Corporation (www.airmonitor.com)
Ametek (www.ametekusg.com)
AVL (www.avl.com)
Barksdale (www.barksdale.com)
Brandt Instruments (www.thermo.com)
Bristol Babcock (www.bristolbabcock.com)
Burkert (www.burkert-usa.com)
CR Magnetics (www.crmagnetics.com)
Danfoss (www.danfoss.com)
Dresser Instruments (www.dresserinstruments.com)
Drexelbrook Engineering (www.drexelbrook.com)

Dwyer Instruments (www.dwyer-inst.com)
Dynisco (www.dynisco.com)
Elan Technical (www.elantechnical.com)
Endevco (www.endevco.com)
Endress + Hauser (www.systems.endress.com)
Fischer & Porter (www.fluidprocess.com)
Fisher Controls (www.emersonprocess.com)
Foxboro (www.Foxboro.com)
GP:50 (www.gp50.com)
Great Lakes Instruments (www.glint.com)
Greyline Instruments (www.greyline.com)
HiTech Technologies (www.hitechtech.com)
Honeywell (www.Honeywell.com/sensing)
Inor Transmitter (www.inor.com)
ITT Industries (www.ittcannon.com)
Johnson Controls (www.jci.com)
Jordan Controls (www.jordancontrols.com)
Jumo Process Control (www.jumousa.com)
Kavlico (www.kavlico.com)
Kulite Semiconductor Products (www.kulite.com)
Love Controls (www.love-controls.com)
Magnetrol (www.magnetrol.com)
Measurement Specialties/Schaevitz Sensors (www.msiusa.com/schaevitz)
Mensor (www.mensor.com)
Minco Products (www.minco.com)
MKS Instruments (www.mksinst.com)
Moore Industries (www.miinet.com)
M-System Technology (www.m-system.com)
MTS Systems (www.mtssensors.com)
Neutronics (www.neutronicsinc.com)
Powers (www.mmcontrol.com)
Pyrometer Instrument (www.pyrometer.com)
RDP Electrosense (www.rdpelectrosense.com)
Revolution Sensor (www.rev.bz)
Ronan Engineering (www.ronan.com)
Rosemount (www.rosemount.com)
Sensidyne (www.sensidyne.com)
Sensitech (www.sensitech.com)
Sensotech (www.sensotech.com)
Setra Systems (www.setra.com)
S-Products (www.s-products.com)
Teledyne Analytical Instruments (www.Teledyne-ai.com)
Teledyne Hastings (www.hastings-inst.com)
Thermo Electric (www.thermo-electric-direct.com)
Transicoil (www.flwse.com)
Vaisala (www.vaisala.com)
Validyne Engineering (www.validyne.com)
Wilkerson Instrument (www.wici.com)
Yokogawa (www.Yokogawa.com)

TRANSMITTER CONFIGURATIONS

Industrial products and processes, commercial products, and automotive products all have requirements to access data that indicate physical parameters either continuously or on demand. Since industrial process control is the target of this work, attention in this chapter is concentrated on the measurement of temperature, pressure, flow, level, and motion. Although many industrial processes require additional measurements, those listed are the most common. To facilitate a closed-loop control system, information from the process must be obtained before a controller can determine what action may be required by a control element. Some popular names for the sensing devices that provide the information are sensors, transducers, and/or transmitters. Other names are also used, but we will define these three for the purposes of this chapter on transmitters.

> A *sensor* is an input device that provides a usable output in response to the input measurand.[1] A sensor is also

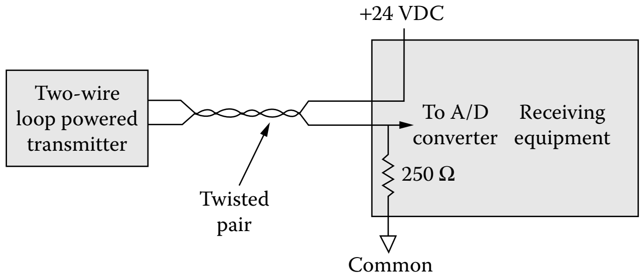


FIG. 3.6a
Wiring configuration of a two-wire, loop-powered 4- to 20-mA transmitter.

> commonly called a sensing element, primary sensor, or primary detector. The *measurand* is the physical parameter to be measured.
>
> An *input transducer* produces an electrical output that is representative of the input measurand. Its output is conditioned and ready for use by the receiving electronics.[1] (The terms "input transducer" and "transducer" can be used interchangeably.)

The receiving electronics can be an indicator, controller, computer, PLC, etc. The term "transmitter," as commonly used with industrial process control instrumentation, has a more narrow definition than those of a sensor or transducer:

> A *transmitter* is a transducer that responds to a measured variable by means of a sensing element and converts it to a standardized transmission signal that is a function only of the measured variable.[2]

Transmitters can have any of several electrical connection schemes. The most common and easiest to use is the two-wire, loop-powered configuration. This is generally the basic configuration assumed by engineers for industrial process control systems when digital communication is not required. As shown in Figure 3.6a, only two wires are used to accommodate both power to the transmitter and output signal from the transmitter.

The loop current is usually 4 to 20 mA, but 1 to 5 and 10 to 50 mA have been used. Important calibration parameters with a current loop are zero, full scale, and span. With the 4- to 20-mA range, the loop current is normally 4 mA when the measurand is at *zero*, and 20 mA when the measurand is at *full scale*. The difference between zero and full scale, 16 mA, is called the *span*. Thus, the span corresponds to the indicated range of the measurand.

When considering a motion transmitter, for example, the range of the measurand could be 0.0 to 100.0 mm, corresponding to a 4- to 20-mA loop current (output span is then 16 mA); the output *scaling factor* is 0.16 mA/mm (which is 100 mm/16 mA).

Two-Wire Loops

The main advantage of a two-wire loop is that it minimizes the number of wires needed to run both power and signal. The use of a current loop to send the signal also has the advantages of reduced sensitivity to electrical noise and to loading effects. The electrical noise is reduced because the two wires are run as a twisted pair, ensuring that each of the two wires receives the same vector of energy from noise sources, such as electromagnetic fields due to a changing current in a nearby conductor or electric motor. Since the receiving electronics connected to the transmitter is designed to ignore common-mode signals, the resulting common-mode electrical noise is ignored.

The sensitivity to loading effects is reduced because the current in the twisted pair is not affected by the added resistance of long cable runs. A long cable or other series resistance will cause a greater voltage drop but does not affect the current level as long as enough voltage compliance is available in the circuit to supply the signal current. The circuit compliance to handle a given voltage drop from additional loop devices depends on the transmitter output circuit and on the power supply voltage.

The typical power supply for industrial transmitters is 24 VDC. If 6 volts, for example, are needed to power the transmitter and its output circuit, then 18 volts of compliance remain to allow for wire resistance, load resistance, voltage drops across intrinsic safety (IS) barriers and remote displays, etc. Where the current loop signal is connected to the main receiving equipment or data acquisition system, a precision load resistor of 250 ohms is normally connected. This converts the 4- to 20-mA current signal into a 1- to 5-volt signal, since it is standard practice to configure the analog-to-digital converter of the receiving equipment as a voltage-sensing input.

Three- and Four-Wire Loops

In contrast to the two-wire current-loop configuration, some current-loop devices require a three- or four-wire connection, as shown in Figure 3.6b. These are not loop powered and

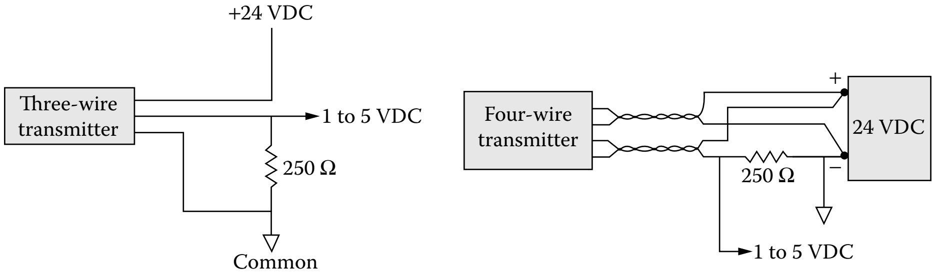


FIG. 3.6b
Three-wire (left) and four-wire (right) current-loop configurations.

therefore have a separate means for providing power by adding one or two more wires.

In a four-wire configuration, the current-loop wires can be a twisted pair, and the power supply wires a separate twisted pair. This preserves the ability to reject electrical and magnetic common-mode interference.

This is not so effective in a three-wire configuration due to the common connection for the return current path. Typically, though, when an instrumentation engineer specifies a current-loop transmitter for industrial process control, it is assumed that a two-wire, loop-powered 4- to 20-mA device is intended. Other data signals may also be impressed upon the same wire pair, or alternatively, various digital communication techniques can be used instead of a current loop. These are described in Section 4.16, which covers field boses and network protocols.

MEASURED VARIABLES

Although almost any type of transducer can be configured as a transmitter, the most common types for industrial process control comprise measurands of temperature, pressure, flow, level, and motion. These are presented here with an explanation of principle of operation. Transmitters for measuring other parameters will have the same possibilities for connection and communication methods, with the main differences being in the sensing element design.

Temperature

A temperature transmitter can utilize a sensing element based on a solid state device (voltage or current output), or thermistor (resistance change) sensing element, but most use either a thermocouple (TC) or resistance temperature device (RTD). Since they are the most popular, thermocouple and RTD-based temperature transmitters are presented here. Thermocouples and RTDs for use with temperature transmitters are available from many sources and will have similar characteristics if ordering the same type.

Thermocouples Thermocouples are one of the most widely used devices for measuring temperature where the temperature probe can contact the body to be measured. They are simple and relatively easy to fabricate. A thermocouple half-circuit is formed by the joining of two dissimilar metals together, usually by welding the tips of the two wires. The point of contact between the two wires is called a thermocouple junction. The metal used for each wire can be a pure element or an alloy.

A voltage potential is developed that can be measured in millivolts across the wire ends opposite the junction. The amplitude of the voltage depends on the difference in electronegativity between the two metals and the difference in temperature between the junction and the other ends of the thermocouple wires. A complete thermocouple measuring circuit includes two junctions. The pair of wires leading back to a voltmeter can be the same metal or alloy as each other, shown in Figure 3.6c.

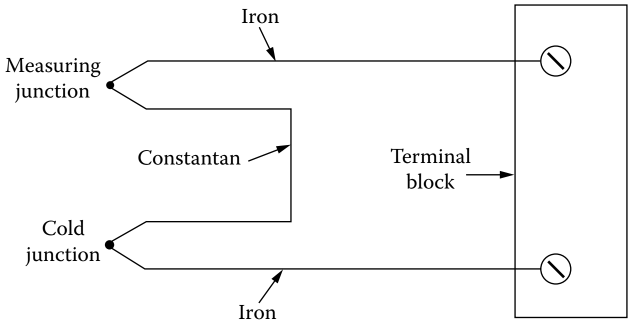


FIG. 3.6c
The measuring and the cold junctions of a thermocouple-type temperature detection circuit.

One junction is the reference, theoretically held at 0°C and called the cold junction. The other junction is incorporated into the measuring probe. Letters such as J, K, and T are used to indicate the metals used. Standard tables list the output voltage generated vs. temperature and thermocouple type, assuming that the cold junction is held at 0°C. (For such tables and much more, refer to Section 4.13 in Volume 1 of this handbook.)

Temperature transmitters do not actually incorporate a “cold” junction but simulate it through the use of an electrical circuit and a special terminal block to accept the thermocouple (TC) extension wires. The terminal block is arranged so that the two terminals will remain at a temperature equal to each other. This provision is needed in order to prevent the introduction of additional thermocouples at the extension wire connection points from affecting accuracy. In the electronic circuit, the cold junction simulator adjusts the zero point for the given thermocouple type, while the gain of an amplifier adjusts the sensitivity or span. In addition, a linearizing circuit is incorporated to improve the accuracy over the full operating range. The most popular thermocouple types are J, K, E, and T, with ranges and conductors as listed in Table 3.6d.

For a tabulation of inaccuracy, repeatability, ambient temperature, and supply voltage effects on a variety of pressure and temperature transmitters, refer to Table 3.6e.

TABLE 3.6d
The Most Common Thermocouple Types, Their Ranges, and Conductors Used

Thermocouple Type	*Conductors Used*	*Temperature Range*
J	Iron-Constantan	–210 to 1200°C
K	Chromel-Alumel	–270 to 1372°C
E	Chromel-Constantan	–270 to 1000°C
T	Copper-Constantan	–270 to 400°C

TABLE 3.6e

Typical Inaccuracies and Other Characteristics of Electronic Pressure and Temperature Transmitters

	All Errors Are in Units of ±% Full Span			
Type of Instrument	*Inaccuracy*	*Repeatability*	*50°F (28°C) Ambient Effect*	*Supply Voltage Change by 1.0 Volt or by 10%*
Absolute PT—				
mmHg range	1%	0.5%	1%	0.1%
PSIA range	0.5%	0.1%	1%	0.1%
Gauge PT—				
below 2000 PSIG (14 MPa)	0.25% or 0.5%	0.1%	1%	0.1%
above 2000 PSIG (14 MPa)	0.5%	0.1%	1%	0.1%
D/P Cell—				
below 500″ H_2O (125 kPa)	0.5%	0.1%	1%	0.1%
below 850″ H_2O (212.5 kPa)	0.75%	0.1%	1%	0.1%
PSID range	0.5%	0.1%	1%	0.1%
Repeaters—				
up to 35 PSIA (242 kPa)	2″ H_2O max (0.5 kPa)	0.2″ H_2O (50 Pa)	1%	N.A.
up to 100 PSIG (690 kPa)	1″ H_2O max (0.25 kPa)	0.3″ H_2O (75 Pa)	1%	N.A.
Filled TT—				
force balance	0.5%	0.2%	0.75%	N.A.
motion balance	0.5%	0.1%	1.0%	0.1%
RTD-Based TT (platinum)	0.15%	0.05%	0.75%	0.02%
TC-Based TT	0.1% or 5 microV	0.05%	See below	0.02%

Ambient error sources include 1) reference junction error of 40 microvolt maximum, 2) span error of 0.5%, and 3) zero error of 0.5%.

PT = Pressure Transmitter; TT = Temperature Transmitter.

RTDs Most RTDs are made with a platinum resistance element and calibrated to have a nominal resistance at 0°C of either 100 or 1000 ohms, although other resistances are available. (For a detailed description of RTDs refer to Section 4.10 in Volume 1 of this handbook).

The resistive element can be made from wire coiled onto a nonconductive form or can be a film deposited onto a nonconductive substrate. The form or substrate is usually made of a ceramic material. The element resistance increases as temperature increases. A constant current is applied, and the voltage drop across the element is measured. The current must be kept low enough so that the power dissipation is not sufficient to appreciably heat the RTD.

If the copper extension wires are long, the resistance reading can be corrupted by the resistance change of the wires due to ambient temperature changes. To prevent this, a four-wire circuit called a Kelvin connection can be used, as shown in Figure 3.6f.

In the four-wire configuration, the current is applied through one pair of wires, and the voltage is measured by the other pair of wires. There is minimal current in the voltage-sensing pair, thus avoiding a voltage drop and preserving the accuracy of the RTD resistance reading. Table 3.6g shows the performance data for standard “smart” RTD transmitters.

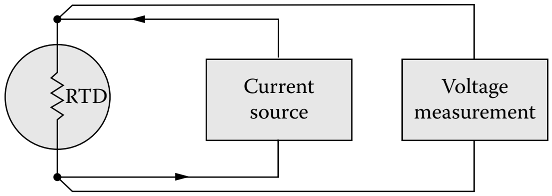


FIG. 3.6f

The four-wire circuit, which is also called the Kelvin sensing circuit, that is used in RTD thermometry.

When using any type of contact probe for the measurement of temperature, it is important to note that this technique does not work well in vacuum. The sensing area (usually the tip) of the temperature probe must touch a thermally conductive medium. Air is acceptable as a thermally conductive

TABLE 3.6g

Performance Data for Standard and "Smart" RTD Transmitters

	*Standard**		*Smart**	
*Performance Criteria**	*Platinum Element*	*Nickel Element*	*DigitalOutput*	*Analog Output (4- to 20-mA DC)*
Inaccuracy	± 0.15% or ±0.15°F (0.08°C)	± 0.25%	± 0.035% or ± 0.18°F (0.1°C)	± 0.05% or ± 0.18°F (0.1°C)
Repeatability	± 0.05%	± 0.05%	± 0.015% or ± 0.18°F (0.1°C)	± 0.025% or ± 0.18°F (0.1°C)
Zero Shift/6 mo.	± 0.1%	± 0.2%	± 0.06% R or	± 0.1% R or
Span Shift/6 mo.	± 0.1%	± 0.4%	0.18°F (0.1°C)	0.18°F (0.1°C)
Supply Voltage Variation	± 0.2% or 0.02°F (0.01°C)		—	(0.005%)/Volt
Ambient Effect (100°F or 55°C)	± 0.75%		Included above	± 0.1%

*When two values are given the error is the higher of the two. When % is given it refers to % of span or % of calibrated span, except if %R is shown, which means % of actual reading.

medium for measuring relatively slow temperature changes, but heat is not conducted in a vacuum. A vacuum chamber is heated by radiation (or conduction through some other medium, such as the chamber walls), so another sensing method should be used in vacuum. One example of a preferred method would be to aim an infrared detector at an interior surface.

Pressure

Industrial pressure transmitters measure fluid pressure, either gas or liquid, while using any of several types of sensing elements. (For a detailed discussion of all the pressure sensors used in the processing industry, refer to Chapter 5 in Volume 1 of this handbook.)

The most popular pressure sensors are the strain gauge, LVDT, capacitive, resonant wire, and inductive or variable reluctance types.

Strain Gauges Strain gauges can be wire, bonded or unbonded foil, thin film, or semiconductor. In any of these types, the resistance varies with strain, although capacitance- and inductance-based gauges have also been built. Figure 3.6h provides a pictorial representation of a bonded foil-type strain gauge and a diffused semiconductor type.

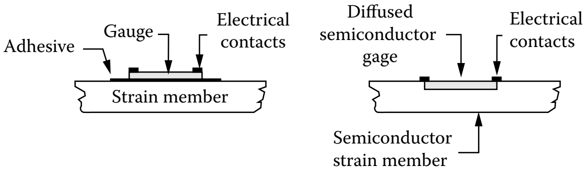


FIG. 3.6h

Strain gauge configurations: bonded foil (left) and diffused semiconductor (right).

In the bonded-foil gauge, the adhesive also provides electrical insulation from the strain member. Due to its small size and relatively low cost, a diffused semiconductor sensing element is the most common strain gauge implementation. Since an adhesive is not required, it is easier to manufacture than the bonded type, but has an intrinsic temperature sensitivity that results in both strong zero and span shifts. Accordingly, effective means for temperature compensation must be employed.

The compensation scheme usually corrects for nonlinearity of the sensing element in addition to correcting temperature errors. A diffused semiconductor strain gauge pressure transmitter can be designed so that the fluid directly wets the semiconductor diaphragm if the measured media are noncorrosive, but an isolation diaphragm of stainless steel or other metal is usually used to ensure compatibility with various measured media. The space between the metal diaphragm and the semiconductor sensing element is filled with a liquid (usually a degassed silicone oil) to transmit the pressure from the diaphragm to the sensing element.

With pressure transmitters utilizing LVDT, capacitive, inductive, or variable reluctance sensing elements, a pressure-sensing metal diaphragm, capsule, or bourdon tube is exposed to the pressure. The diaphragm, capsule, or bourdon tube converts the pressure into motion. Then the motion is measured

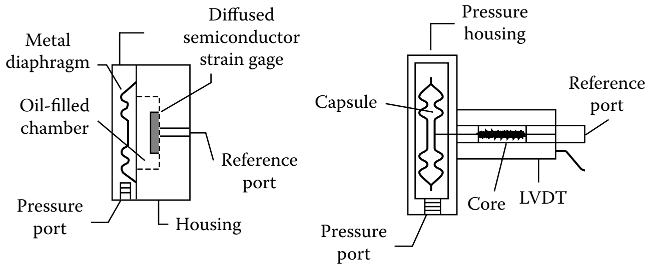


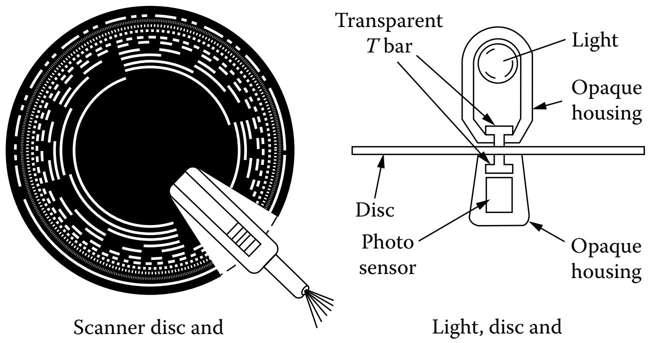
FIG. 3.6i
Mechanical pressure sensor configurations: diffused semiconductor strain gauge (left) and LVDT with pressure capsule (right).

and converted into an electronic signal by the sensing element and associated electronics module. Figure 3.6i shows some representative configurations for pressure sensing.

LVDT The diffused semiconductor strain gauge was shown in Figure 3.6h. A pictorial representation of an LVDT (linear variable differential transformer) is shown in Figure 3.6j. An LVDT normally comprises three coils of wire, arranged in a line as shown. The center coil is the primary and is driven by an AC waveform, usually a sine wave at a frequency of 500 Hz to 10 kHz. The circuit generating the drive for the primary coil and receiving the signal from the secondaries is called a signal conditioner. A ferromagnetic core is positioned within the bore of the bobbin on which the coils are wound.

The two secondaries are connected in series so their voltages subtract. When the core is centered, the voltage of each secondary has the same amplitude but opposite polarity, so the output voltage is approximately zero. As the core moves one way or the other, the output becomes more positive or negative respective to the direction. In order to produce this bipolar output, a demodulating circuit must be used to change the output voltage of the secondaries to DC.

There are many circuits to do this, and so they are not shown here, but they can be passive or synchronous. A passive demodulator uses diodes, capacitors, and resistors and does not rely on the phase of the primary voltage. Synchronous demodulators use semiconductor switches that conduct with a timing depending on the phase of the primary.



Scanner disc and photocells

Light, disc and photocell assembly

FIG. 3.6k
Photoelectric encoder transducer.

A synchronous demodulator has better performance than a passive one when the cable to the LVDT is longer than a few inches from its signal conditioner.

Photoelectric Transducers Figure 3.6k shows a typical schematic of a photoelectric transducer where the position of the photocoder is proportional to the motion of a primary sensing element. Light from the source shines through perforations in the shutter to energize photoelectric cells. The output of these cells is scanned and the pulses are amplified to produce a digital signal or they are rectified to produce a DC analog signal.

Capacitive Transducers Figure 3.6l shows a capacitance-type pressure or differential pressure transducer, which is of the motion-balance type. Positioned between two fixed capacitor plates is a highly prestressed thin metal diaphragm. This forms the separation between two gas-tight enclosures that are connected to the process. The difference in pressure between the two chambers produces a force that causes the diaphragm to move closer to the fixed capacitor plate of the low-pressure chamber. The transducer is excited by a 10-kHz voltage with an amplitude proportional to the difference in pressure.

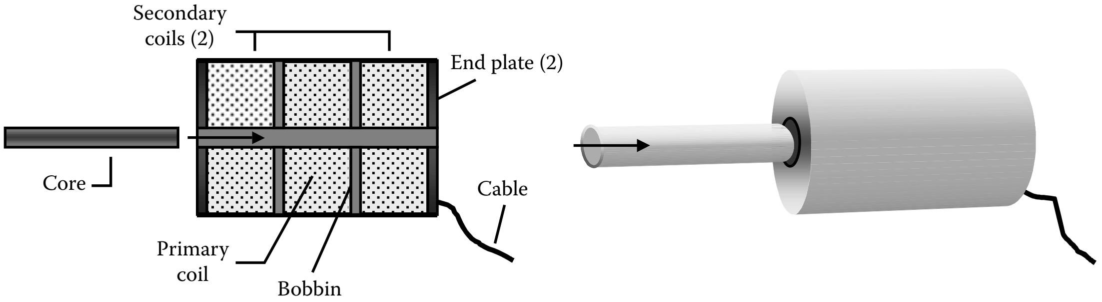


FIG. 3.6j
Cutaway view of an LVDT-type pressure sensor (left), and the outside view of the same sensor showing the core as it is about to enter the bore (right).

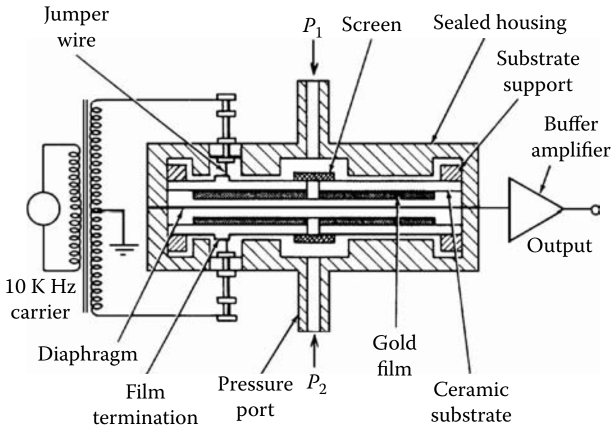


FIG. 3.6l
Capacitance pressure transmitter.

In some designs, the transducer is filled with an inert fluid (usually silicone oil) to prevent contamination of the capacitor plates or transducer interior by the process fluid or gas. A metal barrier diaphragm keeps the fill fluid in the transducer and the process out of the transducer.

Capacitive Semiconductor Transducers The capacitive semiconductor transducer is very similar to the metallic capacitance transducer except that the sensing diaphragm is micromachined from silicon instead of metal. Figure 3.6m shows a silicon capacitive cell. The semiconductor transducer has almost no hysteresis and has excellent elasticity. The sensing diaphragm is sealed between two rigid plates to complete the capacitor cell. Silicon is normally sealed with low-melting-point glass to create a hermetic seal, instead of being welded like the metallic cell.

Potentiometric Transmitters Figure 3.6n shows a potentiometer (resistance) driven by a bourdon tube in a manner similar to the movement of a pressure gauge. Rotation due to a pressure change turns the shaft of a precision potentiometer, and the change in resistance is proportional to the process pressure.

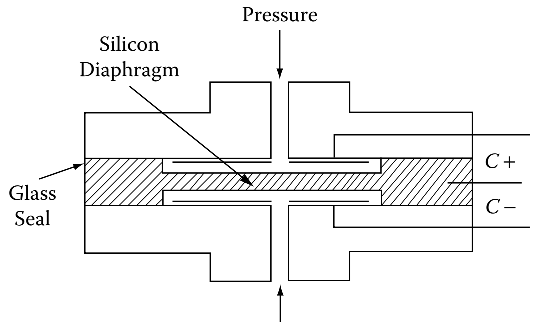


FIG. 3.6m
Silicon capacitive cell.

Pressure Measurement Types Pressure measurements can be gauge, absolute, or differential. A pressure-sensing diaphragm, for example, has two sides. To measure gauge pressure (PSIG), one side (the reference port) is exposed to the ambient atmospheric pressure, and the other side (the measuring port) is exposed to the measurand. Thus, the measurement is made relative to the ambient pressure.

The reference port may or may not be accessible to the user. To measure absolute pressure (PSIA), the reference port is sealed off while containing a vacuum. To measure differential pressure (PSID), both ports are connected by the user and the pressure difference is indicated. A pressure transmitter specially designed to accept a wide common mode pressure range and high overpressure is called a DP cell.

Flow

Flow transmitters are widely used in all types of process control applications. (For a detailed discussion of the various flow detectors, refer to Chapter 5 in Volume 1 of this handbook.)

Airflow is measured to monitor and control the performance of fans and dampers. Liquid flow is measured to control boilers, chillers, heat exchangers, and the supply of liquid components and final product. Common methods employ measuring the pressure drop across a section of pipe, duct, orifice plate, venturi, or pitot, as well as direct methods using a paddle, thermal conductivity, vortex shedding, or turbine.

A paddle-type flow transmitter is shown in Figure 3.6o. The paddle, which is positioned in the flow stream, receives increasing force as the flow increases. This force can be measured by a strain gauge, or the paddle can be allowed to move with the force and then measured by a motion sensor.

Process flow transmitters utilizing a pressure drop combine a standard pressure transmitter with a pressure drop–generating element that is placed in line with the flow. Some of these are depicted in Figure 3.6o. When a fluid flows through an orifice, a pressure differential develops across the two sides of the orifice. The amount of pressure drop varies as the square of the flow rate, so a square root function is added within the transmitter or within the data acquisition system to return a linear output reading vs. flow rate.

In a pitot arrangement, a port perpendicular to the flow measures a reference pressure, and a port angled into the flow measures a pressure that increases with increased flow due to the impact pressure of the fluid. A venturi arrangement provides a more linear change in pressure vs. flow rate and provides less obstruction of flow than would an orifice.

Pitot turbine meters are inserted into the flow to detect its velocity, while full flow turbine meters are more accurate because the full flow passes through the turbine blades. Turbine meters can measure the flow rate by detecting the RPM developed. In this case a small magnet in the impeller can induce voltage pulses into a nearby pickup coil. The pulse rate then represents the flow rate.

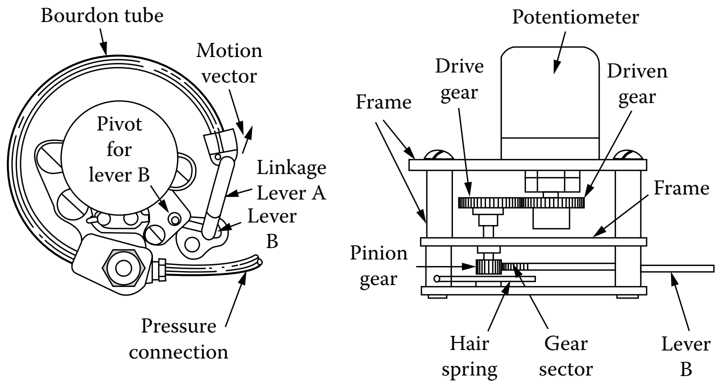


FIG. 3.6n
Potentiometric transmitter.

When using thermal conductivity to measure flow, a reference temperature sensor (nonheated) on the upstream side provides a reading that is subtracted from the heated temperature sensor. The temperature difference decreases with increased flow.

Flow transmitters can also detect the speed of sound in the measured fluid, electromagnetic induction, and the float position of rotameters. To measure fluid flow depending on the speed of sound in a medium (the fluid), a sonic wave is caused to propagate in the fluid along the path of the flow to be measured. With zero flow, a known amount of time will elapse as the wave travels from the sender to the receiver. This time will increase as the flow velocity in the opposite direction to that of the waves rises or will decrease if they travel in the same direction. The amount of time measured indicates the flow rate. This varying rate of wave propagation due to the relative difference in velocity between the sender and the medium is called the Doppler effect.

When using electromagnetic induction (called the Faraday effect) to measure the flow of electrically conductive liquids, the fluid flows in a magnetic field. Electrodes that are electrically insulated from the pipe pick up a voltage potential with a magnitude depending on the flow rate.

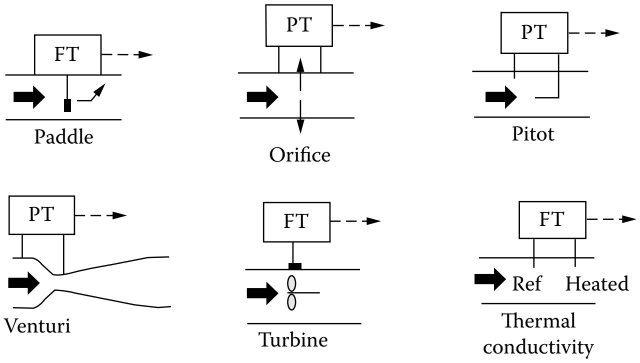


FIG. 3.6o
Six flow transmitter designs, each operating on a different working principle.

Rotameters can be instrumented with magnetic, inductive, optical, or other sensors so that the position of the float within the measuring tube is measured. If a rotameter with a glass measuring tube is used, local visual indication is also available.

Level

Transmitters for measuring liquid level are commonly used for process control, but also for inventory control and custody transfer. (For a detailed discussion of this topic, refer to Chapter 3 of Volume 1 of this handbook.) Level measurement for solids is not as accurate or reliable, due to the nonuniform settling that often occurs.

A liquid level measuring range can be less than 1 meter for a small reaction vessel, or may be more than 60 ft (20 meters) for a petroleum storage vessel. The highest accuracy can be obtained by using a contact method, such as a liquid level float measured by a magnetostrictive transmitter. Sometimes a contact method is not suitable due to the sticky nature of the product being measured, such as tar. Noncontact measurements can be made with transmitters employing radar or ultrasonic waves.

A simple but always popular way to indicate liquid level is through the use of a sight glass, but this typically provides no electronic output for use by automated equipment.

A variation of a sight gauge uses a closed metal pipe that houses a float containing a permanent magnet. As the float follows the liquid level in the vessel, the float position is detected by an externally mounted magnetostrictive transmitter. The transmitter then provides an accurate electronic reading of the level. Several level transmitter designs are shown in Figure 3.6p.

A simple and popular system for measuring liquid level in vessels open to atmospheric pressure is the bubbler shown in Figure 3.6p. Here, a small amount of airflow is introduced into a tube that extends down to the tank level from which the head measurement is to be made. Bubbles slowly escaping from the tube end ensure that the entire tube is full of air at the same pressure as at the bottom end of the tube. A gauge pressure transmitter measures the air pressure and thus indicates the liquid head.

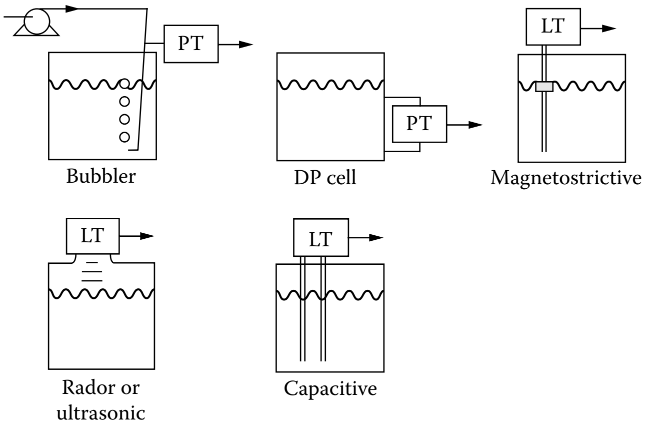


FIG. 3.6p
A number of level transmitter designs.

A differential pressure transmitter can also be connected to pressure taps at different elevations that have a known distance between them. This configuration allows the detection of level if the liquid density is known and that measurement is independent of vessel pressure. This installation has an advantage over the bubbler in that the level can be determined when the vessel is pressurized at various amounts of pressure during normal operations.

A magnetostrictive level transmitter can have a flexible probe, which is easy to carry up a ladder to the top of a large vessel. The probe tip is attached to the vessel bottom by a magnet, weight, or clamp. A float rides on the liquid level. A second float can be weighted to sink through oil, for example, and float on the layer of water near the vessel bottom. This allows measurement of the amount of water, and the amount of oil is the top float position minus the water float position.

These sensors can measure with high resolution. Inside each float is mounted a permanent magnet. The probe contains a magnetostrictive wire, called the waveguide, in which current pulses are periodically introduced. The interaction of the magnetic field from the current pulse with the magnetic fields from the float magnets generates sonic waves in the waveguide at the location of each float. The sonic waves travel to the transmitter end of the waveguide at a velocity depending on the waveguide material.

When a current pulse is introduced into the waveguide, a timer is started. When a sonic wave is detected at the end of the waveguide, the timer is stopped. The measured time indicates the position of the float, and thus the liquid level. If more than one float is used, then multiple timers are used.

In a capacitance-level probe, the relative permittivity (also called the dielectric constant) of the measured medium must be substantially different from the surrounding environment. The probe responds with a variable capacitance, depending on how much of the probe length is submerged or adjacent to the measured medium.

Motion

Popular transducers used in motion transmitters include the LVDT, magnetostriction, and inductive types. The LVDT is good for relatively short strokes but becomes less accurate and more expensive with measuring lengths over a few inches. Magnetostrictive linear motion transmitters are highly accurate and available with lengths of less than a foot, to well over 20 feet. Inductive types are practical for lengths of up to a few feet.

A simple LVDT is constructed with three coils of wire wound onto a bobbin. A ferromagnetic core passes within the bore through the center of the bobbin. The center coil is the primary and is powered from a stable with an AC source, as shown in Figure 3.6q.

The two secondaries are typically connected in series bucking, so the voltages subtract. Synchronous demodulation is also used (with semiconductor switches instead of diodes) for better performance. As the core moves within the bore, the output voltage becomes more positive with movement in one direction and more negative with movement in the other direction. The output voltage vs. input position is relatively linear, in the range of 0.1 to 0.5%.

A magnetostrictive linear position transducer operates as was described earlier. Instead of mounting within floats, the magnet assemblies have means to mount to the member to be measured. They are called position magnets.

At one time, it was not thought possible to operate a magnetostrictive transducer with the low power needed for a

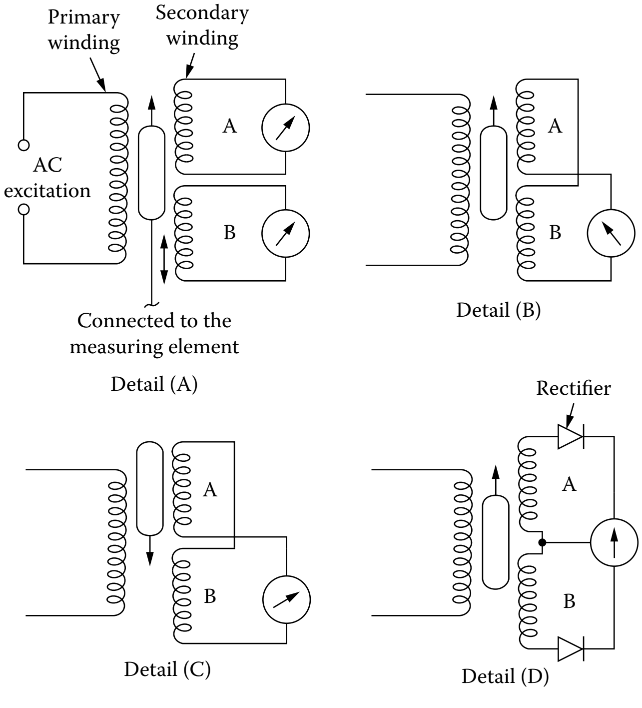


FIG. 3.6q
Linear variable differential transformer.

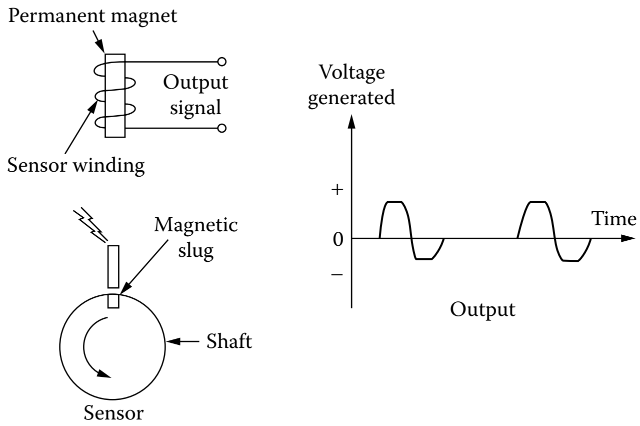


FIG. 3.6r
Magnetic tachometer sensor.

two-wire transmitter. This problem was solved in 1990, and several companies now offer motion transmitters based on this technology.[3] Magnetostrictive linear position transmitters can have a nonlinearity of less than 0.01%.

Inductive motion transducers are constructed similarly to LVDTs, but as few as one coil may be used. As the core moves farther into the coil, the inductance increases. Then, the inductance can be converted into a signal voltage by using an oscillator and measuring the inductive reactance. The nonlinearity error is in the range of 0.2 to 0.5%.

The AC tachometer is an example of a device that generates variable frequency signals for transmission. Here a magnetic slug, gear tooth, key in a shaft, or other device is counted as pulses created in a coil (Figure 3.6r). When the magnetic slug comes across the face of the coil, a voltage is generated in one direction as it leaves.

These pulses are a direct measurement of the speed of shaft rotation. When counting gear teeth where the space between them is similar to the width of the teeth, a smooth AC signal without wide gaps between pulses is generated. The design of the sensor probe has to be made with knowledge of the tooth shape, size, etc., so that good waveforms can be generated for easy measurement.

IMPORTANT FEATURES

Inaccuracy

The basic performance criteria that define the error limitations of a transmitter are contained in a specification called the static error band. This includes the combined effect of nonlinearity, hysteresis, and repeatability. The next group of errors are contributed by calibration error, temperature-induced errors, and line/load regulation. There are many additional performance criteria that can be specified, but they typically are related to the specific type of transmitter, so they will not be presented here.

Nonlinearity *Nonlinearity* is expressed as a percent of full range output (%FRO) and is determined from the maximum difference between a datum on the output vs. measurand plot and a best straight line (BSL) drawn through the data. Nonlinearity error can be specified in several different ways, including independent, zero based, and terminal based.

With independent nonlinearity, the BSL does not have to go through the origin or any particular datum. It can be found using a computer program with an iterative process, adjusting the slope and intercept of a straight-line equation until the errors are minimized. A more popular way to find an independent BSL, while avoiding an iterative process, is to use a least-squares approximation.

This function is available on most calculators and spreadsheet programs. It provides a BSL approximating that achievable with iteration but less perfectly if there are any strong irregularities in the data. The least-squares approximation is the most commonly used method because it can be directly calculated.

In zero-based nonlinearity error specifications, the straight line goes through the origin, but the full-scale point is adjusted (changing the slope of the line) to minimize the error. This is used when it is most important to have an accurate zero indication.

A terminal-based BSL is drawn through both the origin and the full-scale datum. This normally produces the highest nonlinearity error number for a given device but may be preferred when no manual or automated adjustments are possible after installation.

Hysteresis *Hysteresis* in a process transmitter is the maximum difference between readings taken at the same measurand with upscale and downscale approaches to the readings. It is due to inelastic qualities of the sensing element, friction, and backlash. When linkages are used to connect the measurand to the sensing element, hysteresis can be increased. A related property is called deadband, which is the range through which the measurand may be varied by reversing direction without producing an observable response in the output signal.

Repeatability *Repeatability* is the difference between consecutive readings taken at the same measurand, approaching from the same direction, other conditions remaining constant. As long as the repeatability error is low, an intelligent system can reduce the errors from other sources by using the known characteristics of those errors and compensating accordingly.

Zero Suppression and Elevation

There may be some measurement values that are of no interest in a transmitter installation for process control, even though that range falls within the operating range of the transmitter. In some cases it will be desirable to ignore this unimportant area in order to improve the readability within the range of interest.

If, for example, a temperature transmitter is rated for a range of 0 to 1000°C but readings of less than 200°C are of no interest, a *zero suppression of 20%* can be used. Then the 4.0 to 20.0 mA output signal would correspond to a temperature range of 200 to 1000°C, yielding a higher sensitivity (in mA/°C).

Conversely, if the same transmitter were set instead for a *20% zero elevation,* a 7.2- to 20.0-mA output signal would correspond to a temperature range of 0 to 1000°C (the lower end of the current range would be increased by 20% of the 16-mA span). Zero elevation may be used in order to match the output signal current to that expected by another piece of equipment.

Turndown

When one part of a process operates in the 90- to 100-PSIG range, while another part operates within a 15- to 20-PSIG range, an engineer might specify the installation of two different pressure transmitters, each having the appropriate full-scale range.

Ordering the transmitters and then stocking spare parts can become a logistical problem in a processing plant where hundreds of pressure transmitters are installed. A preferred method is to specify a pressure transducer that can provide the required accuracy over several ranges.

This can be done in the cited case if the transducer is designed to operate over a full range of 0 to 100 PSIG but also can be readjusted (turned down) to read accurately at 20 PSIG. In order to do this effectively, the sensing element used in the transmitter must be highly accurate, so that the resulting accuracy after turndown will still be adequate to provide the needed performance.

The ratio of the highest full scale to the least full scale that can still deliver the rated accuracy is called the turndown ratio. Turndown ratios of 10 are common, but some up to more than 100 are also available. This feature can be used with any type of process transmitter and is commonly used with pressure transmitters.

External Calibration

In older equipment, it was often required that the instrument cover be removed to facilitate calibration or other adjustment. For explosion-proof equipment, this required a safety tag and appropriate means to ensure safety during the operation, often meaning that the process had to be shut down completely.

Many modern transmitters allow for calibration without removing the instrument cover. Some have zero and span adjustments under a nameplate, which is easily slid to the side to provide the access. This provides the convenience of allowing adjustment while the only tool needed is a screwdriver.

Some transmitters have provision for the use of a handheld calibrator that plugs into an electrical connector or uses magnetic or optical coupling to enable the digital connection. This method has more flexibility but requires the often-proprietary handheld calibrator.

Still others, such as those using the HART protocol, provide means for uploading or downloading information digitally over the same pair of wires operating the 4- to 20-mA signal. Digital communications will be discussed later in this section.

Intrinsic Safety

When the transmitter will be used in a hazardous location, where flammable or explosive gases, vapors, fibers, or powders may be present, a protection method must be employed. Popular methods include mounting the transmitter within an explosion-proof housing, in a purge cabinet, or as part of an intrinsically safe system.

In an explosion-proof housing, it is possible that the enclosed atmosphere may burn or explode, but the resulting pressure is withstood by the strength of the housing. A flame path allows the high-pressure gas to escape while being cooled. This prevents ignition of the external atmosphere and bleeds away the pressure so the worker is not injured when the housing cover is subsequently removed.

When the transmitter is mounted in a purge cabinet, fresh air flows through the cabinet at a sufficient rate to prevent the buildup of a flammable mixture. Both explosion-proof and purged equipment relies on the installation method to provide the safety features. Intrinsically safe installations, however, require the transmitter itself to be designed within specific safety criteria.

A transmitter can be installed into an intrinsically safe system through the use of an approved safety barrier device. In the U.S., the approval agency is usually Factory Mutual (FM) or Underwriters Laboratories (UL) but can be one of the other smaller agencies if acceptable under local regulations where the equipment will be installed. The safety barrier is installed in the safe area (typically, the control room), in series with the wiring between the safe area and the hazardous area (the process).

The safety barrier limits the electrical current and voltage to levels that are too low to provide energy sufficient to ignite the type of hazardous materials expected to be present. In addition, the transmitter must be designed so that there is no energy storage, arcing, or hotspot sufficient to cause ignition. The specific requirements are listed in the National Electrical Code, and the publications of FM, UL, and NFPA (National Fire Protection Agency).

DIGITAL COMMUNICATION

As described earlier, the 4- to 20 mA two-wire current loop is the assumed standard for analog communication for industrial process control. In addition, there are several popular methods for digital communication from process transmitters. Some of these are HART, CAN, and PROFIBUS.

These and other digital networks and buses are described in full detail in Volume 3 of this handbook.

The HART (highway addressable remote transmitter) protocol was originally developed by Rosemount in 1986. It is now owned by the HART Communication Foundation. It utilizes a standard current loop installation, but allows digital signals to be impressed onto the same pair of wires. The digital signals do not affect other equipment that may be using the 4- to 20-mA signal.

The digital words are placed on the lines using a frequency shift keying (FSK) method according to the Bell 202 telephone standard. A logical zero is 2200 Hz, and a logical one is 1200 Hz. This allows information to be sent from the transmitter to the receiving equipment, as well as allowing the receiving equipment to send signals to the transmitter for calibration, configuration, etc. It is also possible to disable the current loop signal, if desired, and use only the digital communications. Since it uses Bell 202 FSK, it is a relatively slow communication method.

CAN was introduced by Robert Bosch, GmbH in 1986 and originally intended for automotive use. It stands for controller area network and is a much higher-speed system than HART, but it uses a different hardware layer. Power is provided by one pair of wires, and bidirectional signals are sent and received on a second pair.

CAN is a bus system and several devices can be connected to the same sets of wires. It can operate in a master–slave mode, or in a mode where any device can take the bus when needed. Bus contention is solved by a priority level assigned to each device. A popular implementation of CAN is called DeviceNet.

PROFIBUS is an open digital communication system that uses application profiles, thus receiving the PROFI part of its name. The profiles are provided by manufacturers and users to define performance and other features of the device. These profiles are downloaded by the connected equipment to interpret how to interface with the device.

The hardware interface is by an RS 485 connection, using two wires for bidirectional signals. Another pair of wires is used for power. There are several different implementations of PROFIBUS that provide tradeoffs among baud rate, power consumption, transmission distance, fiber optic, and intrinsic safety.

There are many other protocols for digital communications with transmitters and industrial instrumentation. The most popular ones change over time, and new ones are added. The reader can also investigate Foundation Fieldbus, MODBUS, IEEE 1451, and Ethernet, which are all described in detail in Volume 3 of this handbook.

TRANSMITTER SELECTION

In addition to selecting the proper range and accuracy rating for a transmitter installation, it is important to keep in mind a few additional criteria for proper selection. The materials of construction of the transducer elements that will contact the process must be specified to avoid corrosion by or contamination of the measured media.

Sometimes, one must use a transmitter design with a robust housing. For example, it is common for technicians to use pressure transmitters mounted to vessel walls as steps while climbing up to perform routine maintenance and inspection duties.

When the transmitter will be exposed to wide changes in temperature during normal operation, the temperature sensitivity specification of the transmitter may become the most important specification regarding total accuracy of the system.

Many communication protocols are now in common use. Consideration must be given to the cost of providing the hardware connection (wiring) as well as the suitability for providing the desired information. A HART device may communicate more slowly than a CAN device but can use existing current loop wiring.

When using intrinsic safety barriers in a 4- to 20-mA current loop, make sure that the sum of voltage drops for the transmitter, readout devices, IS barriers, and load resistor add to less than the power supply voltage available for the loop.

Intelligent Transmitters

Although Section 3.10 in this chapter deals with smart transmitters in detail, they should also be mentioned in connection with digital communications. The main contribution of the addition of a microprocessor to the transmitter has been the ability to calibrate the unit over a much wider range than the actual span needed for the particular application. This resulted in much increased rangeability without sacrificing accuracy because by memorizing the temperature and pressure effects on zero and span the smart transmitter can automatically correct for these variations, and therefore the performance of the unit is only a function of repeatability, linearity, and hysteresis.

In addition to lower error and higher rangeability, the smart transmitters are also more flexible. Because their calibration curve is in the microprocessor's memory, one can electronically change the zero and the span of the transmitter through the keyboard of a portable terminal (Figure 3.6s), and the microprocessor will automatically match the minimum and maximum output signals to the newly set measurement inputs without affecting instrument calibration.

Desirable Features Checklist

The most desirable features are high resolution, accuracy, rangeability, reliability, and low cost. In order to meet these requirements, a transmitter should be designed to have the following features:

1. Small size and weight for easy installation and maintenance.
2. Rugged design to withstand the industrial environment.

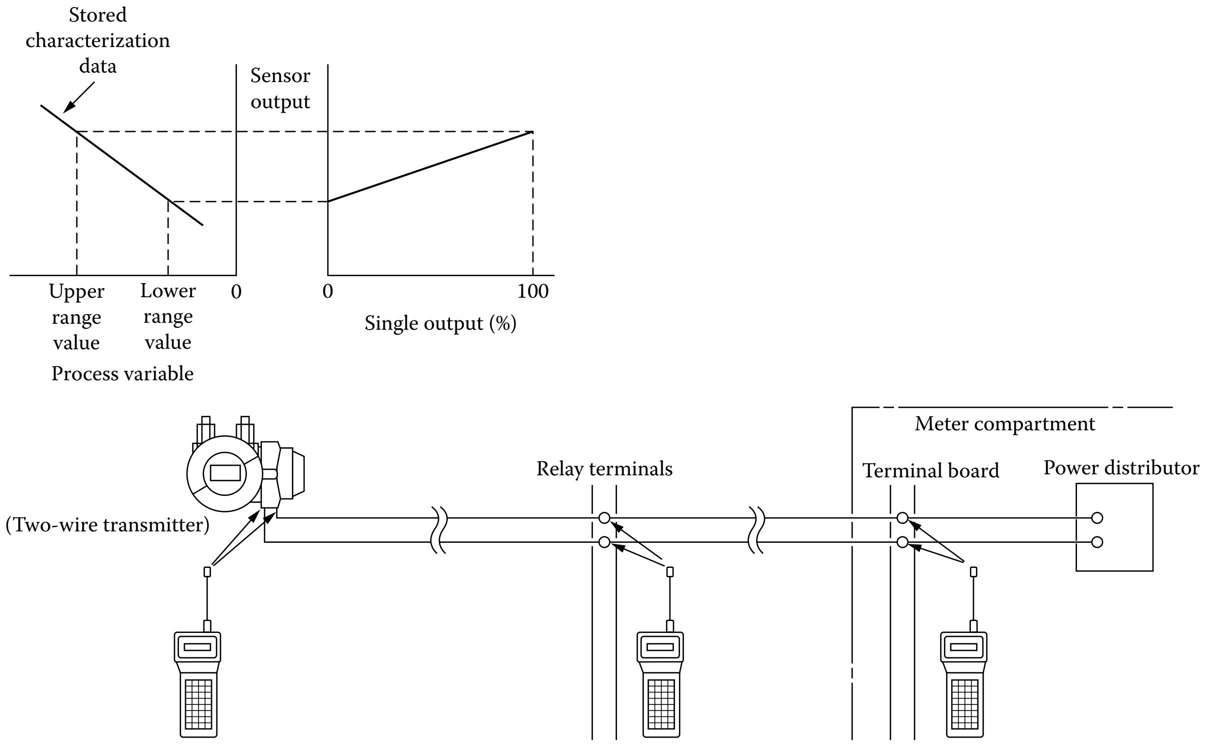


FIG. 3.6s
Portable terminal can be used to re-range transmitters from a number of locations. (Courtesy of Yokogawa Electric Corp.)

3. Minimum dependence on environmental conditions for accuracy; this requires good temperature stability, resistance to barometric change, weatherproofing, etc.
4. No need for adjustment due to load or line resistance variations.
5. No potential hazard to personnel or equipment in explosive atmospheres. Low-voltage operation with limited current capacity assists in eliminating these problems. By definition, intrinsically safe equipment provides all three kinds of protection.
6. Convenient and accurate field calibration and maintenance. In electrical transmitters this usually means conveniently located test terminals. In explosion-proof designs, calibration should be possible without opening the housing.
7. Capacity to operate during voltage dips and power outages. Generally, this is accomplished elsewhere in the control system.
8. Minimum number of transmission and power wires. In most systems installed today, the transmitter is powered from the receiver over the same wires that are used for transmission.
9. Output compatible with both measuring and controlling instruments.
10. Optional local indication of output signal.
11. Circuitry designed to facilitate troubleshooting and maintenance. Most systems today are solid state (no tubes) and are mounted on circuit boards or are encapsulated. Plug-in components make fast repairs easy, but encapsulated modules are considered to be throwaway items.

Ability to be integrated into DCS systems and, in the case of smart transmitters, to communicate over the data highway provided with the particular system.

References

1. Nyce, D. S., “Linear Position Sensors, Theory and Application,” New York: John Wiley & Sons, 2003.
2. ANSI/ISA Standard S51.1-1979, “Process Instrumentation Terminology,” 1993.
3. Nyce, D. S., U.S. Patent 5,070,485, “Low Power Magnetostrictive Sensor,” 1991.

Bibliography

Blake, H., “Transmitters and Transducers with a Purpose,” paper presented at AGA conference on transmission, 1968.

Brindley, K., *Sensors and Transducers,* Oxford, England: Heinemann Professional Publishing, 1988.

Carr, J. J., *Sensors and Circuits,* Englewood Cliffs, NJ: PTR Prentice Hall, 1993.

Carstens, J. R., *Electrical Sensors and Transducers*, Englewood Cliffs, NJ: Regents/Prentice-Hall, 1992.

Demorest, W. J., “There’s More to Transmitter Accuracy Than the Spec,” *Instruments and Control Systems,* May 1983.

Fraden, J., *Handbook of Modern Sensors*, New York: Springer-Verlag, 1996.

Herceg, E., *Handbook of Measurement and Control,* New Jersey: Schaevitz Engineering, 1976.

Magison, E. C., *Intrinsic Safety,* Research Triangle Park, NC: Instrument Society of America, 1984.

Minar, E. J. (Ed.), *ISA Transducer Compendium,* latest edition, Research Triangle Park, NC: Instrument Society of America.

Morris, H. M., "Microsensors Enhance Process Variable Transmitters' Abilities," *Control Engineering,* October 1991.

Norton, H., *Handbook of Transducers,* Englewood Cliffs, NJ: Prentice-Hall, 1989.

Nyce, D. S., "Magnetostriction-Based Linear Position Sensors," *Sensors,* 11(4), 1994.

Nyce, D. S., Position Sensors for Hydraulic Cylinders, Hydraulics & Pneumatics, November 2000.

Nyce, D. S., U.S. Patent 5,070,485, "Low Power Magnetostrictive Sensor," 1991.

Nyce, D. S., U.S. Patent 6,401,883, "Vehicle Suspension Strut Having a Continuous Position Sensor," 2002.

Oliver, F. J., *Practical Instrumentation Transducers,* New York: Hayden Book Co., 1971.

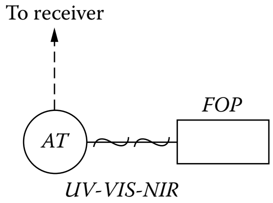
Pallas-Areny, R., and Webster, J. G., *Sensors and Signal Conditioning*, 2nd ed., New York: John Wiley & Sons, 2001.

Webster, J. G., The *Measurement, Instrumentation, and Sensors Handbook*, Boca Raton, FL: CRC Press, 1999.

Winch, R. A., "Utility Puts Smart Transmitters to the Test," *InTech Magazine,* August 1990.

3.7 Transmitters—Fiber-Optic Transmission

V. J. MAGGIOLI (1985) **J. C. HUBER** (1995) **B. G. LIPTÁK** (2005)



Flow sheet symbol

Costs: Transmitters are in the $2500 to $4500 range. Microprocessor-based temperature transmitter with PID included is $5000 to $6000. Same unit with 2, 4, 6, or 8 channels is about $15,000. Wand probe $1000 with tip to $2500 for spectra-caliper probe; fiber-optic spectrophotometer systems range from $2000 to $20,000.

Types:
A. Fiber-optic cable
B. Fiber-optic connectors
C. Fiber-optic receivers, transmitters, probes
D. Industrial fiber-optic modems and networks

Partial List of Suppliers:
ABB-Bomem Inc. (C) (www.bomem.com)
Accufiber Div. Of Luxtron Corp. (C) (www.luxtron.com)
ADC Broadband Connectivity (D)
Alcatel (D) (www.alcatel.com)
Allen-Bradley Co. (D) (www.ab.com)
AMP (A, B, C, D)
AT&T (A, B, C)
Banner Engineering Inc. (C) (www.bannerengineering.com)
Barber-Colman Industrial Instruments (C) (www.barber-colman.com)
Belden Div. Of Cooper Industries (A, B, D) (www.belden.com)
Bran & Luebbe (C) (www.branluebbe.com)
Brinkmann Instruments (C) (www.eppendorfso.com/search.asp)
Cabletec (D) (www.cabletec.com)
Corning (D) (www.corning.com)
Dolan-Jenner Industries (C) (www.dolan-jenner.com)
Edmund Scientific (C) (www.edsci.com)
FLIR (C) (http://www.flir.com/)
Fotec (D) (www.fotec.com)
Guided-Wave (C) (www.guided-wave.com)
Hart Scientific (C) (www.hartscientific.com)
Hewlett-Packard Co. (A, C) (www.hp.com)
Honeywell (C) (www.acs.honeywell.com)
Indigo Systems (C) (http://www.indigosystems.com/)
Ircon (C) (http://www.ircon.com)
Land Instruments (C) (http://www.landinst.com)
LT Industries Inc. (C) (www.LTIndustries.com)
Lucent Technologies (D) (www.lucent.com)
MetriCor Inc. (C) (www.metricorinc.com)
Mikron Instruments Co. (C) (http://www.mikroninst.com/)
Motorola (C) (www.motorola.com)
Northern Telecom (A, B, C) (www.northerntech.com)
Ocean Optics Inc. (C) (www.oceanoptics.com)
Omega/Vanzetti (C) (http://www.vanzetti.com/)
Optical Cable Corp. (A, D) (www.occfiber.com)

Perstorp Group (C) (www.perstorp.ce)
Phoenix Digital (D) (www.phoenixdigital.com.au)
Raytek (C) (http://www.raytek.com/)
Siecor (A, B, C) (www.siecor.com)
Square D, Infrared Measurement Div. (C) (www.squared.com)
Technology Dynamics Inc. (C) (www.technologydynamicsinc.com)
Thermo Electron (C) (www.thermo.com)
Tyco Electronics (D) (www.tycoelectronics.com)
Wahl (C) (http://www.palmerinstruments.com/wahl/wahl.html)
Williamson Corp. (C) (www.williamsonir.com)
3M Company (A, B, D) (www.3mtouch.com)

The topic of this section is fiber-optic transmitters, but there is a fair amount of overlap between this subject and related ones in the three volumes of this set of handbooks. In the first volume Section 4.5 covers fiber-optic temperature transmitters and Section 8.23 discusses fiber-optic probes. In the third volume of this handbook, the subject of fiber-optic networks and cables is discussed in Section 4.18, which also covers the topics of installation, inspection, and testing.

INTRODUCTION

Optical fibers are commonly used in communications because they cost less than copper due to their multiple pathways and because of their good distance and high bandwidth capabilities. Although such characteristics are valuable in most applications, the overwhelming reason for using fiber optics in industrial communications is their immunity to electromagnetic interference (EMI).

Today, few plants can operate without extensive data processing and control. Sensors, solenoids, switches, motors, and pumps are all monitored and controlled by computers and programmable logic controllers. These devices are interconnected by communication cables, which used to be made of copper wire. Often, the copper cables perform adequately. But as data rates increase, the data signals are frequently interrupted by electrical interference. When these interruptions can be very expensive, it is vital to control such interference.[1,2]

Industrial transmitter signals are subject to many sources of electrical noise. Most plants move large quantities of product at high speeds, which requires large electric motors to operate conveyors and pumps. These large electrical loads create electromagnetic waves and ground surges. In addition, many plants are composed of several structures, which usually means separate power distribution systems, resulting in differences in ground potential. Finally, the communication cables between structures tend to attract lightning strikes. These are some of the reasons why alternates to electrical transmitters are being considered.

FIBER OPTICS BASICS

Almost all data processing and control systems use digital signals. The distinguishing feature of fiber optics is that the digital signals are optical pulses rather than electrical. Except under the most intense conditions, light is unaffected by electrical fields; therefore, it is an ideal medium for data transmission in automatic process control. Optical fibers conduct light in a manner similar to the way wires conduct electricity. The physical laws are different, but the results are quite similar. A typical optical fiber is shown in Figure 3.7a.

The center of the fiber is called the core. It conducts the light, similarly to the conductor in a wire. The outer part of the fiber is called the cladding. It keeps the light in the core, similar to the insulator on a wire. Outside the cladding is a protective polymeric coating that gives the fiber greater mechanical strength and endurance.

Both the core and the cladding will pass light, but they have different densities. When the light passes from a higher- to a lower-density material, it changes speed, which in turn changes its direction. This is called *refraction.* This change of direction is a function of 1) the light velocity difference in the two materials (*refractive index*) and 2) the angle at which the light approaches one medium from the other (*angle of incidence*). As the angle of incidence increases, the angle of refraction also rises until the *critical angle* is reached. At this angle, the light is reflected back into the originating medium and does not enter the other medium at all (Figure 3.7b).

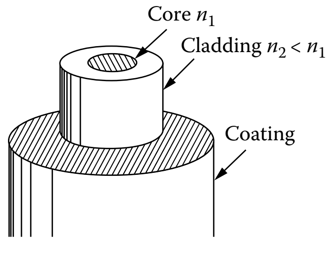


FIG. 3.7a
Optical fiber construction.

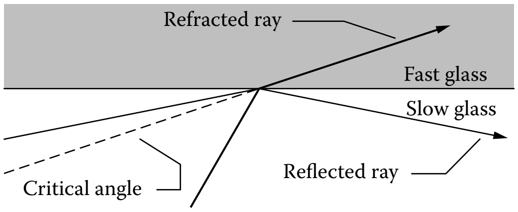


FIG. 3.7b
Refracted and reflected light rays and the critical angle.

Both the cladding and the core are transparent. For cladding with a lower index of refraction than the core, light within the cone will be totally reflected back into the core. The greater the difference between the indexes of refraction of the core (n_1) and the cladding (n_2), the larger is the cone of light (theta) that the fiber will handle. The size of this cone of light is called the numeric aperture (*NA*); it is calculated by Equation 3.7(1) and is shown in Figure 3.7c.

$$\mathrm{NA} = \sin\theta = \sqrt{n_1^2 - n_2^2} \qquad n_2 < n_1 \qquad \mathbf{3.7(1)}$$

Numeric aperture is one of the most important parameters in a fiber-optic system. If light enters the fiber core at an angle greater than the critical, it will be reflected at the core-cladding boundary back into the core and will travel down the core in a zigzag pattern. When this is achieved (Figure 3.7d), the light signal stays within the core of the fiber and travels down to the end of the cable, where a receiver converts the light into an electrical signal.

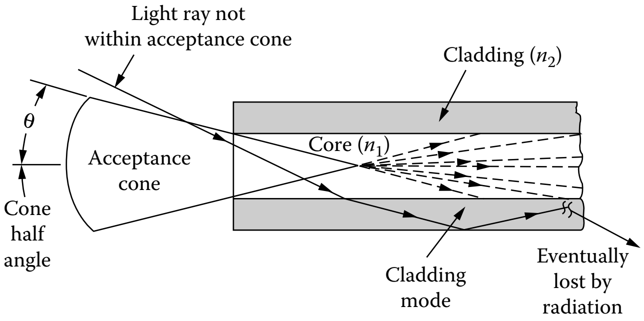


FIG. 3.7c
In order for the light to be fully reflected, its entering angle must be within the "acceptance cone." If the angle exceeds the critical, some of the light will be refracted and lost.

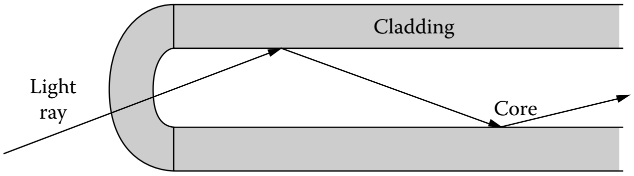


FIG. 3.7d
If the light entered at an angle exceeding the critical, it will travel down the optical fiber without attenuation.

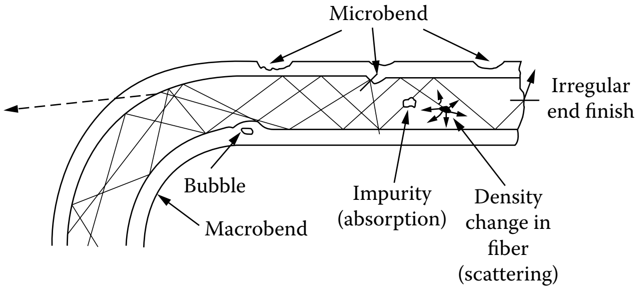


FIG. 3.7e
The sources of light attenuation include macro-bends, micro-bends, bubbles, impurities, fiber density variations, and irregular end finish. (Courtesy of 3M Company.)

All optical fibers are subject to a number of sources of light attenuation. These sources are shown in Figure 3.7e. In industrial applications, the most important causes of light attenuation are macro-bends and dirt. As is shown later in this section, a large numeric aperture helps reduce these effects.

SYSTEM COMPONENTS

The fiber-optic system consists of the transmitter, the cable, and the receiver (Figure 3.7f). The sensor associated with a fiber-optic transmitter can be a temperature probe (Section 4.5 in Volume 1) or a variety of other probe designs (Section 8.23 in Volume 1). The fiber-optic transmitter contains a driver and a light source, and it converts the measured process variable signals into light signals. The driver converts a standard electrical signal to a digital pulse that modulates the light. The light source is an LED (light emitting diode) or a laser, which is turned on and off by the driver.

The data are transmitted in a digital form over the optical fiber by turning the light on to signal 1s and turning it off to signal 0s. The frequency of light is seldom varied because it is less expensive to install more fibers. Because most communications are two-way, most fiber cables are used in pairs, with

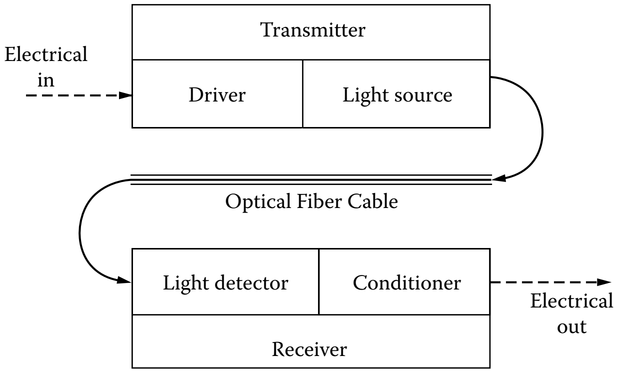


FIG. 3.7f
The three main components of a fiber-optic system are the transmitter, the cable, and the receiver.

one carrying the signal from the transmitter to the receiver and the other carrying another signal back to the transmitter.

The receiver is the interface, which converts the light signal back to an electrical one. It consists of a light detector and a conditioning circuit. The optical-to-current converter is usually a photodiode. The conditioning circuit checks that each bit of information is a binary 1 or 0 and amplifies the signal from the light detector. After that the binary signal is converted to the form required by the control system.

Fiber Sizes and Properties

There are many standard sizes of optical fibers. Sizes are specified in microns by stating the diameter of the core and the diameter of cladding. Standard sizes are 9/125, 50/125, 62.5/125, 85/125, 100/140, and 200/230. All these fibers, including the 200/230 fiber, have been standardized in IEC 793–2 Ammendment 1, "Optical Fibers," Part 2: Product Specifications, Issue 1991–04. The most widely used sizes are 9/125, 62.5/125, and 200/230. Their relative sizes are shown in Figure 3.7g.

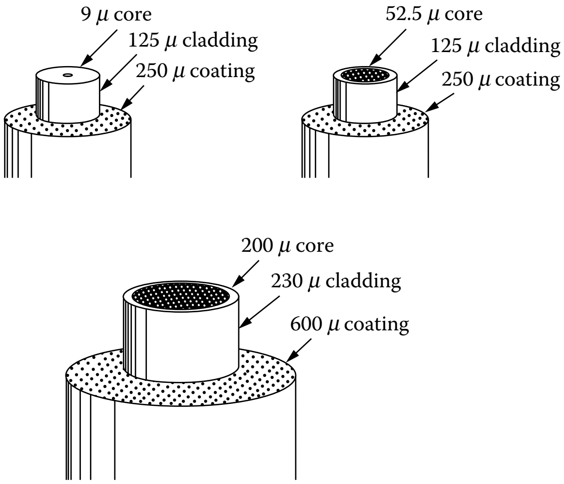


FIG. 3.7g
Sizes of the most widely used optical fibers. (Courtesy of 3M Company.)

When installing an optical fiber network, choosing the fiber size is mostly a matter of balancing the requirements for communications performance and ease of installation and maintenance. Table 3.7h lists the fiber sizes, numeric apertures, most common applications, data rates, and connector installation times.

Generally speaking, fibers with smaller cores tend to have smaller attenuations and larger bandwidths, which are required for their long-distance applications. Fibers with larger cores tend to have larger numeric apertures and greater tolerance to dirt or bending, and the time to install their connectors is faster.

Step index fibers have a constant index of refraction from the center of the core to the edge of the cladding and then a single step down to the index of refraction of the cladding. Graded index fibers have a smooth downward gradation of index of refraction from the center of the core to the edge of the cladding. Step index fibers collect twice as much light as graded index fibers do.[3] This factor is called coupling efficiency. These factors are shown in Figure 3.7i.

Single- and Multi-Mode Fibers

The "mode" is the light path followed by the light as it travels down a fiber. If the light travels straight at the axis of the core, because the core diameter is so small (8 or 9 microns) that the light has only one path, it is called a *single-mode* design. In single-mode fibers the light travels a shorter distance and therefore reaches the receiver sooner. Single-path fibers are not much used in the process industry, because their termination and connection are difficult due to their smallness. Their main applications are in long distance (up to 75-mile) communications at very high bandwidth (10-Gbps) applications.

In *multi-mode* designs, the light is reflected off the core-cladding interface and travels in a zigzag as shown in Figure 3.7j. Here, the different modes cause the rectangular light pulse to spread out, a phenomenon called *dispersion.* The multi-mode fiber has a 50- to 200-μm diameter, which causes dispersion and limits the allowable transmission distance and bandwidth. On the other hand, because of its larger

TABLE 3.7h
Optical Fiber Characteristics as a Function of Their Sizes

Size	*NA*	*Application*	*Data Rate*	*Connector Installation*
9/125	0.14 to 0.16	Telephony	1 to 2 Gbps	40 to 50 minutes
50/125	0.20 to 0.23	Data communications	125 Mbps	15 to 30 minutes
62.5/125	0.27 to 0.29	Data communications	125 Mbps	15 to 30 minutes
85/125	—	Substantially obsolete		
100/140	—	Substantially obsolete		
200/230	0.37 to 0.39	Factory automation	20 Mbps	4 to 5 minutes

*Gbps = 10^9 bits/sec, Mbps = 10^6 bits/sec, each one and zero is called a bit, and a binary number of eight bits is called a byte.

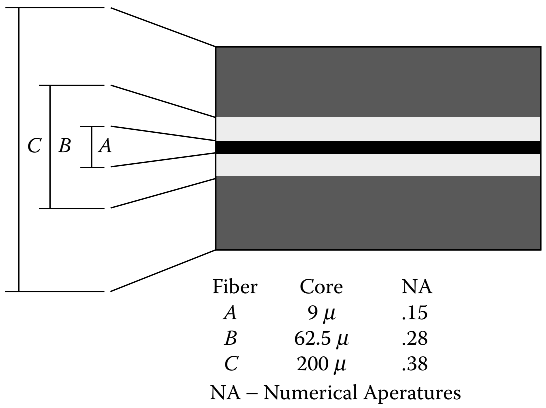


FIG. 3.7i
Fiber size and acceptance cone. (Courtesy of 3M Company.)

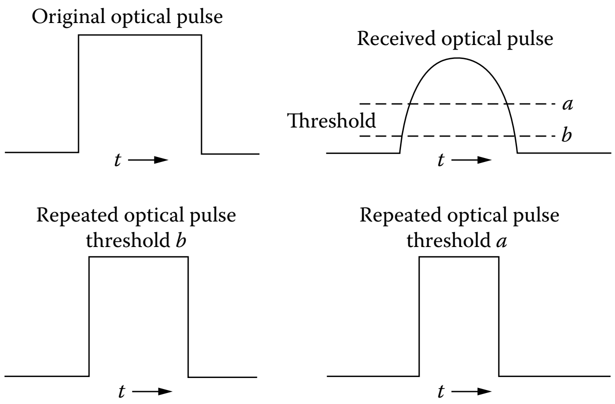


FIG. 3.7k
Pulse width distortion increases as the threshold rises, but so does noise immunity. (Courtesy of 3M Company.)

diameter, termination is easier and the light source is less expensive. The multi-mode fibers used in industry can handle 200-Mbps data rates over distances up to about 1 mile. The most common core-cladding diameter for multi-mode is 62.5/125 μm.

Pulse Width Distortion

In programmable logic controller installations, a common application is to have one master controller with a number of remote input/output (slave) modules. (PLC system design is discussed fully in Chapter 5 of this handbook.) In one common layout, the remote I/O simply taps onto the copper cable. When a fiber-optic system is used, each remote I/O repeats the signal. This means that a remote I/O string of 20 units would thus have 19 repeaters.

Pulse repetition is an important aspect of digital signal transmission. The receiver must be able to identify the point at which the signal has risen sufficiently above the noise to be certain that the leading edge of a pulse has arrived. It must also be able to determine when the signal has fallen enough to be certain that the trailing edge of the pulse has arrived.

Then a new, repeated pulse is placed between these two event times. As shown in Figure 3.7k, if the edges of the received pulse have dispersed from their original square shape, different threshold levels will result in different pulse widths. This is called pulse width distortion.

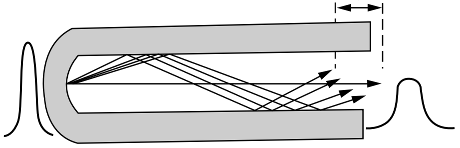


FIG. 3.7j
Modal dispersion and distortion of the optical pulse in a multi-mode fiber.

Although exaggerated in this example, it can be readily seen that a chain of many repeaters can distort the pulse sufficiently to cause communication errors. Many PLCs can tolerate 20% pulse width distortion. If each repeater causes 1% distortion, the maximum number of repeaters is 20. Clearly, pulse width distortion must be kept small.

Fiber Selection

There is no one perfect fiber for all applications. Tradeoffs and compromises must be made for each application. Table 3.7l lists the some of the primary fiber selection criteria. The 200/230 fiber was specifically designed for factory automation applications, while the 62.5/125 μm size is popular for industrial process control applications.

There are substantial differences between factory automation and other data communications applications; the most important difference is that the installation and maintenance in a factory are usually done by onsite personnel, primarily because interruptions in a factory can be very expensive and thus repairs must be performed immediately.

TABLE 3.7l
Fiber Selection Criteria

Fiber Size (μm)	*9/125*	*62.5/125*	*200/230*
Longest distance between repeaters	>3 km	<3 km	<2 km
Greatest bandwidth at longest distance	>1 GHz	<200 MHz	<10 MHz
Environmental tolerance	Very clean	Clean	Dirty
Installation and maintenance skill	Expert	Specialist	Electrician

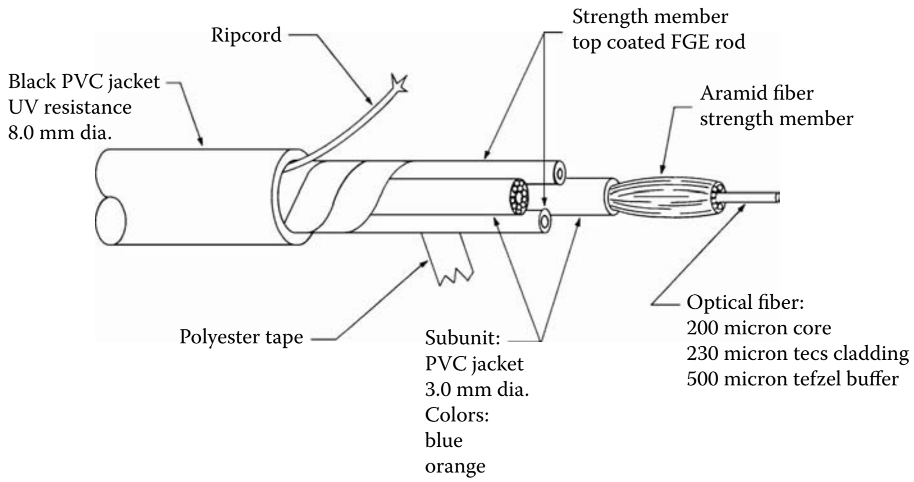


FIG. 3.7m
Tight buffer fiber-optic cable. (Courtesy of Belden Div. Cooper Ind.)

Other reasons for using onsite personnel for maintenance relate to proprietary processes, safety considerations, and union rules.

Onsite personnel usually are electricians by training, and most of their daily work is with large-gauge power and communications wiring. The work site is often cramped, dimly lit, hot or cold, and dirty: a catwalk is a typical example of these conditions.

Fiber-Optic Cable

Two basic types of fiber-optic cable constructions exist: tight buffer and loose tube. Selecting the cable construction depends on the application.[8]

Indoor and short-distance outdoor applications use a tight buffer construction, shown in Figures 3.7m and 3.7n. This construction supports the fiber and therefore usually has greater crush and impact resistance, suitable for cable tray applications. Also, it usually has strength members for each fiber, and connectors can be attached directly. Tight buffer construction cables can be obtained for flame-retardant, aerial, and cable tray installations.

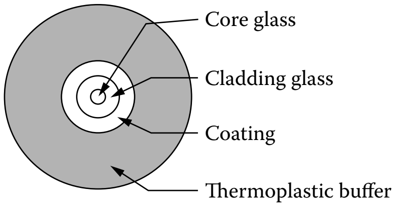


FIG. 3.7n
Cross-section of a tight-buffered cable.

Long-distance outdoor applications typically use a loose tube construction, shown in Figures 3.7o and 3.7p. In the loose tube construction, the fiber is loosely laid inside a plastic tube, which is usually filled with a water-repellent gel. This construction insulates the fiber from stresses on the cable such as thermal expansion and contraction. It usually has less attenuation in outdoor, long-distance applications. Loose

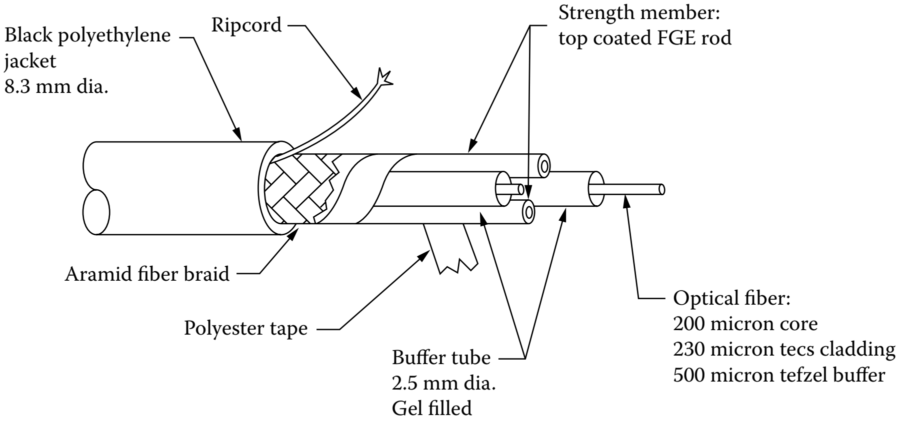


FIG. 3.7o
Loose tube fiber-optic cable. (Courtesy of Belden Div. Cooper Ind.)

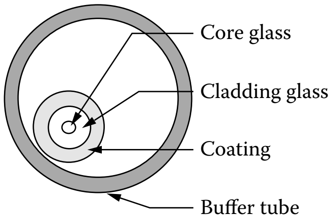


FIG. 3.7p
Cross-section of a loose-tube fiber optic cable.

tube construction cables can be obtained for aerial, conduit, and buried installations. This construction lacks strength members on each fiber and consequently requires accessories to attach connectors.

Fiber-Optic Connectors

Since fiber-optic communications first became practical in the early 1980s, at least six connector designs have come into standard usage. Some of these designs are shown in Figures 3.7q and 3.7r:

The SMA, ST, and SC connectors are commonly used in data communications and are the principal candidates for industrial applications.
ST is a steel bayonet-style connector that is most widely used in industrial plants.
SC is a plastic rectangular connector that is officially recommended to replace the ST style.
MT-RJ is a duplex-style connector that contains both transmit and receive fibers in a single, small form-factor connector.
VF-45 is a competing duplex-style connector that also contains transmit and receive fibers in a single connector similar to an RJ-45 phone jack.
The biconic, FC, and D4 connectors were designed for long-distance telephony and are rarely used in data communications.

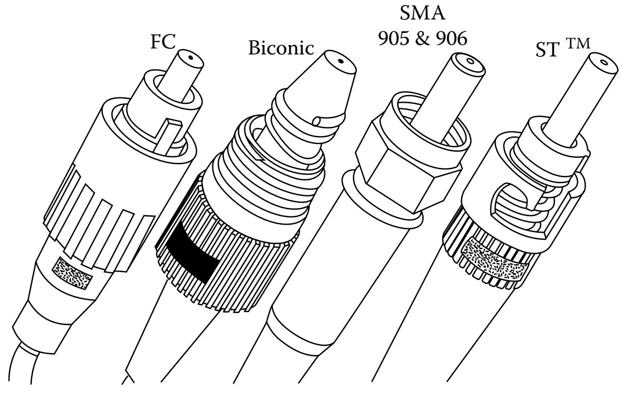


FIG. 3.7q
Fiber-optic connectors. (Courtesy of GTE Products Corp.)

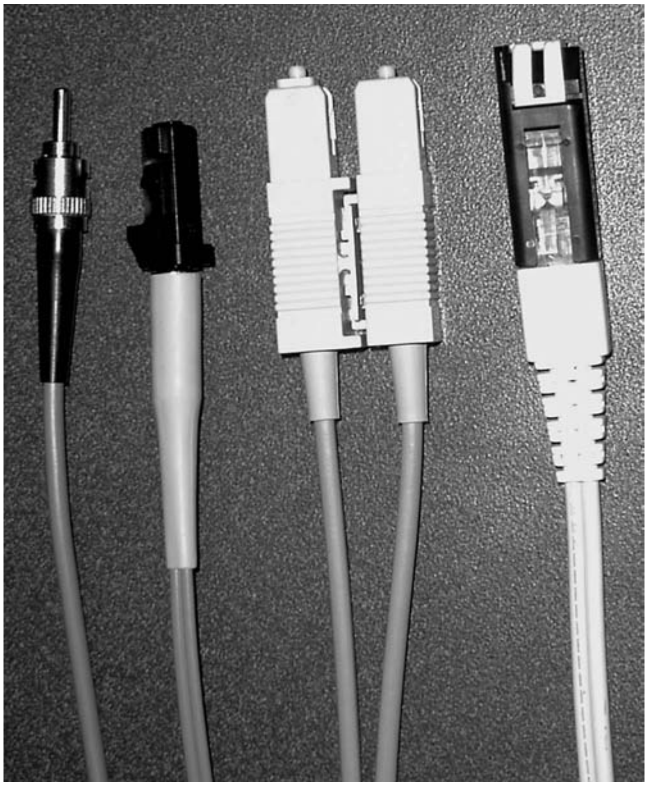


FIG. 3.7r
ST, MT-RJ, SC, and VF-45 fiber-optic connectors.

The typical method of mounting connectors on the other fibers uses adhesives and polishing. These techniques require substantial training and experience to be achieved in the installation times shown in Table 3.7h.[5] Further, they require at least 2 square feet of work space to lay out the equipment.

Each of these connector designs has advantages and disadvantages; no single design is the best for all applications. The automated factory and process control environment is quite different from the office or telephone-related environment. The SMA connector was designed with a dirty environment in mind, and its performance in industrial applications has made it one of the common connectors. Connector design selection criteria are summarized in Table 3.7s.[6]

Attaching the Connectors For 200/230 fiber, the combination of large core and thin cladding allows the use of connectors that can be installed with simple tools. Their crimp-and-cleave operation requires no adhesive and no polishing. It takes less than 5 minutes to install a connector, including the time to prepare the cable. All the installation operations are handheld and do not require work space to lay out equipment.[7]

There are basically two methods of attaching the connectors to the end of each fiber. One is the splicing of factory-made pigtails. These are short cable segments that have factory-finished fiber-optic connectors on one end and are not terminated

TABLE 3.7s
Connector Design Selection Criteria

	Unit	*SMA*	*ST*	*SC*
Maximum clearance (ferrule to receptacle)	μ	28	17	10
Loss due to 10-μ particle on ferrule side				
62.5/125 fiber	DB	0.35	0.70	1.00
200/230 fiber	DB	0.06	0.10	Not available
Loss due to 20-μ particle on fiber core				
62.5/125 fiber	DB	1.03	1.03	1.03
200/230 fiber	DB	0.10	0.10	Not available

at the other end. The unterminated end is field spliced to a cable needing termination. If loose-tube fiber is being used, pigtail splicing is the fastest way to terminate a cable.

The other method of attaching connectors is field connectorization. In this case the connectors are directly attached to the fiber. Consequently there is no splice and no need for a splice tray. This method is less expensive than buying preconnectorized pigtails, but it requires that the installer learn how to cleave, polish, and fasten connectors. The techniques of field connectorization include three easily learned methods: the use of crimp, epoxy, and hot-melt.

RECEIVERS AND TRANSMITTERS

Receivers are integrated packages containing both the light detector and signal amplifier circuit. Detectors for optical fibers are almost exclusively semiconductor diodes, either PIN or avalanche design. Avalanche diodes are commonly used with 9/125 fiber. PIN diodes are commonly used with 62.5/125 and 200/230 fiber due to their lower cost, wider temperature range, and greater stability. An example is shown in Figure 3.7t.

Light sources for optical fibers are almost exclusively semiconductor light-emitting diodes and lasers. Lasers are commonly used with 9/125 fiber. Light-emitting diodes (LEDs) are commonly used with 62.5/125 and 200/230 fiber for their lower cost, wider temperature range, and longer lifetimes. Transmitters are similar to the detector shown in Figure 3.7t, but with the emitter integrated circuit (IC) replacing the detector IC. The light from a typical LED is imaged into a spot 100 to 250 microns in diameter with a numeric aperture of .31 to .49.[4]

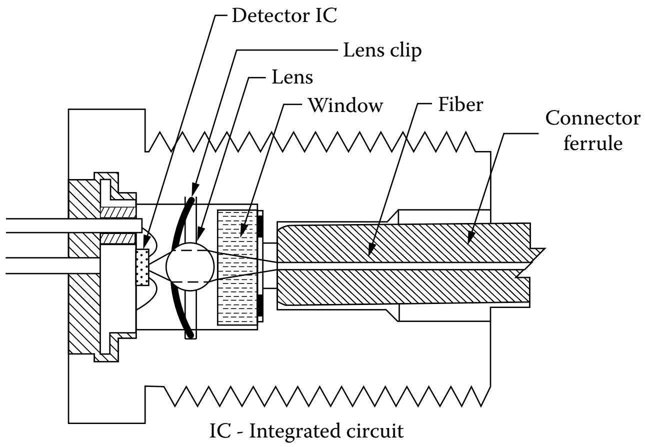


FIG. 3.7t
Fiber-optic receiver. (Courtesy of Hewlett-Packard Co.)

For large-core fibers, the core is about as big as the spot of light at the window of the LED, so very little light spills over into the cladding and is lost. For uniform brightness, the light collection efficiency is proportional to the area of the core.

For large numeric aperture fibers, the maximum ray from the LED is about the maximum ray the fiber can capture, so very little light is lost. For uniform angular distribution, light collection efficiency is proportional to the square of the numeric aperture.

Modems

In its simplest form, a modem translates the communication signals from one kind of device into the signals for another kind of device. The basic principles of electronic transmission were described in Section 3.6. The most common form of modem is one that converts the digital electrical RS-232 communication signals from a personal computer into the analog electrical signals required for the telephone network. Thus, modems allow use of the telephone network for long-distance communications between computers.

In the case of fiber-optic modems, the operation is somewhat similar. For industrial communications, digital computers, programmable logic controllers (PLCs), and their remote I/O drops often need to communicate over optical fiber. Thus, the fiber-optic modem converts digital electrical signals into digital optical signals to be transmitted by optical fibers.

In its simplest form, a PLC modem converts each electrical pulse into an equivalent optical pulse. There is no sampling or multiplexing of the pulse; thus, no change in the pulse-to-pulse timing occurs. The attached PLCs and remote I/O drops operate exactly as if they were linked by their ordinary cabling. The major difference is that they experience virtually no electrical noise — radio frequency interference, ground loops, lightning surges, and so forth — so they have virtually no errors, outages, or downtime due to interrupted communications.

Modem Selection Criteria

Until recently, PLC modems were designed as single-purpose devices. However, increased use of optical fiber networks, advances in fiber-optic technology, and advances in PLC communication technology have resulted in changing requirements for new modems.

Even more difficult were changing requirements for modems that were already installed. For example, a simple point-to-point link would need to be extended, thereby requiring one or more of the modems to become a repeater. Another example is that the fiber-optic link would need to be extended and a longer distance needed, thereby requiring higher-power transmitters and/or more sensitive receivers. Yet another example is the need to add fault tolerance to an existing noise-immune fiber-optic network.

A modular fiber-optic modem's capabilities can be tailor-made to the original application, yet additional capabilities can be added later if the application changes. However, if it were not properly designed, such a modem could be hard to install in the first place and hard to upgrade thereafter. In selecting modular modems, it is important to consider the following:

Standardization—The modem must support all the standard PLCs. Even though some of the signaling methods are proprietary, the modem must function with them. Many plants use more than one brand of PLC yet need the benefits of fiber optics throughout the plant. The modem must be transparent to the electrical signaling by the attached devices. Specifically, design circuitry must minimize pulse width distortion.

Compactness—PLCs and modems are usually mounted together in racks or cabinets. Space is usually at a premium, especially vertical space.

Robustness—In factory automation and process control, the environment is often hostile: wide temperature swings, vibration, dust and dirt, oil mist, and corrosive vapors are among the perils. The modem design must have proven field reliability.

Ease of use—Selecting the modem modules to fit common user requirements must be simple and trouble-free. A simple step-by-step process with a few choices would be best. The actual installation should be as simple as "plug-in and play."

Upgradeability—Upgrading existing modules should be a simple matter of replacing the old with the new. There should be no further adjustments necessary. Adding new capabilities should be a simple matter of plugging the new module into an empty space. This is especially true when adding fault tolerance to an existing network.

Modularity—The main functions of optical communication, electrical communication, and power supply must be in separate modules so that each can be changed without the cost of changing the others.

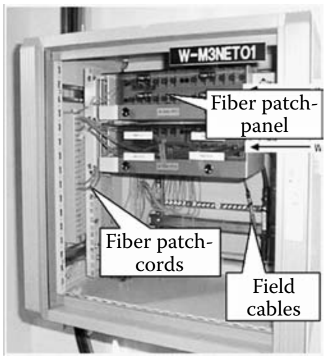


FIG. 3.7u
Wall-mounted cabinet with terminal tray and patch panel.

INSTALLATION AND TESTING

Cabinets

It is recommended that each plant design and standardize its communication cabinets, where all field cabling is terminated and where splices and connectors are protected. Usually, two cabinet sizes are used, a full-size, freestanding cabinet and a smaller wall-mounted one (Figure 3.7u).

Fiber cables usually are brought to the back of the cabinet and are terminated on the fiber patch panels in the rear. One should never connect the fiber cables from the field directly to any electrical device. Patch cords are used to make the connections between the field cables and the electronics. For future expansion, at least 30% spare capacity should be provided in each cable and in each cabinet.

Most of the wear and tear on fiber-optic cables occurs at the end points. One should be careful in avoiding the risk of cutting back the field cables for repair until they become too short and the whole cable has to be reinstalled.

Splicing

The ends of the cable can be terminated by connectors, which are metal or plastic fasteners (Figures 3.7q and 3.7r) used to mate two fibers or to attach fibers to equipment. The other means of termination is by splices, which are permanent connections between two fibers made by welding them in an electric arc (fusion splicing) or aligning them and gluing them together (mechanical splicing).

Fiber termination is the most critical part of the installation process because most failures occur there and because the greatest losses occur at terminations. Cables are run continuously from area cabinet to area cabinet because that results in the least attenuation (signal loss) and are patched through the patch panels, which makes cable numbering easier to follow. The two methods of splicing, mechanical and fusion, both provide excellent long-term reliability when the splices are protected by splice closures.

Fusion Splicing Fusion splicing uses an electric fusion splicer, which is easy to operate. Here, the cables to be connected are fused together with a high-voltage electrical arc. For single-mode splicing, fusion splicing is preferred because this connection results in lower losses. The disadvantage is that the fusion splice machine that is required costs from $20,000 to $40,000.

Mechanical Splicing In this type of splicing the two cable ends are first aligned to a common centerline, thereby aligning their cores. The ends are butted together and permanently secured by adhesive or other means. For multi-mode fiber, this is an effective and easy method. It requires only a few special tools and is ideal for performing emergency field splices. On the other hand, it can be expensive if many splices need to be made.

Pull Strength and Bending Radius

Manufacturers of fiber-optic cables will specify two stress values that must not be exceeded. The higher value is the installation or short-term stress limit, which must be observed while pulling the cable around stationary objects, such as corners, ducts, and conduits. The lower stress limit applies during operation and is called the long-term static or operating load limit. Neither should be exceeded.

Bending fiber-optic cables can seriously affect their performance because it causes increased attenuation of the optical signal in the bends. The minimum bend radius recommended by the manufacturers also has two limit values. The higher value is associated with temporary bending during installation, while the lower limit applies for final training. Naturally, these limits change with cable size, construction, and application.

In addition, repeated impact causing crushing forces must also be avoided, and much heavier power cables should not be placed on unarmored fiber-optic ones. Also, the fiber-optic cable should not be spun off the spool end because that will put a twist in the cable for every turn of the spool. Instead, the cable should be rolled off the spool.

Testing

After the cables are installed, the integrity and performance of the cable system is tested. Visual testing uses 30 to 100 magnification power to inspect the end surface of the connector, which should be smooth, polished, and free of scratches. Only defects that affect the fiber core are a problem. Scratches that affect only the cladding or chipping of the glass outside the cladding will have no effect on the ability of the connector to couple light in the core.

Attenuation Testing Here a light source of known strength is injected into the cable, and the resulting strength is measured at the other end. The optical power loss is measured in decibels (dB) and is a good overall indication of the quality of splices, connectors, and cable. Attenuation is increased by badly finished connectors, tight bends, and excessive strain on the cables. No installation should be accepted without attenuation testing in both directions.

Figure 3.7v shows the main components used in attenuation testing. The optical light source provides the input light at a set wavelength. The launch cables or test jumpers are short fiber-optic cables that connect the light source and the optical power meter to the cable system under test. The optical power meter is set to the same wavelength as that of the light source and contains a photodiode detector and a meter that indicates the amount of signal loss in the cable system. In case of Figure 3.7v, that loss is 21.7 dBm. Such losses are compared to the budgeted design loss, and unacceptable losses are rectified.

Optical Time Domain Reflectometers (OTDR) When superior information is desired about the cable, and it is necessary to know the type of flaw and the location of the flaw along the cable, Optical time domain reflectometers (OTDRs) are used. The OTDR works by injecting a light pulse into one end of the fiber and by detecting the reflected light that is reflected back to the OTDR from any imperfections along

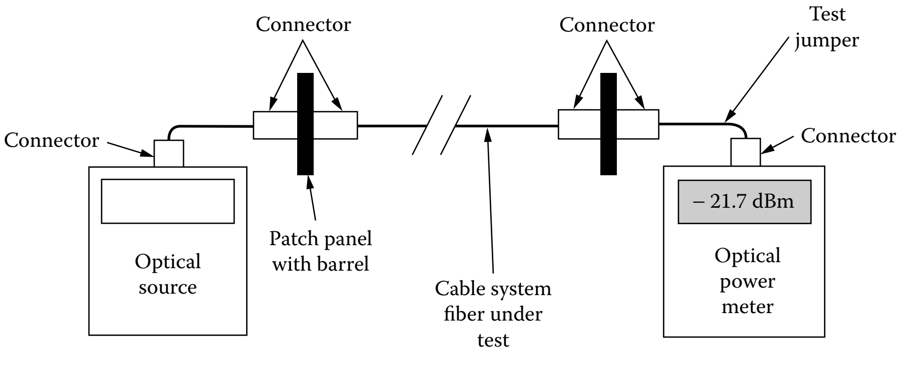


FIG. 3.7v
Attenuation in fiber-optic cables is tested by connecting a light source to one end and an optical power meter to the other end of the cable under test.

the cable. Because of the high cost of OTDRs, their use is limited to testing long-distance cable runs.

CONCLUSIONS

Fiber-optic networks are increasingly used in factory automation and process control applications. These applications have some similarities and significant differences from other fiber-optic data communications networks. These differences affect the criteria for selecting a fiber-optic supplier. The following should be considered:

Experience—No university teaches industrial fiber networking. All the knowledge must be learned from real-world field experience. A supplier should be able to perform network designs and have a strong working knowledge of product specifications and their system compatibility.

Standards conformance—Networks are a utility, similar to the electric power or telephone system. The network must support any connected device. Thus the supplier must use standard fiber, use standard connectors, and support standard industrial communication protocols used by the major equipment suppliers.

Field support—Lightning will strike. Forklift trucks will ram columns. Welders will cut cables. A fiber-optic supplier must have field support personnel readily available to consult or to travel to a field location to assist in disaster recovery. Support is also needed to confirm network layout, product selection, and installation practices.

Full system offering—Especially for an initial fiber-optic network installation, one-stop shopping for all the components in the system and an end-to-end warranty is important for a successful installation. Most factory and process plant communication managers do not want to become experts in fiber-optic technology. They want a network that is painless to install, operate, and maintain so they can concentrate on their own business—manufacturing products.

Supplier selection is an important task. The above criteria must be carefully considered and the supplier's ability to meet them must be carefully evaluated. A mistake in supplier selection can be very expensive and time-consuming to correct.[9]

References

1. Hoaglund, J. B., "Fiber Optics Experience in Paper Manufacturing," ISA, 1990, Paper #90-646.
2. Byers, P., "Landfill Monitoring with Fiber Optics," ISA, 1990, Paper #90-648.
3. Krohn, D. A., et al., "Technology-Enhanced Polymeric-Clad Silica Fibers for Industrial Applications," ISA, 1990, Paper #90–595; and Keck, D. B., and Love, R. E., "Fiber Optics for Communications," Chapter 11 in Volume VI of *Applied Optics and Optical Engineering,* edited by J.C. Wyaut, New York: Academic Press, 1980, p. 476.
4. *Optoelectronics Designer's Catalog 1991–1992*, Hewlett-Packard Co., pp. 5–47, 5–99; *Optoelectronics Device Data*, Motorola, Inc., pp. 5–42 through 5–60, 1989.
5. Huber, J. C., "Installed Cost of Fiber Optic Connectors on Various Fiber Sizes," International Wire & Cable Symposium Proceedings, November 1991, pp. 848–852.
6. Mettler, S. C., "A General Characterization of Splice Loss for Multimode Fibers," *The Bell System Technical Journal,* December 1979, pp. 2163–2182; Chu, T., and McCormick, R., "Measurements of Loss Due to Offset, End Separation, and Angular Misalignment in Graded Index Fibers Excited by an Incoherent Source," *The Bell System Technical Journal,* March 1978, pp. 595–602.
7. Landis, C. S., "New Crimp and Cleave Connector for Large Core Optical Fibers," ISA, 1990, Paper #90-586.
8. Traskiewicz, P., "Fiber Optic Cable in the Industrial Marketplace," ISA, 1990, Paper #90-598.
9. Huber, J. C., "Industrial-Strength Fiber Networks," *Photonics Spectra,* August 1990, p. 107.

Bibliography

AMP Incorporated, "Introduction to Fiber Optics and AMP* Fiber Optic Products," HB5444.

Andreiev, N., "New Fiber Optic Data Highway Is a Masterless Token-Passer," *Control Engineering,* June 1982.

Bliss, J., "A Designer's Guide to System Budgeting," *Electro-Optical Systems Design*, August 1981.

Blume, D., and Myers, J., "Smart Digital Pressure Transmitters for Innovative Tank Gauging," *InTech,* December 1986.

Brambley, J. H., "Fiber Optics: A Bright Future," *InTech,* May 1982.

Byres, E. J., "What Superintendents and Engineers Need to Know About Fiber Optics," *Pulp and Paper Canada,* Volume 8, 1999.

Cole, J. A., and Ellis, J. L., "Specifying Optical Fibers and Cables," *Instruments and Control Systems,* November 1981.

Cowen, S. J., "A High Performance Fiber Optic Pressure Penetrator for Use in the Deep Ocean," NTIS Rep. AD-A100 732/7, 1981.

Fayfield, R. W., "Role of Fiber Optics in Photoelectric Sensing Applications," 33rd IEEE Conference, April 6, 1981.

Geuchalla, M. E., "Specifying Critical Components for Fiber Optic Data Collection Systems," *InTech,* September 1983.

Global Photonics Technology Forecast, *Photonics Spectra*, January 1998.

Hage, J., "Smart Temperature Transmitter Accents Accuracy," *Control,* September 1999.

Hayes, J., *Fiber Optics Technician's Manual,* Florence, KY: Delmar Publishers, 2000.

Hecht, J., *Understanding Fiber Optics*, Indianapolis: Howard W. Sams & Co., 1988.

"Instrument Matchmaker," *Instrumentation Technology,* April 1970.

International Electrotechnical Commission, Safety and Laser Products, IEC-60825, November 2001.

Intrieri, A. J., "Optical Fibers Look Around Obstacles to Measure Temperature," *Control Engineering,* December 1997.

Jaeger, R. E., "Optical Communication Fiber Fabrication," Electro Optics/Laser 76 Conference, New York City, 1976.

Jäger, R., "A Smart Transmitter for Differential-Pressure Measurement," *M & C News,* September 1992.

Johnson, R., "Measuring the Hot, Difficult, and Inaccessible," *Control Engineering,* June 2001.

Keiser, G., *Optical Fiber Communications*, New York: McGraw-Hill, 1983.

Klaassen, M. P., "Fiber Optics in Industry," ISA, C.I. 82–864.

Klaassen, M. P., "Fiber Optics: Ready for Industry," *InTech,* May 1983.

Kleekamp, C., and Metcalf, B., "Designers' Guide to Fiber Optics," *EDN Magazine,* December 1977–March 1978.

Krohn, D. A., "Fiber Optics: New Sensors for Old Problems," *InTech,* May 1983.

Krohn, D. A., "Inside Fiber Optics Sensors," *Instruments and Control Systems,* September 1983.

"Lennie Lightwave's Guide to Fiber Optics," Fotec Inc., 1998.

Mahlke, G., and Gossing, P., *Fiber Optic Cables*, New York: John Wiley, 1987.

Makuch, J. A., "Connector Standards for Fiber Optic Applications," *Insulated Circuits,* June 1981.

Marcuse, D., *Light Transmission Optics,* New York: Van Nostrand Reinhold, 1982.

Mirtich, V. L., "Designer's Guide to Fiber-Optic Data Links," *EDN Magazine,* June 20, 1980, July 20, 1980, August 20, 1980.

Montgomery, J. D., "Choosing Components for Industrial Fiber Optic Systems," ISA, C.I. 82–978.

Ohlhaber, R., "Budgeting Optical Power for Fiber Optic Systems," *Instruments and Control Systems,* September 1979.

Ormond, T., "Fiber Optic Components," *EDN Magazine,* March 20, 1979.

Paros, J. M., "Fiber-Optic Resonator Pressure Transducer," *M & C News,* September 1992.

Pearson, E. R., *Complete Guide to Fiber Optic Cable System Installations*, Florence, KY: Delmar Publishers, 1996.

Personick, S. D., *Fiber Optics: Technology and Applications*, New York: Plenum Press, 1985.

Personick, S. D., *Optical Fiber Transmission Systems*, New York: Plenum Press, 1981.

Sahm, W. H., III, "The General Electric GFOD/E Series Key to Flexible, Low-Cost Fiber Optic Systems," 200.89, October 1981.

Seitz, W. R., "Chemical Sensors Based on Fiber Optics," *Analytical Chemistry,* 56(1), 16A–34A, 1984.

Weik, M. H., *Fiber Optics and Lightwave Communications Standard Dictionary,* New York: Van Nostrand Reinhold, 1980.

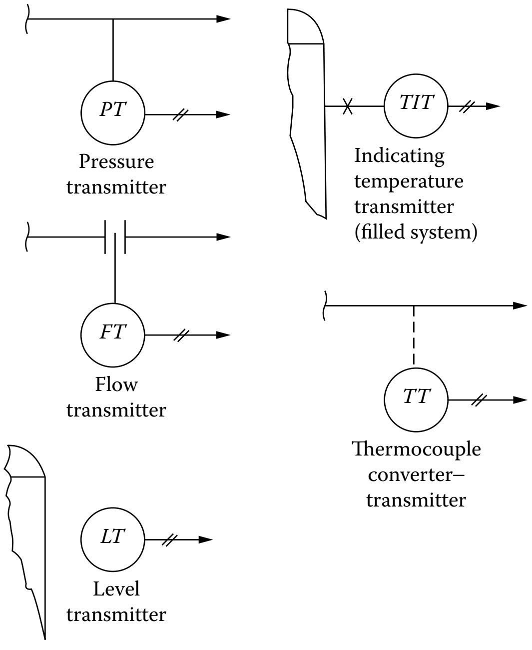
Williams, D. N., "Fiber Optics for Data Transmission," *InTech,* September 1976.

Wohlstein, S., "Fiberoptics," *Measurements and Control,* September 1991.

Workman, J., "Imaging, Chemometrics and New Developments in Sensor Technologies," *Standardization News,* October 2002, pp. 23–25.

3.8 Transmitters—Pneumatic

C. L. MAMZIC (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbols

Ranges:	Force: 15 to 180,000 lbf (66.75 to 801,000 N) Level (buoyancy: 1 in. to 60 ft (25 mm to 18 m) Motion: $^1/_8$ in. to 60 ft (3 to 1500 mm) Pressure: 0.2 in. H_2O to 80,000 PSIG (0.05 KPa to 552 Pa) Pressure differential: 0.01 in. H_2O to 1500 PSID (2.5 Pa to 10.4 MPaD) Temperature span (filled): 50 to 1000ºF (28 to 555°C) within the range of absolute zero to 1400°F (760°C)
Inaccuracy:	Generally $^1/_4$ to 1% of full scale
Costs:	From \$300 for an HVAC-quality transmitter up to \$3000 for some of the noncorrosive special units used in process industry applications. As the cost of many pneumatic transmitters is also affected by line size (variable area flow transmitters) or by materials of construction (displacement-type level transmitters), for detailed pricing information, please refer to the corresponding section in Volume 1 (*Process Measurement*) of this handbook.
Partial List of Suppliers:	ABB Automation Inc. (www.abb.com/processautomation) Acromag Inc. (www.acromag.com) Air Monitor Corp. (www.airmonitor.com) Ametek-APT (www.ametekapt.com) Barton Instrument Systems LLC (www.barton-instruments.com) Bristol Babcock Inc. (www.controlwave.com) Devar Inc. (www.devarinc.com) Dymec Inc. (www.dymec.com) Emerson Process Management (www.easydeltav.com) Fairchild Industrial Products Co. (www.fairchildproducts.com) Invensys Process Systems (www.invensysips.com) Johnson Controls Inc. (www.johnsoncontrols.com) MKS Instruments Inc. (www.mksinst.com) Moore Industries International Inc. (www.miinet.com) MTL Inc. (www.mtl-inst.com) Robertshaw (www.robertshawindustrial.com) Rochester Power Instruments (www.rochester.com)

Ronan Engineering Co. (www.ronan.com)
Rosemount Inc. (www.rosemount.com)
Transmation Products Group (www.transmation.com)
United Electric Controls (www.ueonline.com)
Wilkerson Instrument Co. Inc. (www.wici.com)

INTRODUCTION

The transmitter senses a process variable and generates an output whose steady-state value varies only as a predetermined function of the process variable. The sensor can be an integral part of the transmitter, as in the case of a direct-connected pressure transmitter, or can be a separate part, as in the case of a thermocouple-actuated temperature transmitter.

Electronic, fiber-optic, self-checking, smart, and multivariable transmitters are discussed in other sections of this chapter. The design of traditional pneumatic transmitters has not changed much during the last couple of decades. For this reason, the material that follows is little changed from the section that was prepared by C. L. Mamzic 20 years ago.

HISTORY

A pneumatic transmitter is a device that senses some process variable and translates the measured value into an air pressure that is transmitted to various receiver devices for indication, recording, alarm, and control. Pneumatic controllers date back to the end of the 19th century. Pneumatic transmitters, however, did not make their appearance until the late 1930s—some 25 years after electric telemetering had become an established practice.

Before pneumatic transmitters were introduced, controllers were all direct connected, i.e., they contained a measuring element that was connected to the process. This meant that the controllers and control boards had to be located close to the process.

Pneumatic transmitters were first developed as an alternative to expensive explosion-proof electric transmitters for use in medium-range signal transmission systems in refineries and chemical plants. It was quickly recognized that transmitters offered many advantages over the use of direct-connected controllers, recorders, and indicators, such as safety, economy, and convenience. Transmitters eliminate the need for connecting flammable, corrosive, toxic, and pressurized fluids into the control room.

Furthermore, since controls can be located remotely, centralized control rooms become practical and such elements as long, gas-filled, temperature-sensing bulbs with expensive armored capillary and with attendant ambient temperature errors and sensing lags become unnecessary. As a result, the process variable could be conveniently indicated, recorded, and controlled on relatively inexpensive standardized receiver devices. Once introduced, transmitters caught on quickly, and when miniature pneumatic controls were introduced in 1948, the concept was based on the use of pneumatic transmitters in all remotely controlled loops.

GENERAL FEATURES

Signal Ranges

At first, each supplier settled on its own standard transmitter output range. Generally, the span was selected to be compatible with commonly available pressure sensors and in some cases with the then commonly used operating ranges of pneumatic valve actuators. Too high a pressure would have placed extra demands on the piping and air supply system, whereas too small a span would have meant a sacrifice in resolution or accuracy.

Also, the minimum pressure had to be some value above 0 PSIG (0 bar), since the nozzle backpressure in a transmitter does not (theoretically) drop to atmospheric. In fact, the closer the nozzle backpressure approaches zero, the more critical the seating and the greater the change in baffle clearance per increment of output becomes. This, in turn, results in greater error due to hysteresis and nonlinearity. Using a "live" zero provides an added benefit in that if a transmitter failed, the reading would drop below zero on the scale and give immediate evidence of failure.

Originally, ranges such as 2 to 14, 3 to 18, and 3 to 27 PSIG (0.13 to 0.93, 0.2 to 1.2, and 0.2 to 1.8 bar) were used. The benefits from complete standardization became obvious, and by 1950 a 3- to 15-PSIG (0.2- to 1-bar) range was fast becoming the accepted standard. The 3- to 27-PSIG (0.2- to 1.8-bar) range had been used mostly in power plants and combustion-control systems, and it continued to be used as one of the standard ranges. Formal recognition of standard ranges appeared in 1958 with the issuance of SAMA (Scientific Apparatus Makers Association) Standard RC2-11958. Three ranges were listed as standard: 3 to 15, 3 to 27, and 6 to 48 PSIG (0.2 to 1, 0.2 to 1.8 and 0.4 to 3.2 bar). Today, however, 3 to 15 PSIG (0.2 to 1.0 bar) is overwhelmingly the most accepted.

Baffle-Nozzle Error Detector

The heart of practically all pneumatic transmitters and controllers is a baffle-nozzle–type error detector (Figure 3.8a). The circuit consists of a restriction, detecting nozzle, connecting chamber, and baffle. The baffle is effective as long as the clearance is within one-fourth of the inside diameter of the nozzle. Beyond this clearance, the annular escape area is greater than the area of the nozzle itself, and the baffle no longer provides a restrictive effect.

The restriction must be small enough with respect to the nozzle so that when the baffle is wide open, the resultant nozzle backpressure will be practically atmospheric. The restriction size also determines the continuous air consumption of the circuit, and for that reason it should be kept small.

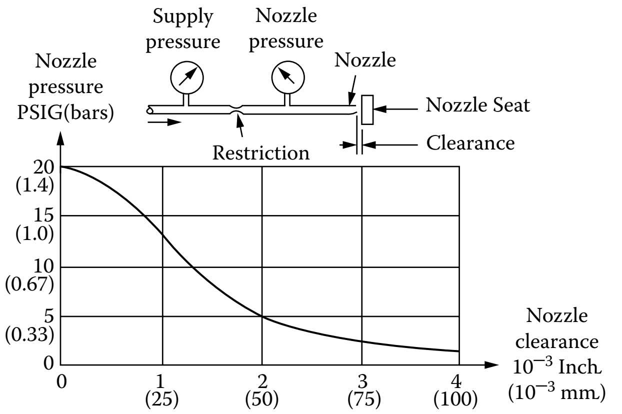


FIG. 3.8a
The relationship between nozzle seat clearance and nozzle backpressure.

On the other hand, the restriction cannot be too small because then it could clog easily with dirt or foreign matter that can be present in the supply air.

Conversely, the nozzle should be large enough to give the proper minimum backpressure but not so large that the corresponding clearance change for full-scale operation becomes so small as to require near-perfect seating. With this in mind, a typical restriction size might be 0.012 in. (0.3 mm) ID and nozzle size 0.050 in. (1.3 mm) ID. For such a general configuration, a plot of nozzle backpressure versus baffle clearance is given in Figure 3.8a.

Pneumatic detector circuits also take some other forms, some of which will be covered in the descriptions of individual transmitter types. The emphasis in this section is on the functioning of the pneumatic transmitter circuitry, since application and measurement details are covered in the *Process Measurement* volume of this handbook. It is not possible to cover all varieties of pneumatic transmitters, but the ones selected are typical and cover the basic types, so that collectively, they represent a good cross-section of the various design features that are available.

FORCE-BALANCE DEVICES

One-to-One Repeaters

The force-balance principle is commonly used in pneumatic transmitters. A very basic 1 : 1 force-balance transmitter is shown in Figure 3.8b. Other repeater designs are discussed in connection with level and pressure detectors in the *Process Measurement* volume of this handbook.

In Figure 3.8b the process pressure acts downward on the flexible diaphragm, and the force resulting therefrom is counterbalanced by the force of nozzle backpressure acting upward on the diaphragm. Air, at a supply pressure slightly higher than the maximum process pressure to be measured, flows through the restriction to the underside of the diaphragm and bleeds through the nozzle to atmosphere. At equilibrium, the nozzle seat clearance is such that the flow of air in the nozzle is equal to the continuous flow through the restriction. If the process pressure increases, the diaphragm closes (baffles) off the nozzle, causing the backpressure to increase until a new equilibrium is achieved. If the process pressure drops, the diaphragm moves away from the nozzle, causing the backpressure to drop until equilibrium is reestablished. Since nozzle backpressure is directly related to process pressure, the signal can be used remotely for indication, recording, or control.

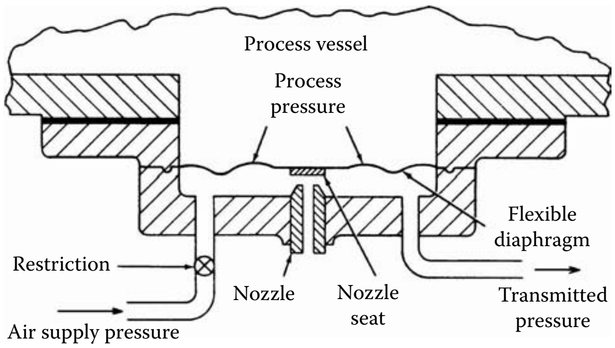


FIG. 3.8b
Simplified drawing of a 1:1 force-balance pressure transmitter with direct nozzle circuit.

Volume Booster Though such a transmitter is accurate, it has some limitations. Namely, since all the air must flow through the restriction, the speed of response will be slow, particularly if there is much volume in the output side of the circuit. Also, a leak in the output side will cause an error or even make the unit completely inoperative. Figure 3.8c shows

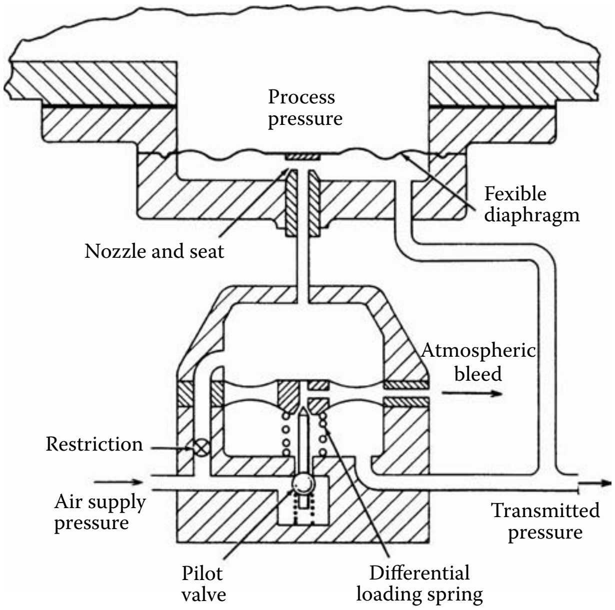


FIG. 3.8c
One-to-one force-balance pressure transmitter with volume booster and constant differential nozzle circuit.

essentially the same repeater as did Figure 3.8b, but with refinements. This system employs a volume booster that has considerable air-handling capacity, so that it speeds up the response, minimizes transmission lag, and copes with leakage.

In addition, it provides a constant pressure drop across the nozzle regardless of the level of operation. This improves accuracy with regard to linearity and hysteresis as the total diaphragm travel is lessened, and the relationship of nozzle seat position to nozzle backpressure is more linear than that of the direct nozzle circuit in Figure 3.8b.

The booster contains an exhaust diaphragm assembly consisting of two diaphragms that move integrally and that provide a bleed to atmosphere. On the underside of the diaphragm is a differential spring, which at balance exerts a force equivalent to 3 PSID (0.2 bar) acting on the diaphragm. Therefore, the nozzle backpressure that acts above the exhaust diaphragm is always nominally 3 PSI (0.2 bar) higher than the transmitted pressure. Since the nozzle bleeds into the transmitted pressure, the pressure drop across the nozzle is *always constant* regardless of output pressure. As in Figure 3.8b, the air pressure on the underside of the transmitter diaphragm counterbalances the process pressure.

If the process pressure increases, the diaphragm moves the seat closer to the nozzle, increasing the nozzle backpressure. This moves the exhaust diaphragm downward, closing off the exhaust port as it contacts the pilot valve, and moves the pilot valve downward to open the supply port. The result is an increase in transmitted pressure until the transmitted pressure, which is fed back to the underside of the diaphragm, equals the process pressure. A decrease in process pressure causes the nozzle backpressure to drop, and the exhaust diaphragm moves upward so that the pilot valve closes off the supply port while the diaphragm opens the exhaust port, causing the transmitted pressure to drop until equilibrium is established. At balance, there is a continuous flow of air passing through the restriction, the detection nozzle, and out to atmosphere via the exhaust diaphragm.

Membrane d/p Cells

Transmitters developed for the detection of very low pressure differentials (Figure 3.8d) detect by pressure balance across a membrane, which has a negligible mass and practically no movement. The output signal (Po) of the membrane d/p cell is related to the pressure difference between the high-pressure (Ph) and the low-pressure (Pl) inputs but is amplified by 1000 times or more. The membrane d/p cell is continuously purged and is capable of detecting down to a range of 0 to 0.01 in. H_2O (0 to 0.25 mm H_2O) differential.

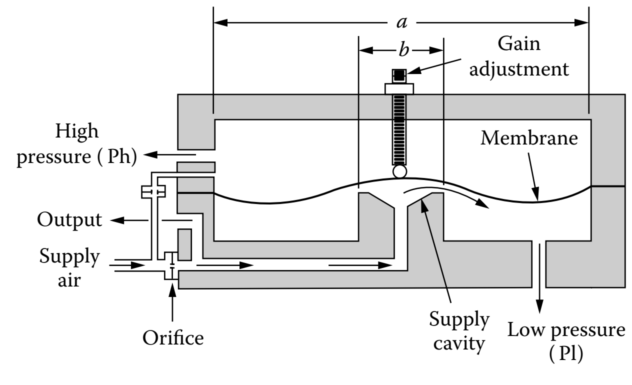


FIG. 3.8d
Membrane type low d/p transmitter. (Courtesy of Brandt Industries Inc.)

Advantages Since the membrane does not move or displace during measurement, the valve has no moving or mechanical parts to wear or contribute to hysteresis. Thus, the device will accurately sense very low pressures at its input ports, while being insensitive to shock, vibration, and mechanical problems normally associated with high-mass pressure-sensing devices. It consumes less than $^1/_5$ the amount of instrument air of a traditional d/p cell.

Because this transmitter does not have bellows or diaphragms that must fill with air, it responds much faster to an input change than any other device measuring pressure changes at very low levels. This means that in an application measuring turbulent air or gas flow, the transmitter will exhibit a higher output than manometers or similar devices would because it will respond to true average flow, which has short-term peaks not "seen" by a manometer or mechanical device.

This transmitter is not "wetted" by the process, but rather continuously purges the high and low ports with a low flow of air. This prevents the entrance of water, dirt, or process gases into the sensitive measurement cavities of the transmitter. As a result of this constant purge, some backpressure or resistance to flow in the connecting tubing, will develop. Therefore, this tubing must never be blocked, and corrections must be made for the pressure drop in the connecting tube.

Operation Referring to Figure 3.8d, supply air enters through an orifice. When there is no pressure difference between the high and low ports, the membrane does not exert a force on the supply cavity, and output pressure does not appear at the output port. If the output is dead-ended, all the supply flow is relieved through the ports.

When a difference exists between the high and low ports, the membrane exerts a force on the supply cavity, which also appears at the output port. This pressure is a function of the gain, which is a function of dimensions *a* and *b*. When the pressure in the supply cavity satisfies the gain equation, supply air flows out of the supply cavity into the low port. At this point, the valve is in equilibrium. If the applied pressures decrease, the reverse occurs. The valve has a stable gain and no measurable hysteresis; therefore, the valves can be cascaded (output of one fed into the input of the next valve) for very high total gains.

Pressure Balance Membrane technology is based upon pressure-balance theory rather than force balance or diaphragm displacement. By definition, the membrane is virtually mass free and will not support a bending moment.

Movement is limited to billionths of an inch (millionths of a millimeter).

A membrane-type differential pressure transmitter utilizes three high-gain amplifiers, each with a specific gain according to the required span of the transmitter (Figure 3.8e). The first stage senses the input differential pressure to be measured, which may be as low as 0 to 0.01 in. W.C. (0 to 2.5 Pa) full scale. It amplifies this pressure, and its output is conveyed to the second-stage valve, which amplifies the signal further and conveys it to the third stage. This third stage amplifies the signal by a small amount, but its primary purpose is to provide "driving" power to supply a standard 3- to 15-PSIG (0.2- to 1.0-bar) transmitter output.

Pressure Transmitter

The applications for pressure repeaters are limited. The more common transmitter is one that converts the measured process variable range into a standard 3- to 15-PSIG (0.2- to 1.0-bar) output pressure signal (Figure 3.8f). It operates on a principle of force balance, or more precisely, on a principle of moment balance.

Process pressure acts on the input bellows and applies a force on the balancing beam, which rotates on the flexure pivot. A change in input pressure results in a moment that is counterbalanced by an equivalent change in moment as output pressure changes in the feedback bellows. An increase in process pressure rotates the beam counterclockwise and moves the nozzle seat closer to the nozzle, increasing the backpressure and booster output until the moments come into balance. The biasing spring ensures sufficient counterclockwise moment on the balancing beam so that, even with zero process pressure, it is

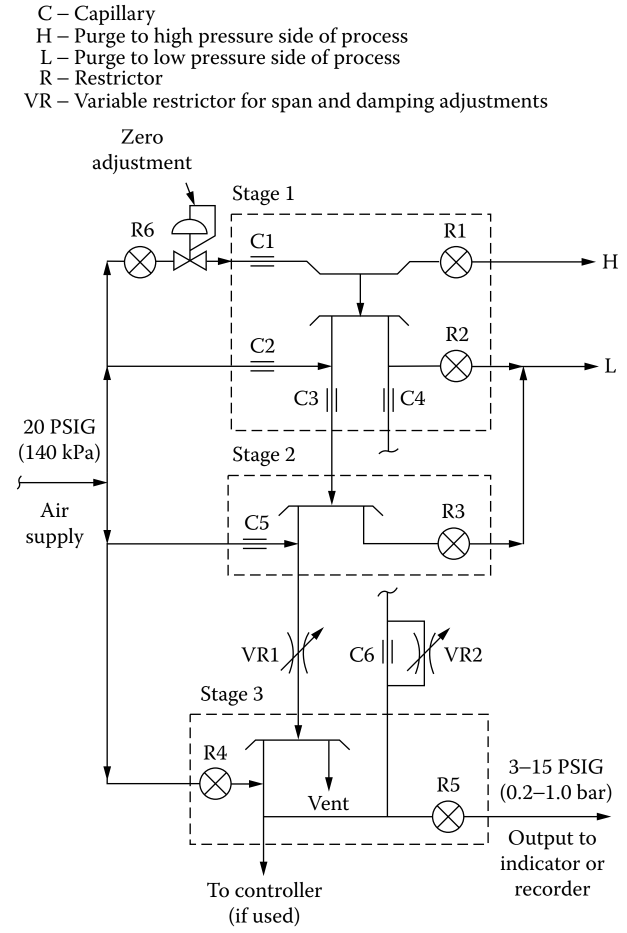


FIG. 3.8e
Membrane-type differential pressure transmitter.

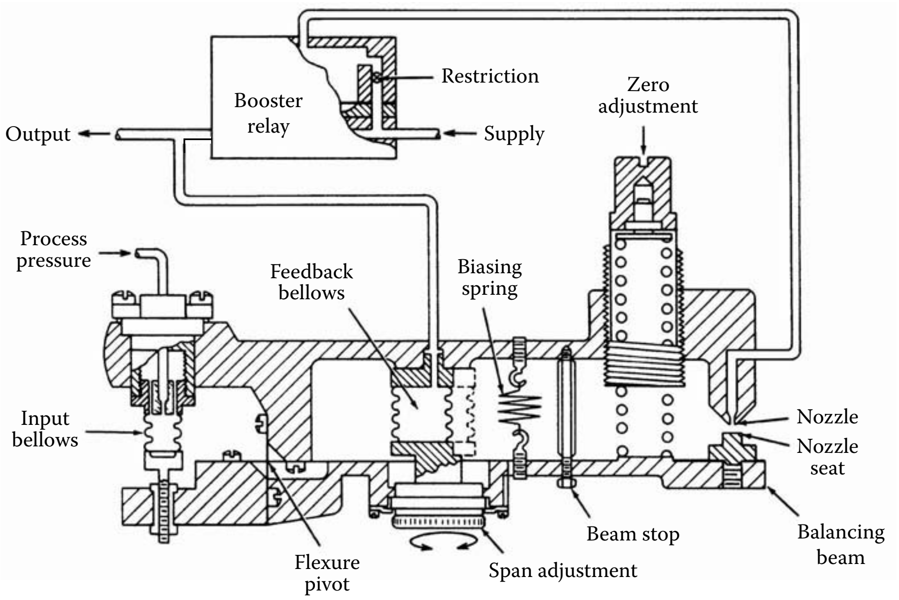


FIG. 3.8f
Force-balance pressure transmitter with variable range.

possible to set the output at 3 PSIG (0.2 bar) by adjusting the compression of the zero spring. Added compression of the zero spring gives an elevated zero, or range suppression.

The feedback bellows is eccentrically mounted on a rotatable seat. Rotating the assembly changes the moment arm, hence the span. The span is basically a function of the relative ratio of effective areas of the input bellows and the feedback bellows. The booster relay is an amplifying type that minimizes total nozzle clearance and improves accuracy.

MOTION-BALANCE TRANSMITTERS

The pressure transmitter in Figure 3.8g consists of a pressure measuring element and a motion transmitter. Instead of a conventional baffle-nozzle, this transmitter employs an annular orifice with a variable restrictor called the wire pilot. Supply air passes through the fixed restriction into the follow-up bellows and out the detector nozzle. The wire pilot throttles the exhaust from the nozzle. It has a sharply tapered step so that when the large-diameter wire restricts the orifice, the backpressure rises to a maximum, and when the small diameter is effective the backpressure drops to atmospheric. At balance, therefore, the follow-up bellows move to position the detector nozzle in line with the tapered step of the wire pilot.

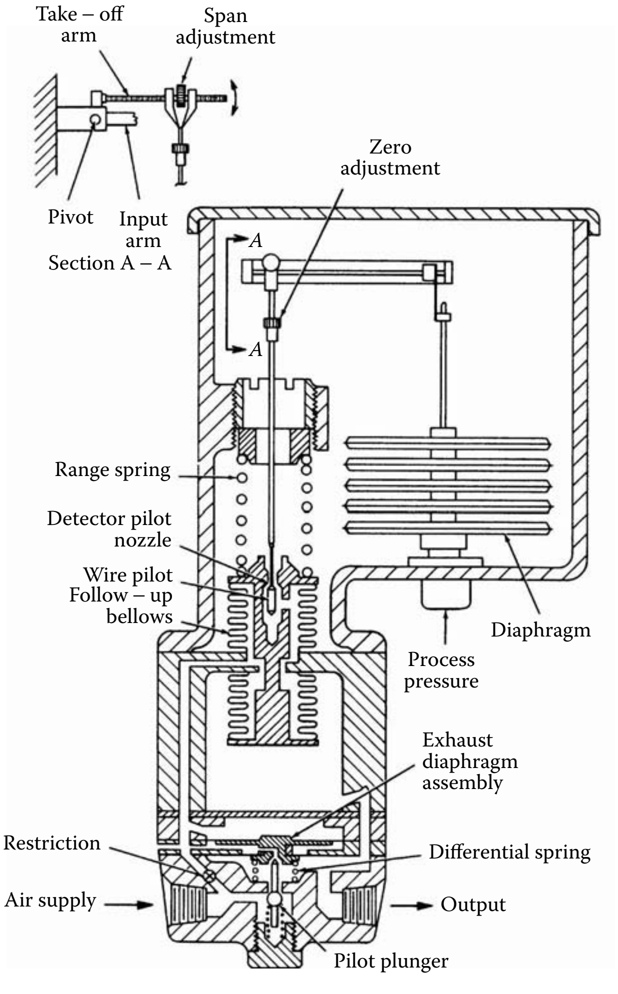


FIG. 3.8g
Motion-balance-type pressure transmitter.

Process pressure acts on the measuring diaphragm. An increase in process pressure moves the diaphragm upward, which, via the U-shaped linkage, moves the wire pilot upward. The wire pilot restricts the annular orifice and the backpressure increases. The two bellows that make up the follow-up bellows system have the same area and are connected rigidly by a center post so that nozzle backpressure has no effect on movement of the bellows assembly. The nozzle backpressure is connected to the top of the exhaust diaphragm assembly.

This is an amplifying-type diaphragm assembly since the upper diaphragm has six times the effective area of the lower. Therefore, as the nozzle backpressure increases, the output increases in a 6:1 ratio. The output feeds back to the underside of the follow-up bellows and pushes it upward. Upward motion is resisted by the range spring. The spring constant is such that a 12-PSIG (0.8-bar) change on the follow-up bellows moves the bellows assembly through the nominal full-scale travel of the wire pilot.

Zero is adjusted by setting the initial operating position of the wire pilot. Span is adjusted by varying the radius of the takeoff arm shown in section A-A. The total force required to operate such a transmitter is only 2 grams. Therefore, any type of primary element can be used with it—from low-force draft elements to various types of high-pressure elements. The basic range is determined by the spring rate of the measuring element.

TRANSMITTERS GROUPED BY MEASURED VARIABLE

Differential Pressure Transmitter

A typical force-balance differential pressure transmitter is shown in Figure 3.8h. The high and low pressures act on opposite sides of a diaphragm capsule, and the resulting pressure differential exerts a force on the force bar. The force bar pivots on the diaphragm seal. The external end of the force bar pulls on one end of the range rod. The range rod, with its integral flapper, pivots about the range wheel. A feedback bellows acts on the opposite side of the range rod. A change in differential results in a changed flapper position, which alters the nozzle backpressure, relay output, and feedback force until all moments come to balance.

Zero is adjusted by adjusting the zero spring tension. Span is adjusted by moving the range wheel, which changes the relative input/output moment arm ratio.

Since the output of the transmitters is linearly proportional to pressure differential, if the unit were used on an orifice-type measurement, the flow would have to be read on a square root calibrated scale.

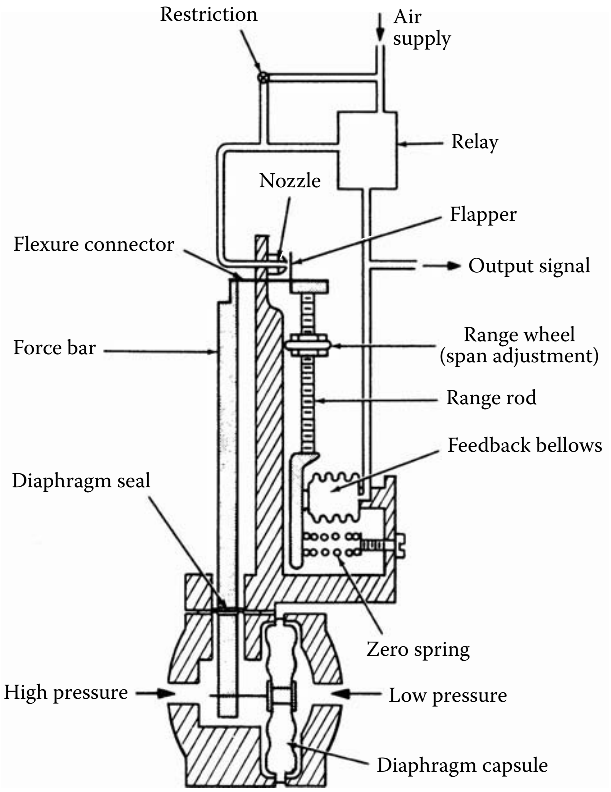


FIG. 3.8h
Force-balance–type differential pressure transmitter.

Square Root–Extracting d/p Transmitter

Differential pressure elements, such as in Figure 3.8i, provide a motion output. A motion transmitter, as in Figure 3.8g, could be used with such a meter to give an output linearly proportional to the differential. However, the motion transmitter shown in Figure 3.8i converts linear motion of the meter into a square root–related output so that orifice-type flow can be read on a linear scale.

Linear signals are preferred when flows are added, subtracted, or averaged and when other analog computing and characterizing requirements exist. They are often specified in order to give better readability and control rangeability.

The meter in Figure 3.8i consists of a high- and a low-pressure bellows joined by a common center shaft. The differential causes the bellows to move a linearly proportional amount depending upon the total spring rate of the range spring, the bellows, and torque tube. Movement of the bellows twists the torque tube, causing rotation of the torque tube shaft. It is this motion that actuates the transmitter. The bellows are filled with liquid (usually ethylene glycol), and as the bellows move, liquid is transferred from one bellows to the other via the pulsation dampener needle valve. If the normal differential pressure range is exceeded, the bellows move until an O-ring on the center shaft seals off the liquid in the bellows. The pressure can then build up to the body rating of the meter without damaging the unit.

The torque tube rotates the lever connector of the motion transmitter. A floating pilot link is socketed in the connector at one end and baffles off the detector nozzle at the other. The transmitter is calibrated so that at zero differential the output is at 3 PSIG (0.2 bar) and the pilot link is horizontal. An increase in differential rotates the connector counterclockwise. This restricts the nozzle and causes a buildup in transmitted pressure, which in turn acts on the large bellows, moving it downward until equilibrium is attained.

Since the pilot link has a fixed length and because the nozzle clearance is constant for all practical purposes, the pilot nozzle must be moved downward according to the cosine law. This means that at minimum input with link horizontal, there must be considerable motion of the pilot nozzle as compared with the lever connector, and therefore the gain is practically infinite. As the included angle increases, considerably less nozzle travel is required to offset lever connector motion. For small angles, the relationship is almost exactly square root.

Another transmitter having the same function employs a varying spring rate with travel. This is accomplished by picking up added leaf springs as the travel increases.

Variable Area Flow Transmitter

When a variable area flowmeter is used as a primary flow-measuring device, the float position is magnetically detected. The motion of the magnetic follower mechanism then is converted into a 3- to 15-PSIG (0.2- to 1.0-bar) signal. (Volume 1 of this handbook contains a detailed discussion of the operation of rotameters.)

In Figure 3.8j, a permanent magnet is embedded either in the float or in an extension of the float. A magnetic steel helix, which is supported in an aluminum cylinder, is mounted between bearings. The leading edge of the helix is constantly attracted to the magnet. The vertical position of the float results in a corresponding radial position of the helix.

A cam is attached to the helix follower assembly. A pneumatic circuit consisting of a transmitting and a receiving nozzle detects the cam profile. The receiving nozzle pressure serves the same function as nozzle backpressure in a conventional baffle-nozzle circuit. When the cam does not interrupt the flow from the transmitting to the receiving nozzle, the receiver pressure is at maximum. When it is fully interrupted, the receiver pressure is 0 PSIG (Pa). At balance, the detector system is throttled by following the cam profile.

As the float rises, the cam rotates in a direction of decreasing displacement. As the cam edge moves away from the detector nozzles, the increasing receiver nozzle pressure acts on the relay and results in an increase in transmitted pressure. The transmitted pressure acts on the feedback capsule and moves the flexure-mounted detector nozzles toward the cam edge until equilibrium is established. The spring rate of the range spring is set so that a 12-PSI (0.8-bar) change in output is required to track the full displacement of the cam.

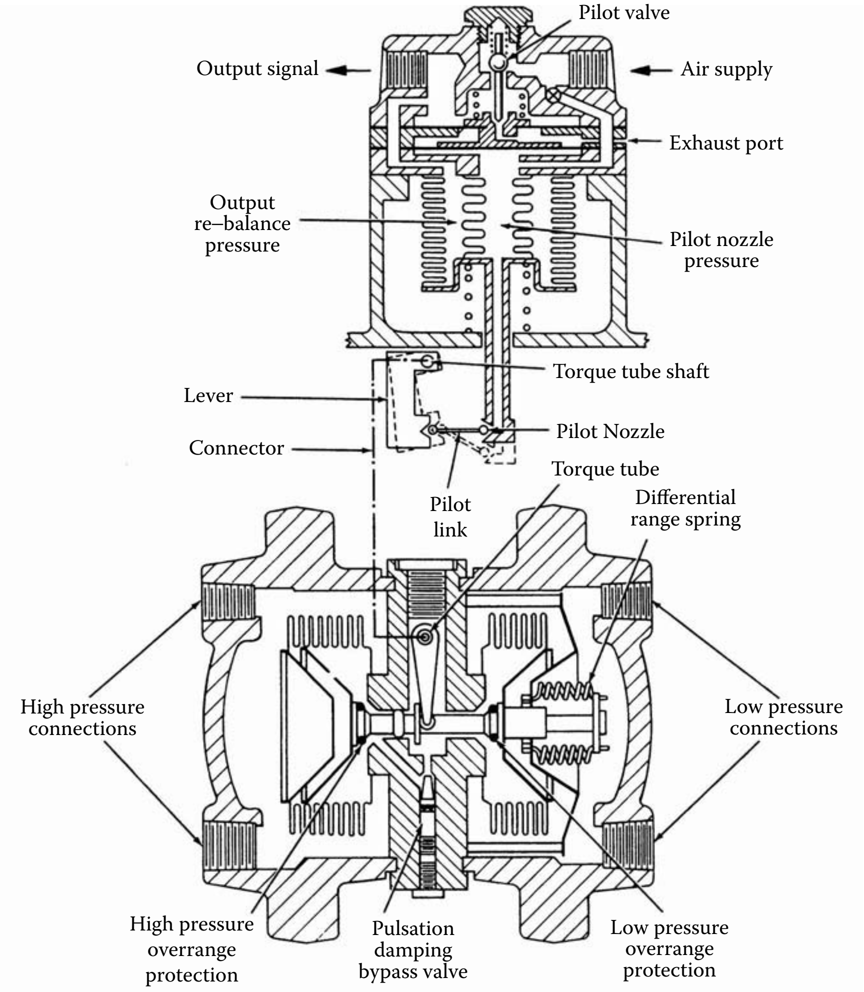


FIG. 3.8i
Linear flowmeter consisting of a combination of differential meter and square root–extracting motion transmitter.

Span is adjusted by turning a screw that takes up coils in the range spring, thus changing its spring rate. Zero is set by adjusting a second screw that sets the initial spring tension.

Filled Bulb Temperature Transmitter

Pneumatic temperature transmitters are almost exclusively the force-balance types. Figure 3.8k shows a transmitter with a sealed, gas-filled bulb (Class III system). (A detailed discussion of filled thermal systems can be found in Chapter 4 of the *Process Measurement* volume of this handbook.).

Except for a negligible volume of gas in the thermal system bellows and in the connecting capillary, practically all the fill gas is in the bulb. The volume of the thermal system is constant, and as the bulb senses the process temperature, the pressure of the fill gas varies according to the gas laws. The fill gas pressure creates a force downward on the thermal system bellows. This force acts through the thrust rod and is counterbalanced by the force resulting from transmitted air pressure acting upward on the transmitter bellows.

An increase in process temperature increases fill gas pressure, which pushes down on the thrust rod baffling off the nozzle. The consequent increase in nozzle backpressure pushes the exhaust diaphragm and pilot plunger down, closing the exhaust seat and opening the supply port until the increase in transmitted pressure acting on the transmission bellows results in force balance.

The temperature range is a function of the initial fill pressure (the higher the pressure, the narrower the span) and the ratio of the effective areas of the thermal system and transmission bellows. The zero spring acts counter to the thermal system and establishes the low end of the measured temperature range.

Buoyancy Transmitter (Level or Density)

Figure 3.8l illustrates a motion-balance buoyancy transmitter. (A more detailed discussion of displacement type level instruments is presented in Chapter 3 of the *Process Measurement* volume of this handbook.) Changes in level directly affect the net weight of the float as the float displaces the liquid.

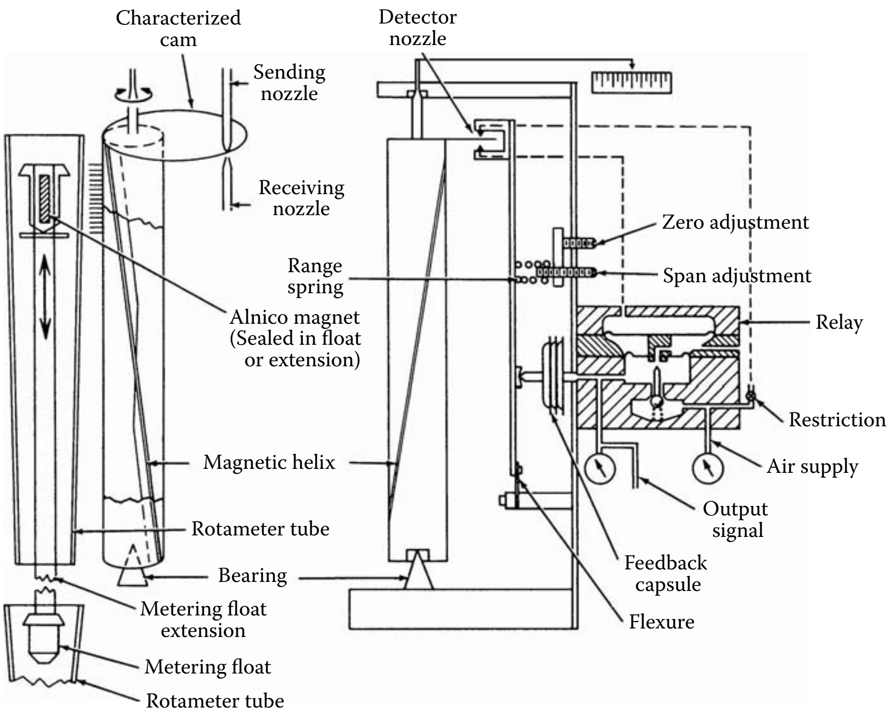


FIG. 3.8j
Rotameter transmitter with magnetic take-out and pneumatic cam follower mechanism.

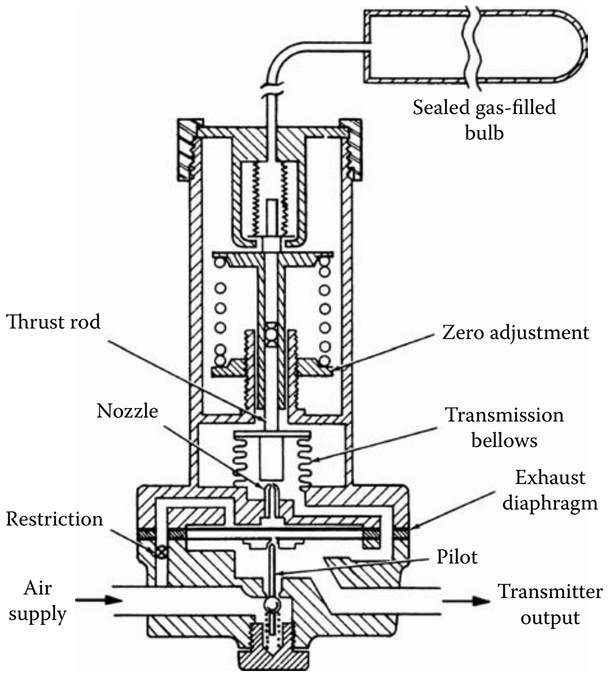


FIG. 3.8k
Filled bulb–type force-balance temperature transmitter.

The float lever is connected to a torque tube. As the torque tube twists, it rotates a center shaft to which the flapper is attached.

Supply air flows through a restriction to the top of the booster relay and through a small tube inside the bourdon to the detecting nozzle. The booster relay, which is amplifying with a 3:1 ratio, provides an output pressure, a portion of which is fed back to the bourdon via a three-way valve. The three-way valve provides for span adjustment. If its plunger is moved up, closing off the exhaust seat, full transmitted pressure feeds back to the bourdon.

The result is that when the level rises and moves the flapper closer to the nozzle, the consequent increase in transmitted pressure makes the bourdon move away from the nozzle (negative feedback) so that total flapper travel and hence measuring range is large. Adjusting the three-way valve in its other extreme position where it closes off transmitted pressure results in practically on/off action at the nozzle, representing the narrowest range of measurement. Normally the three-way valve is adjusted somewhere between these limits. By rotating the bourdon tube with respect to the flapper, the zero is changed.

Buoyancy transmitters can be used to measure density or specific gravity as well as level and level interface. Force-balance designs are also available.

Force Transmitter

Force transmitters can serve as load cells in weighing applications or in the measurement of variables such as web tensions. The transmitter in Figure 3.8m operates on a force-balance principle. (For a more detailed discussion of pneumatic load cells refer to Section 7.25 in the *Process Measurement* volume.)

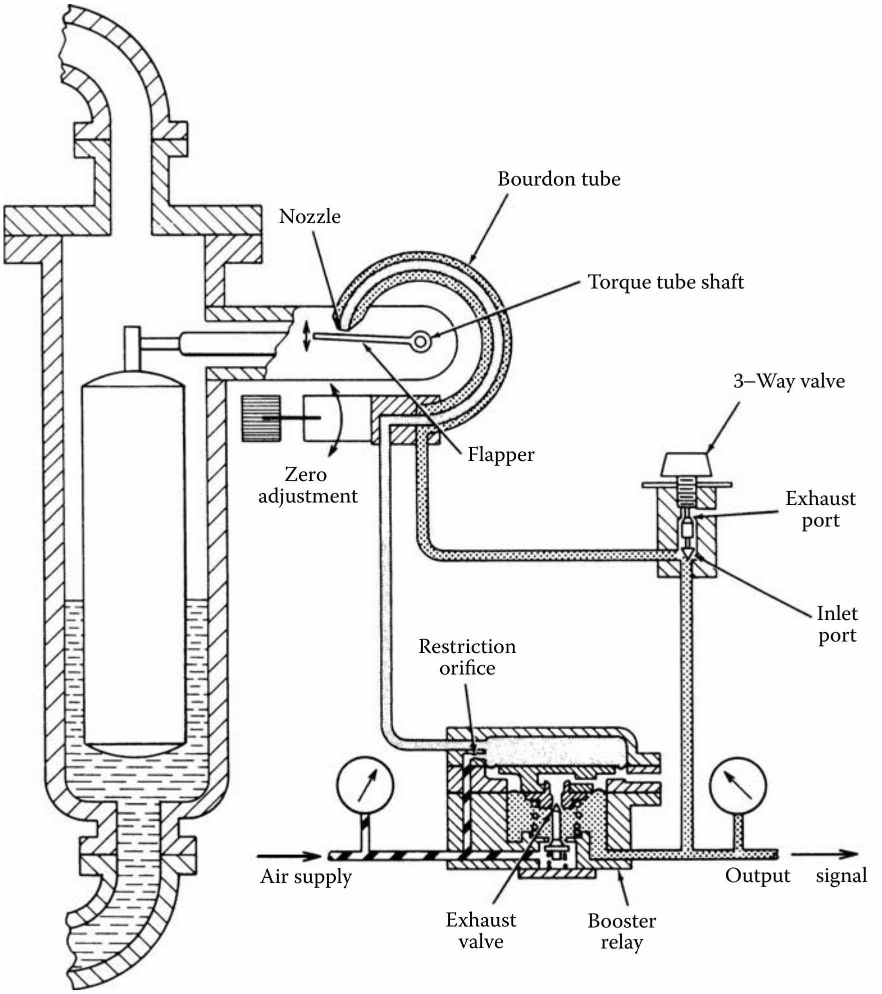


FIG. 3.8l
Level transmitter with torque tube–type spring and pneumatic follower system.

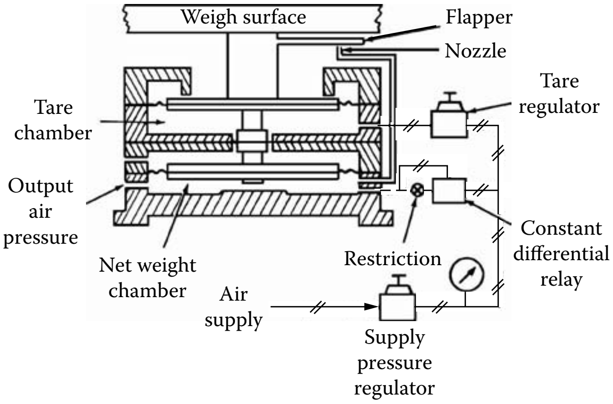


FIG. 3.8m
Force-balance weight transmitter.

Pressure from the tare regulator acts upward on the top diaphragm and is set to counterbalance the fixed weight of a hopper or tank and support structure. The net weight in the tank is counterbalanced by the output air pressure, which is acting under the bottom diaphragm. Air flows from supply through a constant differential relay, which maintains a relatively constant flow across a restriction and into the net weight chamber. The net weight chamber is connected to the detector nozzle. The nozzle baffle is attached to the supporting platform. If net weight increases, the platform and flapper move down, baffling off the nozzle and causing the backpressure to increase until force-balance is restored.

Smaller pneumatic force transmitters are also available. For stabilizing potentially noisy systems, some of these designs incorporate hydraulic pulsation dampening.

Motion Transmitter

A variety of approaches are used to detect motion. Some of these are described in Sections 7.14 and 7.19 of Volume 1, the *Process Measurement* volume of this handbook.

A motion balance transmitter as shown in Figure 3.8g can be adapted to measure total motions from $^1/_8$ in. (3 mm) to approximately 1 in. (25 mm). This design is particularly useful where only a low force is available. Ordinary pressure

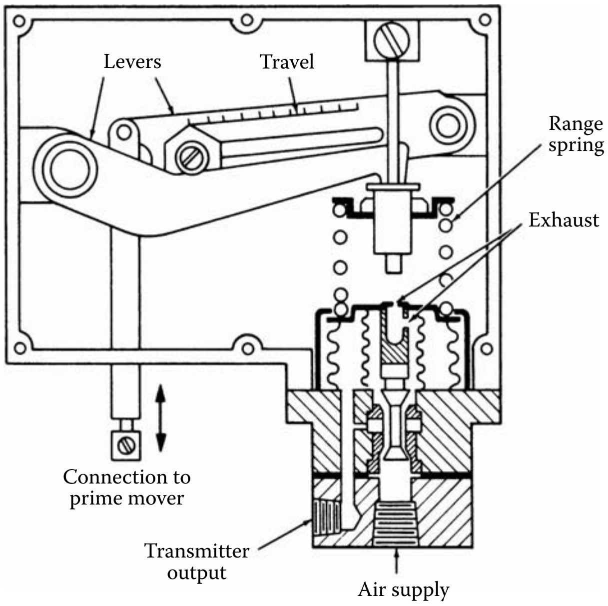


FIG. 3.8n
Valve positioner connected to serve as a position transmitter.

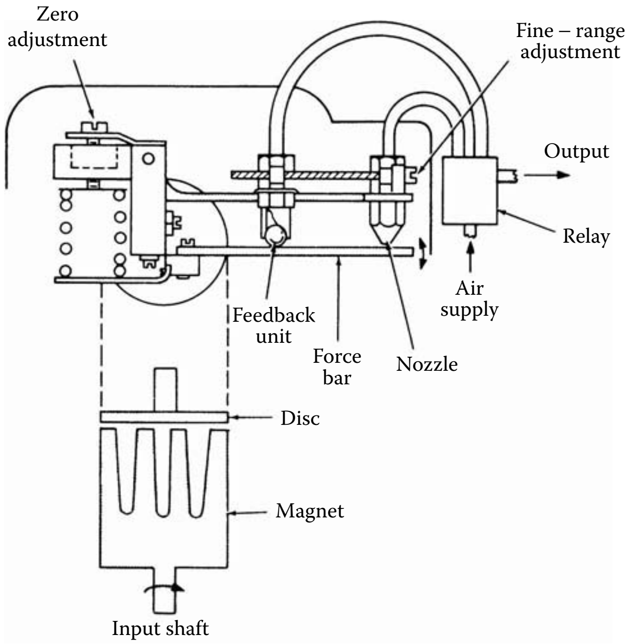


FIG. 3.8o
Pneumatic, force-balance–type speed transmitter.

regulators can also be adapted as motion transmitters by substituting a sliding thrust rod for the lead screw that normally adjusts the pressure-setting spring.

The most common approach, however, is to use a valve positioner and, in effect, reverse its functioning. This way, instead of controlling position or motion, it will transmit a signal that is proportional to position or motion. Figure 3.8n shows a valve positioner that is connected as a motion transmitter. The pilot valve is reverse-acting, and the output feeds back into what is normally the input bellows.

If in Figure 3.8n the prime mover pulls the connector downward, the parallel lever system pushes down and compresses the range spring. This opens the supply port and closes off the exhaust port of the three-way pilot, causing the output to increase. The output applies an upward pressure on the outer bellows to counteract the increase in spring force. A roll-type contact point establishes the gain of the lever system. Hence, its adjustment determines the span of measurement.

Speed Transmitter

In Figure 3.8o, a prime mover drives an input shaft that carries a multi-pole permanent magnet. The combination of magnetomotive pull and rotation of the magnet tends to turn the disc on its flexure pivot. The torque on the flexure is proportional to input shaft speed. Attached to the flexure is a radial force bar that doubles as a flapper and as a rebalancing lever. Output pressure feeds back to a ball-type piston, and the derived force counterbalances the input torque.

As speed increases, an increase in counterclockwise torque results. This moves the flapper toward the nozzle. The relay amplifies the nozzle backpressure change, which serves as the output and as the feedback to the ball piston. The pressure increases until the feedback moment balances the input torque.

Pneumatic speed transmitters are also obtained by combining an electrical speed detector–amplifer combination with an electric-to-pneumatic converter. A typical detector is one that integrates the rate of magnetic pulses generated as gear teeth cut across the field of a magnetic pick-up head, illustrated in Figure 3.8p. (This design is explained in more detail in Section 7.19 of the *Process Measurement* volume.)

TRANSMISSION LAG

With pneumatic transmitters, there is a transmission lag that increases with the length of transmission tubing.

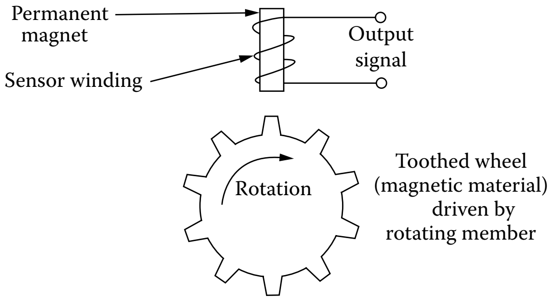


FIG. 3.8p
Schematic of an induction-type speed sensor.

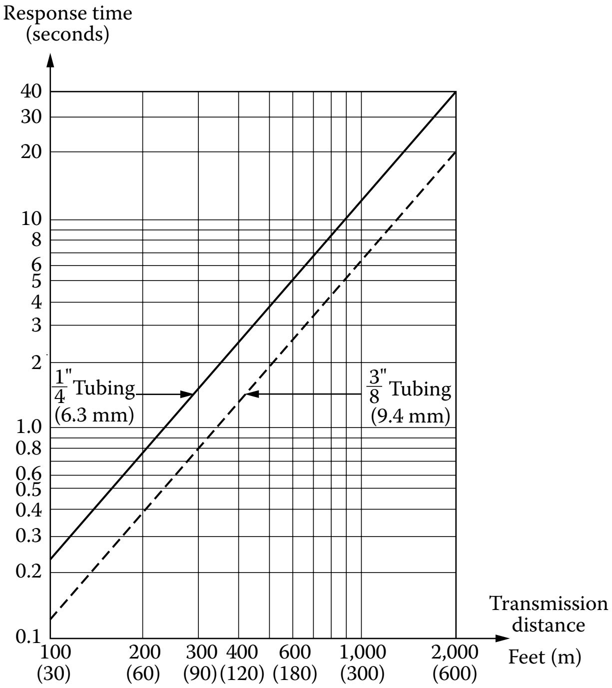


FIG. 3.8q
Response time (time for 63.2% of the complete response) versus transmission distance.

The transmission lag is described as the time that is required to obtain 63.2% of a step input. The 63.2% figure defines the time constant in first-order systems, but since transmission systems do not behave as first-order systems, the figure cannot be so interpreted. Nevertheless, it does serve as an arbitrary benchmark for comparison.

A plot of response time versus transmission distance for 1/4 and 3/8 in. (6.3 and 9.4 mm) tubing is given in Figure 3.8q. Response is faster with 3/8 in. tubing, but 1/4 in. tubing is more often used. At 300 ft (90 m), for example, the response time for 1/4 in. tubing is 1.5 seconds, and for 3/8 in. tubing, it is 0.8 seconds.

The effect of transmission distance on pneumatic control is discussed in more detail in Sections 1.4 and 3.1 of this volume.

Bibliography

Adams, L. F., *Engineering Measurements and Instrumentation,* The English Universities Press, 1975.

"Air Pressure for Pneumatic Controllers for Transmission Systems," ISA Standard S7.4, 1981.

Brindley, K., *Sensors and Transducers,* Heinemann Professional Publishing, 1988.

Buckley, P. S., "Dynamic Design of Pneumatic Control Loops," *InTech,* April 1975.

Buckley, P. S., *Techniques of Process Control,* New York: John Wiley & Sons, 1964.

Buckley, P. S., "A Modern Perspective on Controller Tuning," Texas A&M 30th Annual Symposium on Instrumentation for the Process Industries, January 1973.

Buckley, P. S., and Luyben, W. L., "Designing Long-Line Pneumatic Control Systems," *Instrumentation Technology,* April 1969.

Demarest, W. J., "There's More to Transmitter Accuracy Than the Spec," *Instruments and Control Systems,* May 1983.

Doebelin, E. O., *Measurement Systems — Application and Design,* rev. ed., New York: McGraw-Hill Book Co., 1975.

Dyer, S. A., *Wiley Survey of Instrumentation and Measurement,* New York: Wiley, 2001.

Eckman, D. P., *Automatic Process Control,* New York: John Wiley & Sons, 1958.

Gore, F., "When to Use a Square Root Extractor," *Instruments and Control Systems,* April 1977.

Harriott, P., *Process Control,* New York: McGraw-Hill Book Co., 1964.

Houghen, J. O., and Martin, O. R., "Dynamics of Pneumatic Transmission Lines," *Control Engineering,* September 1963.

Jones, B. E., *Instrumentation, Measurement, and Feedback,* New York: McGraw-Hill Book Co., 1977.

Luyben, W. L., et al, *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

"Quality Standard for Instrument Air," ISA Standard S7.3, 1976.

Sandell, R. P., and Ceaglske, N. H., "Frequency Response of Pneumatic Transmission Lines," *ISA Journal,* December 1956.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, Inc., 1985.

"Transmitters," *Measurements and Control,* June 1994.

Vannah, W. E., and Catheron, A. R., "Improved Flow Control with Long Lines," Paper 51-6-2, *Proceedings of the Sixth National Instrument Conference and Exhibit,* September 1951.

Warnock, J. D., "How Pneumatic Tuning Size Influences Controlability," *Instrumentation Technology,* February 1967.

Wightman, E. J., *Instrumentation in Process Control,* London: Butterworth, 1972.

3.9 Transmitters: Self-Checking and Self-Validating

J. BERGE (2005)

Communications Infrastructure (Permanently Connected):	Transmitters with 4- to 20-mA output shall conform to NAMUR NE-43 HART handheld such as the Smar HPC301 (about U.S. $1000) Foundation™ fieldbus host such as Smar SYSTEM302 (from U.S. $4000)
Costs (Typical base prices for pressure transmitters):	Microprocessor-based 4 to 20 mA, such as SMAR LD290, U.S. $500 HART, such as SMAR LD301, U.S. $700 *Foundation Fieldbus, such as SMAR LD302, U.S. $800*
Suppliers:	Refer to Sections 3.6, 3.7, and 3.8 in this volume and to Volume 1 of this handbook.

INTRODUCTION

Self-diagnostics are of great importance for both operation and maintenance. Self-diagnostics are important because the reliability of the measurements is essential for proper control. Control and alarm systems using invalid inputs are a safety hazard. Measurement validation is therefore paramount. Indications of invalid measurements can be used to shut down the loop or to activate backup systems.

Conventional and more sophisticated software tools such as statistical process control (Section 2.34), model-based predictive control (Section 2.14), and optimization (Section 2.20) should all work only with validated data and must know whether a measurement is invalid.

Until recently only two types of maintenance strategies were used in the processing industry: reactive maintenance (its response is usually too late) and preventive maintenance (too early). Both are costly and ineffective. The recommended maintenance scheme is a proactive one, which responds to the actual device status.

Such condition-based maintenance strategies rely on self-diagnostics to report on the health of field instruments to an asset management software system. Self-diagnostics detect and immediately signal the failure of a device. Diagnostics in conjunction with appropriate means of communication and advanced software tools permit remote troubleshooting.

Measurement validation is also critical in the proper operation of safety control systems. This requirement resulted in a trend that because switches provide no diagnostics, low-cost transmitters are taking their place in critical applications, where self-diagnostics are used.

LEVELS OF DIAGNOSTIC INFORMATION

Self-validation methods vary by both the type of measurement and by the supplier involved. In the past, most transmitters only had a single "general failure" indication for all faults. Today's transmitters, however, increasingly provide detailed diagnostics. The level of diagnostics available varies greatly by manufacturer.

Indeed, some transmitters are equipped to provide diagnostics that go down to the chip level, even if repair can only be performed at the board level. Because component level repairs are currently rare and are subject to approval from both a certification agency and the manufacturer, for most practical purposes, board-level diagnostics are sufficient.

Transmitters have diagnostics for the transmitter itself, i.e., the main circuit board, as well as diagnostics for its sensor or sensors. Transmitters can be categorized into two groups with respect to the sensor: those with integral sensors such as most pressure, flow, and level transmitters, and those with external sensors such as most temperature, pH, conductivity, etc. Devices without microprocessors have no diagnostics at all.

How Diagnostics Are Performed

Self-diagnostics of transmitters are active when the power supply is on. When this is the case, the integrity of data in the different nonvolatile memories is checked to ensure that they have not been corrupted (Figure 3.9a). Memory checks are also performed periodically while the device is in operation. Other diagnostics include consistency of configuration and calibration. If a device is in simulation mode, this is also reported by the diagnostics software.

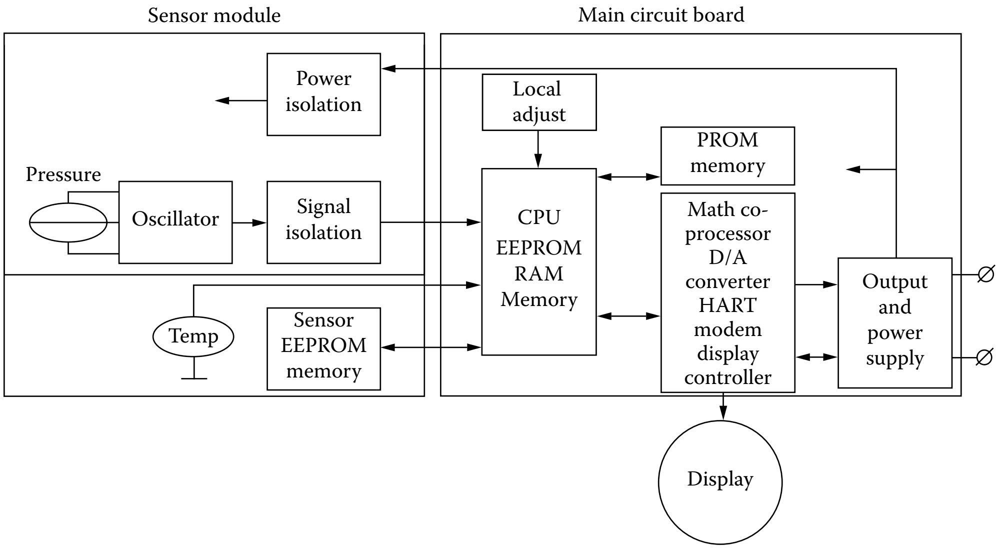


FIG. 3.9a
Diagnostics are performed by CPU firmware in conjunction with sensor's application specific integrated chips (ASIC).

Many types of device failures can manifest themselves in similar ways and it may not be possible to distinguish among them. Moreover, a failure of the main circuit board often results in the failure of a local display and communication channel, thereby making easy diagnostics impossible. Therefore, the user may not know exactly what is wrong with the sensor module. However, it suffices to know that the main circuit and not just the sensor module has failed because the main circuit has to be replaced in either case. Detailed analysis can be performed once the faulty circuit board has been replaced.

Transmitters with Integral Sensors Transmitters with integral sensors continuously check the sensor signal and compare it to expected readings to detect abnormal conditions based on manufacturer experience with the particular measurement method used. Pressure, level, and flow transmitters are all available with integral sensors. In order to allow for easier repair, the sensor is usually detachable from the main circuit board (Figure 3.9b).

The device or its sensor module provides temperature monitoring for temperature compensation of the measurement in addition to registering the violation of temperature limits. If a low limit is violated, this can be used to alert operators that the process fluid may be solidified or frozen and therefore the operation of the heat tracing system should be checked. Similarly, if a high-temperature limit is violated, it is desirable to check whether the sensor has been affected, degrading its accuracy or requiring recalibration. Such condition monitoring is required for proactive maintenance, which can ensure that small transmitter problems are corrected before they can cause large plant problems.

The symptoms of several types of sensor failures are similar and may not be distinguishable. For example, the leaking of pressure sensor diaphragms caused by corrosion

FIG. 3.9b
Pressure sensor module contains application specific integrated chips (ASIC) that handle diagnostics. (Courtesy of SMAR.)

may result in the same symptoms as some other type of sensor failure. In this case, the user may not know exactly what is wrong, but it suffices to know that the problem is with the sensor module and not the main circuit board and therefore the sensor has to be replaced. Detailed analysis can be performed after the faulty sensor has been removed. What is most important is to know that it was indeed a genuine sensor failure and not some kind of process upset or blocked impulse line that created the indication of failure.

Transmitters with External Sensors Transmitters utilizing external sensors include temperature, pH, and conductivity as well as several other types. A self-diagnosing transmitter periodically checks the external sensor to determine its health as well as checking the integrity of the wiring. A variety of tests can be performed, depending on the measurements and on the type of the primary sensor.

For example, in the case of thermocouple-type (TC) temperature sensors, the test may involve the sending of current through the leads to verify the continuity of the wires and to detect "burn-out" of the TC junction. Another test is to check the plausibility of the cold-junction temperature sensor reading.

For an RTD (Resistance Temperature Detector) sensor, the transmitter may measure the resistance of the individual sensor wires because excessive resistance can signify poor or wrong connections. Furthermore, the transmitter may contain internal comparison circuits, which can detect drift in its own internal secondary measurement circuitry.

More expensive but also more comprehensive diagnostics can be provided if two temperature sensor elements are used. If their readings excessively deviate from each other, that is used as an indication of failure.

pH is a notoriously difficult measurement and is perhaps one of the best examples where good use can be made of measurement validation. A modern pH transmitter (Figure 3.9c) continuously monitors both the measurement and the reference electrodes to detect mechanical damage of the sensor, contamination or blockage of the diaphragm, and aging or defects in the cabling. As a means to support proactive maintenance, the transmitter may even include a timer that alerts the operator when it is time to recalibrate. Furthermore, the transmitter may contain internal gain check to detect drift in its own internal secondary measurement circuitry.

FIG. 3.9c
Self-checking pH transmitter. (Courtesy of Mettler-Toledo.)

DIAGNOSTICS TRANSMISSION

The output signal of a transmitter can indicate its own health and the validity of its measurement. In case of analog transmitters, if the output is outside the normal operating range, that usually signals some type of failure. Intelligent transmitter communication indicates the status by using codes and parameters.

Analog Transmitters

Transmitters without microprocessors have practically no real diagnostics capability. However, thermocouple temperature transmitters may still have a "pull-up" resistor that prevents the input from floating in case of thermocouple burnout, which otherwise can drive the output to either of the following extremes: above 20 or below 4 mA.

Other analog transmitters work in a similar fashion. This scheme of protection is available for "live zero" signals such as 4 mA or 1 V, but for "dead zero" (zero-based) signals, the scheme does not work because it is usually not possible to go below 0 mA or 0 V. In these cases, an output that is below 1% is usually considered to be a failure indication.

Microprocessor-Based Transmitters

Microprocessor-based transmitters with 4- to 20-mA output and with or without highway addressable remote transducers (HART) have sensor diagnostics and can manipulate their outputs intelligently.

The NAMUR NE-43 (Normen-Arbeitsgemeinschaft für Meß-und Regulungstechnik in der Chemischen Industrie) standard defines the signal levels that indicate the health of instruments (Figure 3.9d).

To indicate that the measurement is "Good," the transmitter uses a signal in the range 4 to 20 mA. The wider range of 3.8 to 20.5 mA indicates that the measurement is outside the set range but probably still useful. This status may be considered "Uncertain." If the signal is between 3.6 and 3.8 mA or between 20.5 and 21 mA, the transmitter is "Bad." So, when the signal rises higher (20.5 to 21 mA) or drops lower (3.6 to 3.8 mA), a set of user-defined safety actions should be initiated.

HART Transmitters HART (highway addressable remote transducer) transmitters are smart instruments (see Section 4.11 in Volume 3 of this handbook) that provide slow digital communication in addition to their simultaneous 4- to 20-mA analog signals. The device status is included in all their communication responses.

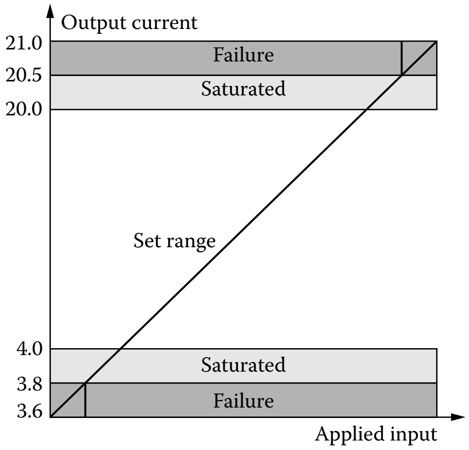


FIG. 3.9d
A failure indication recommendation by a German standard, NAMUR NE-43 (Normen-Arbeitsgemeinschaft für Meß-und Regulungstechnik in der Chemischen Industrie).

Because HART is relatively slow, the control loops rely on the 4- to 20-mA analog signal for control. Therefore, in most installations the HART communication capability is only utilized occasionally, by connecting a handheld tool. Therefore, in most plants the HART device rarely communicates the measurement validity digitally. Consequently, when the HART communication is not continuous it is even more important that the control system should detect any fault indication by the analog signal from the transmitter and bring it to the attention of the operator (Table 3.9e).

TABLE 3.9e
Descriptions of HART Errors

Error	*Description*
Field Device Malfunction	The device has failed. The measurement is invalid.
Configuration Changed	The device configuration has been changed, possibly affecting the measurement.
Cold Start	The device has restarted.
More Status Available	Additional detail status about the device health or measurement validity is available.
Analog Output Current Fixed	The device is in simulation mode. Output does not reflect measurement.
Analog Output Saturated	The output is out of range. The output does not reflect the measurement.
Nonprimary Variable Out of Limits	Auxiliary measurement, e.g., sensor temperature, is out of range. The measurement may be uncertain.
Primary Variable Out of Limits	The measurement is out of range. The output does not reflect the actual value.

However, in such control systems where the communication is "always on" and continuously polls the transmitters, a faulty sensor is reported instantly. In order to provide this mode of operation, it is necessary that the DCS/PLC systems use input modules with HART communication or an auxiliary HART multiplexer.

Foundation Fieldbus Transmitters Foundation Fieldbus transmitters are intelligent instruments (see Section 4.12 in Volume 3 of this handbook) with pure digital communication. Fieldbus communication is "always on." The health of the device and the validity of the measurement are continuously communicated. The extensive diagnostics capabilities and the ability to effectively report the health and measurement validity of the transmitted data are among the primary reasons for choosing Fieldbus.

In addition to the diagnostic and validity information listed in Table 3.9f, every transducer block has detailed diagnostic

TABLE 3.9f
Diagnostics and Validity Information Provided by Fieldbus Transmitters

Parameter	Description
*.status	All input and output parameters, including the measurement, as well as some contained parameters have a status associated with the value (see Table 3.9g)
BLOCK_ERR	All blocks have a summary of faults. In the resource block, this parameter reflects the health of the device as a whole. In the AI block, it represents the associated measurement (see Table 3.9h).
XD_ERROR	All transducer blocks have more detailed information about the fault.
MODE_BLK	All blocks have a mode. If the actual mode of the resource block does not match the target mode, this is an indication of some sort of problem with the device as a whole. If the actual mode of a transducer block does not match the target mode, this is an indication of a problem with the associated measurement.
RS_STATE	The resource block indicates the overall health of the device. If it is "failure," the memory or other hardware has a fault.

TABLE 3.9g
Fieldbus Measurement Status Attributes and Their Descriptions

Status Attribute	*Description*		
Quality	The validity of the measurement value may be Good, Bad, or Uncertain. There are two forms of Good; the one associated with measurements is "Good (Noncascade)."		
Substatus	Additional details hinting why the quality is Bad or Uncertain. For Good it contains alarm summary or other information used by the internal workings of the block.		
	Bad	Nonspecific	
		Configuration error	Some parameter is incorrectly configured.
		Not connected	Input is not linked.
		Device failure	Output has failed.
		Sensor failure	Sensor has failed.
		No communication—last usable value	Input is not being received. The value remains since last communication.
		No communication—with no usable value	Input is not being received. No earlier value is available.
		Out of Service	The block is out of service.
	Uncertain	Nonspecific	
		Last Usable Value	Input is disconnected. The value remains since earlier on.
		Substitute	The value is entered manually
		Initial Value	Value entered while in out-of-service mode.
		Sensor Conversion Not Accurate	Out of range or the sensor may have fouled.
		Engineering Unit Range Violation	Out of range.
		Subnormal	Auxiliary or redundant sensors have failed or are not in agreement
Limit condition	The limit condition for the value may be either High, Low, Constant, or none at all. High, low, and constant mean that the measurement does not represent the actual value, e.g., due to over range.		

parameters that are specific for the particular transmitter type, technology, and manufacturer.

The BLOCK_ERR parameter is found in all Foundation Fieldbus function blocks. It gives a summary of all faults in the device (Table 3.9h).

DIAGNOSTIC INFORMATION DISPLAYS

The fact that a transmitter failed or that it needs attention must be indicated both locally and in the control room in order to bring this information to the operator's attention. Fieldbus and HART configuration tools allow for effective management of failures, as was discussed in Section 1.6 of the first volume of this handbook.

The local indicator on the transmitter can display status, such as Bad, and can provide direct failure messages, such as sensor "burnout," both textually (Figure 3.9i) and symbolically (Figure 3.9j).

Health indication is very helpful for troubleshooting in the field. For this reason it is a good idea to use transmitters that are provided with local digital displays.

Usually, the operator in the control room is the first to notice that invalid measurements or transmitter failures have occurred. In order for the total process of transmitter self-checking and validation to be fully effective, the chain, consisting of failure detection in the transmitter, transmission of that information to the control system, and its presentation for the operators, must be fully integrated (Figure 3.9k).

In addition to the displaying of the status on the faceplate, any Uncertain or Bad status should also be logged and alarmed. Once operators detect an invalid measurement they can initiate the process that will determine the actual cause.

OPC (Object link embedding for Process Control) is a key technology serving to get data to the operator's workstations in the control room. This software architecture was described in Section 5.4 of the third volume of this handbook. It is recommended to use OPC in conjunction with HART or Foundation Fieldbus.

Portable and Handheld Displays

On the displays of handheld tools, technicians can see the detailed diagnostics of the transmitter. In the case of HART

TABLE 3.9h

Types and Descriptions of Universal Fieldbus Errors

Error	*Description*
Block configuration error	One or more parameters are wrongly configured, preventing the block from operating properly. The measurement may be invalid.
Link configuration error	One or more of the links for the block are wrongly configured.
Simulate is active (enabled)	For the resource block this means that simulation is permitted for the transmitter inputs. In an analog input it means that the input is actually being simulated and does not represent the actual measurement.
In Local Override (LO) mode	The block is in local override mode.
Device fault state is forced	"Fail safe" is forced in the device.
Device needs maintenance soon	The predictive diagnostics in the device indicates that it may soon be in need of service. The device may e.g., require calibration, cleaning, or some other service.
Input failure	The measurement has failed. The measurement may not be valid.
Output failure	The output has failed.
Memory failure	The device has a problem with one or more of its memories.
Lost static data	The device has lost its configuration. The measurement may be affected.
Lost nonvolatile data	The device has lost its configuration. The measurement may be affected.
Readback check failed	The actual output may not match the desired output.
Device needs maintenance now	The predictive diagnostics in the device indicates that is now in need of immediate service. The device may e.g., require calibration, cleaning, or some other service.
Powering up	The device is starting up
In Out-of-Service (OOS) mode	The mode of the block is set Out-of-Service.
Others	Additional device-specific status is available in other parameters.

systems, the portable handheld tool is brought out into the field and is connected at the transmitter (Figure 3.9l).

In a Fieldbus system the technician can "drill down" into the transmitter the detailed diagnostics information that can be helpful in the troubleshooting effort. Because the communication in Fieldbus systems is "always on," there is no need to locate and connect a handheld tool to obtain the diagnostics; they are available from the engineering station at any time (Figure 3.9m).

FOUNDATION Fieldbus transmitters are provided with more diagnostics, and the information provided is easier to access from the control room. These extensive diagnostics and the effective reporting of measurement validity are primary reasons for choosing Fieldbus.

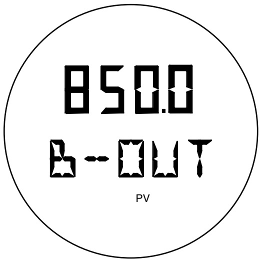
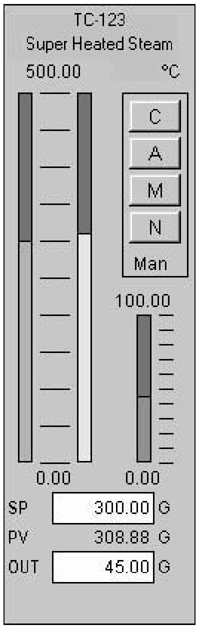


FIG. 3.9i

Failure message provided textually in a temperature transmitter display. (Courtesy of SMAR.)

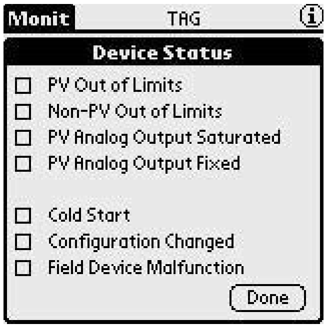
FIG. 3.9j

Failure message provided symbolically in a pH transmitter display. (Courtesy of Mettler-Toledo.)



TC-123
Super Heated Steam
500.00 °C
Man
SP 300.00 G
PV 308.88 G
OUT 45.00 G

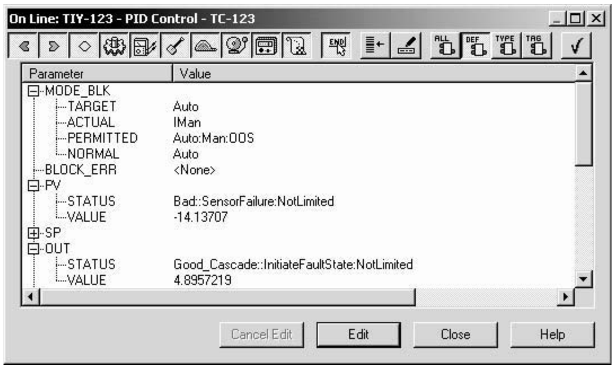
FIG. 3.9k
On the faceplate of the controller, next to the values of setpoint (SP), process variable (PV), and output signal, the letter "G" displays the good status of the complete system.



Device Status
☐ PV Out of Limits
☐ Non-PV Out of Limits
☐ PV Analog Output Saturated
☐ PV Analog Output Fixed

☐ Cold Start
☐ Configuration Changed
☐ Field Device Malfunction

FIG. 3.9l
The transmitter diagnostics information, which is presented on a HART handheld tool display.



On Line: TIY-123 - PID Control - TC-123

```
Parameter         Value
⊟-MODE_BLK
   ├-TARGET       Auto
   ├-ACTUAL       IMan
   ├-PERMITTED    Auto:Man:OOS
   └-NORMAL       Auto
 ├-BLOCK_ERR      <None>
⊟-PV
   ├-STATUS       Bad::SensorFailure:NotLimited
   └-VALUE        -14.13707
⊞-SP
⊟-OUT
   ├-STATUS       Good_Cascade::InitiateFaultState:NotLimited
   └-VALUE        4.8957219
```

FIG. 3.9m
The transmitter diagnostics information, which is presented on a Fieldbus software-supported display.

To effectively manage large numbers of installed transmitters and other instruments, the Fieldbus and HART network infrastructure should be complemented with powerful online plant asset management (OPAM) software (Figure 3.9n).

If the asset management software is Web-based, the maintenance system can securely be connected to the enterprise-wide intranet or the public Internet using appropriate firewalls and other means of protection. This permits diagnostics to be carried out from just about anywhere where it is possible to establish an Internet connection. For example, experts can access it from their homes, or access can be granted to the manufacturer's support center.

ACTING ON THE DIAGNOSTIC DATA

It is not possible to control a process if the transmitted information is invalid or Bad. However, it may be possible to maintain control while using uncertain measurements, such as readings that are slightly out of range. In general, plant safety is improved if transmitter self-diagnostics are utilized to improve the validity of measurements.

Failsafe and Alarm Actions

Even the simplest analog control loop can be designed to fail safely. For example, in case of level control, a failsafe transmitter will generate a high analog output (say 21 mA), if the sensor fails. Therefore, the controller or alarm system will interpret a 21-mA signal the same as if the tank is overfilled and thus will automatically close the filling valve.

Sophisticated DCS and PLC may interpret NAMUR NE-43 signal levels and thus determine if signal quality is Good, Uncertain, or Bad.

HART communication is too slow for closed-loop control or shutdown interlocks and therefore both controls and alarms utilize 4- to 20-mA analog signals.

Converters exist that can tap the HART communication from the signal lines and can activate relays in case of failure. Such relays can tie to control systems, which do not communicate HART but do need to know the transmitter status.

In a Fieldbus-based control system, safe loop action is part of the IEC 61804-1 function block diagram language for building control strategies, which is an integral part of the FOUNDATION Fieldbus system architecture. Values communicated between function blocks, such as from an analog input (AI) block in a transmitter to a PID block in a control valve positioner, are accompanied by their status.

A "Bad" measurement status from the transmitter can automatically switch the loop to a manual mode of operation or optionally, the PID control block can bring the control valve to its predetermined safe position (Figure 3.9o).

An advantage of the Fieldbus function block language is that the interlocks are built into the control blocks. Therefore there is no need to configure and validate additional logic to implement the interlocks. Moreover, because Fieldbus is a standard, the interlocks work across all devices conforming

Device Status

MAXIMUM MEASURED PRESSURE: 3598
MAXIMUM MEASURED TEMPERATURE: 84.7
ACTUAL OFFSET: 0
ACTUAL SPAN: 1

Diagnosis

- Power Up
- Sensor Failure
- Memory Failure
- Out of Service
- Device Needs Maintenance Soon
- Device Needs Maintenance Now
- Calibration Error
- Block Configuration Error
- Data Integrity Error
- Software Error
- Electronics Failure
- General Error

FIG. 3.9n
Transmitter diagnostics using Web-based asset management (OPAM) software.

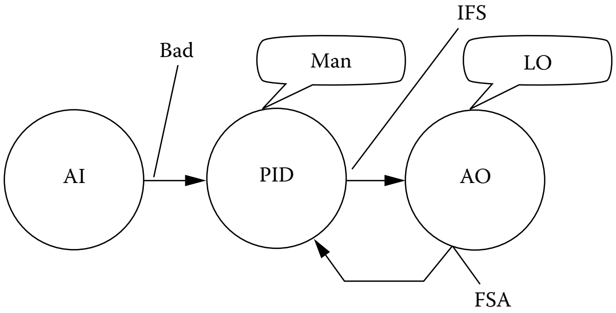


FIG. 3.9o
Status and operating mode propagation in a FOUNDATION™ fieldbus control system, using function block language.

to the standard. Therefore it is advisable to use transmitters, valve positioners, and central controllers that are based on the FOUNDATION fieldbus blocks rather than using proprietary languages. This can ensure that the measurement validity and other status information is propagated throughout the control strategy and not lost along the way.

Within the fieldbus PID block it is possible to set whether the status "Uncertain" shall be treated as "Good" or as "Bad." This makes it possible to be selective when balancing production availability against plant safety on a loop-by-loop basis. For loops that require high availability, an uncertain status is configured as good, thus permitting control to continue under such conditions. For loops where safety is the primary concern, the uncertain status can be treated as bad, thus shutting the loop down.

Reference

1. Berge, J., *Fieldbuses for Process Control—Engineering, Operation, and Maintenance*, Research Triangle Park, NC: ISA, 2002.

3.10 Transmitters: Smart, Multivariable, and Fieldbus

J. BERGE (2005)

Communications Infrastructure (Permanently Connected):	Transmitters with 4- to 20-mA output shall conform to NAMUR NE-43 HART handheld such as the Smar HPC301 (about U.S. $1000) Foundation™ fieldbus host such as Smar SYSTEM302 (from U.S. $4000)
Costs (Typical base prices for pressure transmitters):	Microprocessor-based 4- to 20-mA, such as SMAR LD290, US$ 500 HART, such as SMAR LD301, U.S. $700 Foundation fieldbus, such as SMAR LD302, U.S. $800
Suppliers:	Refer to Sections 3.6, 3.7, and 3.8 in this volume and to Volume 1 of this handbook.

INTRODUCTION

Transmitters are referred to as “analog,” “microprocessor based,” “intelligent,” “smart,” “fieldbus,” etc. depending on the technology used.

An “analog” transmitter has no microprocessor and therefore its linearization, temperature compensation, and diagnostics are rudimentary. An analog transmitter is configured by switches and jumpers and adjusted using potentiometers.

A “microprocessor-based” transmitter is capable of much more sophisticated linearization, temperature compensation, and diagnostics. However, a microprocessor-based transmitter still only has an analog output, such as 4- to 20-mA, and therefore only transmits rudimentary diagnostics. A microprocessor-based transmitter is configured from the local display and buttons.

An “intelligent” transmitter too has a microprocessor, but it also has digital communication capability, which permits remote diagnostics and configuration.

A “smart” transmitter is an intelligent transmitter that has analog output but that also permits simultaneous digital communication. The HART protocol is a typical example. Not all intelligent transmitters are smart because some do not provide simultaneous digital communication and analog output, as the analog output is disrupted while communicating digitally.

A “fieldbus” transmitter has only digital communication. Today, the same supplier often markets microprocessor-based, smart, and fieldbus transmitters that are externally identical but differ in their internal electronics and software (Figure 3.10a). In general, a smart or fieldbus transmitter consists of the sensor or input circuitry, the microprocessor, the communication block and, in the case of smart transmitters, an analog output (Figure 3.10b)

OPERATION AND PERFORMANCE

Analog circuits are subject to drift over time and are affected by ambient temperature variations. Moreover, there is no obvious way to check whether or not an analog signal is invalid because signal distortion just changes one valid value to another valid value, as all signal levels are valid. For example,

FIG. 3.10a
Microprocessor-based, smart, fieldbus, or Profibus-PA transmitters can be externally identical and differ only in the electronics and software used. (Courtesy of SMAR.)

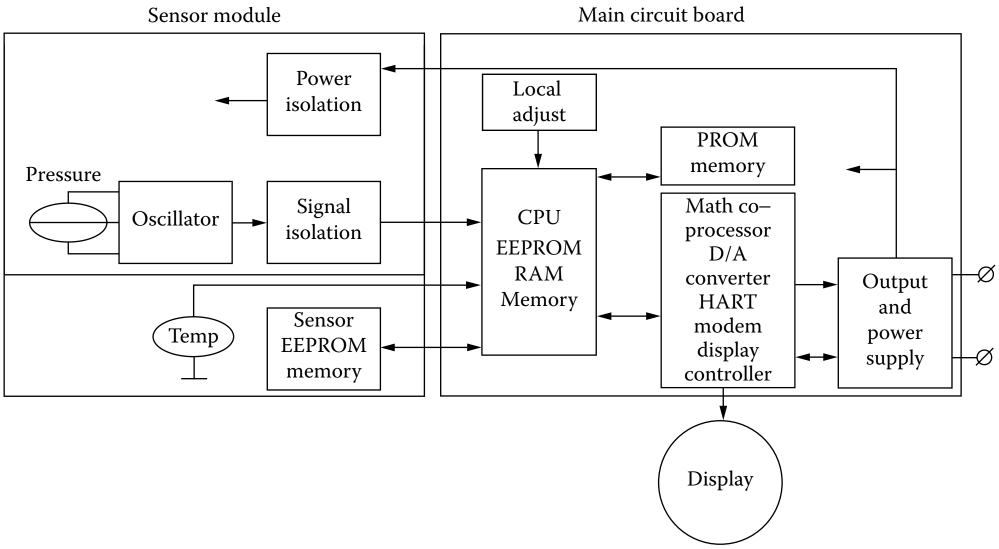


FIG. 3.10b
Smart pressure transmitter with auxiliary cold junction temperature. (Courtesy of SMAR.)

a signal that should be 5 V may vary between 4.85 and 5.15 V due to noise, but there is no way to tell the value is true because 4.85 or 5.15 V are also valid signals.

Similarly, if, because the supply voltage is insufficient, a signal that should be 19 mA is limited to 18 mA, there is no way to tell that a distortion has occurred because 18 mA is also a valid signal. Distorted or limited signal values cannot be detected in analog circuitry because these signals cannot be distinguished from the genuine process measurements.

For these reasons, there is a trend to obtain greater accuracy and stability in transmitters by eliminating more and more analog components in the measurement circuitry and aiming to convert it to a completely digital system that has digital sensors, processing, and communication.

Digital Sensors

In smart and fieldbus transmitters the sensor is usually digital. Keeping the sensor signal digital from the very beginning provides for more precise signal processing and lowers noise pickup and signal degradation.

The signal from a digital sensor gives the value of the measurement as a function of frequency or time. Such time-based functions do not depend on the signal amplitude. For example, a capacitance-type pressure sensing element can be a part of an oscillatory circuit with a frequency output. Similarly, an ultrasonic flow meter detects the time of flight. Other digital sensors operate on frequency phase-shift.

Timers in the microprocessor easily measure frequency and duration using precise crystal clocks. In such direct digital measurements there is no need for an A/D converter, which also eliminates the conversion error and lowers inaccuracy by eliminating quantization error.

In earlier “semismart transmitter” designs with analog sensors, the outputs from resistive strain gauges or piezo or conventional capacitance sensors were first converted to a voltage before being converted to the digital format. Most of these transmitter designs have been replaced by digital sensing.

When the detector is inherently analog, such as in case of some temperature sensors, A/D converters are required, and their resolution determines the accuracy of the conversion.

Sensor Compensation and Characterization

Firmware in the microprocessor provides a variety of functions, including sensor temperature chatacterization and linearization, communication, self-diagnostics and sensor diagnostics, as well as flow totalization, PID control capability, and other functions depending on type of the transmitter.

No sensor has perfect linearity or temperature stability. It is for that reason that the microprocessor must characterize and compensate the sensor signal. For external sensors such as thermocouples and RTD (Resistance Temperature Detectors), the relation between detected and sensor output is well established. In these cases, standard tables or polynomials are used to convert the thermocouple’s millivolt or the RTD’s ohm signals to temperature in the transmitter firmware. Similarly, the temperature effect on a pH sensor is well understood, and its compensation is preprogrammed in a pH transmitter.

The characteristics of many pressure and flow transmitters are not predetermined but must be obtained by testing. In case of these units, all sensors are tested in the factor by exposing them to their operating range of inputs at different temperatures (Figure 3.10c).

The test data obtained are used to compute the characterization coefficients that are unique to the particular sensor, and these coefficients are stored in the memory of the sensor module. When the power supply is turned on, the transmitter loads these characterization coefficients from the sensor module into

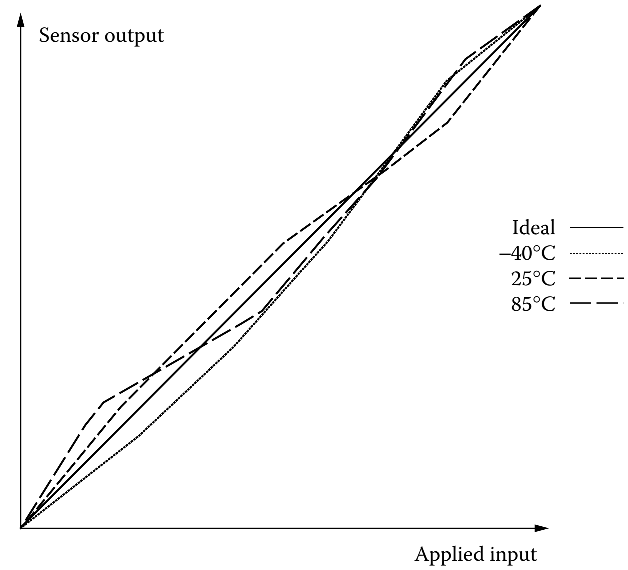


FIG. 3.10c
Exaggerated illustration of sensor characterization test results.

its characterization algorithm. The goal of this procedure is a temperature-compensated and linear measurement.

Because the sensor characterization coefficients, along with other pertinent information such as sensor serial number, range, and materials of construction, are stored in the sensor module itself, when the sensor is replaced the new characterization coefficients are automatically loaded into the microprocessor.

By reducing the measurement error, it becomes possible to increase the rangeability of smart transmitters and provide turndowns, sometimes as high as 120:1, while still keeping the error within acceptable limits. The increased turndown makes the transmitter suitable to a wider range of applications. Two of the main limitations to accuracy is the hysteresis and the thermal hysteresis of the sensor because these cannot be corrected.

Multivariable and Inferential Sensors

Once the signals of individual sensors have been characterized, they can be combined in a multivariable transmitters to compute such process variables as density or concentration (Figure 3.10d). Multivariable transmitters perform several measurements and use these to obtain an inferential reading of a single property and generate a single 4- to 20-mA output. For example, a multivariable transmitter may measure density and process temperature in order to compute concentration or density based on the Brix or API tables.

Similarly, a multivariable transmitter may measure the differential pressure drop across and orifice plate, static line pressure, and process temperature in order to compute either standard volumetric flow or mass flow. Such a transmitter also has auxiliary sensors to compensate the differential pressure and static pressure sensors for temperature effect before they are used in the AGA (American Gas Association) equations to compute standard volumetric gas flow or in the AIChE (American Institute of Chemical Engineers) equations to compute mass flow for gas or liquid flows. (These equations can be found in Volume 1 of this handbook.)

FIG. 3.10d
Multivariable transmitter for the detection of density. (Courtesy of SMAR.)

Other multivariable transmitter applications include the use of multichannel temperature transmitters in order to obtain differential, average, minimum, or maximum temperature readings. A multichannel transmitter is connected to several sensors that are installed in various points in the process piping and equipment.

In addition to multivariable transmitters, particularly in the case of FOUNDATION fieldbus, transmitters can also perform calculations such as flow totalization and PID control. These capabilities are also present in some smart transmitters.

Some transmitters can also perform process linearization such as the use of hydrostatic tank gauging (HTG) lookup tables to relate the level in cylindrical, spherical, and odd-shaped tanks to volume. Fieldbus transmitters, particularly those supporting instantiable function blocks, may also support signal selection, limiting, and arithmetics. Today, it is up to the system design engineer to decide whether the above-mentioned calculations should be performed in the field instruments or in the central computers.

HART Communication

Digital communication such as HART and FOUNDATION fieldbus permits remote monitoring, diagnostics, configuration, and calibration. In the case of fieldbus networks, the digital communication can also be used to perform closed-loop control.

In digital transmission there are only two valid states for the signal (zero and one). Therefore, a large amount of interference is required to distort the signal by changing one to the other. This makes the digital signal very robust. In addition, error detection techniques are provided to detect errors due to noise. Errors due to D/A conversion are eliminated by not requiring such conversions at all.

HART transmitter communicates at 1200 bit/s, which is too slow for closed-loop control but adequate for the technician to perform interrogation using a handheld tool. In order to guarantee fast loop response, in most applications the analog 4- to 20-mA signal is used for control (see Section 4.11 in Volume 3 of this handbook). HART transmitters are typically wired in a point-to-point fashion. Multi-dropping up to 15 transmitters on the same pair of wires is possible but is rarely used as it is slow and not well supported in control systems.

Fieldbus Transmitters

FOUNDATION fieldbus transmitters use the H1 version of fieldbus communicaton at 31.25 kbit/s, which is fast enough for closed-loop control. This speed eliminates the need for using analog signals (see Section 4.12 in Volume 3 of this handbook).

Typically 12 to 16 devices, which can be a mix of types and suppliers, are multi-dropped on the same pair of wires. Fieldbus is much more than a digital equivalent of 4- to 20-mA transmission. Fieldbus can be used in place of Distributed Control Systems (DCS) because the devices in the field can also perform the control functions. This reduces the controller and I/O-subsystem complexity because the majority of controls tend to be simple monitoring and basic control functions, computation, or logic, which all can be done in the field instruments.

Fieldbus transmits the measurement status information along with the sensor's reading. The status information includes measurement quality and limit condition information. The measurement quality is usually stated as Good, totally Bad, or Uncertain, e.g., having an error of not exceeding a few percent. Other function blocks use this measurement quality information to automatically initiate safety steps or to exclude that measurement when doing selection or calculating an average.

Smart and fieldbus transmitters use their auxiliary sensors to provide temperature and static pressure compensation of the primary sensor's reading. Such compensated smart transmitters still have only one output because the other measurements only serve to detect the existing conditions under which the primary sensor operates.

BENEFITS OF ADVANCED TRANSMITTERS

Many benefits resulted from the use of digital circuitry, microprocessors, and digital communications in smart and fieldbus transmitters. These include greater accuracy and stability, lower temperature sensitivity, greater flexibility due to higher turndown ratio, remote parameterization, calibration, diagnostics, and monitoring. Online diagnostics also improve plant safety.

Savings

Savings can be achieved in both capital expenditures (CAPEX) and operational expenditures (OPEX). Because fieldbus is a system architecture and not just a 4- to 20-mA replacement, fieldbus savings exceed the savings associated with transmitters because they also improve the operation of the process.

Although a smart transmitter costs about $100 more than its analog counterpart, the savings in commissioning expenses due to quicker parameterization and loop checking result in a lower installed cost for a smart transmitter.

Maintenance costs are also reduced because of the easier and faster diagnostics. In addition, the monitoring of ambient conditions allows plants to switch to a proactive maintenance program. In general, effort and resources required for maintenance are low in comparison with those of analog instruments.

If "always on" communication is used, the ability to perform remote diagnostics and other functions can reduce the number of required field trips as the causes of trouble can often be determined from the control room. For example, it is possible to determine that loss of control is caused not by transmitter failure but by a process upset, and thus one can avoid unnecessary maintenance. Also, because failures can be pinpointed more accurately, repairs tend to be faster. In addition, the availability of user-friendly handheld tools will make calibration easier and faster, while improved calibration will improve the performance of the transmitter.

Improved diagnostics guarantee that failures can be detected and corrected faster so that production can get back online sooner, thus reducing downtime. By monitoring sensor temperature, it is possible to protect against overheating- or freezing-related failures. By monitoring the health of the transmitters, it is possible to schedule maintenance shutdowns so that many instruments are simultaneously maintained, rather than experiencing several unexpected shutdowns due to surprise instrument malfunctions. In comparison with analog transmitters, the resources required for maintenance are also lower.

More accurate measurements result in better product quality, lower risk of off-spec product, and reduced costs of disposal or rework. Higher accuracy measurement also allow

FIG. 3.10e
HART to current converter. (Courtesy of SMAR.)

one to keep set points closer to values corresponding to maximum production efficiency.

Installation and Commissioning

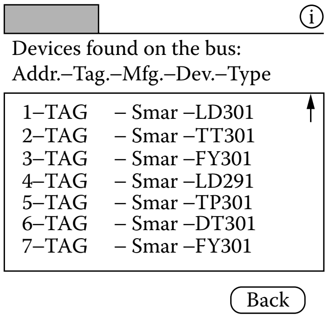
Smart transmitters are usually loop powered and are wired the same way as their analog counterparts are. One should make sure that a typically 250-ohm shunt resistor for the minimum current loop load is installed. That resistor is usually already provided on the input of the controllers, I/O-modules, recorders, or indicators.

As was described in Section 4.11 of Volume 3 of this handbook, because HART operates in a frequency range that is above the registered range of most process instrumentation, smart transmitters are completely compatible with analog instrumentation. Naturally, in order to guarantee intrinsically safe operation, special safety barriers are required.

Because the smart multivariable transmitters provide only a single analog output, only the calculated inferential measurement is available in the control room; the multivariable inputs, such as temperature, are not. However, HART-to-current converters are available, which use HART communication to interrogate multivariable transmitters and generate 4- to 20-mA signals based on their digital inputs (Figure 3.10e).

Using a HART-to-current converter with one or three outputs, it is possible to send all the process measurement signals to the control system even if it does not communicate HART.

Fieldbus transmitters connect using H1 wiring based on the IEC 61158-2 type 1 standard. Fieldbus transmitters are wired in parallel, just like other fieldbus devices (see Section 4.12 in Volume 3 of this handbook). Note that power supply impedances are required for regular installations, and special safety barriers are required to provide intrinsic safety (see Section 2.2 in Volume 3 of this handbook). For fieldbus systems, transmitters with low power consumption should be selected so that more devices, longer wires, and reduced numbers of safety barriers can be used.



Devices found on the bus:
Addr.–Tag.–Mfg.–Dev.–Type

1–TAG – Smar –LD301
2–TAG – Smar –TT301
3–TAG – Smar –FY301
4–TAG – Smar –LD291
5–TAG – Smar –TP301
6–TAG – Smar –DT301
7–TAG – Smar –FY301

FIG. 3.10f
Handheld tools can be used to identify which HART transmitters are connected to a bus.

By using handheld tools, one can remotely detect which transmitters are connected (Figure 3.10f).

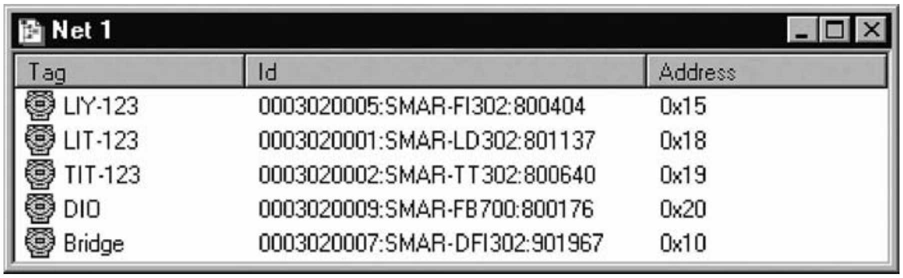
It is also possible to check a loop by switching the transmitter into the simulation mode and setting its output current to any value in the 4- to 20-mA range, independent of the measurement. This makes it possible to verify the reading at the workstations and on the panel.

When a fieldbus transmitter is connected to a HSE (High Speed Ethernet) linking device or to a proprietary interface module, this will be automatically detected and shown in the live list on the configuration tool (Figure 3.10g).

Parameterization

The term "parameterization" implies the process of setting the operating parameters, such as the range of the transmitter.

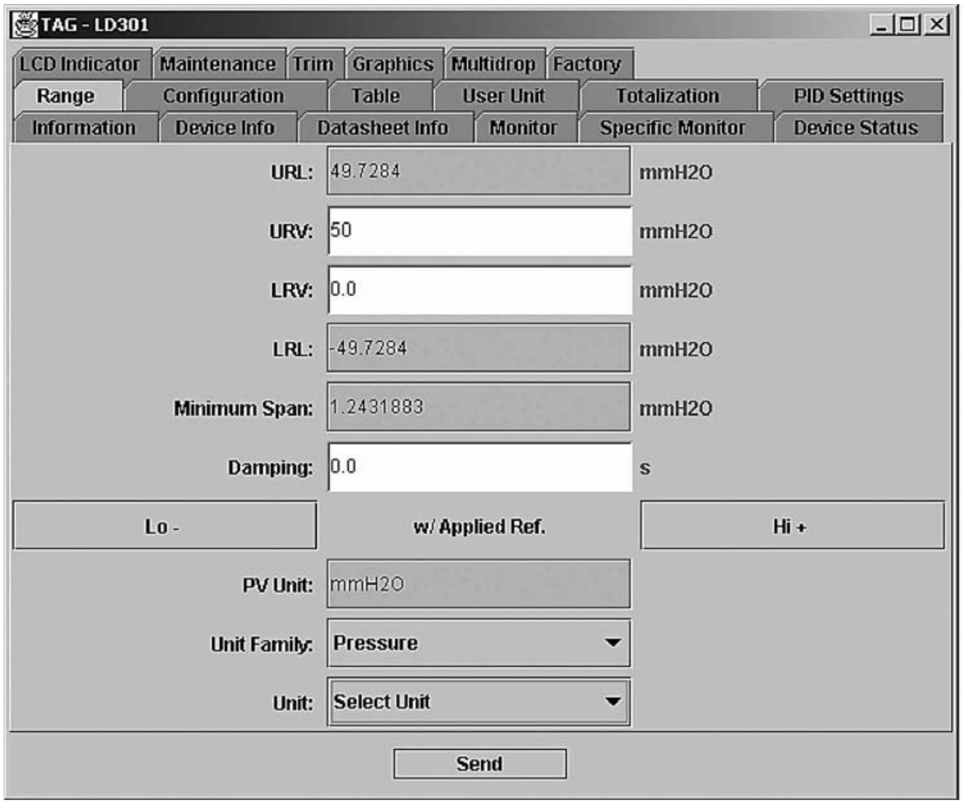
Analog transmitters are configured by using switches or jumpers in order to select sensor type as well as elevation and suppression. These are finely adjusted using potentiometers. To get square root extraction it is necessary to change the



Net 1

Tag	Id	Address
LIY-123	0003020005:SMAR-FI302:800404	0x15
LIT-123	0003020001:SMAR-LD302:801137	0x18
TIT-123	0003020002:SMAR-TT302:800640	0x19
DIO	0003020009:SMAR-FB700:800176	0x20
Bridge	0003020007:SMAR-DFI302:901967	0x10

FIG. 3.10g
Fieldbus transmitters are automatically detected in the live list on the configuration tool.



URL: 49.7284 mmH2O

URV: 50 mmH2O

LRV: 0.0 mmH2O

LRL: -49.7284 mmH2O

Minimum Span: 1.2431883 mmH2O

Damping: 0.0 s

PV Unit: mmH2O

Unit Family: Pressure

Unit: Select Unit

FIG. 3.10h
The means of setting such parameters as the upper and lower range values (URV and LRV) of a smart pressure transmitter.

circuit board. Smart and fieldbus transmitters can be parameterized from a handheld tool or by the use of software by keying in the desired sensor type, range, transfer function, etc.

Because the sensor and input circuitry errors are absolute, the percentage error gets worse as the span is reduced. This limits the transmitter's turndown, particularly those that use low-resolution A/D converters. Most smart transmitters maintain their nominal accuracy even if their span is turned down by a ratio of 10:1. When the spans are reduced further (greater turndown), the accuracy becomes progressively worse.

Unfortunately, many transmitter manufacturers do not state this clearly, but it is always the case. Therefore, transmitters should be specified by stating their minimum and maximum span requirement and the maximum allowable error limit (inaccuracy), in units of percent of span, which must hold anywhere within that range. The minimum span limit prevents the inadvertent setting of an excessively small span, resulting in gross errors. Some transmitters may permit turndowns as high as 400:1, but the corresponding error can become very high.

Perhaps the most important setting for a smart transmitter is its range, i.e., the process variable measurements corresponding to the outputs of 4 and 20 mA (0 and 100% of scale). The range can be set remotely without requiring an input just by keying in the desired range value (Figure 3.10h). While this is possible, it is not always desirable. In case of level loops, for example, it might be desirable to "zero" the transmitter by bringing the actual process level to an elevation that should be treated as zero and use the corresponding transmitter output as the "zero" for the loop. The HART protocol supports both methods of setting the range, which should not be confused with calibration trim.

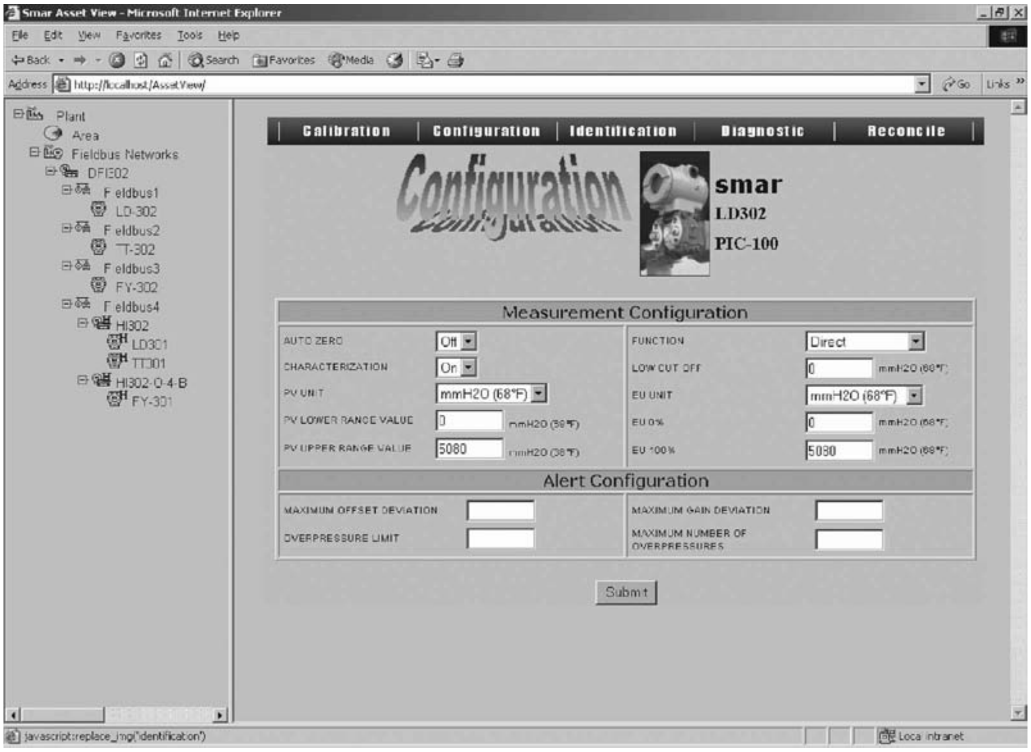
Because fieldbus communicates the measured process variable values in engineering units, it is often not necessary to set a range in the transmitter. When the transmitter measures 5080 mmH_2O, it can transmit that value. On the other hand, range needs to be set in the transmitter for inferential measurement applications, such as when using a differential pressure transmitter on an orifice-type flow element (Figure 3.10i).

For fieldbus, instead of setting the range in the transmitter, it is often set in the controller by adjusting a scale in the PID control function block.

Integration into DCS and PLC Systems

Smart and fieldbus transmitters provide new capabilities but also place new requirements on the control system software and networks. Most DCS and PLC systems that are in use today were designed when analog transmitters with single outputs were used. Today, there is a tremendous increase in diagnostic and other data, which the existing control system networks and software are not able to manage.

Smart transmitters can be connected into traditional DCS and PLC systems or to single station controllers by



Measurement Configuration

AUTO ZERO: Off
CHARACTERIZATION: On
PV UNIT: mmH2O (68°F)
PV LOWER RANGE VALUE: 0 mmH2O (68°F)
PV UPPER RANGE VALUE: 5080 mmH2O (68°F)
FUNCTION: Direct
LOW CUT OFF: 0 mmH2O (68°F)
EU UNIT: mmH2O (68°F)
EU 0%: 0 mmH2O (68°F)
EU 100%: 5080 mmH2O (68°F)

Alert Configuration

MAXIMUM OFFSET DEVIATION:
OVERPRESSURE LIMIT:
MAXIMUM GAIN DEVIATION:
MAXIMUM NUMBER OF OVERPRESSURES:

FIG. 3.10i
Asset management software.

their 4- to 20-mA output signals, in which case they are configured by handheld tools. However, in order to fully benefit from smart instruments, the communication channels must be "always on," meaning a permanent HART communication connection. This can either be achieved by using a separate multiplexer, which is a gateway, in parallel with the analog I/O-subsystem that communicates with the smart transmitters and links to the host computers, which are using a faster protocol and media such as Modbus or Profibus. A neater alternative is a system where HART communication is built into the AI module.

Fieldbus is a system architecture in itself, complete with field-level networking (H1) and host-level networking (HSE), as well as a graphical function block programming language to build the control strategy. Fieldbus transmitters sit on the H1 network that connects to the HSE network through standard linking devices. However, some systems still use a proprietary host-level network and controllers and connect to H1 fieldbus through interface modules.

Although many controller and system manufacturers also make transmitters, none of them make all the different kinds of transmitters used in the processing industries, and none of them are best in all categories. Therefore, there is always a need to mix and match transmitters from different suppliers in the overall control system.

When equipment from different suppliers is combined, interoperability becomes an issue. It should not be confused with interchangeability. Interoperability is the ability of different equipment to work together, allowing a single configuration tool to be used for all brands of transmitters. Control loop integration when using smart transmitters is easy because smart transmitters use the universally accepted, but functionally limited, 4- to 20-mA signal.

Building Control Loops

The FOUNDATION fieldbus technology includes not only a communication protocol but also a graphical programming language for building control strategies. This language is based on function blocks and includes an AI block, which is the principal block in a transmitter. Since many function blocks are standardized, especially in terms of their input and output parameters as well as the way they are linked and executed, a fieldbus transmitter can easily operate in a control loop with devices from other manufacturers.

For a fieldbus configuration tool to interact with a fieldbus transmitter, it is necessary for the standard Device Description (DD) and capabilities files to have been loaded. These files are provided by the transmitter manufacturer and contain all the information the host software requires in order to access all the features and capabilities the transmitter manufacturer has put into the transmitter and arranged logically to reflect typical use of the instrument.

However, some systems do require additional proprietary files for each device type, but in these cases, they must be obtained from the systems manufacturer and may not support all kinds of devices. Because fieldbus devices are very sophisticated and extremely rich in features, there is essentially no way a configuration tool can work in a meaningful way without the DD files.

HART also uses DD, and it is implemented in universal handhelds that can be used to fully configure any HART transmitter. However, HART functionality is divided into groups of commands. Universal commands are supported by all HART devices, and most support common practice commands. Therefore a handheld tool and PC software or even a smart calibrator can access many functions in a HART transmitter even without using DD.

Reference

1. Berge, J., *Fieldbuses for Process Control: Engineering, Operation and Maintenance*, Research Triangle Park, NC: ISA, 2002.

Control Room Equipment

4

4.4
CONTROLLERS—ELECTRONIC ANALOG AND DIGITAL 633

4.5
CRT DISPLAYS 650

4.6
DCS: BASIC TRENDS AND ADVANCES 663

4.7
DCS: CONTROL AND SIMULATION ADVANCES 677

4.8
DCS: INSTALLATION AND COMMISSIONING 687

4.9
DCS: INTEGRATION WITH BUSES AND NETWORKS 693

4.10
DCS: INTEGRATION WITH OTHER SYSTEMS 700

4.11
DCS: MANAGEMENT OF ABNORMAL CONDITIONS 706

4.12
DCS: MODERN CONTROL GRAPHICS 720

4.13
DCS: OPERATOR'S GRAPHICS 727

4.14
DCS: SYSTEM ARCHITECTURE 739

4.15
DIGITAL READOUTS AND GRAPHIC DISPLAYS 757

4.20
RECORDERS, OSCILLOGRAPHS, LOGGERS, TAPE RECORDERS 818

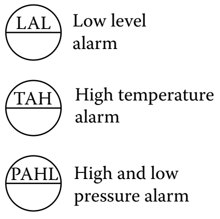
4.21
SWITCHES, PUSHBUTTONS, KEYBOARDS 829

4.22
TOUCH-SCREEN DISPLAYS 845

4.23
UNINTERRUPTIBLE POWER AND VOLTAGE SUPPLIES (UPS AND UVS) 854

4.24
WORKSTATION DESIGNS 868

4.1 Annunciators and Alarms



Flow sheet symbol

J. A. GUMP (1972, 1985) **B. G. LIPTÁK** (1995, 2005)
E. M. MARSZAL (2005)

Types:

A. Audiovisual Annunciators: integral, remote, and semigraphic systems with audible and visual display and electromechanical (relay) or solid-state (semiconductor) designs
B. Recording Annunciators: integral, solid-state systems with recorded printout
C. Bargraphs
D. Vocal Annunciators: integral, solid-state systems with audible command message

Cost per Alarm Point:

Integral cabinet costs $75 to $175; remote system, $125 to $250; semigraphic system, $125 to $250; recording annunciator or annunciator with communications will add about 30% to the cost per point. These figures are budgetary in nature (± 20%), and a number of variables can affect the price. These factors include system size, window size, number of alarm points per board, field contact voltages, type of lamp, communications options, and required certifications (e.g., Class I, Div. 2, etc.).

Partial List of Suppliers:

4B Components Ltd. (A) (www.go4b.com)
Acromag Inc. (A) (www.acromag.com)
Adaptive Micro Systems Inc. (A) (www.adaptivedisplays.com)
Advotech Inc. (D) (www.advotechcompany.com)
Ametek Power Instruments (A, B, mosaic graphic) (www.ametekpower.com)
Barnett Engineering Ltd. (A) (www.barnett-engg.com)
Beta Calibrators div. Hathaway Process Instrumentation Corp. (A, B) (www.the-esb.com)
CAL Controls Inc. (www.cal-controls.com)
CEA Instruments Inc. (www.ceainstr.com)
Cole-Parmer Instrument Co. (www.coleparmer.com)
CTC Parker Automation (www.ctcusa.com)
Daytronic Corp. (A) (www.daytronic.com)
Devar Inc. (A) (www.devarinc.com)
Draeger Safety Inc. (www.draeger.com/gds)
Druck Inc. (www.pressure.com)
Fisher Controls International Inc. (A) (www.fisher.com)
Flow Tech Inc. (www.flowtechinc.com)
Fluid Components International (www.fluidcomponents.com)
Foxboro Co. (A) (www.foxboro.com)
GE Kaye div. General Electric Co. (A)
General Monitors (www.generalmonitors.com)
Graybar Electric Co. (www.graybar.com)
Honeywell Industry Solutions (www.iac.honeywell.com)
ImageVision Inc. (www.imagevisioninc.com)
Mauell Corp. (A) (www.mauell-us.com)
Matrikon Inc. (www.matrikon.com)
Metrix-PCM/Beta (A) (www.metrix1.com)
Moore Industries Inc. (www.miinet.com)
North American Manufacturing Co. (A) (www.namfg.com)
Oceana Sensor (www.oceanasensor.com)
Phonetix Inc. (www.sensaphone.com)
Powers Process Control, A Unit of Mark Controls Corp. (A) (www.powerscontrols.com)
Precision Digital Corp. (www.predig.com)

ProSys Inc. (www.prosysinc.com)
Puleo Electronics Inc. (A) (www.annuciator.com)
Raco Manufacturing and Engineering (www.racoman.com)
Robicon div. High Voltage Engineering Corp. (A) (www.robicon.com)
Ronan Engineering Co. (A, B, mosaic graphic) (www.ronan.com)
Schneider Electric/Square D (www.squared.com)
Scott Aviation div. Tyco Inc. (A) (www.tycoelectronics.com)
Seekirk Inc. (D) (www.seekirk.com)
Sierra Monitor (www.sierramonitor.com)
Swanson Engineering & Manufacturing (A) (www.amtonline.org)
Texmate Inc. (A) (www.texmate.com)
Thermo Brandt Instruments (www.brandtinstruments.com)
Tips Inc. (www.tipsweb.com)
Transmation Inc. (D) (www.transmation.com)
Trip-A-Larm (A) (www.modicon.control.com)
Visi-con div. Visicomm Industries (A) (www.alarmpanels.com)
Vorne Industries Inc. (A) (www.vorne.com)
Western Reserve Control (www.wrcakron.com)
White Electronic Designs Corp. (C) (www.motionnet.com)
Wilkerson Instrument Co. (A) (www.wici.com)
Zetron Inc. (A) (www.zetron.com)

In addition to this section, safety alarm systems are discussed in several other parts of this handbook, particularly in connection with DCS and CRT systems in Chapter 4 and PLCs in Chapter 5. The subject of process alarm management is separately covered in Section 1.6. This section concentrates on dedicated, conventional annunciators and other alarm devices.

INTRODUCTION

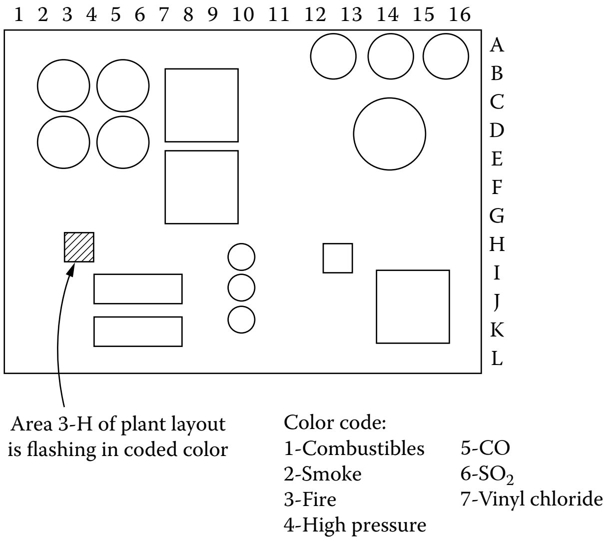
The purpose of an alarm system (annunciator) is to bring attention to an abnormal or unsafe operating condition in the plant. Traditional annunciators used discrete alarm modules for this purpose. These dedicated hardware units are diminishing in numbers yet are still used in installations where simplicity is desired or where separation from the basic process control system is required for safety reasons.

In some installations where traditional units have been replaced by PLC- or DCS-based annunciators, the recognition of and response to alarm conditions have deteriorated because on computer screens they are not very visible and can go unnoticed. In addition, because of the low incremental cost of adding new alarm points, excessive numbers of alarms been configured. Because of the floods of alarms, an important new component of safety system design is alarm rationalization and alarm management (Section 1.6).

It is possible to connect conventional annunciators as front-end devices to DCS systems through various communication links. There is a wide variety of such links available, ranging from serial links employing MODBUS protocol to Ethernet links utilizing Object Linking and Embedding (OLE) for Process Control (OPC). This hybrid solution adds the visibility, reliability, and built-in redundancy of dedicated annunciators to the flexibility and record-keeping convenience of DCS-based systems.

More sophisticated annunciator designs can incorporate bargraph-type displays, color computer graphics, and event-recording or data-logging systems. Much of the new development in annunciator system designs involves enhanced methods of communication and reporting. As a consequence, annunciator status can be logged and used for tasks such as alarm management and abnormal event analysis.

Graphic displays can be dynamic, where flow in pipes is shown by actual movement, and CRT displays can concentrate large amounts of information into a single display. Figure 4.1a illustrates such a display, where the CRT displays



Area 3-H of plant layout is flashing in coded color

FIG. 4.1a
The overall safety status of the plant can be displayed on a single CRT.

the plot plan of the plant as the background. Such a plot plan can be separated into small square segments, so that if an unsafe condition is detected in a particular segment, the corresponding square can start flashing in the color that corresponds to the type of safety problem detected. This type of annunciator display is easily and quickly comprehended and can provide a summary report on a large number of safety conditions in an efficient manner.

HISTORY

The term "drop" was initially applied to individual annunciator points, from which we may infer that annunciator systems developed from paging systems of the type used in hospitals and from call systems used in business establishments to summon individuals when their services are needed. These systems consisted of solenoid-operated nameplates that dropped when deenergized. The drops were grouped at a central location and were energized by pressing an electrical pushbutton in the location requiring service. The system also included an audible signal to sound the alert.

Similar systems were used for fire and burglar alarms. The drops were operated either by manual switches or by trouble contacts that monitored thermal and security conditions in various building locations. The use of these systems in the chemical processing industry was a logical development when alarm switches became available.

This development, however, was preceded by explosion-proof, single-station annunciators that were designed to operate in the petroleum and organic chemical process plants constructed immediately before, during, and after World War II. They were usually installed on control panels located either outdoors near the process unit or in local control houses. A drop-type system could not be used in these locations because they were electrically hazardous.

By the late 1940s, centralized control rooms were introduced. Drop-type annunciators were suited for these general-purpose central control rooms. However, more compact, reliable, and flexible annunciators were subsequently introduced.

In the early 1950s, the plug-in relay annunciator was developed. Instead of utilizing solenoid-operated drops, it used electrical annunciator circuits with small telephone-type relays to operate alarm lights and to sound a horn when abnormal conditions occurred. The alarm lights installed in the front of the annunciator cabinets were either the bull's-eye type or backlighted nameplate designs.

The annunciators were compact, reliable, and because of the hermetically sealed relay logic modules, they could also be mounted in certain hazardous areas in addition to the general-purpose control rooms. In order to be mounted in Class 1 explosion-proof areas, they required purging (Figure 4.1b). Miniaturization of instruments and the use of graphic control panels initiated the development of remote annunciator systems, consisting of a remotely mounted relay cabinet connected to alarm lights installed at appropriate points in the graphic or semigraphic diagram.

Solid-state annunciator systems with semiconductor logic modules were developed in the late 1950s. These permitted additional miniaturization and lowered both the operating power requirements and the amount of heat generated. The semigraphic annunciator was introduced in the late 1960s and fully utilized the high-density capabilities of solid-state logic. It has permitted the designs of very compact and flexible semigraphic displays in control centers.

With the spread of digital communication networks and microprocessor-based smart instruments, the number of alarm points increased, their management as a function of importance (Section 1.6) became a separate field, and their displays were further miniaturized. These days, with the greater availability and reliability of integrated circuit logic components, alarms can be displayed on handheld tools in the field, on computer screens at workstations, or on CRTs in DCS systems.

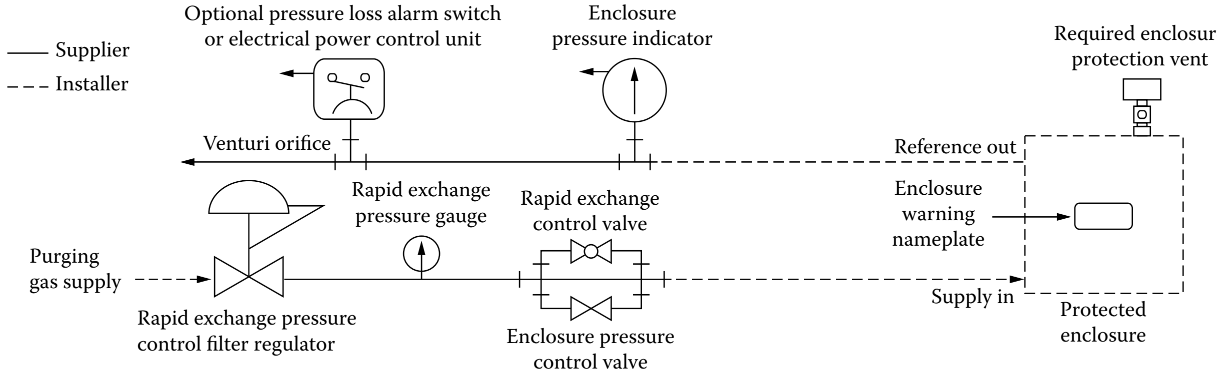


FIG. 4.1b
Local annunciators were available in explosion-proof designs or were mounted in air-purged enclosures when mounted in Class 1 areas. (Courtesy of Bebco Industries.)

PRINCIPLES OF OPERATION

The annunciator system consists of multiple alarm points. Each alarm circuit includes a trouble contact (alarm switch), a logic module, and a visual indicator (Figure 4.1c). The individual alarm points are operated from a common power supply and share a number of annunciator system components, including an audible signal generator (horn), a flasher, and acknowledge and test pushbuttons. In normal operation the annunciator system and individual alarm points are quiescent.

The trouble contact is an alarm switch that monitors a particular process variable and is actuated when the variable exceeds preset limits. In electrical annunciator systems it is normally a switch contact that closes (makes) or opens (breaks) the electrical circuit to the logic module and thereby initiates the alarm condition. In the alert state, the annunciator turns on the visual indicator of the particular alarm point, the audible signal, and the flasher for the system. The visual indicator is usually a backlighted nameplate engraved with an inscription to identify the variable and the abnormal condition, but it can also be a bull's-eye light with a nameplate. The audible signal can be a horn, a buzzer, or a bell.

The flasher is common to all individual alarm points and interrupts the circuit to the visual indicator as that point goes into the alert condition. This causes the light to continue to flash intermittently until either the abnormal condition returns to normal or is acknowledged by the operator.

The horn acknowledgment pushbutton is provided with a momentary contact: when it is operated, it changes the logic module circuit to silence the audible signal, stop the flasher, and turn the visual indicator on "steady." When the abnormal condition is corrected, the trouble contact returns to normal, and the visual indicator is automatically turned off.

The lamp test pushbutton with its momentary contact tests for burned-out lamps in the visual indicators. When activated, the pushbutton closes a common circuit (bus) to each visual indicator in the annunciator system, turning on those lamps that are not already on as result of an abnormal operating condition.

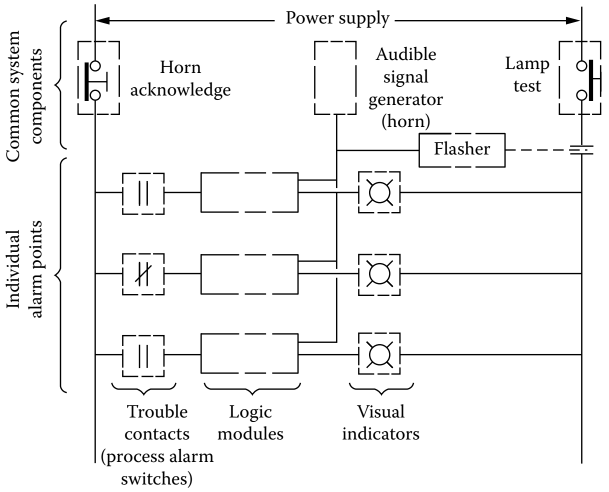


FIG. 4.1c
The main components of a traditional annunciator system.

Operating Sequences

A wide variety of sequences are available to define the operation of an individual alarm point in the normal, alert, acknowledged, and return-to-normal stages in the annunciator sequence. The five most commonly used annunciator sequences are shown in Table 4.1d, identified by the original code designation of the Instrumentation, Systems, and Automation Society (ISA). These sequences were specified by the ISA-recommended practice RP-18.1, which has since been revised and updated into standard ISA 18.1.

Because the old sequence designations are still used in some plants, some of their more common versions are listed in Table 4.1d and also described below. The sequence designations of the present standard ISA 18.1 will also be discussed below.

The Old ISA Sequence Designations ISA Sequence 1B, also referred to as flashing sequence A, is the one most frequently used. The alert condition of an alarm point results in a flashing visual indication and an audible signal. The visual indication turns off automatically when the monitored process variable returns to normal.

ISA Sequence 1D (often referred to as a dim sequence) is identical to Sequence 1B except that ordinarily the visual indicator is dim rather than off. A dimmer unit, common to the system, is required. Because all visual indicators are always turned on—for dim (normal), flashing (alert), or steady (acknowledged)—the feature for detecting lamp failure is unnecessary.

ISA Sequence 2A (commonly referred to as a ring-back sequence) differs from Sequence 1B in that following acknowledgment the return-to-normal condition produces a dim flashing and an audible signal. An additional momentary contact reset pushbutton is required for this sequence. Pushing the reset button after the monitored variable has returned to normal turns off the dim flashing light and silences the audible signal. This sequence is applied when the operator must know if normal operating conditions have been restored.

ISA Sequence 2C is like Sequence 1B except that the system must be reset manually after operation has returned to normal in order to turn off the visual indicator. This sequence is also referred to as a manual reset sequence and, like Sequence 2A, requires an additional momentary contact reset pushbutton. Sequence 2C is used when it is desirable to keep the visual indicator on (after the horn has been silenced by the acknowledgment pushbutton) even though the trouble contact has returned to normal.

ISA Sequence 4A, also known as the first-out sequence, is designed to identify the first of a number of interrelated

TABLE 4.1d
The Old ISA Designations of Annunciator Sequences

ISA Code for the Sequence	*Annunciator Condition*	*Process Variable Condition (Trouble Contact)*	*Visual Indicator*	*Audible Signal*	*Use Frequency*
IB	Normal	Normal	Off	Off	55%
	Alert	Abnormal	Flashing	On	
	Acknowledged	Abnormal	On	Off	
	Normal again	Normal	Off	Off	
	Test	Normal	On	Off	
ID	Normal	Normal	Dim	Off	1%
	Alert	Abnormal	Flashing	On	
	Acknowledged	Abnormal	On	Off	
	Normal again	Normal	Dim	Off	
2A	Normal	Normal	Off	Off	4%
	Alert	Abnormal	Flashing	On	
	Acknowledged	Abnormal	On	Off	
	Return to normal	Normal	Dim flashing	On	
	Reset	Normal	Off	Off	
	Test	Normal	On	Off	
2C	Normal	Normal	Off	Off	5%
	Alert	Abnormal	Flashing	On	
	Acknowledged	Abnormal	On	Off	
	Return to normal	Normal	On	Off	
	Reset	Normal	Off	Off	
	Test	Normal	On	Off	
4A	Normal	Normal	Off	Off	28%
	Alert	Abnormal			
	Initial		Flashing	On	
	Subsequent		On	Off	
	Acknowledged	Abnormal			
	Initial		On	Off	
	Subsequent		On	Off	
	Normal again	Normal	Off	Off	
	Test	Normal	On	Off	
All others					7%

variables that have exceeded normal operating limits. An off-normal condition in any one of a group of process variables will cause some or all of the remaining conditions in the group to become abnormal. The first alarm causes flashing, and all subsequent points in the group turn on the steady light only. This sequence monitors interrelated variables. The visual indication is turned off automatically when conditions return to normal after acknowledgment.

The New ISA Sequence Designations In the updated annunciator standard ISA 18.1, the sequence designations are different, as shown in Table 4.1e. The most widely used sequence, the basic flashing sequence, is now designated as sequence A. The sequence designations in ISA 18.1 use the following letter codes:

A = Automatic Reset
M = Manual Reset
R = Ringback
F = First-out

Therefore, using the ISA 18.1 sequence designations A-13 means that the annunciator has automatic reset and is provided with Option 13, which suggests the presence of a dim lamp monitor. For definitions of less frequently used sequences, refer to ISA 18.1.

TABLE 4.1e

The New ISA Designations of Annunciator Sequences as Defined by ISA Standard ISA 18.1

Sequence	*Signal Device*	*Normal*	*Alert*	*Condition-sensing Returns to Normal Before Acknowledge*	*Acknowledge*	*Condition-sensing Returns to Normal*	*Return to Normal Reset*	*Remarks*
A	Visual	Off	Flash	Flash	On	Off	—	Flasher memory
	Audible	Off	On	On	Off	Off	—	
A-5	Visual	Off	On	On	On	Off	—	Memory
	Audible	Off	On	On	Off	Off	—	
A-4	Visual	Off	Flash	Off	On	Off	—	Flasher
	Audible	Off	On	Off	Off	Off	—	
A-4-5	Visual	Off	On	Off	On	Off	—	
	Audible	Off	On	Off	Off	Off	—	
A-13	Visual	Dim	Flash	Flash	On	Dim	—	Memory—flasher
	Audible	Off	On	On	Off	Off	—	Continuous lamp test

Optional Operating Features Annunciator sequences may be initiated by alarm switch trouble contacts that are either open or closed during normal operations. These are referred to as normally open (NO) and normally closed (NC) sequences, respectively, and the ability to use the same logic module for either type of trouble contact is called an *NO-NC option.* It is important because some alarm switches are available with either an NO or an NC contact but not with both, and therefore without the NO-NC option in the logic module two types of logic modules would be required. The logic module is converted for use with either form of contact by a switch or wire jumper connection.

The relationship between the NO and NC sequences required in the logic module to match the various trouble contacts and analog measurement signal actions is shown in Figure 4.1f. A high alarm in a normally closed annunciator system requires a normally closed trouble contact operated by a direct-acting analog input.

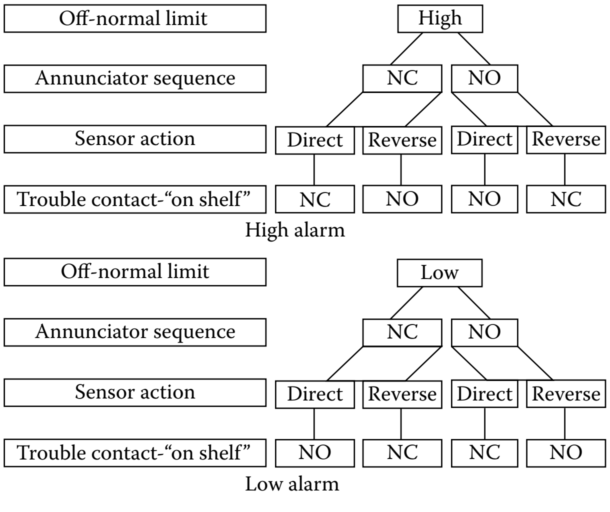


FIG. 4.1f

Logic trees for the NO and NC annunciator sequences.

If an increase in the measured variable results in an increased output signal, the detector is direct acting; if the output signal is reduced, it is a reverse-acting sensor. If the trouble contacts in all alarm switches in the plant are standardized such that normal operating conditions will cause all trouble contacts to be NC (or NO), the required annunciator sequence is also NC (or NO), and Figure 3.1f need not be consulted.

Annunciator systems are fail safe or self-policing if they initiate an alarm when the logic module fails because of relay coil burnout. The feature is standard for most NO and NC annunciator sequences; annunciators using NC trouble contacts are also fail safe against failures in the trouble contact circuit.

The lock-in option locks in the alert condition initiated by a momentary alarm until the horn acknowledgment button is pushed, preventing loss of a transient alarm condition until the operator can identify it. The logic module is usually changed from lock-in to non-lock-in operation by either addition of a wire jumper or operation of a switch. The lock-in feature is useful for monitoring unstable or fluctuating process variables.

Test and Repeater Features The test feature in a standard annunciator serves only to test for burned-out lamps in the visual indicators. The operational test feature provides a test of the complete annunciator system, including logic modules, lamps, flasher, audible signal, and acknowledgment circuits. The operational test circuit usually requires an additional momentary contact pushbutton, which can replace the regular lamp test pushbutton.

The logic module of relay-type annunciators may have spare (electrically isolated) auxiliary contacts that can operate shutdown and interlock systems when alarm conditions occur. The auxiliary contacts are wired to terminal blocks in the annunciator cabinet for connection to external circuitry.

Repeater lights may be located away from the common logic module and serve to alert operators in other areas. Annunciator cabinet terminals for connecting these repeater lights in parallel with the annunciator visual indicator are available.

It may also be desirable to actuate a horn in more than one location. The electrical load of multiple audible signals requires an interposing relay, called a horn-isolating relay, operated by the logic modules. This relay has contacts of adequate capacity to operate multiple audible signals. Horn-isolating relays may be installed either in the annunciator cabinet or in a separate assembly.

Annunciator systems can be used for several operational sequences without changing system wiring, and many logic modules can supply more than one operational sequence. This multiple sequence capability is sometimes useful when the sequence has not yet been determined.

ANNUNCIATOR TYPES

The audiovisual annunciator can be packaged as an integral, remote, or semigraphic annunciator.

Integral Annunciator

The integral annunciator, a cabinet containing a group of individual annunciator points wired to terminal blocks for connection to external trouble contacts, power supply, horn and acknowledge and test pushbuttons, is the most economical of the various packaging methods available in terms of cost per point. It is also the simplest and cheapest to install.

Two methods of packaging integral annunciators are illustrated in Figure 4.1g. In the nonmodular type, plug-in logic modules are installed inside the cabinet and connected to alarm windows on the cabinet door through an interconnecting wiring harness; in the modular type, individual plug-in alarm point assemblies of logic module and visual indicator are grouped together.

The nonmodular and modular cabinet styles are both designed for flush panel mounting with the logic modules and visual indicators accessible from the front. Electrical terminals for the external circuitry are located in the rear of the cabinet and are accessible from the back.

Integral annunciators are used on nongraphic and on semigraphic control panels in which physical association of the visual indicators with a specific location in the graphic

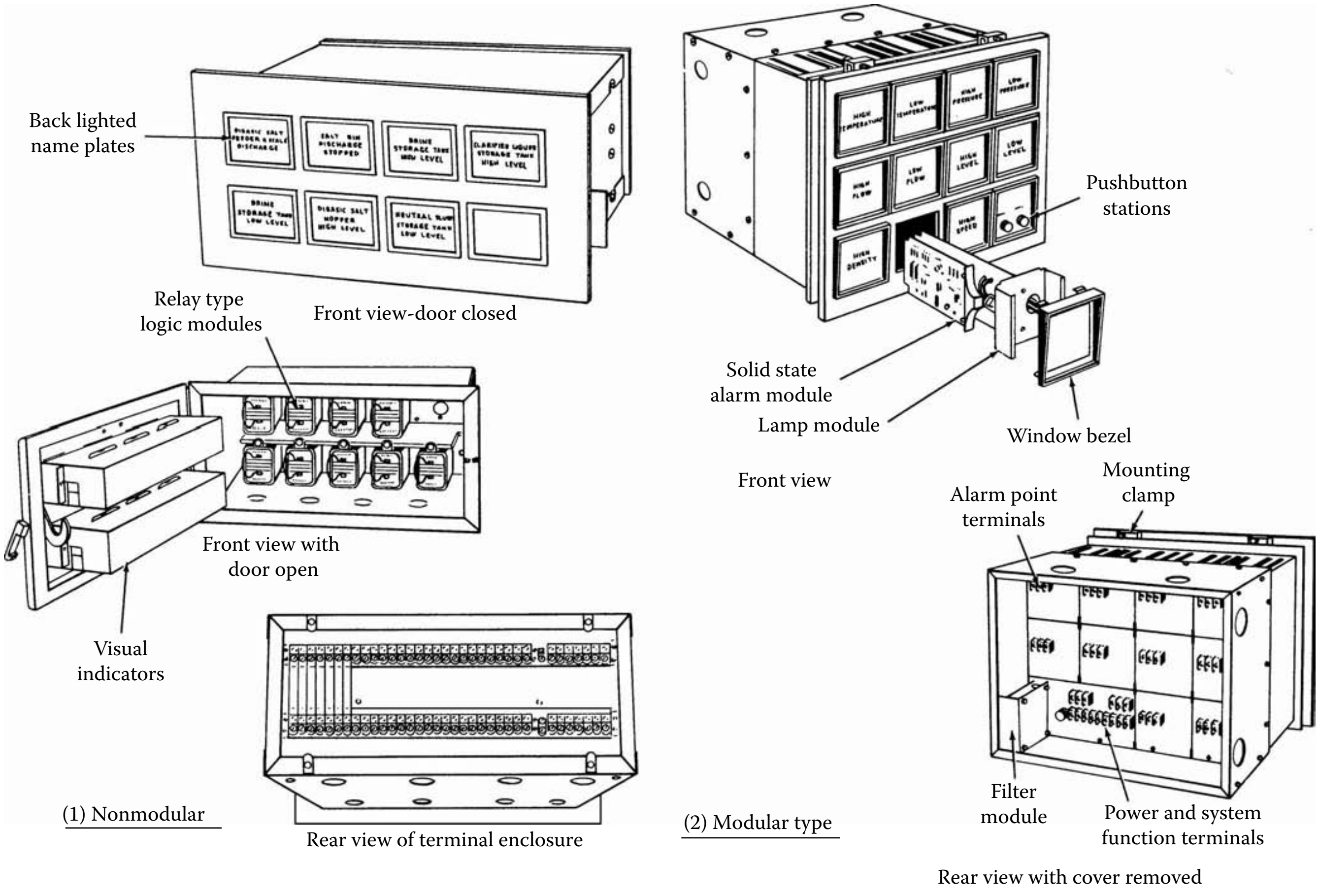


FIG. 4.1g
Integral annunciator cabinets are available in modular (right) and nonmodular (left) construction.

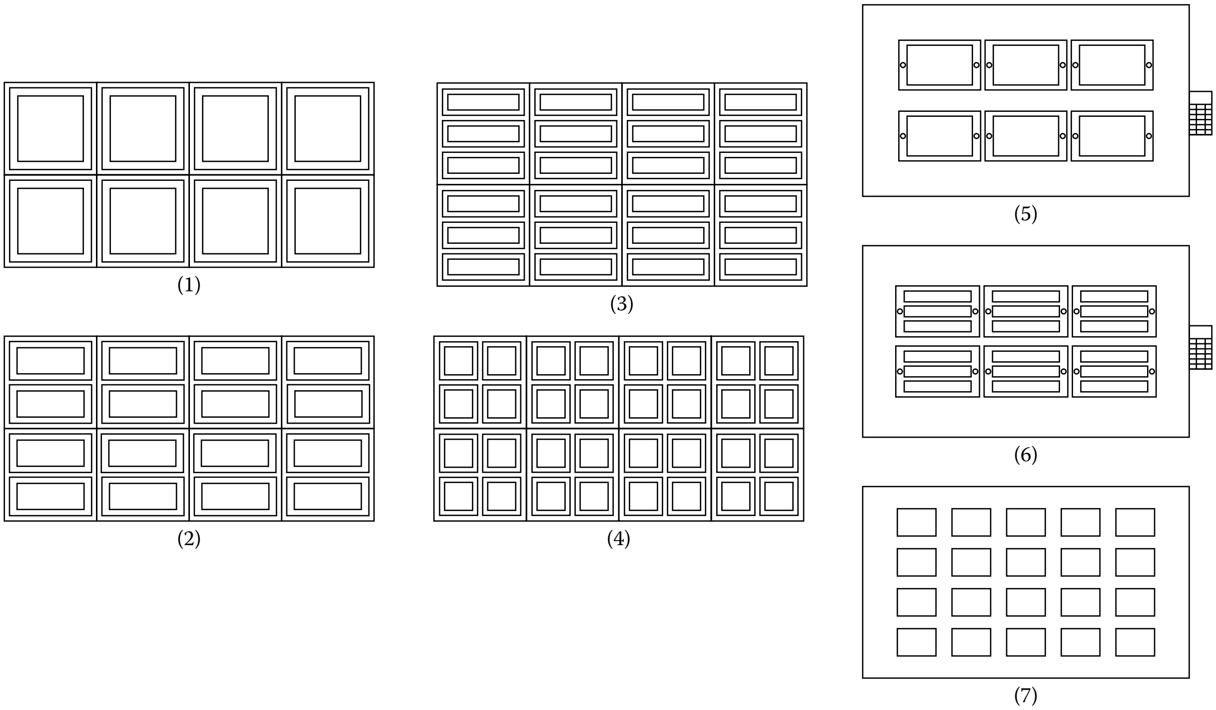


FIG. 4.1h
Integral annunciator window configurations. 1. Modular single-point annunciator. 2. Modular double-point annunciator. 3. Modular triple-point annunciator. 4. Modular quadruple-point annunciator. 5. Nonmodular single-point. 6. Nonmodular triple-point. 7. Nonmodular single-point with small nameplate.

process flow diagram is not required. Integral annunciator cabinets occupy more front but less rear panel space than the equivalent remote designs. The electrical terminals are in a general-purpose enclosure at the rear of the cabinet, and trouble contacts can be wired directly to them, thus eliminating the need for and resultant costs of intermediate terminal blocks for trouble contact wiring.

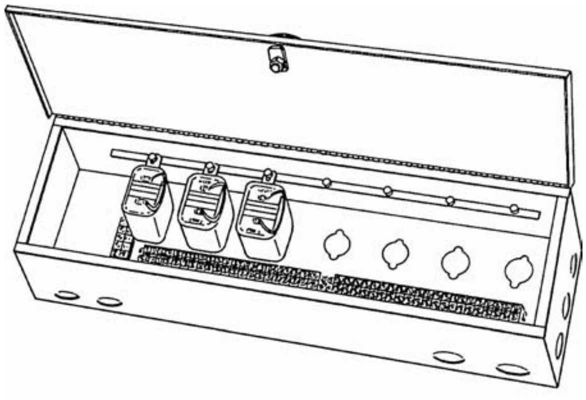
An advantage of the modular-type cabinet is that it can be expanded by enlargement of the panel cutout and by the addition of modular alarm point assemblies. Nonmodular cabinets cannot be expanded, and new cabinets must be installed to house additional alarm points. Consequently, one should include more spare points when specifying the cabinet size for a nonmodular system. The modular cabinet is also more compact, takes up less panel space, and has a greater visual display area per point than the nonmodular design.

Figure 4.1h illustrates various configurations of visual indicators that can be supplied with integral annunciator cabinets. Many of these groupings are also available in single-unit assemblies for remote annunciator systems.

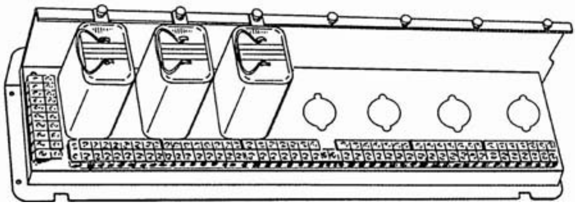
Remote Annunciator

The remote annunciator differs from the integral annunciator in that the visual indicators are remote from the cabinet or chassis containing the logic modules. Remote annunciators were developed to allow the visual indicators to be placed in their actual process location in the graphic flow diagram.

They are used with full and semigraphic control panels and in nongraphic applications in which an integral annunciator cabinet may require too much front panel space. Figure 4.1i



(1) Chassis with general purpose enclosure



(2) Chassis

FIG. 4.1i
General purpose, remote annunciator cabinets.

illustrates a remote annunciator chassis with optional cabinet enclosure. The chassis contains spare positions for plug-in logic modules and a system flasher. Auxiliary system modules, such as horn-isolating relays, may also be plugged into the logic module chassis positions.

The chassis and cabinet enclosure are designed for wall or surface mounting behind the control panel. Each chassis position has terminal points for connecting the visual indicator and trouble contact. In addition, the chassis has a system terminal block for connecting electrical power, horn, flasher, and acknowledge and test pushbuttons.

The disadvantages of remote annunciators include higher equipment and installation costs and an increased requirement for back panel space. In addition, the wiring connections from field trouble contacts must be made to intermediate terminal blocks rather than directly to the cabinet terminals, as with the integral annunciator.

These terminal blocks, the terminal enclosure, and the required wiring result in higher installation costs and extra space requirements. Finally, the remote annunciator is difficult to change, and modification costs of remote systems are substantially higher than those of the integral type, partially because spare visual indicators cannot be installed initially.

Semigraphic Annunciator

The semigraphic annunciator developed in the late 1960s combines some of the advantages of the integral annunciator with the flexibility to locate visual indicators at appropriate points in a graphic flow diagram. Figure 4.1j illustrates a semigraphic annunciator.

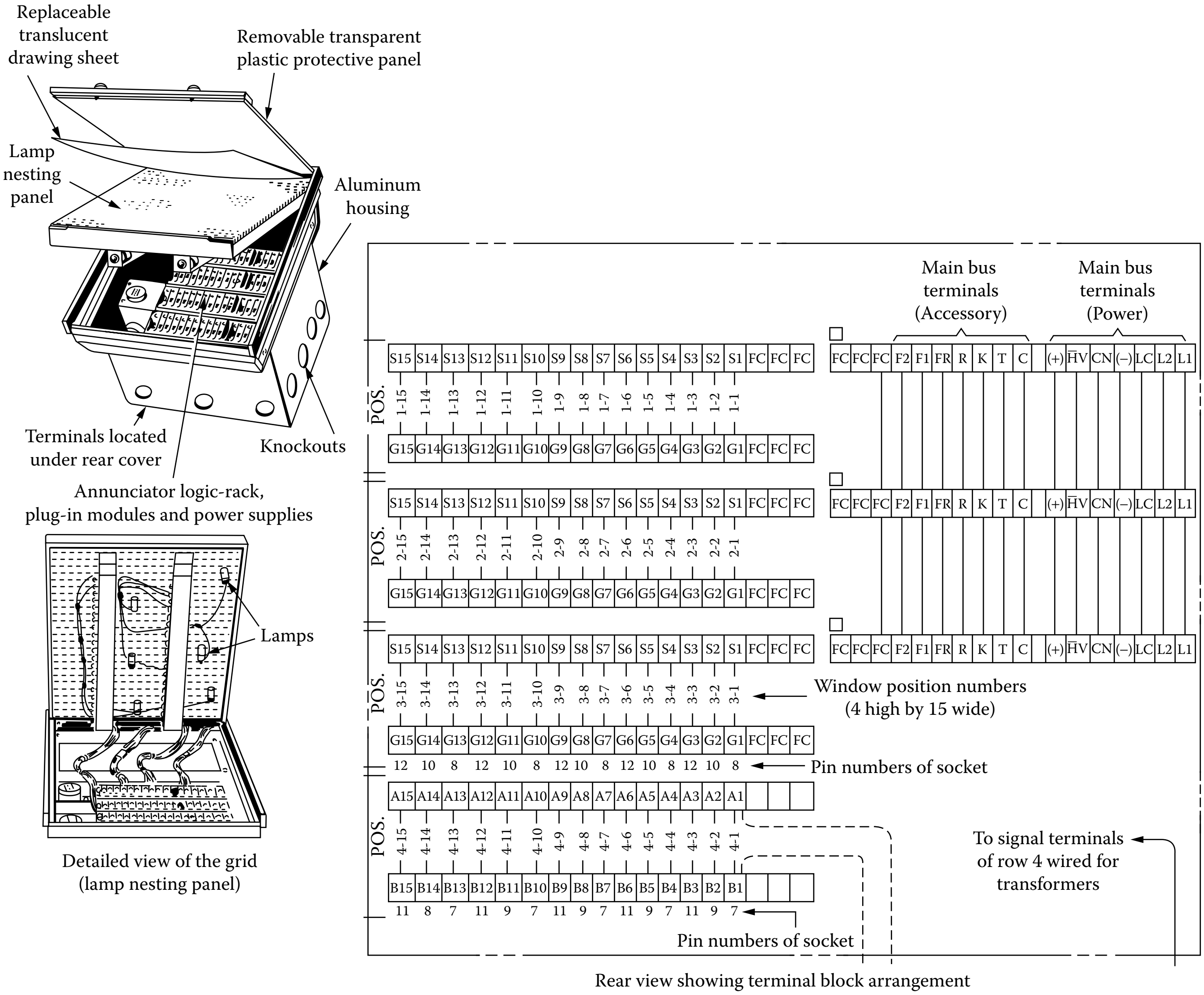


FIG. 4.1j
Construction of a typical semigraphic annunciator, showing the plug-in modules,the lamp nesting panel, and the arrangement of the terminal blocks on the rear of the unit.

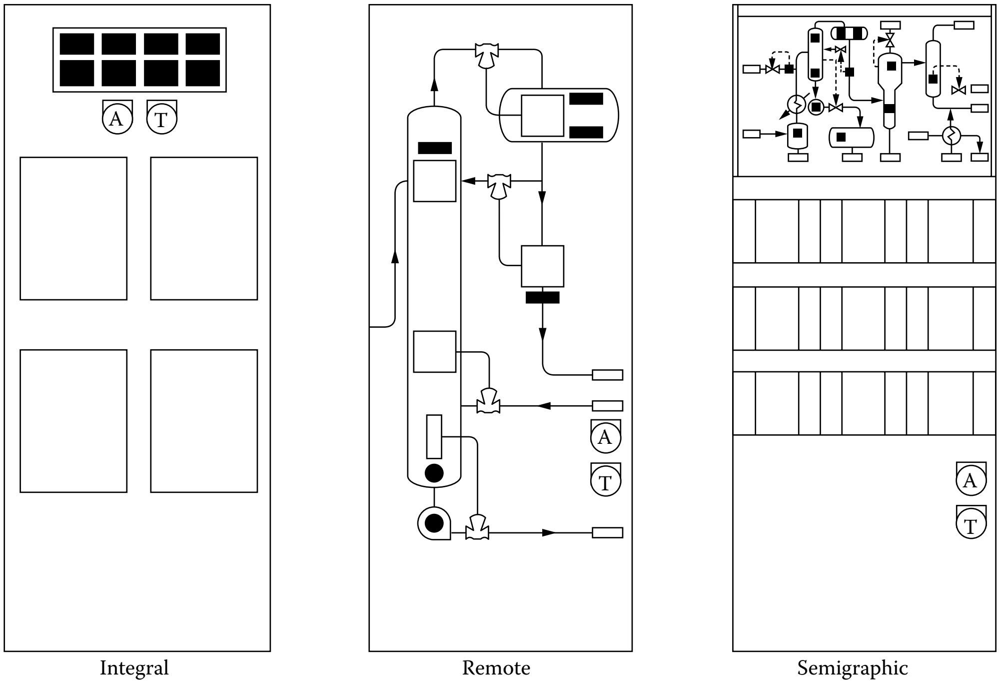


FIG. 4.1k
Control panels with integral, remote, and semigraphic annunciators.

It consists of a cabinet containing annunciator logic modules wired to visual indicators placed in a 3/4-inch (18.75-mm) lamp insertion matrix grid forming the cabinet front. The semigraphic display is placed between the lamp grid and a transparent protective cover plate, and the visual indicators are positioned to backlight alarm name plates located in the graphic display.

The protective cover and lamp grid are either hinged or removable to provide access to the logic module and lamp assemblies. The lamp assemblies are connected to intermediate terminals located behind the lamp grid, and the terminals in turn are connected to the logic modules. Terminal points for trouble contact wiring are in the back of the cabinet.

The semigraphic annunciator is flexible, and changes in the annunciator system, graphic display, and related panel modifications can be made easily and cheaply. It is practical to prepare the graphic displays in the drafting room or model shop, thus protecting proprietary process information of a confidential nature.

The graphic display has little or no effect on completing either the annunciator or the control panel because it can be installed on site or at any time. The semigraphic annunciator has a high density of 40 alarm points per linear foot (0.3 m) and a solid-state rather than relay-type logic design. Power supplies are self-contained in the semigraphic annunciator cabinet.

Front panel layouts illustrating integral, remote, and semigraphic annunciators are shown in Figure 4.1k. Integral systems similar to the one shown at the left in the figure are normally specified on nongraphic control panels. The graphic panel in the center contains a remote annunciator with backlighted nameplates (shaded rectangles) and pilot lights (shaded circles) for visual indication. The remote system may also be used with miniature lamps in a semigraphic display similar to the one shown at the right.

Recording Annunciators

Solid-state, high-speed, recording annunciators are available to amend or substitute a printed record of abnormal events for only visual and/or audible alarms. These systems print out a record of the events and identify the variable, the time at which the alarm occurs, and the time at which the system returns to normal. They can also discriminate among a number of almost simultaneous events and print them out in the time sequence in which they occurred. A number of optional features, including secondary printers at remote locations, supplementary visual indication, and computer interfacing, are also available. The typical unit consists of logic, control, and printer sections.

The input status is continuously scanned. If a change in the trouble contacts has occurred since the preceding scan cycle, the central control places the exact time, the alarm point identification, and the new status of the trouble contact (normal or abnormal) into the memory and initiates the operation of the output control unit (Figure 4.1l).

The output control unit accepts the stored information and transfers it to the printer, which logs the event. Following this, the memory is automatically cleared of the data and is ready to accept new information. In addition to or in place of the printer (if a permanent record is not required), a CRT display can serve as the event readout. Trouble contacts are

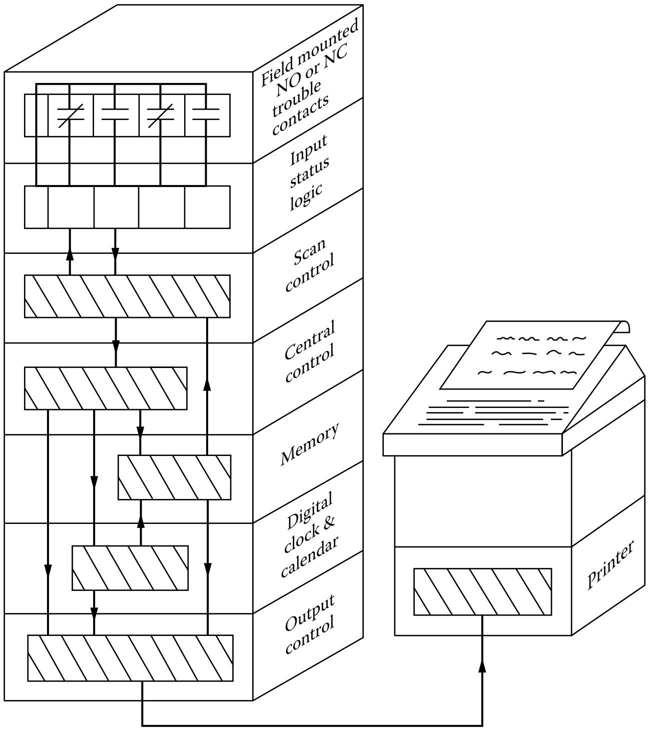


FIG. 4.1l
Functional block diagram of a digital, recording annunciator with its printer, which can be replaced by a CRT.

connected to terminal points in the logic cabinet, and a cable connects the cabinet and the printer.

A recording annunciator can perform more sophisticated monitoring than an audiovisual annunciator and is correspondingly more expensive on a per-point basis. System cost per point decreases as the system size increases. Higher equipment cost, however, is offset in part by savings in control panel space and in installation costs. Recording annunciators are frequently used by the electrical power generating industry but may be applied to advantage in any industrial process that must monitor large numbers of operating variables and analyze abnormal events efficiently.

Vocal Annunciators

Vocal annunciators are unique in the type of abnormal audible message they produce. The audible output is a verbal message identifying and describing the abnormal condition when it occurs and repeating the message until the operator acknowledges the difficulty. The system continuously scans the trouble contacts, and when an abnormality is found, it turns on a flashing visual indicator and selects the optional proper verbal message for broadcast.

The visual indicator is turned off by the system when the point returns to normal. The control unit also arranges the messages to be broadcast in the order in which the difficulties occur. In the event of multiple alarms, the second message is played only after the first has been acknowledged. The flashing visual indicator for each point, however, turns on when the point becomes abnormal. The verbal message may be broadcast simultaneously in the control room and related operating areas, thus permitting personnel at the operating unit to correct the problem immediately.

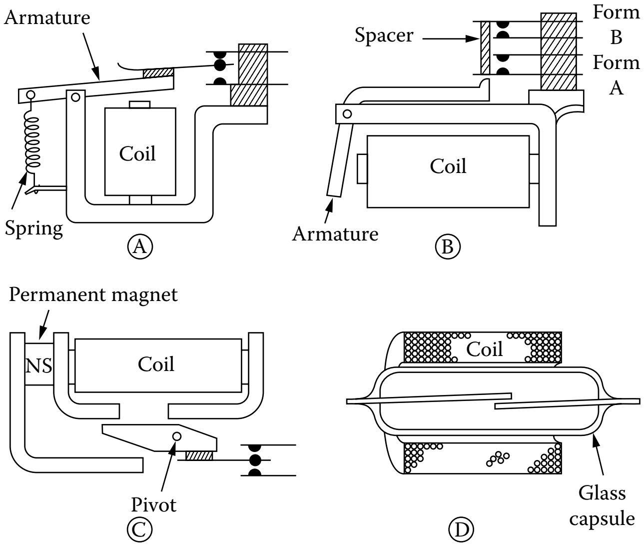


FIG. 4.1m
Standard electro-mechanical relay structures include: A) the clapper-type, B) the phone-type, C) the balanced-force, and D) the Reed relay.

Relay-Type Annunciators

The basic element of this annunciator is an electrical relay wired to provide the logic functions required to operate a particular sequence. Figure 4.1m illustrates some of the basic relay designs.

At least two relays per logic module are necessary for most sequences. The relays are installed and wired in a plug-in assembly, which is the logic module for a single alarm point. The plug-in module assembly is usually hermetically sealed in an inert atmosphere to prolong the life of the relay contact. The sealing also makes the logic module acceptable in certain hazardous electrical areas.

Figure 4.1n is a semischematic electrical circuit for a remote system with sequence operation according to old ISA Sequence 1B (new Sequence A): Two logic modules are shown; one is in normally closed operation and the other is in normally open operation. The remote visual indicators for each alarm point, the horn, the flasher, and the acknowledge and test pushbuttons common to the system are also shown. Each logic module has two relays, A and B, shown in their deenergized state according to normal electrical convention. The operation of these circuits during the various stages of sequence operation is as follows:

Normal The trouble contact of the NC alarm point is wired in series with the A relay coil at point terminals H and NC. In the normal condition, the trouble contact is closed, relays

Operational sequence

Condition	Trouble contact	Signal lamps	Audible signal	"A" relay	"B" relay
Normal	Normal	Off	Off	Energized	Energized
Alert	Abnormal	Flashing	On	Deenergized	Energized
Audible silenced (acknowledged)	Abnormal	Steady–on	Off	Deenergized	Deenergized
Normal again	Normal	Off	Off	Energized	Energized
Lamp test	Normal	Steady–on	Off	Energized	Energized

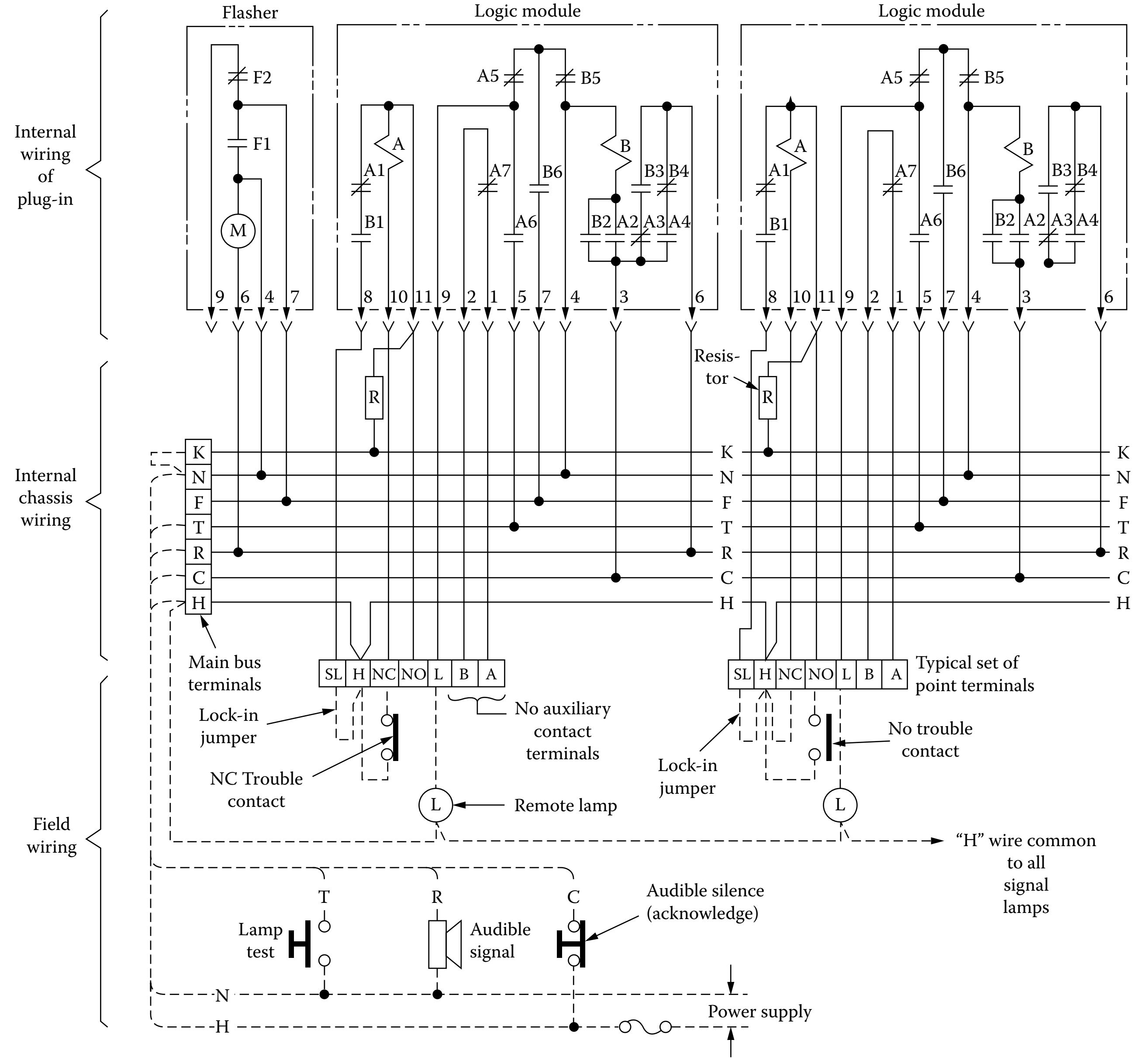


FIG. 4.1n
Semischematic diagram of a relay-type annunciator designed for ISA new sequence A (old sequence 1B), which sequences are described in Tables 4.1d and 4.1e.

A and B are energized, and all A and B relay contacts are in the state opposite that shown.

Relay A is energized from power source H through the closed trouble contract, relay coil A, resistor R, terminal K, and jumper to the neutral side of the line N. Relay B is also energized from H through the normally closed acknowledge pushbutton to terminal C, closed contact A2, and relay coil B, to N. Relay B is locked in by its own contact B2, which closes when relay B is energized. The visual indicator is turned off by open contact A5. The audible signal and the flasher motor are turned off by open contacts A3 and B4 in the same circuit.

Alert The trouble contact opens, deenergizing relay A and returning all *A* relay contacts to the state shown in Figure 4.1n. The visual indicator is turned on, flashing through circuit H, lamp filament, terminal *L*, closed contacts A5 and B6, bus F, and flasher contact F1, to N. The flasher motor is driven through circuit C, closed contacts A3 and B3, R bus, and the flasher motor to N, and the audible signal is turned on by the same circuit.

Acknowledged Relay B is deenergized by operating (opening) the momentary contact horn acknowledgment pushbutton and is locked out by open contact A2. All B relay contacts are returned to the state shown in Figure 4.1n. The visual indicator is turned on steady through closed contact *B*5 to *N* and is disconnected from flasher contact F1 by open contact B6. The flasher motor and audible signal are turned off by the horn acknowledgment pushbutton and remain off as a result of open contact B3.

Normal Again When the variable condition returns to normal, the trouble contact closes to energize relay A, the visual indicator is turned off by open contact A5, relay B is energized by closed contact A2, and all circuits are again in the state described under normal.

Lamp Test The lamp test circuit operates the visual indicators of only those alarm points that are in the normal condition. The circuit is completed through power source H, lamp filament, terminal L, closed contact A6, bus T, and normally open momentary contact lamp test pushbutton, to N. Closing the lamp test pushbutton to N completes the circuit and lights the visual indicators. Alarm points that are in the off-normal condition (either alert or acknowledged) do not operate because their A relays are deenergized and the A6 contact is open. The visual indicators of these abnormal alarm points are already turned on (either flashing or steady) through the operation of the alarm sequence.

Lock-In The lock-in feature operates to prevent an annunciator alert condition (caused by a momentary alarm) from returning to normal until the horn acknowledgment button is pushed. Point terminals H and SL are jumpered to provide the lock-in feature (see Figure 4.1n). When the trouble contact opens, the A relay is deenergized, and power source H is applied to the N side of the relay through closed contacts A1 and B1.

The power is dissipated through resistor R, terminal K, and jumper to N. If the trouble contact returns to normal, relay A will remain deenergized because potential H is on both sides of the coil. If the acknowledgment button is pushed before the trouble contact closes again, relay B will be deenergized, opening contact B1 and the lock-in circuit, thus permitting the system to return automatically to normal when the trouble contact closes.

If the acknowledgment button is pushed after the trouble contact has reclosed, contact B1 opens momentarily, allowing the A relay to reenergize. Contact A1 opens, and the circuits are reestablished in their normal operating state.

Operational Test Full operational test is incorporated in the annunciator sequence shown by replacement of the jumper connection between main bus terminals K and N with a normally closed momentary contact pushbutton, which when pushed opens all annunciator circuits, thus initiating the alert condition of all alarm points in the normal condition.

Auxiliary Contacts Normally closed contact A7 connected to point terminals A and B is available for auxiliary control functions.

Relay Fail-Safe Feature Two parallel circuits, one consisting of closed contact A3 and open contact B3 and the other of open contact A4 and closed contact B4, operate an alert signal when there is a failure of either the A or the B relay coil. A failure of the former initiates a normal alert in the same way as the trouble contact. A failure of the B relay turns on the audible signal through closed contacts A4 and B4.

Normally Open Trouble Contacts The annunciator sequence and features described for NC trouble contacts operate in essentially the same way when NO contacts are used. In the NO system, however, the trouble contact is wired in parallel with the A relay coil to point terminals H and NO, and a wire jumper is installed between point terminals H and NC. Normally, the trouble contact is open and the A relay is energized from terminals H and jumper to terminal NC. In the alert condition, the trouble contact closes to deenergize the A relay by applying power source H to the N side of the relay.

Electromechanical relays are available for use with a variety of AC and DC voltages, but 120 AC, 50 to 60 Hz, and 125 DC are the most popular. Power consumption of the logic modules is normally less than 10 voltamperes (AC) and 10 watts (DC). Special low-drain and no-drain logic modules are available; these consume no power during normal operation. Visual indicators consume different amounts of power, depending on the type. Small bull's-eye lights and backlighted nameplates use approximately 3 watts, whereas large units require 6 to 12 watts, depending on whether one or two lamps are used.

Electromechanical annunciator systems are reliable and may be used at normal atmospheric pressures and ambient temperatures in the 0 to 110°F (–17.8 to 43.3°C) range. They are not position sensitive. They will generate a substantial amount of heat during plant shutdown when a large number of points are askew, and therefore power should be disconnected during these periods. The principal disadvantages of the relay-type annunciator are size, power consumption, and heat generation.

Solid-State Annunciators

A solid-state logic module consists of transistors, diodes, resistors, and capacitors soldered to the copper conductor network of a printed circuit board supplying the required annunciator logic functions. The modules terminate in a plug-in printed circuit connector for insertion into an annunciator chassis; they may also contain mechanical switching or patching devices to provide lock-in and NO-NC options.

Figure 4.1o is a semischematic electrical circuit for a remote system with ISA Sequence 1B. The logic module shown is in normally closed operation. Remote lamps for two points and a flasher-audible module, speaker, and acknowledge and test pushbuttons common to a system are also included. Switch S1 is the NO-NC option switch and is shown in the NC operating position. Switch S2 is the lock-in option switch and is shown in the lock-in position. The following description uses negative logic, i.e., a high equals a negative voltage, whereas a low is approximately 0 volts.

Normal The trouble contact of the NC alarm point is connected to an input filter circuit consisting of resistors R13 and R50 and capacitor C1. This provides transient signal suppression as well as voltage dropping. The slide switch S1 connects the trouble contact and filter network to resistor-R14. In this state transistor T1 is conducting, causing the full negative voltage to be dropped across resistor R17, resulting in a low voltage at the bottom end of resistor R20. Transistors T2 and T3 are the active elements of the input memory and are roughly equivalent to the A relay of an electromechanical module (see Figure 4.1n).

The base of T2 has four inputs, including resistor R20, either directly from the trouble contact in NO operation or from the collector of T1 in NC operation; resistor R19 with a locking signal from the alarm memory transistor T5; resistor R28 and capacitor C2, which form a regenerative feedback from T3; and resistor R15 from the test circuit. The base of T3 has one input, resistor R29 from the collector of T2. In normal operation, all four inputs to the base of T2 are low, T2 is not conducting, and T3 is conducting. Conversely, when a high signal is present at any one of the four inputs to the base of T2, T2 conducts and T3 turns off.

Transistors T4 and T5 are the active elements of the alarm memory and are approximately equivalent to the B relay of an electromechanical module. T4 and T5 together with bias resistors R30 and R33 and cross-coupling resistors R31 and R32 form a bi-stable (flip-flop). In normal operation T4 is off and T5 conducts. The upper end of capacitor C4 is connected to the collector T2. When T2 is off, its collector is at a high and capacitor C4 will change from top to bottom, minus to plus. Transistor T7 is a high-capacity lamp amplifier and T6 is its preamplifier. In normal operation the base of T6 is high and T6 is on, T7 is off, and the visual indicator is off.

Before completion of the description of the normal condition, the operation of the flasher-audible module in Figure 4.1o will be described. The module has two oscillators. The first is a 3-Hz unit generating a signal that is amplified and supplied to the logic modules through the Fl bus. The second is a 700-Hz oscillator generating a signal that is amplified and supplied to the audible signal through the R bus.

The audible signal is a permanent magnet-type transducer (speaker) that converts the electrical energy into sound. Initiated by an audio oscillator, the active elements of which are transistors T1 and T2, these transistors together with passive components (capacitors C1 and C2 and diodes D3 and D4) form an unstable multivibrator when an input is present on the FR bus. In the normal conditions there is no FR signal, the voltage necessary to turn on transistor T1 is missing, and the oscillator will not operate. This is the normal, or quiescent, condition.

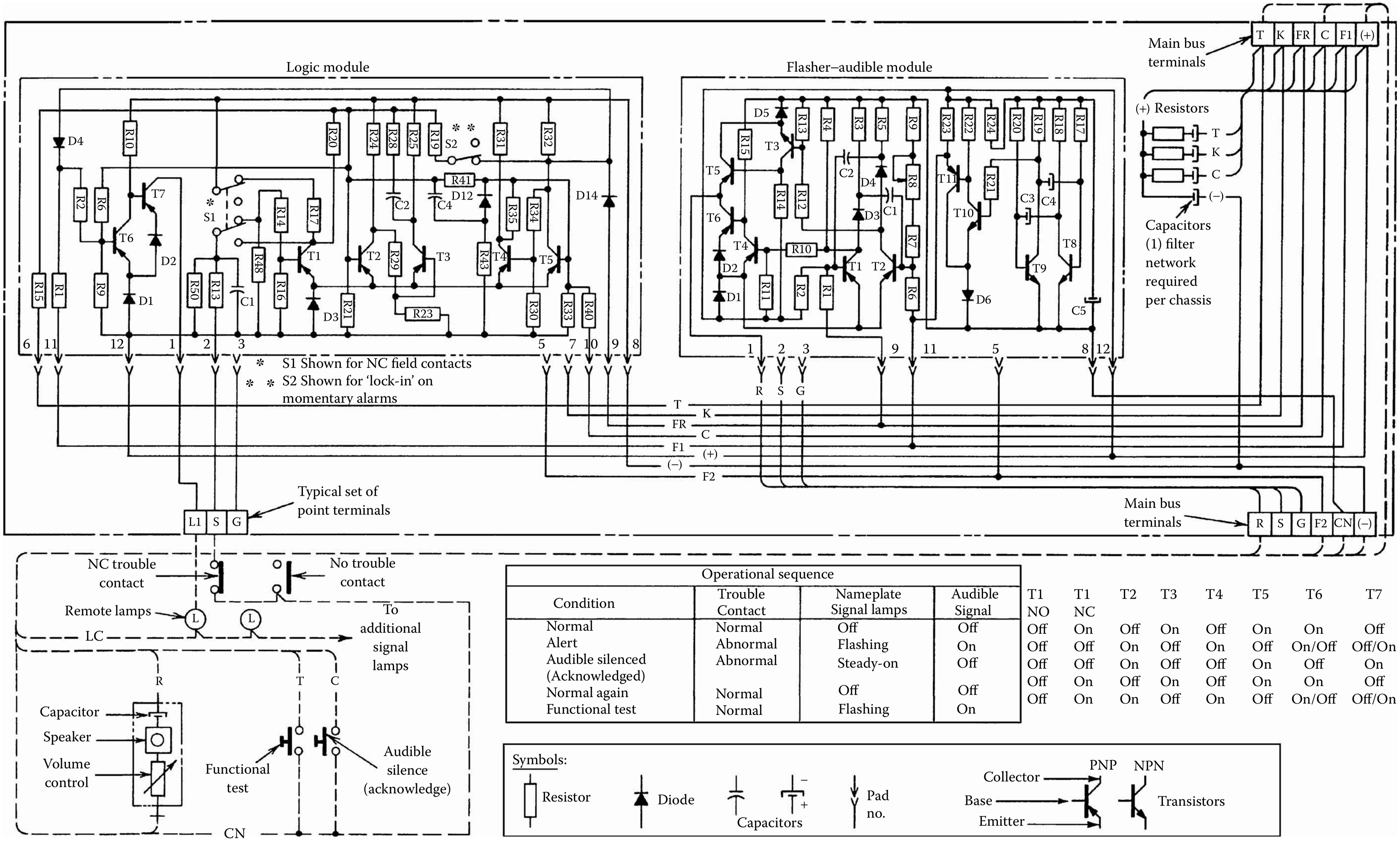
Alert The trouble contact opens and the base of T1 becomes low, turning off T1. This action produces a high on the base of T2 through R20, which turns on T2 and turns off T3. The negative end of C4 is clamped to common through T2, causing a positive pulse at the base of T5 through diode D12, turning T5 off and T4 on. With T2 conducting, the base of T6 becomes low through resistor R6, and with T5 off the clamp on the flasher signal is removed through diode D4 at the junction of R1 and R2. The flasher source provides an alternating high and low voltage at the Fl bus, which turns T6 on and off, which in turn turns T7 and the light off and on.

The flasher signal is generated by transistors T8 and T9, which are the active elements of an unstable multivibrator used as an on/off signal to the output driver stage. Resistor R24 and capacitor C6 decouple the oscillator from the power lines so that its frequency is not affected by that of the other oscillators. The output driver stage consists of transistors T10 and T11, a switching inverter, and an emitter follower stage, which produces an alternating high and low voltage of F1 bus through R23. Transistor T11 is a high-current transistor capable of driving a multiple lamp load.

The audible signal is initiated by a high on the FR bus, which turns on an audio oscillator, the elements of which are transistors T1 and T2. The audio oscillator output is amplified in an audio amplifier stage composed of transistors T3, T4, T5, and T6 connected in two pairs—one T3 and T5; the other T4 and T6.

The input components to the stage from the audio oscillator are opposite each other, i.e., whenever one is high (negative voltage) the other is low (near zero), causing only one pair of transistors to conduct at a time. When T5 is off, T6 is on and the capacitor is discharged. This alternating action causes an alternating current to flow in the speaker coil, giving an audible signal.

Acknowledged A negative voltage is applied to the base of T5 through resistor R40 by closing the acknowledge pushbutton. This turns T5 on and T4 off, and the FR bus becomes



Operational sequence			
Condition	Trouble Contact	Nameplate Signal lamps	Audible Signal
Normal	Normal	Off	Off
Alert	Abnormal	Flashing	On
Audible silenced (Acknowledged)	Abnormal	Steady-on	Off
Normal again	Normal	Off	Off
Functional test	Normal	Flashing	On

T1 NO	T1 NC	T2	T3	T4	T5	T6	T7
Off	On	Off	On	Off	On	On	Off
Off	Off	On	Off	On	Off	On/Off	Off/On
Off	Off	On	Off	Off	On	Off	On
Off	On	Off	On	Off	On	On	Off
Off	On	On	Off	On	Off	On/Off	Off/On

FIG. 4.1o
Semischematic diagram of a solid-state annunciator designed for ISA new sequence A (old sequence 1B), which sequences are described in Tables 4.1d and 4.1e.

low, thus silencing the audible signal. When the point is acknowledged, T5 conducts; this restores the clamp at R1 and R2, which removes the flash source voltage. T6 is off all the time, T7 is on all the time, and the light is on steady.

Normal Again When the variable condition returns to normal, the trouble contact closes and the base of T1 becomes high, turning T1 on. This produces a low on the base of T2, which turns T2 off and T3 on. All circuits are again in the state described under the normal condition.

Lamp Test No separate lamp test is normally provided. One initiates a full operational test by pushing the test button, which applies an alternate input signal through resistor R15 to the base of transistor T2. This turns T2 on, initiating a full operational test of the system as already described.

Lock-In The lock-in feature is provided by a switch S2. If the switch is in the lock-in position (see Figure 4.1o), when T5 turns off (on the alarm condition), a high at the collector of T5 is coupled to the base of T2 through R19, which keeps T2 turned on even if the trouble contact returns to normal. Transistor T5 will remain off and keep T2 turned on until the acknowledgment button is pushed.

If the switch is in the non-lock-in position, the circuit between the collector of T5 and the base of T2 is open; therefore, T2 will turn off if the trouble contact returns to normal before acknowledgment and return all circuits to normal.

Operational Test See description under lamp test.

Auxiliary Contacts Auxiliary contacts are not supplied as part of the logic module. Adapter assemblies consisting of relays operated by the semiconductor logic, however, are available.

Relay Fail-Safe Feature Not available in solid-state circuits.

Normally Open Trouble Contacts The annunciator sequence and features already explained for NC trouble contacts operate in essentially the same way for NO contacts. The NO-NC option switch bypasses inverter stage transistor T1. When the contact closes on an abnormal condition, it turns on T2 though R20 and the sequence operation proceeds exactly as described for the NC operation.

Solid-state annunciators are for use with DC voltages ranging from 12 to 125 DC. Power consumption of the logic modules ordinarily is less than 5 watts. Visual indicators consume different amounts of power, depending on the type. Bull's-eye lights use approximately 1 watt, whereas back-lighted nameplates use from 1 to 6 watts, depending on the number and wattage of lamps.

Solid-state annunciators are very reliable and are not position sensitive. They offer the advantages of compactness, low power consumption, and little heat generation, factors that make them particularly useful in large integral annunciators. The per-point cost of solid-state systems is slightly higher than that of their relay-type equivalent, due to the cost of power supplies and interfacing accessories that may be required with solid-state systems. The cost of the logic modules, visual indicators, and cabinets themselves is not excessive. Integrated circuit components using recently developed microcircuits will most likely reduce size, power consumption, and heat dissipation of annunciator systems.

ANNUNCIATOR CABINETS

Annunciator systems are installed in areas ranging from general purpose to hazardous. Annunciator cabinets are installed indoors and outdoors in a variety of dusty, moisture-laden, and other adverse environments. Industrial annunciator cabinets are usually designed for general-purpose, dry indoor use. Special cabinets and enclosures are used in hazardous, moist, and outdoor locations.

Hazardous Area Designs

The requirements of Class 1, Division 2 hazardous locations as defined in Article 500 of the National Electric Code (NEC) are satisfied by the visual indicators and logic modules (either relay or solid state) of most annunciator systems. A manually operated or door-interlocked power disconnect switch is used with annunciator cabinets in those locations to turn off power when logic modules are relamped or changed.

Annunciator equipment for Class 1, Division 1 areas is installed in cast steel or aluminum housings approved for the hazardous environment. The housings are expensive to purchase and install. These annunciators are available in both integral and remote configurations. The remote type is generally wired to explosion-proof bull's-eye lights. Annunciator power must be disconnected either manually or automatically before the enclosures are opened to prevent an accidental arc or spark when logic modules are relamped or changed.

One can weatherproof annunciator cabinets installed in either general-purpose or hazardous areas (class 1, division 2) either by housing them in a suitable enclosure or by covering the exposed cabinet front with a weatherproof door. Housings that comply with class 1, division 1 requirements are also weatherproof. Figure 4.1p illustrates several weatherproof and hazardous-area enclosures. For the design of purging systems, refer to Figure 4.1b.

Intrinsically Safe Designs

Annunciators are classified as intrinsically safe if they are designed to keep the energy level at the trouble contact below that necessary to generate a hot arc or spark. Care must also be taken in installing the system to place wiring so as to prevent a high-energy arc or spark at the trouble contact caused by accidental short circuit or mechanical damage. Thus, general-purpose trouble contacts may be used with

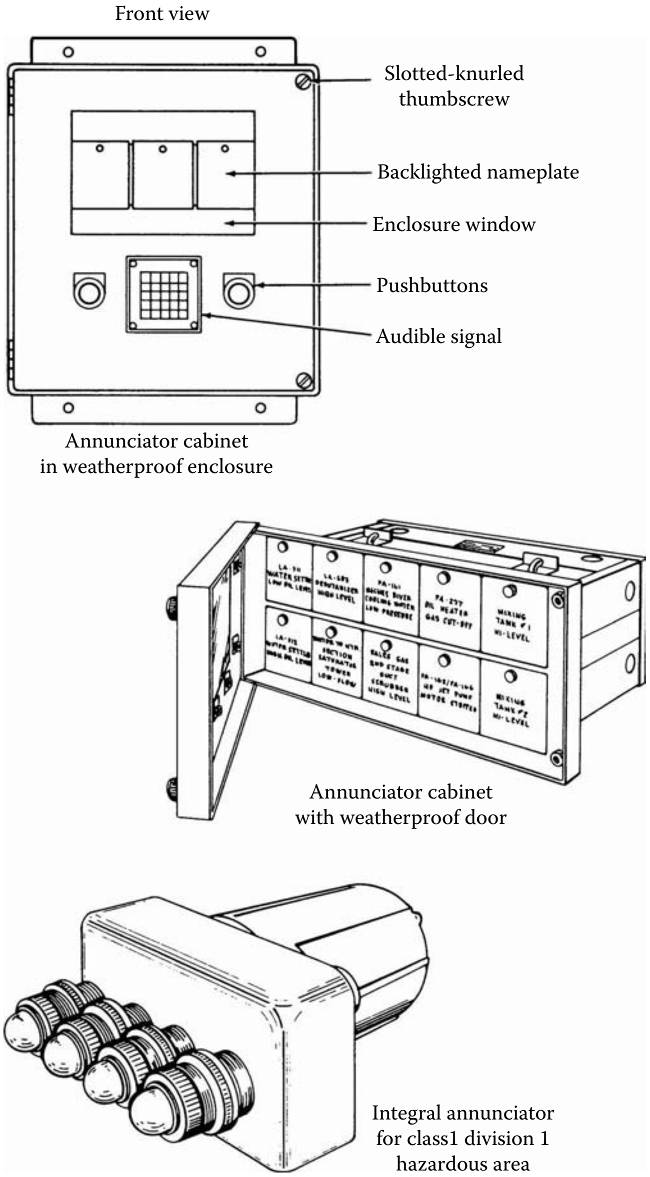


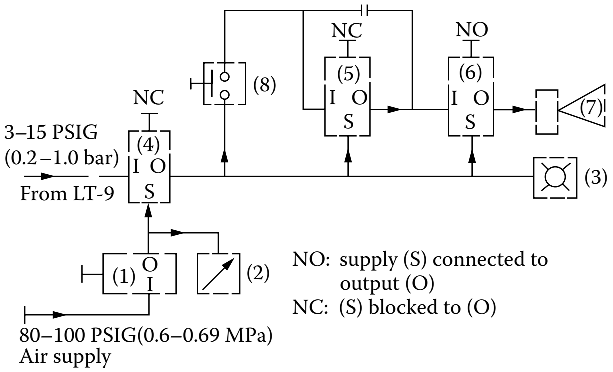
FIG. 4.1p
Annunciator enclosures for weatherproof and hazardous areas.

intrinsically safe annunciator systems even though they are installed in a hazardous area.

The annunciator logic modules and visual indicators, however, must conform to the electrical classification of the area in which they are installed. (For more on intrinsically safe designs, see Section 7.2 in Volume 1 of this handbook.)

PNEUMATIC ANNUNCIATORS

Pneumatic annunciators consist of air-operated equivalents of the trouble contact, logic module, and visual indicator stages of an electrical annunciator system. A single-point system furnishing high tank level monitoring is shown in Figure 4.1q.



Condition	Relay (4)	Relay (5)	Relay (6)	Horn (7)	Indicator (3)
Normal	Closed	Closed	Open	Off	Off
Alert	Open	Closed	Open	On	On
Acknowledge	Open	Open	Closed	Off	On
Normal again	Closed	Closed	Open	Off	Off

FIG. 4.1q
The tubing configuration and components of a pneumatic annunciator circuit.

Power supply to the system is instrument air at 80 to 100 PSIG (0.6 to 0.69 MPa), which is reduced to the required operating pressure by pressure regulator (1). The operating pressure is indicated on pressure gauge (2). A 3 to 15 PSIG (0.2 to 1.0 bar), an analog input signal from a direct-acting level transmitter (LT-9) enters high-pressure limit relay (4), which is normally closed and set to open when the high level limit is exceeded. When this happens (alert condition), an input at supply pressure from (4) turns on a pneumatic visual indicator (3), and a normally open high-pressure limit relay (6) allows supply air flow to air horn (7), turning it on. Simultaneously, the air output from (4) enters normally closed high-pressure limit relay (5) and momentary contact pushbutton (8), which is a normally open acknowledgment pushbutton for the system. In the alert condition, the pneumatic indicator and horn are both on.

One acknowledges the alert condition by pushing button (2), closing it, and thereby opening high-pressure limit relay (5). Supply air pressure from (5) closes high-pressure limit relay (6), which cuts off the operating air to the horn, thereby turning it off. Simultaneously, operating air pressure from (5) is fed back to the inlet of (5).

The feedback pressure locks up (5) so that it will not close when the acknowledgment pushbutton (8) is released. In the acknowledged condition, the pneumatic indicator is on and the horn is off. The system returns to normal when the 3- to 15-PSIG (0.2- to 1.0-bar) analog input falls below the setpoint. This closes high-pressure limit relay (4) which turns off the pneumatic indicator (3). It also closes high-pressure limit relay (5) by venting the lock-in circuit through relay (4).

Pneumatic annunciators are used when one or two alarm points are needed but electrical power is not readily available and in hazardous electrical areas where an electrical annunciator might not be practical. Pneumatic annunciators require a substantial amount of installation space and are expensive to manufacture.

Bibliography

"Alarm Systems, a Guide to Design, Management and Procurement," Engineering Equipment and Materials Users Association Publication No. 191.

Attwood, D., and Fennell, D., "Cost-Effective Human Factors Techniques for Process Safety," Paper Presented at CCPS International Conference and Workshop, Toronto, 2001.

Belke, J., "Recurring Causes of Recent Chemical Accidents," Paper Presented at MAPP, 1999.

"Better Alarm Handling," *The British Health & Safety Executive,* 1999.

Bullemer, P., "Managing Abnormal Situations: A New Operations Paradigm," Paper Presented at Honeywell North American Users Group Meeting, Phoenix, 2001.

Campbell Brown, D., and O'Donnell, M., "Too Much of a Good Thing? Alarm Management Experience in BP Oil, Part 1: Generic Problems with DCS Alarm Systems," Paper Presented IEE Colloquium on "Stemming the Alarm Flood," London, 1997.

Cochran, E., and Bullemer, P., "Abnormal Situations Management: Not by New Technology Alone," Proceedings of the 1996 AIChE Safety Conference, Houston, TX, 1996.

Cochran, E., Miller, C., and Bullemer, P., "Abnormal Situations Management in Petrochemical Plants: Can a Pilot's Associate Crack Crude?" Proceedings of the 1996 NAECON Conference, 1996.

Davis, R. C., "Vocal Alarms," *InTech,* December 1969.

Errington, J., and Bullemer, P., "Designing for Abnormal Situation Management," Proceedings of the 1998 AIChE Conference on Process Plant Safety, Houston, TX, 1998.

Fitzpatrick, B., "Alarms in the Real World," Paper Presented at Honeywell North American Users Group Meeting, Phoenix, 2001.

"Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems," IEC 61508, 1998.

"Functional Safety: Safety Instrumented Systems for the Process Industry Sector," IEC 61511, 2002.

Gangaloo, N. R., "New Annunciators," *InTech,* March 1989.

Gerold, J., "Managing Abnormal Conditions Pays Process Dividends," *Automation World,* September 2003.

Ghosh, A., *Abnormal Situation Management Is Your First Line of Defense,* Dedham, MA: ARC Advisory Group, 2000.

Hendershot, D., and Post, R., "Inherent Safety and Reliability in Plant Design," Proceedings of the Mary Kay O'Connor Process Safety Center Annual Symposium, College Station, TX, 2000.

ISA TR91.00.02, "Criticality Classification Guideline—Enunciator Sequences and Specifications," ANSI/ISA S18.1, 1992.

Jamieson, G., and Vicente, K., "Modeling Techniques to Support Abnormal Situation Management in the Petrochemical Processing Industry," Proceedings of the Symposium of Industrial Engineering and Management, Toronto, Ontario, Canada, 2000.

Jutila, J. M., "Guide to Selecting Alarms and Annunciators," *InTech,* March 1981.

Mattiasson, C., "The Alarm System from the Operator's Perspective," Paper Presented at IEE People in Control Meeting, Bath, U.K., 1999.

Mostia, B., "How to Perform Alarm Rationalization," *Control,* August 2003.

National Electrical Code, Articles 725 and 760, National Fire Protection Association, 1981.

Nimmo, I., "Abnormal Situation Management," *Chemical Engineering Progress,* September 1995.

Nimmo, I., "The Importance of Alarm Management Improvement Project," Paper Presented at ISA INTERKAMMA, 1999.

Nochur, A., Vedam, H., and Koene, J., *Alarm Performance Metrics,* Singapore: Honeywell Singapore Laboratory, 2001.

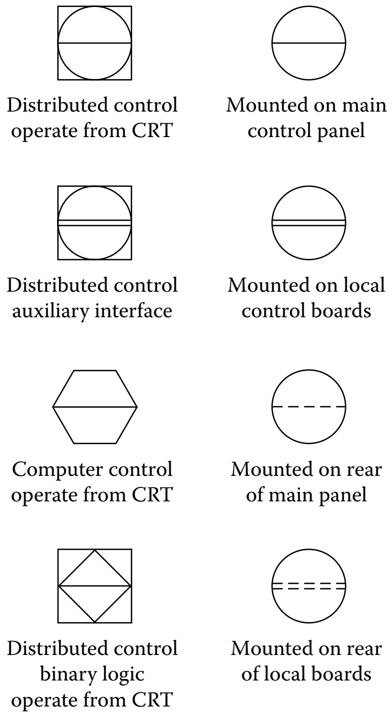
PAS (Plant Automation Services), *The Cost/Benefit of Alarm Management*, Houston, TX: Plant Automation Services, 2000.

Poole, A. D., "Design Considerations for Discrete Alarm Systems," *InTech,* September 1992.

Smith, W., Howard, C., and Foord, A., *Alarm Management—Priority, Floods, Tears or Gain?* www.4-sightconsulting.co.uk: 4-sight Consulting, 2003.

4.2 Control Centers and Panels—Traditional

R. W. BORUT (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbols

Costs:	The cost of conventional panels ranges from \$2000 to \$8000 per linear foot (\$6000 to \$25,000 per meter) depending on type and concentration of instruments and graphics. The costs of the instruments mounted on the panel are not included in these figures. For estimating the cost of distributed control systems (DCS), the cost of workstations ranges from \$15,000 to \$30,000 depending on the number of displays, computers, and input devices. These costs do not include the expenses associated with the integration of hardware with software. Refer to the later sections of this chapter for a more detailed treatment of DCS costs.
Note:	Because many process units are constructed by members of building trade unions, prefabricated control boards will not be permitted on the site unless they have labels certifying that they were constructed by members of the required unions. The required labels are United Association of Journeymen and Apprentices of the Plumbing and Pipe Fitting Industry of the United States and Canada (UA) and International Brotherhood of Electrical Workers (IBEW).
Partial List of Suppliers:	ABB Inc. (www.abb.com) Bebco Industries (www.okbebco.com) Daniel Measurement and Control (www.danielind.com) Emerson Process Management (www.mdctech.com) Eurotherm Controls Inc. (www.eurotherm.com) Foxboro-Invensys (www.foxboro.com) Honeywell Industrial Control (www.iac.honeywell.com) Johnson Controls Inc. (www.johnsoncontrols.com) Moore Industries International Inc. (www.miinet.com) Rockwell Automation (www.rockwellautomation.com) Siemens Energy and Automation (www.sea.siemens.com) Westinghouse Process Control (www.westinghousepc.com) Yokogawa (www.yokogawa.com/us)

INTRODUCTION

The operator should be the focal point of any control room design because operators are the real experts on the daily operation of the plant. In the traditional control room the control panel was the operator's window on the process, through which the operator judged how the process was doing and based on that information, determined the way to control it. The operator makes hundreds of decisions every day as problems with supply materials and equipment operation develop or when schedules change, tests and reports are needed, etc.

In traditional control rooms (Figure 4.2a), the operators adapted the control panel to their needs by attaching notes to the various instruments or by marking critical conditions and the proper responses to them. This was easily done because the instruments associated with the various unit processes were mounted on the panel in logical clusters. In DCS-based systems one must be careful to provide the operator with the same logical convenience for alarm handling, status display review, and navigation. Particular care is required to make start-up and shut-down controls convenient and not to force operators to move between various interface systems to do their jobs.

This section describes the design of both conventional and DCS-based control rooms and a variety of traditional panel designs that were used in connection with analog and pneumatic control systems. DCS-based control rooms are also discussed, but more briefly because the details of many aspects of DCS systems are discussed in other sections of this chapter and the reader is therefore advised to also refer to those sections.

The design, wiring, and tubing of conventional control panels are discussed in this section. ISA, The Instrumentation, Systems, and Automation Society, has prepared a number of standards and recommended practices on the design of both traditional and DCS-based control centers. These are recommended reading for those who are responsible for the detailed design of such facilities. The documents that have been prepared include the following:

RP60.1-1990	Control Center (CC) Facilities
RP60.2-1995	CC Design Guide and Terminology
RP60.3-1985	Human Engineering for Control Centers (updated in 1985)
RP60.4-1990	Documentation for Control Centers
RP60.6-1984	Nameplates, Labels and Tags for Control Centers (updated in 1984)
RP60.8-1978	Electrical Guide for Control Centers
RP60.9-1981	Piping Guide for Control Centers
RP60.1-1991	Crating, Shipping and Handling of CC

For other ISA standards and recommended practices refer to http://www.isa.org.

TRADITIONAL CONTROL ROOMS

Conventional control panels often consist of several sections, each corresponding to a different processing step or unit operation (Figure 4.2b). The control room itself must be so designed that only those operations necessary for the control of the plant are performed there. The operators must not be distracted by unassociated functions. The room should have limited access and should not act as a passageway. Equipment must be arranged in such a way that unauthorized personnel cannot tamper with the instruments or with the auxiliaries mounted close by. Figure 4.2c shows a traditional control room layout for analog instruments.

In the control room, air conditioning and room pressurization must be provided. Aside from ensuring operator comfort, maintaining a constant ambient temperature at the instruments will also minimize signal drift. Room pressurization is used where the plant atmosphere is explosive or flammable. The control room is pressurized by admitting into it fresh and clean air from a safe area. This permits the reduction of the area

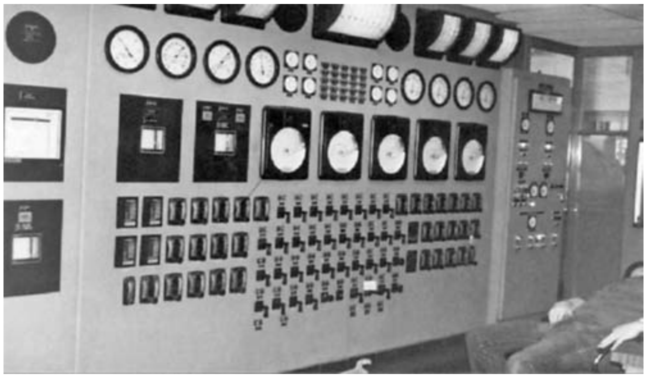


FIG. 4.2a
A traditional boiler control panel from the 1980s with a mix of pneumatic and electronic instruments. (Photo courtesy of Jim Mahoney.)

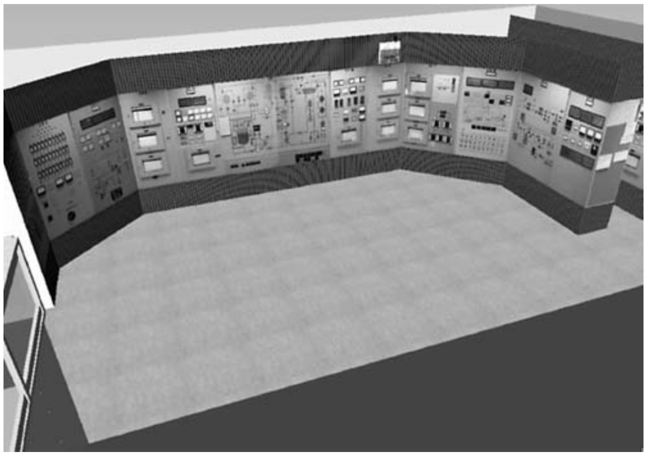


FIG. 4.2b
Traditional analog control panels were configured from several segments, each assigned to a particular unit operation.

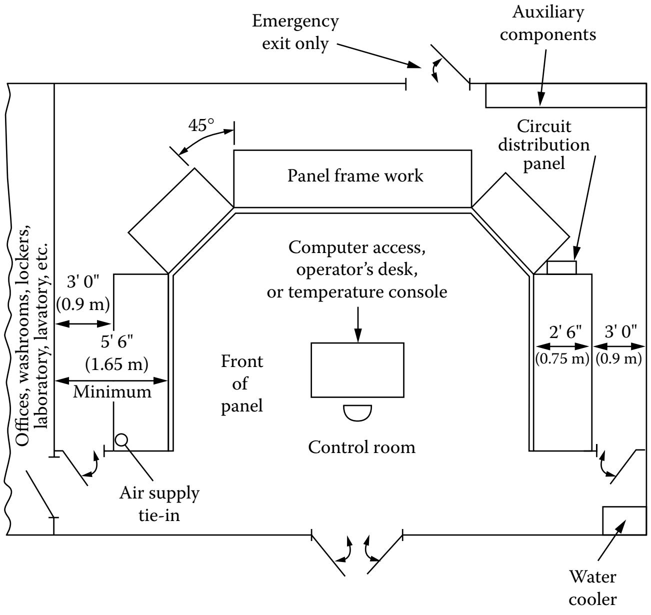


FIG. 4.2c
Traditional control room layout.

classification from either "hazardous" or "semihazardous" to unclassified, with commensurate savings in instrument and installation costs. (Section 7.2 in the *Process Measurement* volume of this handbook gives details on area classification.)

The illumination in the control room must be of a level consistent with close work. The lighting intensity of the panel should average 75 foot-candles (807 lx) across its face. The back of the panel area should be lighted to 30 foot-candles (322.8 lx). The lighting system should be designed to minimize reflections on instrument cases, and point sources of light should be avoided. Continuous fluorescent lighting, placed behind egg crate-type ceiling fixtures, will give adequate light and will minimize annoying highlights.

The most advantageous ratio of panel length to control room area is obtained by bending the panel to a U shape. Right-angled bends of the panel, as opposed to 45-degree bends, should be avoided. The slightly increased panel length that could be gained by the use of right angles is negated by the interference to opening instrument doors or withdrawing the chassis. Also, operators can monitor a greater length of panel if it bends around them.

CONTROL ROOMS FOR DCS SYSTEMS

The upgrading of traditional control rooms is discussed in the next section, while other aspects of DCS-based and other digital control systems are covered in other sections of this volume.

In physical appearance the control equipment used in DCS systems is similar to modern office equipment. CRT-based consoles, keyboards, printers, and personal computers are some of the basic high-visibility components in these control centers. The design of these workstations is modular and can be configured to match the number of operators and other features of the particular plant (Figures 4.2d and 4.2e).

DCS-based control rooms require significantly less space than did traditional control rooms. On the other hand, DCS-based control rooms do require additional rooms for auxiliary equipment and other purposes. These rooms should be located adjacent to the control room itself (Figure 4.2f). In these rooms computer-type floors are used to allow for the routing of the cables between units.

A well-designed control center of this type requires a significant amount of advance planning. The input/output (I/O) and peripheral equipment cabinets are usually located in rooms adjacent to the control center. These adjacent rooms also include the central processing units (CPUs) and their supporting gear.

The overall control house might consist of two rooms (Figure 4.2f), one for the control room operators and the other for the auxiliary equipment, or might consist of several rooms, as shown in Figure 4.2g. In either configuration it is recommended to provide adequate spare cabinet space and display area to accommodate later plant enlargement.

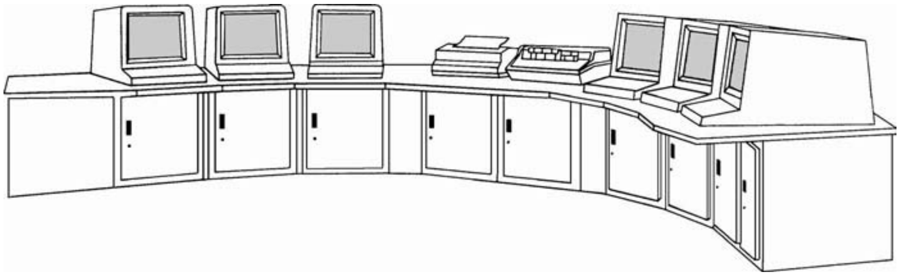


FIG. 4.2d
CRT-based workstations can be configured like office furniture to match the operating philosophy of the plant. (Courtesy of Siemens, formerly Moore Products Co.)

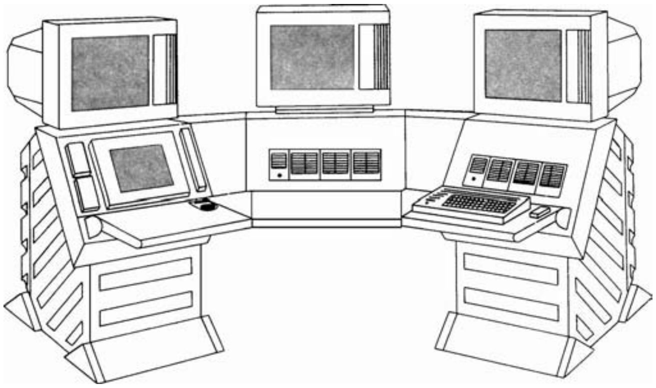


FIG. 4.2e
The workstations can be configured for a single-bay or multi-bay configuration to match the operating strategy and the number of control room operators. (Courtesy of Invensys, formerly The Foxboro Co.)

TRADITIONAL CONTROL PANELS

There are three basic shapes of control panels—straight, breakfront, and console. Each type has an accompanying family of variations. The dimensions shown in the various illustrations are only typical, and instrument heights and spacing must be adjusted to suit the particular manufacturer and application.

Flat Panels

The flat panel shown in Figure 4.2h is the least expensive, the easiest to construct, and the simplest to design. The straight, vertical plane of the panel allows an orderly layout of tubing, electrical ductwork, and miscellaneous equipment. Instruments and auxiliary components can be arranged so that all are accessible for maintenance and calibration. The lower row of instruments is approximately 3 feet 3 inches (0.975 m) from the floor and should be used to mount recording or indicating instruments only because this low elevation is rather inconvenient for controller operation. In this layout a maximum of four horizontal rows of miniature instruments (recorders and controllers) can be used, which is less than with some of the breakfront console configurations.

Breakfront Panels

Breakfront panels allow greater use of the front face of the board because the instruments located in the lower rows are swung upward to a convenient height (Figure 4.2i). The top portion of the panel is swung downward to an angle normal to the line of sight, allowing better visibility. The additional rows obtained with this layout tend to reduce the overall panel length requirement. However, the higher the instrument density, the less room is left for maintenance and for the mounting of auxiliary components in the back of the panel.

Consoles

Consoles are aesthetically pleasing to most people. They are often used with high-density instrumentation in control rooms of limited size. Normally, their length is a function of operator responsibility and convenience. Usually there is one operator

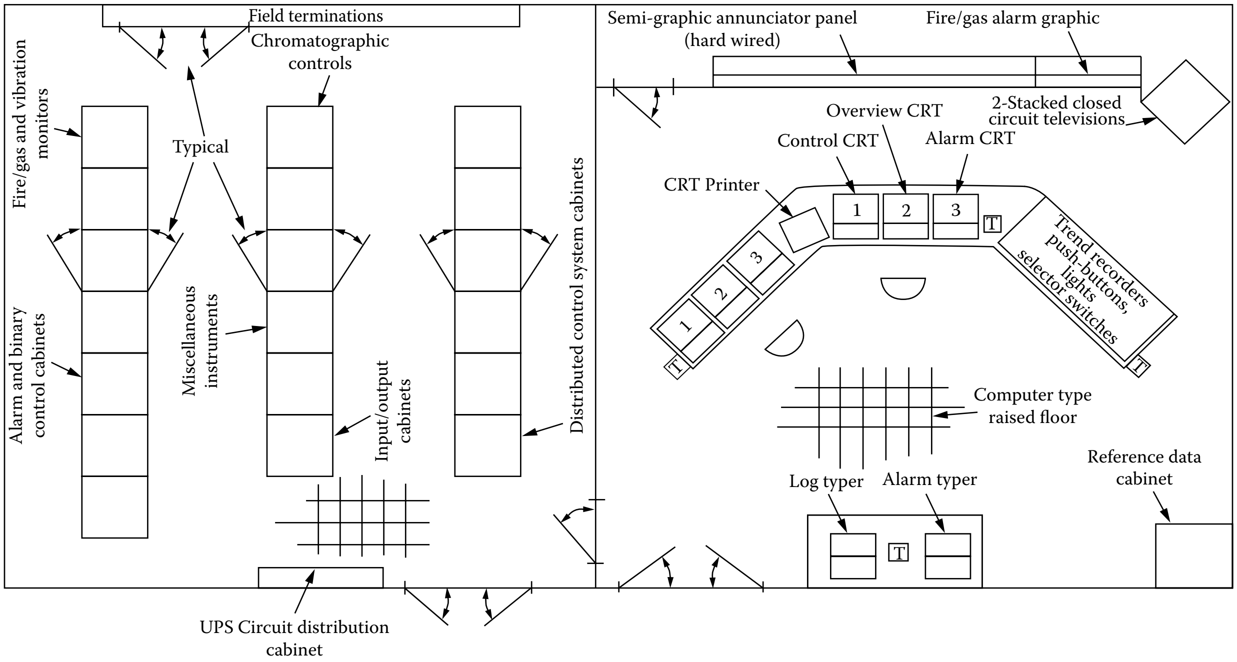


FIG. 4.2f
Typical control house arrangement for distributed control instrumentation.

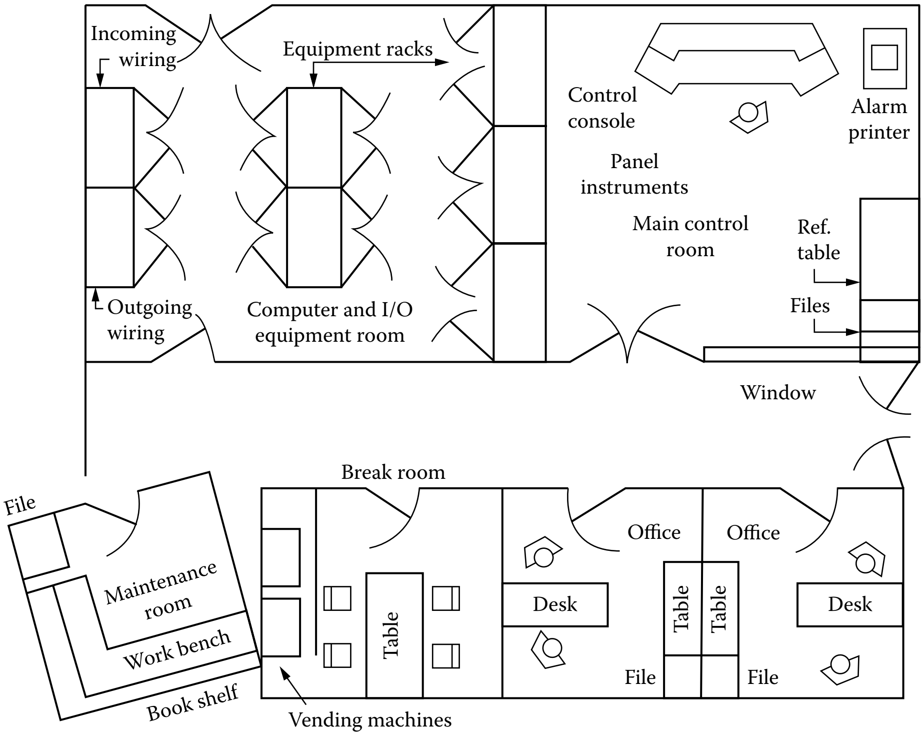


FIG. 4.2g
The ideal control house layout can be arrived at by trying a variety of configurations using 20:1 scale cutouts of the required equipment.[1]

per console and one console per operator. The lengths vary from 4 to 12 feet (1.2 to 3.6 m).

Auxiliary equipment, such as transducers and pressure switches, can be installed inside the console cabinet, but the arrangements of flush-mounted instruments, as shown in Figure 4.2j, severely limit the available free space in the back of the console. Performing instrument maintenance within the consoles can therefore become a real problem.

Consoles are often coupled with a conventional flat "backup" panel, which contains the larger instruments and auxiliary components. Because the average operator is less than 5 feet 9 inches (1.7 m) tall, the "see over" console should not rise above 5 feet 0 inches (1.5 m). This will allow the operator to see the backup panel over the top of the console.

TRADITIONAL FRONT PANEL LAYOUTS

Analog instrument sizes and shapes include the large-case conventional type (nominally 18 inches wide by 24 inches high [450 mm wide by 600 mm high]), the miniature type (nominally 6 inches square [150 mm square]), and the high-density type (2 inches wide by 6 inches high [50 mm wide by 150 mm high]). The same considerations that determined the panel type often also determined the selection of the instrument size.

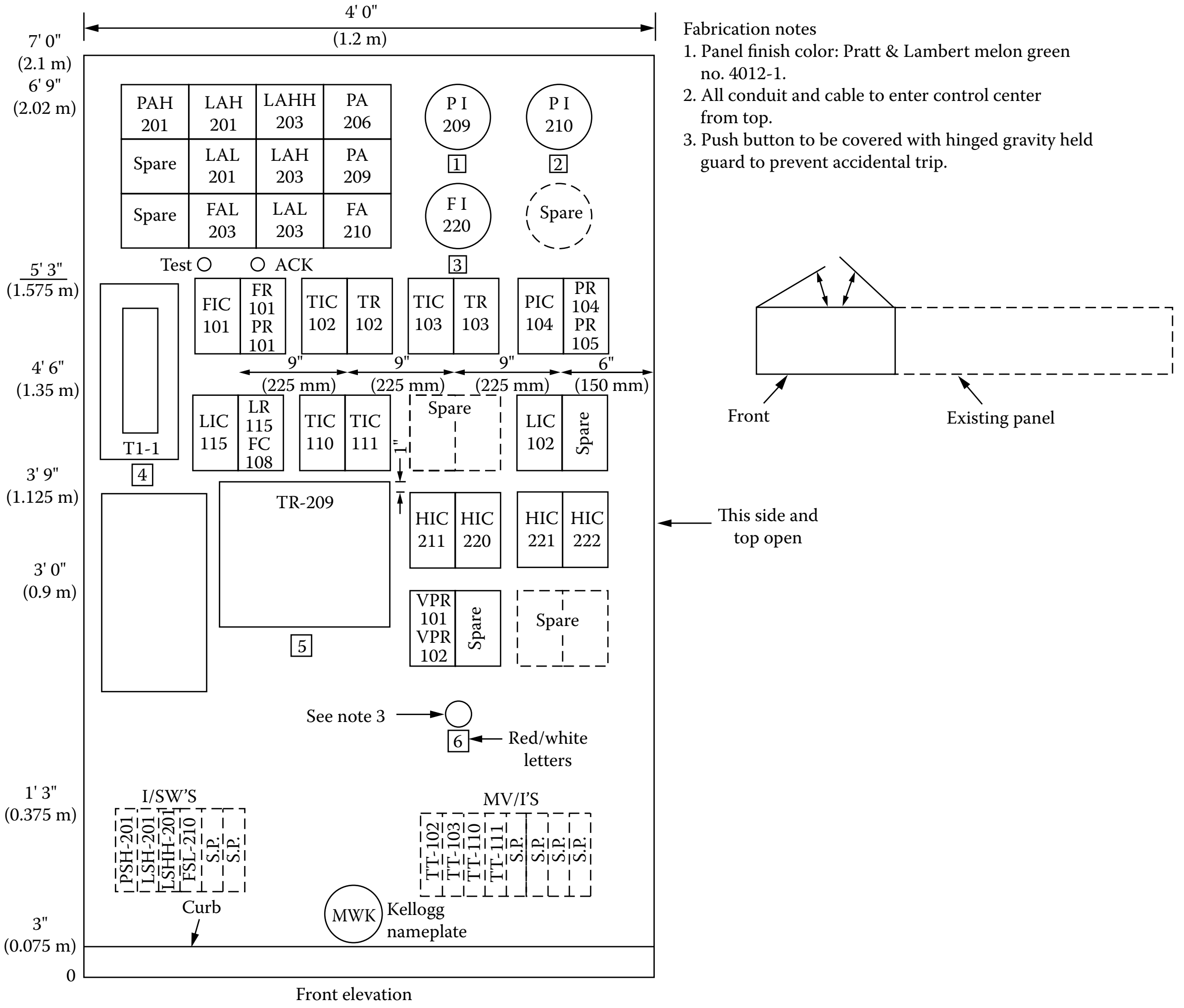
Large-Case Instruments

Today, seldom seen are the large-case, conventional instruments, which were installed on old panels, mounted two rows high (Figure 4.2k). They did not permit a great deal of flexibility in layout, but they were rugged and suitable for almost any outdoor location as local control stations. The control systems for most of these applications were simple with few instrument components.

Miniature Instruments

Prior to the digital age and DCS systems, miniature instruments were the most widely used process control devices. Relative to large-case instruments, they allowed a moderate reduction in panel length (Figure 4.2l). They were used for most indoor installations. These devices can be used in almost all configurations of flat panel, breakfront, or console designs, but there are some mounting angle limitations that should be kept in mind when the panel is designed.

For estimating the length of control panels, a horizontal spacing of 9 in. (225 mm) between vertical center lines of miniature instruments was used. Ten to 12 inches may be used between horizontal center lines. Instrument manufacturers have not fully standardized the overall dimensions, cutouts, connection locations, and so forth. Therefore, the manufacturer's spacing and installation recommendations for each instrument had to be checked and followed.



Fabrication notes
1. Panel finish color: Pratt & Lambert melon green no. 4012-1.
2. All conduit and cable to enter control center from top.
3. Push button to be covered with hinged gravity held guard to prevent accidental trip.

FIG. 4.2h
Typical flat control panel layout.

High-Density Instruments

There are several designs of high-density-type miniature instruments available on the market. They may be mounted in groupings to suit a particular processing unit, as shown in Figure 4.2m, or in long rows to condense panel length, as illustrated in Figure 4.2n.

This type of instrument layout requires additional space on the rear of the panel for auxiliary equipment. Additional devices can also be mounted on the wall behind the panel or in a peripheral equipment room remote from the panel. High-density instruments tend to have a longer chassis, requiring about 6 additional inches in panel depth. Some of these instruments require amplifiers or converters, which must be mounted in close proximity to the primary instrument.

Some high-density instrument lines did not include a similarly sized recorder. In these cases, a standard miniature strip chart recorder (6 by 6 inches, or 150 by 150 mm) was utilized (Figure 4.2l). In some cases, instead of permanently installed recorders, trend recorders were used to reduce the overall recorder requirements.

Graphic Panels

A graphic control panel depicts a simplified flow diagram of the processing unit and describes its control philosophy. Prior to DCS systems, the most common material used for this depiction was colored plastic or melamine. The lines and symbols were affixed to a removable steel, aluminum, or plastic plate.

The extra expense of a graphic panel was justified with the following reasons:

1. To enable the panel operator to visualize a complex process flow pattern
2. To make a sophisticated control philosophy with complex interrelationships between variables understandable

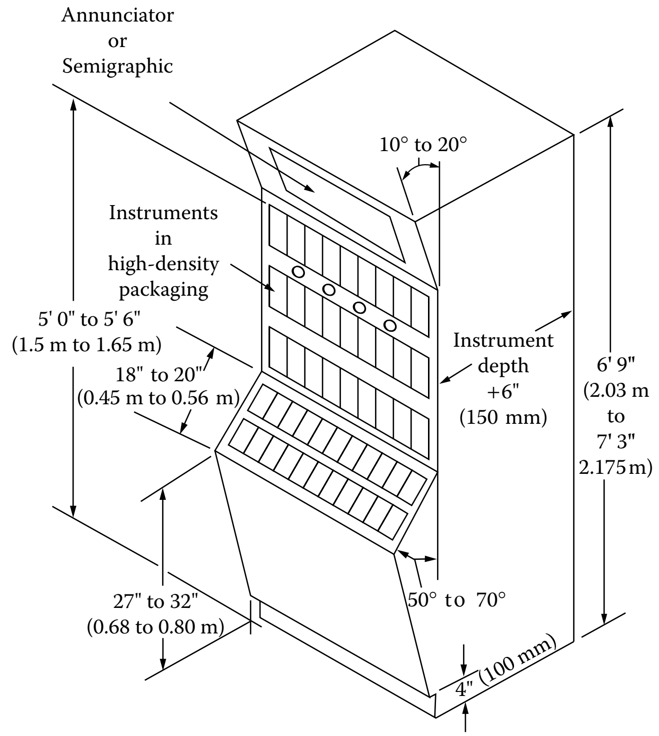


FIG. 4.2i
Breakfront console design.

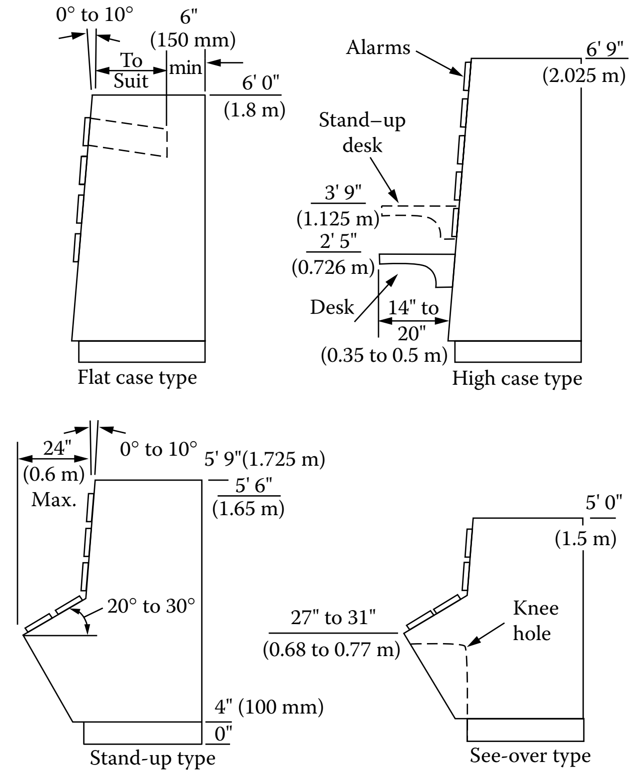


FIG. 4.2j
Variations of console shapes.

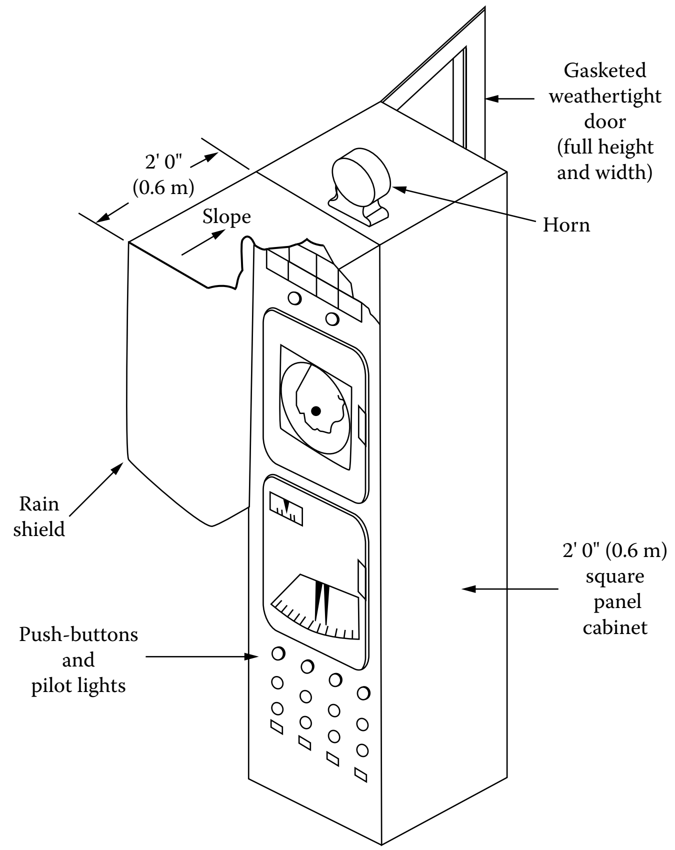


FIG. 4.2k
Old local panel cabinet, which was used to house large-case instruments.

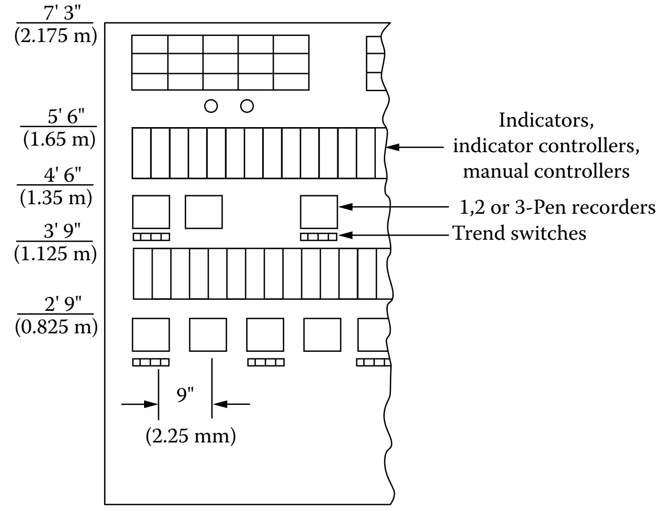


FIG. 4.2l
Panel layout for miniature analog instruments.

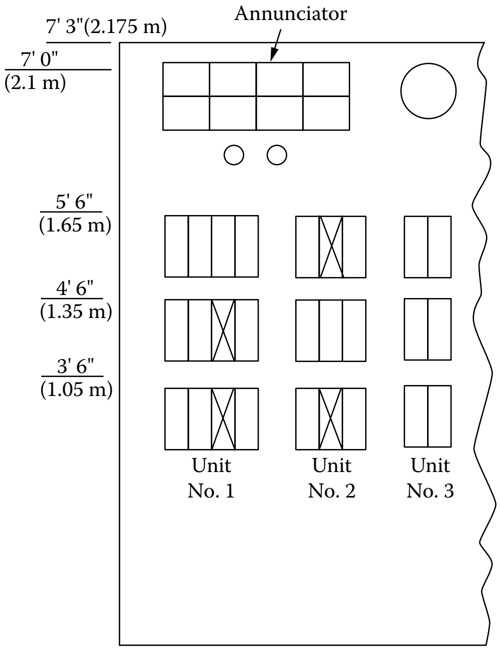


FIG. 4.2m
Miniature analog instruments in a grouped layout.

3. For the training of new operators
4. For aesthetic enhancement of the control room (a motivation seldom admitted but frequently present)

Although the shapes or dimensions for equipment symbols were not standardized, there were some agreements in the adoption of some common symbols for graphic layout purposes, such as the ones shown in Figure 4.2o. Here, the dimensions shown are typical and can be adjusted to suit individual requirements, as long as they are consistent throughout the panel. Symbols representing equipment, such as furnaces, vessels, drums, pumps, and compressors, are usually shown in silhouette. Internals are shown only when necessary to ensure complete understanding of equipment functions.

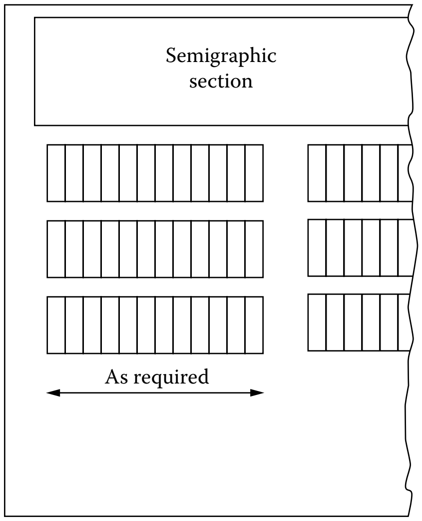


FIG. 4.2n
Semigraphic panel with high-density layout.

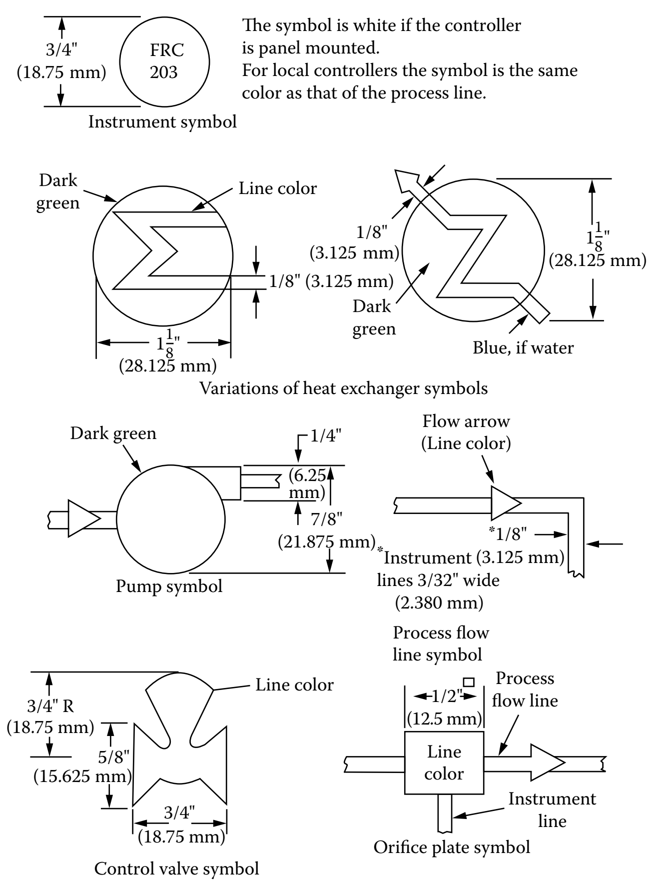


FIG. 4.2o
Symbols used on graphic panels.

In general, the graphic display showed the process flow from left to right in a step-by-step sequence, where fresh feed enters from the left and the product stream exits to the right. Only the process streams and those utilities that are required for clarifying unit operations were normally shown. Similarly, only the most important local instrumentation was depicted.

The flow diagrams were laid out with horizontal and vertical lines. The pattern and its density was as consistent as possible. Vessels usually encompassed the middle three fifths of the graphic section, with a common bottom line whenever practical. Equipment in similar service often occupied similar positions. For example, reflux drums would be aligned along the upper three-fifths division line, while reboilers would rest atop the lower two-fifths line, and so forth. This leaves the top and bottom fifths of the panel for long horizontal lines. The graphic diagram was designed so that as few lines as possible would cross each other. Where a crossing was necessary, the horizontal line was broken.

Semigraphic Panels Semigraphic panels were either flat or of the breakfront type. The top portion of the panel was occupied by a flow diagram of the process (Figures 4.2p).

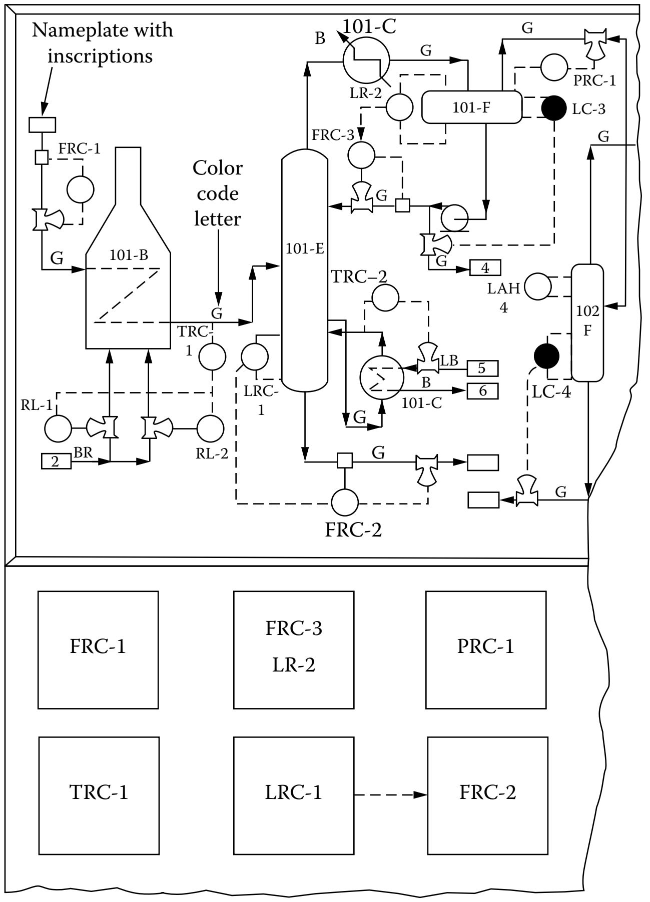
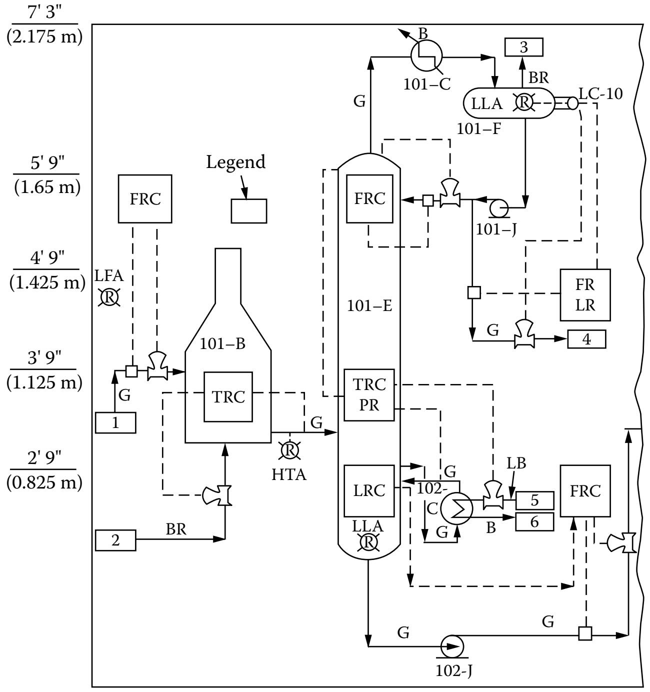


FIG. 4.2p
Partial front view of semigraphic panel with miniature instruments.



Legend
G – Process line (green)
BR – Vent or fuel gas (brown)
LB – Steam (light blue)
B – Water (blue)
S – Instrument line (silver)
No. – Nameplate with inscription

FIG. 4.2q
Full graphic panel with miniature instruments incorporated into the graphics.

The location of each instrument was so selected that it was installed directly underneath its corresponding symbol in the graphic diagram (or as close to this point as was feasible). The instrument item number was clearly noted in both locations. In some instances, alarm or running lights were also installed within the graphic section in the appropriate locations. The instrument density was reduced when using this type of panel because of the desirability of locating the instruments in relative proximity to their graphic symbols.

Full Graphic Panels In this design, the graphics cover the complete front face of the control panel, as shown in Figure 4.2q. On these panels the instruments were positioned at the points that corresponded to their location in the process. As in all other panels, the instruments were aligned in horizontal and vertical rows for ease in conduit, duct, and tube layout. The instrument density for this type of panel is extremely low and can vary from one to three instruments per linear foot of board length. This low density significantly extends the overall length of the panel.

Graphic Panel Constructions In addition to the glued-on plastic-type graphic panels, other designs exist. If there is a likelihood of process changes, the degree of difficulty in making changes to the graphic diagram should be considered. The aforementioned plastic strips are relatively easy to change. If the lines and symbols are painted on the panel, changes can be made easily when the panel is new, but as the paint ages and fades, colors become difficult to match.

A flow diagram can also be drawn full scale on reproducible paper and placed in a semigraphic frame. It may be held in place and protected with a glass or clear rigid plastic cover. This method of graphic presentation is simple to change by revising the drawing master and acquiring a new print. If this type of panel is made in only black and white, it is not as aesthetically pleasing as other types.

Another method of graphic process display is by the use of automatic slide projectors. This system uses a rear-projection-type, translucent screen. In order to gain the necessary focal length distance, a set of mirrors is strategically placed on the rear of the panel. This approach to graphics allows a great deal of flexibility. As many slides can be prepared as desired, showing the process or instrument systems or subsystems in color and in as much detail as is deemed necessary. Slides can be prepared to give instructions for emergency procedures, unit operating parameters, and optimum set points for various product choices.

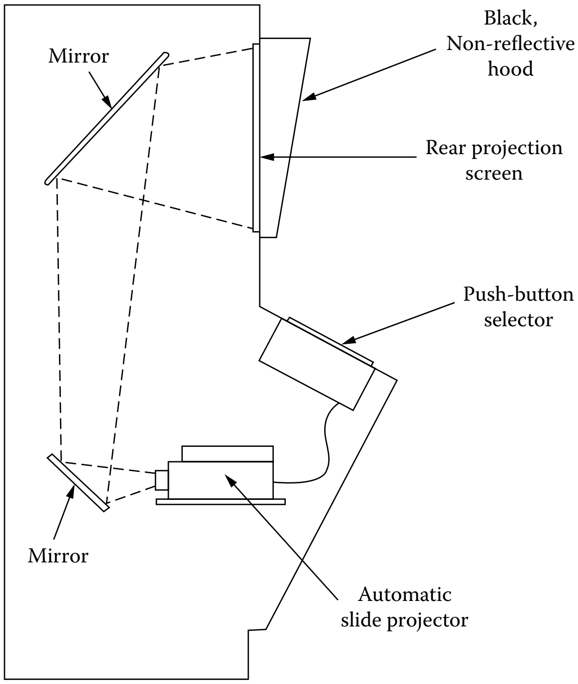


FIG. 4.2r
Breakfront console with rear projection system for graphic presentation.

One serious drawback to the system is that the bright light in the control room tends to "wash out" the image and reduce legibility. Therefore, the lighting level must be reduced, and the fixtures must be so located as to minimize reflections on the screen. In addition, the screen should be furnished with a black shield to reduce side lighting (Figure 4.2r).

Yet another type of graphic panel design uses the back-engraving of a sheet of clear plastic. The plastic lines are then filled with the selected colors. This type of panel is easy to maintain because only the smooth surface is exposed. Changes can be made by reengraving and by refilling. This is extremely difficult, and even when it is expertly done, telltale signs of the change remain.

The front-engraved and enamel-filled graphic line-work allows vivid coloring and sharp lines, making this panel one of the most impressive and pleasing to the eye. It is not feasible to make changes on this panel, and maintenance is somewhat more difficult, since the engraved lines tend to fill with dust.

Back-of-Panel Layout

In order for the overall panel design to be well executed, sufficient attention must be given to the back-of-panel arrangement. It must be verified that there is sufficient room to run the conduits, ductworks, air headers, tube leads, and so forth. There must also be enough room to mount the various switches, relays, converters, amplifiers, and other auxiliary components. A back-of-panel profile will sometimes assist in this (Figure 4.2s). Auxiliary equipment must be located such that its connecting wiring or tubing does not obstruct the maintenance or calibration of, and accessibility to, the other instruments.

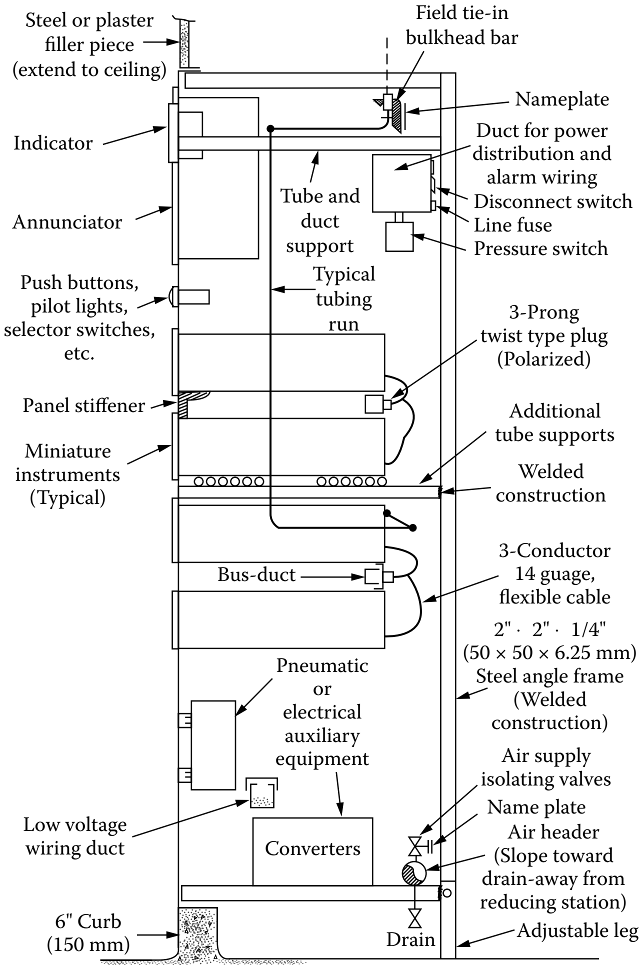


FIG. 4.2s
Back of panel arrangement for miniature analog instrumentation.

Panel Materials of Construction

The materials used most frequently for the construction of panel boards are steel and various plastics. Materials less frequently used are stainless steel, aluminum, and fiberglass.

Steel The advantages of steel are as follows:

1. *Strength.* At instrument cutouts, where the panel is weakened, steel requires less stiffening than do other materials. Steel stands up better to the cantilever effect of the deep bezel-supported, flush-mounted instruments.
2. *Ease of construction.* Holes can be drilled or flame-cut. Auxiliary equipment supports can be welded at any point. Bend-backs on straight panels and the breakfront shapes add significantly to panel stiffness.
3. *Safety.* When grounded, the panels offer an excellent path to ground if an instrument chassis or case

becomes energized as a result of a short circuit or mistermination.

4. *Attractive finish.* Long lengths can be finished in any color or hue desired, without visible seams.
5. *Ready availability.* Sheet steel of good quality and with a minimum of surface pitting is readily available.

Disadvantages of steel panels include the following:

1. *High susceptibility to corrosion.* In corrosive atmospheres, the finish is only as durable as the paint.
2. *Difficulty in adding cutouts.* Once the panel has been constructed and instruments are installed, cutting the steel to accommodate new instruments scatters steel particles that may interfere with the mechanical or electrical operation of the surrounding instruments.

Plastics The plastics available for panel boards are all melamines: Formica, Peonite, Micarta, Textolite, and so forth. They are used either as sheets 1/2 in. (12.5 mm) in thickness or as laminates. The cores of the laminates are materials such as aluminum, flakeboard, steel, or plywood.

Plastic panels, like those of steel, have both favorable and unfavorable aspects:

1. *Great durability.* The finish is very resistant to scratching and heat.
2. *Availability in colors.* Sheets come in many colors, but if the desired color is not in stock, delivery can be delayed. A color may vary slightly from one run at the factory to the next.
3. *Seam required every 4 feet (1.2 m).* This is true regardless of panel length because sheets are usually 4 by 8 feet (1.2 by 2.4 m).
4. *Steel frame required.* The plastic panel must be bolted to a skeletal steel frame for support.
5. *Routing of holes may be necessary.* Many switches, pilot lights and other flush-mounted instruments are designed for a 3/8-in. (9.36-mm) maximum panel thickness. Solid Formica and some laminates will require routing in the back of the panel, around the cutout for bezel-locking rings or locknuts, because of their thickness.

PANEL SPECIFICATIONS

In addition to the drawings and diagrams described earlier, a written specification covering other important aspects of panel manufacturing must be developed. This specification is the document that precisely instructs the panel manufacturer as to design options and the materials to be used to fulfill the contract.

A panel specification should include a delineation of at least the following requirements:

1. *General.* Definition of the design drawing specification and codes furnished by the purchaser that the panel manufacturer is to follow.
2. *Engineering.* Description of the extent and type of engineering drawings to be developed by the panel manufacturer, including whether "as-built" drawings are required. It also includes the number of prints and reproducible masters required and the approval or review requirements for preliminary designs.
3. *Construction.* Description of the type of panels and their fabrication. This includes National Electric Code (NEC) area classification, ambient conditions, and similar requirements.
4. *Design.* Specification of methods of installing wiring and piping systems. This includes a listing of materials of construction for wire, pipe, tubing, ducts, nameplate inscriptions, and so forth.
5. *Materials.* Complete description of all materials to be used. A generic description is usually sufficient.
6. *Cost.* The specification should direct the bidder to delineate various costs so that additions and deletions to the contract can be negotiated more easily. A Quotation form (such as is shown in Figure 4.2t), which is completed by the control center manufacturer, should be included with the bid.
7. *Inspection.* Delineation of the number and types of inspections planned, which may include preliminary inspections during specific stages of construction. This section of the specification should also describe the extent of inspection required, such as visual, point-to-point checks or functional testing.
8. *Shipping.* Specification of the type of conveyance used to ship the panel to the plant site, type of crating, and protection requirements.
9. *Guarantees.* Conditions under which a panel or equipment may be rejected and the length of time during which the panel is covered by the manufacturer's warranty.

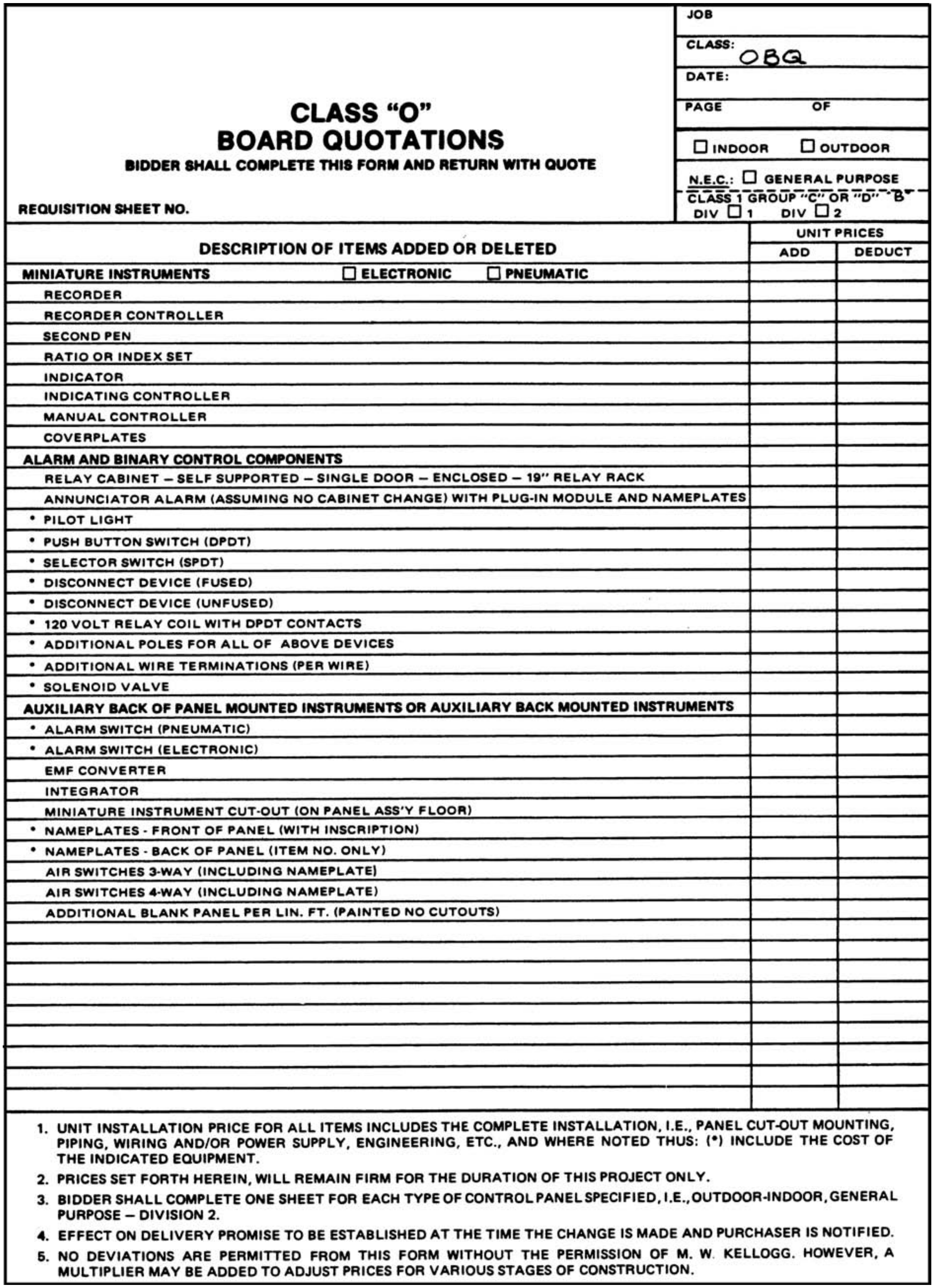
HUMAN ENGINEERING

Human engineering or ergonomics has already been discussed in full detail in Section 1.5, so it is only briefly covered here. Ergonomics has proved invaluable in the design of all panels and consoles.

Figure 4.2u describes the reach of 95% of the people in the United States when conveniently operating controls. Median heights and reaches of people from other nations should be used if the control center is to be used in another country.

Studies performed in nuclear energy power plant control rooms have shown that psychological as well as physical considerations are important. Many of the rules listed below are obvious, yet they are often overlooked:

1. Indicating lights and pushbuttons should be colored so that their relative importance is immediately recognized: i.e., green = go or safe, red = danger or stop, yellow = caution or slow.



CLASS "O" BOARD QUOTATIONS

BIDDER SHALL COMPLETE THIS FORM AND RETURN WITH QUOTE

REQUISITION SHEET NO.

JOB

CLASS: OBQ

DATE:

PAGE OF

☐ INDOOR ☐ OUTDOOR

N.E.C.: ☐ GENERAL PURPOSE

CLASS 1 GROUP "C" OR "D" "B"

DIV ☐ 1 DIV ☐ 2

DESCRIPTION OF ITEMS ADDED OR DELETED	UNIT PRICES	
	ADD	DEDUCT
MINIATURE INSTRUMENTS ☐ **ELECTRONIC** ☐ **PNEUMATIC**		
RECORDER		
RECORDER CONTROLLER		
SECOND PEN		
RATIO OR INDEX SET		
INDICATOR		
INDICATING CONTROLLER		
MANUAL CONTROLLER		
COVERPLATES		
ALARM AND BINARY CONTROL COMPONENTS		
RELAY CABINET – SELF SUPPORTED – SINGLE DOOR – ENCLOSED – 19" RELAY RACK		
ANNUNCIATOR ALARM (ASSUMING NO CABINET CHANGE) WITH PLUG-IN MODULE AND NAMEPLATES		
• PILOT LIGHT		
• PUSH BUTTON SWITCH (DPDT)		
• SELECTOR SWITCH (SPDT)		
• DISCONNECT DEVICE (FUSED)		
• DISCONNECT DEVICE (UNFUSED)		
• 120 VOLT RELAY COIL WITH DPDT CONTACTS		
• ADDITIONAL POLES FOR ALL OF ABOVE DEVICES		
• ADDITIONAL WIRE TERMINATIONS (PER WIRE)		
• SOLENOID VALVE		
AUXILIARY BACK OF PANEL MOUNTED INSTRUMENTS OR AUXILIARY BACK MOUNTED INSTRUMENTS		
• ALARM SWITCH (PNEUMATIC)		
• ALARM SWITCH (ELECTRONIC)		
EMF CONVERTER		
INTEGRATOR		
MINIATURE INSTRUMENT CUT-OUT (ON PANEL ASS'Y FLOOR)		
• NAMEPLATES - FRONT OF PANEL (WITH INSCRIPTION)		
• NAMEPLATES - BACK OF PANEL (ITEM NO. ONLY)		
AIR SWITCHES 3-WAY (INCLUDING NAMEPLATE)		
AIR SWITCHES 4-WAY (INCLUDING NAMEPLATE)		
ADDITIONAL BLANK PANEL PER LIN. FT. (PAINTED NO CUTOUTS)		

1. UNIT INSTALLATION PRICE FOR ALL ITEMS INCLUDES THE COMPLETE INSTALLATION, I.E., PANEL CUT-OUT MOUNTING, PIPING, WIRING AND/OR POWER SUPPLY, ENGINEERING, ETC., AND WHERE NOTED THUS: (•) INCLUDE THE COST OF THE INDICATED EQUIPMENT.
2. PRICES SET FORTH HEREIN, WILL REMAIN FIRM FOR THE DURATION OF THIS PROJECT ONLY.
3. BIDDER SHALL COMPLETE ONE SHEET FOR EACH TYPE OF CONTROL PANEL SPECIFIED, I.E., OUTDOOR-INDOOR, GENERAL PURPOSE – DIVISION 2.
4. EFFECT ON DELIVERY PROMISE TO BE ESTABLISHED AT THE TIME THE CHANGE IS MADE AND PURCHASER IS NOTIFIED.
5. NO DEVIATIONS ARE PERMITTED FROM THIS FORM WITHOUT THE PERMISSION OF M. W. KELLOGG. HOWEVER, A MULTIPLIER MAY BE ADDED TO ADJUST PRICES FOR VARIOUS STAGES OF CONSTRUCTION.

FIG. 4.2t

Quotation form, which the bidder is to complete and attach to the control panel bid package.

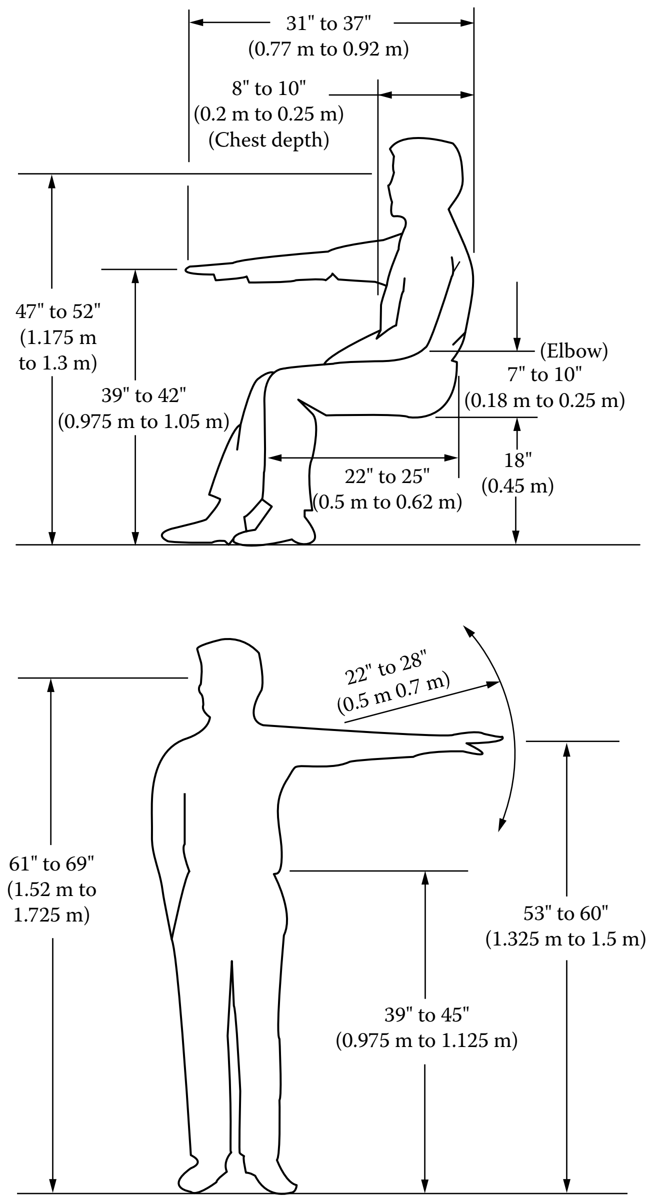


FIG. 4.2u
Anthropometric figures. The mid-range of these dimensions should cover the 95th to 5th percentiles of American operators.

2. Indicating scales, selector switches, and other instruments should push in or up, rotate clockwise, or slide to the right for the "increase" or "on" setting. The opposite position should result in "decrease" or "off" settings.
3. The style of emergency operating devices should reviewed. Emergency trips must be designed so that they are easy and fast to operate, but they should also be located and designed to prevent accidental trips.

Other less obvious ideas include the following:

1. *Annunciator horns.* These can be set to increase in pitch or to go on and off with increasing frequency as the emergency increases in severity.
2. *Annunciator lights.* These can be made to flash brighter and at a higher frequency.
3. *Instrument arrangement.* Instruments should be located in a panel or CRT display as though they were in a graphic panel, that is, following the process from left to right. They should be grouped in logical sets. For example, a distillation column instrument arrangement would have the column feed to the left, the overhead pressure instrument high, the reflux drum instruments high and to the right, and a reboiler instrument low and to the right. It is not possible to follow these recommendations exactly in all cases, but the devices should be located in these relative positions. Additionally, these instruments should be separated from the next column or other unit operation by a space or a line.

A key rule of human engineering is that the operator must be allowed to concentrate on running the plant. Distractions such as performing a complicated or confusing instrument procedure should not take the operator's attention away from this primary responsibility. The primary goal of the control center design is to make plant operation convenient.

PANEL TUBING AND WIRING

Tubing

Most panel manufacturers used to stock an acceptable line of industrial-grade tubing, pipe, and fittings. One of the most commonly used tubing materials behind the panel was copper. It is relatively resistant to corrosion, readily available, and rigid enough to require only a minimum of support, yet it is sufficiently ductile to bend to precise measurements. It is available with a tightly extruded PVC sheath for corrosive atmospheres, and with cadmium or tin plating for damp locations. Commonly used sizes are 1/4 in. (6.25 mm) OD with 0.030 in. (0.75 mm) wall thickness, ASTM B68.

Another commonly used tubing material is aluminum (1/4 in. [6.25 mm] OD with 0.032 in. [0.8 mm] wall thickness), federal specification WWT-700/4c. Aluminum tubing offers good resistance to chloride- or ammonia-containing atmospheres. It is somewhat softer to work with than copper and requires significantly more support. It has a tendency to harden, which makes it particularly susceptible to vibration failures. This tubing is also available with a PVC sheath.

The most frequently used plastic tubing material is polyethylene. This tubing should be of a material that has undergone environmental stress testing in accordance with ASTM D1693. Although polyethylene is markedly less expensive than metallic tubes, additional costs are incurred because it requires an extensive support network, which consists of plastic, slotted, or sheet metal ducting. The ducting should extend to within 1 or 2 inches of the connected instrument. Unsupported lengths are to be avoided because the plastic tubing has a tendency to kink and because it is almost impossible to

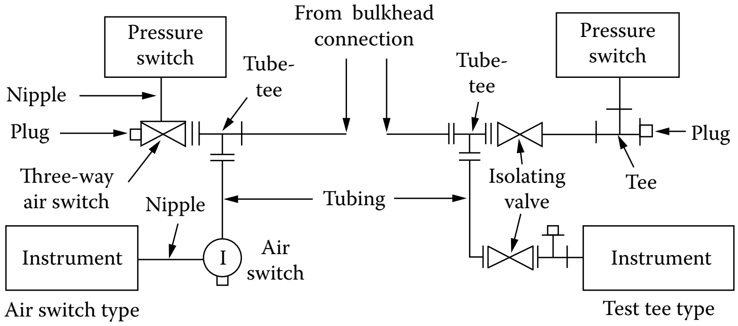


FIG. 4.2v
Test or calibration connections for pneumatic instruments.

run this soft tubing neatly. Tubing may be color coded to conform to the ISA recommended practice ISA RP-7.2. Common sizes are 1/4 in. (6.25 mm) OD, 0.040 in. (1 mm) wall thickness.

Other tubing materials are available, such as stainless steel, nylon, polyvinyl, rubber, and glass, but they are not used as frequently as the aforementioned three materials.

Fittings

Fittings are available for all tubing materials. For control panel applications standard compression-type tube fittings were used rather than flared, soldered, or other specialty types. Pipe material should be seamless red brass, threaded in accordance with USAS B.2.1 and soldered or brazed. The fittings should be rated to at least 125 PSIG (865 kPa). Air supply isolating valves should be either the straight-line needle type or the packless diaphragm two-way type.

The panel air supply should consist of two sets of parallel-piped reducing valves and filters. Each should be sized and manifolded so that either set can supply the total panel requirements. In general, 0.5 SCFM (0.014 m^3/m) of air flow per air user is sufficient. The main air header must be sized so that the pressure drop to the farthest instrument is under 1 PSID (6.9 kPa).

The use of a plugged tee and isolating valve or a three-way switch allows for calibration of instruments or components without disconnecting them or losing the measurement signal to other loop components (Figure 4.2v). The isolating valve need be used only when the measurement signal is branching to another instrument or when the instrument is not furnished with an integral spring-loaded air cutoff valve, as in plug-in type devices. The location of the instrument air supply tie-in to the board is usually noted on the panel drawings, as depicted in Figure 4.2w.

The pneumatic signal lead tie-in location usually requires only a general description since there is enough flexibility in the tubing installation for adjustment. Notes such as “top of panel” or “bottom of panel” are usually sufficient.

Each tubing termination for field leads should be identified with the tag number of the instrument it serves and its function. For example, FRC-10-V could indicate that the field lead terminates at the control calve of FRC-10, whereas FRC-10-T would mean that the air signal tube is received from the so-identified flow transmitter.

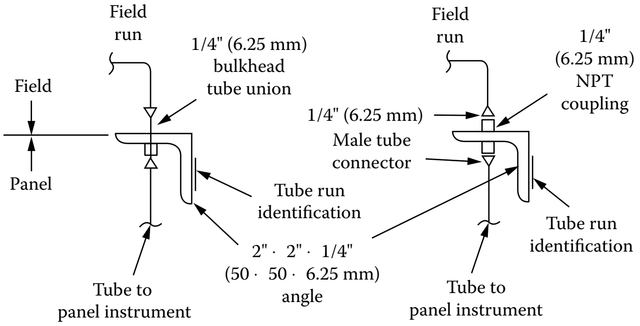


FIG. 4.2w
Typical field tie-ins to pneumatic panels.

There are several different types of interface connections on the panel, some of which are shown in Figure 4.2w. To allow for future changes or additions, at least 10% additional tie-in points should be provided.

Panel Wiring

The first step in deciding what equipment to use is to determine the nature and the degree of hazard. The National Electrical Code[*] describes hazardous locations by class, group, and division.

The class defines the physical form of the combustible material mixed with air:

Class I: Combustible material in the form of a gas or vapor
Class II: Combustible material in the form of a dust
Class III: Combustible material in the form of a fiber, such as textile flyings

The group subdivides the class:

---

[*] NFPA 70, ANSI/NFPA-70 National Electrical Code. Canadian equivalent: CSA Standard C22.1 Canadian Electrical Code Part 1.

Group A: Atmospheres containing acetylene
Group B: Atmospheres containing hydrogen, gases, or vapors of equivalent hazard, such as manufactured gas
Group C: Atmospheres containing ethyl ether vapors, ethylene, or cyclopropane
Group D: Atmospheres containing gasoline, hexane, naphta, benzene, butane, propane, alcohol, acetone, benzol, lacquer, solvent vapors, or natural gas
Group E: Atmospheres containing metal dust, including aluminum, magnesium, and their commercial alloys or other metals with similar hazardous characteristics
Group F: Atmospheres containing carbon black, coal, or coke dust
Group G: Atmospheres containing flour starch or grain dusts

The division defines the probability that an explosive mixture is present. For instance, a hazardous mixture is normally present in a Division 1 area but will be only accidentally present in a Division 2 area.

In addition to knowing the area classifications, one should also be aware of the National Electrical Manufacturers Association (NEMA) terminology for classifying equipment enclosures:

NEMA 1 General-purpose
NEMA 2 Drip-tight
NEMA 3 Weatherproof
NEMA 4 Watertight
NEMA 5 Dust-tight
NEMA 6 Submersible
NEMA 7 Hazardous (Class I, Groups A, B, C, or D)
NEMA 8 Hazardous (Class I, Groups A, B, C or D) — oil-immersed
NEMA 9 Hazardous (Class II, Groups E, F, or G)
NEMA 10 Explosion proof—Bureau of Mines
NEMA 11 Acid- and fume-resistant, oil-immersed
NEMA 12 Industrial

Most panels are enclosed by or parallel to a wall, with a door on either or both ends to limit unauthorized access. Under such conditions and when the area is general-purpose and nonhazardous (see Figure 4.2c), it is permitted to reduce the mechanical protection requirements for the wiring. Electronic transmission, power, and signal wiring need not be enclosed in conduit or in thin-walled metallic tubing.

All of the wiring may be run in a sheet metal or slotted plastic duct. The insulated wire may be run exposed, from the duct to the instrument, an inch or two (25 to 50 mm) without the necessity for a conduit nipple. Bare or exposed terminals, however, are not permitted.

Panels installed in hazardous or semihazardous areas must be installed in strict adherence to the National Electrical Code requirements. Because the code does not allow much flexibility in these cases, this discussion will be limited to the nonhazardous applications, in which the designer has some flexibility. For requirements in hazardous locations, refer to Section 7.2 in the *Process Measurement* volume of this handbook.

Power Distribution Instrument power supplies should be taken from a reliable source, with automatic switchover capability to an alternate power supply, which is to be used upon failure of the main source. The two typical standby power supplies are a separate supply bus (fed from batteries or from a different source) and a steam- or gasoline-powered generator. A detailed coverage of backup power supply systems is given in Section 4.23.

To reduce the cross-sectional area of the power feeders, it is often expedient to mount a three-phase (440-, 208-, and 120-volt) transformer directly in the control house. Then only a 440-volt power supply is provided from the switch gear to the transformer (Figure 4.2x).

A conventional lighting-type circuit breaker panel may be installed on the back of the panel to provide the necessary circuit distribution. This permits the panel manufacturer to install the complete system and significantly reduces field tie-in time. Breakers (sized to trip above 15 amperes) and AWG #14 wire gauge are generally used. Three-wire power circuits having hot, grounded neutral, and ground leads are frequently used.

To avoid the possibility of overloading, circuits should be lightly loaded to approximately half their rated capacity or to a maximum of 850 volt-amperes.

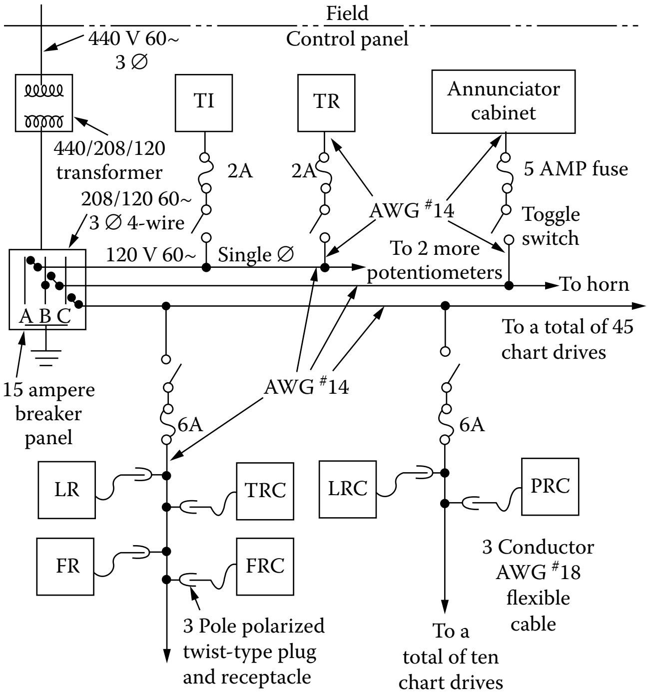


FIG. 4.2x
Typical power distribution system.

Fuses The following groups of instruments will keep the load of 15-ampere circuits within acceptable limits:

4 Potentiometer-type temperature instruments
1 Annunciator cabinet with horn
2 Analyzer circuits (900 volt-amperes maximum)
45 Miniature (pneumatic) recorder chart drives
5 Miscellaneous auxiliary components, at 800 volt-amperes maximum
10 Electronic instrument loops (500 volt-amperes maximum)
2 Emergency trip circuits

Secondary subcircuits must be used so that a short or ground at one instrument does not rip out the 15-ampere circuit breaker but only the associated fuse and so that each instrument (or group) can be isolated for maintenance or replacement. This isolation is accomplished with a fused disconnect device. The fuses must be coordinated so that it is not possible for that 15-ampere breaker to trip before a fuse blows. The fuse must also blow before a significant voltage dip occurs. An exception to the individual fusing isolation rule is chart drive power to pneumatic instrumentation. Here, ten drives (or some similar number) may be connected to one common fuse. A three-pronged, polarized, twist-type plug can serve as disconnect.

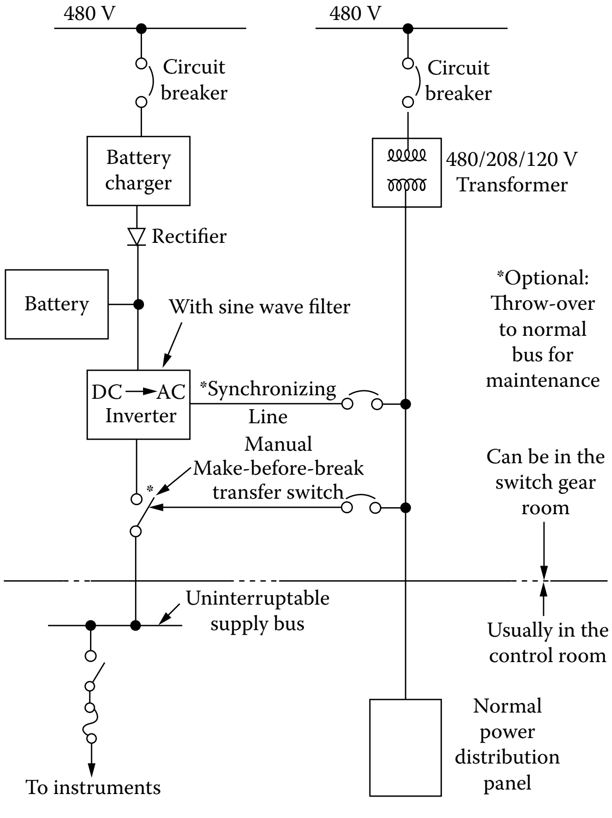


FIG. 4.2y
Battery backup type standby power supply system.

When instruments are internally fused, that fuse may be utilized, but the instrument circuits must be checked to verify that the fuse protects the complete chassis and not just a single critical component.

Battery Backup Several factors must be considered in deciding whether a standby power supply is required. These are as follows:

1. The power source that is used has a history of failures with duration of up to one half hour.
2. The unit uses normally energized solenoids or relays that must be manually reset.
3. There is a flame safety system.
4. The process is extremely fast-acting with electronic controllers in critical services, in which a dip in power supply could send the unit into uncontrollable cycling.
5. The process control system is computerized.

A typical uninterruptable power supply (UPS) is shown in Figure 4.2y, having a battery backup system. This system consists of a battery charger, a bank of batteries, and an inverter. One of the main advantages of this system is that the AC power input phase does not have to be synchronized with the output phase. The instruments are normally powered directly through this system. Upon failure of the mains, the batteries (which have been floating on the charging current) start feeding the inverter.

The battery ampere-hour capacity should be sized using one or more of the following considerations:

1. Length of time of average power outage × 1.5
2. Length of time plant will remain operable after power mains fail × 1.5
3. Length of time it takes to switch to alternative power supply × 2
4. Length of time the functioning of instruments will be required to bring about an orderly shutdown

The charger must be sized so that it can simultaneously operate the unit and recharge the battery system after a discharge. A recharge time of 8 hours is reasonable. Circuit breakers and fuses downstream of the inverter must be sized and coordinated so that the available current will trip them out before a significant voltage disturbance could occur.

Isolation Transformers Digital equipment is usually very susceptible to spikes in its power supply. The UPS system will filter out spikes from the main power source. If equipment other than the digital system is connected to the UPS bus, an isolating transformer should be used. A transformer that will filter a 50-volt peak and is furnished with a Faraday shield will often be adequate. The actual instrument manufacturer's recommendations must govern the final choice of specifications.

Grounding Each instrument case, control panel, auxiliary cabinet, and instrument system must be safely wired to the control house ground. Low-level instrumentation and DDC equipment require their own separate grounding net. A 1-ohm resistance to ground is usually a safe number. Here again, the individual equipment manufacturer's recommendations must be followed.

Wiring and Terminal Identifications Most analog electronic control loops are of the "two-wire" type. This means that the locally mounted transmitter or control element does not require a separate power supply but takes its actuating energy from the signal wires. Therefore, the installation requires only a two-conductor cable for each transmitter or final control element. To simplify field wiring and minimize overall field installation costs, auxiliary components are usually mounted on the back of the panel. In this way the complex interconnecting wiring is installed by the panel manufacturer.

For flexibility in making loop changes and additions, one method of installation is to use a centralized terminal block for each complete instrument loop on the back of the panel. In this system, each transmission and control loop is assigned a set number of terminals in the field tie-in junction box. Each group of terminals is identically marked, and the terminal marking strip carries the instrument loop number. Each component of the loop is then wired to this terminal, as shown in Figure 4.2z. Spare terminals should also be included for future instruments and components.

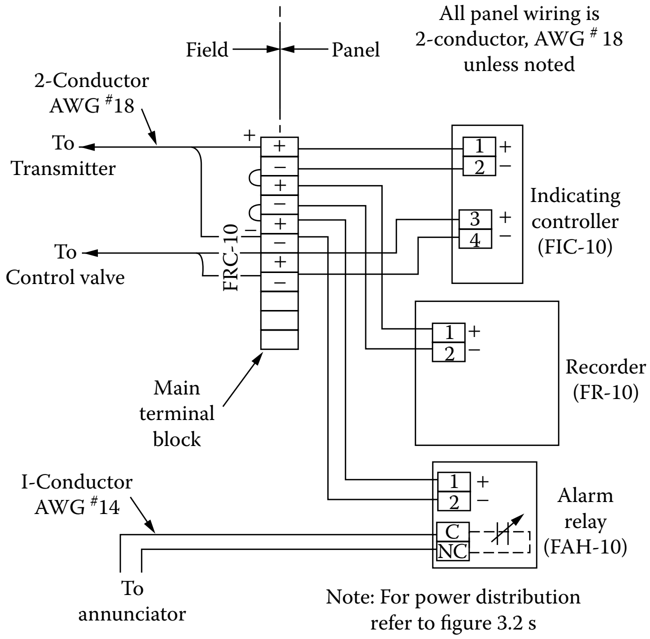
Terminal strips are available with white plastic or painted marking surfaces, suitable for penciling in identification. At all electronic instrument field tie-ins, the junction blocks should be identified with the instrument loop number, its function, and its polarity in the following manner; FRC-10-T(+) FRC-10-T(–). All other terminals should be identified similarly, except that they are marked with the instrument terminal designation instead of the polarity. For equipment such as relays, switches, and other components without terminal designations, the terminal should duplicate the identification shown in the wiring diagrams, such as SV-10-VS (SV-1) or annunciator A (S1-1).

Wire identification data should duplicate the information shown on the terminal marking strips: instrument item number, function, and polarity. Function and polarity identification may be replaced with color-coded wire insulation. To be effective, the color code must be simple and consistent. A typical color code recommendation is given in Table 4.2aa.

In addition to the color coding, the wire may be identified with the instrument item number by means of a preprinted marker tape or by a plastic sleeve. The sleeves either are sized to a snug fit over the wire insulation or are of the shrink type.

Wire terminations should be made with crush-type wire lugs. Special lugs are available for solid wire. A good lug to use is flanged spade lug. This type combines the ease of installation of a spade lug with flanges that hold the lug in place if the terminal screw becomes loose.

Where wires are run within a duct, it need not be laced. Although it gives a neater appearance, lacing is time-consuming and is a nuisance when wires must be frequently added or rerouted.



Note: For power distribution refer to figure 3.2 s

FIG. 4.2z
Typical wiring diagram for an electronic instrument loop.

TABLE 4.2aa
Wiring Color Code Used on Control Panels

Color of Wire Insulation	*AC Service (120-volt, 60-cycle Supply Using AWG #14 wires)*	*DC Service (Low Voltage Using 2-conductor AWG #18 wires)*
Black	"A" phase — hot	Positive power, transmission or control signal
Red	"B" phase — hot	Negative power or control
Blue	"C" phase — hot	—
White	Neutral	Negative transmission signal
Brown	Annunciator common (II)	Annunciator signal
Orange	Annunciator signal	Annunciator signal
Green	Ground	Ground
Gray	Miscellaneous interconnections and jumpers	

CONTROL CENTER INSPECTION

A complete panel inspection at the manufacturer's plant will usually pay dividends in ease of installation, field tie-ins, loop checkouts, and in a smoother plant start-up with fewer field man-hours expended.

The control center designer, familiar with the overall instrumentation and operating philosophy of the unit, can visually inspect and functionally check a control panel most expeditiously. Together with a pipe-fitter and an electrician, the designer can locate piping errors or wiring mis-terminations in a fraction of the time that it would take in the field.

Each panel has its own peculiarities. The following examples may be useful as a guide for formulating a checklist for inspection:

1. Panel construction dimensions should be evaluated, for example:
   a. Overall dimensions
   b. Thickness of panel
   c. Size of framing
2. Construction materials to be assessed include:
   a. Panel and framing material
   b. Panel finish (e.g., smooth, unblemished, correct color)
   c. Piping materials (copper, brass, PVC, etc.): sizes, correct valves, fittings, and so forth
   d. Wiring, proper wire gauge, type, and insulation
   e. Hardware (acceptable industrial grade, rated equal to or better than service requirements)
3. A properly designed panel will have the following construction features:
   a. The finish and appearance of the overall panel will be workmanlike.
   b. All instruments and equipment will be properly aligned.
   c. Tubing, piping, and wiring will be neatly laid out and adequately supported and will not interfere with instrument maintenance.
   d. All equipment will be rigidly mounted.
   e. The back-of-panel auxiliaries and miscellaneous hardware will be properly identified by item numbers.
   f. All field tie-ins will be identified.
4. The following rules regarding instrumentation should also be followed:
   a. All instruments should be installed in their proper location on the panel.
   b. Correct instruments should be furnished and installed including the charts, scales, model or type numbers, and instrument nameplate inscriptions.
5. Preliminary checks are performed as follows:
   a. Power distribution is checked using these steps:
      i. Verify that no one is working on the panel.
      ii. Check that the panel is securely grounded.
      iii. Put all disconnect switches and circuit breakers in the "off" position.
      iv. Pull all polarized plugs.
      v. With a high resistance (light or other) across the input terminals, energize the panel.
      vi. If the light dims, find and remove the ground or short circuit.
      vii. Energize each circuit and check each power supply subcircuit sequentially. Deenergize the circuit after checking and before energizing the next one.
      viii. Check for proper voltage.
   b. Air supply is checked using the following steps:
      i. Close all instrument air supply isolating valves.
      ii. Close reducing station gate valves.
      iii. Connect clean, dry air supply to the panel at the specified pressure.
      iv. Open reducing station blocks and check downstream air pressure of each reducing station (set at 20 PSIG, or 138 kPa).
      v. Blow down the filters and the header drain valves.
      vi. Increase the header pressure until the relief valve pops.
      vii. Individual air supplies may be checked as each loop is operated.
      viii. Bubble-test the main air header connections for leakage.
6. Functional tests may include the following examinations, which should be performed in the order listed:
   a. Assessment of pneumatic instruments:
      i. Simulate the input signal at the bulkhead fitting with a 3- to 15-PSIG (21- to 104-kPa) regulator.
      ii. Attach a 0- to 30-PSIG (0- to 207-kPa) gauge to controlled output at the bulkhead.
      iii. Turn on the air supply.
      iv. Verify that the bulkhead and air supply nameplates are correctly inscribed.
      v. Vary the input signal and watch the output gauge for proper response.
   b. Evaluation of alarm and 120-volt control circuits:
      i. "Jumper" the input terminals one at a time to verify that the correct annunciator light flashes. The horn may be disconnected after the first alarm checkout.
      ii. Energize the relay circuits on the panel by simulating the input signals.
      iii. Connect a pilot light to the output terminals of outgoing signals actuating remote solenoid valve or relay. Be sure to verify output voltage so that pilot lamp will match.

   - iv. As each item is checked, verify the tagging of equipment and of field tie-ins.
   - v. Energize all chart drives, place mark on roller, and check after 1 hour for movement.
- c. Assessment of analog electronic instruments:
   - i. Energize the loop for checking and deenergize when checked.
   - ii. Simulate the input signals at field tie-in points. Check input signal type, level, and voltage. This is particularly important for special instruments.
   - iii. Put proper resistance across output terminals.
   - iv. As each instrument is checked, verify the identification of equipment and of field tie-in terminals.

DCS or other digital instrumentation is checked as follows:

1. Simulate the proper field input signal at the I/O cabinet.
2. Add resistance across output terminals at the I/O cabinet.
3. Check the CRT for proper response.
4. Verify that all incoming square roots have been converted to linear signals within the distributed system.
5. Verify that all control algorithms and PID settings are correct.
6. Check the groupings to ensure that each loop appears in the proper place.
7. As each loop is checked, verify that the I/O terminals are correct.

If the inspector is unable to stay at the shop and verify that all mis-terminations and errors have been corrected prior to panel shipment, then a “punch list” is prepared. One copy is left with the panel manufacturer, another is kept as a record, and a third copy is sent to the field so that the panel can be checked upon arrival at the job site.

After a proper panel inspection there should be no difficulty in hooking up the field tie-ins. Any problems in the loop checkout and calibration will most likely be external to the panel. This will significantly ease the troubleshooting in the field.

PANEL SHIPMENT

A panel should be handled as little as possible because the chance of damage is much higher during loading and unloading and when the device is in motion on the carrier.

When the panel is to be shipped by truck and installed immediately upon arrival at the job site, only skids with a light framework holding a tarpulin are necessary. To save time and handling, the panel should travel via an air ride van. The van should be “exclusive,” that is, reserved for transportation of the panel. This will ensure a direct route to the plant, without stopovers at trucking terminals, and will reduce handling. The van should also be furnished with a removable top so that the panel can be lifted out.

If the panel cannot be installed immediately and must be stored at the plant site, heavier crating is required and a thicker plastic sheeting should be used. Because time is not critical and the panel is better protected, an exclusive van need not be used in this case. Shipment by train, although less expensive for long distances, requires additional handling and moving. Some trains are also severely jostled during make-up and routing.

When shipped by boat, the panel should be sent as below-decks cargo. The panel crating must be especially heavy and must be cushioned within the case. The wrapping should effectively seal out the salt air. Prior to sealing, the voids inside the wrapping should be liberally loaded with a dessicant, such as silica gel. Heavy, impregnated, water-resistant paper or 5-mil-thick polyethylene can be used for wrapping. All seams should be covered with waterproof tape. When possible, smaller shipping units should be used to ease handling. All panel equipment must be securely braced.

Air freight does not require any particular crating or wrapping other than that required for a nonexclusive van. The particular airline must be contacted and questioned regarding weight and overall size limitations for each panel and crate. The plant site airport may also be checked to verify that it is capable of receiving the type of airplane required for the shipment.

CONCLUSIONS

Control center designs need not be limited to the basic examples discussed in this section. There is no limit to the number of design variations. Each center may be formulated of new and different shapes specifically conceived and adapted to its own unique application.

The multiplicity of design parameters is such that drawings and specifications cannot cover every particular feature. The only reasonable way to ensure the development of the exact control panel desired is by close cooperation of the panel manufacturer and the panel user.

Reference

1. Farmer, E., “Design Tips for Modern Control Rooms,” *Instruments and Control Systems,* March 1980.

Bibliography

Callisen, F. I., “Control Rooms of the Future,” *Chemical Engineering,* June 2, 1969.

Collier, S. G., and Green, M., “Verification and Validation of Human Factors Issues in Control Room Designs and Upgrades (HWR-598),” Halden, Norway: OECD Halden Reactor Project, 1999.

“Control Center Design,” ISA Publications RP 60.1 through 60.11, Research Triangle Park, NC: ISA, Farmer, E., "Design Tips for Modern Control Rooms," *Instruments and Control Systems,* March 1980.

"Flow Diagram Graphic Symbols for Distributed Control/Shared Display Instrumentation Logic and Computer Systems," ISA Draft Standard dS5.3, Instrument Society of America, Research Triangle Park, NC, 1982.

Gadberry B. E., "Designing Integrated Control System Displays," 1991 ISA Conference, Paper #91-0354, Anaheim, CA, October 1991.

Green, M., et al., "A Method for Reviewing Human Factors in Control Centre Design," Oslo: Norwegian Petroleum Directorate, 2000.

"Human Factors Affecting the Reliability and Safety of LNG Facilities," Volumes 1 and 2, Final Report, Gas Research Institute, January–September 1982.

"Human Factors Evaluation of Control Room Design and Operators Performance at Three Mile Island — 2," U.S. Department of Commerce *NUREG/CR-2107,* Volumes 1, 2, and 3, January 1980.

"Human Factors Review of Nuclear Power Plant Control Room Design," ERPI NP-309, Palo Alto, California, March 1977.

ISO, "Ergonomic Design of Control Centres–Part 1: Principles for the Design of Control Centres," ISO 11064-1:2000, Brussels: ISO, 2000.

Kwok, T., "Advanced Workstations as DCS Operator Consoles," 1992 ISA Conference, Houston, October 1992.

Military Standard, "Human Engineering Design Criteria for Military Systems, Equipment and Facilities," MIL-STD 14728, U.S. Department of Defense, 1974 and later addenda.

"Pneumatic Control Circuit Pressure Test," ISA Publication RP 7.1, Pittsburgh, PA: ISA, 1956.

Thompson, B. J., "Preparing for Computer Control," *Instruments and Control Systems,* February 1980.

Wells, S. E., "Organizing Tie-ins for Modernization/Expansion," ISA/94 Technical Conference, Anaheim, CA, 1994.

4.3 Control Center Upgrading

B. J. GEDDES (2002)*

Cost:	Workstation costs range from $10,000 to $30,000 depending on the number of flat panel displays, video display units, computers, and input devices. Where new screens replace analog instruments in a conventional panel, cost is typically $5,000 per screen. The above costs do not include hardware/software integration.
Partial List of Suppliers:	ABB Instrumentation (www.abb.com/us/instrumentation) Barco Projection Systems (www.barco.com) Clarity Visual Systems (www.clarityvisual.com) Emerson Process Management (www.easydeltav.com) Framatome ANP (www.framatome.com) General Electric (www.ge.com) Invensys Process Systems (www.invensysips.com) Raytheon ELCON Digital Display Group (www.ifp-beisch.com) Rockwell Automation (www.rockwellautomation.com) Siemens Energy & Automation Inc. (www.sea.siemens.com/ia) Westinghouse Electric Co. (www.westinghouse.com)

As process plants age, owners must consider a variety of asset life cycle management issues, including instrument and control (I&C) system obsolescence. Plants constructed prior to the early 1970s were predominantly supplied with analog instruments and controls, either electronic or pneumatic.

It is likely that the vendors who supplied these original analog systems no longer provide parts or services to keep them running. As computer chips and information systems became available through the advent of the information age, instrument and control system vendors found innovative ways to apply this digital technology across their product lines. Systems and components are now smaller, more reliable, and fault tolerant, and they can perform multiple functions that were previously spread across multiple, discrete components.

Today, process engineers are presented with a variety of digital components and systems ranging from transmitters, switches, relays, breakers, controllers, recorders, indicators, video displays, distributed control systems, fieldbuses, and networks to connect them all together. The control system vendor may supply any or all of these new components and systems as replacements for the original equipment.

Digital information and control technology offers a variety of entry points in the analog process plant. The entire plant can be upgraded at once in one outage, or the owner and designer can choose to upgrade one system or component at a time. The technology is flexible enough to consider any upgrade path, which almost always has an effect on the control room.

The designer has two basic choices:

1. Building a New Control Room—If the plant is still operating, but has excessive cost or obsolescence problems in the I&C system or in the control room, then a new control center (building) may be economically justified. A new control center will be built as close as possible to the old control room or control center, so operations and maintenance tasks that require proximity to plant equipment can be maintained. This option also has the advantage of allowing construction to proceed while the plant is operating.
2. Upgrading the Control Room as Required (Phased Approach)—If the plant can still meet its business goals safely and effectively, and obsolescence is not yet a significant safety, reliability, or cost issue, then the designer can select an upgrade path for the plant on a system-by-system basis that accounts for time-to-obsolescence. System upgrades can be performed while the plant is in production, or during routine maintenance outages. Over time, the control room will have the same basic layout as that supplied during original construction, but the panels and furniture will be configured with modern technology where it is needed. Original indications and controls that are not obsolete and can still be easily maintained may remain as-is, but they will be reintegrated in a hybrid

---

[*] This section been reprinted with minor additions and changes, from Volume 3 of the *Instrument Engineers' Handbook*, which is titled: "Process Software and Digital Networks."

analog/digital design with modern human machine interfaces.

Designers must consider the impact on the control room when applying new technology in any part of the plant. Even a seemingly innocuous plant change can impact the control room. For example, a digital, fieldbus-compatible process transmitter may offer a local upgrade path for a given plant system. But this transmitter now supplies much more information than the original 4–20 mA DC transmitter, and it should be applied in the context of a total plant architecture that will impact the control room as it evolves.

It is critical to see beyond the immediate needs of the plant. Plant owners will naturally tend to apply funding and resources to the I&C systems that cause the most trouble and stop until the next acute need arises. Acute system problems usually have a significant impact on plant safety or production, and the designer owes a more proactive and comprehensive approach to the owners.

BASELINE EVALUATION

The goal of the baseline evaluation is to ensure the upgraded control room still functions well within the original plant. The functional allocations between humans and machines should be well understood and documented. Operations and maintenance tasks should be understood and documented, and information zones should be mapped and correlated with these tasks. More importantly, the baseline will also ensure that the control room operating staff's sense of how the plant is designed and operated will be maintained with the modified or new control room.

The designer must first understand how and why the original control room is designed the way it is, then understand the basic operating philosophy of the control room staff. The operating philosophy can be defined as the set of rules and procedures matched with the set of roles and responsibilities of operating staff resources that are applied during all modes of operation, including normal, abnormal, and emergency operations. Operating staff (control room supervisors, control room operators, and system operators) roles and responsibilities in an existing plant are based on the functional allocations between humans and machines, which then determine the tasks assigned to humans in all phases of plant operations.

The baseline evaluation should begin with an evaluation of the basic design of the existing plant information and control systems by reviewing loop diagrams, schematics, wiring diagrams, and panel layout/cutout drawings. Figure 4.3a represents an architectural perspective for the analog I&C systems and components in a typical older process plant. Each set of systems and components can be categorized by its basic function.

The cloud at the bottom of Figure 4.3a comprises the basic sensors such as pressure, temperature, flow, level, and

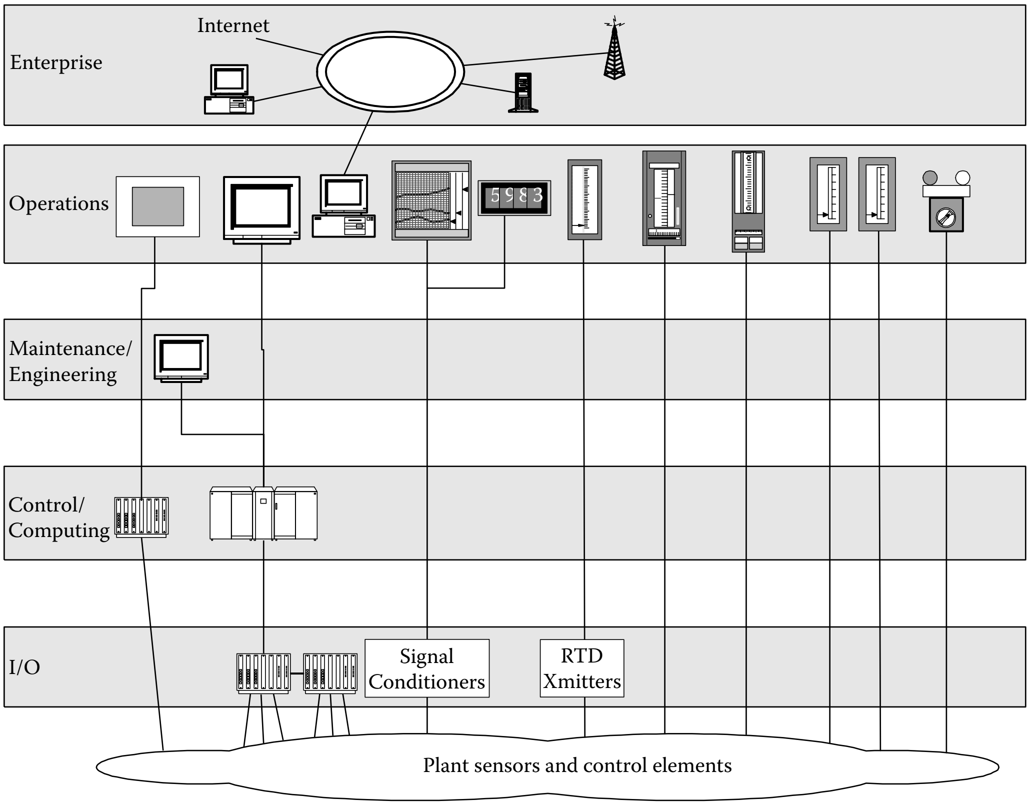


FIG. 4.3a
Analog I&C architecture.

position switches, various measurement elements, and analog process transmitters. It also represents final control elements such as motor control centers, motor-operated valves, air-operated valves, dampers, and machines. This population of sensors and control elements usually numbers in the thousands for large process plants.

The input/output (I/O) layer just above the cloud at the bottom of Figure 4.3a represents the basic I/O systems that may be used by what few digital platforms are employed, such as a mainframe plant process computer.

The mainframe computer, in the control/computing layer, typically runs a real-time database program for those plant variables that need to be stored and retrieved in group displays or reports, or used as input data for complex performance calculations. The number of data points assigned to the plant process computer may number in the thousands, but its role is limited to information-only applications.

The maintenance layer represents any interfaces made available to maintenance and engineering. The only interfaces that are available to maintenance and engineering in the analog plant are confined to the one or two computers that may be used in specialized applications.

The operations layer represents the devices used by the plant operator, and it can be seen that the majority of the devices in Figure 4.3a are in this layer. These devices include trend recorders, digital indicators, vertical and horizontal meters, a variety of hand switches, single-loop controllers, lamps, and annunciator light boxes. By inspection, most of the analog plant sensors in Figure 4.3a terminate directly on these devices used by the operators. Any maintenance required on these analog loops will usually take the affected devices out of service and result in a negative impact on the operation of the plant due to disruptions and loss of information and control.

Finally, the enterprise layer illustrates the basic plant or company enterprise where information technology (IT) systems are deployed. Operators may have a company PC at their disposal in the control room for retrieving enterprise information such as drawings or accounting data, or sending and receiving e-mails. This PC is only connected to the office local-area network (LAN) and is not used directly to operate or maintain the plant.

Changing any component in the lower layers of Figure 4.3a can impact the operations layer, and vice versa. For example, the plant may need to upgrade a set of plant sensors due to obsolescence. The signals from the old sensors were processed by the mainframe computer. The new sensors require a new software code to properly acquire and convert their signals to the engineering units required by the operator. The new software cannot run on the old mainframe computer, so a new computing platform is needed in the form of a server, which comes with a client software package for use on a PC or workstation in the control room. The server can also be integrated with a DCS (distributed control system). Now we have a complete system upgrade that affects every functional layer in Figure 4.3a, with significant control room impact.

FIG. 4.3b
Analog control room.

ORIGINAL DESIGN PERSPECTIVE

Figure 4.3b shows an older control room, typical of power plants and other process plants designed and built between the late 1960s and the early 1980s. The majority of the indications and controls are from original construction and are almost all analog instruments. Some plant systems have been upgraded since original construction, but only on an as-needed, system-specific basis. No serious consideration has been given yet to a total control room modernization effort.

The operator interfaces are typically located on “stand-up” control panels and are used by control room operators during normal and emergency operations. Desks are provided for senior control room operators and control room supervisors, with telephone and radio equipment and terminals connected to a mainframe plant process computer. The panels are laid out by major plant functions such as electrical systems, cooling water systems, major machine controls (e.g., turbine/generators or compressors), reactor controls, and emergency systems.

On each panel, indications and controls are grouped by specific plant systems or functions. For example, Figure 4.3c shows a feedwater control system panel from the same control room shown in Figure 4.3b. Annunciator windows are on the top plane of the panel. Feedwater heater level indications are grouped on the left-hand side of the vertical plane. Steam generator temperature and level indications are grouped on the right-hand side. Across the bottom planes are indications and controls associated with the control elements for feed pump turbines, condensate pumps, and various valve position indications and controllers. Groups of instruments and controls are bordered or separated by a thick black demarcation line called “zone banding.”

Each group of instruments and controls can be associated with a set of operator tasks for each mode of plant operation (start-up, normal, shutdown, and emergency operations).

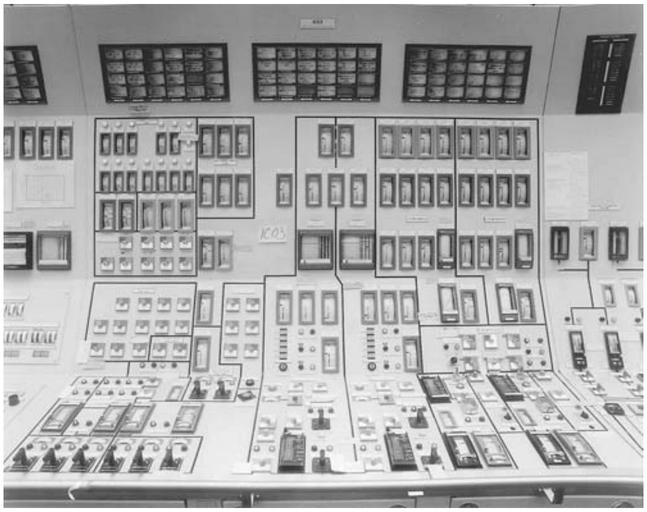


FIG. 4.3c
Analog control panel.

Written procedures describe operator actions required for responses to various alarms or during various controlled transients. Operators are trained in a full-scale simulator for various operating modes and equipment failure scenarios, and they are trained to perform periodic visual scans, or surveillances, of each group of indications and controls either close up or from a “total control room” perspective. The baseline evaluation should capture all of these operator tasks and their associations with each group of indications and controls.

INFORMATION ZONES AND TASK ANALYSIS

Perhaps the most significant design aspect of an older, analog control room is that all plant indications and controls are in discrete locations on the panels. Most plant variables are only readable from one instrument in a “fixed position” on a panel, and plant system variables are grouped in “information zones” where operators are trained to look for patterns in the indications to detect, diagnose, and maintain or mitigate various plant conditions. The information zone human–machine interface (HMI) design concept is a very strong paradigm for operators, and should not be significantly modified or corrupted unless a new control room is built, complete with a new task analysis, operating procedures, and operator retraining and qualification program. In the phased approach, information zones are kept intact as each zone is upgraded over time.

Each information zone is associated with a set of operator tasks, which should be studied and evaluated using simulators or mockups. Tasks will vary with modes of plant operation. Most operator tasks in a continuous process plant will be associated with surveillance of the process systems. During routine surveillances, an operator will visually scan the control panels looking for patterns in the indications and alarms that indicate plant systems are operating within specifications. Some manual manipulations of multichannel meters or displays may be required.

In an upset mode, such as a sensor or equipment failure, or a trip of a major plant machine or the plant itself, an operator’s task loading will shift dramatically, and will be focused on key plant variables and controls. Alarms will be prioritized by the operator per procedures and training plans, and the event will be diagnosed and mitigated through the cognitive strengths of the human operator. The HMI will be designed to accommodate this shift in task loading with an emphasis on decision making.

Once the task evaluation is captured, operating experience should be evaluated. Operating logs and event reports should be reviewed for positive or negative trends in plant or human performance. Operators should be interviewed to see if there are any HMI issues that need to be improved or corrected. Are there any recurring situations where operators are tasked with too many simultaneous inputs requiring concurrent actions? How do operators prioritize their actions under these scenarios? Are there any operator tasks where there is a deficiency in available information, such as resolution, timeliness, or readability? Are there any indications and controls that are too difficult to read or understand, or are simply not available under certain conditions where they are required for an operator task?

If an operating experience review indicates deficiencies, event scenario walkthroughs should be conducted in a full-scale simulator if one is available. If a full-scale simulator is not available, a full-scale mockup of the control room should be built. At first, a mockup of the old control room might seem like a waste of time, but it can support an engineering analysis of the strengths and weaknesses of the old control room without disturbing plant operations, and it can support experimentation and evaluation of new design concepts as they are integrated in the original control room if the phased approach is used.

Figure 4.3d shows a 1/8 scale mockup of the same control system panel shown in Figure 4.3c. It was made by reviewing photographs and layout and cutout drawings of the panel, then making a scale template of each plane on a personal computer using Visio™, a commercial graphics software package. Objects for each type of control panel device (recorders, indicators, controllers, etc.) were made and stored in a Visio stencil, then applied to the scale template using a “drag and drop” technique. Each panel was printed on a color printer or plotter, spray-glued to a piece of cardboard, then cut out and assembled using a hot glue gun. The total cost of the software and materials for this mockup was less than $100.

The engineering and technician time invested in building the object library, or stencil, and preparing this first panel was over 100 hours, but subsequent panels take less time and the library is now available for finishing a mockup of the entire control room. This mockup was prepared at full scale in the computer, then printed at 1/8 scale for this version. The same Visio files can be printed at full scale on

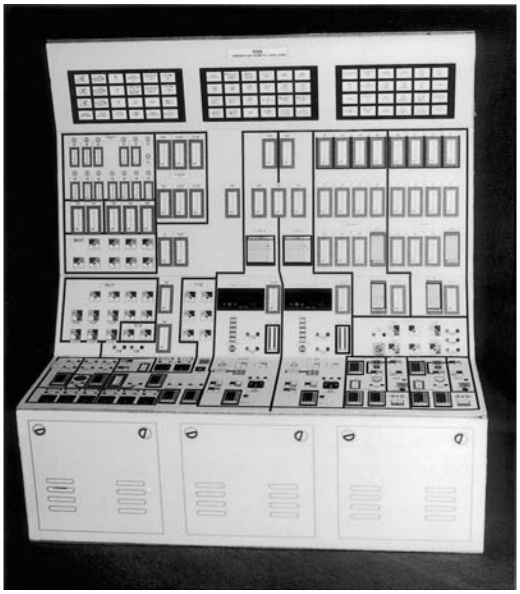


FIG. 4.3d
Panel mockup.

an engineering plotter, and applied to a full-scale plywood mockup, which is useful for walkthroughs with operators.

Finally, the baseline evaluation should examine all of the operating and maintenance costs associated with the original I&C systems in the plant. If an older, analog plant is expected to have a long operating life, many if not all of the I&C systems will need to be replaced. These replacement activities will be necessary to maintain continued plant operation, to take advantage of current technology, to improve plant availability, and to facilitate further reductions in operating and maintenance costs.

The most significant economic driver is usually obsolescence, which can be changed from a liability to an asset in considering control room upgrades. When obsolescence is considered in aggregate, a strong economic case for an integrated, modern control room can be made. A system should be considered obsolete when any of the following criteria are met.

1. The system will not meet the operational demands required of it during all phases of plant operation, including normal, abnormal, offline, online, and emergency situations.
2. The parts manufacturer has formally announced that it will no longer directly support the system or component, or will require a significant increase in component cost to continue support.
3. The system or component will not interface with new components that are required in other systems. It is virtually assured that if the operational life of a process plant is extended, all system components will require some form of major component replacement.
4. The system or component cannot support plant operations that are desired by the business plans.

Once the baseline evaluation is completed, the plant owner and the designer will have an accurate assessment of the functional, performance, and economic state of the plant I&C systems and, in particular, the control room. Goals and objectives can be clearly described, and upgrade options can be estimated and undergo a total cost–benefit analysis.

BUILDING A NEW CONTROL ROOM

Building a new control room all at once allows the full potential of modern I&C systems to be immediately realized in an older process plant. A new control room in this context means the complete retirement and replacement of the original analog panels, I&C systems and HMI components with new digital systems, using a DCS or PLC (programmable logic controller) platform. The replacement systems may be installed in the original room, or may be installed in a new building or otherwise spare plant location.

If the plant can no longer meet its business goals safely and effectively because of control system degradation and obsolescence, it probably needs a major overhaul, and an entirely new control room may be in order. If a plant is being overhauled for a major capital upgrade (new reactors, boilers, compressors, turbine generators, or other major machinery), it will probably include a new process or machine control system. The designer can engineer a new control house and demolish or retire the old one in place, or gut and rebuild the existing control room. The new control room should be implemented with a workstation/console arrangement designed to maximize safety, quality, reliability, and productivity with a heavy emphasis on human engineering.

The decision to build an entirely new control room all at once has profound implications for the entire process plant. The designer can demolish the old control room by removing all components and structures, including panels and furniture, leaving floor and ceiling slabs and four bare walls, with field cables pulled back or left coiled near their penetrations.

Using this approach, an analog control room is replaced with a fundamentally new “glass control room,” or one that consists of video display units (VDU) located on operator workstations or consoles and can also include one or more large displays. The control system cabinets will typically be located in a computer room or cable spreading room adjacent, above, or below the new control room, with data communications interfaces serving the consoles and workstations.

The ISO 11064 standards (Parts 1 through 8) offer a sound methodology in the design of a new control room. Their emphasis is on the design of a new control room in the context of a new plant, but they also describe design processes that can be used in designing a new control room as a retrofit in an existing facility.

Using a standard design methodology, as shown in Figure 4.3e, will improve safety, quality, and reliability when applying modern digital system technology. The roles and responsibilities of operators and machines can be modified

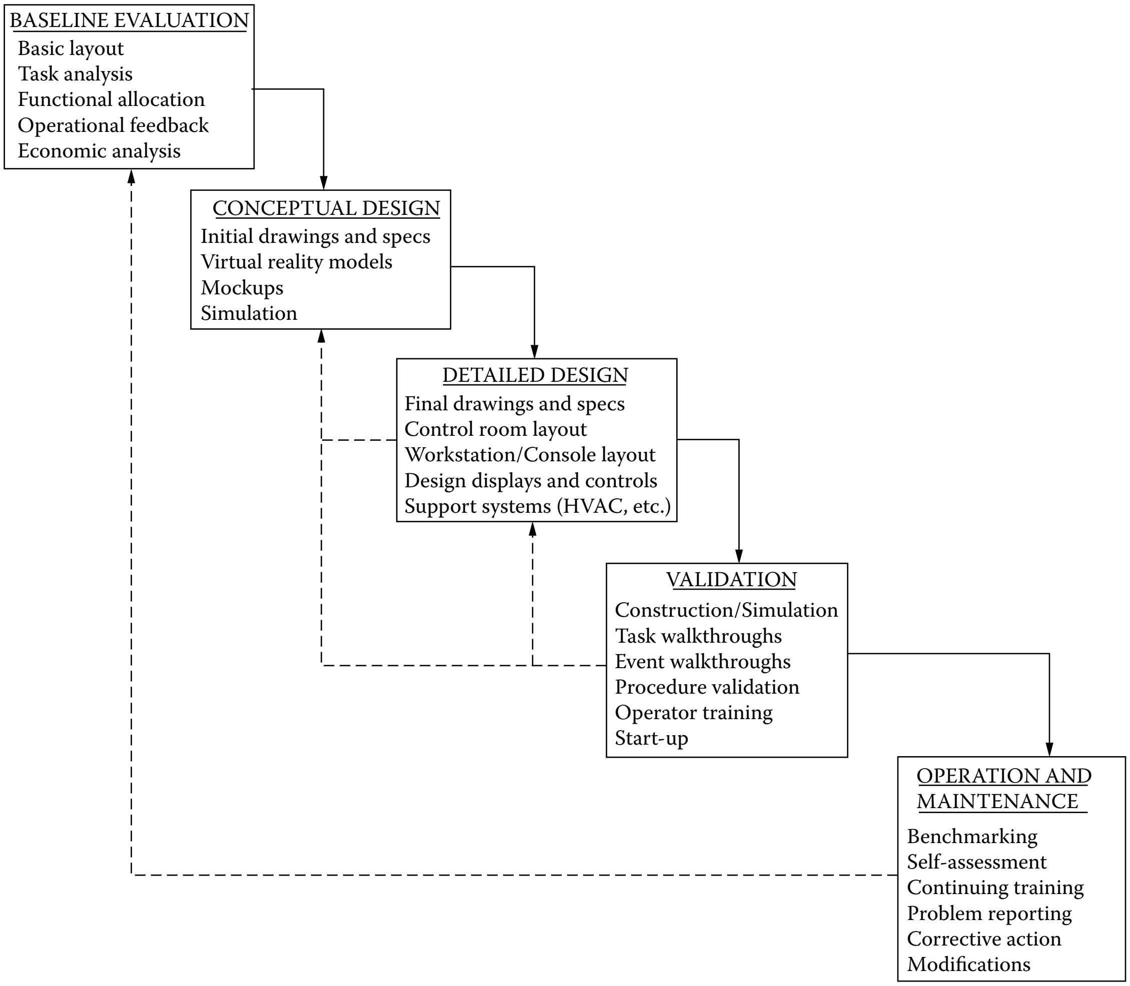


FIG. 4.3e
General design process for new control rooms.

and tasks and functions reallocated to achieve a higher performance standard. An increased reliance on automation can significantly alter the role of operators by reducing their task to vigilance over a constant process, while also greatly expanding their role during system upsets to one of corrective action through cognition, diagnosis, and appropriate manual responses. If a new control room is under consideration, particular attention has to be paid to ensuring that any changes in roles and responsibilities of machines and operators from the old control room to the new one continue to meet or exceed the goals and objectives outlined in the baseline evaluation. Without a careful design process, unintended functions can be created with potentially negative consequences.

Basic ergonomic principles should be applied in the design of control rooms. Ergonomic design principles include:

- Use a human-centered design approach. Design for basic physical limitations, but also emphasize human cognitive strengths such as perceptual, problem-solving, and decision-making abilities.
- Improve design through iteration. Repeat design evaluations until functional and performance goals are achieved.
- Start with the baseline evaluation performed on the original control room, and revise through each design iteration.
- Design error-tolerant systems. Allow for human error, and provide interlocks, alarms, and overrides where appropriate.
- Ensure user participation. The most knowledgeable asset on the design team can be the control room operator. Add system and process engineers, ergonomists, architects, and industrial designers to form an interdisciplinary design team.

TECHNOLOGY ADVANCES

Technology advancements in recent years have enabled a state-of-the-art approach to control room design. Compact workstations, large display panels, soft controls, fixed and selectable displays, advanced alarm systems, and computerized procedure systems greatly enhance safety, reliability, and efficiency.

A modern control room includes compact workstations, a safety console for plants that require one, and furniture and rest areas. A large display panel (LDP) can also be provided. Key features that support the ability of the operating crew to maintain efficient and safe plant operation include:

- Full-function workstations supporting direct plant control and monitoring by one or more system operators
- An identical workstation supporting normal monitoring and crew coordination functions of a control room supervisor and serving as a backup to the operator workstations
- An LDP providing overall plant operational and safety assessment
- A safety console (for those industries with safety requirements) providing control capability for all safety-related components independent of the workstations supporting safe plant shutdown even in the event of complete workstation failure

Advantages of this control room layout, shown in Figure 4.3f, include enhanced communication between operators, operational facilities for all expected crew members, good visibility of the LDP, ease of accommodating design and job allocation changes, and convenient access and egress routes. Layouts will vary depending on the number of systems in the process plant that are monitored and controlled in a main control room. If a site has multiple process plants that are independent islands, but still controlled from one central control room, then the control room layout may employ multiple sets of compact workstations, depending on operator and supervisory tasking.

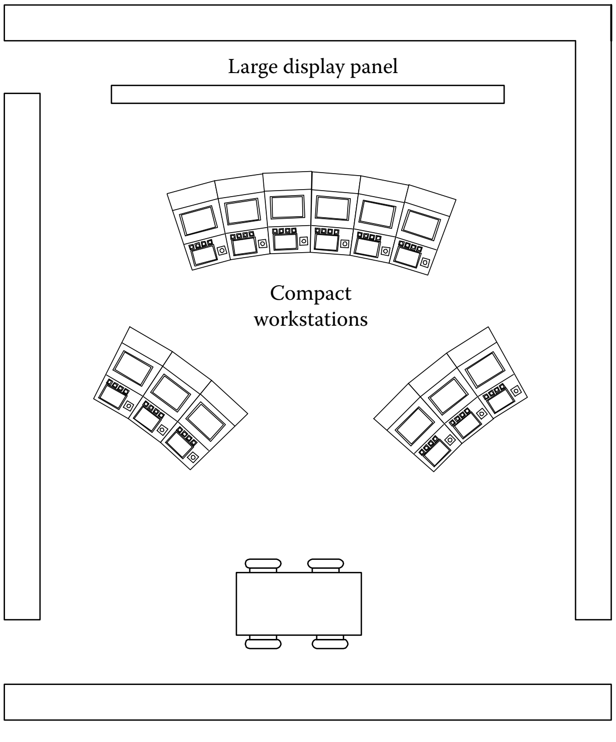


FIG. 4.3f
Sample control room layout.

The LDP is a wall-mounted overview display, which includes static and dynamic elements. High-resolution, tiled projection systems are available today with seamless borders between tiles. The fixed display section of the LDP provides continuous, parallel display of key alarm, component, system, and parameter information. This complements the workstation HMI with a spatially dedicated graphical depiction of the plant. A variable display section allows operators to display pertinent information selectively to support crew coordination. A sample LDP design is shown in Figure 4.3g.

Each main control room workstation provides devices for access to all information and controls necessary for one person to monitor and control all processes associated with plant operation and safety. This includes both safety and nonsafety systems. A sample workstation layout is illustrated in Figure 4.3h. Each workstation contains the following:

- One alarm VDU with trackball user interface
- Multiple VDUs supporting process monitoring or electronic procedures with trackball user interface
- Multiple flat panel displays used as soft controllers for process and component control; each works in conjunction with one VDU, using a touch-sensitive user interface
- Dedicated, diverse push buttons for manual safety system actuation
- Laydown area for logs, drawings, backup paper procedures, etc.

The major advantages of the compact workstation approach are its (1) operational and design flexibility, (2) compactness and simplicity, (3) ability to accommodate changes cost-effectively, and (4) provision of an enhanced integrated environment for a computerized procedure system (CPS) and operator aids.

Workstations should allow simultaneous access to plant information through selectable displays on multiple VDUs per workstation. A wide variety of display formats should support system mimics, major plant functions or conditions, technical data sheets, trends and graphical information, and application program access. All are designed to support specific operator functions. Multiple methods should be provided for convenient access to the display set, including navigational access through menus, direct access through format chaining from other displays (or alarms and procedures), and a dedicated mechanism such as function buttons or voice entry. A major function of the VDU displays is to provide a soft control link allowing the operator quickly to select a

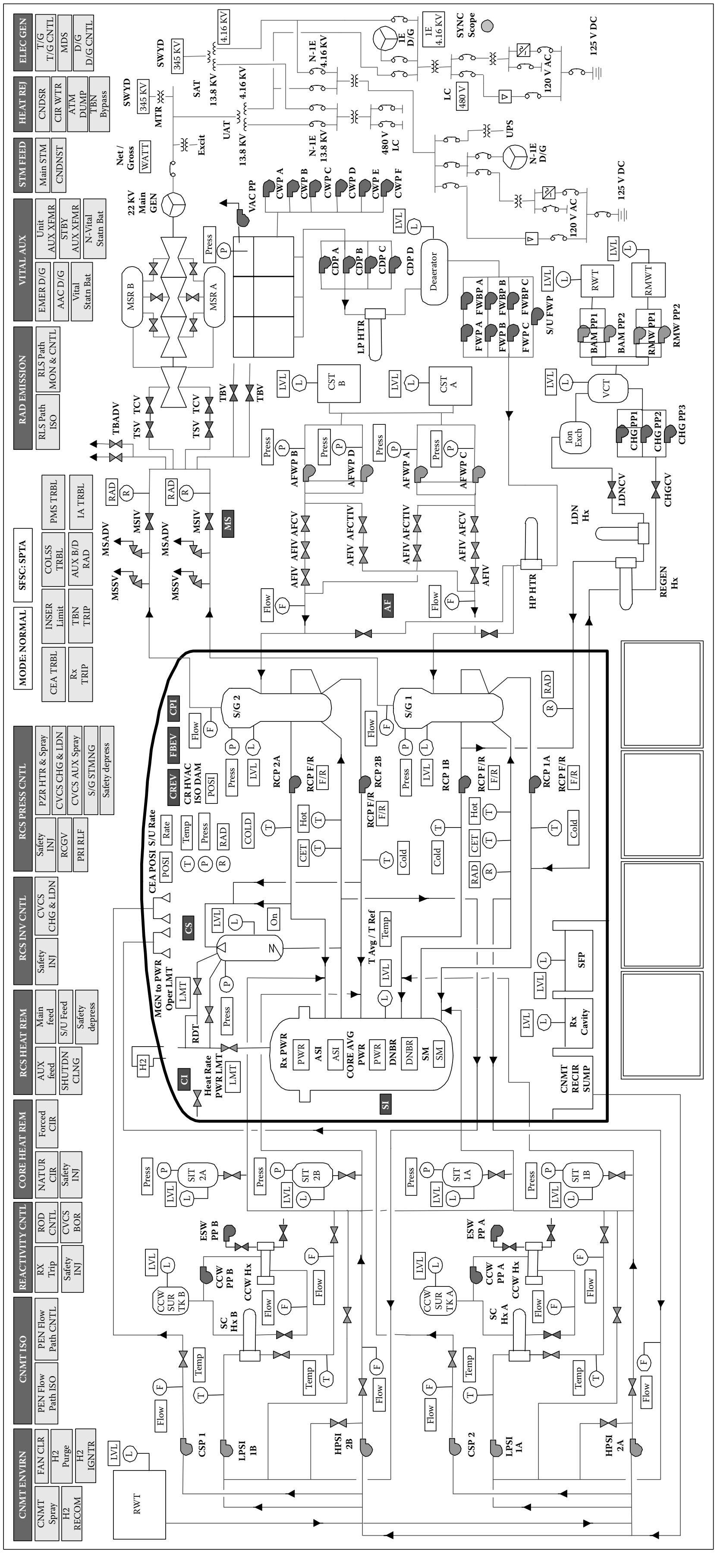


FIG. 4.3g
Large display panel. (Courtesy of Westinghouse–CENP & KEPRI.)

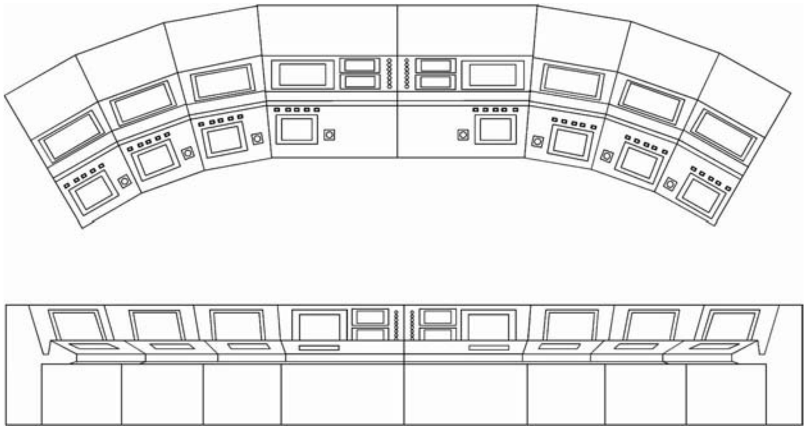


FIG. 4.3h
Typical workstation design. (Courtesy of Westinghouse–CENP & KEPRI.)

component or process control on the soft controllers directly from display pages.

"Soft controls" should utilize flat panel displays to emulate the physical switches and manual/auto stations that populate conventional plant control panels. Use of software-based control allows a standard interface device to assume the role of numerous physical devices. This has the advantage of allowing operator access to all plant controls from a single workstation, design flexibility, the ability to accommodate changes easily, and simplification of hardware procurement and maintenance. Various design concepts can support control of multiple safety and nonsafety divisions from the same workstation device, while maintaining the single-failure-proof reliability of conventional channel separation and independence to support safety system requirements such as those called out in IEEE-603 (1998) "Standard Criteria for Safety Systems for Nuclear Power Generating Stations" and ISA 84.01 (1996) "Application of Safety Instrumented Systems for the Process Industries."

An advanced alarm system can be provided to improve the annunciation process, by incorporating methodologies that:

- Reduce the total number of alarms with which the operator must cope
- Distinguish between true process deviations and sensor failures
- Minimize the occurrence of "nuisance" alarms
- Prioritize the relative importance of alarms so the operator can focus on the most critical alarm conditions first while deferring less critical alarm conditions
- Determine the impact of alarms on plant operations and distinguish these from lower-level system alarms

Highest priority alarms, such as those for critical functions, are presented in fixed locations on the LDP. All alarms are presented in list form on a dedicated workstation VDU as well as through relevant locations in the VDU display hierarchy.

COMPUTERIZED PROCEDURE SYSTEMS

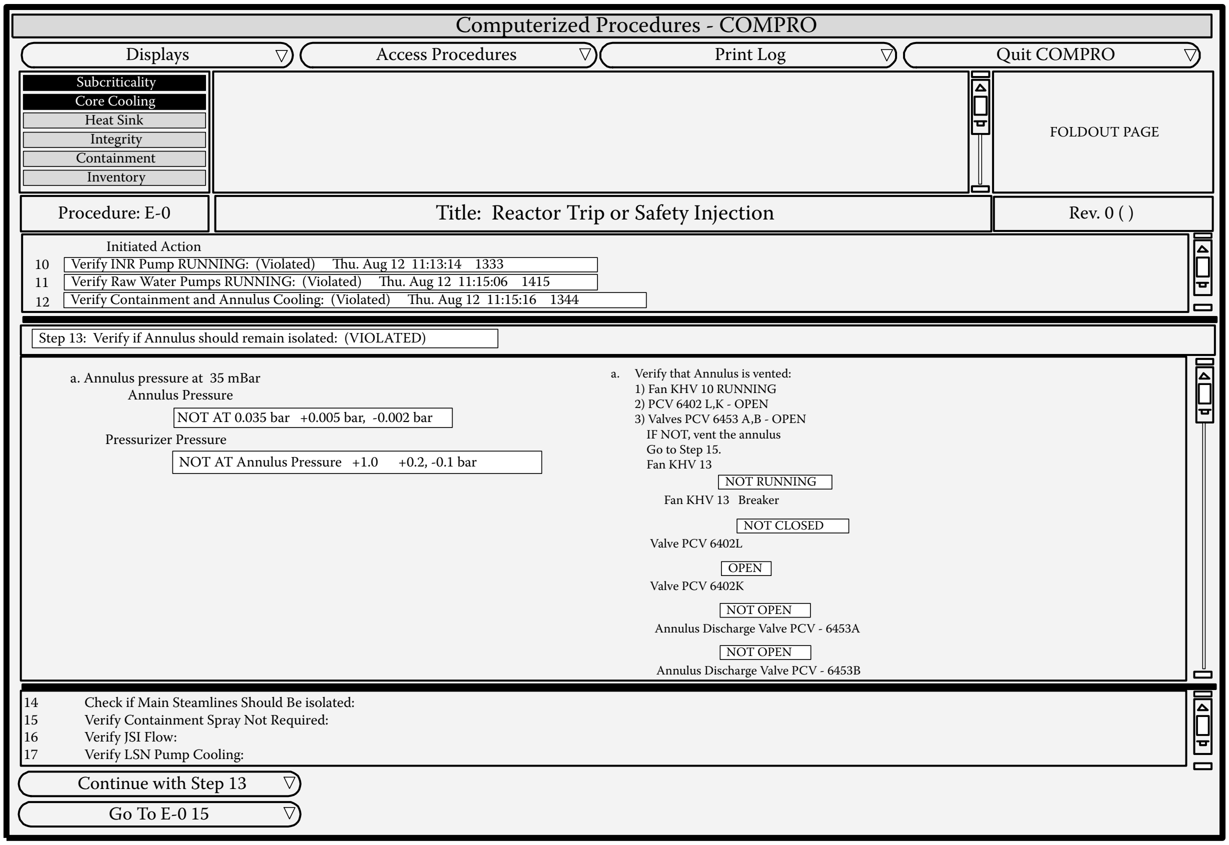
Computerized procedure systems (CPSs) are another significant technology development in modern control rooms, which take advantage of state-of-the-art technology to provide benefits in the human–machine and control room integration areas (Figure 4.3i).

The purposes of an online, data-driven, cathode ray tube (CRT)-based CPS are the following:

- To guide the user step by step through the procedures by monitoring the appropriate plant data, by processing the data and by identifying the recommended course of action.
- To provide the parallel information necessary to allow the user to assess other plant conditions that may require attention, such as, for example, notes, cautions, and foldout page items.

The computer monitors and evaluates large amounts of information quickly and efficiently. Procedure information is online and is updated essentially continuously. Hence, the operator becomes more vigilant because a large amount of procedurally required information is immediately available. The operator's mental loading is reduced because all required procedural information is displayed, including, for example, the status of the current high level step.

The human user's primary role in a CPS environment is to monitor the progression through the plant procedures while maintaining a clear picture of plant state, to take control actions on the control board when they are required and to watch for unsafe plant conditions. The user should retain both authority and responsibility for plant operation. A CPS is user paced, that is, the system should not advance to a procedure step, a note, caution, or foldout page item, or a procedure unless instructed to do so by the user.



Procedure: E-0

Title: Reactor Trip or Safety Injection

Rev. 0 ( )

Initiated Action

10 Verify INR Pump RUNNING: (Violated) Thu. Aug 12 11:13:14 1333
11 Verify Raw Water Pumps RUNNING: (Violated) Thu. Aug 12 11:15:06 1415
12 Verify Containment and Annulus Cooling: (Violated) Thu. Aug 12 11:15:16 1344

Step 13: Verify if Annulus should remain isolated: (VIOLATED)

a. Annulus pressure at 35 mBar
Annulus Pressure
NOT AT 0.035 bar +0.005 bar, -0.002 bar
Pressurizer Pressure
NOT AT Annulus Pressure +1.0 +0.2, -0.1 bar

a. Verify that Annulus is vented:
1) Fan KHV 10 RUNNING
2) PCV 6402 L,K - OPEN
3) Valves PCV 6453 A,B - OPEN
IF NOT, vent the annulus
Go to Step 15.
Fan KHV 13
NOT RUNNING
Fan KHV 13 Breaker
NOT CLOSED
Valve PCV 6402L
OPEN
Valve PCV 6402K
NOT OPEN
Annulus Discharge Valve PCV - 6453A
NOT OPEN
Annulus Discharge Valve PCV - 6453B

14 Check if Main Steamlines Should Be isolated:
15 Verify Containment Spray Not Required:
16 Verify JSI Flow:
17 Verify LSN Pump Cooling:

FIG. 4.3i
Computerized procedure system. (Courtesy of Westinghouse–CENP.)

A CPS can operate on a two-CRT workstation such that the required procedural information always appears on the first screen and supporting or supplementary procedural information appears on the second screen. Access to the supporting information should be through pull-down menus located along the top of the main screen.

The benefits of a CPS include:

- The computer and the operator complement each other for a more accurate implementation of the procedures, resulting in enhanced situation assessment by the operator.
- The system simultaneously monitors multiple plant parameters.
- The system brings all procedural information to one location.
- The system provides detailed record-keeping capability of the procedure execution.

PHASED APPROACH

A phased approach to upgrading the control room may be economically attractive. In this approach, it is important to recognize whatever design and operating constraints may be in place. The control room to be upgraded is assumed to be in service, and should continue to meet the business goals set by the owner.

The entire control room can be upgraded in phases, over time, according to whatever drivers are selected. The scope and schedule of each phase is optimized when it is consistent with the overall cost–benefit analysis of the entire upgrade, which may take years. If obsolescence issues are not acute, but are slowly accumulating, and the plant is running safely and meeting business objectives, then upgrades can be applied to individual systems or groups of systems while the plant is running or during routine maintenance outages.

If a phased approach is used, then a long-range strategic I&C modernization plan should be developed before any system or component upgrades can significantly impact the control room. Failure to do so will not have any impact on a particular system upgrade, but in the overall scheme, it can increase the total cost of the finished control room. The long-range plan will provide a conceptual design for the finished plant and control room, and it will provide a scope, schedule, and budget for each phase so that they are scheduled in the most cost-beneficial sequence. Each phase will account for the set of systems and components that should be upgraded, taking into account time-to-obsolescence and grouping of systems and components so duplication or rework on later phases is avoided. Obsolescence can be successfully applied in arguing for a comprehensive upgrade plan. Not all the original analog I&C systems will become acutely obsolete at the same time.

In the original analog control room, a unique treatment of indications and controls for each plant system was the norm. Panels were typically designed to support collections of meters, recorders, switches, and controllers grouped by plant system. It is important to avoid unique upgrades for each plant system. If unique digital solutions are applied over time, the result will be a muddled set of DCS, PLC, and other operator interfaces in the control room. All the operational goals of the plant upgrades will not only go unfulfilled but will worsen when realized in total.

Figure 4.3j illustrates the modernized plant. The functional layers are still the same, but there is a significant change in the technology applied in each layer. Figure 4.3j can be considered a hybrid digital/analog plant. The left side of this figure shows digital systems; the right side shows the original analog components. As more plant systems are upgraded, more of Figure 4.3j becomes digital. Hardware and software diversity may be applied for safety systems in each layer where required by regulatory agencies.

The key point in Figure 4.3j is that new I&C systems are expandable horizontally in this perspective, using a building-block approach that applies DCS or PLC technology linked by network communications technology, such as Ethernet, fieldbus, or any other standard that is supported by the manufacturer. This point is essential for implementing a cost-effective phased control room upgrade, where systems are expanded until obsolescence issues are resolved and modernization goals are realized.

The I/O layer employs subracks that can be field-mounted or mounted in cabinets or panels in the control room. The location of the new subracks depends on a number of factors, such as communication constraints (distance, bandwidth, number of allowable nodes) and the impact on cable costs. To optimize cable costs, the best place to mount I/O subracks is within reach of the field cables that were originally terminated on or near the old analog devices.

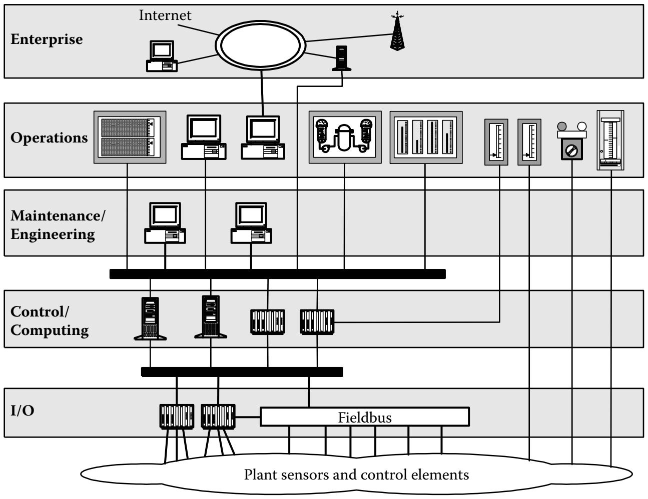


FIG. 4.3j
Hybrid I&C architecture.

The control/computing layer employs DCS or PLC processors or a combination of both. Equipment selection should allow data archiving and computational servers to be added to this functional layer and integrated into the whole control and computing platform system. Hardware and software integration can be expensive and time-consuming, so it should be consistent with one set of standards.

Note that original control room meters can still be used with a new DCS/PLC platform using the new I/O subsystems. This design feature is useful when an entire process system is upgraded, where video display units cannot replace every system switch, light, meter, or annunciator. For example, a small-scale DCS can replace a turbine/generator electrohydraulic control (EHC) system with one or two main VDUs in the place of the original operator interface and still allow connections to original devices in other locations. If all original EHC interface devices in this example are scheduled for replacement with VDUs, the potential exists for other plant systems to be caught up in the EHC upgrade simply because the new VDUs take up more panel space than the original meters and switches.

If any special applications are required, such as a data archiving agent or any plant performance calculation packages, plan ahead and select a DCS/PLC hardware and software platform that can support the special application hardware and software. Pay close attention to database integration issues, because a central database is the strength of a DCS. If an original plant data acquisition system can be connected to a new DCS and the data integrated into the central database, then it will be possible to display and archive the data seamlessly, transparent to the operator, without a new front end for those data points.

Network communications is utilized between all layers. The network topology is optimized for cost and connectivity options, and the first phase of the modernization effort should be carefully designed so that future phases are accounted for with at least 100% spare connectivity. Spare connectivity should support copper and fiber media options. When pulling new fiber-optic cables, select termination locations that support the long-range plan, and pull as many fibers as possible in each cable run (limited by budget or raceway fill).

The operations layer in Figure 4.3j is the most constrained layer in the phased upgrade approach. This layer is constrained by the physical layout and fundamental design of the existing control room. For example, the breakfront panels shown in Figure 4.3b are designed for a standing operator, and indications and controls are laid out by groups associated with major plant system functions. HMI resources such as flat panel displays have to be designed so that the basic information zones of the control panels can remain intact. If a particular information zone contains vertical meters and trend recorders, then flat panel displays should be sized to fill that zone (with redundancy if possible), and the DCS or PLC system should be

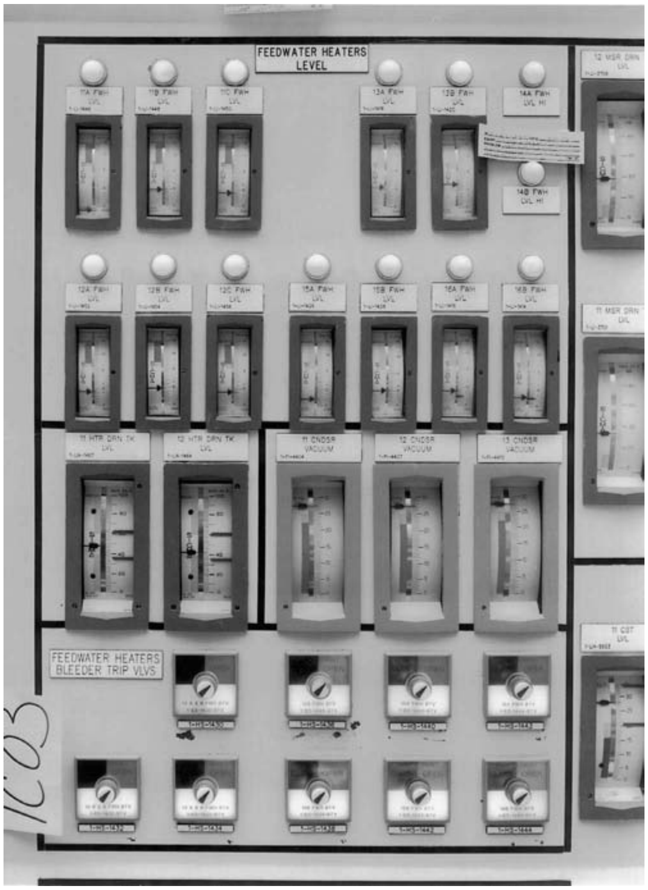


FIG. 4.3k
Feedwater heater indicators and controls.

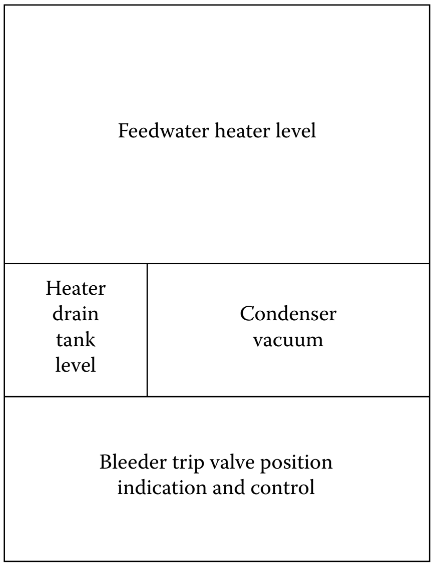


FIG. 4.3l
Feedwater heater information zone.

programmed with a display page that carries the same information with the same readability of the replaced components.

Figure 4.3k is a good example of some information zones on the control room panel shown in Figure 4.3c. Shown here are four zones. The zone on top is a group of feedwater heater level indicators. The middle-left zone is a pair of heater drain tank level indicators while the middle-right zone consists of three condenser vacuum indicators. The bottom zone is a set of indicating hand switches that indicate and control valve position. All of these instruments are connected to obsolete transmitters in the field, which are being upgraded using fieldbus technology. New controllers are being installed, and an integrated display can be connected to the system for use by operators in the control room. The original indicators in Figure 4.3k can be maintained as-is using analog and digital I/O points connected to the controller I/O subrack, but this upgrade provides a point of entry into the control room for a modern HMI interface, which is inevitable given the I&C obsolescence issues in the plant. The feedwater heater system is a good pilot project for this technology, where its information zone remains intact using flat panel displays.

Note the size of the meters in Figure 4.3k. The character size, color, luminosity, and arrangement are designed to allow readability from 15 ft away, where the control room supervisor sits. This is a firm constraint for the system, because operator tasking is designed to allow one control room supervisor to stand back and take in all of the control panel indications simultaneously from a central location in the control room, while an operator can stand close to these indicators and manipulate them manually.

The information zone "map" for these instruments is shown in Figure 4.3l. The information is of relatively low density. These parameters are used by control room operators during normal operations, including start-up and shutdown. The operators' task during normal operations is to monitor these variables during routine watch-standing operations. Their task during alarm conditions is to diagnose the cause of the alarm and verify automatic control actions have taken place and/or to take manual action as required by written procedures, so that the system parameters are returned within normal operating limits.

Figure 4.3m shows the same information zone in a modern display example, which can also be applied in a mockup (any scale). This figure shows two flat panel displays, designed for installation in the place of the original instruments. The concept is to program each flat panel with a "default page" that arranges the same variables in a similar fashion to the original arrangement, with font sizes as large as or larger than the characters on the original meters. Because this is a redundant design, each flat panel can be set to display any other page in the display hierarchy.

This design maintains the same readability and operator tasking as the original design, while offering powerful new interfaces to the operator. The control room supervisor can continue to stand back and take in all indications during normal and emergency operations, while the control room operator can manually page through the various display pages

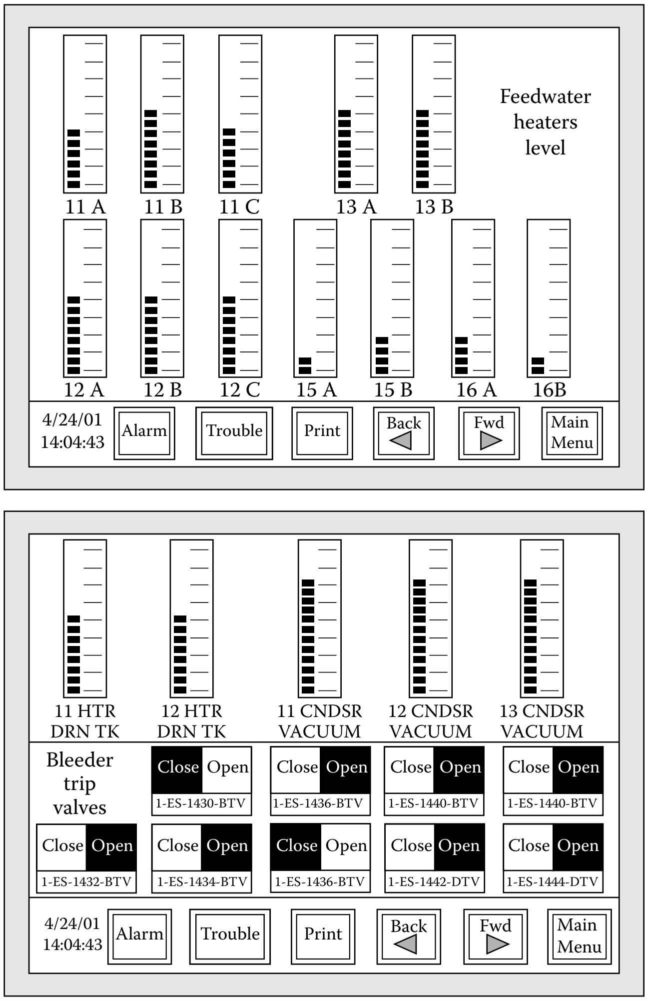


FIG. 4.3m
Upgraded feedwater heater displays.

available on the flat panel using a touch or trackball interface. Also, applying DCS technology for this interface supports additional flat panels or VDUs in other parts of the control room, such as an auxiliary panel or on the control room supervisor's desk. Although the default page is designed to be simple and readable consistent with the original panel design and operator tasking, the system also supports a full range of object-oriented graphic displays, event logs, archives, alarm management, and testing provisions.

Maintaining the original information zone design concept is only the start of any control room upgrade effort. Improvements in safety, reliability, readability, and a reduction in the potential for operator error should also be goals of the project. The use of a digital DCS or PLC platform enables the realization of these additional goals. Variations in colors, fonts, reverse video, and flashing objects are available with digital technology. New systems can detect an alarm on a failed channel (out of range high or low). Also, technology developments such as automatic alarm prioritization and computerized procedures may be implemented in a new system such as this one.

The feedwater heater displays in the above example are a starting point for the entire panel in Figure 4.3c. As more channels of information are added to the DCS/PLC system, more displays can be added to the same panel, or larger displays can be installed with more information available per display.

PLANNING AHEAD

The first phase in a phased control room upgrade project should prove that the overall conceptual design for the control room can be implemented safely and effectively. Once it is proven that the upgraded system has a positive impact on control room operations with improved reliability and cost-effectiveness, then the system can be expanded to all systems in the control room. The pilot approach allows training of engineers, technicians, and operators to commence without a massive undertaking, and valuable lessons learned can be applied to all disciplines on future phases (design, procurement, start-up, and operations and maintenance). The key is to purchase a digital I&C platform that can be applied on a smaller-scale project then expanded into the rest of the I&C systems in the plant.

HMI STANDARDS

It is essential that the long-range plan account for the impact on the operator, and that an HMI standard be applied consistently from start to finish. Flat panels, video display units, workstations, and consoles should be uniformly designed, and software should be standardized so that all system indications and controls are seamless and transparent in a common interface style. If an operator has to move to a specific piece of HMI hardware for each plant system or point-and-click on different icons for each plant system, then the opportunity to improve safety and reliability while reducing operating costs is diminished. The only exception to this rule is when diverse indications and actuations are required, such as for manual safety system actuations (e.g., reactor trip).

HMI attributes such as color codes, abbreviations, fonts, graphic elements, display sizes, and readability should all be consistent. A human-factors engineering (HFE) program should develop standards and procedures that will be coupled with an operator and maintenance technician training program, consistent with the long-range plan.

While the goals of a control room upgrade are to eliminate obsolescence, reduce maintenance and operating costs, and improve safety and reliability, the basic control room panel and furniture layouts and operating and maintenance procedures will be constrained throughout the phased approach. However, the control room panels and furniture can still be altered in a local sense. It is acceptable to introduce different HMI methods in clearly distinct contexts. For example, it is acceptable to add modern soft controls

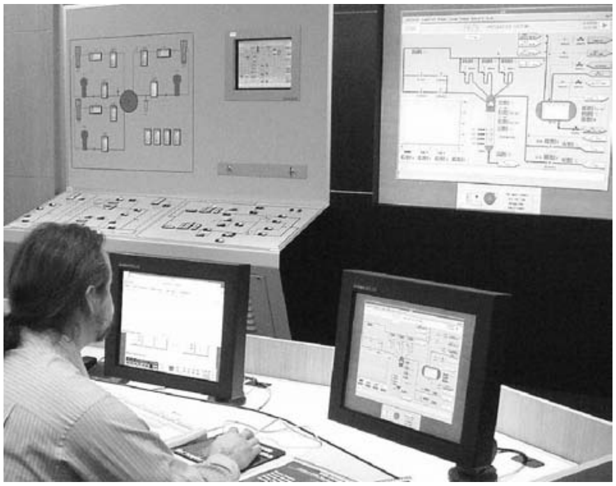


FIG. 4.3n
Demonstration control room. (Courtesy of Westinghouse Electric Company.)

techniques into conventional control rooms for distinct operations such as water treatment. As operators become familiar with the new HMI, the designer can be confident they will accept transitioning to the modern HMI for more critical systems. At any point in this transition process, it is important to ensure that operators are within the same HMI context (conventional or soft) for any particular plant system.

Figure 4.3n illustrates design concepts for panels, furniture, and a large display that can be evaluated for use in any analog control room. Panels can be modified and still mix manual controls, analog meters, and flat panel displays. Furniture can be modified to mount flat panels on arms that can be manipulated to suit operator tasks. For example, if an operator needs a line of sight to the control panels during routine operations, then the flat panel can swing out of the way. If an alarm requires some diagnosis, then the operator can swing the flat panel back into view and perform a specific task, or use it in combination with the large display panel in a supervisory role.

Gross design problems in the control room should not be applied as constraints in the phased approach. For example, if a critical process variable is indicated on one panel and controlled on another in a way that does not allow the operator to observe and control the process simultaneously within acceptable HFE practice, even if the operator has learned to work around the problem, then the control and indication should be colocated in the new interface. This is a good example of how a VDU in each information zone can resolve old design issues.

MAINTAINABILITY

Improved maintainability of the plant and control systems in particular should be a significant goal of a control room upgrade. Information and control systems should be purchased with diagnostic aids and functions that support the ability to store short-term, local data logs and the ability to off-load data to archival systems. In addition to normal process information, there should be pages of information related to process events (warnings and alarms), diagnostic events (system and component faults), and maintenance information. This information should enable an operator to detect and mitigate an event immediately, and call the maintenance department to investigate, troubleshoot, and perform corrective actions if necessary. Corrective actions should be more timely and effective on the control system itself as well as on plant equipment.

Networks allow multiple video displays in and around the control room. When an analog plant was constrained to field- or panel-mounted indications and controls, the only data available to the maintenance or engineering department were real-time data in the control room, or sparse, historical data by way of manual operator logs. Intermittent problems were usually very difficult to repeat, diagnose, and correct. A control room upgrade should fill this void in the maintenance and engineering departments. Displays should be set aside for these departments in or adjacent to the control room, with secure access. Control system access should be controlled first by physical security, then by password security.

Maintenance and engineering displays should provide enough data to understand a plant process event or a control system event. Digital inputs and outputs should be captured in a sequence of events (SOE) log. Maintenance and engineering displays should also allow users to navigate all available information and settings on the system, including the ability to change settings, start and stop programs or tasks, and run system diagnostics in an offline mode. Modes should be controlled by password and/or key switch inputs to the system.

Displays on the main control panels should be limited to information and controls that are related only to normal and emergency plant processes. Physical security should limit access to operator displays to qualified plant operators only.

Bibliography

Harmon, D. and Shin, Y. C., "Joint Design Development of the KNGR Man-Machine Interface," Taejon, Korea: KEPRI & Combustion Engineering, 1998.

IEEE-603, "Standard Criteria for Safety Systems for Nuclear Power Generating Stations," Piscataway, NJ: IEEE Press, 1998.

IEEE 1289, "IEEE Guide for the Application of Human Factors Engineering in the Design of Computer-Based Monitoring and Control Displays for Nuclear Power Generating Stations," Piscataway, NJ: IEEE Press, 1998.

ISA 84.01, "Application of Safety Instrumented Systems for the Process Industries," Research Triangle Park, NC: Instrument Society of America, 1996.

ISO 11064-1, "Ergonomic Design of Control Centres, Part 1: Principles for the Design of Control Centres," Geneva, Switzerland: International Standards Organization, 2000.

ISO 11064-2, "Ergonomic Design of Control Centres, Part 2: Principles for the Arrangement of Control Suites," Geneva, Switzerland: International Standards Organization, 2000.

ISO 11064-3, "Ergonomic Design of Control Centres, Part 3: Control Room Layout," Geneva, Switzerland: International Standards Organization, 1999.

Ivergard, T., *Handbook of Control Room Design and Ergonomics,* London: Taylor & Francis, 1989.

Lipner, M. H., "Operational Benefits of an Advanced Computerized Procedures System," Pittsburgh, PA: Westinghouse Electric Company, 2001.

Modern Power Station Practice, 3rd ed., Vol. F, *Control & Instrumentation,* Oxford: British Electricity International, Pergamon Press, 1991.

NUREG/CR-6633, "Advanced Information Systems Design: Technical Basis and Human Factors Review Guidance," USNRC, March 2000.

NUREG/CR-6634, "Computer-Based Procedure Systems: Technical Basis and Human Factors Review Guidance," USNRC, March 2000.

NUREG/CR-6635, "Soft Controls: Technical Basis and Human Factors Review Guidance," USNRC, March 2000.

NUREG/CR-6636, "Maintainability of Digital Systems: Technical Basis and Human Factors Review Guidance," USNRC, March 2000.

NUREG/CR-6637, "Human Systems Interface and Plant Modernization Process: Technical Basis and Human Factors Review Guidance," USNRC, March 2000.

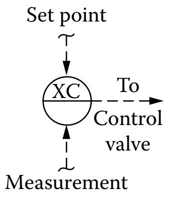
NUREG/CR-6684, "Advanced Alarm Systems: Revision of Guidance and Its Technical Basis," USNRC, November 2000.

NUREG/CR-6691, "The Effects of Alarm Display, Processing and Availability on Crew Performance," USNRC, November 2000.

Technical Report 1001066, "Human Factors Guidance for Digital I&C Systems and Hybrid Control Rooms: Scoping and Planning Study," Palo Alto, CA: Electric Power Research Institute (EPRI), November 2000.

4.4 Controllers—Electronic Analog and Digital

J. E. TALBOT (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

Types:	A1. HVAC quality and or 1/8 Deutsche Institut fuer Normung (DIN) case for direct RTD, TC, mV, or mA DC service. Can be analog or microprocessor-based A2. Process control quality electronic analog controller for 4 to 20 mA DC service B. Digital, microprocessor-based (“smart”), configurable controllers
Standard Input/Outputs:	A. 4- to 20-mA DC B. No standard
Other Input/Outputs:	A. 1 to 5, 10 to 50 mA DC; 1 to 5, 0 to 10 V DC; ±2 mA DC; ±10 V DC B. Binary, binary-coded decimal, or American Standard Code for Information Interchange (ASCII) data with communication formats per RS232, RS422, RS485, and IEEE488 in addition to TC and resistance temperature detector (RTD) inputs and analog I/O
Repeatability:	A. 0.25 to 0.5 %; A2 designs give better repeatability B. 0.1 to 0.5 %
Inaccuracy:	A. Generally around 1%; can be better with A2 designs and is worst with A1 designs B. Generally 0.1% on input (except TC and RTD, which are 1°F or 0.5°C), set point, alarm, calculations, and displays, except for bar graphs, which are 2%. Outputs to valves are 0.5% and mode settings are about 5% inaccurate. A 55°F (30°C) ambient temperature change will cause an error of 0.1% in local set points and 0.5% in remote set points, and an error of 0.5% in analog inputs and valve outputs. The error due to humidity variations or to a 10% change in supply voltage is about 0.1%. While the different error sources can be additive, it is also possible to limit the overall system error to less than 0.1%.
Control Modes:	A. Manual, on/off, proportional, integral, PI, PD, PID B. Same as analog plus error squared, sample and hold, ratio, linearization, arithmetical and logical operations, counter-timer-selector or limiter functions, external feedback, feedforward compensation including lead/lag, and many other software options
Control Mode Adjustment Ranges (Standard):	A. Gain = 0.1 to 50, proportional band (PB) = 2% to 1000%, integral = 0.04 to 100 repeats/minute, derivative = 0.01 to 20 minutes B. Gain = 0 to 128, PB = 0.8% to unlimited, integral = 0 to 293 repeats/minute, derivative = 0 to 895 minutes
Displays:	A. Analog indications of process variable, set-point, output, deviation, and balance B. Hybrid display of above data in digital and bar graph form plus a variety of status displays including lights and letter code abbreviations
Costs:	A. Direct TC, mV, RTD input controllers in 1/4 or 1/8 DIN case can be obtained for $500 or less; process control quality electronic PID controllers range from $1000 to $2500 B. $1500 to $7500; the higher-cost units can handle more loops and can be provided with I/O for both logical status and multi-loop control

Partial List of Suppliers:

ABB Instrumentation (A2, B) (www.abb.com/us/instrumentation)
Action Instruments (A1) (www.actionio.com)
Analogic Corp. (B) (www.analogic.com)
Athena Controls Inc. (A1) (www.athenacontrols.com)
Barber-Colman Industrial Instruments (A1) (www.barber-colman.com)
Bristol Babcock Inc. (B)(www.bristolbabcock.com)
Chromalox (A1, B) (www.mychromalox.com)
Emerson Process Management (B) (www.EasyDeltaV.com)
ExperTune Inc. (A2) (www.expertune.com)
Fisher Controls International Inc. (B) (www.fisher.com)
Foxboro-Invensys (A2, B) (www.foxboro.com)
Honeywell Sensing and Control (A1, A2, B) (www.honeywell.com/sensing)
ICS Triplex (B) (www.icstriplex.com)
Jumo Process Control Inc. (A1) (www.jumousa.com)
Love Controls Corp. (A1) (www.love-controls.com)
Matrikon Inc. (B) (www.matrikon.com)
Omega Engineering Inc. (A1) (www.omega.com)
Powers Process Controls (A1) (www.powerscontrols.com)
Robertshaw IDP, an Invensys Co. (A1) (www.robertshawindustrial.com)
Siemens Energy & Automation (A2, B) (www.sea.siemens.com)
Smar International Corp. (B) (www.smar.com)
Thermo Electric Co., Inc. (A1) (www.thermo-electric-direct.com)
Toshiba International Corp. (A2, B) (www.tic.toshiba.com)
Triplett Corp. (A1) (www.triplett.com)
United Electric Controls (A1) (www.ueonline.com)
Watlow (A1) (www.watlow.com)
Westinghouse Process Control (A2, B) (www.westinghousepc.com)
Wilkerson Instrument Co. (A1) (www.wici.com)
Yokogawa (A2) (www.yokogawa.com/us)

INTRODUCTION

This section describes on/off, direct-connected, and transmitter-based controllers of both the analog and the digital types. In the first part of this section, the discussion of electronic controllers begins a description of on/off and direct-connected analog controllers. This is followed by a more detailed treatment of transmitter-based analog control (Figure 4.4a), which is followed by a description of the features and capabilities of digital controllers.

Analog vs. Digital Controllers

As of this writing analog controllers are still more familiar to plant operators and maintenance personnel, and even to some engineers, than are their digital counterparts, and therefore require less time to specify or to program. Analog controllers can also be less expensive, although some low-cost microprocessor-based controllers are also available (Type A1 in the Feature Summary at the front of this section).

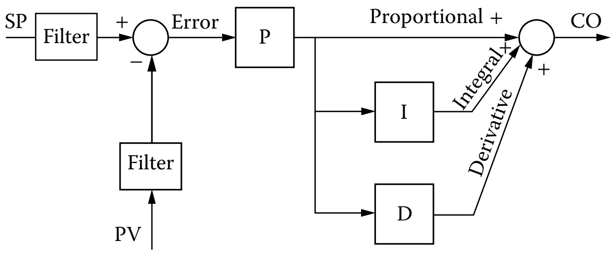


FIG. 4.4a
Configuration of a transmitter-based analog PID loop (PV = Process Variable, CO = Controller Output).

Analog controllers are also not only dedicated and hard-wired but are continuous devices, while microprocessor-based controllers, even when dedicated and hard-wired, are still intermittent units in the sense that they do not look at their inputs continuously, but on a 0.05- to 0.5-second time cycle. Analog controllers do not require the use of analog-to-digital (A/D) and digital-to-analog (D/A) converters. Another advantage of analog controllers is that they are easily interfaced with any other analog control system, due to their standardized signal level of 4- to 20-mA DC. In contrast, the interfacing of digital controllers of different suppliers can require the use of interface cards.

The advantages of digital controllers are also many. Their accuracy and stability are fundamentally unlimited. Error sources such as hysteresis, nonlinearity, or thermal drift do not exist in a digital system. In addition, the digital systems are capable of more sophisticated calculations and algorithms without recalibration (see Figure 4.4b showing a model-based feedback controller). Digital controllers can also be easily and quickly reconfigured without requiring any change in hardware. They can provide more information to the operator on their displays and can transmit data to other

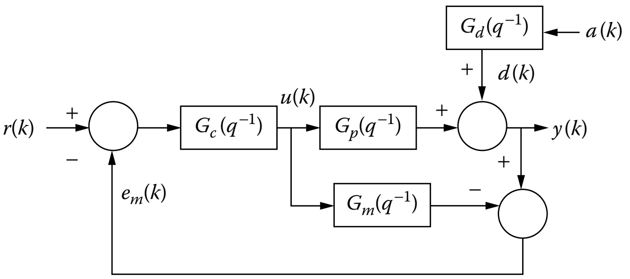


FIG. 4.4b
The configuration of a model-based feedback controller.

digital devices faster and more accurately than analog systems can.

Digital systems can reduce wiring costs and control room sizes and are also less subject to drift, which occurs in analog devices due to component leakage. Digital controllers can also include such functions as alarm, square root extraction, timing, sequencing, totalization, and specialized displays, which in analog configurations would require separate components.

The Controller's Function

Most industrial processes require that certain variables, such as flow, temperature, level, and pressure, remain at or near some reference value, called a set point. The device that serves to maintain a process variable value at the set point is called a controller. The controller looks at a signal that represents the actual value of the process variable, compares this signal to the set point, and acts on the process to minimize any difference between these two signals.

Any simple process control loop (Figure 4.4c) contains the equivalent of a sensor, transmitter, control element (usually a valve), and controller. The sensor measures the actual value of the process variable. The transmitter amplifies this sensed signal and transforms it into a form suitable for sending to the controller. The controller really has two inputs: a measurement signal and a set-point signal. The set point can also be internally generated. The controller subtracts the two input signals and thereby produces a deviation or error signal.

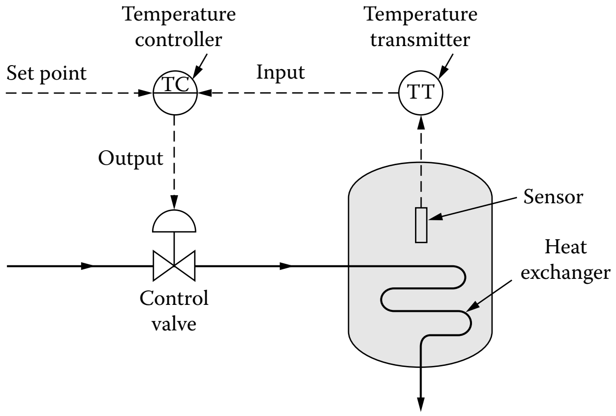


FIG. 4.4c
Standard temperature control loop.

Additional hardware or software is provided in the controller to act on this deviation or error signal and thereby to generate the controller's output. For the various control algorithms, including PID, refer to the sections in Chapter 2. The controller output then typically sets the position of a pneumatic or electrical final control element or control valve in a direction to decrease the error.

Although the function of the controller is easy to define, it may be implemented in many different ways. For example, the controller may work on pneumatic, fluidic, electric, magnetic, mechanical, or electronic principles, or on combinations of these. This section, however, will confine itself primarily to electronic controllers.

The general nature of their inputs, outputs, and control modes makes controllers applicable to a wide variety of process control situations. The same controller may control flow in a pipe, which responds nearly instantly to valve movements, or control a temperature in a distillation column, which often takes hours to respond to disturbances. Any process variable is a candidate for remote control as long as it can be converted to a transmitter signal. The process control engineer must choose and tune the control modes to fit the dynamics of the variable to be controlled (see Chapter 2).

Feature Checklist

The following is a summary of the relevant features that an instrument engineer must consider when picking a controller:

Control mode selection
Input and output ranges (including set point)
Output load resistance range
Emergency service provisions
Maintenance accessibility and convenience
Panel readability and accuracy
Control repeatability and accuracy
Tuning ranges and resolution
Electrical classification
Power requirements; need for regulation
Mounting flexibility and density
Switches for local/remote set point, direct/reverse action, and manual/automatic operation
Balancing procedures and accuracy
Alarm modules and lights
Output limits
Antiwindup
Computer compatibility

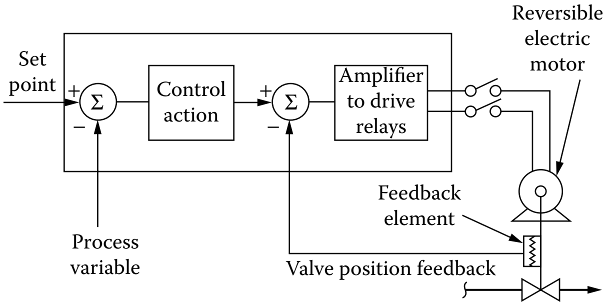


FIG. 4.4d
Electronic positioner controller with relay outputs.

ON/OFF AND DIRECT-CONNECTED CONTROLLERS

On/Off Relay Outputs

A common example of an electronic controller with relay outputs is the positioner controller (Figure 4.4d). Here the controller energizes one of two relays to operate an electric motor in either direction. The motor drives some final element to a desired position. A feedback signal from the final element tells the controller when the desired position has been reached.

Essentially, the controller acts as a master controller in a cascade system, where the relays, motor, final element, and feedback signal constitute the slave loop, and the master controller generates the set point for the slave by conventional control action.

A meter–relay combination provides an extremely simple way of implementing two- and three-position, high-gain control. Basically, the input signal, such as a thermocouple, drives a galvanometer's pointer to indicate temperature. Another pointer (or index) is manually set to the desired temperature on the same indicating scale. The idea is to operate a relay or a silicon controlled rectifier (SCR) when the two pointers are in the same position.

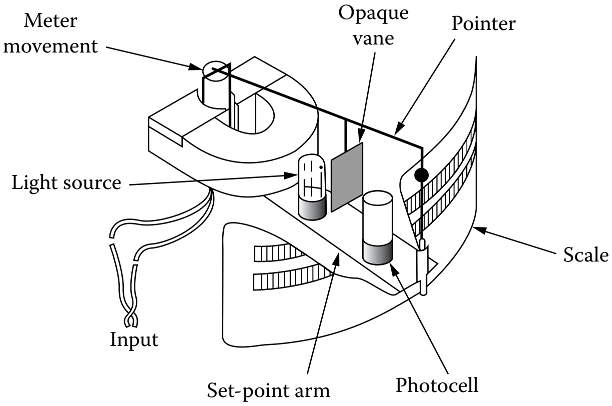


FIG. 4.4e
Optical meter relay.

Optical Relays The system shown in Figure 4.4e optically senses the correspondence of the two pointers. The set-point arm carries a small photocell and light source. As long as light strikes the photocell, amplified current energizes the relay to supply power to a load. When the indicated temperature reaches the set-point temperature, a vane on the indicating pointer breaks the light beam, and the relay drops out.

Some meter relays are equipped with two set-point arms and relays to give three-position control. In a meter with an SCR output, the vane can be shaped to provide some degree of proportional control over a narrow temperature band. The optical system avoids any interaction between the process and set-point pointers, as would be present in mechanical or magnetic systems. Obviously the idea can be extended to other displays where the variable is indicated by a pointer's position. For example, it has been applied to electronic circular and strip-chart recorders.

Electronic Relays Meters can also be purchased with built-in electronic amplifiers to boost the input signal, as in Figure 4.4f. In one meter–relay package of this type, the meter's set-point arm wipes a resistance element. This resistance element, connected as a voltage divider, provides a voltage reference signal, which is compared to the amplified input. The deviation, after further amplification, drives a transistor that operates a control relay. The meter and resistance element can be encapsulated in an inert atmosphere if the device must be installed in a relatively hostile environment.

Direct-Connected Controllers

A wide variety of controllers on the market are tailored to specific kinds of control by specialized input or output schemes, or both. The most common examples by far are electronic temperature controllers for driving electric heaters and motor speed controllers. The temperature controllers are designed to directly accept inputs from sensors like thermocouples, resistance bulbs, or thermistors. Their outputs may drive relays or contactors, silicon-controlled rectifiers (SCRs), or saturable reactors.

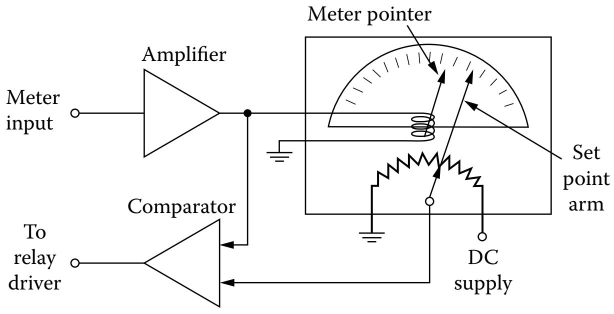


FIG. 4.4f
Meter-relay package with electronic input and output circuits.

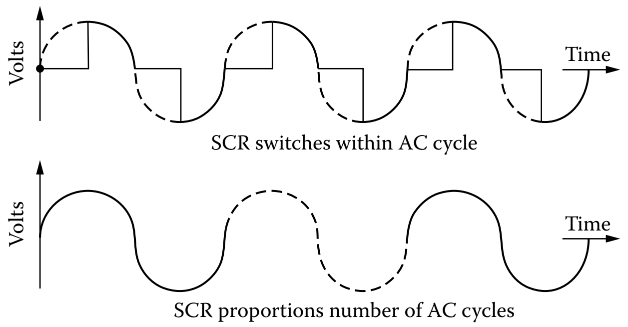


FIG. 4.4g
Two SCR switching techniques that are used for driving electrical heaters.

Since the sensor connects directly to the controller, the input signal lines handle very low power levels and the lines are kept short to avoid interference from other sources. In these designs the input circuit is usually for a particular kind of sensor and a certain range of temperature. Common thermocouple inputs include iron-constantan, copper-constantan, chromel-alumel, and platinum/rhodium-platinum pairs. The controller's input circuit compensates for variations in the thermocouple's cold junction temperature.

Resistance bulb and thermistor inputs generally feed into resistance bridge circuits.

As in all feedback controllers, the input circuit compares the incoming signal with a reference or set-point signal, and control circuitry acts on the resulting deviation. The form of control ranges from simple two- (on/off) or three-position (high-low-off) control to sophisticated three-mode control. Some manufacturers offer three-mode control on one set-point position and on/off control on another.

Time-Proportioning Control Proportioning, however, is generally on a time basis, called time-proportional control. Here the output is driven full on or off, but the control signal varies the percentage of on-to-off time in some duty cycle, such as 10 seconds. The *average* power level determines the load temperature. This kind of control applies to both relay and SCR outputs. However, the SCR is capable of faster switching than the relay, permitting proportioning periods of a fraction of a second and thereby giving smoother control.

Some controllers trigger the SCR to chop up each AC cycle; in others the SCR proportions the number of AC cycles that are on and off in a certain period (Figure 4.4g). Available SCRs can handle loads up to 300 kW in three line phases.

ANALOG CONTROLLERS

An electronic process controller consists of six basic parts or sections: input, control, output, display, switching, and power supply (Figure 4.4h). The input section comprises the hardware for generating the set-point signal, for accepting and conditioning the process signal, and for comparing the two signals to produce a deviation signal. The heart of the control section is generally a chopper-stabilized AC amplifier that acts on the deviation signal. Associated circuitry amplifies, integrates, and differentiates the deviation signal to produce an output with the necessary proportional gain, integral (reset), or derivative (rate) control actions. (In some controllers the derivative circuit acts on the process signal rather than the deviation signal.)

The controller's output has two forms: automatic and manual. When in the manual mode, the controller's output remains at a steady level, often dependent on the shaft position of a potentiometer. Switching and balancing hardware on the front panel selects the automatic or manual output. In some cases the source of the manual output is the control amplifier (modified by switching), while in other cases the source is entirely separate, requiring only the operation of a power supply.

The display section on the front panel carries information about the set point, process variable value, deviation, and controller output. The front panel also contains mechanisms for switching between manual and automatic modes and for adjusting the value of the set-point input and the manual output. A secondary indication, balance, allows the operator to equalize the manual and automatic outputs before transferring from the automatic to the manual mode.

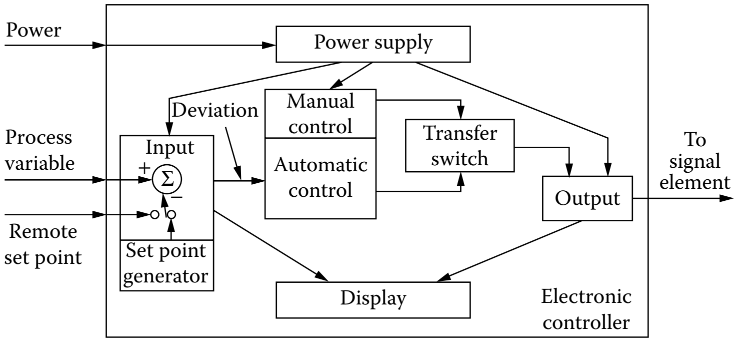


FIG. 4.4h
The basic components of a controller.

The power supply section serves to transform the incoming AC line voltage to the proper DC levels to operate the other controller sections. Some controller power supplies also serve to energize instrumentation external to the controller. For example, the controller may power the transmitter in the same loop. In other cases the power supply itself may be external to the controller, energizing many controllers and other instruments.

In addition to these basic parts, the controller may have various special optional or standard features. Some examples include alarms, feedforward inputs, output limits, and batching aids (described later).

Input Variations

Input and output signals for analog electronic controllers have more-or-less been standardized, through general usage, to three ranges: 1- to 5-, 4- to 20-, and 10- to 50-mA DC, with 4 to 20 by far the most common. Nearly all the controllers on the market will accommodate at least one of these standard input signals. Many offer all three, either as standard or optional features. On the output side, the manufacturers generally offer only one of the signals, the most popular being 4- to 20-mA Dc.

The use of a DC current amplifier for transmitting signals over great distances has one primary advantage over a voltage signal: small resistance that develops in the line (from switches and terminal connections and from the line itself) do not alter the signal value. Once the input signal reaches the controller, it is generally sent across a suitable resistor to produce a voltage that matches the working voltage range of the controller's amplifier.

The signal's live zero aids in troubleshooting since it differentiates between the real signal zero and a shorted or grounded conductor. Besides these three signal ranges, some manufacturers offer uncommon inputs such as 1- to 5- and 0- to 10-V DC, and outputs such as 1- to 5-V DC; centered-zero signals (±10 V DC and ±2 mA DC). Direct TC, RTD, or mV signals are also available.

Set-Point Variations The controller's set-point input is either internally (local) or externally (remote) generated. Manufacturers generally provide a switch to let the user choose between the two, but other suppliers require the user to choose between two different models. In some electronic controllers the two signals can be displayed and "nulled" to equalize them before transferring from one to another. It is also possible, but more costly, to have the unused signal automatically track the other, so that the two are always equal.

If the remote and local set-point signals are not equal before transfer, the controller will see a sudden set-point change. The output will therefore also change abruptly, unnecessarily disturbing or "bumping" the process.

The local set-point is often merely the output of a voltage divider that is connected to the set-point adjustment and display mechanism. The divider consists of either a multi-turn variable resistor (potentiometer) or a large fixed circular slide-wire with wiper. The divider's output is compared to the voltage drop produced by the process input signal, yielding a deviation signal that corresponds to the difference between the two.

The remote set-point signal must have the same range as the process input signal. Manufacturers provide terminals at the back of the controller to accept this signal. Common examples in which a remote set-point signal comes into play are the cascade and ratio control systems (Figure 4.4i).

In the cascade system the source of the set-point signal is the output of another controller, called the primary, or master, controller. The controller that accepts this remote set point is the secondary, or slave, controller. Another common source of remote set point is a function generator or a programmer, set up to make the controlled variable follow a prescribed pattern (Section 3.4).

Blending systems are a form of ratio control. Here the set-point signal is really a second process variable from a flow transmitter. The controlled stream is forced to follow the flow rate of the uncontrolled (or wild) stream in some adjustable proportion (ratio). So the signal from the flow transmitter on the wild stream serves as a remote set point. The means for adjusting the desired ratio between the two streams is usually incorporated in the front panel of the controller, making it a special model. Calibrated ratios generally run from 0.3 to 3.0.

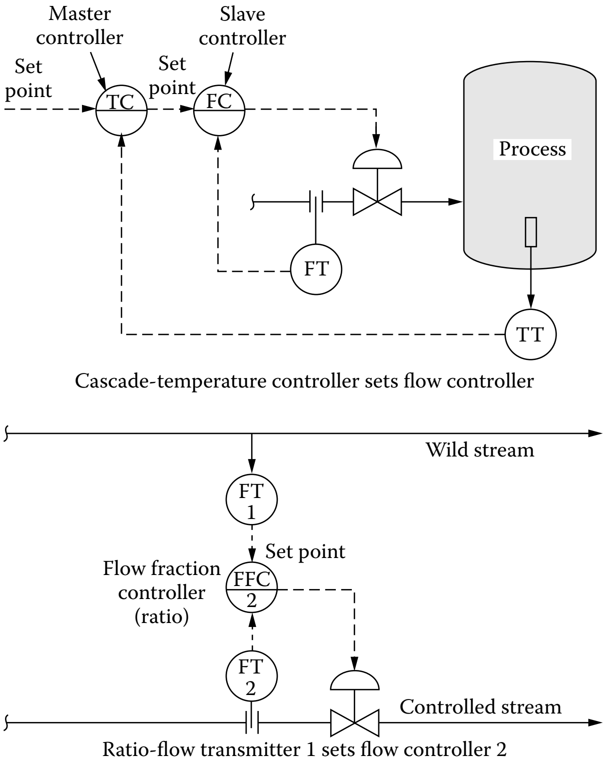


FIG. 4.4i
Cascade and ratio control systems.

An optional feature pertaining to a remote set point is a servo system (motorized set point) that makes the set-point display mechanism track the remote set-point value. Otherwise the display should be blanked or marked in some way because it corresponds to the local rather than remote set point. In some controllers the set point is displayed on a galvanometer, and then the above comments do not apply. Motorized set points often find application in supervisory control systems where the remote set-point source is a computer.

Control Modes

The controller must deal with processes having wide ranges of dynamic characteristics. Adjustable control modes provide the needed flexibility. These modes act on the error signal, producing changes in controller output to compensate for process disturbances that have affected the controlled variable. Electronic hardware modules provide one or more of the following control actions, which are discussed in detail in Chapter 2:

1. Proportional gain
2. Proportional gain plus integral (reset)
3. Proportional gain plus derivative (rate)
4. Proportional gain plus integral plus derivative
5. Integral (proportional speed floating)

The control action with the most widespread use is proportional gain plus integral. Only a few manufacturers offer proportional gain alone, and even fewer offer pure integral.

In most cases the manufacturer provides one basic controller that has all three modes. In that case the user can switch off the unneeded mode(s). In other designs, the user can pick from a variety of models, depending on desired modes, fast or slow integral rates, fast or slow derivative rates, and other options. The choice depends on the dynamics of the process to be controlled.

The proportional mode varies the controller output in proportion to changes in the error. High gains mean that the controller output varies greatly for small changes in deviation from set point. Extremely high gains result in nearly on/off control. In processes requiring relatively low gains, the importance of the integral mode increases because at low proportional gains the controller will be satisfied even though the controlled variable and the set point are relatively far apart (offset). The integral mode continues to drive the output until the deviation goes to zero.

The derivative mode is a refinement that produces an output component proportional to the rate of change of either the deviation or the process input signal. Its effect is to anticipate changes in the process variable under control. Figure 4.4j shows a simplified circuit for implementing three-mode control.

The advantage of having the derivative mode act on the process input rather than on the deviation is that the controller does not respond directly to a set-point change. If it acts on the deviation, the derivative component of the output depends, for example, on how fast the operator turns the set-point knob. Of course one could always switch to manual before making a set-point change and then switch back to automatic. A few manufacturers can provide both kinds of derivatives as optional designs, but most offer only one or the other.

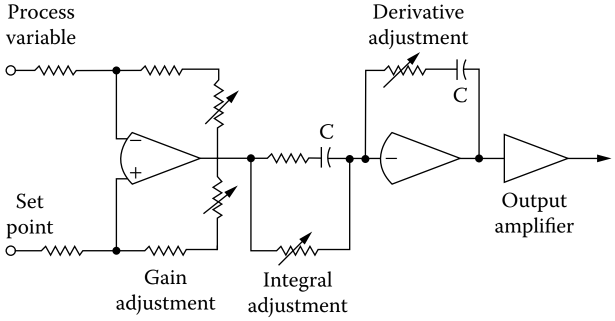


FIG. 4.4j
Three-mode control circuit.

Tuning As was discussed in Chapter 2, the adjustments of each control mode allow the operator to tune the controller so it will match the process dynamics. Continuous or multiposition knobs are provided, usually inside the controller case. These knobs are calibrated in units of the corresponding control mode. Typical ranges and units are:

1. Proportional gain, 0.1 to 50 (proportional band 1000 to 2%)
2. Integral, 0.04 to 100 repeats per minute
3. Derivative, 0.01 to 20 minutes

Higher and lower values for each mode are available, depending on the manufacturer. The term proportional band (PB) is also used; it is related inversely to proportional gain:

$$\text{PB} = \frac{100}{\text{proportional gain}}\% \qquad \textbf{4.4(1)}$$

One repeat per minute means that the integral mode will ramp the output up or down, repeating the proportional action that corresponds to a particular deviation every minute. Ramping continues until the deviation is zero. A derivative setting of 1 minute means that the controller output should satisfy a new set of operating conditions 1 minute sooner than it would without the derivative mode.

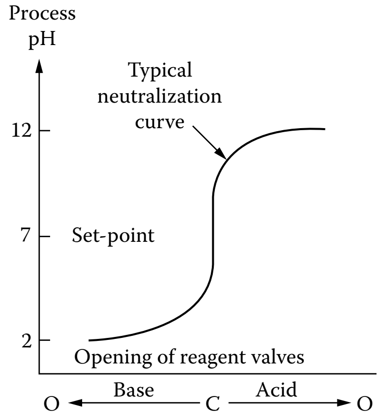
Many manufacturers also have scaling switches or jumpers that multiply the effects of the integral and derivative modes by factors of 2 or 10. Also, the ability to switch the derivative and integral modes completely out of the control circuit is an aid for troubleshooting and calibrating.

The direction of the control action depends on the set point and the process variable input. Direct control means that as the process variable increases, the controller output also increases. Reverse control means that the controller output responds in a direction opposite to that of the process variable. A direct-reverse switch in the controller allows the operator to reverse the controller action. If the derivative action is on the process measurement rather than the error, at the time of switching, this connection must also be reversed.

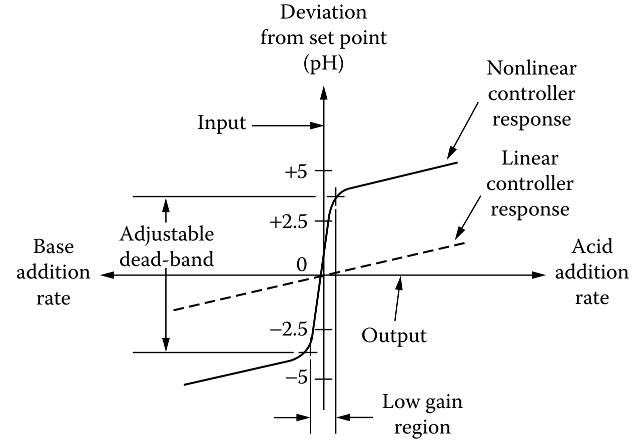
Some circuit designs permit changes in the control settings while the controller is operating in automatic, without disturbing the output. Sometimes only the proportional gain can be so adjusted, without having an effect on the output. Otherwise, the controller must be switched to manual before changing the settings. Usually the settings are somewhat interacting, meaning that a change in one affects the value of another. However, certain controllers incorporate features such as separate amplifiers for each mode to eliminate or minimize this interaction. These controllers have an advantage when tuning is critical.

Although the process controllers have standard output signal ranges, limitations on the maximum load resistance in the output circuit vary by manufacturer. The load resistance is the sum of the series resistance of all the devices driven by the controller's output current. Typically, the output current drives a converter that connects to a diaphragm-actuated control valve. These varying load requirements among manufacturers complicate matters somewhat if the user wants to buy parts of the control loop from different manufacturers.

A sampling of actual specifications for maximum load resistance includes 75, 500, 600, 800, 1500, and 3000 ohms. Higher maximums, of course, give the user greater flexibility. A few manufacturers require a minimum load resistance as well. At least one vendor asks the user to adjust the load resistance in the output circuit to an exact value, which is a severe limitation.



The nonlinear process of neutralization



Nonlinear controller response

FIG. 4.4k
The nonlinear controller response required for the control of a nonlinear acid-base neutralization process.

Nonlinear Controllers

If the process tends to be sensitive (fast) at certain measurement values, while at other measurement values it is sluggish (slow to respond), then that process is nonlinear. A nonlinear process can be properly controlled only by a nonlinear controller. In order to arrive at a uniform loop gain at all measurement values, it is necessary to change the controller gain from high to low as the process measurement moves from its sluggish region to a more sensitive zone.

The best known of the nonlinear processes is acid-base neutralization, which is sensitive in the region near neutrality and therefore requires a low controller gain in that region. Figure 4.4k illustrates this nonlinear process together with the nonlinear controller response it requires. This controller has an adjustable nonlinear relationship between measurement and output, with low gain surrounding the set point and with normal gain outside of this band. The width of the "dead band" (band of low gain) is usually adjustable both manually and remotely from 0 to ±30% of the measurement range, and its gain relative to the normal gain of the controller is adjustable manually from 0.02 to 0.2. If the dead band is set at zero, the result is a linear controller, illustrated by the dotted line in Figure 4.4k.

The standard control modes of proportional, reset, and derivative are superimposed on the nonlinear function. For example, if the normal gain is set at 2 and the minimum gain within the dead zone is 0.2, then the effective gain inside that zone will be 0.04. This means that while the process measurement is inside the dead band, a 25% change in measurement will result in only a 1% change in controller output, due to the proportional mode response.

The width of the dead band can be adjusted manually or automatically in direct or reverse proportion to an external signal. For example, the control valve gain can increase as the valve opens. Widening the controller's dead band can compensate for the increase in loop gain.

Nonlinear controllers can also be used to allow the levels to float in surge tanks, while preventing the tanks from flooding or draining. The goal of such a loop is to keep the controlled flow reasonably constant while allowing it to be smoothly modified when necessary.

A nonlinear controller can also be used to filter out the noise or pulsations from flow loops, while permitting effective control over major disturbances. In order to eliminate valve cycling during normal operation, the dead band should equal the band of noise or pulsation. When used for flow control, the nonlinear controller gain inside the dead band should *not* be set at very low values because this could cause regenerative oscillation.

Special Features

Manufacturers usually offer special features, either optional or standard, to enhance the convenience and usefulness of controllers. The most common features involve alarm modules and output limits in some form. Electronic alarms can be set up on either the process measurement, the deviation, or both, depending on the supplier. Also, they can be actuated on either the high or low side of the alarm point. Some controllers have panel lights for connecting to the alarm contacts as well as terminal connections for external alarms such as lights, bells, horns, or buzzers.

Nearly all manufacturers of electronic controllers limit the high side of the output current range to the maximum standard value. Often a zener diode arrangement serves to clamp the current at this value. Some manufacturers go a step further and place a fixed limit on the low end of the range. Still others offer adjustable output limits on both ends. Output limits serve only to constrain the range of the manipulated variable (valve stroke).

Antireset Windup Certain limits can prevent a phenomenon called reset (integral) windup (discussed in detail in Chapter 2). Windup occurs if the deviation persists longer than it takes for the integral mode to drive the control amplifier to saturation. This might occur, for example, at the beginning of a batch process, if the controller input signal is temporarily lost, or if a large disturbance occurs in either the set point or the measurement.

Only when the deviation changes sign (process variable value crosses set point) does the integral action reverse direction. So the process variable will most likely overshoot the set point by a large margin.

If the limit acts in the feedback section of the control amplifier's integral circuit, the controller output will immediately begin to drive in the opposite direction as soon as the process signal crosses the set point. This approach is commonly referred to as *antireset windup.* On the other hand, if the limit acts directly on the output as discussed earlier, it essentially diverts excess current coming from the control amplifier, and therefore the output will remain at a high value for some time after the process crosses the set point. The output limit will simply divert less and less excess current. This extends the time that the output remains at a saturated value, aggravating the overshoot due to windup. Antireset windup circuits are usually offered only to eliminate saturated output currents at the high end of the signal range.

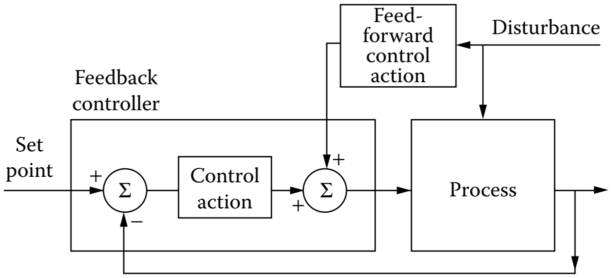


FIG. 4.4l
Electronic controller configuration in a feedforward application.

A few manufacturers go so far as to provide special batch controller models to minimize reset windup. Here, special circuit modifications actually begin to drive the output out of saturation *before* the process variable value crosses the set point. Reset windup can be avoided in batch applications if the operator equalizes the process variable value and set point in the manual mode before switching from manual to automatic operation.

Feedforward and Special Features Another powerful control configuration is feedforward. In these controllers, a feedforward input signal influences the value of the controller output. Feedforward control (covered in detail in Chapter 2) assumes that the user knows exactly how much the control valve (or other final element) should change to compensate for the changes in the feedforwarded process input variable (Figure 4.4l).

This relieves the controller of compensating for this particular process disturbance by normal feedback control. A few manufacturers offer gain and bias modules for conditioning the feedforward signal before it sums with the controller output signal. A direct and reverse switch is necessary to match the feedforward signal to the controller action.

Other special features offered by various manufacturers include:

Communication jack. Lets signal wires between the controller and transmitter serve also as an audio communication link; an AC carrier system does not disrupt the DC process variable signal. This eases calibration and troubleshooting of the control loop.

Trend record. Provides jacks or pins for patching a recorder to the process variable input signals available in voltage form in the controller.

Output tracking. Locks the controller output to a remote signal. This feature is useful, for example, when an electronic controller backs up a computer control installation.

Process retransmission. Converts a process current signal to a low-impedance voltage signal for distribution to other parts of the control panel. Other devices can then be connected in parallel to this signal without loading down the controller input signal.

Special input modules. Extract square root, select inputs, accept resistance inputs, excite strain gages, convert pulse inputs, integrate, isolate, condition, and limit inputs. Some manufacturers offer many more options than others.

Battery backup. Automatically takes over when regular power mains fail.

Displays

As mentioned earlier, the controller's front panel indicates four basic signals: set point, process variable, error deviation, and controller output. The deviation meter sometimes doubles as a balance meter for equalizing the manual and automatic signals or local and remote set-point signals when the operator wants to switch from automatic to manual output. The front panel has one small meter to show the controller output. The forms of display for the other three basic variables differ among manufacturers and sometimes differ among models by the same manufacturer. Controllers without any panel displays are called blind controllers.

The majority of process controllers merge the set point, deviation, and process variable indications into one display (Figure 4.4m). This display consists essentially of a deviation meter movement combined with a long steel tape or drum scale behind the meter's pointer. The scale length ranges from 6 to 10 in. (150 to 250 mm), depending on the manufacturer. Only a portion of this scale shows, but the tape or drum can be moved so that the scale value corresponding to the desired set point rests exactly at the center of the display, marked by a hairline or a transparent green band. If used, the green band masks the red pointer for small deviations. For larger deviations, the red pointer peeks glaringly from behind the band to alert the operator.

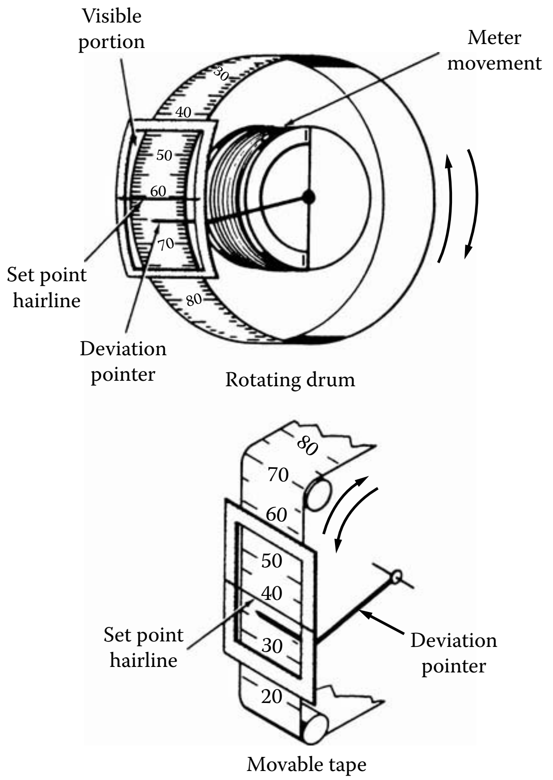


FIG. 4.4m
Process variable, set point, and error can all be shown on expanded scale displays.

The deviation meter movement corresponds to the visible part of the scale. For example, if 50% of the scale shows, the deviation movement is ±25 % of full scale. So the pointer indicates the process variable value as long as it is on scale, the hairline at the center gives the set-point value, and the distance between the hairline and pointer indicates the deviation.

The mechanical set-point scale movement also drives components for generating the corresponding electrical set-point signal. Commonly, the scale movement drives a multi-turn potentiometer. In some controllers, a wiper attached to a movable scale drum contacts a large fixed circular slide-wire connected as a voltage divider. Either design should avoid gearing between the set-point drive and electrical components; otherwise, backlash would destroy their one-to-one correspondence.

Controllers that have a large deviation display with a long, expanded scale offer high resolution and readability of each signal value. Of course, the deviation must be on scale to be able to read the process value at all. Also, the electrical and mechanical set-point values must be calibrated carefully or the process reading will be inaccurate.

In controllers without this kind of display, some compromise is usually made in the indication of set point, deviation, and process variable value. It is impractical and confusing to put a display for each of these signals on the controller's narrow face. The set point must be displayed, but either the process variable or deviation display can be sacrificed without great loss, since one is implied by the other once the set point is known.

The set point is either read out directly on a meter or mechanically displayed with a calibrated dial and index arrangement (Figure 4.4n). Sometimes the process variable and set-point value are shown with two pointers on a single scale. This lets the operator easily compare the two for estimating the deviation. In another case it is the process value that must be estimated from markings that relate the deviation meter readings to a calibrated set-point dial.

One way to circumvent the limited space on the controller's front panel is to have one meter serve to indicate more than one variable. The operator then must position a switch to choose the variable he or she wants to read. One manufacturer indicates five basic controller signals (including a balance indication) with one meter. To avoid confusion, a set of switches inside the case determines the direction of meter movement for different switch positions so that the meter movement always drives upscale.

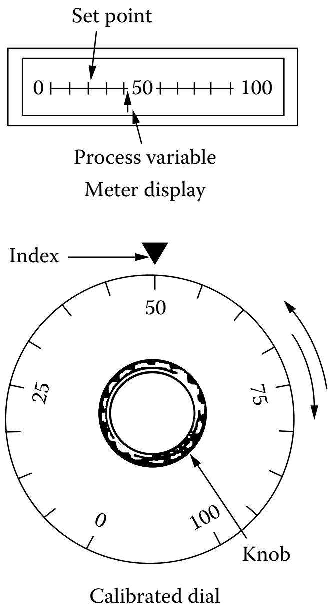


FIG. 4.4n
Conventional set-point setting knob and display.

The controller's output meter shows either the manual or the automatic output signal, depending on the transfer switch position. Interchangeable tags at the side of the meter remind the operator of the valve positions (open or closed) for the extreme controller outputs. In certain controllers the output meter can be physically inverted so that the pointer always drives upscale as a valve opens. Occasionally the output meter may act as a deviation meter to null the automatic and manual outputs before transfer.

The front panel may also contain alarm light displays. These are often connected to alarm modules behind the panel that monitor deviation or process variable input. Some lights may merely run to an external connection at the rear of the controller. The position and form of the lights vary greatly. One manufacturer backlights legend plates. Another backlights the top and bottom of the deviation meter to show high and low deviation. In one case, deviation alarm lights replace the deviation meter itself. A few manufacturers provide a small standard panel light to indicate whether the local or remote set point is implemented.

Balancing Methods

All electronic controllers have a switch on the front panel for transferring between automatic and manual outputs. The switch takes many forms, including lever, toggle, rotary, and pushbutton. It is often important to balance the two output signals before transferring from automatic to manual to avoid a sudden change in controller output that might disturb or "bump" the process. Controllers that feature "bumpless transfer" have some provision for balancing the two output signals just before switching.

No balancing is necessary when switching from manual to automatic output because the unused automatic controller signal takes the same value of (or tracks) the output when the controller is in the manual mode. So the operator merely switches from manual to automatic and need not be concerned with balancing procedures. However, if a deviation exists between process and set-point values at the time of transfer, the integral action in the controller (if present) will immediately begin to drive the controller in a direction to eliminate this deviation. For large deviations the integral action will often produce a controller output change fast enough to bump the process.

This balance-less, bumpless transfer from manual to automatic output depends on a modified integral circuit to make the automatic signal track the manual output. So the integral function must be present if balance-less transfer is desired.

Nearly all manufacturers require a balancing procedure for bumpless transfer when the controller is switched from automatic to manual output. The transfer switch has a balance position to accommodate this procedure. In the balance position, the controller retains its automatic output, but the functions of the display meters change. The automatic and manual output signals usually feed to either side of a deviation meter. This allows the operator to equalize the two signals by adjusting the manual output until the meter reads zero at mid-scale.

This nulling method has a high sensitivity for accurate balancing. Usually the large deviation meter, if present, serves to compare the output signals, but some manufacturers connect the output meter as a deviation meter for the controller's balance state. At least one supplier places a small separate deviation meter on the panel that always shows the condition of balance.

Still other manufacturers provide a momentary pushbutton on the front panel that, when depressed, connects the manual signal to the output meter in place of the automatic signal. To balance the two signals, the operator repeatedly presses the button while adjusting the manual output. When the pointer on the output meter remains motionless, the two signals should be about equal.

A unique design features balance-less, bumpless transfer in both directions. This means that the unused output signal tracks the other signal at all times. In this case the control amplifier is switched into a simple integrator configuration when the operator transfers from automatic to manual mode. This circuit holds the latest value of the automatic output as a manual output.

To change the manual value, the operator throws a momentary switch that introduces a DC input signal to the integrator circuit input. This causes the output to ramp up or down, depending on whether the input signal is positive or negative. So the operator jogs the manual output to the desired value. The control amplifier also holds the latest manual output signal at the instant the operator switches back to automatic operation, but integral action begins immediately to eliminate any existing deviation between the set point and process variable values.

Mounting

Miniature electronic controllers are designed for panel mounting in a relatively clean and safe environment such as a control room. The room should be free from corrosive vapors and excessive vibration. A few controllers can work at ambient temperatures from 0 to 130°F (0 to 72°C), but others are more restricted. Some models can be mounted in areas designated by the National Electrical Code as Class 1, Group C, Division 2. These areas require that equipment have no open sparking contacts but do not require the instruments to have explosion-proof cases. In these designs the manufacturers provide hermetically sealed on/off power switches to avoid open sparking.

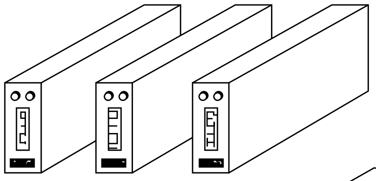
The controller's small frontal size allows many controllers to be installed in a small panel space, which is called a high-density configuration. The controller's depth (over 20 in. or 500 mm) and weight (often over 15 lb or 6.8 kg) usually require additional support at the rear of the panel for multiple mountings. The controllers can usually be mounted in multiple arrays in four different ways, as shown in Figure 4.4o. They can be in individual adjacent panel cutouts, side by side in a single large panel cutout, in large cases or packs that accept some multiple of controller chassis, or on shelves that accept some number of individual controller cases.

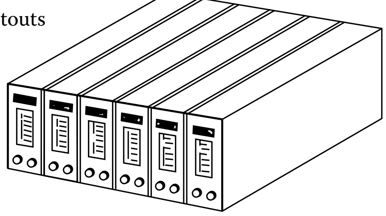
The mounting approach depends on the particular manufacturer, but some companies offer more flexibility than others. A few require individual panel cutouts for each controller. This increases panel fabrication costs. When the controllers are placed side by side in a single cutout, the manufacturer may offer outside trims to frame the installation. The company will also suggest ways to support the back ends.

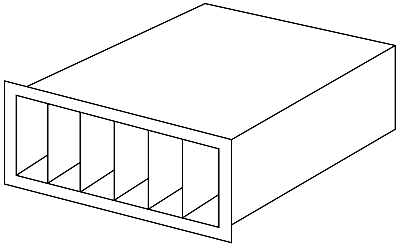
Packs sometimes have both vertical and horizontal capacity, while shelves are limited to horizontal mountings. Some horizontal packs may be specified for any number of controllers, but others are limited either by a maximum number or by certain multiples of controllers. Mounting in packs or shelves generally simplifies power and signal connections. Controllers designed for 1/2 and 1/8 DIN cases are lighter and require less panel space.

If the panel tilts too far from the vertical, the controller may not operate properly. For example, one manufacturer specifies a maximum of 60-degree panel tilt, and another stops at 75-degree tilt from the vertical. Still others say that the controller works in any position. But some companion products, such as recorders, have limitations on their mounting positions.

In all these electronic controllers, the chassis slides partway out of the case without interrupting the control signal. A retractable plug connects the chassis with all the necessary power, signal input, and signal output leads. This permits tuning, calibration, and realignment, while the controller operates in place. In a few controllers, the panel display section can be mounted in a location different from that containing the control and output circuitry.


Individual cutouts


Single cutout–side by side


Pack for controller chassis

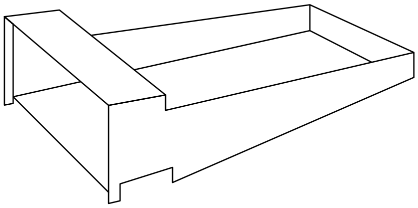

Shelf for controller cases

FIG. 4.4o
Alternate methods for the panel mounting of miniature instruments.

Although the controller is mounted in a relatively safe area (the control room), its signal must eventually go into the field to operate some final element, such as a control valve. Some control systems have been approved by either Factory Mutual or by Underwriters Laboratories as intrinsically safe for hazardous field areas (see Section 7.2 in the *Measurement* volume of this handbook). This means that the output signal in the field cannot release sufficient electrical or thermal energy, under defined normal or abnormal conditions, to ignite a specific hazardous atmospheric mixture.

Special barrier circuits and energy limiters, sometimes with zener diodes (Figure 4.4p), limit the output signal energy, even if the full voltage of the power main is somehow connected to the signal lines. Some suppliers put the barrier in the controller case; others prefer to place a barrier at the location where the signal lines leave the control room. Also, distinct mechanical separation of energy levels in the controller will add to its intrinsically safe properties.

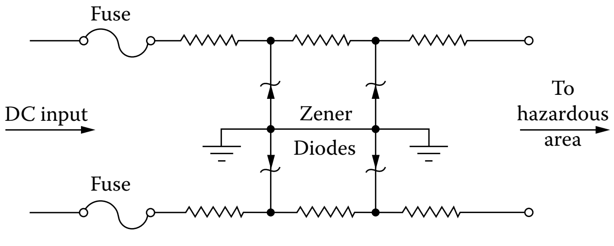


FIG. 4.4p
Barrier circuit approved in Great Britain.

Servicing

The electronic controller frequently is the first to receive the blame for loop malfunctions because its panel tends to give the first indication that something is wrong. Before removing the controller, however, the user should check other loop components and the operation of main plant equipment and should also make sure that the controller output is not shorted to ground in the field.

If the controller is at fault, the course of action depends on the manufacturer's provisions for emergency servicing. In many cases the manual section of the controller will continue to work, so the operator can set the controller output at a desired DC current value. But the instrument maintenance technician must somehow remove the automatic section to troubleshoot it. The following items highlight the basic approaches taken by manufacturers:

1. Manual and automatic sections are completely independent, plug-in modules. One may be removed without affecting the other. They may even have separate power supplies to ease testing of the faulty unit. Sometimes only the automatic section is removable for emergency servicing.
2. An adjustable plug-in power module takes the place of manual output when the controller is removed from its case.
3. A battery-operated or auxiliary manual station replaces the controller to hold output current. In some cases the substitution can be made without interrupting the signal to the process and (by a balancing procedure) without abruptly changing (bumping) the value of the controller output signal.

A few manufacturers make no provisions for emergency servicing. The faulty unit must simply be replaced regardless of signal interruption and output bumping.

Once the controller or its automatic section is in the shop, the technician can try to isolate the trouble by hooking up suitable input signals and a dummy load to the controller, closing a mock control loop. Some suppliers offer a jig that plugs into the controller to simulate a closed loop. Terminals are provided for input signals and for monitoring the deviation and output signals. Most companies at least provide easily accessible test pins or jacks for checking and calibrating key controller signals. Instruction manuals contain guides and schematics for diagnosing the fault from the symptoms. Multimeters used for voltage signal measurements should have at least 20,000 ohms-per-volt sensitivity.

The best buy is the controller that requires little maintenance and can be quickly replaced and repaired when it does need servicing. All electronic controllers will perform satisfactorily in a control loop, at least in the short run. But those that frequently break down may result in costly process downtime, a factor that tends to override any differences in initial controller purchase and installation costs. Highly modular controller designs and circuit accessibility simplify maintenance. Rugged, high-quality, solid-state circuit designs along with dust-tight encapsulation of key parts, including meter displays, help to prevent controller breakdowns. Silicon transistors are more rugged and reliable than those made of germanium. Much depends on the quality control practices implemented by the particular supplier.

DIGITAL ELECTRONIC CONTROLLERS

Later sections in this chapter will cover a variety of DCS and other computer-based digital control systems. Here our focus in on the first generations of microprocessor-based digital controllers (see Figure 4.4q).

Advantages and Disadvantages

The main disadvantage of microprocessor-based shared controllers is an inherent consequence of their being shared among several control loops. This shared nature results in a potential problem of reduced overall system reliability because "many eggs are placed in one basket."

Single-loop digital controllers have overcome this limitation because they can be dedicated to serve only a single output, while retaining most of the capabilities of shared digital controllers. This gives the instrument engineer a very powerful and flexible tool for fully distributed optimal process control because all control loops can be configured and reconfigured to any desired level of sophistication while using the same controller.

Because the control functions are in software, all the control hardware can be purchased and installed before the control system itself is finalized, or an existing control system can be reconfigured without adding or revising any hardware or wiring.

The capabilities of single-loop digital controllers are very similar to the shared controllers also discussed in other sections of this chapter, and therefore they will be discussed here only briefly, emphasizing those features that are not available in analog controllers.

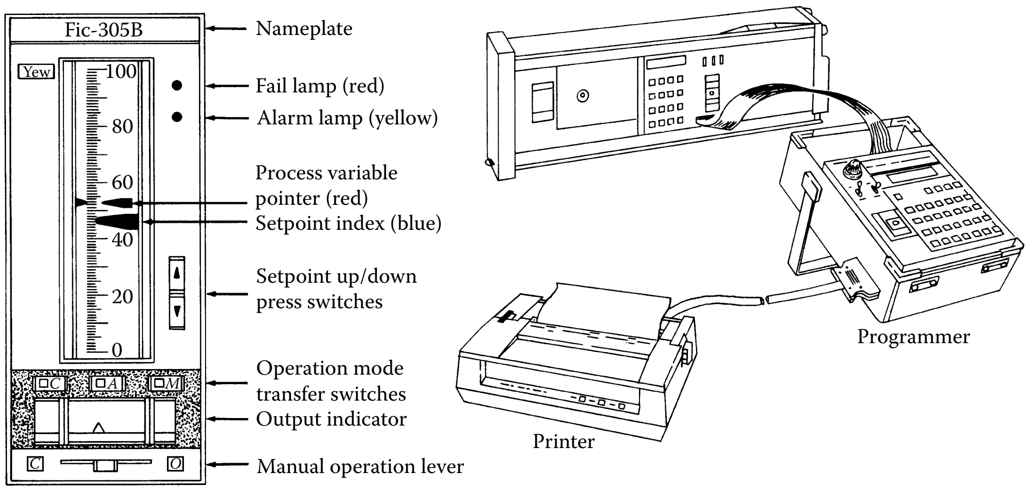


FIG. 4.4q
The faceplates of earlier digital controllers resembled the analog faceplate, and the units were programmed by separate programmers. (Courtesy of Yokogawa Corp.)

Hardware Components

A typical digital controller consists of a D/A and A/D converter, a microprocessor, a display, and control sections, in addition to communications ports (Figure 4.4r). The microprocessor section includes the computational element, the read-only memory (ROM) for program storage, and the operating or random-access memory (RAM).

The A/D converters on the input must have high resolution and high noise tolerance, which can be obtained from voltage-to-frequency converters with counters or from dual-slope integrators. Some controllers operate on pulse train inputs, where the frequency indicates the value of the input. The D/A converters are usually implemented from resistor ladder networks.

The microprocessor calculates the engineering equivalents of set-point, measurement, and output signals and operates the display. The cycle time of the microprocessor varies from 0.05 to 0.5 seconds. During this cycle the controller is not aware of alarm or other conditions that might have evolved until the processor again comes around to evaluate the condition that has become abnormal.

Communications with other instruments are usually handled on an interrupt basis. As digital controllers process data in this "batch" manner, the sequence usually starts with reading all the inputs, linearizing them, and checking them for alarms. This is usually followed by calculating the control algorithm, which can be velocity or a positional configuration (for the actual PID algorithms see Section 2.4 in Chapter 2), and only after that are the displays updated and the other communications outputted.

The potentiometers that are used on analog controllers to set the various alarm points, tuning constants, etc., are replaced by switch matrices, pushbuttons, or keyboards and are retained in RAM on the digital controllers. Provisions are required to retain this memory when the power supply fails,

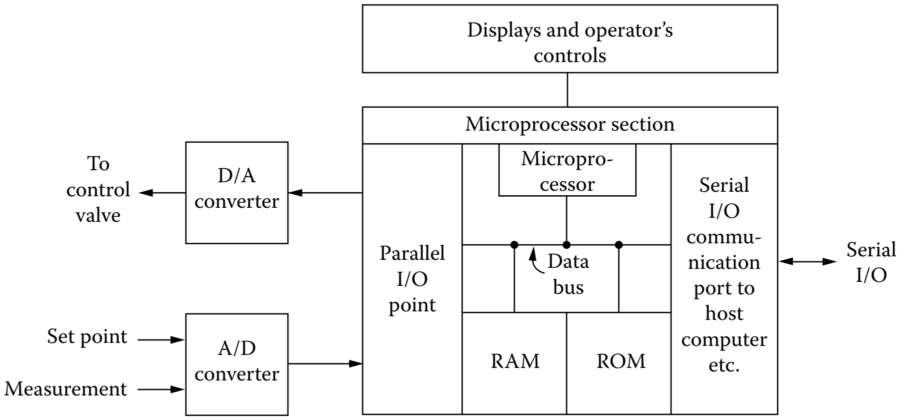


FIG. 4.4r
Block diagram of the hardware components of a typical digital controller.

so that these parameters will not need to be reentered after the power interruption. This is usually accomplished by the use of rechargeable, long-life batteries.

The communication formats used with other digital equipment have not been standardized (RS232, RS422, RS485, IEEE488, etc., may be used), and the coding of the data also varies from manufacturer to manufacturer, using systems such as ASCII, binary, and binary coded decimal. Therefore, special provisions can be needed before digital instruments can be interconnected. International standardization of digital protocols is much needed but is progressing rather slowly. (See Volume 3 of this handbook.)

These communication formats are usually serial in nature to minimize the wiring requirements, they can be unidirectional (single wire) or bi-directional (two conductors), and their data rates vary from kilobauds to several megabauds. In setting up communications with a host computer, the user usually needs to perform programming to code and decode the data communicated by the computer.

Software Capability

A library of control algorithms is available, including many of the more sophisticated functions such as lead-lag, feedforward, dead time compensation, program pattern control, sample and hold, inverse derivative, gap control, auto select or limit functions, error squared, PID, and multiple inputs and/or outputs for cascade, ratio, or batch control. Many designs are able to handle just about all the different types of input signals (including low-level RTD and TC-type measurements with table lookup) and are able to generate any of the various outputs.

Signals can be linearized, pressure- or temperature-compensated, converted to any engineering unit, totalized, time delayed, or arithmetically manipulated.

TABLE 4.4s
*Programming an Auto-Manual PID Controller by Answering a Series of Questions**

Step No.	*Description*	*Parameter*	*Value*
1.	Engineering Units 0% value	EU 0%	–99999 to 99999 (E.U.)
2.	Engineering Units 100% value	EU 100%	–99999 to 99999 (E.U.)
3.	Gain	GAIN	0 to 128
5.	Reset	RESET	0 to 293 (repeats/minute)
6.	Rate	RATE	0 to 895 (minutes)
8.	Process variable filter time constant	PV FTIM	0 to 1.8 (minutes)
9.	Reverse control action	REV ACT?	YES or NO
10.	Increase closes	INC CLO?	YES or NO
11.	Alarm A trip point	ALM A TR	0 to 136% of E.U. span (E.U.)
12.	Alarm B trip point	ALM B TR	−16 to 136% of E.U. span (E.U.)
13.	Alarm C trip point	ALM C TR	−16 to 136% of E.U. span (E.U.)
14.	Alarm dead band	ALM DBND	0 to 136% of E.U. span (E.U.)
16.	Setpoint low limit	SP LO LM	−13.97 to 113.97% of E.U. span (E.U.)
17.	Setpoint high limit	SP HI LM	−13.97 to 113.97% of E.U. Sspan (E.U.)
18.	Valve output low limit	VO LO LM	−13.97 to 113.97(%)
19.	Valve output high limit	VO HI LM	−13.97 to 113.97(%)
20.	Antireset windup low limit	ARW LO LM	−13.97 to 113.97(%)
21.	Antireset windup high limit	ARW HI LM	−13.97 to 113.97(%)
22.	Feedforward gain	FF GAIN	0 to 1
23.	Feedforward reverse action	FF REV?	YES or NO
24.	Feedforward filter time constant	FF FTIM	0 to 112 (minutes)
25.	Track filter time constant	TK FTIM	0 to 112 (minutes)
28.	Restart mode	RST MD	1 to 5
29.	Restart valve output	RST VO	−13.97 to 113.97(%)
30.	Restart set point	RST SP	−13.97 to 113.97% of E.U. span (E.U.)

*Courtesy of Honeywell Inc.

The logic interlocks associated with the loop can also be implemented because many single-loop digital controller designs have both continuous and sequential control features, utilizing all the basic logic functions. These functions are in addition to the standard alarm features.

Programming and Capabilities The programming of a particular controller is usually done by sequentially entering the answers to a number of questions (Table 4.4s). In this example, the process control engineer would, in response to the first question, identify the engineering units of the controlled variable and would indicate the reading that the controller should consider as a 0% reading.

In step 2, the full span (upper range limit = 100%) value would be similarly identified, and in step 3, the gain (proportional band) setting would be entered. As the engineer goes through these steps, the control algorithm and related features are all specified and the controller is ready to take control. In most PID configurations one can vary the basic algorithm (velocity, positional, sample-and-hold, error squared, noninteracting, external-reset, derivative and/or proportional, acting on measurement instead of error, nonlinear, etc.)

In addition to configuring the control algorithm (Table 4.4s), one might also provide linearization functions (such as square root extraction, TC, RTD, or irregularly shaped tanks), counting or totalization, batching, or other types of logic and safety interlock functions. Some digital controllers also offer multivariable control or self-tuning and self-optimizing capability.

In general, digital controllers are much more flexible in tailoring the controller to the particular application than are analog units. Table 4.4t gives a summary of the capabilities of a particular single-loop programmable controller.

TABLE 4.4t
*Features and Capabilities of a Programmable Single-Loop Digital Controller**

Item	*Specifications*	
Control	Loop	Single PID loop Cascade loop Autoselector loop Two PID loops
	Control Types	Nonlinear PID control Feedforward control Batch PID control Sample-and-hold PI
	Self-tuning	
	SVF (Set Variable Filter)	
Signal Computation	Four-function math, square root, selector, limiter, characterizer alarms, first-order lag, derivative, dead time, velocity, moving average, timer, program set unit, counter, pulse output, logical calculations, comparison, branch, subprogram, switch, stack-register operations	
Control period	0.05 sec, 0.1 sec	
User program	Main program and subprogram, 200 steps total	
Analog signals	Inputs	5 points (1- to 5-V DC; direct input (mV/RTD/pulse etc.) possible for one point)
	Outputs	3 points (1- to 5-V DC; two 4- to 20-mA)
Status signals	Inputs Outputs	6 points (user-selectable as inputs or outputs)
Communications	Communicates with DCS or via RS-485) PC	

*Courtesy of Johnson Yokogawa Corp.

Faceplates and Programmers

Loop configuration and programming are usually done through pushbutton operations. The problem of maintenance and spare parts storage is minimized by the complete interchangeability between units. Single-loop controllers can be integrated into large systems through data bus communications between them (networks and buses are covered in Volume 3 of this handbook), and they can also be connected to CRTs for display purposes.

The data highway can also be used to provide a communication link for the plantwide optimizing computer, which can reset the individual loops as required. Such remote signals can not only modify set points, but can also initiate auto/manual transfer, disable displays for information security, switch set points between local and remote sources, change controller actions, etc. In addition, some of the more advanced units also provide self-calibrating, self-diagnostics, and self-tuning functions.

Evolution of Faceplate Designs The earlier designs of digital controllers attempted to imitate the appearance of analog controller faceplates (Figure 4.4q) in order to gain better operator acceptance. For programming these early digital controllers a separate programmer unit was also required, and this device had to be reconnected whenever reprogramming was needed.

Later, in many designs the D’Arsonval meter movement was replaced by digital displays using gas discharge, light emitting diode (LED), or liquid crystal bar graphs. These “quantitized” images convey the innate value of the variable, similarly to analog displays, but have the added capability of flashing to gain the attention to the operator.

Figure 4.4u illustrates the faceplate of a more recent digital controller design in which bar graphs are augmented with digital readouts and the display itself is a bitmap LCD. The programming of most of the more recent digital controllers can be performed through a keyboard, which is integral with the controller. Some of the older designs still require separate plug-in programmers or can be integrated into a DCS system and programmed through the keyboard of the central console, as discussed in later sections of this chapter.

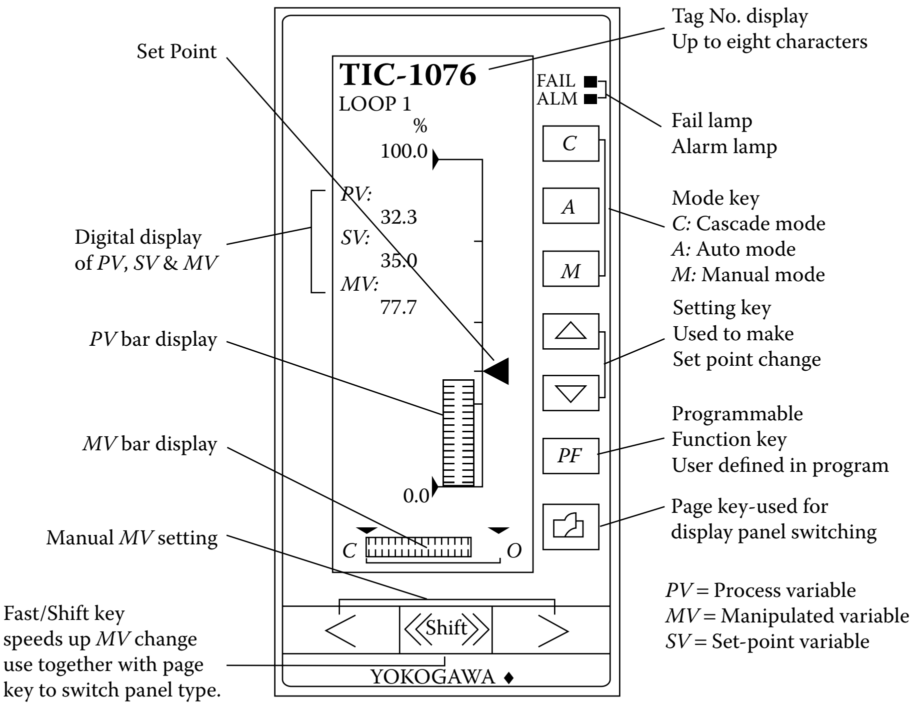


FIG. 4.4u

Faceplates of digital controllers changed in the mid-1990s, when these units became directly programmable. (Courtesy of Johnson Yokogawa Corp.)

Bibliography

Andreiev, N., "A Process Controller that Adapts to Signal and Process Conditions," *Control Engineering,* December 1977.

Andreiev, N., "A New Dimension: A Self-Tuning Controller that Continually Optimizes PID Constants," *Control Engineering,* August 1981.

Astrom, K. J., *PID Controllers: Theory, Design, and Tuning,* 2nd ed., Research Triangle Park, NC: ISA, 1995.

Beard, M. J., "Analog Controllers Develop New Wrinkles," *Instruments and Control Systems,* November 1977.

Buckley, P. S., *Techniques of Process Control,* New York: John Wiley & Sons, 1964.

Butterworths, W. R., *Advanced Process Control,* New York: McGraw-Hill, 1981.

Corripio, A. B., *Tuning of Industrial Control Systems,* Research Triangle Park, NC: ISA, 1990.

Goldfeder, L. B., and Deuschle, R. C., "Analog Control Architecture: Integral vs. Split," *InTech,* April 1977.

Harriott, P., *Process Control,* New York: McGraw-Hill Book Co., 1964.

Huang, N., and Shah, S. L., *Performance Assessment of Control Loops,* London: Springer-Verlag, 1999.

Johnson, C. D., *Process Control Instrumentation Technology,* latest edition, New York: Wiley & Sons, 1977.

Jones, B. E., *Instrumentation, Measurement, and Feedback*, New York: McGraw-Hill Book Co., 1977.

Kaminski, R. K., and Talbot, T. E., "Electronic Controllers for Process Industries," *InTech,* October 1968.

Korn, G., *Interactive System Simulation,* New York: McGraw-Hill, 1989.

Lipták, B. G., "How to Set Process Controllers," *Chemical Engineering,* November 23, 1964.

Luyben, W. L., "A Proportional-Lag Level Controller," *Instrumentation Technology,* December 1977.

Luyben, W. L., *Plantwide Process Control,* New York: McGraw-Hill Publishing Co., 1998.

MacFarlane, A. G. J., and Hung, Y. S., "Quasi-Classical Approach to Multivariable Feedback Systems Design." 2nd IFAC Symposium, Purdue University, September 1982, pp. 39–48.

McCauley, A. P., Jr., and Persik, S.R., "From Linear to Gain-Adaptive Control ... A Case History," *Instrumentation Technology,* November 1977.

Merrit, R., "Electronic Controller Survey," *Instrumentation Technology,* June 1977.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Ruel, M., "How Valve Performance Affects the Contol Loop," *Chemical Engineering,* Vol. 107, No. 10, 2000.

Ruel, M., "PID Tuning and Process Optimization Increased Performance and Efficiency of Paper Machine," Presented at 87th Annual Meeting, PAPTAC, February 2001.

Seborg, D., et al., *Process Dynamics and Control,* New York: Wiley, 1989.

Shah, S., et al., "Multivariable Controller Performance Analysis: Methods, Applications, and Challenges," in Preprints of Chemical Process Control — VI, Tucson, AZ, 2001.

Shinskey, F. G., "Control Topics," Publ. No. 413-1 to 413-8, Foxboro Co.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill Book Co., 1996.

Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: John Wiley & Sons, Inc., 1985.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, Inc., 1984.

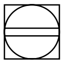
Tsuda, K., et al., "New Single Loop Controller," 1992 ISA Conference, Paper # 104, Houston, October 1992.

4.5 CRT Displays

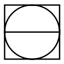
N. O. CROMWELL,
D. D. RUSSELL (1972, 1985)

B. G. LIPTÁK (1995, 2005)

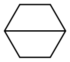
General symbol for recorder or printer. "XX" designates the measured variable and modifier. The letter K is substituted for R to indicate control station.



An auxiliary operator's interface device for distributed control, normally panel-mounted with an analog faceplate. Normally not mounted on main operator console. Can be a backup controller or manual station.



Distributed control/shared display for an indicator/controller/recorder or alarm point, usually for a video display. Access is limited to the communication link.



This is a computer symbol, usually a video display, normally accessible to the operator, which is an indicator controller/recorder or alarm point in a distributed control scheme.

Flow sheet symbols

Designs:	Monoaccelerator and postdeflection accelerator (PDA) tubes
Types:	Storage tubes, character generation tubes, refreshed raster (TV) scan or refreshed X-Y position designs
Operating Voltages:	3 to 4 kV for monoaccelerator and 10 to 14 kV for PDA designs
Screen Size:	Standard sizes are 14, 17, 19, 20, and 21 inches, but units are available from 6× 8 in. (150 × 200 mm) to 23 × 30 in. (575 × 750 mm).
Refresh Rate:	40 to 60 Hz
Character Capability:	500 to 4800 characters
Characters per Line:	64 to 128 characters
Number of Character Lines:	12 to 74 lines
Character Set:	64 to 96 characters
Vector Modes:	Relative or absolute, or both
Vector Capability:	500 to 5000 per frame
Cost:	$3,000 to $10,000, higher for multiple CRT workstations
Partial List of Suppliers:	ABB Group (www.abb.com) Aydin Displays Inc. (www.aydindispalys.com) Emerson Process Management (www.emersonprocess.com) Fisher Controls International Inc. (www.fisher.com) Foxboro-Invensys (www.foxboro.com) GE Fanuc Automation (www.gefanuc.com) Honeywell Automation and Control (www.honeywell.com/acs) Nematron (www.nematron.com) Ronan Engineering (www.ronan.com) Rosemount Inc. (www.rosemount.com) Siemens Energy & Automation (www.sea.siemens.com) Toshiba International Corp. (www.tic.toshiba.com) Westinghouse Process Control (www.westinghousepc.com) Xycom Automation Inc. (www.xycom.com) Yokogawa (www.yokogawa.com/us)

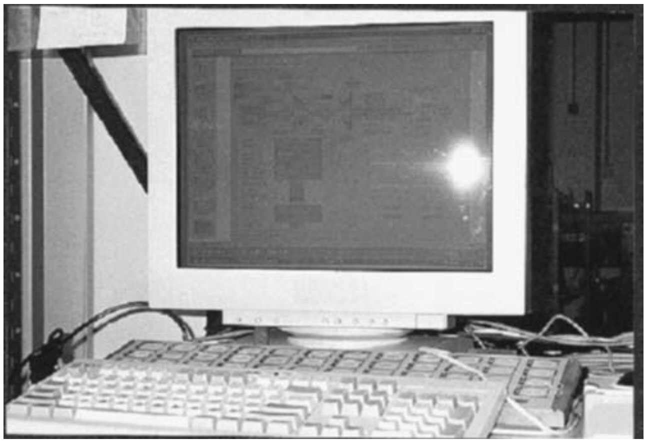


FIG. 4.5a
Workstation with single CRT and keyboard.

INTRODUCTION

The cathode ray tube (CRT) has become an important component in all workstations, DCS-based process control systems, and other human-machine interfaces (HMI). The various workstation, HMI, and control center designs have been discussed in detail in Chapter 3 of Volume 3 of this handbook. DCS systems in general and DCS graphics in particular are discussed in this chapter. The topic of this section is the CRT tube itself.

The CRT tube belongs to the family of emissive displays. Another main family of displays is the passive displays, such as liquid crystal devices. Other emissive displays include incandescent lamps, gas discharge lamps, electroluminescent devices (such as LEDs), and cathodoluminescent displays, including vacuum fluorescent devices.

The first CRT tubes were used on oscilloscopes and in industrial television applications. When, in the 1970s, large and reliable color tubes became available, the CRT tube became the prime means of process display in DCS systems (Figure 4.5a). Compared to the "nonvideo" type displays, its main advantage is its compatibility with computers and with digital electronics. The CRT can serve as an indicator, recorder, alarm, monitor, or logger; it also can provide instructions, display trends, or store data. Most DCS systems (Figures 4.5b and 4.5c) utilize several CRT displays.

DISPLAY OPTIONS

Some of the available display categories are also shown in Section 4.6. In addition, other types of displays can provide management with status reports, provide diagnostic displays to assist plant personnel in their troubleshooting efforts, provide coordinated and high-resolution bar graphs for analog readings, or display status or alarm conditions on a display whose background is a diagrammatic overview of the plot plan of the plant. Figure 4.5d shows such alarm displays.

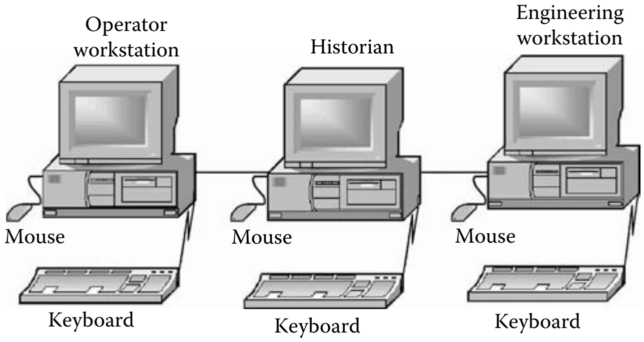


FIG. 4.5b
Intelligent workstations allow for each station to perform its own task while sharing the same database.

When such a display is used to report the safety status of the plant, the nature of the problem is identified by the color of a flashing square (Figure 4.1a), while the location is made obvious by the plot plan of the plant in the background. These types of displays are quickly and easily comprehended and thus are useful for conveying safety-related information. Such displays are often used in addition to, but not in place of, conventional annunciators.

In some applications it is useful that CRTs can generate three-dimensional displays that can be rotated and viewed from all sides. The CRT display can be used in the graph mode, interrelating two or more variables with each other (*X*–*Y* recorder) or with time.

As larger and more complex plants are built with central control rooms containing greater numbers of instruments, the man–process communication problems grow. When digital computer control was first applied in the process industry, alphanumeric control panels were used in addition to the conventional analog instrumentation displays. With these systems the user was able to display and manipulate the control loop parameters (usually, however, only singly).

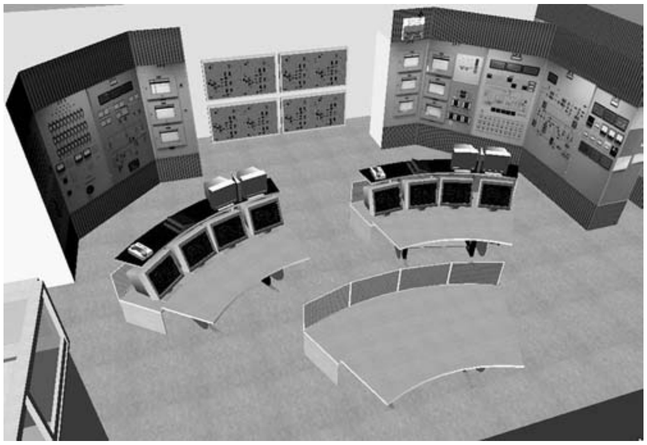


FIG. 4.5c
New workstations are often added to existing control rooms with analog instrumentation.

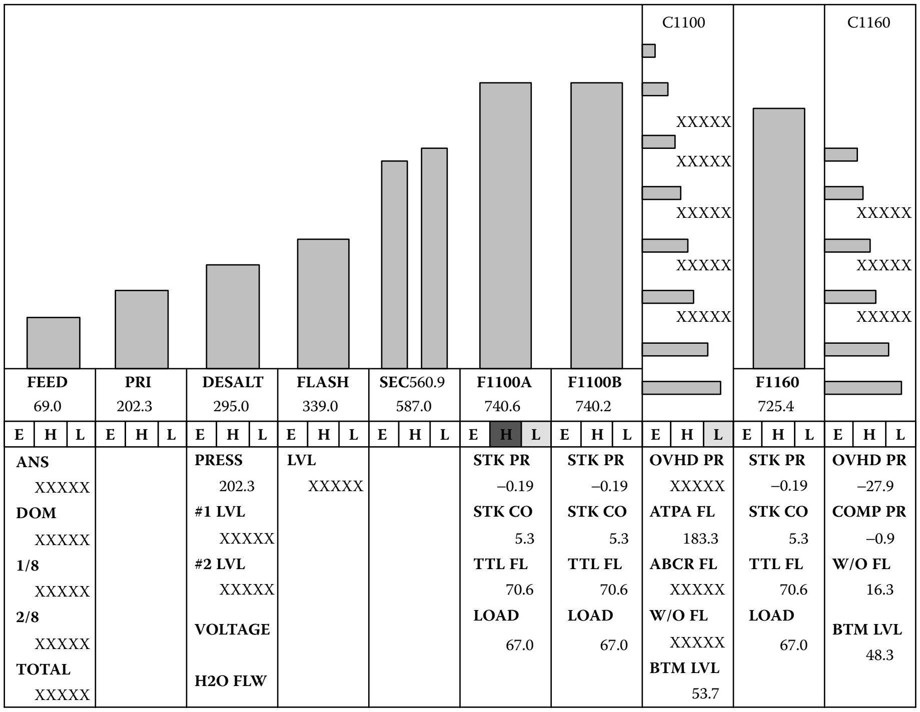


FIG. 4.5d
Integrated CRT display of alarm status overview.

More recently, the man–process communication requirement has been expanded to include not only the needs of the process operator but also all communication between the process and plant personnel. Table 4.5e defines seven levels of communication in a process plant involving process operating, engineering, programming, and management personnel. A process operator, for example, would use information from levels 1 and 3, an instrument engineer would require information from level 2, and a systems engineer would require information from level 4. A manager needs information from level 7.

TABLE 4.5e
Communication Levels in a Process Plant

Level 1:	*Emergency Indicators and Alarms*	Includes both indicators and safety alarms that warn of impending difficulty; assists operator either in moving process to a safe operating point or in shutting it down.
Level 2:	*Component Diagnosis and Maintenance*	Includes information required to diagnose plant and computer system component failure; makes maintenance checks and assists maintenance personnel in finding the corresponding faults.
Level 3:	*Process Operation Information*	Includes all information required by operating personnel to keep the plant running safely and as close to economic optimum as possible.
Level 4:	*Process Evaluation and Diagnosis*	Includes all information needed to determine how well the process is operating, to investigate potential technical or efficiency problems, and to diagnose rapidly complex process failures when they occur.
Level 5:	*Process Supervision Information*	Includes plant parameters that affect overall economy and efficiency; e.g., information on current schedules, feedstock availability and quality, utility usage, and product qualities and costs required to make day-to-day or minute-by-minute adjustments of operating conditions to achieve optimum plant operation.
Level 6:	*System Maintenance and Improvement Information*	Includes program information needed to derive the most from the computer system and to make online system modifications as better operating methods are developed by plant personnel.
Level 7:	*Process Accounting and Scheduling*	Includes information on quantities of production, feedstock supplies, shipping, and labor to assist in establishing production schedules.

THE TOTAL SYSTEM

Although there are multichannel bar graph instruments utilizing CRTs for displaying analog inputs, most CRTs are used in DCS-controlled plants and are operated by digital logic devices. An overall block diagram of a typical digital control system is shown in Figure 4.5f.

The CRT hardware is contained in the two consoles and consists of a CRT display, a keyboard (containing alphanumeric, functional, and cursor control keys), alarm light switches, a refresh memory, an alphanumeric character generator and format control, and associated control logic. A vector generator can be supplied (optional) for graphic displays. Figure 4.5g illustrates the interconnections within the CRT hardware components.

The digital computer memory stores the operating system and data lists. An auxiliary bulk storage device (drum or disk) is sometimes used to store additional programs and data files, and the computer uses a priority interrupt scheme and two bidirectional information channels for communication with other devices. One of these channels, commonly referred to

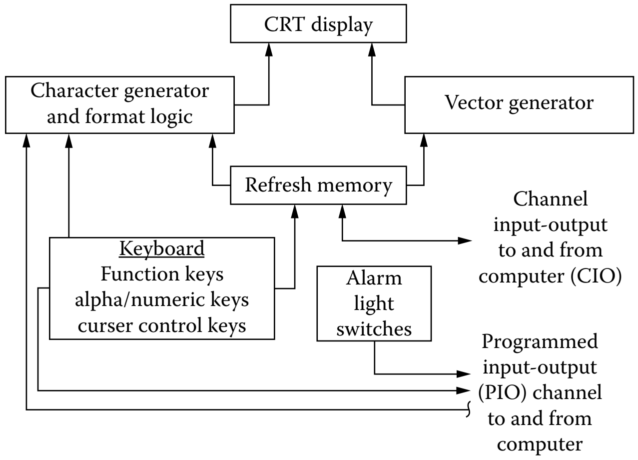


FIG. 4.5g
The roles and interconnections within the CRT hardware components.

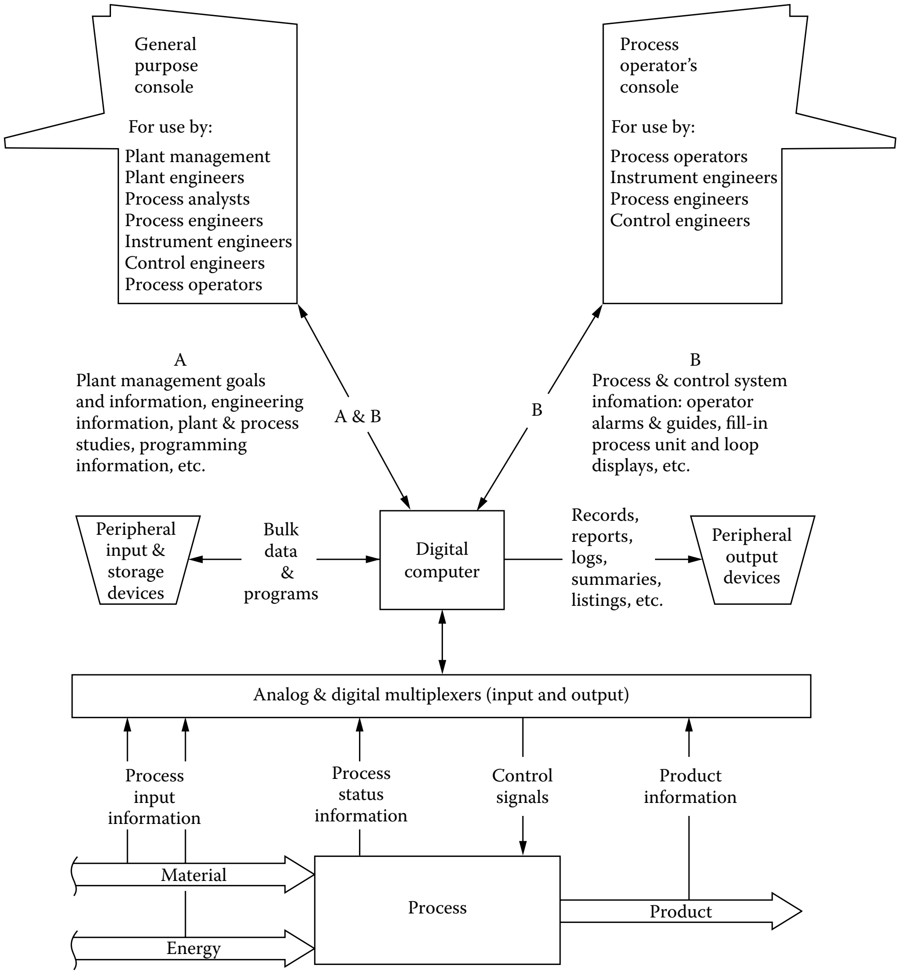


FIG. 4.5f
The components of a plantwide computer control system with CRT display.

as the programmed input–output (PIO) channel, transfers control information and single data words to and from a specified register (usually an accumulator) in the computer. The second channel transfers multiple words or blocks of data to and from the computer memory and is usually referred to as the direct memory access (DMA) channel, or simply the channel input–output (CIO).

The process control program requires a list of control tasks to be performed at specified intervals. These tasks include acquiring process data through analog and digital inputs and computation of appropriate control or alarm actions, or both, based on the input data. The task list contains the necessary data for the process control programs to carry out the desired control operations. These data include input and output addresses, constants, point names, digital status information, and current input–output values. This task list is called the *process database* and is the prime source of process data for display.

Servicing console data transmission requests from programs operating in the central processing unit and handling console keyboard interrupt requests are functions of the real-time executive system. Programs performing tasks requiring data to or from a process display console pass their requests for service to the real-time executive function programs. The calling programs receive acknowledgment of successful completion of the requested operation or indicators describing an aberrant condition.

Data Display Options

Cathode ray tubes can display large amounts of information and can be selective in displaying only relevant parameters or complex relationships between parameters. By fully exploiting the alphanumeric and graphic capabilities, the CRT is more efficient and economical than other methods of data display.

Several choices of CRT implementation include a storage tube display, a raster (TV) scan display, and a random *X–Y* positioned, refreshed display. What follows is a description of the random *X–Y* positioned, refreshed display.

X-Y Positioned Displays The size of the usable display area, and hence the size of the CRT, is determined primarily by the size of the character, the number of characters per line, and the number of lines required. A secondary consideration is the required amount of graphic display. A typical display of a process plant unit is shown in Figure 4.5h. As discussed in Section 4.15, in order that this display be legible from a distance of 5 to 10 feet, the character height should be between 3/8 and 1/4 in. (9 and 6.3 mm).

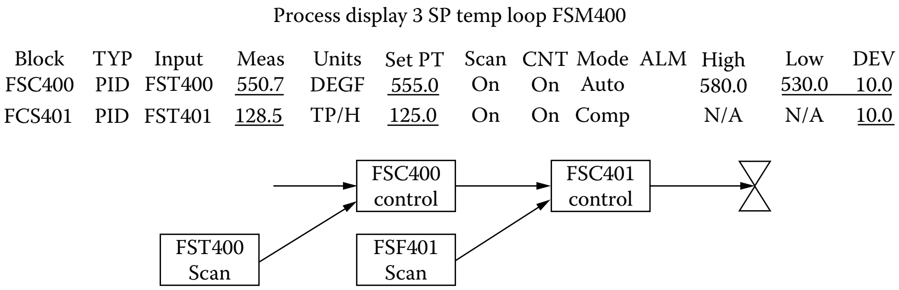
To display the information shown in Figure 4.5h, a character format of 80 to 96 characters per line and 20 to 30 lines are required. These characteristics dictate a diagonal measurement on the CRT of at least 19 or 21 in. (475 or 525 mm). The body of data necessary to initiate a display of the type shown in Figure 4.5h includes:

1. The names of the blocks or loops, or both, to be displayed
2. The format of the display
3. For each block in the display:
    a. The symbolic references to values in the block record, e.g., MEAS (measurement), ABS (absolute)

PROCESS DISPLAY 2 FEED SPLITTER DATE 3-1-93 TIME 1515

											Alarms	
Loop		Block	Input	Meas	Units	Set Point	Output	Scan	CNT	Mode	ABS	DEV
FSM100	→	FSC100	FSP100	340 5	PSIG	300.0	.....	ON	ON	Auto	HI	
				.....		.....	.....					
		FSC101	FSF101	5.8	TCFT/H	5.6	.....	ON	ON	Comp		
FSM300	→	FSC300	FSL300	12.8	FT	13.0	.....	ON	ON	Auto		
				.....		.....	.....					
		FSC301	FSF301	120	TGPH	124	.....	ON	ON	Bkup		
				.....		.....	.....					
FSM400	→	FSC400	FST400	550.7	DEGF	555.0	.....	ON	ON	Auto		
				.....		.....	.....					
		FSC401	FSF401	128.5	TP/H	125.0	.....	ON	ON	Comp		
FSM450		FSC450	FSF450	132.7	TGPH	134.0	.....	ON	ON	Comp		
	→↑	:Cursors		.....		.....	.....					

FIG. 4.5h
Typical CRT display of the process data that relates to the operation of a processing unit.



Process display 3 SP temp loop FSM400

Block	TYP	Input	Meas	Units	Set PT	Scan	CNT	Mode	ALM	High	Low	DEV
FSC400	PID	FST400	550.7	DEGF	555.0	On	On	Auto		580.0	530.0	10.0
FCS401	PID	FST401	128.5	TP/H	125.0	On	On	Comp		N/A	N/A	10.0

FIG. 4.5i
Mixed alphanumeric and graphic display on a CRT.

b. For each symbolic block reference value, its relative or absolute display address
c. Whether or not the value is to be updated in real time, whether or not the operator is to be allowed to modify the value, and how many characters are to be displayed

The mixed display of alphanumeric characters and graphics shown in Figure 4.5i would also fit comfortably on a 21 in. (525 mm) CRT.

Operator Interaction Operator interaction with a process display consists of manipulating the blocks displayed, e.g., changing a block from ON to OFF, and of entering new numerical data, e.g., changing a set point.

Figure 4.5h shows a process display. In the column labeled "set point" there is (for each block) a numerical value, and directly under the value a line of underscores that shows the operator where a new value for the set point is to be entered. The cursor is moved to the underscore field and the new value is entered, using the numerical keys. The underscores are unprotected characters and can be overwritten from the keyboard.

The operator may then either move the cursor to other underscores and enter more data or press the ENTER key, which causes an interrupt to occur in the CPU. In response to the interrupt a program is called for execution that will service the operator's console. The program will in this case read the unprotected characters in the display memory and attempt to modify the appropriate values in the process database block records.

The functional key service program uses data supplied by the display initiator to determine the set of data that goes with the blocks on display. Appropriate visual feedback to the operator is obtained by overwriting the existing value (the one above the entered value) with the new value and restoring the data entry area to unprotected underscores. Should the new value be unacceptable, the underscores are not restored and the offending value may be set to blink. An error diagnostic message may also be displayed.

The operator may accomplish block state changes by pointing the cursor to the block name and pressing one of the measurement and control status function keys. These keys have been appropriately labeled control on, control off, and so forth (see Figure 4.5i). The same interrupt response takes place as already described.

Interrupt Response The program responding to operator requests for service must be able to activate and deactivate function keys. Also, the interrupt-causing keys must be identified. Activating and identifying keys is performed by a key mask table for each console keyboard. It contains the code for each key that is currently active. An active key is one that has been put in the table by the program servicing the console.

Thus, the servicing program can at any time determine the function keys that the operator may legitimately use. The key mask table is used by the console-interrupt-handler segment of the real-time executive to determine whether the servicing program has to be called.

Multiple Workstations If many consoles are operating concurrently, the servicing program attends to each as requests occur. There need be only one servicing program. The information that it requires for each console that has a currently active process display includes:

1. The total number of unprotected characters on display
2. A sequential list of the block name and symbolic value name for each data entry field
3. The length of each data entry field and the display address of its related protected value
4. Access to the same data as the display update program

This set of data is needed to ensure that visual feedback for every requested state change is available. If visual feedback is not possible, the requested state change is erroneous, and an appropriate diagnostic measure is displayed.

Components of a CRT Display In a block diagram of a typical CRT display unit (Figure 4.5j), electromagnetic deflection and low-voltage electrostatic focus maintain display quality at all locations on the CRT screen. P-31 phosphor (green) is usually preferred over P-1 phosphor (white) because it is more durable. The block of input data to the display unit shown in Figure 4.5j (X and Y position data and

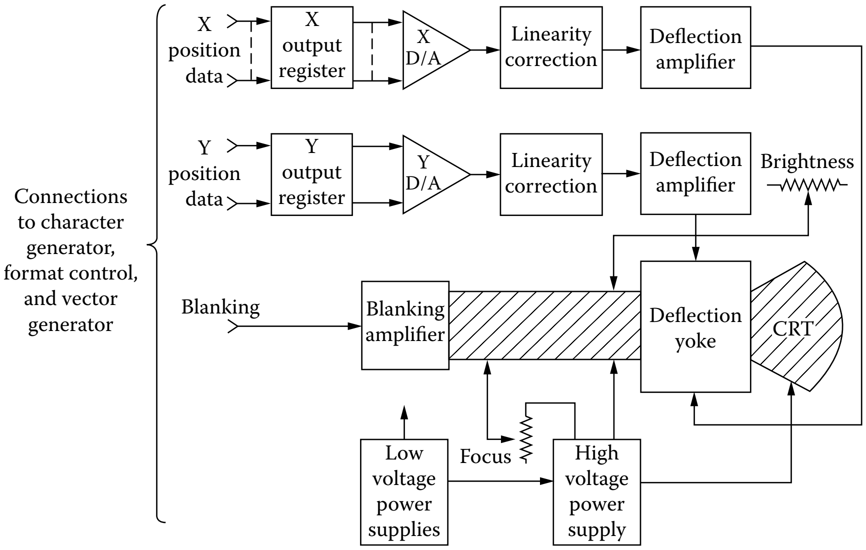


FIG. 4.5j
CRT display components.

blanking) is supplied by a character generator, format control, or vector generator. The data are digital in nature.

The body of *X* and *Y* position data is loaded into output registers connected to high-speed, digital-to-analog (D/A) converters, the output of which drives a linearity corrector and deflection amplifier. The linearity corrector compensates for geometrical distortion in the CRT, and the deflection amplifier must be capable of furnishing as much as 5 amperes of current to the deflection yoke. The blanking amplifier provides a signal to turn the electron beam in the CRT either on or off.

The information supplied to the CRT display unit must be continually repeated or refreshed. So that flickering or a "swimming" effect does not occur on the display screen, the refresh rate should be synchronous with the power line frequency—ordinarily 60 (or 50) Hz.

Keyboard

When a CRT display is used in process control, a pointer, or cursor, is required to indicate the parameter upon which the action is to occur. Cursors (see Figure 4.5h) are manipulated from a keyboard so that they are beneath the line, value, or character to be selected for the next operation.

Cursor Control Cursor control keys (Figure 4.5k) include the four arrow keys (←, →, ↑, ↓) for movement in one of the four primary directions. In addition there is a FAST key to increase the rate of movement, a HOME key to return the cursor to the upper left corner of the CRT screen, and a JUMP key, which will be subsequently explained together with the "protect" feature. The position of the cursor can also be controlled or questioned from the computer.

It is often undesirable to enable a user to modify values or characters on the CRT screen. A "protect" feature protects characters specified by a program in the digital computer from being modified. This feature might be implemented by a bit associated with each character, such that when it is set to a "1" state, the character can be modified only by the computer, not directly by the user. This feature also enables the computer to read selectively only unprotected information in the refresh memory.

The cursor control JUMP key allows the cursor to move from a current position to the next unprotected character following a protected one, thus bypassing (protected) characters that cannot be changed by the user—a very useful feature in a fill-in-the-blanks operation.

A "blink" feature permits individual characters displayed on the CRT to be blinked on and off several times per minute; this is useful for special conditions such as alarm indication. This too is controlled by a bit associated with each character in the refresh memory. Supplying solid, dashed, and dotted lines is useful for graphic displays. For example, a solid line and a dashed line might differentiate between a measurement and a set point when trend information is displayed.

Alphanumeric Keys Alphanumeric keys (see Figure 4.5k) modify or make additions to the display on the CRT screen. Entries can be made only into unprotected locations and are themselves unprotected. The operations are performed by the hardware associated with the CRT and require no response from the computer. The keys resemble those commonly found on a typewriter.

When they are depressed, a code (usually USASCII-8) corresponding to the key legend is entered into a refresh memory location corresponding to one directly above the cursor on the CRT screen, and the cursor is incremented by one location. With the key code entered into the refresh memory, the character is displayed at the corresponding location on the CRT screen.

A depression of the SPACE bar (key) causes a space (blank) character to be entered into the refresh memory and the cursor

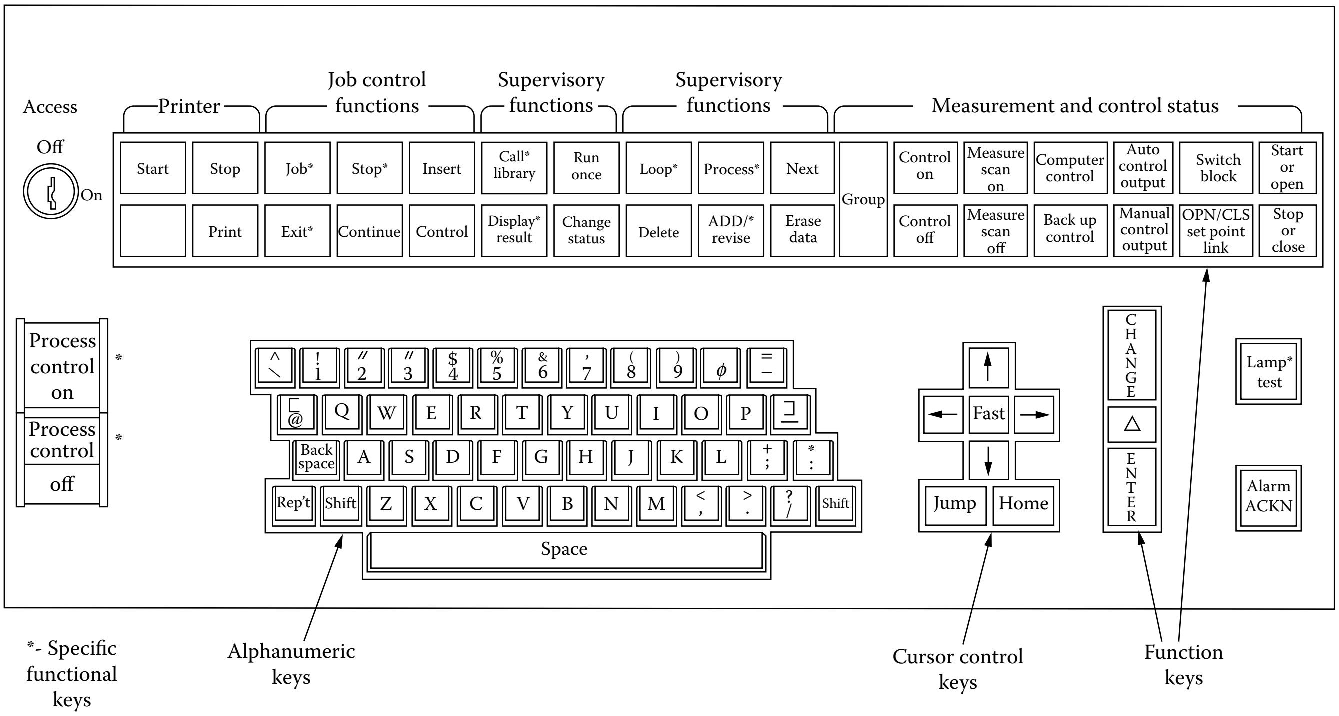


FIG. 4.5k
General-purpose keyboard adapted for process control.

to be incremented by one location. A BACKSPACE key, when depressed, causes a space character to be entered into the refresh memory and the cursor to be decreased by one location. By depression of the repeat key and a character key, the normal operation of the character key is repeated at a predetermined rate.

The function keys (see Figure 4.5k) request a specific action of the digital computer. When a function key is depressed, a priority interrupt signal is sent to the computer. The computer reads a code on the PIO channel corresponding to the depressed key and executes the request, which might be to place all the control loops displayed on the CRT on manual control or to show a directory of the display library on the CRT. In other words, each key requests a unique function that is programmed in the computer.

Alarm Light Switches Alarm light switches operate very much like function keys, with one notable exception—the former are lighted pushbutton switches whose light is controlled either from the computer or by an external (field) contact closure. When depressed, the buttons primarily request new displays; when lighted, they indicate alarm conditions associated with the corresponding display. Depressing an alarm light switch causes a hardware action identical to a function key depression, when the computer program detects conditions that should turn an alarm light on the PIO channel. By setting a unique bit to 1 or 0, one can turn the light on or off, respectively.

DISPLAY CAPABILITIES

Refresh Memory

The refresh memory stores information (in coded form) displayed on the CRT screen. Since the duration of the CRT phosphor is several hundred microseconds, the displayed information must be regenerated and displayed at a nominal rate of 60 times per second. The refresh memory may consist of magnetic or acoustic delay lines, semiconductor shift registers, magnetic cores, magnetic disk or drum, or semiconductor memory cells.

The particular size, organization, and bit coding can vary. It can furnish information to a computer or to a character and vector generator. It can also accept information from a computer and a keyboard.

For example, a refresh memory associated with a 2000 alphanumeric character display (80 characters per line, 25 lines) or with a display having 3000 in. (76 m) of vectors (straight line segments for graphic displays) may consist of semiconductor memory cells, which are organized into 2000 words, with each word containing 12 bits of information. For display generation, each word is sequentially accessed and sent either to a character generator or to a vector generator. For a memory word that stores a character code, the bit structure shown in Figure 4.5l might be used.

The mode bits differentiate among characters and several types of vectors. For example, when the mode bits are

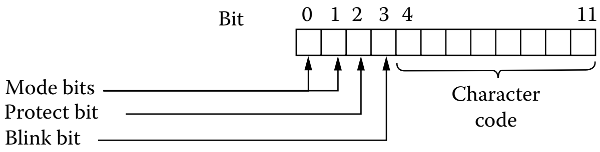


FIG. 4.5l
Bit structure for character code.

logical 00, the word is defined as containing alphanumeric character information. The protect bit determines whether or not the character code can be changed from the keyboard or selectively accessed from the computer. The blink bit determines whether or not the character will blink. The character code defines the alphanumeric character (usually in USASCII code) that will be accessed from this memory location and displayed by the character generator.

If the mode bits of a memory word are logical 10 or 11, the current word and the word in the next memory location are defined as containing either relative or absolute vector information, respectively. A relative vector is a straight line; its origin is the current beam position on the CRT screen and its endpoint is defined as a change in X and Y position with respect to this origin.

An absolute vector is a straight line. Its origin is the current beam position on the CRT screen and its end point is defined as an X and Y position in a fixed grid with the grid origin of $X = 0$ and $Y = 0$ at the lower left corner of the CRT screen. The bit structure in Figure 4.5m might be used to define a vector.

If the mode bits (in word 1) are logical 10, the two words define a relative vector, and therefore the X displacement contains a ΔX value and the Y displacement contains a ΔY value. If the mode bits are logical 11, the two words define an absolute vector, and therefore the X displacement contains an X value and the Y displacement contains a Y value. The line type determines whether the vector to be generated will be a solid, dashed, dotted, or invisible line (blanked movement).

Character and Format Control

Alphanumeric characters may be generated by means of several techniques. Analog stroke, "race-track," character mask scanning, and read-only memory character generation are a few examples. The following example is based on read-only memories.

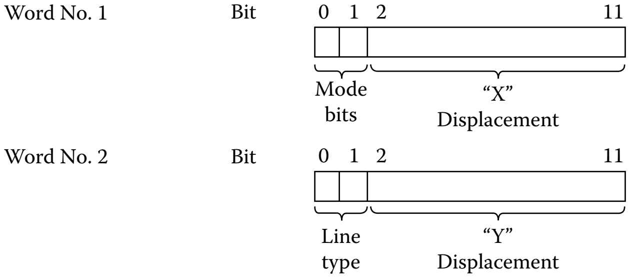


FIG. 4.5m
Bit structure for vector code.

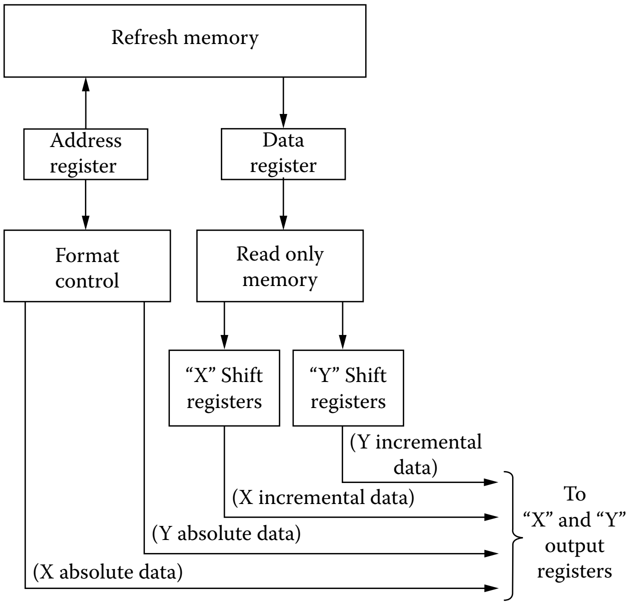


FIG. 4.5n
The logic of generating characters and controlling the format.

Figure 4.5n illustrates a character generator and format control logic. Since there are 2000 memory locations containing character codes for each of 2000 character positions on the CRT screen, the value contained in the refresh memory address register (Figure 4.5n) must be unique for each character position.

The refresh memory is accessed at the location specified by the contents of the address register, and data are loaded into the data register from this location. The format control accepts the contents of the address register as an input and generates absolute values of X and Y data, which positions the CRT beam to the starting position of the appropriate character location on the CRT screen.

The contents of the data register are then used as an address for the read-only memory, and the body of X and Y data is accessed and loaded into the appropriate shift registers (see Figure 4.5n). As the information is shifted out of these registers bit by bit, it is decoded and sent to the X and Y output registers (see Figure 4.5j). This mass of decoded data causes the output registers to increment (or count) up or down, which in turn causes the appropriate character (specified by the contents of the data register) to be written on the CRT screen.

Vector Generator

Vectors are straight line segments used to construct graphic displays. Typical methods of vector generation use analog ramp generators, binary rate multipliers, or digital arithmetic units. The example to be described uses digital arithmetic units.

The vector generator receives data from the refresh memory (see Figures 4.5m and 4.5n) and based on this body of data provides incremental data to the X and Y output registers (see Figure 4.5j).

When the vector generator receives the X and Y displacement information from the refresh memory, if the vector was specified as a relative vector, the information will consist of a ΔX and a ΔY value. The vector generator operates directly on this body of data. If the information received from the refresh memory has been specified as an absolute vector, an auxiliary operation of computing ΔX and ΔY will occur, by subtraction of the current beam coordinates (X and Y values) from those obtained from the refresh memory.

The operation (algorithm) of the vector generator is such that an incremental step (or movement) is made to minimize the value of ΔX and ΔY. A new value of ΔX and ΔY is then computed, and another incremental step is made to minimize the value of ΔX and ΔY. This process is repeated until the computed values of ΔX and ΔY are both zero. The result is a best fit to the desired straight line segment displayed on the CRT screen. The solid, dashed, or dotted lines are generated by turning the CRT beam on or off (blanking) at desired intervals.

Display Initiation

Process display initiation comprises a chain of events that begins with an operator key action and ends when the selected display has been transmitted to the console refresh memory and real-time update has commenced. From the operator's point of view, one or more key actions are required to fetch a display. From the point of view of program or software, these key actions identify what the operator wants to see.

The operator must have a method of observing both the process variables and the response to actions taken by the control programs in the computer. The operator also needs to be notified of alarm conditions and requires a method of communicating directives to the process control program.

To accomplish these objectives the process display programs must allow the operator to:

1. Initiate process data display requests, which will be updated to reflect process variable changes
2. Manipulate reference values and states (on/off or automatic-manual) of block or loop records in the process database
3. Terminate a process display
4. Request other relevant programs, such as directories or plant efficiency calculations
5. Respond directly to an alarm

Function Keys Inherent in each of the operations just mentioned is the process display console keyboard. Considerations of key sequence include:

1. How much data must be entered from memory
2. How many keystrokes are required to achieve a display response
3. How many keystrokes are required to recover from a data entry error
4. How many operator decisions (choices) are required to proceed through a desired sequence

The function keys (see Figure 4.5k) may be divided and arranged in groups as shown. From the point of view of software, the keys are also arranged by purpose. Keys supplying a constant response can be grouped by key code. All other keys are conditioned response keys. It is useful to think of these two groups as specific (constant-response) keys and conditional (sequence-dependent) keys. Alarm key lights are specific keys. Conditional keys manipulate process reference values, control states, and select data from a recipe.

Specific functional keys are indicated by an asterisk in Figure 4.5k; all other keys are conditional. Although in general it is desirable to minimize key operations that serve to initiate process displays, it does not always follow that every action that an operator might take should have minimum key activity. On one hand, operations such as modification of numerical values may require visual verification before entry into the process database is attempted. On the other hand, state changes of process data blocks or loops should require minimum key actions. Thus, the design of operator key activity and key sequencing must be related to the display tasks and to the keyboard design.

Key Sequencing The key sequence for any operation may be constructed as follows: First a specific key is used. This produces a fixed (by key code) visual response. Operator data are entered by the alphanumeric keys followed by a conditional key. The alphanumeric keys transmit data only to the display memory rather than to the central processing unit (CPU). The cursor is manipulated by the operator to enter data at appropriate display locations. Key sequence diagrams are useful in planning process display–process operator interaction.

Figures 4.5o and 4.5p show two typical sequences. The alarm key sequence (Figure 4.5o) is used when an alarm key light comes on owing to a process upset. The operator presses the key, initiating a process unit display (see Figure 4.5h). The process loop display sequence (Figure 4.5p) requires entry of data before the loop to be displayed can be selected.

The sequence begins with a specific key operation (loop key) and continues through entry of data (Figure 4.5q) and initiation of loop display (see Figure 4.5i). These examples are initiating sequences. Operator interaction with a live loop or process display is a continuation of the techniques described, using conditional keys.

Format Construction Display formats consist of the fixed or static information (column titles, headings, operator instructions, and recipes) and the address for each piece of data to be retrieved from the process database, displayed, and appropriately updated on the display. Static format data may be conveniently separated from the process database-related information, allowing independent modification of titles and headings.

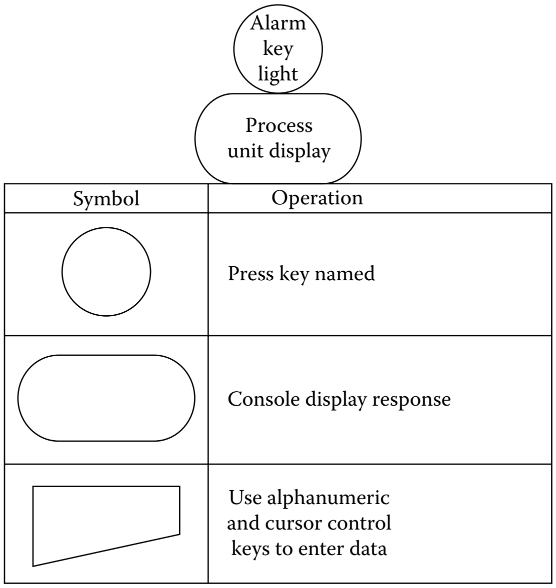


FIG. 4.5o
Alarm key sequence: When a process upset causes an alarm key light to actuate, the operator presses it, which in turn initiates the process display shown in Figure 4.5h.

The references to process database information should be symbolic. Usually, the process database is referenced by block or loop name, which points to a complete set of measurement and control data about one process control input or output or both. Within this set of data the references to particular information, such as a set point or measurement, should be symbolic.

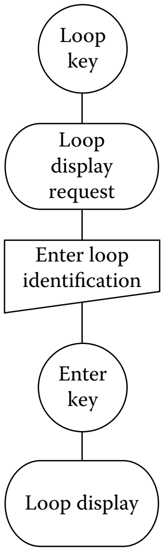
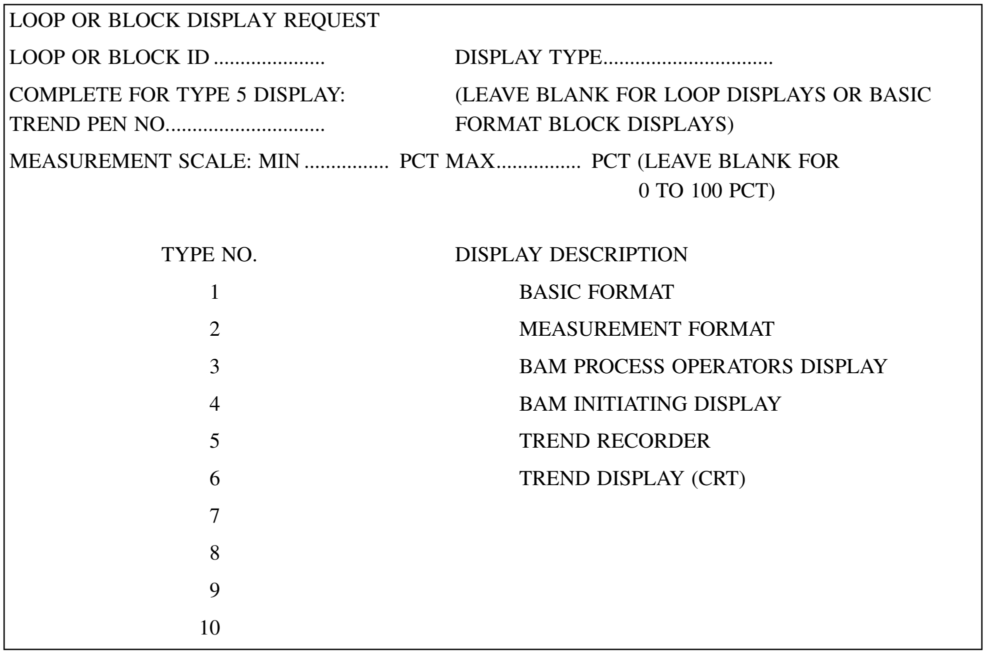


FIG. 4.5p
Process loop display sequence: before the loop display in Figure 4.5i can be displayed, the operation of the loop key is required.

Symbolic references to data items in each block record simplify the display-initiating program. The references are passed as arguments to a subroutine set that locates the appropriate item and performs internal-to-external format conversions.

Process Data Retrieval The CRT-based process display allows considerable flexibility about what process data are to be shown. Various special-purpose displays may be constructed



LOOP OR BLOCK DISPLAY REQUEST

LOOP OR BLOCK ID .... DISPLAY TYPE...............................

COMPLETE FOR TYPE 5 DISPLAY: (LEAVE BLANK FOR LOOP DISPLAYS OR BASIC
TREND PEN NO............................. FORMAT BLOCK DISPLAYS)

MEASUREMENT SCALE: MIN .... PCT MAX................ PCT (LEAVE BLANK FOR 0 TO 100 PCT)

TYPE NO.	DISPLAY DESCRIPTION
1	BASIC FORMAT
2	MEASUREMENT FORMAT
3	BAM PROCESS OPERATORS DISPLAY
4	BAM INITIATING DISPLAY
5	TREND RECORDER
6	TREND DISPLAY (CRT)
7	
8	
9	
10	

FIG. 4.5q
Loop display request.

PROCESS DISPLAY: 8 PLANT FLOW MONITOR DATE: 3-1 TIME: 1515

Point ID	Description	Status	Value	Alarm
FSF101[10]	Plant 2 feed flow	On	5.8	TCFT/H
FSF301	Splitter flow to heater	On	120.8	TGPH
FSF401	Splitter steam flow	On	128.5	TP/H
FSF450	Acc flow to splitter	On	132.7	TGPH
HRF502	Heater fuel flow	On	2500	CFT/H
PFF600	Product A flow	On	60.7	TCFT/H
PFF550	Rct. feed flow to fractionator	On	106	TGPH
PFF701	Fractionator steam flow	On	107.4	TP/H
PFF750	Product B flow	On	75.5	TGPH

FIG. 4.5r
Display of the status and readings of all flow monitoring loops in the plant.

to suit individual processes and operating policies. Individual blocks or a single piece of information for several loops (Figure 4.5r), groups of connected blocks (a control loop—see Figure 4.5i) or sets of data related to process unit performance (see Figure 4.5h) may be displayed.

If each of these displays is considered a standard type, one may then construct displays using different sets of block or loop names for each process unit display, the format of which will remain constant—the block or set of loop data, or both, is changed to reflect the set of names chosen. Each set of blocks or loops is referred to by a identity code.

Process Data Display Process unit displays may be automatically shown in response to an alarm key action by construction of a list of alarm key codes and association of a "process unit display" identity code with each. The identity code retrieves the list of block or loop names to be displayed. At the same time that the name list is retrieved, the identity of the format (both the fixed part and the database-related part) is also retrieved. Thus, a process unit display identity is used to point to a predefined display format and a set of process data.

Propagation and Termination

The display-initiating program retrieves the appropriate data for building the display and supplies real-time display control and functional key service data through files or lists to the respective programs. The updating program is responsible for maintaining the displayed measurements and other values in a current or real-time state. The functional key service program is responsible for all operator-requested modifications of the data display. Typically, these changes are of two types: data entry or value manipulation, and state changes (e.g., on/off or automatic/manual).

In Figures 4.5h, 4.5i, and 4.5r, unprotected underscores define the appropriate areas for data entry for the operator. All other data displayed are protected and cannot be modified or changed by the operator at the console. The display-initiating program also must set a flag or bit in each block record requiring update; this bit or flag is referred to as the display capture bit.

Data Capture and Routing In display propagation the values shown are changed to reflect the variations in the controlled process. Measurements, alarm states, internally modified set points, and reference values are examples of data requiring continuous updating. Update frequency may be other than the normal processing interval, which is inconvenient and requires additional program logic.

If the update frequency is the same as the processing (scan) interval, the process control program may be constructed to examine the data capture bit in each block record. When the bit is on, the block record data are set aside in a temporary display file or list. When the process control program has completed its tasks for the current interval, it calls on the display update program for execution. It should be noted that display update is called on only when data have been captured for update.

The display update program finds the block record in the display file or list and with the display control information assembled by the display initiator converts the appropriate mass of data in the captured block record to external format and transmits it to the display.

Typically, for each block record the display update program includes:

1. Block name
2. Display console number if more than one
3. Symbolic data names for all items to be updated
4. Display memory address (where the data are to be displayed)

Terminating the Display After observation and manipulation, the operator indicates that the operation is complete by requesting another display by a specific function key or alarm key light, or both. The console-interrupt-handler segment of the real-time executive determines whether process display termination is necessary by keeping track of real-time update

TABLE 4.5s
Comparison between a 13.5-in. Passive Matrix LCD (PMLCD) and Active Matrix LCD (AMLCD) and a 15-in. CRT Monitor

Display Type	*Viewing Angle*	*Contrast Ratio*	*Response Speed*	*Brightness (foot-lamberts)*	*Power Consumption (watts)*	*Life*
PMLCD	49 to 100 degrees	40:1	300 ms	70 to 90	45	60K hours
AMLCD	>140 degrees	140:1	25 ms	70 to 90	50	60K hours
CRT	>190 degrees	300:1	n/a	220 to 270	180	4 to 5 years

operations on a console-by-console basis. If the current display on a console is not being updated in real time, termination is unnecessary; the requested program is responsible for clearing the display.

The process display termination program determines which blocks in the process database were being "captured" for display on the console and resets or stops their ensuing capture and display by resetting the display capture bit in the block record. The program also purges the data files supplied by the display initiation to the update program. When termination is complete, the operator-requested function is allowed to proceed. If many process displays on many consoles are being updated in real time, the termination must take care not to terminate capture of blocks that are being displayed on other consoles.

CONCLUSIONS

For a detailed discussion of a number of new developments in the field of graphic displays, please refer to Section 4.15. For a comparison between the features and capabilities of a CRT monitor and two LCD monitor designs, refer to Table 4.5s.

Bibliography

Anderson, L., "X-Y Recorders," *Measurements and Control,* October 1992.

Arsonson, R. L., "CRT Terminals Make Versatile Computer Interface," *Control Engineering,* April 1970.

Bailey, S., "Recorders and Indicators Exert a Cohesive Force in Online Quality," *Control Engineering,* January 1991.

Bailey, S. J., "Non-Video Display Rounds Out Vital Man-Machine Interfaces," *Control Engineering,* May 1980.

Compton, K., *Image Performance in CRT Displays*, Bellingham, WA: SPIE Press, 2003.

Davis, R. C., "Vocal Alarms," *Instrumentation Technology,* December 1964.

Ellis, R. K., "Color Graphics CRTs Provide Window into Factory Operations," *Instruments and Control Systems*, February 1983.

Gadberry, B. E., "Designing Integrated Control System Displays," Paper #91-0354, 1991 ISA Conference, Anaheim, CA, October 1991.

Guerlain, S., and Bullemer, P., "User-Initiated Notification: A Concept for Aiding the Monitoring Activities of Process Control Operators," Honeywell Technology Center, 1996.

IEEE 1289, "IEEE Guide for the Application of Human Factors Engineering in the Design of Computer-Based Monitoring and Control Displays," Piscataway, NJ: IEEE Press, 1998.

ISA-TR77.60.04-1996, "Fossil Fuel Power Plant Human–Machine Interface CRT Displays," 1996.

Jutila, J. M., "Guide to Selecting Alarms and Annuciators," *Instrumentation Technology,* March 1981.

Keller, P. A., *Electronic Display Measurement: Concepts, Techniques, and Instrumentation,* New York: John Wiley & Sons, 1997.

Lanier, G. W., "Developing CRT Graphics Displays," 1992 ISA Conference, Houston, TX, October 1992.

Lueder, E., *Liquid Crystal Displays: Addressing Schemes and Electro-Optical Effects,* New York: John Wiley & Sons, 2001.

Lundstrom, J. E., "CRT Terminal Uses Mosaics to Build Process Graphics," *Control Engineering,* February 1981.

MacDonald, L. W., and Lowe, A. C., *Display Systems: Design and Applications,* New York: John Wiley & Sons, 1997.

McCready, A. K., "Man–Machine Interfacing for the Process Industries," *InTech,* March 1982.

Neumann, P. A., "Recorders with Touch-Screen Displays," *Measurements and Control,* February 1993.

NUREG-0700, "Human-System Interface Design Review Guideline," Vol. 1, Washington, DC: The NRC Public Document Room, 1996.

NUREG/CR-6637, "Human Systems Interface and Plant Modernization Process," USNRC, March 2000.

Reinhart, S., "Are Your Control Displays User Friendly?" ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Rumph, P., "Improved Information Interfacing Turns Operators into Process Managers," *Pulp & Paper,* August 1998.

Shirley, R. S., et al., "What's Needed at the Man/Machine Interface?" *InTech,* March 1981.

Smith, G. M., "Data Management Systems," *Measurements and Control,* September 1991.

Stubler, W. F., et al., "Human System Interface Modernization Process," NUREG/CR-6637, Washington, D.C., Office of Nuclear Regulatory Research, 2000.

Technical Report 1001066, "Human Factors Guidance for Digital I&C Systems in Hybrid Control Rooms," Palo Alto, CA, EPRI, November 2000.

Thumm, M., et al., "Displays and Vacuum Electronics," April 29–30, Garmisch-Partenkirchen, VDE-Verlag, Berlin, 1998.

4.6 DCS: Basic Trends and Advances

B. G. LIPTÁK (2005)

Hardware Suppliers:

DCS Packages:	ABB (www.abb.com)
	Allen-Bradley (www.ab.com)
	Emerson (www.EasyDeltaV.com)
	Honeywell (www.honeywell.com)
	Invensys (www.invensys.com)
	Schneider (www.schneider.com)
	Siemens (www.siemens.com)
	Yokagawa (www.yokogawa.com/us)
Touch Screen Displays:	3M United States (www.3m.com)
	Elo TouchSystems Inc. (www.elotouch.com)
	GVision Inc. (www.gvision-usa.com)
	Posiflex Inc. (www.posiflexusa.com)
	Touch International Inc. (www.touch-international.com)
Annunciator and Safety Packages:	4B Components Ltd. (www.go4b.com)
	Acromag Inc. (www.acromag.com)
	Adaptive Micro Systems Inc. (www.adaptivedisplays.com)
	Ametek Power Instruments (www.ametekpower.com)
	Barnett Engineering Ltd. (www.barnett-engg.com)
	CAL Controls Inc. (www.cal-controls.com)
	CEA Instruments Inc. (www.ceainstr.com)
	Cole-Parmer Instrument Co. (www.coleparmer.com)
	CTC Parker Automation (www.ctcusa.com)
	Draeger Safety Inc. (www.draeger.com/gds)
	Druck Inc. (www.pressure.com)
	Fisher Controls International Inc. (www.fisher.com)
	Foxboro Co. (www.foxboro.com)
	General Monitors (www.generalmonitors.com)
	Graybar Electric Co. (www.graybar.com)
	Honeywell Industry Solutions (www.iac.honeywell.com)
	ImageVision Inc. (www.imagevisioninc.com)
	Matrikon Inc. (www.matrikon.com)
	Metrix-PCM/Beta (www.metrix1.com)
	Moore Industries (www.miinet.com)
	North American Manufacturing (www.namfg.com)
	Oceana Sensor (www.oceanasensor.com)
	Phonetix Inc. (www.sensaphone.com)
	Powers Process Control (www.powerscontrols.com)
	Precision Digital Corp. (www.predig.com)
	ProSys Inc. (www.prosysinc.com)
	Raco Manufacturing and Engineering (www.racoman.com)
	Robicon div. High Voltage Engineering Corp. (www.robicon.com)
	Ronan Engineering Co. (www.ronan.com)
	Schneider Electric/Square D (www.squared.com)
	Scott Aviation div. Tyco Inc. (www.tycoelectronics.com)
	Sierra Monitor Corp. (www.sierramonitor.com)

Thermo Brandt Instruments (www.brandtinstruments.com)
Tips Inc. (www.tipsweb.com)
Transmation Inc. (www.transmation.com)
Vorne Industries Inc. (www.vorne.com)
Western Reserve Control (www.wrcakron.com)
Wilkerson Instrument Co. (www.wici.com)
Zetron Inc. (www.zetron.com)

Networks and Buses:

Most fieldbuses or control system communication systems that have reached the stage of international or industrial standard are supported by user groups who maintain Internet sites as listed below. These sites contain lists of products and suppliers:
AS-Interface (www.as-interface.com)
ControlNet (www.controlnet.com)
Ethernet TCP/IP (www.industrialethernet.com)
Foundation Fieldbus (www.fieldbus.org)
HART (www.hartcomm.org)
MODBUS (www.modbus.org)
PROFIBUS (www.profibus.org)

Software for Tuning:

Tuning Controllers:

ExperTune (www.expertune.com/)
ProTuner (www.protuner.com/)

Self-Tuning Controllers:

ABB Process (www.abb.com/us/instrumentation)
Asea Brown Boveri (www.abb.com)
CAL Controls (www.cal-controls.com)
Control Techniques (www.controltech.co.uk)
Eagle Controls (www.applegate.co.uk)
Emerson Process Measurement (www.emersonprocess.com)
Eurotherm (www.eurotherm.com)
FGH Controls (www.fgh.co.uk)
Fuji Electric (www.fujielectric.co.jp)
Honeywell (www.iac.honeywell.com)
Invensys, Foxboro (www.foxboro.com)
Jumo (www.jumousa.com)
Philips (www.philips.co.uk)
Self-Tuning Friend (www.csc.umist.ac.uk)
Siemens (www.sea.siemens.com)
Toshiba (www.tic.toshiba.com)
Yokogawa (www.yokogawa.com/us)

Control Algorithms, Advanced:

Model Predictive Control:

Adersa (www.adersa.asso.fr)
AspenTech (www.aspentech.com)
DOT Products (www.dot-products.com)
Emerson Process Management (www.easydeltav.com)
Honeywell (www.honeywell.com)
Invensys (www.invensys.com)
Pavilion Technologies (www.pavtech.com)
Shell Global Solutions (www.shellglobalsolutions.com)

Statistical Process Control:

ABB Inc. (www.abb.com)
Aspen Technology Inc. (www.aspentech.com)
CyboSoft (www.cybosoft.com)
Foxboro Co. (www.foxboro.com)
Honeywell (www.acs.honeywell.com)
JMP Software (www.jmp.com)
Lifestar (www.qimacros.com)
Minitab Inc. (www.minitab.com)

Northwest Analytical (www.nwasoft.com)
Oil Systems Inc. — PI (www.osisoft.com)
Pavillion Technologies Inc. (www.pavtech.com)
Rosemount Inc. Emerson Process Management (www.rosemount.com, www.emersonprocess.com)
Siemens (www.sea.siemens.com/ia)
Software Toolbox (www.softwaretoolbox.com)
StatPoint LLC (www.statpoint.com)
StatSoft (www.statsoft.com)
Stochos Inc. (www.stochos.com)
Wonderware (www.wonderware.com)
Yokogawa (www.yokogawa.com/us)
Zontec Inc. (www.zontec-spc.com)

Graphics Software:

ABB (www.abb.com)
Emerson (www.EasyDeltav.com)
ESA Technology Inc. (www.esa.int)
GE Fanuc Automation (www.faunc.com)
Honeywell (www.honeywell.com)
ICONICS (www.iconics.com)
Invensys (www.invensys.com)
Kessler-Ellis Products (www.kep.com)
National Instruments (www.ni.com)
Nematron (www.Nematron.com)
OMRON Electronics (www.omron.com)
Reichard Software (www.reichard.com)
Rockwell (www.rockwell.com)
Schneider Electric (www.Schneider.fr)
Siemens (www.siemens.com)
XYCOM Automation (www.xycom.com)
Yokogawa (www.yokogawa.com/us)

Modeling Software Suppliers:

Process Control Modeling:

ASPEN Dynamics (Aspen Technology Inc./www.aspentech.com)
cc-Dynamics (Chemstations Inc./www.chemstations.net)
DynaChem (OLI Systems Inc./www.olisystems.com)
DYNSIM (SIMSCI-ESSCO division of Invensys /www.simsci.com)
gPROMS (Process Systems Enterprise Ltd./www.psenterprise.com)
HYSYS Dynamics (Hyprotech /www.hyprotech.com)
VisualModeler (Omega Simulation Co./www.omegasim.co.jp)

Batch Process Modeling:

BatchCAD (Hyprotech Ltd./www.hyprotech.com)
Batchfrac and Batch plus (Aspen Technology Inc./www.aspentech.com)
cc-Batch (Chemstations Inc./www.chemstations.net)

Systems for DCS Configuration Checkout and Operator Training:

Autodynamics (Trident Computer Resources Inc./www.tridentusa.com)
DYNSIM (SIMSCI-ESSCO division of Invensys /www.simsci.com)
HYSYS OTS+ (Hyprotech Ltd./www.hyprotech.com)
MIMIC (MYNAH Technologies division of Experitec Inc./www.mynah.com)
XANALOG NL-SIM (Xanalog Corporation /www.xanalog.com)

Patent Offices:

Canadian Patents Database (patents1.ic.gc.ca/intro-e.html)
European Patent Office (EPO) (www.european-patent-office.org/espacenet/info/access.htm)
German Patent Office (www.dpma.de/index.htm)
Japan Patent Office (www.jpo.go.jp)
United States Patent Office (USPTO) (www.uspto.gov)
World Intellectual Property Organization (WIPO) (www.wipo.org/about-wipo/en/index.html)

INTRODUCTION

This section gives a bird's eye view of where we stand in utilizing the distributed control systems (DCS) in the process control field. This section is followed in this chapter by more detailed discussions of fieldbus and network protocols (Section 4.16) and the various aspects of DCS design, installation and start-up.

Before going into the specifics of the progress made and problems remaining in the field of DCS-based control, I would like to illustrate the general trend through two examples.

In the 1970s I designed the controls of the new IBM headquarters so that the building would be self-heating and without chimney effect. The building had 43 floors and the control room was in the third basement. During startup, communication was an immense task. Just visualize that you have experienced a "hang-up" of the herding-optimization scheme because of a sticking control valve on the top floor and you constantly need information from the control room in the third basement. Well, in respect of the ease of communications, tremendous progress been made.

Today, one can plug in a PC laptop or use a wireless hand tool (see Figure 4.6a) and one can have instant access to all the data, displays and intelligence that reside anywhere in a DCS network. This capability, in combination with the self-tuning, self-diagnosing, and optimizing features, makes startup and operation much more efficient.

Operator training for startup, operation and shutdown can also be simplified because of the process simulator. If the simulation model integrates the process and its control system, so that it can train the operators without risking the consequences of learning while the plant is running, everybody benefits.

FIG. 4.6a
Mobile communication system can provide interface capability for roving operators. (Courtesy of Iconics.)

But this is only part of the picture. On the other side of the coin are the problems associated with the fact that some DCS vendors are selling their "violin" without strings (software is extra) and expect the users to put on these strings (prepare the control algorithm, faceplate, and graphic displays) on their own. In addition to these hidden costs due to "outsourcing," the problems include lack of standardization due to the competing vendors' commercial interests and limited connectivity.

If one reads a bid that lists the costs as "hardware with software license," one might think that the operating software is included. Sometimes it is not, only its license. Similarly, when one reads that an analyzer or an optimization or simulation package needs "layering" or is in the "8th layer," that can mean the software has not been integrated into the total system and to do so is an extra.

In this section, I will concentrate on three areas: First, I will talk about connectivity system integration to obtain good communications among the components of the total process control system of the plant. After that I will discuss the various new and advanced control algorithms and strategies, and finally I will give some data and advice on the costs of DCS.

CONNECTIVITY AND INTEGRATION

Automation is the only means of making American industry globally competitive, while tariffs are lifted to guarantee free trade. The only way to eliminate unemployment is to fully automate our industries and to make sure that the children of the blue-collar steel workers of Pennsylvania grow up to be well-educated process control engineers, computer programmers, intelligent device designers, and optimization specialists. This will take education, and the contents of this handbook should help in providing that education.

At present, 90% of industrial production is still controlled by analog systems, but modern control systems of newly built plants consist of intelligent and self-diagnosing field instruments (sensors, valves, motors, safety devices) and a number of data highways or network buses, which serve to integrate these field devices with the DCS workstations (serving control/operation, engineering/historian, maintenance), the plant/enterprise-wide network serving business functions, and external PCs serving process modeling and simulation functions (Figure 4.6b).

The trend seems to be that HART is becoming the standard for interfacing with analog systems and Ethernet is handling most office solutions. SCADA serves to combine field and control data to provide the operator with an overall view of the plant.

Protocol is the language computers speak. If two black boxes in a refinery do not speak the same language, the consequences can be serious. Yet the commercial goal of

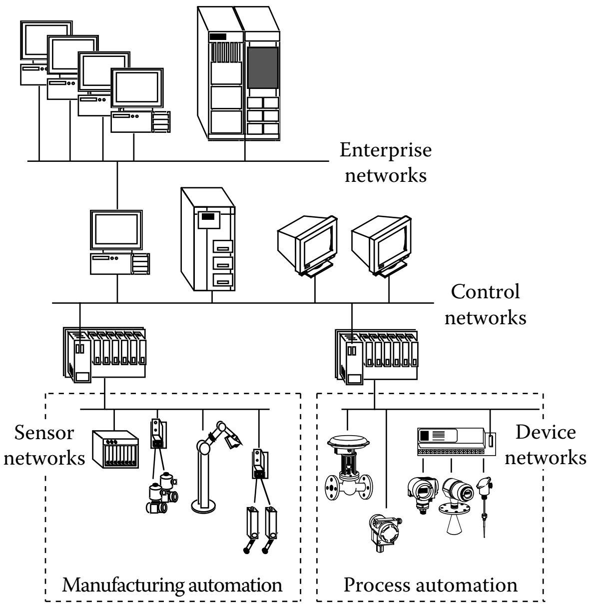


FIG. 4.6b
Connectivity requires that the intelligent devices of the sensor and device networks be provided with unlimited two-way communications not only with DCS workstations connected to the control network but also with sophisticated (external) process simulation models.[1]

creating captive markets is strong, and IEC has standardized on eight protocols (Table 4.16i).

While there is no common DCS fieldbus protocol, all protocols use Ethernet at the physical and TCP/IP at the Internet layer. MODBUS TCP is used to interface the different DCS protocols.

The layers of the communication pyramid are defined several ways. As described in Figure 4.16c, OSI defines seven (7) layers, #1 being the physical and #7 the application layer (with #8 being used by some to refer to the "unintegrated" components). The IEC-61512 standard also lists seven levels, but they base them on physical size: 1) control module, 2) equipment, 3) unit, 4) process cell, 5) area, 6) site, and 7) enterprise. In the everyday language of process control, usually four layers are discussed in the automation pyramid: Here, #1 is the level of sensors and actuators in the field, #2 is control, #3 is operations, and #4 is the level of business administration.

When one is designing the overall control system of a new plant or when one is updating the control system of an existing facility, there is a tendency to make a quick I/O count and, based on that preliminary information, to obtain bids and select the DCS supplier. This is the wrong approach because it will result in endless extras and delays.

Organizing the Project

Field Instrument Integration In a new plant, the first step should be to specify all the sensors, transmitters, control valve actuators, positioners, local controllers, electric motors, solenoid valves, and all other microprocessor-operated devices in the field that need to communicate information to or receive commands from the DCS system over the networks.

When the project consists of additions to an existing plant, all the interfacing between the existing and new facilities must also be specified (Figure 4.10a). The three types of interfaces most often used are object linking and embedding (OLE), serial-to-highway gateway and serial-to-serial port. Because digital transmitters, smart analyzers, digital actuators, and motor starters utilize a variety of communication protocols, the manufacturers provide interface cards to the common buses. Such cards are available for HART, DeviceNet, Foundation Fieldbus, AS-inetface, Profibus, Modbus, or AB's Data Highway Plus.

This integration is not as simple as one might think because there is no international standardization and some smart field devices are proprietary, or cards can be mixed up. Another roadblock to DCS system integration is that manufacturers implement the same task in different manners. For example, the controller status is described by some manufacturers by using two parameters: Auto/Manual and Remote/Local, while others use a single parameter: Manual/Auto/ Cascade/Remote-Cascade/Remote-Out.

Safety Integration The second step is to specify the network requirements in terms of the needs for physical layer redundancy and also in terms of the number and type of network levels required (Figure 4.6c). It is at this point, where the type of safety instrumented system (SIS) is selected for the plant and it is decided whether the plant's SIS and the control systems will be separated (Figures 4.10e and 4.10f) and, if they will be separate, what kind of interfacing will be used. If safety and control will not be separated, decisions are required on prioritizing, redundancies, hard wiring, etc.

Some suppliers provide redundancy by building two of everything in the system as standard; others provide redundancy on an optional basis for power supplies, data highways, traffic directors, remote controller electronics, and workstation video terminals. Yet others provide one for one, one for four, or one for eight backup of controller file cards. In addition to the backup method, there are also differences to backup diagnostics and method of switchover, which can be automatic or manual.

Maintenance and Simulation Integration The third step is to specify the requirements for interfacing between maintenance and control systems and to decide on the optimization, simulation, training and commissioning tools that will be used and therefore need to be integrated into the total plant automation package.

When it comes to process simulation, the process modeling software packages can be very large and might reside in separate PCs or might even be obtained from the Internet (Figure 4.6c). Therefore, it is essential that the communication

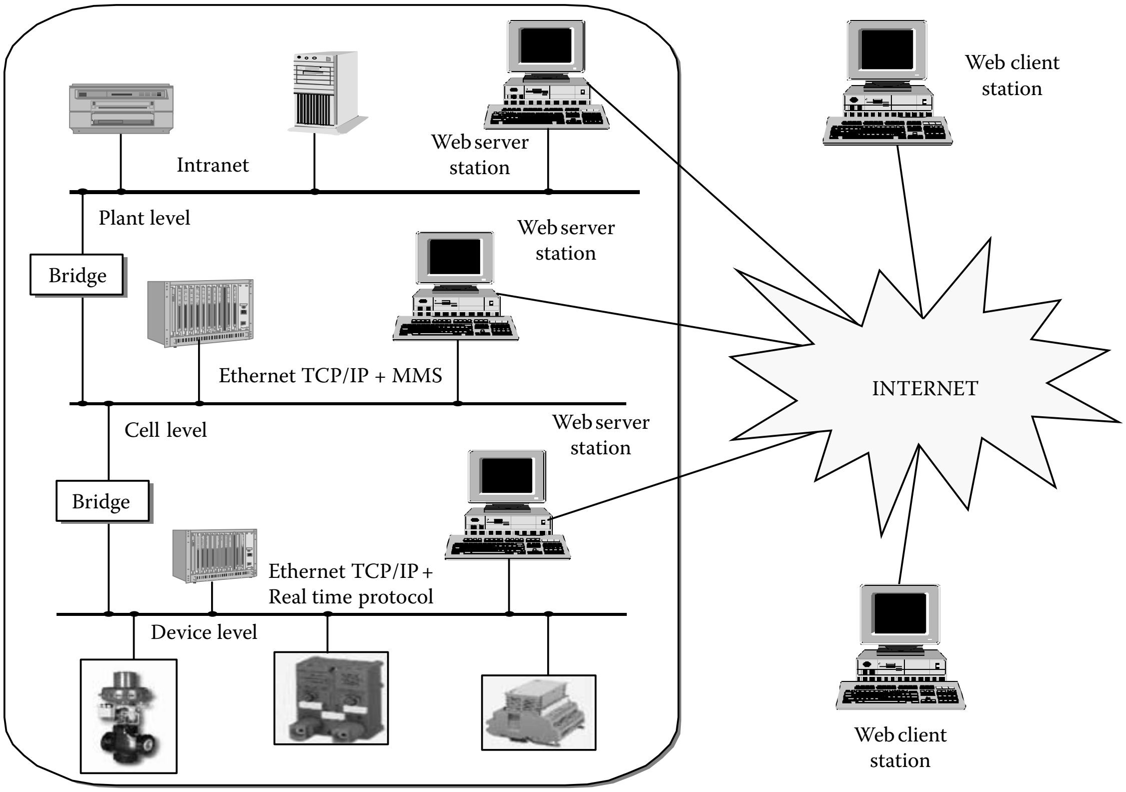


FIG. 4.6c
It is important to integrate the DCS system with intelligent instruments at the device level and provide communication bridges both between the networks and with process-specific dynamic models used for simulation and training.[2]

bus network be able to interface with them and to handle the information flow without interrupting the control functions.

Network and Bus Integration The fourth step is to select the network to be used. Profibus is supported by Siemens, Endress & Hauser, Yokogawa, Rosemount, Invensys, and ABB; Fieldbus Foundation is supported by Emerson, Honeywell, and Telemechanique; while Controller Area Network (CAN)/DeviceNet is supported by Allen-Bradley, Modicon, Numatics, and ProLog.

The major DCS manufacturers accepted the IEC1158-2 fieldbus standard for integrating their systems by providing the means of communication between them. IEC-61804 specifies function block interfaces between technologies, while others can be integrated "proxy" function blocks, which can "map" the data from the systems of the "other" suppliers. The digital communication of these devices should be specified to include not only the requirements of control but also those of self-diagnostics, maintenance, and safety. MODBUS interfacing can be used to integrate existing and new DCS systems.

In this fourth step the user should evaluate which bus network is best capable of meeting the requirements of the previously listed three steps. For the capabilities and features of some of the above-mentioned networks, refer to Table 4.6d.

TABLE 4.6d
Features and Characteristics of Some Field Automation Networks[3]

Network Name	*Type of Protocol*	*Speed*	*Type of Wiring*	*Bus Power*
PROFIBUS-DP	Master/slave and token passing	12 Mbps	UTP	No
PROFIBUS-PA	Master/slave	31,250 bps	STP	Yes
Foundation Fieldbus H1	Arbitrated	31,250 bps	STP	Yes
Foundation Fieldbus HSE	Arbitrated	100 Mbps	UTP	No[a]
DeviceNet	Master/slave	500 Kbps	UUP(2) or flat	Yes
AS-interface	Master/slave	500 Kbps	Flat	Yes
ControlNet	Timed slot	5 Mbps	CATV coaxial	No
Interbus	Slot-ring	500 Kbps	UTP	No

[a] IEEE802.3af will define power from switch.

UTP = unshielded twisted pair, STP = shielded twisted pair, UUP = unshielded untwisted pair, FSK = frequency shift keying, HSE = high-speed Ethernet, CATV = community antenna television (cable).

Selecting the DCS Supplier The fifth and final step in this process of designing a plant control system is to send out the whole automation package for bids. Only DCS suppliers who support the selected network should be invited to bid (for the supplier's web addresses, refer to the Feature Summary at the beginning of this section).

In the specifications, it is essential to list all the intelligent field devices and external systems that must be integrated into the total control system. The bid specification should require the DCS supplier to include all that is needed, both in terms of hardware and in terms of supporting software to integrate the external devices and systems into the overall control system. Similarly, it is essential to clearly specify what software packages (advanced control, maintenance, graphics, simulation, and training tools) must be included in the bid.

In terms of the required graphics, the bid package should state which of the three IEC 61131-3-approved graphic control languages are to be used. The specification should also list the displays (graphic, detail, faceplate, alarm and event summary, trend, tuning, diagnostic, and report), which the DCS supplier is responsible for preparing, and the types of dynamic elements (dynamos), background colors, and aliases that are to be used.

The specification should also state the division of responsibilities between the user, engineering firm, system integrator, and DCS supplier for advanced control, graphics or other software development, simulation, training, and commissioning. In the area of commissioning, the DCS supplier's responsibilities should be clearly stated in connection with training, preliminary checkout, fieldbus testing, safety system and advanced control testing, startup, and controller tuning.

If the above outlined procedure is followed, the extras and delays will be fewer. If they are not and if the DCS supplier is selected before its scope is clearly specified, the consequence could not only be a doubling of the initial bid, but could also cause construction delays, startup problems, and reduced safety.

The Future

It will take a couple of decades, but it is hoped that the day will come when the proprietary networks have disappeared and there is only one worldwide communication network standard. It is hoped that the time is not too far off when all digital devices can connect to all networks and all software packages can be "embedded" into any DCS systems.

From the experience with the PC and the Internet, we can see that if it is in the interest of suppliers, we can quickly arrive at global standardization and any software package can be "embedded" into all PCs. What is needed to bring about global standardization in the process control field is commercial pressure by the users. Professional societies and user groups can play an important role in speeding up this development.

CONTROL ADVANCES

Murphy's second law is that one cannot control a process without understanding it. Others will argue that no computer will ever beat the world champion in bridge or chess because the programmers are not as good at bridge or chess as these champions. There is much validity to these views, yet it is also true that computers need not be limited by the knowledge of the programmers but can be provided with the ability to learn. After all, once even Einstein was only a baby and knew absolutely nothing. Yet, through some process of neural fuzzy logic, he managed to build a pretty good model of the world we live in and learned much more than just to speak and walk.

The purpose of this writing is not to predict the future but only to give a bird's-eye view of where we stand in the field of process control. I will first discuss the progress made in the configuration of PID control algorithms and in their autotuning, both in the "on demand" mode and in the self-tuning or adaptive mode. Next, I will mention the Smith predictor and discuss the various non-PID based model predictive control (MPC) schemes, including the Dahlin variety and adaptive model controllers (AMC). I will conclude this discussion by describing the operation of the fuzzy logic and artificial neural networks (ANN).

Basic PID Algorithms

The main advantage of the computer is the speed and convenience with which one can select and change the best choice of a PID algorithm to match the dynamics of the controlled process. Figure 4.6e shows some of the basic PID configurations from which the DCS system can select the best choice either automatically or with the operator's approval. In this figure, configuration 1 is the basic ISA PID algorithm, where all three modes (P, proportional; I, integral; and D, derivative) act on the error, which is the difference between the process variable (PV) and the set point (SP).

Configuration 2 is a better selection when it is desirable to reduce the frequency of changes to the controller's output. This algorithm is applied either to increase the life expectancy of the control valve or in "herding" control, where the error is considered to be zero as long as it is inside a predetermined dead zone or gap.

Configurations 3 and 4 give different options on which control modes should act on the error and which on the process variable only. In my view, with the exception of cascade slaves, 3 should be used for all controllers that have a derivative mode because there is no advantage to D acting on SP changes. If one wants to completely eliminate the overshoot on set point changes, 4 should be used, where both D and P act on PV only.

If the set point of a loop is changed frequently, it is recommended to filter the set point as in configuration 5. This will reduce the overshoot while allowing the use of a narrower P-band. In configuration 6 the controller responds to

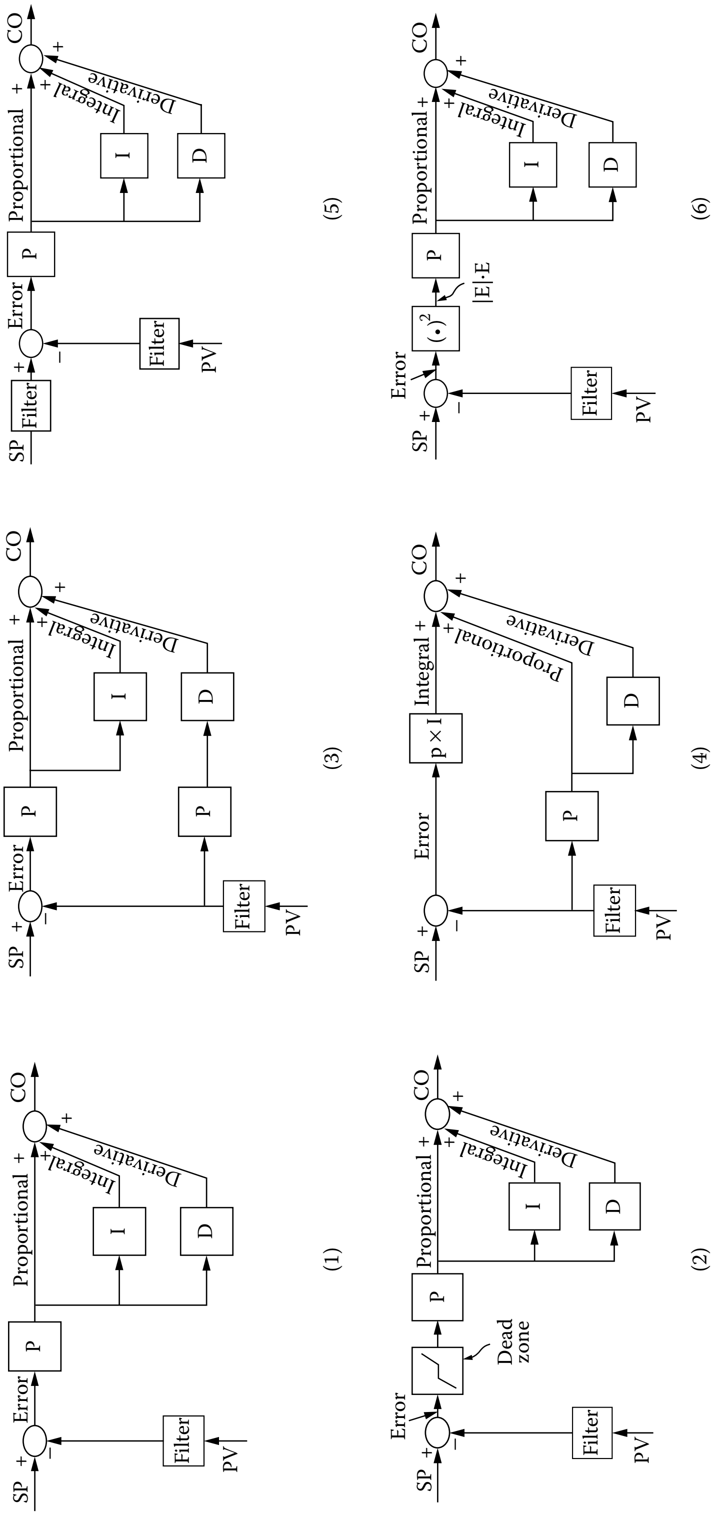


FIG. 4.6e
The basic PID algorithms include the 1) ISA, 2) dead band, 3) D on PV, 4) D and P on PV, 5) SP filtering, 6) error squared, and 7) integral squared.[4]

the square of the error, which is desirable on surge level control, where we want stable operation in near the steady state, but aggressive correction of large errors. In configuration 7 only the integral is squared, which is good on variable dead time processes.

Naturally, there are many variations on these basic themes. For example, if the controller output can be blocked, such as in selective or in cascade configurations, the integral action should act not on the output but on some external signal, such as the opening of the valve. This is called external reset. In other cases, such as on extremely fast flow loops, one might use inverse derivative to stabilize the loop.

Auto-Tuning

Once the appropriate PID algorithm is selected for the particular process, the next task is to tune the controller to match the dynamics (gain, time constant, and dead time) of the process. The various PID algorithms are described in full detail in Section 2.3 and the equations to be used to tune them are defined in Sections 2.35 to 2.37. Section 2.38 explains how the loops are optimized and tuned by computers.

There is an inherent contradiction between the goal of tuning, which is to keep the controlled variable constant, and the method by which the tuner must learn the dynamics of the process, which is to intentionally upset the controlled variable. Finding an acceptable balance between these considerations is the criterion for a good self-tuning controller.

Self-tuning or auto-tuning can be done either on demand, where it is initiated by the operator, or automatically. In “on-demand” tuning, it is up to the operator to decide when the loop needs retuning, but once that decision is made, the DCS system performs the retuning automatically. If “adaptive” tuning is used, the DCS system is programmed to initiate an automatic retuning cycle after changes in set point or when significant noise or upsets are experienced.

Tuning on Demand This method of tuning involves the switching of the controller to manual, while maintaining its output signal to the manipulated variable (control valve) as it was before the switching. After switching to manual, a small increase (1 to 3%) step change is made in the output signal, then it is held and after that a step change is made down (Figure 4.7b), and this sequence is repeated in an oscillatory manner.

The above-described oscillatory step changes in the manipulated variable cause a sinusoidal response in the controlled variable. This response is analyzed by the DCS system to determine the ultimate gain, ultimate period, dead time, time constant, etc. of the process. For an explanation of the determination of ultimate gain and time period, refer to Figures 2.38c and 2.38i in Chapter 2.

Once the process dynamics are determined, the operator can input the kind of tuning rules to be used (Section 2.35) and the DCS software will use the corresponding equations to calculate the corresponding controller settings.

This oscillatory method of open-loop tuning usually gives acceptable results on self-regulating or integrating processes to obtain the initial controller settings during plant startup. Because such open-loop tests disregard the dynamics of the controller itself and because they tend to be rather inaccurate when the measurement is noisy, it is advisable to retune the loop using a closed-loop method (Section 2.35) after startup.

Adaptive Tuning Different DCS suppliers use different methods to continuously and automatically adjust the tuning constants of the control algorithm in response to changes in the dynamics of the process. Similarly, they use different methods of injecting excitation into the loops prior to detecting the process response. These excitations can be small changes to the set point or to the controller output, and they can occur either in the manual or the automatic (open or closed) modes.

The adaptation cycles continue until the controlled variable response is sufficiently excited for analysis. Adaptation of feedback and feedforward portions of loops can be done separately or in combination. Once the process dynamics or the total loop dynamics are identified and, based on those, the tuning constants for standard PID can be calculated (1 in Figure 4.6e), the adaptive tuning controls can proceed to select the most appropriate control algorithm and switch to it.

While theoretically this sounds fine, in reality, we have a long way to go before adaptive tuning and automated control algorithm switching on the basis of self-diagnostics will be fully developed. This is because on process control applications, the presently used DCS systems have a tendency to remodel continuously. This is due to the compromises made between the size of the excitations applied vs. the effects of noise, hysteresis, valve sticking, and nonlinearity in our control loops.

Model-Based Control Algorithms

Chapter 2 of this volume discusses a number of control strategies (Sections 2.13, 2.14, 2.17, and 2.19), where the controller is loaded with the model of the process and therefore it can compare the response of the real process to that of the model and act on the difference similarly to how a PID controller acts on the measurement error.

Smith Predictor On processes with long dead times, this method of control gives good set-point response without overshoot, while it is only slightly better than regular PID on responding to load disturbances. It operates by comparing the controlled variable from a model of the process, which

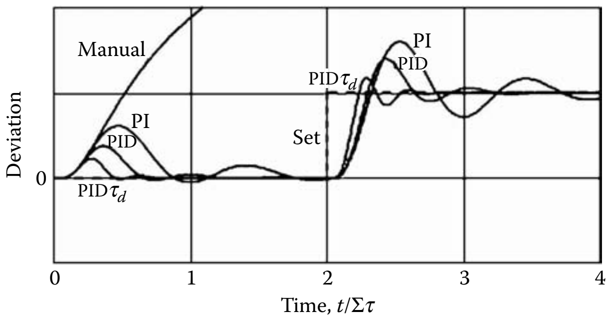


FIG. 4.6f

The response of the $PID_{\tau d}$ *controller to load and set point changes is superior to those of both the conventional PI and PID controllers.*[5]

does not have dead time with the controlled variable measurement from the real process.

Shinskey[5] has shown that the regular $PID_{\tau d}$ controller (which is easier to tune than the Smith predictor) can give similar performance under some conditions. Figure 4.6f shows the improvement provided by the interacting $PID_{\tau d}$ controller, compared with regular PI and PID control.

Model Predictive Control (MPC) These controllers, where the controller model is the inverse of the dynamic process model, are called by different names, including internal model controllers (IMC) and lambda, Dahlin, or pole cancellation controllers. From most DCS suppliers, only small MPC packages (eight inputs by eight outputs) can be embedded; larger models have to be external. As to who calls what, an MPC application varies. Some DCS suppliers consider such simple tasks as set-point limiting, rate of change limiting of controller output, or the application of high and low constraint limits as MPC.

Most larger MPC controllers have to be external (layered) and integrated into the DCS system. If the process model is very accurate (usually it is not) and if the dynamics of the process do not change (usually they do with load variations), MPC performance will be superior to PID. In most cases the process dynamics do change with load changes, and in such cases the MPC response will be sluggish unless the model itself is updated (adapted) periodically.

When the MPC model is used in a simulator, it is possible to see the actual process response before it happens. The ability to be used for operator training and the ability to adjust the speed of execution are some of the advantages of MPC. One of its limitations is the high cost of developing an accurate process model that will self-adapt to the changing dynamics of the process.

Fuzzy Logic What a good operator does is hard to describe by a mathematical equation. It is for this reason that very complex processes (say the landing of a helicopter) cannot be easily controlled by conventional means and are often left to be controlled manually. This is because operators aggregate a variety of data and combine them into strategies that cannot be integrated into a single control law.

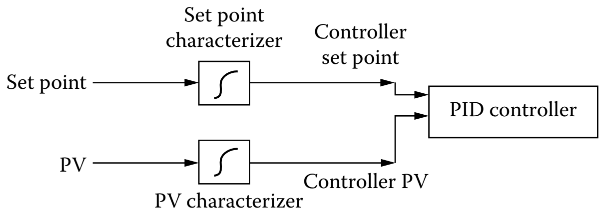


FIG. 4.6g

If the nonlinearity is known and does not change with time, conventional controls can be used if a characterizer is applied to both the set point and the measurement (PV) signals.[6] *(Courtesy of ExperTune Inc.)*

One reason why conventional controls cannot be used on some processes is that they are nonlinear. If the nonlinearity is accurately known and does not change with load or time, conventional PID controls can be used after the measurement and set point are both characterized (Figure 4.6g). When this is not possible, one of the options to consider is fuzzy logic.

The first step in developing a fuzzy model is to talk with the plant engineers and operators and learn about the process variables that have an influence (positive or negative) on the controlled variable and to approximate the size of their influence by some weighing factors. What is required for fuzzy logic to work is that the process be controllable in the manual mode by an operator and that great quantities of data be available to show how that manual control is achieved (see Section 2.31 for details).

Based on all the data, rules are established to determine the controller output on the basis of a number of input functions, and each of these functions is scaled to describe its relative influence on the total output.

ANN Control Artificial neural networks (ANN) are explained in some detail in Section 2.18. They are similar to fuzzy logic controllers to the extent that the mathematical model that relates their inputs to their outputs need not be known. The main difference is that an ANN can only be "trained"; the gains of its functions cannot be changed, they are fixed.

Some successful applications of back-propagation type ANN control involve the prediction of pH values in stirred tanks and the prediction of product boiling points in distillation processes. ANN has shown its ability to learn the dynamics of a process well if sufficient amounts of training data are available.

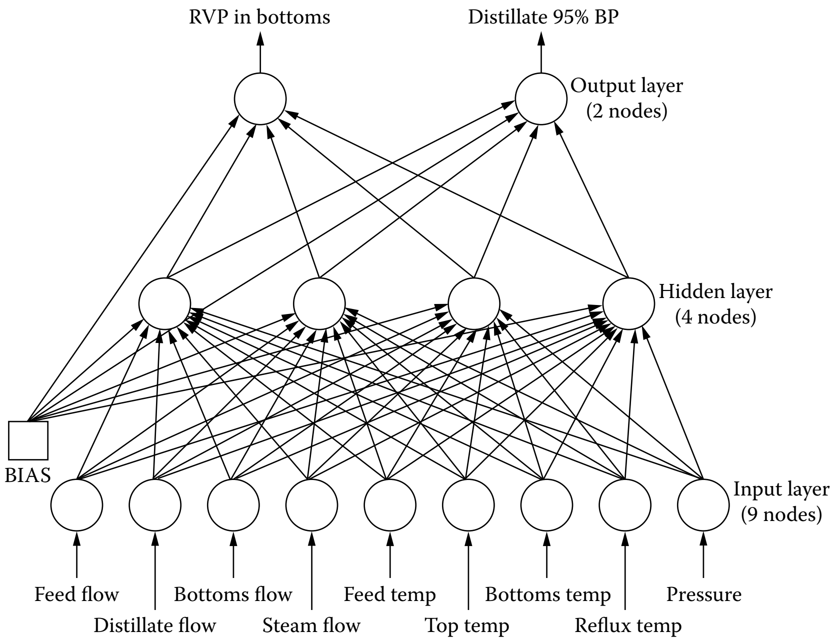


FIG. 4.6h
A three-layered ANN can predict the Reid vapor pressure (RVP) of the bottom product and the 95% boiling point of the distillate.

An example of a three-layer ANN network is illustrated in Figure 4.6h. Here the goal is to control the Reid vapor pressure of the bottom products and the 95% boiling point of the distillate without measuring it directly. Instead, a variety of other properties are measured (nine input nodes) and, based on large amounts of past historical data, the actual real-time values of the two controlled variables are detected. Between the in- and output layers in this example there are four additional nodes in a hidden layer and bias.

The main advantage of ANN is the ability to predict the controlled variables when otherwise they are either unmeasurable online or the analyzers (online or in the laboratory) take too long to analyze the sample. The ANN model can have more than three layers and more than nine inputs. The smaller ANN applications can be embedded in most DCS systems, while the larger ones are usually interfaced to the DCS from an external eighth layer.

DCS Bid Package

When specifying the control system, it is important to obtain both the PID and the control algorithm library from each of the bidding DCS suppliers (Tables 4.6i and 4.6j). At the time of analyzing the bids, it is equally important to check whether all the costs associated with the building and embedding the control algorithms are included. These costs should also cover the associated graphics and faceplate preparation for the required PID. Advanced control costs also should be included and fully guaranteed, because otherwise startup problems can result.

TABLE 4.6i
Typical Algorithm Library

- PID Algorithms
  - PID
  - PID—ratio
  - PID—gap
  - Auto/Manual
  - Supervisory control
  - Adaptive tuning
  - PV vs. setpoint error
  - Ramp generation
- Computation
  - Summer
  - Divider
  - Mass flow
  - Function generators
  - High and low selectors
  - Dead time
  - Comparator
  - Median
  - Lead/Lag
  - Square root
  - Exponential function
  - Log function
- Input/Output
  - Analog input
  - Analog output
  - Digital input
  - Digital output
- PID Functions
  - Output limiting
  - Output tracking
  - Feedforward
  - Setpoint ramping
  - Anti-reset windup
  - Setpoint clamping
- Logical Functions
  - *And, or, exclusive or not*
  - Switch—latched, unlatched
  - Timer
  - Counter
  - Sequencer
  - Flip-flop
  - Motor control

TABLE 4.6j
*A Partial List of the Constantly Growing List of Available PID Algorithms**

Proportional (Gain)
Integral (Reset)
Derivative (Rate)
P + Adaptive Gain
I + Adaptive Gain
D + Adaptive Gain
P + Bias
P + I
P + I + Adaptive Gain
P + I + Reset Feedback
P + D
P + D + Adaptive Gain
P + D + Bias
I + D
P + I + D (PID)
PID + Adaptive Gain
PID + Ratio
PID with Bumpless Transfer (Auto Ratio)
PID + Bias with Bumpless Transfer (Auto Bias)
PID + Error^2 on Gain
PID + Error^2 on Bias
PID + Gap (Adjustable Deadband)
PID + Ratio − Ratio Forceback
PID + Ratio − Bias Forceback
PID + Reset Feedback

*See Section 2.3 for a discussion of available PID variations.

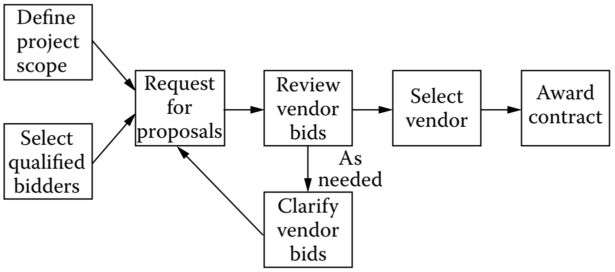


FIG. 4.6k
The traditional bidding process.

COSTS

When it comes to DCS-based total control systems, one has to be careful with implementing the traditional bidding process (Figure 4.6k). This is because the expenses associated with the previously discussed interfacing and integrating problems and the costs of special software requirements for control, simulation, and optimization can sometimes exceed the total hardware and labor cost of the project.

Table 4.6l gives the costs associated with a small control system addition consisting of one analyzer, seven transmitters, some thermocouples, and three computer supported touch-screen monitors. In this $100,000 project, the installation labor cost was only 10% and the software cost was only 20%. These percentages can be so small only if no custom software is used and if most of the sensor/transmitter installation, commissioning, operator training, and startup is performed by the user. If that is not the case and some advanced control algorithms, custom graphics, and training or simulation is required from the DCS supplier, the cost can easily double.

Table 4.14n describes the hardware cost of a medium-size DCS project with a fieldbus and about 200 analog and 1,000 on/off I/O. The hardware cost of this project (including licensed software) is $250,000. What is not so obvious from that hardware list is that very little software is included. Yet, if some of the I/O is from intelligent field devices, which need to be integrated (layered) into the total system, and if advanced controls, process simulation, and operator's training are needed, the installed final cost can approach a million dollars.

It is the software and labor costs that often determine the total DCS project cost. In order to control these costs, it is essential to fully specify all the interfacing and all the custom software requirements of the project and obtain the initial bid on that basis. This means that all field components of the system, including intelligent field devices, must be selected and their interfacing included in the DCS bid package before bids are obtained. Similarly all the interfacing and software support for the required advanced controls, process simulation, and operator's training must also be specified before bids are obtained. If this is not done, substantial delays and extras will result.

CONCLUSIONS

Figure 4.6m illustrates a DCS workstation, the end result of a long and difficult design effort. Designing a good process control system takes a lot of effort. The tools of process control — the DCS and PLC systems, and the network buses, sensor, and valves — are like the instruments in an orchestra. The engineer who designed the control strategy is the composer, and his job is to make sure that the instruments of the

TABLE 4.6l
Cost Estimate of a Small DCS Project.[7]

Materials
Instrumentation **Sub-Total $27,311**

No.	Item	Cost per	Total	Source/Assumptions
17	Type J Thermocouples, modified	$195	$3,315	Quote from XYZ Distributor
3	Smart Pressure Transmitters	$2,475	$7,425	Quote from ABC Company
4	Level Sensors	$924	$3,696	Similar to project X
1	Analyzer	$12,875	$12,875	Competitive bid

User Interface **Sub-Total $26,941**

No.	Item	Cost per	Total	Source/Assumptions
3	Computers	$1,774	$5,322	800 MHz, 128M RAM, 80GBHD
3	Touch-Screen Monitors	$795	$2,385	Quote from DCS vendor
1	Color Laser Printer	$1,879	$1,879	Can we do better?
3	Software User Licenses	$5,785	$17,355	Discuss corporate discount

I/O **Sub-Total $6,375**

No.	Item	Cost per	Total	Source/Assumptions
5	T/C Input cards, 4 channels each	$1,275	$6,375	Vendor price list

Materials
Instrumentation **Sub-Total $27,311**

No.	Item	Cost per	Total	Source/Assumptions	Capital	Expense
17	Type J Thermocouples, modified	$195	$3,315	Quote from XYZ Distributor	$3,315	
3	Smart Pressure Transmitters	$2,475	$7,425	Quote from ABC Company	$7,425	
4	Level Sensors	$924	$3,696	Similar to project X	$3,696	
1	Analyzer	$12,875	$12,875	Competitive bid	$12,875	

Labor **Sub-Total $11,025**

No.	Item	Cost per	Total	Source/Assumptions	Capital	Expense
8	Days Site-Clearance	$480	$3,840	T&M Contractor		$3,840
3	Days Wiring Install	$795	$2,385	T&M Contractor	$2,385	
4	Days Production Support	$1,200	$4,800	Required by Op. Dept.		$4,800

orchestra are fully coordinated to produce smooth and optimized control.

Another important aspect to remember is that in the absence of strong and competent user associations, the (natural) commercial interests of suppliers tend to conflict with standardization. When the system supplier says that this or that optimization or other software package needs to be layered (referring to the eighth layer of the ISO model protocol), the resulting problems and expenses of integration can be substantial. The difficulty in integrating systems that were not designed to "speak the same language" is like flying an airplane with pilots who do not speak the same language.

It should also be remembered that DCS costs have not dropped. The case is that the costs of unconfigured hardware have dropped, but total system costs have risen. The

FIG. 4.6m
Modern DCS workstation. (Courtesy of Honeywell.)

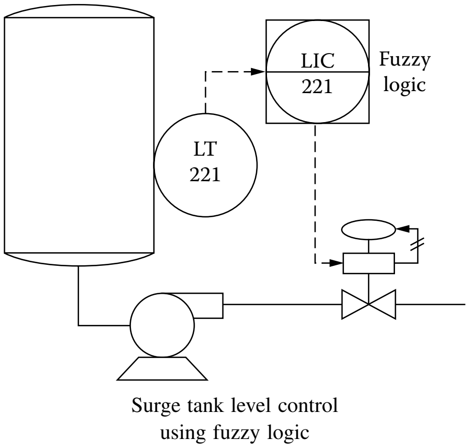
limitations of new software do not end with their connectivity and their hidden costs; an added serious source of problems is that they are only as good as the programmer who configured them, and the Greg Shinskeys of this world do not do programming.

References

1. Pratt, W. A., *Instrument Engineers' Handbook*, Vol. 3, Section 4.4, Published jointly by CRC Publishing and ISA, 2002.
2. Vitturi, S., *Instrument Engineers' Handbook*, Vol. 3, Section 4.5, Published jointly by CRC Publishing and ISA, 2002.
3. Caro, R. H., *Instrument Engineers' Handbook*, Vol. 3, Section 4.9, Published jointly by CRC Publishing and ISA, 2002.
4. Ruel, M., *Instrument Engineers' Handbook*, Vol. 3, Section 5.6, Published jointly by CRC Publishing and ISA, 2002.
5. Shinskey, F. G., *Process Control Systems*, 4th ed., New York: McGraw-Hill, 1996.
6. Gerry, J. P., *Instrument Engineers' Handbook*, Vol. 3, Section 5.1, Published jointly by CRC Publishing and ISA, 2002.
7. Buckbee, G. C., *Instrument Engineers' Handbook*, Vol. 3, Section 1.10, Published jointly by CRC Publishing and ISA, 2002.

4.7 DCS: Control and Simulation Advances

T. L. BLEVINS, M. NIXON (2005)



Surge tank level control using fuzzy logic

Flow sheet symbol

Advanced Control Software Packages That Are Embedded in Modern DCS Systems by Emerson:	A. Autotuning B. Fuzzy Logic C. Model Predictive Control D. Neural Network–Based Variable Estimation E. Performance Monitoring F. Process and Control Simulation
Partial List of Software Suppliers:	Adaptive Resources (A, C) (www.adaptiveresources.com) Allen-Bradley (A, B) (www.ab.com) Aspentech (C, D, E, F) (www.aspentech.com) ControlSoft (A) (www.controlsoftinc.com) Emerson (A, B, C, D, E, F) (EasyDeltaV.com) Expertune (A, E) (www.expertune.com) Gensym Corporation (D, E) (www.Gensym.com) Honeywell (A, B, C, D, E, F) (www.honeywell.com) Invensys (A, B, C, D, F) (www.invensys.com) Matrikon (D, E) (www.matrikon.com) Neuralware Inc. (D) (www.neuralware.com) Pavilion Technologies (C, D) (www.pavtech.com) Siemens (A, B, C, D) (www.siemens.com) Universal Dynamics Technologies (A, C) (www.unadyn.com) Yokogawa (A, B, C) (www.yokogawa.com/us)

This section discusses some of the advanced software packages that are embedded in modern distributed control systems (DCS). They serve control, estimation, monitoring, simulation, and tuning functions. For the advanced control strategies that are not discussed in this section, refer to the following sections in this volume:

- Section 2.20: “Optimizing Control”
- Section 2.33: “State Space Control”
- Section 2.34: “Statistical Process Control”

For in-depth discussions of auto-tuning, fuzzy logic, and model predictive control (MPC), which are also discussed in this section, the reader is referred to the following sections:

- Section 2.13: “Model-Based Control”
- Section 2.14: “Model-Based Predictive Control Patents”
- Section 2.16: “Modeling and Simulation of Processes”
- Section 2.17: “Model Predictive Control and Optimization”
- Section 2.18: “Neural Networks for Process Modeling”
- Section 2.19: “Nonlinear and Adaptive Control”
- Section 2.27: “Sampled Data Control”
- Section 2.29: “Self-Tuning Controllers”
- Section 2.31: “Fuzzy Logic Control”
- Section 2.38: “Tuning by Computer for Optimizing”

INTRODUCTION

The computational requirements of software products for advanced control have traditionally limited their implementation. For that reason, their associated software resided either in a host computer or in dedicated control units that were external to the basic DCS package and were “layered” onto the plant’s control system. This situation is changing, as some of the more advanced software packages can be embedded in the basic DCS system.

Recent advances in the processing power available within some DCS control systems have allowed them to include some advanced control capability. These embedded software packages include:

- Performance monitoring
- Loop tuning
- Fuzzy logic control
- Model predictive control
- Neural network-based variable estimation
- Control simulation

Most of the successful advanced control applications, such as MPC, that are in use today are used on high-throughput processes, where the savings justify the high cost of developing, installing, and maintaining these layered control applications. This new generation of advanced control is often designed so that the average process engineer can use these features with the same level of confidence as he or she used the more traditional controls. Fuzzy logic control, model predictive control, and property estimation may be implemented as a function block that can be assigned to execute in the standard redundant controllers.

This operating environment has improved security and performance in comparison to the traditional approach of leaving the advanced controls outside of the basic control system, which is referred to as “layering.” For this embedded implementation, the engineering tools for the configuration and troubleshooting of traditional control may also be used with these advanced software blocks.

Also, the standard dynamos of these blocks support operator displays that include advanced controls without the need to map parameters. Thus, the engineer and the operator can have consistent interfaces to both the advanced and the traditional control applications.

Embedding advanced control applications has many other advantages as well. Probably the most important is that the advanced control application software runs inside the embedded controllers where the applications can fully participate in controller redundancy, alarming, configuration downloads, and online upgrades. This also means that all parameter references are managed by the runtime and the configuration system, and such terms as “Mode” and “Status” have meanings consistent with the terminology of other software packages in the system. To the operator and other operations personnel, the advanced control applications are largely invisible — they are just other function blocks and parameters.

PERFORMANCE MONITORING

The maximum production rate and efficiency of a plant are determined by the process design and the capacity of the equipment used. However, many plants never reach the production capacity that is inherent in the plant design and equipment. In addition, there usually is little time to study the process operation to determine whether it could be improved because attention is concentrated on maintenance and on abnormal conditions that can become major sources of process disturbances.

To improve on this situation, some distributed control systems are now designed to support automatic and continuous monitoring and detection of abnormal operation of control and field devices. When control is based on the Foundation fieldbus architecture, the status associated with each function block output gives a direct indication of the quality of the measurement of control signal. Also, the actual and normal mode attributes of these blocks may be used to determine whether a function block is operating in its designed mode. Thus, by continuously monitoring the status and mode supported by fieldbus function blocks, it is possible to automatically determine the following abnormal conditions in control and input-output (I/O) function blocks:

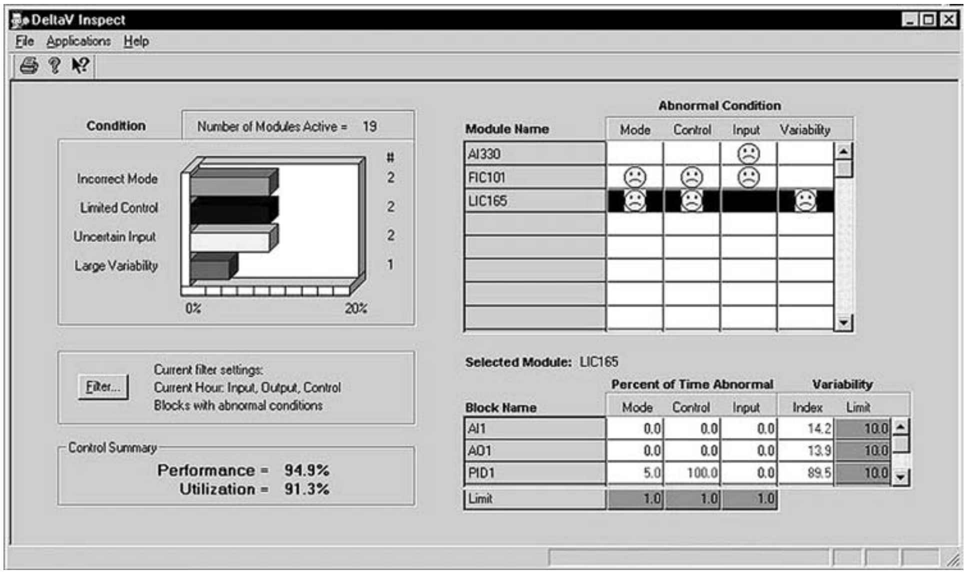
- A condition exists that could be reducing the accuracy of the measurement.
- Control effectiveness is limited by a downstream condition.
- The block is not running in its designed mode of operation.

Based on the inherent accuracy of the sensor and its maximum error, it is possible to compute a variability index that compares the actual performance to the best achievable. These calculations can be built into the control and I/O function blocks to support control analysis. By this approach, the control performance of even the fastest processes can be evaluated.

Within the control system, the status and mode values are automatically processed to determine the percent of time when an abnormal condition existed during the current and previous hour, shift, or day. If this percentage exceeds a preset limit, then the condition is flagged as abnormal and the module containing the function block will show up in the monitoring interface. An example of an abnormal condition monitoring interface is shown in Figure 4.7a.

CONTROLLER TUNING

Loop tuning can be done either when needed (on demand) or continuously by adaptive tuning. On-demand tuning is initiated by the operating personnel. Continuous adaptive tuning is performed automatically after set-point changes or significant disturbances, or when noise is experienced. On-demand tuning is often used during startup to obtain approximate tuning constants during the initial commissioning of a control system. Adaptive tuning corrects the tuning in response to changes in process dynamics.



Condition Number of Modules Active = 19

Condition	#
Incorrect Mode	2
Limited Control	2
Uncertain Input	2
Large Variability	1

Current filter settings:
Current Hour: Input, Output, Control
Blocks with abnormal conditions

Control Summary
Performance = 94.9%
Utilization = 91.3%

Module Name	**Abnormal Condition** Mode	Control	Input	Variability
AI330			☹	
FIC101	☹	☹	☹	
LIC165	☹	☹		☹

Selected Module: LIC165

Block Name	**Percent of Time Abnormal**			**Variability**	
	Mode	Control	Input	Index	Limit
AI1	0.0	0.0	0.0	14.2	10.0
AO1	0.0	0.0	0.0	13.9	10.0
PID1	5.0	100.0	0.0	89.5	10.0
Limit	1.0	1.0	1.0		

FIG. 4.7a
Operator's display for the monitoring of abnormal conditions.

On-Demand Tuning

One of the most successful techniques for providing on-demand tuning is based on using the relay oscillation tuning method to identify the process ultimate gain and ultimate period. (For a detailed discussion of tuning by computer, refer to Section 2.38 and for an explanation of the determination of ultimate process gain and ultimate time period, see Figures 2.38c and 2.38i in that section.)

This "relay oscillation" tuning method can be used on both self-regulating and integrating processes but is not suited for pure dead time processes, which require sample and hold control. Once the ultimate gain and ultimate time period of a loop are known, the tuning rules (see Sections 2.35 to 2.38) may be automatically applied to obtain the initial controller settings.

The relay oscillation technique of tuning is performed in an open loop under two-state control. Each time the controlled variable (process output) crosses the set point or initial value of the controlled variable, the relay is switched. Relay action causes the loop to oscillate at its ultimate period, T_u. The ultimate gain, K_u, is determined as the ratio between the amplitude of the two-state controller output and the amplitude of the oscillations in the controlled variable.

Upon completion of the relay oscillation test, the autotuner calculates the ultimate gain, ultimate period, dead time, time constant, static gain, and integrating gain of the process. The user can input the type of process (only self-regulating or integrating), the type of tuning rules to use, and perhaps an additional tuning factor such as the closed-loop time constant. The tuning rules are then applied and the calculated settings are provided for the controller, as illustrated in Figure 4.7b.

Adaptive Tuning

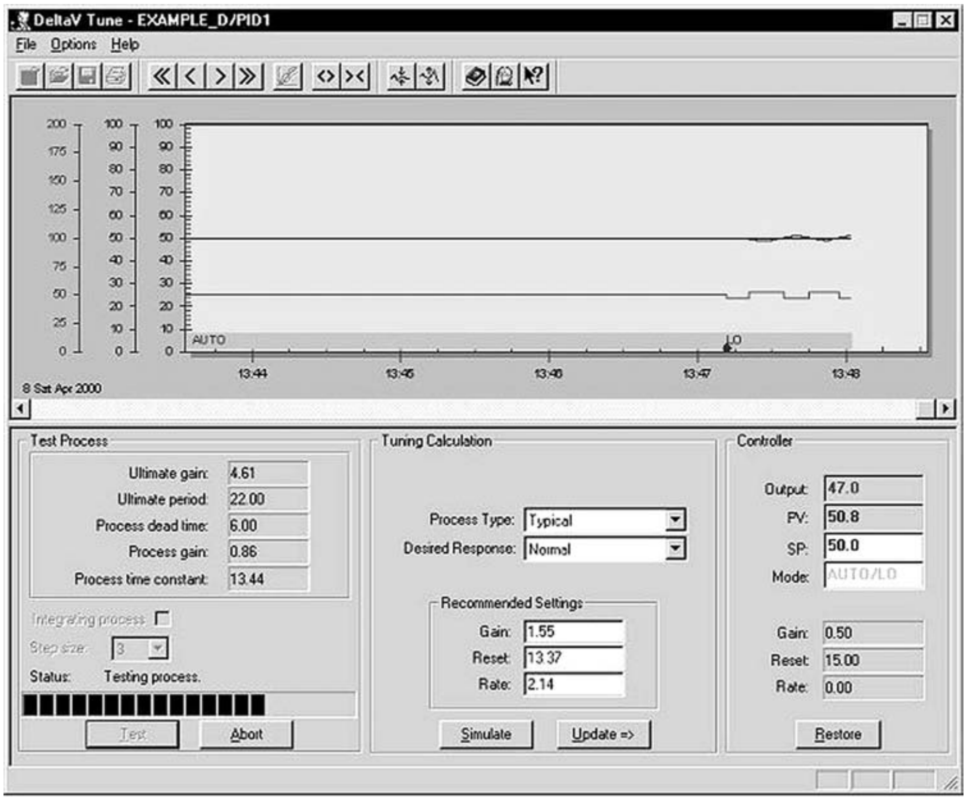
Adaptive tuning can be used to tune both feedback and feedforward controllers. There are two ways to design an adaptive PID controller: direct and indirect or identifier based. An identifier-based approach is advantageous when model switching and parameter interpolation are used. In this case a set of models is available together with a switching strategy among them, and parameter interpolation is based on the integrated squared error assigned to every value of the model parameter.

The adaptation cycle continues through a declared number of scans or until there is enough excitation on the inputs. As soon as a model has been updated, controller redesign takes place based on the updated model parameters. Adaptation can proceed for the whole model or separately for the feedback or feedforward portions of the model. The external excitations can be injected into the feedback loop automatically. The applied excitations can be a small change of the set point or controller output in either the Manual or in the Automatic modes.

Since the process model obtained is first-order plus dead time, any tuning rules can be used, including applied, typically lambda or IMC tuning. A general adaptive PID controller structure with model parameter interpolation is shown in Figure 4.7c.

FUZZY LOGIC CONTROL

Many industrial processes operated by humans cannot be automated using conventional control techniques since the



Test Process

Ultimate gain: 4.61
Ultimate period: 22.00
Process dead time: 6.00
Process gain: 0.86
Process time constant: 13.44
Integrating process ☐
Step size: 3
Status: Testing process.

Tuning Calculation

Process Type: Typical
Desired Response: Normal

Recommended Settings
Gain: 1.55
Reset: 13.37
Rate: 2.14

Controller

Output: 47.0
PV: 50.8
SP: 50.0
Mode: AUTO/LO

Gain: 0.50
Reset: 15.00
Rate: 0.00

FIG. 4.7b
Operator's display after "on-demand" tuning been performed. The lower line in the graph shows the oscillating step changes that are made in the manipulated variable (usually the control valve opening), and the upper line shows the response of the controlled variable.

performance of these controllers is often inferior to that of the operators. One of the reasons is that linear controllers, which are commonly used in conventional control, are not appropriate for the control of nonlinear processes.

Another reason is that humans aggregate various kinds of information and combine control strategies, which cannot be integrated into a single analytic control law. The underlying principle of knowledge-based (expert) control is to capture and implement experience and knowledge available from experts (e.g., process operators).

One would usually consider fuzzy logic control only after all attempts at controlling the process by optimized PID

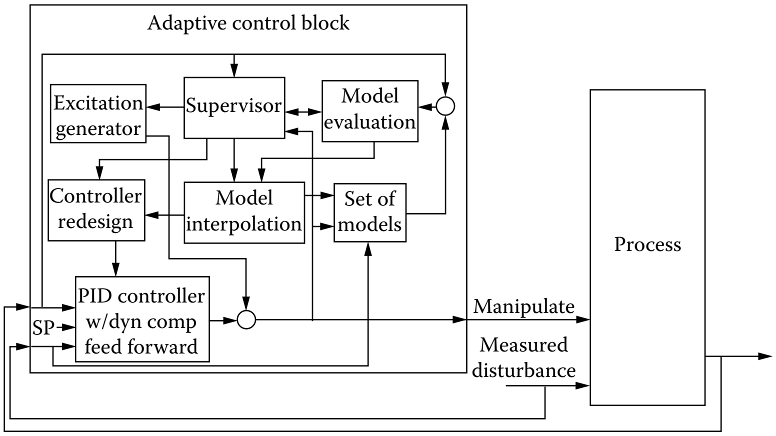


FIG. 4.7c
The configuration of the blocks of software in an adaptive PID controller, which includes both feedback and feedforward parts.

controllers have failed. Because fuzzy logic control requires exhaustive amounts of process data, the other essential requirement is that the process be somehow manually controllable. For a detailed discussion of fuzzy logic control, refer to Section 2.31.

The concept of fuzzy logic originated in 1965 when Lofti Zadeh of the University of California at Berkeley proposed what he called "fuzzy set theory." Theoretically, fuzzy logic-based controllers can provide significant performance improvements over conventional PID-based controllers. The nonlinearity introduced in the fuzzy PID controller can reduce overshoot significantly without compromising disturbance rejection capability, but fuzzy logic has not been widely adopted in the process industries.

One reason is that the tuning is difficult since no standard guidelines exist for the establishment of membership function scaling. To address this issue, many manufacturers provide preengineered versions of fuzzy control that require little user input to commission. In addition, manual tuning may be substituted by relay-based automatic tuning. Such controllers can be used as nonlinear single-loop replacements for conventional PID controllers.

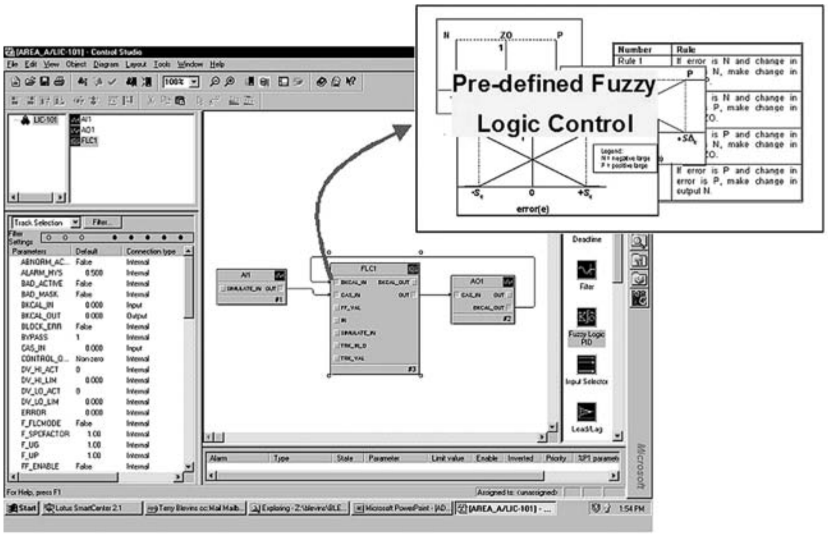
Fuzzy logic controllers combine the concept of membership functions with *rule inference* (see Section 2.31 for definitions and detailed explanations). The rules are used to determine a controller output based on the input membership functions. Tuning is accomplished through adjustment of the scaling associated with the membership functions. Figure 4.7d illustrates the software block components of a fuzzy logic controller that has already been defined.

The fuzzy logic controller in many cases provides increased performance and improved response with little or no overshoot relative to PID control. According to some sources, the improvement can range from 20% on the basis of integral of absolute error (IAE) to almost 50% for integral of absolute error multiplied by time.

MODEL PREDICTIVE CONTROL

In internal model controllers (IMC), MPC, and pole-cancellation, lambda, and Dahlin controllers, the model used by the controller is based on the controlled process (see Sections 2.13, 2.14, 2.17, and 2.19 for details). (Editor's note: According to some users, the MPC model relates to the inverse of the process model. Therefore, when the process dynamics are unchanged (set point changes), these controllers are effective; if the model is exactly known, they are better than PID. Therefore, people use them, *when the process cannot tolerate set-point overshoots*. Their main limitation is their sluggish response to changes in process dynamics (process load changes). In some applications it is the response to load upsets that is critical.)

Since the introduction of MPC in the 1980s until 2004, over 5000 plant sites have utilized this technology. Modern DCS software packages include embedded MPC control for small applications (models of eight inputs by eight outputs in size) as well as larger applications, which traditionally were handled externally to the DCS (layered MPC). Fast execution rates and ease of use in commissioning allow the new embedded MPC control to be used where previously only traditional control techniques based on PID control had been applied. Improvements in the user interface and in the tools for commissioning and operating the MPC allow some experienced process or control engineers to apply this technology without the assistance of outside experts.



Parameters	Default	Connection type
ABNORM_AC...	False	Internal
ALARM_HYS	0.500	Internal
BAD_ACTIVE	False	Internal
BAD_MASK	False	Internal
BKCAL_IN	0.000	Input
BKCAL_OUT	0.000	Output
BLOCK_ERR	False	Internal
BYPASS	1	Internal
CAS_IN	0.000	Input
CONTROL_O...	Non-zero	Internal
DV_HI_ACT	0	Internal
DV_HI_LIM	0.000	Internal
DV_LO_ACT	0	Internal
DV_LO_LIM	0.000	Internal
ERROR	0.000	Internal
F_FLCMODE	False	Internal
F_SPCFACTOR	1.00	Internal
F_UG	1.00	Internal
F_UP	1.00	Internal
FF_ENABLE	False	Internal

FIG. 4.7d
Application of preengineered fuzzy control.

In some simple MPC applications, the controller configuration might consist of only a few parameters. Examples of such parameters include:

- Controlled variable — Set-point limits
- Manipulated variable — Maximum rate of change, controller output limits
- Constraint parameter — Constraint high and low limits

The MPC controller is based on a dynamic model of the process. Traditionally, the dynamic model of the process is determined by manual manipulation of the control valve openings (process inputs). When the model is small enough for embedded MPCcontrol, automated testing of the process is provided as part of the MPC engineering tools.

Once data on the dynamics of the process is collected, the process model and controller the MPC, can be generated automatically. Both the step response of the MPC model and of the actual process can be displayed. An example of the overview step response model and a detail of an individual output response for a particular input are shown in Figure 4.7e.

MPC identification tools compare the measured process variables with those calculated based on the model, if the manipulated variables to both the model and the actual process are the same. This comparison may be displayed in different formats to allow the user to determine how good the model is, i.e., how closely the model dynamics match those of the actual process.

A simulated environment is used to evaluate the control response to both set-point and load changes before the MPC control is commissioned. Using the model of the process for simulation, it is possible to see the actual process response in advance of real time. Similarly, on very fast processes, the ability to simulate the control and process response slower than real time can also be valuable. The ability to adjust the speed of execution allows the control response to be quickly verified for a variety of operating conditions. Figure 4.7f illustrates the testing of the MPC model by viewing the simulated controlled variable response of the MPCmodel to a change in the manipulated variable.

Once the MPC control has been tested offline using simulation, then the control can be commissioned.

(Editor’s note: If the process model accurately reflects the process, the control response to set-point changes is likely to be good. The response to load disturbances is usually poorer if the load disturbances result in substantial changes in process dynamics; in such cases, adaptive control should be considered as a possibly better option.)

NEURAL NETWORK APPLICATIONS

Artificial neural networks (ANNs) can learn complex functional relations by generalizing from a limited amount of training data. Hence they can thus serve as black-box models

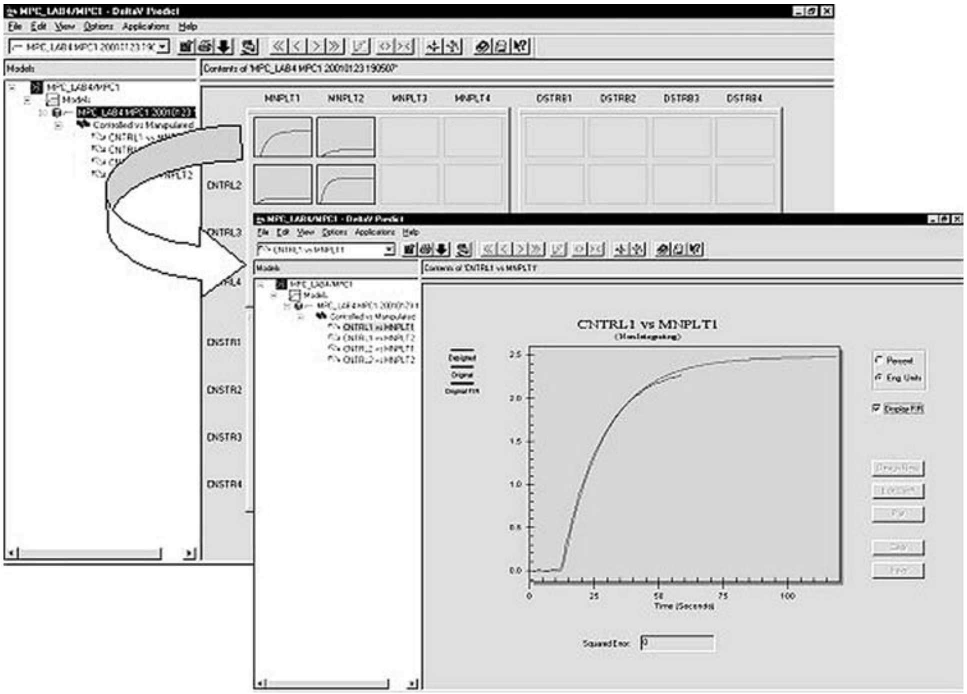
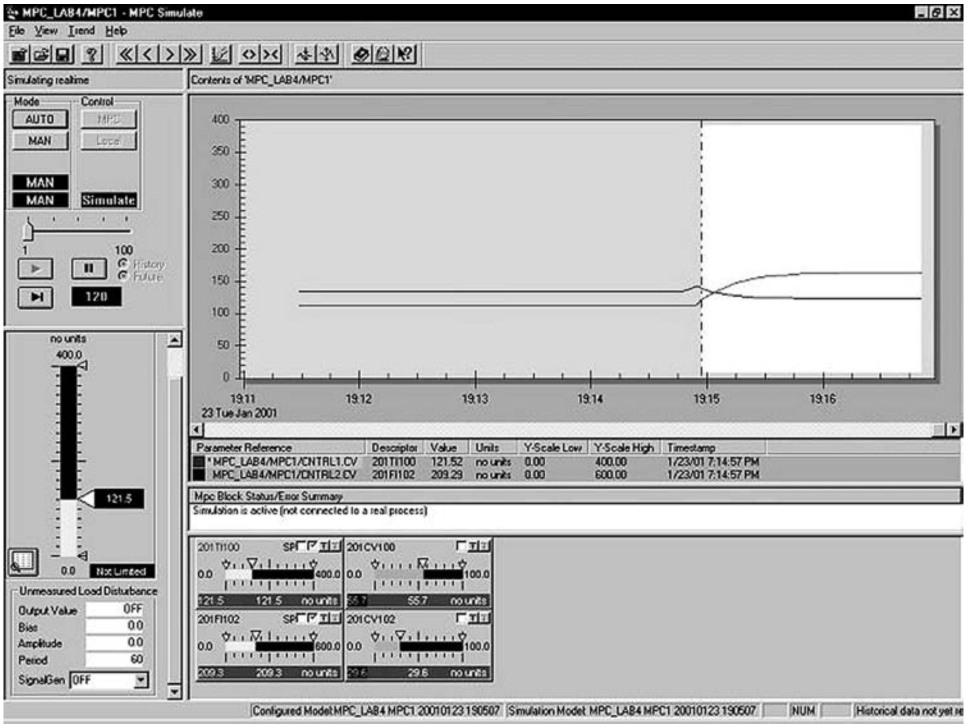


FIG. 4.7e

The step response of an accurate MPC model and of the actual process can be very similar if the process load is the same as it was at the time when the model was developed.



Parameter Reference	Descriptor	Value	Units	Y-Scale Low	Y-Scale High	Timestamp
* MPC_LAB4/MPC1/CNTRL1.CV	201TI100	121.52	no units	0.00	400.00	1/23/01 7:14:57 PM
MPC_LAB4/MPC1/CNTRL2.CV	201FI102	209.29	no units	0.00	600.00	1/23/01 7:14:57 PM

Mpc Block Status/Error Summary
Simulation is active (not connected to a real process)

Unmeasured Load Disturbance
Output Value OFF
Bias 0.0
Amplitude 0.0
Period 60
SignalGen OFF

Configured Model:MPC_LAB4 MPC1 20010123 190507 | Simulation Model: MPC_LAB4 MPC1 20010123 190507 | NUM | Historical data not yet re

FIG. 4.7f
The MPC model of the process is tested by introducing a change in the manipulated variable and viewing the resulting response of the controlled variable of the simulated process.

of nonlinear, multivariable static and dynamic systems and can be trained by the input–output data of these systems. Applications that are suitable for ANN solutions and the detailed theory of ANN are discussed in Section 2.18.

ANN is similar to fuzzy logic in that the mathematical model that relates the inputs and the outputs of the model does not need to be known. To use ANN, it is sufficient to know the behavior or response of the model. The main difference between fuzzy logic for property estimation and ANN is that the gains and functions cannot be modified in ANN; they can only be “trained” with data.

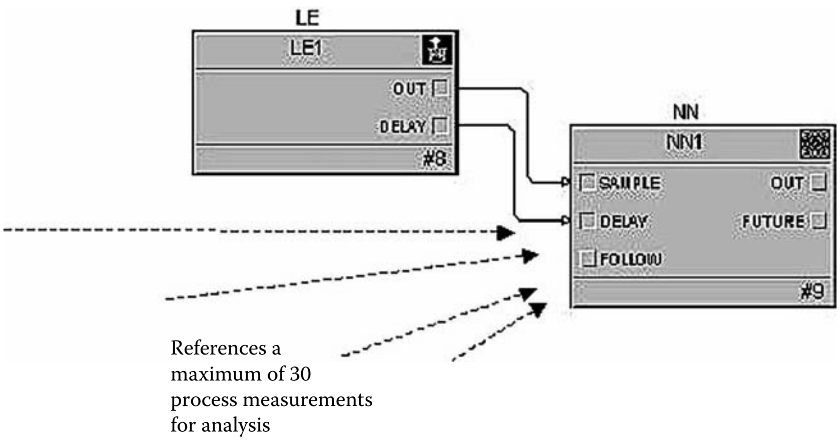
The use of estimated process variables can improve control and monitoring when critical measurements are only available through laboratory analysis**.** The intermittent laboratory analysis results can be used as inputs to provide online inferred or estimated variable values when actual analytical readings are not available. Neural networks are often used for process variable estimation even if the inferring relationships are nonlinear. ANN provides a nonlinear model of a process, which uses those input variables that indirectly influence the composition that is intermittently detected by laboratory analysis.

After a neural network has been commissioned, the offline laboratory analysis is still required to correct for unmeasured disturbances and changes in the process. Without this correction, the prediction provided by the model of the neural network will drift away from the true measurement value as time passes. For example, the laboratory analysis may be used as an input to the ANN model to obtain automatic adaptation of its prediction in response to changes in process.

Modern tools embedded in DCS control systems have reduced the complexity of designing and training neural networks. ANN can be integrated as a function block into a distributed control system and that block can be configured similary to other function blocks. The tools provided to train the neural network take into account the dynamic nature of the processes for which a neural network is developed.

Once a potential application for a neural network is identified, the first step in the ANN development is to configure the function block. For example, as a first step in configuring the ANN block, the user may browse the past history of the process to identify the variables that tend to influence composition and use them as function block parameters. As actual laboratory analysis takes place, the results are provided to the ANN model for updating the model of that analytical variable that is being estimated by ANN (Figure 4.7g).

Once the neural network block is configured and downloaded, all measurements used in the ANN block are assigned to a historian for collection. After a sufficient amount of process data have been collected, the operator can graphically view the historic data and make corrections by adding/removing inputs. Once the input data have been screened, the ANN model can be generated. During this generation process, the dynamics of the input delays and output sensitivity to every input are calculated. To help the user determine the accuracy of the neural



References a maximum of 30 process measurements for analysis

FIG. 4.7g
Neural networks can be used to estimate analytical laboratory measurements on the basis of the past history of such readings by sending the associated laboratory readings to the ANN block. Later composition measurements from the laboratory are used to update the ANN model.

network, the value calculated by the neural network may be compared to the lab analysis, as illustrated in Figure 4.7h.

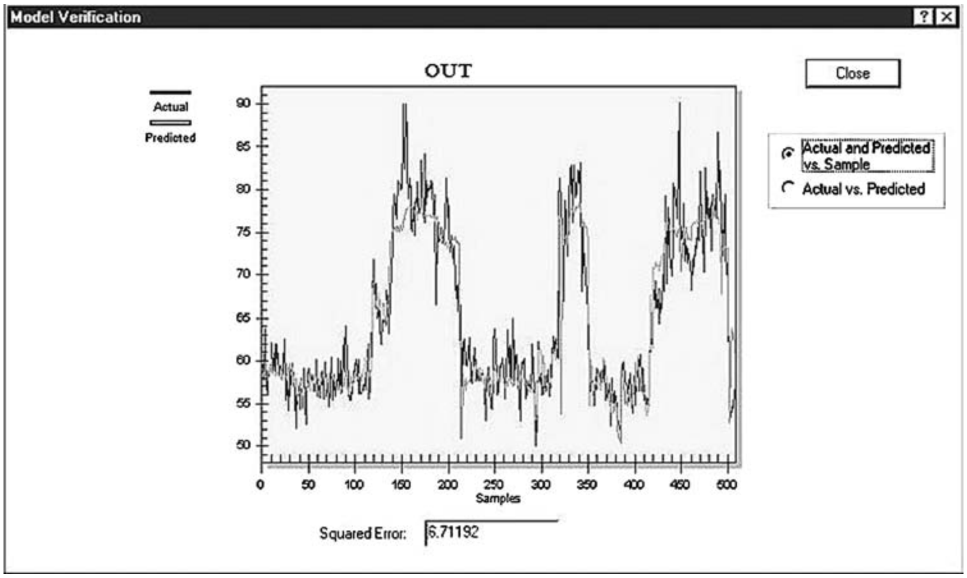
PROCESS AND CONTROL SIMULATION

Process simulation packages are used to train operators prior to plant startup; they are also used to examine the dynamic response and performance of a process during operation, and finally, they are used to prepare the operators for complex and critical shutdown procedures.

The cost of developing the simulation software rises exponentially with the required accuracy of the process model. Yet, it is desirable that the model reflect not only the process dynamics but also the operation of the actually implemented control system and that it incorporate the simulation of process upsets to train the operators in a real-time setting. It is also important that the keyboard, strokes, sequences, mouse, and touch screen used during simulation be identical to the actual ones.

Some commercial systems allow all features of the process control system to be executed on a single PC platform or distributed between multiple computers. This high-fidelity process control simulation capability may be used to check out control system logic and operator interface as well as to train operators on the continuous and discrete control, diagnostic, and alarming under a variety of conditions.

The software and the associated configuration of a control system may be designed to be used in multiple operating environments without change. When this approach is taken,



● Actual and Predicted vs. Sample
○ Actual vs. Predicted

Squared Error: 6.71192

FIG. 4.7h
Comparison between the actual laboratory readings and the ANN-predicted values of composition.

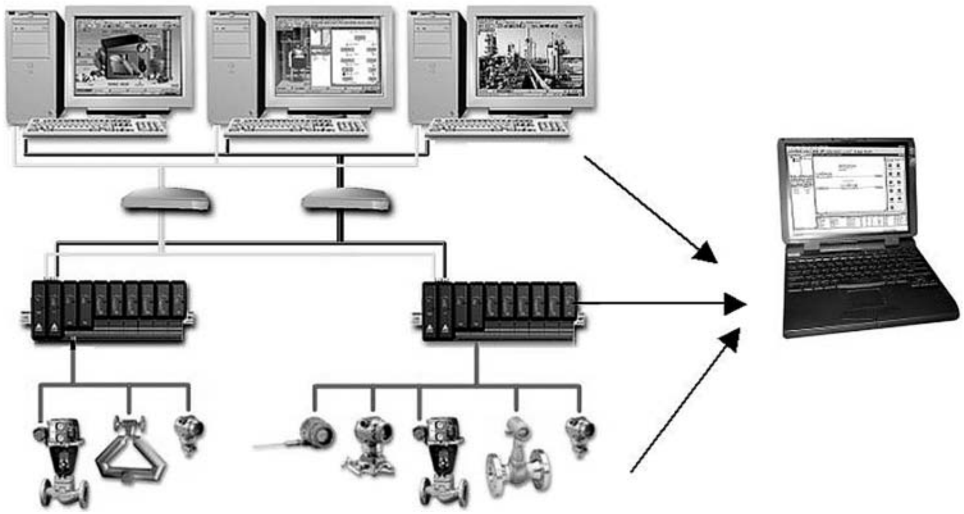


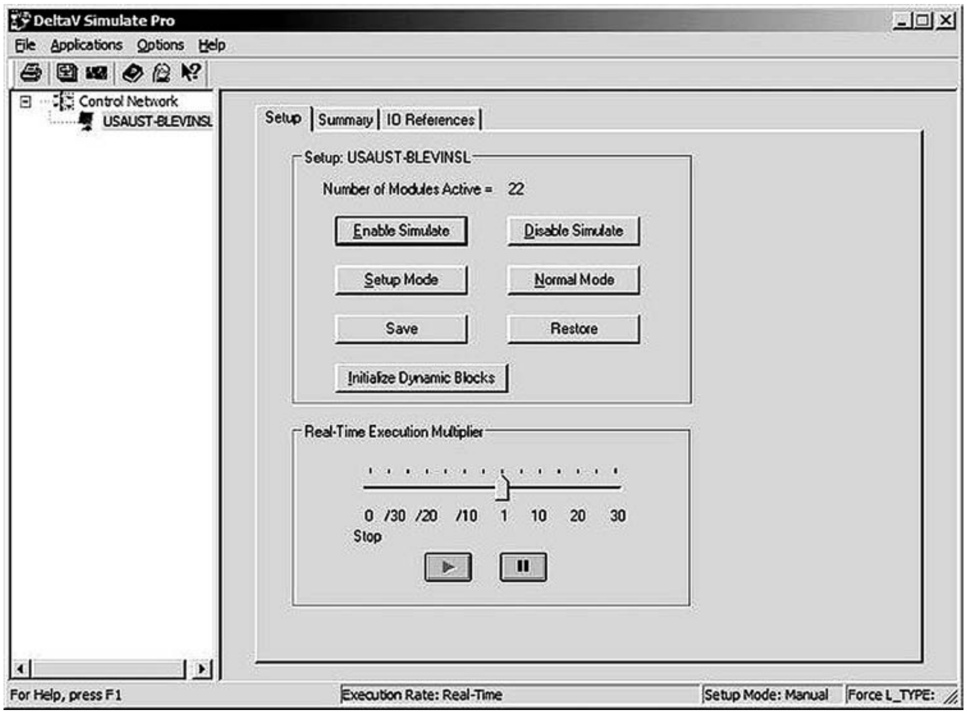
FIG. 4.7i
Operator training can utilize single or multiple PCs loaded with the same software package representing the process and its control system. (Courtesy Emerson Process Management.)

it is possible to distribute control system functionality in a variety of ways without needing to reengineer the software or reconfigure the associated applications.

All features of the control system can be combined on a single platform or distributed between multiple PCs to support configuration and dynamic checkout of the controls and interface associated with a process control system, as illustrated in Figure 4.7i.

When a control system utilizes the Fieldbus Foundation function block architecture, a SIMULATE parameter is available in each I/O function block. Through this parameter, a value and status may be provided to I/O function blocks to simulate the field measurement or actuator in the control system.

This capability may be used to check out control logic and operator displays. Also, where the control systems support



Number of Modules Active = 22

Execution Rate: Real-Time

Setup Mode: Manual

Force L_TYPE:

FIG. 4.7j
Display interface for simulation coordination.

the object linking and embedding (OLE) interface for processing control (OPC), the process simulation package may write the simulated value and status. This capability, combined with the ability to redistribute control functionality without reconfiguration, provides a strong foundation for the integration of high-fidelity simulation with control system simulation in a single or multi-PC environment.

Depending on the modeling rigors and computing resources, the process simulation can potentially run faster or slower than real time. To allow the control simulation to match the process simulation execution, the control system executes function blocks faster or slower. In addition, a trainer or an application may coordinate the execution of control and process simulation. An example interface for such coordination is shown in Figure 4.7j.

CONCLUSION

Modern distributed control systems can provide embedded advanced control solutions. The DCS system supports the advanced control capability similarly to the support provided for the traditional control tools. Among the major advances in DCS system designs are the easy-to-use engineering and commissioning tools.

Reference

1. Ruel, M., *Instrument Engineers' Handbook,* Vol. 3, Section 5.6, Jointly published by CRC Press and ISA, 2002.

Bibliography

Astrom, K. J., and Hagglund, T., *PID Controllers: Theory, Design, and Tuning,* Research Triangle Park, NC: ISA, 1995.

Blevins, T., McMillan, G., and Wheatley, R., "The Benefits of Combining High Fidelity Process and Control System Simulation," in Proceedings of ISA Expr/2000, Chicago, October 2000.

Blevins, T., and Ochoa, D., "Statistical Analysis Rides the Bus," *Intech,* pp. 48–59, March 2000.

Bonavita, N., "Benchmarking of Advanced Technologies for Process Control: an Industrial Perspective," Proceedings of the American Control Conference*,* Vol. 6, pp. 4328–4329, 2001.

Bonavita, N., Martini, R., and Grosso, T., "A step-by-step approach to advanced process control," *Hydrocarbon Processing,* Vol. 82(10), pp. 69–74, 2003.

Canney, W. M., "Special Report: The Future of Advanced Process Control Promises More Benefits and Sustained Value," *Oil and Gas Journal,* Vol. 101(16), pp. 48–55, 2003.

Hokanson, D., Mun, H.C., and Rebeck, J., "Integrating Advanced Process Control on a Grassroots Project," *Hydrocarbon Processing,* Vol. 78(6), p. 63, 1999.

Kuchle, D., "Multivariable Control on a Single-Loop Budget," *Control Solutions,* December 2001.

Narendra, K. S., and Balakrishnan, J., "Adaptive Control Using Multiple Models," *IEEE Transactions on Automatic Control,* Vol. 42, No. 2, pp.177–187, February 1997.

Macias-Hernandez, J. J., and Feliu, J. A., "Dynamic Simulation and OPC," *Hydrocarbon Engineering,* Vol. 8(9), pp. 53–58, 2003.

Mehta, A., and Tzovla, V., "Automating the Approach to Development and Online Operation of Soft Sensors," *Intech*, pp. 30–33, September 2002.

Qin, S. J., "Auto-Tuned Fuzzy Logic Control," American Control Conference, Baltimore, 1994.

Wang, H., "Introduction: Advanced Process Modeling and Control in Paper-Making," *Measurement and Control*, Vol. 32(8), p. 228, 1999.

Xionglin, L., "Problems and Countermeasures for Engineering Application of Advanced Process Control," *Petroleum Refinery Engineering,* Vol. 101(16), pp. 48–55, 2003.

4.8 DCS: Installation and Commissioning

T. L. BLEVINS, M. NIXON (2005)

Installation Costs:	Based on the local labor market, costs vary widely. If the prevailing labor rates are not known, union halls and engineering contractors in the area can help. Instrument installation requires high-skilled tradesmen, some training and good supervision, which can range from 15 to 30% of labor costs.
Fieldbus Costs:	Installation and commissioning costs are reduced by 20 to 50% because fieldbus devices can identify themselves, require fewer terminations, and can be managed remotely.
Types of Services:	A. Installation B. Commissioning
Partial List of Service Suppliers:	ABB (A, B) (www.abb.com) Anvil (A, B) (www.anvilcorp.com) Bay Tec (A, B) (www.bay-tec.com) Biwater (B) (www.biwater.co.uk) Control Automation Technologies (A,B) (www.plantservice.com) Emerson (A, B) (www.EasyDeltaV.com) Energy Control Systems (A) (www.Energycontrosystems.com) Honeywell (A, B) (www.honeywell.com) Invensys (A, B) (www.invensys.com) MPR Technology (A, B) (www.mpr.com) O'Brien & Gere (B) (www.obg.com) Premier System Integrators (B) (www.premier-systems.com) R.G. Automation (A, B) (www.rgautomation.uk.co) Siemens (A, B) (www.siemens.com) Yokogawa (A,B) (www.yokogawa.com/us)

INTRODUCTION

The installation and commissioning of a distributed control system impact the cost and schedule of a new plant or the expansion of an existing plant. Also, the work done during this part of construction can have a long-term impact on the plant operation and system reliability. The steps associated with installation and commissioning are much the same whether the control system is provided for new plant construction or as an upgrade to an existing system. However, the implementation details of installation and commissioning are dependent on the technology incorporated in the control system, for example, fieldbus vs. traditional actuators. Also, in some cases the manner in which the implementation is done will be dependent on the design of the control system and field equipment. Thus, the steps advised in this section may be used as a guideline in examining the detailed requirements at each key step of installation and commissioning.

In most cases the system configuration is completed before installation begins. Even if the system hardware is not available, modern control systems allow the complete system configuration and execute simulation including input-output (I/O). Using this capability, the control logic, alarm setup, and operator displays may be checked out in advance.

Also, using process and control simulation, it is possible to train plant operators on the system interface, the value of which will much depend on the quality of the process model. Thus, all efforts should be made that by the time the system is ready to install, the operators are familiar with the control system and the operator interface. Also, there should be good confidence that the system configuration is complete except for the I/O assignment, which must be verified during installation using the actual field devices.

Commissioning and startup will benefit if the design is well documented, particularly in the area of electrical and instrument (E&I) drawings and supervision, and if the commissioning team is well trained and the operators are experienced on the particular process. It is particularly important that the various software packages used in the intelligent workstations be fully understood and that the communication between intelligent devices from different manufacturers, including analyzers, be fully tested.

INSTALLATION

System installation is of prime importance for personnel safety and system availability. The steps involved in the installation of a distributed control system include the following:

- Power and grounding
- System assembly
- Heating, ventilating, and air conditioning (HVAC) and heat tracing
- Field wiring and checkout
- Bus installation

During control system installation, multiple steps are executed in parallel to reduce the time required for installation. The teams assembled for electrical and instrumentation checkout work closely with the engineering firm responsible for system design and documentation. Work during installation acts as further training for the operators as they participate in the control system checkout and correction of any problems found in the installation. During commissioning, the loop sheets, piping and instrumentation diagrams (P&IDs), and documents showing further details on the installation will be marked up. Where the changes impact the system configuration, these will also be noted and implemented so installation may be verified.

Power and Grounding

If a process is critical to plant operation and process interruptions must be minimized, secondary power sources should be included in the plant design. Electrical noise effects can be minimized by using isolated AC power sources, grounding at single points, minimizing undue influence on signal wiring from stray magnetic fields, and selecting appropriate cables and pathways, including adequate separation. Proper earth grounding is important to user safety and efficient operation of a control system. Installation methods to obtain good power and grounding are defined by industry-accepted standards.[1] A typical ground network is shown in Figure 4.8a.

During the installation, checks should be done to ensure the quality of the power and ground for the control system. This will include verifying the voltage levels, designed load on isolation transformer, and noise level.

System Assembly

In many cases the components of the distributed control system are shipped directly to the construction site for assembly. The control system may be designed to allow controller, I/O, and associated wire trays to be wall mounted, see Figure 4.8b.

Also, specific varieties of enclosures are used when the control system is installed in harsh environmental conditions. System assembly should follow manufacturer guidelines on heating considerations and device placement in the enclosure. In particular, placement of the enclosure or equipment on a wall should take into account the space needed to service, remove, or replace components. During installation, the temperature in the enclosure or room in which the equipment is mounted should be monitored to ensure proper dissipation of internal heat.

System cabinets and components assembled during installation should be labeled for quick identification during

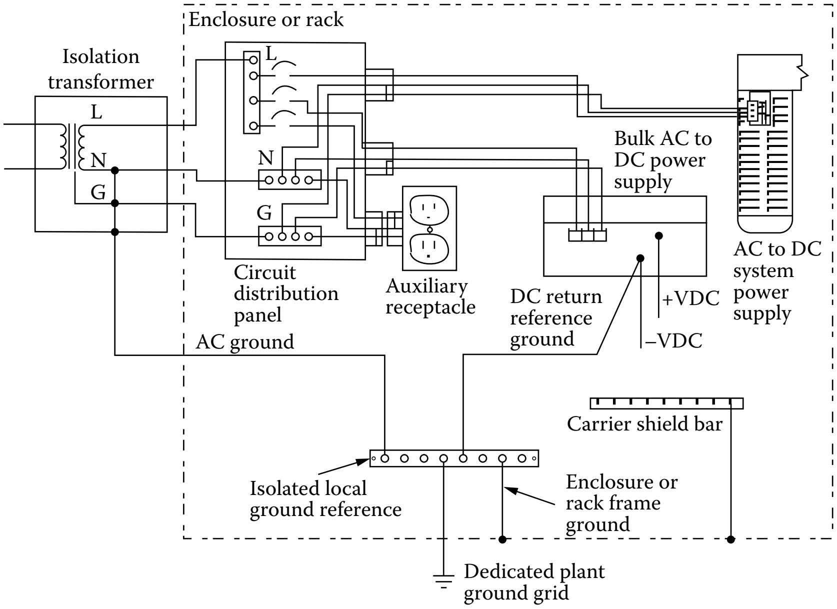


FIG. 4.8a
Typical, single-point ground network for a single enclosure.

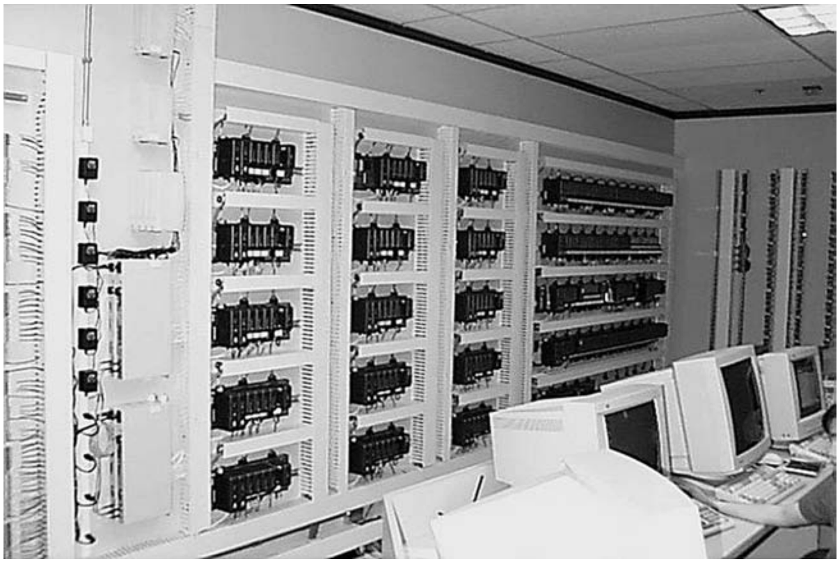


FIG. 4.8b
Example of an installation utilizing wall mounting of I/O, wire trays, and other system components.

installation. Once the system is fully assembled, the correct operation with the plant configuration should be verified. Any system problems should be addressed to avoid delays in using the system for I/O checkout.

HVAC and Heat Tracing

Environmental conditions can affect the operation of electronic equipment. Temperature, humidity, dust (including carbon), and corrosive vapors can cause gradual performance degradation, intermittent failures, and malfunctions.

To ensure maximum system efficiency and reliability, the air conditioning system including air filtering or scrubbers, if required, should be fully in place before the control system is installed. During installation, care should be taken to avoid open doors to the process, which could allow contaminates to enter the rack room or control room area. All environmental conditions must be maintained in accordance with the manufacturer's recommendations.

Where controllers are to be mounted in the field, care should be taken to ensure that enclosures remain closed and that any purge air or heat tracing required to maintain environmental conditions are in correct working order.

The standard described in Reference 2 defines the environment for control system operation. Such conditions include ambient temperature, relative humidity, and electromagnetic interference.

Field Wiring and Checkout

Field wiring should be properly installed to negate electromagnetic interference (EMI) and other electrical noise that can adversely affect the instruments controlling the process. Signal cabling should be kept away from AC power lines, transformers, rotating electrical equipment, or other high-power machinery to reduce the possibility of electromagnetic interference being induced on analog and discrete signals. Industry standards[3] describe noise identification and recommended wiring practices.

After field wiring is installed, continuity checks should be made from the field device to the I/O terminal strip. In addition, electrical isolation checks should be performed between the signal and ground wiring.

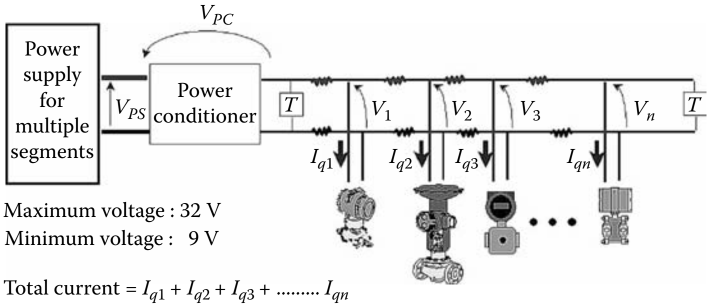
Injecting a signal at the transmitter end and observing the measured value at the operator screen may check the operation of traditional transmitter wiring. For traditional valves, the specified valve position should be verified in the field to ensure the software setting for increase open/close is correctly set to deliver implied value position.

As part of the instrument checkout, it should be verified that the alarm setting and the configured control setting match those specified on the loop sheet. Also, as part of the control checkout, the operation of the control after power failure or system download should be verified.

Bus Installation

To achieve optimum system operation, digital I/O bus systems, such as AS-Interface, DeviceNet, Foundation Fieldbus, and Profibus, must be properly installed. Otherwise, EMI and other electrical noise can adversely affect the capabilities of the instruments controlling the process.

Once a bus wiring is installed, its operation should be verified before commissioning. Most manufacturers provide a verification checklist that includes such things as a measuring of cable voltage and current capacity and verifying of



Maximum voltage : 32 V
Minimum voltage : 9 V

Total current = $I_{q1} + I_{q2} + I_{q3} + \dots\dots\dots I_{qn}$

Maximum current limited by the power conditioner.
Example : 400 mA

FIG. 4.8c
Voltage and current limits of fieldbus installation.

terminators on each end of the segment (Figure 4.8c). After these checks, each fieldbus device may be quickly commissioned and its operation status determined at the operator station using the device diagnostics.

Installing HART Networks

The HART signal is a modulated analog signal, and HART communication is designed to be compatible with existing or new 4- to 20-mA analog process control systems. The HART signaling occurs in a separate frequency band from the 4- to 20-mA analog signal and therefore, in most cases, does not interfere with the operation of the analog control loop. While HART uses modems designed for analog telephone networks and therefore is forgiving, the capacitance of the wire does limit the maximum allowable cable lengths (Table 4.8d).[4]

Depending on the type of transmitter, final control element, or handheld tool used, the HART signal is modulated by varying the loop current or by directly modulating the voltage of the power supply into the loop, as shown in Figure 4.8e.

On a three-wire or multi-drop transmitter loop the current sense resistor is 250 ohms. For DCS, final control elements and handheld tools, the power supply voltage drop is used as the HART signal. For impedance and other requirements of the process control system installation to guarantee reliable HART operation, refer to Table 4.8f.

TABLE 4.8d
Allowable Cable Lengths for 1.02 mm (#18 AWG) Shielded Twisted-Pair Cable[4]

	Cable Capacitance, pf/m			
No. Devices	*65*	*95*	*160*	*225*
1	2769 m	2000 m	1292 m	985 m
5	2462 m	1815 m	1138 m	892 m
10	2154 m	1600 m	1015 m	769 m
15	1846 m	1415 m	892 m	708 m

COMMISSIONING

The commissioning of the control system occurs once process construction is complete. Often this work is staged to each process area or piece of major equipment to be commissioned. The steps involved in the commissioning of a distributed control system generally include the following:

- Training and preliminary checkout
- Fieldbus testing
- Process startup
- Tuning process control loops
- Safety systems
- Advanced control

The plant operators play a key role in commissioning. During this phase of the startup, the engineering firm that designed the plant, the E&I contractor, plant engineering, the operators, and equipment suppliers work as a team to find and correct any process or instrumentation problems. Control loop commissioning early in the startup plays an important part in getting the plant to a sustainable level of operation. Any mistakes in system configuration should be corrected during commissioning and reflected in the final documentation of the system. In some cases, these configuration changes are automatically traced to allow the time and person who made the change to be documented.

Training and Preliminary Checkout

Training is important, particularly if digital communication systems are used and the E&I contractor and/or the commissioning team has no experience in troubleshooting a fieldbus system. In that case, they have to learn unfamiliar definitions, understand unfamiliar drawings, devise check procedures, and devise configuration software and the approach to troubleshooting a communication network. Training is the only way to overcome the defensiveness of the inexperienced, which in the long run will save a lot of rework. On some critical processes, it is

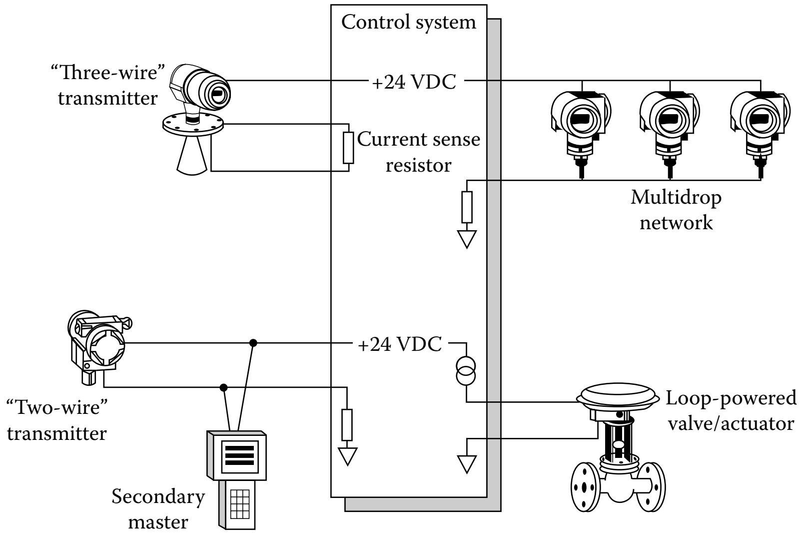


FIG. 4.8e
For multi-drop networks and three-wire transmitters, the loop current (for other devices, the voltage of the power supply) is modulated in a HART installation.[4]

possible to initially test the system using water as the process fluid. During this time, initial loop tuning constants should be determined, and control loop operation and interlock or permissive limit functioning should be verified. Such testing may reveal process or control problems that can be corrected before startup.

Fieldbus Testing

The testing of fieldbus-based process control systems is done in three steps. The first level of testing is called bus segment testing. This testing requires specialized equipment and troubleshooting skills and serves to check the physical installation of the system, including power supplies, terminations, and wire lengths. Prior to starting the testing, it is advisable to identify a qualified contact at each hardware vendor whose field device is part of the system, so that when (not if, when) integration issues requiring expertise arise, one knows whom to call.

The second step is to confirm that the field devices are correctly wired and addressed, their I/O is functioning, and they are communicating on the bus. If some of the field devices perform control or logic functions based on the application software, these too should be checked. When final control elements or other field devices are supposed to provide such information as on/off status and fault or diagnostic information using a data map, which must be unmapped by the other devices on the bus, data-map testing is also part of this second step. The third and last step is operational testing, which confirms that the bus and the PLC, DCS, and other devices that are connected to it are capable of operating as a unit and performing the required control and logic functions. The operational test should also include the checking of cycle times and the testing of all backup systems, data server connections, and bus redundancy.

TABLE 4.8f
Installation Requirements of a HART Network[4]

A network must have at least one, typically only one, low impedance device. Total loop resistance must be between 170 and 600 Ω.
A network must have no more than one device varying the 4–20 mA signal.
Only one secondary device is allowed.
Cable run lengths to 3000 m for single-pair cable and 1500 m for multiconductor cables are typical. Actual length depends on the number of multidropped field devices and the quality of the cable used (see Table 4.8d).
Low capacitance shielded twisted pair cable is recommended. However, HART has been successfully used over poor quality, unshielded wiring. Do not replace wiring until HART communication has been attempted.
Linear power supplies should be used.
HART is compatible with intrinsic safety (IS) rules and HART communicates across most IS barriers. In general, zener diode barriers are acceptable as they do not prevent two-way communication. However, isolating barriers must be HART compatible (i.e., some isolating barriers support one-way communication only).

Process Startup

The startup of a process requires that a person, such as the chief operator, together with a process engineer who is fully familiar with the process and the control system, are both present to coordinate the start-up activities. During the initial phase of a startup, the production rate is usually low and process limitations may dictate that operating conditions be gradually ramped while the process control loops are in manual. During this phase, it is possible to once again verify the operation of the process I/O and the associated field measurements and actuators.

If the quality of prior testing was inferior, at the time of control system startup it is possible that a safety subsystem or control interlock will initiate false trips or that interlocks may hamper the startup. In such cases the source of the problem should be addressed rather than attempting to bypass the input that initiated the trip or interlock.

For this reason, fieldbus devices are designed to allow I/O simulation to be enabled only if a physical key is present on the device. Similar protection is provided in many process control systems through the administration of system security that requires a certain level of authorization to change a critical parameter.

Commissioning of Control Loops

During startup, the process control loops may be commissioned once the process loads reach normal operation and therefore the controller set point and dynamics are close to normal. Before switching any loop to automatic or initiating self-tuning in the automatic mode, it is essential to check the response of the final control element. Therefore, before switching to on-demand tuning, it should be checked that the final control element responds to a change in the controller output when initiated in the manual mode.

Also, such a test may be used to verify that the PID Direct/Reverse setting is correct. Similarly, it is advisable to verify that the final control elements are not sticking and that their hysteresis is not excessive.

Advanced control tools for process and control monitoring are available to determine which loops have been commissioned and have performed as expected. Information collected on control utilization and variability index may be used to determine which loops may need to be closely examined to detect why the loop is unstable, is noisy, does not hold its set point, or is not being used as designed.

Advanced Control

Only after a process has been successfully started up, base loops fully commissioned, and production fully established, is it possible to commission the advanced control features of the system used for process control. A key step in commissioning various optimized, multi-variable or model-based control (MBC) systems is to make sure that the process variable measurements used are reliable and repeatable. Similarly, it is important to verify that the speed of response, characteristics, and dead time of the final control elements are all acceptable for proper loop performance.

CONCLUSION

Long-term reliability of DCS systems may be improved by paying attention to the power and grounding, environmental, and field installation requirements. Industrial standards or manufacturer's guidelines exist for such purposes. Advanced control tools such as on-demand self-tuning and process and control monitoring can be used to quickly commission the control loops and to identify loops that require closer examination.

References

1. IEEE Standard 1100-1992, *Recommended Practice for Power and Grounding Sensitive Electronic Equipment*.
2. ISA Standard S51.1-1979, *Process Instrumentation Terminology.*
3. IEEE 518-1982, *IEEE Guide for the Installation of Electrical Equipment to Minimize Noise Inputs to Controllers from External Source*.
4. Pratt, W. A., *Instrument Engineers' Handbook*, Vol. 3, Section 4.11, Boca Raton, FL: CRC Press, 2002.

Bibliography

Byeon, H. W., and Ulrey, J. L., "Designing a Standard Thermal Power Plant for Daily Startup/Shutdown: The HP Bypass Control and Safety Function," *ISA Transactions,* Vol. 36(1), p. 71, 1997.

DeltaV Books Online, Site Preparation and Hardware Installation, 1996.

Harrold, D., "How to Accelerate Commissioning: Under Fire at Startup?" *Control Engineering,* Vol. 50(6), pp. 42–47, 2003.

Hokanson, D., Mun, H. C., and Rebeck, J., "Integrating Advanced Process Control on a Grassroots Project," *Hydrocarbon Processing*, Vol. 78(6), p. 63, 1999.

Nisenfeld, A. E., *Batch Control: Practical Guides for Measurement and Control,* Research Triangle Park, NC: ISA, 1996.

Rautenbach, A., "Site Acceptance Testing and Commissioning of Process Control Systems," *Elektron Journal — South African Institute of Electrical Engineers*, Vol. 19(5), p. 40–41, 2002.

Weaver, M. D., Hilder, S., and Brown, D. R., "Installation of an Integrated Turbine-Generator Control System for a Pulp Mill," *IEEE Transactions on Industry Applications,* Vol. 39(3), pp. 899–909, 2003.

Xiaojian, L., Mengjun, Y., Zhenren, Z., and Xiaonian, H., "Design and Realization of a Control System for Petroleum Product Distillation Based on Control Local Area Network Bus," *Petroleum Processing and Petrochemicals*, Vol. 34(2), pp. 62–65, 2003.

4.9 DCS: Integration with Buses and Networks

T. L. BLEVINS, M. NIXON (2005)

Industry Fieldbus Standards:	A. OPC Foundation (www.opcfoundation.com) B. Fieldbus Foundation (www.fieldbus.org) C. PROFIBUS (www.profibus.com) D. AS-Interface (www.as-interface.net) E. DeviceNet (www.odva.org) F. MODBUS (www.modbus.org) G. WorldFIP (www.worldfip.org)
Partial List of Suppliers:	ABB (A, C, D, E, F, G) (www.abb.com) Allen-Bradley (A, B, C, E, F, G) (www.ab.com) Emerson (A, B, C, D, E, F) (www.EasyDeltaV.com) Honeywell (A, B, C, E, F) (www.honeywell.com) Invensys (A, B, C, F) (www.invensys.com) National Instruments (A, B, E) (www.ni.com) Schneider Electric (A, D, F, G) Siemens (A, C, D, F, G) (www.siemens.com) Yokogawa (A, B, C, E, F) (www.yokogawa.com/us)

INTRODUCTION

The adoption of the IEC1158-2 fieldbus standard by the major DCS manufacturers has ushered in the next generation of control and automation products and systems. Based on this standard, fieldbus capability may be integrated into a DCS system to provide:

- Advanced functions added to field instruments
- Expanded view for the operator
- Reduced wiring and installation costs
- Reduced I/O equipment by one-half or more
- Increased information flow to enable automation of engineering, maintenance, and support functions

Similarly, the Ethernet and OPC industry standard have provided DCS manufacturers with a standard means to access information in a process control system. Through the active use of control system data in a plant information system, the operation benefits that may be achieved are:

- Improvement in production scheduling
- Better inventory control
- Consolidation of maintenance- and operation-related information from multiple sites

However, the integration of networks and fieldbus has presented DCS manufacturers with many technical and commercial challenges. Many DCS systems on the market today were designed before the creation of these standards for bus and network integration. Also, the integration of bus technology is complicated by the fact that the IEC1158 standard covers eight fieldbus architectures. Each of these fieldbus architectures differs significantly in physical interface, communication protocol, and the way information is modeled within the field device.

Several approaches have been taken by DCS manufacturers on the integration of buses and network technology. The manner in which this is done can influence the engineering effort that is required to implement and maintain such interfaces.

Figure 4.9a illustrates the control and automation communication network of an industrial plant, which connects the components of the total system.[1] The device-level fieldbuses connect the process control sensors, the final control actuators, and their controllers. The cell-level fieldbuses coordinate, trend, monitor, calibrate, and reconfigure the process controllers, while at the plant level, the overall production strategy is planned. At the device level, a variety of standards of real-time protocols exists.

BUS INTEGRATION

(Editor's note: A single, internationally accepted fieldbus protocol has not yet evolved. As shown by the listing of standards A to G above, there are a variety of technologies and interfaces

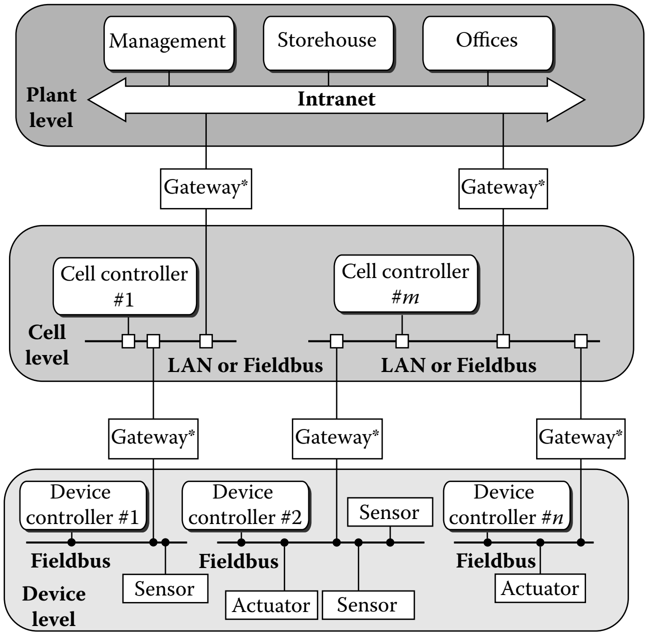


FIG. 4.9a
*The three levels of communication networks used in the automation of industrial plants are 1) the device level, 2) the cell level, and 3) the plant level.[1] *Sometimes not required (see Figure 4.9g).*

that serve the communication among intelligent field devices. Some users refer to this state as a “digital Babel” because in addition to these listed major protocols, a number of proprietary protocols also exist.

Some proprietary networks were designed for specific purposes and therefore lack flexibility in equipment selection and command capabilities. Some were introduced to integrate bus architectures with older DCS architectures. Such linking devices might connect fieldbuses on one side and MODBUS and/or 4-20ma interfaces on the other side. This style of interface tends to be limited by bandwidth, parameter type matching, and, in some cases, the meaning of the parameter value itself.

While much progress has been made, the goal of international standardization is yet to be reached. Today, for example, the digital representation of a 0% or a 50% measurement or the indication of the manual-automatic-cascade status of a controller is still not identical within the various protocols.)

The bus design largely determines the types of devices and applications that may be effectively addressed to implement an application. For example, the Profibus™, AS-Interface, and DeviceNet bus technologies are optimized for fast access of discrete information and best address applications such as motor control and discrete part manufacturing.

Conversely, buses such as Foundation Fieldbus® and Profibus PA address the specific requirements associated with continuous control applications. As such, bus technologies were initially integrated into older DCS as subsystems to address dedicated applications such as motor control interfacing.

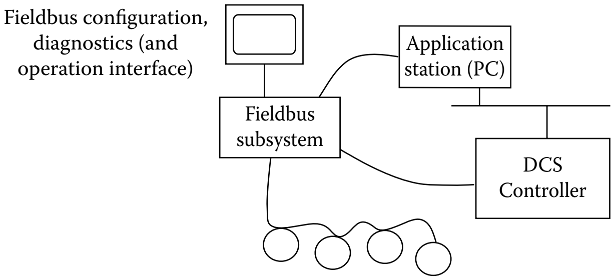


FIG. 4.9b
Integration of open fieldbus subsystems into a traditional DCS.

In such cases, the manufacturer of the fieldbus subsystem was different from that of the DCS supplier. Where such fieldbus subsystems are integrated with the DCS, this is typically done utilizing a serial interface to the DCS controller or interface station, as illustrated in Figure 4.9b.

The integration of fieldbuses often limit the effective use of fieldbus technology because of the differences in design as summarized below:

- Different engineering interfaces must be used to configure and diagnose operation problems.
- It is necessary to engineer how fieldbus system information will be mapped into the DCS controller. Any change in the DCS or fieldbus system may impact this design.
- One interface is not available to view the control and monitoring functions that are split between the DCS and the fieldbus devices. This can complicate engineering, commissioning, and troubleshooting of the control system.

DCS Fieldbus Support

Because of the limitations associated with the integration of fieldbus as separate subsystems, most modern DCS systems provide native support for a number of fieldbus technologies. Such support typically consists of standard interface cards that may be used with DCS controllers. These cards are designed to support the physical and communication interface associated with each fieldbus technology.

In taking this approach traditional I/O can be intermixed on an I/O card basis with standard device buses such as AS-Interface bus, Profibus DP, DeviceNet, Foundation Fieldbus, and HART. Furthermore, the same configuration, diagnostic, and operator interface techniques can be used to configure the system, enabling users to match bus technologies to application requirements.

Field Networks

Field networks reside at the device level (Figure 4.9a), below a PLC, DCS, or other host controller. They serve to replace point-to-point wiring of sensors, actuators, and other I/O devices with an open, multi-drop, bus-type wiring scheme. Sensors, actuators, and other devices formerly hardwired to the host or remote I/O racks now become nodes on a device-level network.

Fieldbuses can be of three categories, depending on the application requirements and device types. These three categories are at the bit, byte, and block levels and are sometimes referred to as sensor-bus, device-bus, and fieldbus, respectively.

Sensor-Bus Sensor-bus networks tend to be driven by the need to move a small number of bits very fast between devices and controllers. Many discrete devices need to transmit only one or perhaps a few bits of information. Examples are proximity switches, pushbuttons, and motor starters. A sensor-bus network can meet this need and transmit bit-level messages to the controller.

Often an I/O brick, such as AS-Interface, which allows several sensors or actuators to be attached to one node, is used in sensor-bus networks, thus distributing the already low interface cost across several physical devices. This also results in a lower node count relative to the number of physical devices. Given the small message size and limited diagnostics, reduced wiring costs are the primary benefit of using sensor-bus networks.

AS-Interface, or Actuator Sensor Interface, is a bit-level network that was introduced in late 1994. Along with Profibus-DP and WorldFIP-I/O, AS-Interface is one of the four technologies included in EN 50254 (High Efficiency Buses for Industrial Automation). Gateways are available for the connection of AS-Interface networks to Profibus-DP, DeviceNet, and ModBus.

Device-Bus Device-bus networks have message capacities ranging from several bytes to over 200 bytes. Device-bus networks are designed to meet the communication needs of higher-level devices. The device-bus networks are typically employed where more information is transferred or where diagnostics are required.

Profibus, or PROFIBUS, was a German national standard for digital communication at the fieldbus level, and is now a European standard as part 3 of EN50170. It is also one of the parts of IEC61508. DP, short for decentralized peripherals, was designed to optimize the network throughput to meet the networking requirements at the device-net level. Profibus-DP is an RS-485-based protocol. Applications range from material handling to control of paper and packaging machines.

Since its 1994 launch by Allen-Bradley, the DeviceNet technology has been turned over to the independent Open DeviceNet Vendor Association (ODVA). DeviceNet offers master/slave and peer-to-peer capabilities with devices from a number of vendors. The goal of DeviceNet is to provide a low-cost communications link eliminating expensive hard wired I/O. The communication profile for DeviceNet is shown in Figure 4.9c.

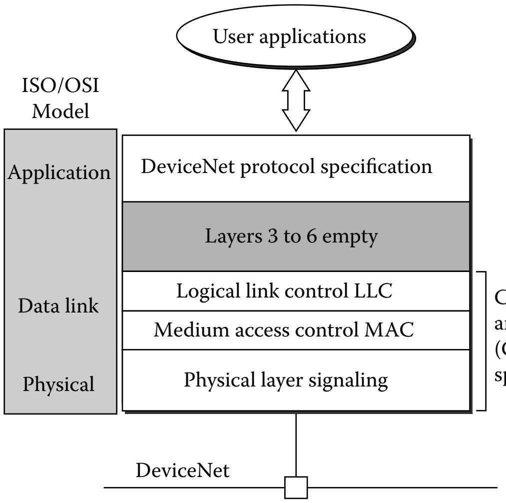


FIG. 4.9c
DeviceNet communication profile.[2]

Fieldbus Fieldbus networks are designed to provide highly reliable bidirectional communications between "smart" sensors and actuators and a control system in time-critical applications. Fieldbus messages contain the readings of several floating point process variables all sampled at the same time and their respective status.

One of the main objectives for fieldbus is to become the digital replacement for analog 4- to 20-mA transmission of process variables in the process industries. Examples of implemented low-speed fieldbus networks include Foundation Fieldbus H1 and Profibus-PA. Due to the differing requirements for process versus discrete installations, fieldbus networks typically have slower transmission rates than device-bus or sensor-bus networks. Another key differentiator is the addition of a user layer, or layer 8, on top of the typical three-layer communications stack. The user layer portion of Foundation Fieldbus includes standard and open function blocks that can be used to implement distributed field control systems.

Fieldbus Devices

Fieldbus devices used in manufacturing automation may in some cases support a limited number of parameters associated with basic measurement or actuator functions. Fieldbus devices designed for use in the process industries may support numerous parameters for diagnostics, measurement, calculation, and control functions. For either type of field device, the manufacturer must provide information that may be used in the DCS to automatically locate information within the field device.

To address this requirement, an Electronic Device Description (EDD) language has been defined by the IEC 61804 standard. This description method allows the DCS to use devices based on different technologies and platforms. The

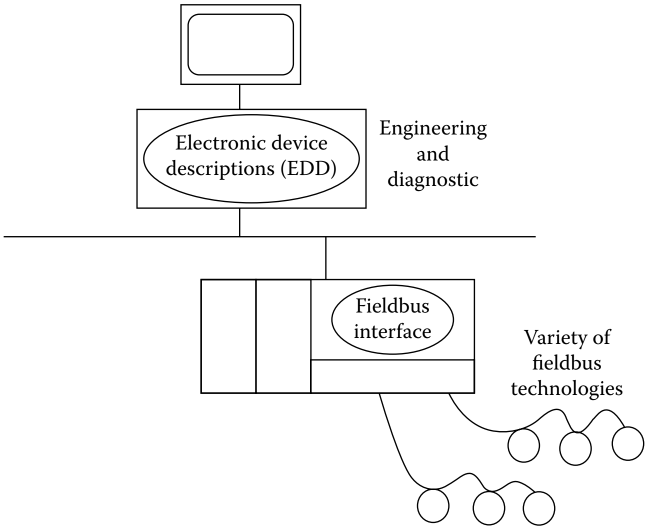


FIG. 4.9d
Native DCS integration of fieldbus technology.

IEC 61804 standard contains profile conformance statements for numerous technologies including PROFIBUS and Foundation Fieldbus. A modern DCS system that fully integrates fieldbus devices is illustrated in Figure 4.9d.

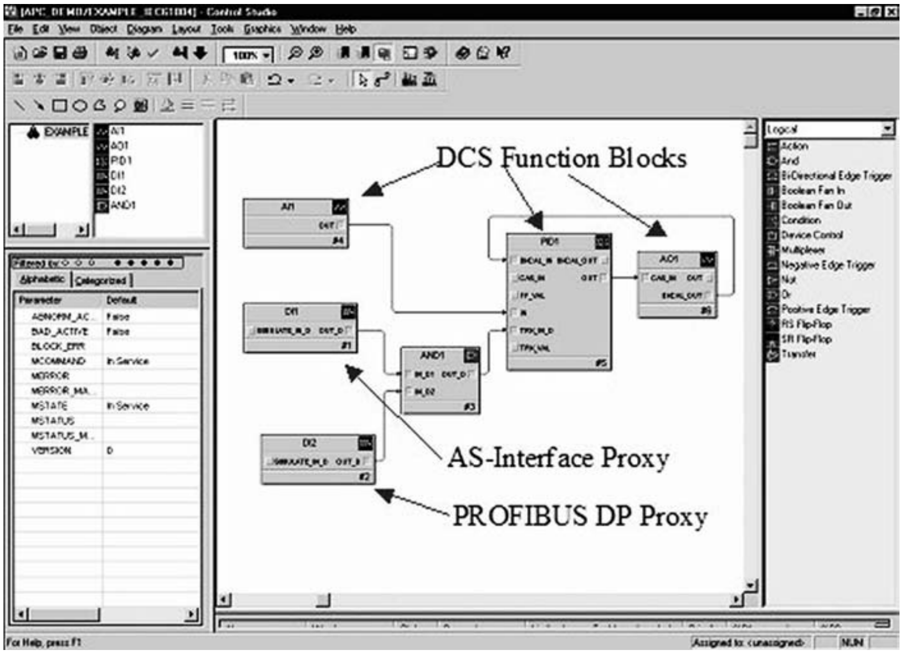
Function Blocks

A block of parameters for measurement, control, and calculation functions may be represented in process control systems as function blocks. The IEC61804 defined the abstract architecture and basic set of function blocks parameters that can be implemented, for example, in field devices or DCS controllers.

Applications within the DCS to system engineering, operation interface, and maintenance have to interact with the function blocks. Functions defined for a function block in the conceptual model are not necessarily mapped one-to-one to the parameters included in a field device. In such cases, these functions can be mapped to a proxy if a device does not support the same capability as included in the DCS.

The IEC 61804 specification may be used in a DCS to give a consistent function block interface for the different technologies. Fieldbus devices that do not support function blocks such as those based on AS-Interface, Profibus DP, and DeviceNet may be brought into the DCS using the concept of a proxy. Through a proxy, function block capability may be provided in the controller for the measurement or actuator value provided by a device.

To the user, these proxy function blocks may be configured for monitoring and calculation application in the same manner as the function blocks supported by the DCS. For these cases, the advanced signal processing and alarming that would normally be done by function blocks in the fieldbus device are done instead in the controller. Some differences in the dynamics associated with the signal processing exist, but in many cases they will have no impact on the application. An application example that utilizes proxy function blocks to provide a consistent interface to AS-Interface and Profibus DP devices is shown in Figure 4.9e.



Parameter	Default
ABNORM_AC...	False
BAD_ACTIVE	False
BLOCK_ERR	
MCOMMAND	In Service
MERROR	
MERROR_MA...	
MSTATE	In Service
MSTATUS	
MSTATUS_M...	
VERSION	0

FIG. 4.9e
Proxies are used to represent information in fieldbus devices.

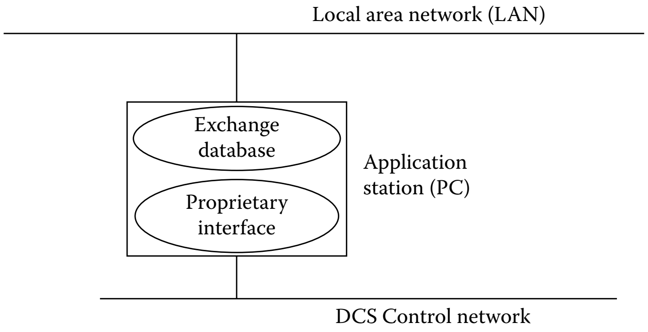


FIG. 4.9f
Traditional method of network integration.

NETWORK INTEGRATION

Some of the initial DCS systems installed in process control provided little or no support for the integration to plantwide networks. When such support was provided by the DCS manufacturer, the interface may have used a common, low-speed physical interface, e.g., IEEE485. However, the protocol associated with the interface in most cases was proprietary to the manufacturer. Information access was often based on the manner in which data were organized within the DCS. Thus, custom applications were required to access information in the DCS and to save this information in a database that could be accessed by other applications (Figure 4.9f).

The cost of implementing and maintaining proprietary interfaces for network access limited their use in early DCS applications. However, the common adaptation of Ethernet technology in the early 1990s for plantwide local area networks (LANs) helped to create a demand for better integration and exchange of information for production planning, remote diagnostics, and monitoring.

Modern DCS systems are designed to support plantwide access to real-time DCS information to support business planning and remote system monitoring. To provide an open interface, this capability is typically based on international and industry standards. In most cases, these will consist of support for:

- OPC to allow plant information to be accessed in a consistent manner
- Ethernet physical communication layer for high performance at low cost
- Internet connection support for remote access to plant information
- Wireless fidelity (WIFI) access within the plant to operation data
- Modern operating system, such as Microsoft XP and .Net framework, to provide extensible markup language (XML) data exchange and support for Terminal Server Capability

These technologies may be combined in a modern control system to address information access requirements. Also, manufacturers of older DCS architecture may offer increased capability by adding interfaces based on these technologies. For either case, the network integration is illustrated in Figure 4.9g.

Through the use of OPC and XML technology, a common means of information access is provided within the control system at remote locations. The capabilities of modern operating systems such as Microsoft XP allow the same interface to be utilized to display real-time information.

The advent of wireless communication standards, such as IEEE 802.11b, WIFI (which is defined by IEEE 802.116), and the use of inexpensive handheld devices based on this technology is being quickly adopted. This technology allows real-time data and configuration information within the control system to be accessed throughout the plant. In such an environment, strict implementation of security to authenticate user access to the control system is crucial to maintain system

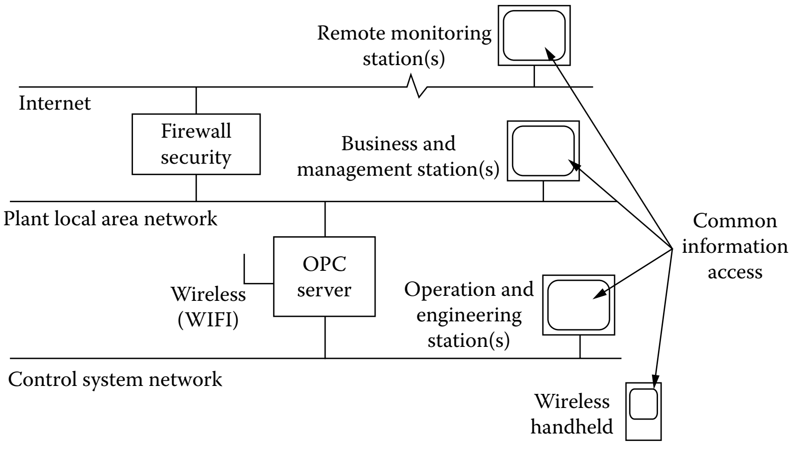


FIG. 4.9g
Present-day method of network integration.

security. Also, any provision that the plant makes for Internet access to the control system usually includes some kind of firewall, a secured connection, and in some cases, data encryption. Virtual private network (VPN) connections are often used to provide a high level of security.

Ethernet routers and switches may be incorporated into the communication system to limit information. Routers and switches are also used to ensure that when plant information is accessed via the LAN or remote applications, then it has no impact on control system communication and data updates at the operator interface.

Plant network interfaces are being used for a wide range of applications including material and resource planning, maintenance, scheduling, remote access, and many other applications. As plant networks become more prevalent, a fundamental shift is occurring in how manufacturers approach production automation.

Historically, the focus was on individual unit operations; the process unit definition and associated equipment defined the overall system architecture. The user definition for the process unit instrumentation determined the plant automation strategy. The expected benefits related primarily to improved unit repeatability and reductions in head count—goals narrowly focused on localized operational improvement.

Recent Integration Trends

In recent years, the needs of the manufacturing industry have changed. Minimizing the total time from conceptual development of a process through implementation and startup is necessary to ensure maximum return on research and development investment. The pressure of competition limits the supplier's ability to increase price. In conjunction with price constraints, increasing regulatory compliance, reporting requirements, and growing raw material costs are also squeezing profitability.

In response, automation users and suppliers have developed an expanded approach, called Manufacturing Execution Systems (MES), for production automation systems. The premise of MES is simple:

- Provide open information exchange across manufacturing production and business planning systems.
- Provide integrated real-time manufacturing applications.

MES has great potential. By integrating today's sophisticated enterprise resource planning systems with real-time production data, process plant managers can now more easily schedule production, manage raw materials, and optimize equipment use. Closing the "information gap" is key for achieving an agile manufacturing environment.

In practice, however, MES systems have been anything but simple. To date, MES solutions have required a significant investment in additional hardware, software, and application engineering. Essentially, users have been required to implement an additional system layer on top of existing systems in order to support the information routing and numerous interfaces required across functional areas. As may be expected, initial user adoption of MES has been slow due to the added cost, complexity, and inflexibility of this approach.

Another area where plant networks are improving is in the area of maintenance. The performance of the processing industries improves by the integration and automation of operation and maintenance functions. A key element in this integration is the quality and availability of maintenance and process data. It is for this reason that most suppliers are integrating their computerized maintenance management systems (CMMS) with their process control products.

Foxboro, for example, integrated its MAXIMO and Invensys' Avantis with its I/A DCS. Similarly, Honeywell integrated its EHM into both MAXIMO and SAP, Emerson's AMS is interfaced with MAXIMO, Wonderware's InTouch was integrated with Avantis, etc.

In the various maintenance packages, alerts are initiated by vibration monitoring (CSi and Bentley Navada), corrosion detection, equipment performance monitoring, production monitoring, and many types of detectors and are routed back into the DCS systems using Internet-based services such as Web services, XML, and SOAP.

Modern DCS systems treat these alerts just like other alarms. They can be prioritized, acknowledged, used to drive animations in process graphics, and combined with control strategies. More advanced features such as combining their alerts with process conditions are being created and presented as "Smart Tags" or "Smart Process Objects."

Plant networks will continue to play an important role picking up features from Web service architectures as they evolve. Standards bodies such as OPC recognize this and are busy developing the next level of interfaces.

CONCLUSIONS

Modern process control systems support a variety of field devices that are based on different fieldbus technologies. These systems use Electronic Device Descriptions and proxies to provide the user with a common interface to all field devices. Information within the control system can be accessed in a common fashion for operations and engineering. Also, control system information may be accessed through local area networks and the Internet. For any network connection to the Internet, protection should be provided by the DCS manufacturer and the plant network to ensure the security of such access.

References

1. Vitturi, S., *Instrument Engineers' Handbook,* Vol. 3, Section 4.5, Boca Raton, FL: CRC Press, 2002.
2. Vitturi, S., *Instrument Engineers' Handbook,* Vol. 3, Section 4.6, Boca Raton, FL: CRC Press, 2002.

Bibliography

Diedrich, C., Russo, F., Winkel, L., and Blevins, T., "Function Block Applications in Control Systems Based on IEC 61804," ISA Conference, September 10–13, Houston, TX, 2001.

Lewis, J., "Control Systems Open Up," *Design News,* Vol. 54(14), p. 61, 1999.

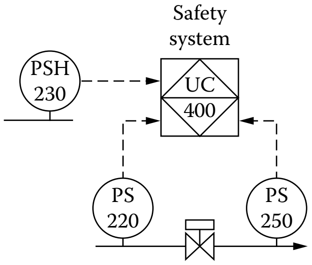
Lian, F.-L., Moyne, J. R., and Tilbury, D. M., "Performance Evaluation of Control Networks: Ethernet, ControlNet, and DeviceNet," *IEEE Control Systems Magazine,* Vol. 21(1), pp. 66–83, 2001.

Manolescue, P., "How XML Is the Key Technology to Integration of Building Systems via the IT Network," *Premises and Facilities Management,* Vol. 31(5), pp. 441–445, 2002.

Xiaojian, L., Mengjun, Y., Zhenren, Z., and Xiaonian, H., "Design and Realization of a Control System for Petroleum Product Distillation Based on Controller Area Network Bus," *Petroleum Processing and Petrochemicals*, Vol. 34(2), pp. 62–65, 2003.

4.10 DCS: Integration with Other Systems

T. L. BLEVINS, M. NIXON (2005)



Flow sheet symbol

Types of DCS System Interface:	A. OPC server B. Serial interface C. DCS gateway
Partial List of Suppliers:	ABB (A, B, C) (www.abb.com) Allen-Bradley (A, B, C) (www.ab.com) Emerson (A, B) (www.EasyDeltaV.com) Honeywell (A, B, C) (www.honeywell.com) Invensys (A, B, C) (www.invensys.com) Schneider (A, B, C) (www.schneider.com) Siemens (A, B, C) (www.siemens.com) Yokogawa (A, B, C) (www.yokogawa.com/us)

INTRODUCTION

The reinstrumentation of an existing process or the construction of a new plant often involves interfacing to intelligent external devices or subsystems. For example, the safety system that protects a critical piece of equipment may be an integral part of and is provided by the equipment manufacturer.

Devices as a motor starter may include embedded logic, diagnostics, and support for digital communications to access this added information. When an existing plant is being expanded, the process control system selected for the expansion may not be the same as that used in the existing plant. To provide safe and efficient plant operation, it is important to fully integrate these external devices and subsystems into the new control system.

Integration provides the plant operator with a single window interface into all functions of the complete process. Related process information may be included in the operator displays. This way, the operator is provided with consistent presentation of process alarming and trending and consistent means of changing set points or modes of operation. This makes it possible to provide a single login and span of control.

EXISTING SYSTEMS

When a new plant area is added or expanded, the operators often need to know the conditions in the existing plant section in order to maintain a coordinated operation. Similarly, the operators of the existing plant will also need to have feedback from the new process area to enable them to make decisions on how best to run the balance of the plant. In most cases, only a small fraction of the information in either system must be communicated to support such coordination between process areas or plant sections.

Three techniques that have been successfully used to support the exchange of such information serving plant integration (Figure 4.10a) are:

1. OPC interface
2. Serial interface to highway gateway
3. Serial interface to serial port

A fourth technique is to hardwire the measurement signals to more than one system. This method is not illustrated here, although it is often used when only a small number of signals are involved in the interface.

MODBUS Interface

MODBUS, a specialized network for use with other open networks, was developed for use on the Modicon Company proprietary network. MODBUS can be implemented over any transmission medium but is most often seen on RS-232 and less often on RS-422 or RS-485. It is a serial transmission technique that uses master/slave arbitration and is discussed in detail in Volume 3 of this handbook. MODBUS can transfer 300 registers per second at 9.6 kB and can communicate

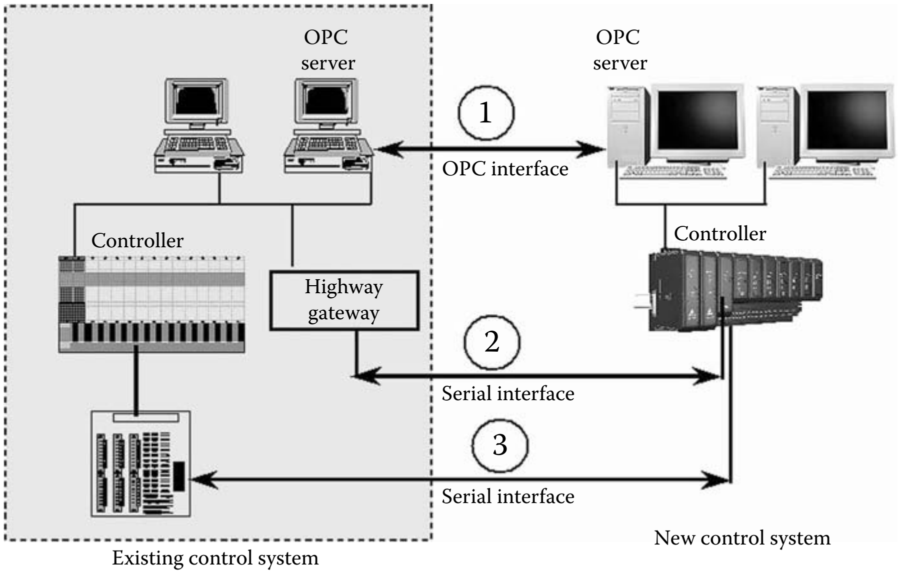


FIG. 4.10a
Three methods of integration of an existing DCS system and a new control system can be obtained by the use of three types of interfacing: 1) OPC interface, 2) serial interface to highway gateway, or 3) serial interface to serial port.

in half-duplex style with up to 247 slaves. MODBUS Plus is an RS-485 based, peer-to-peer network protocol that transmits at a rate of 1 mB.

Many new and existing DCS systems include a hardware interface that utilizes a common communication protocol such as MODBUS block transfer of information from other devices. In some cases, this serial interface may be redundant to improve system availability. If such interface exists, it may be the least expensive method of exchanging information between the systems.

MODBUS is a popular method for moving data values between systems, but if the updating frequency cannot be longer than 1 second, it can handle only a few hundred signals. It is an application layer messaging protocol, positioned at level 7 of the open system interconnection (OSI) model, which provides client/server communication between devices connected on different types of buses or networks. MODBUS support is provided across both serial and transmission protocol/Internet protocol (TCP/IP) networks, as illustrated in Figure 4.10b.

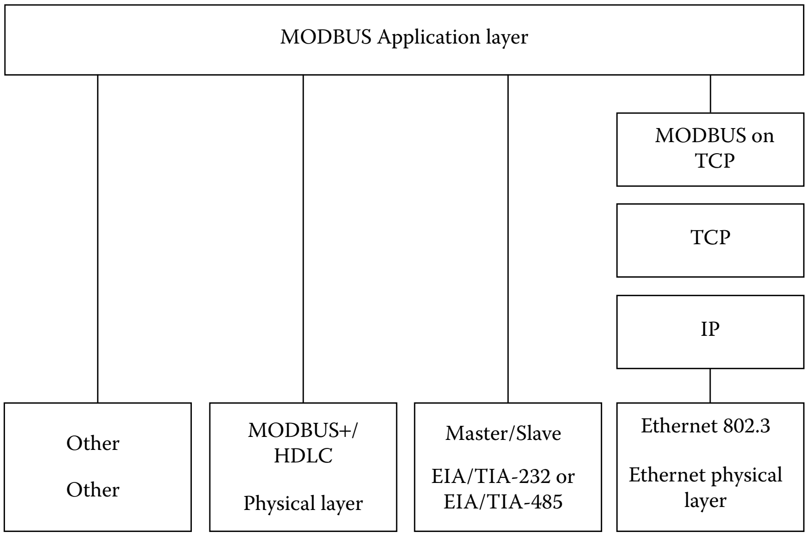


FIG. 4.10b
Integration of existing and new DCS systems using MODBUS interfacing.

Support for MODBUS continues to be strong. Most vendors have built-in support for master/slave stacks over serial and/or TCP/IP. The Internet community can access MODBUS at a reserved system port 502 on the TCP/IP stack.

Some manufacturers may also provide a gateway device that can be used to interface to external systems. These devices are often provided to connect different revisions of equipment from the same manufacturer. Thus, such interfaces may rely on a proprietary communication protocol and limit the types of information that may be accessed through this interface.

OPC Interface

OPC (OLE for Process Control, where OLE stands for Object Link Embedding) is based on Microsoft Windows COM (Component Object Model) architecture. OPC was developed by the OPC Foundation and provides standard software interface to hardware databases.[1] The three basic types of servers it provides are for 1) data access, 2) alarm and events access and 3) historical data access, as shown in Figure 4.10c.

Most modern process control systems support information exchange based on the OPC Foundation's Data Access Specification Version 3.0, March 2003. This process industry standard was created through a collaborative effort of leading automation suppliers and Microsoft. The specification defines a standard set of objects, interfaces, and methods that facilitate data interoperability. Many manufacturers offer OPC servers that may be added to older DCS systems. Where such capability exists, it is to exchange a large number of readings between control systems.

On the other hand, the OPC approach has several disadvantages. One is that the addition of the OPC servers may be more costly than using a serial interface (Figure 4.10a). These interfaces tend to be time consuming to configure—often requiring duplicate configuration to map parameters from one system to another. The interfaces also tend to lack basic features such as holding last value on configuration download, upload, and integrated parameter security. Also, such an interface often is not redundant and thus should be used only where temporary loss of communication will not impair plant operations.

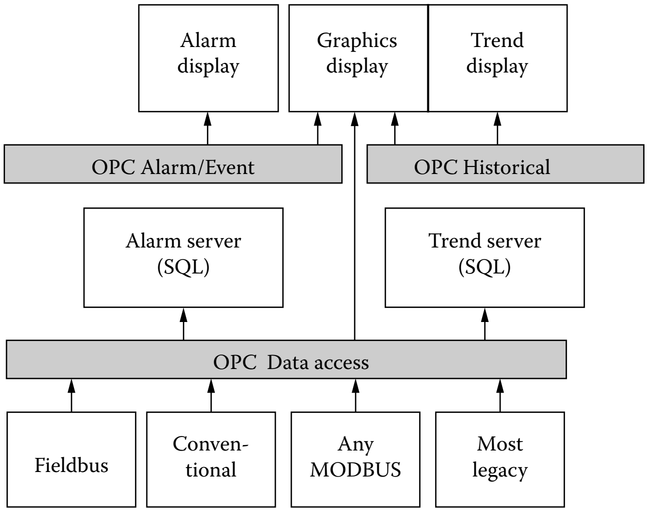


FIG. 4.10c
The three types of OPC servers are: 1) DA, data access, 2) A&E, alarm and events, and 3) HAD, historical data access.[1]

The latest OPC standardization efforts are making important improvements to the data model, in integrating data sources such as alarms, and in supporting a much more capable Web Service Style interface.

One of the biggest challenges in DCS system integration is the fact that each control implementation may represent control parameters in a different manner. Take for example the representation of the operating mode of a controller. Some control systems implement the control mode as two separate parameters, i.e., local/remote and auto/manual, while other systems represent the control mode as a single parameter, i.e., manual/auto/cascade/remote-cascade/remote-out.

Where such differences exist, the basic calculation features of the "old" controller can be used to create a proxy in the "new" controller to map these parameters to their equivalent representation in the new control system.

Looking ahead, several developments are underway that promise to make the integration easier. Perhaps the most significant of these is again OPC. The OPC standardization efforts, recognizing the importance of extensible markup language (XML), software componentization, and Web services, are moving forward with new standards that combine existing data models on top of new service frameworks.

For example, real-time, alarms and events, and history models are being combined. Once combined, a single client will be able to browse to locate the data of interest, subscribe, and then retrieve real-time data, alarms, and events, as well as history. These newer models also address problems related to consistent naming and consistent use of the status/quality attribute.

Consistent/standardized naming would allow clients to reconcile the naming differences between systems — the same OPC path would always refer to the same data. The standardization would also define how status is assigned. OPC servers would be expected to use a consistent status when communicating and receiving data from devices. Uniformity in the assignment of status by OPC servers would improve the consistency of clients.

Other standardization efforts, including the Fieldbus High-Speed Ethernet and PROFIBUS PA, are extending the reach of device networks. These standards provide capabilities to integrate devices across much higher-speed networks.

MOTOR CONTROLS

Analog measurements and actuator signals represent only a fraction of the total inputs and outputs of process control systems. The remaining inputs and outputs are discrete. They

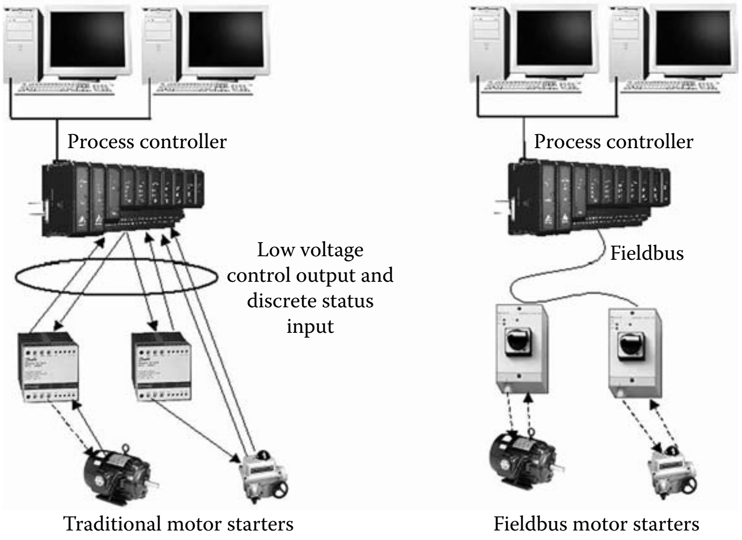


FIG. 4.10d
Traditional and fieldbus-based interfacing of motor starters.

often are associated with motor control, position sensing, and the operation of on/off valves.

The motor control interface for a DCS system often consists of low-voltage discrete outputs from a controller wired to motor starters. Controller feedback of the motor start relay status (M contact) and associated conditions are provided through discrete inputs. The installation cost of this approach may be quite high because of the requirement for the installation and checkout of the large number of wires. Also, diagnostic information that may be available locally at the motor starter may not be remotely accessible, or may be too costly to access as discrete inputs.

Significant advances have been made in motor starter designs that allow more compact packaging. New fault diagnostic capability is also often available for the detection of power loss, sensor input circuit short and phase imbalance, phase currents, full-load current, and thermal capacity for easy troubleshooting and quick restarting.

Also, most manufacturers provide integrated add-on communication interfaces for the motor starter and local input hand-off-auto selectors and pushbuttons. For many applications, this can reduce the quantity of wiring plus simplify connections and thus can reduce installation and maintenance costs, as illustrated in Figure 4.10d.

Fieldbus Interface

The fieldbus communication interface with motor control equipment is often specific to the manufacturer. Some of the most popular intelligent motor starters offer either Profibus DP, AS-Interface, or DeviceNet communication. Thus, modern process control systems are often designed to support a variety of communication interfaces.

The logic associated with motor start, stop, interlocks, and permissive constraints is typically implemented in the process controller. Such logic may have been designed and implemented using ladder logic. However, it is possible to standardize motor control logic as a function block. One example of this is the Device Control block as defined by the Fieldbus Foundation Function Block Application Process Specification, Part 3. A variety of applications including single- and multiple-speed motors with permissives and interlocks may be addressed using this the Device Control block.

Integrating devices such as motor starters with control systems has become considerably easier with standards such as Fieldbus and Profibus. More recent standards such as FDT/DTM (see below) are further simplifying the integration of device configuration and operation by defining a common parameterization format. Together, the Field Device Tool (FDT) and Device Type Manager (DTM) provide a configuration front-end and driver-like back-end, similar to printer drivers today, to support the configuration and communication interfaces to devices.

SAFETY SYSTEMS

Safety instrumented systems (SIS) is responsible for taking the process to a safe state when predetermined conditions are violated. The safety system includes the hardware, software, and all field equipment necessary to perform the desired operations to reach a safe state. Because of the critical role these systems play in plant operation, their design and implementation are normally based on accepted standards, such as IEC 61508, IEC 61511, and ANSI/ISA 84.01. In such designs, the process control system is typically separated from the

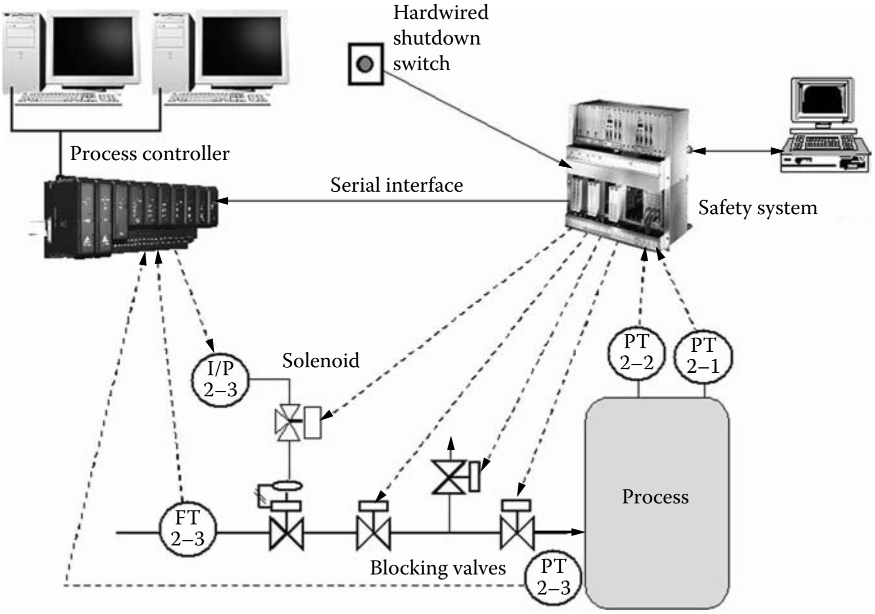


FIG. 4.10e
The integration of the safety instrumented system (SIS) into the overall process control system.

SIS (Figure 4.10e) to avoid the need for all process control systems to meet SIS requirements.

As it is illustrated in this example, the process input signals to the safety system are separate and independent from those used by the process control system. The outputs of the safety system are also designed to act independently from the process control system. For example, a solenoid valve may be placed into the air tube between the I/P converter and the control valve actuator.

In this configuration, when the SIS system requires that the solenoid be energized, it will cause the control valve to take up a safe position, i.e., open or closed. Also, the safety system output may activate blocking on/off valves that stop or divert process streams from their normal flow-paths, regardless of the signals received from the process control system.

As the safety system overrides the normal operation of the throttling control valves of the process control system and prevents the controllers from returning their controlled variables to set point, the integral action is likely to saturate the controller's output. The consequence of this is that when the SIS system returns to normal, the control loop output is still saturated and is not ready to take control.

To avoid any upsets this could cause, when the safety system returns to normal, it also provides a status input to the process control system. If the process control system has maintained or memorized the controller output that existed before the SIS episode, that output value can be automatically set to provide smooth recovery.

The safety system status may be simply provided to the process control system as a discrete input. However, a serial link is typically used to provide status and other information concerning the safety system. Through this link, such information as first-out indication of the cause of action, status of blocking valves, flame sensor indication, etc., may be communicated to the process control system and included in the operator interface.

Operator requests such as "initiate shutdown" or "reset the safety system" may also be added to the operator interface and communicated to the safety system. However, even if this capability is provided, an independent hardwired interface is normally required for the operator to manually initiate shutdown.

To reduce the effort required to integrate information from a safety system into the process control, some process control system manufacturers offer special controllers designed for the execution of safety system logic. Alternatively, independent hardware components may be added to a standard controller for the execution of the SIS logic as illustrated in Figure 4.10f.

When a manufacturer provides SIS controller support as part of the process control system, information and alarms from the SIS controller are available for use in the operator interface without the need to map such information. Also, this approach may allow the same tools that were used to engineer the control system to also be used for the configuration and checkout of the SIS system. Using a common set of tools allows the user to easily integrate parameter access, alarming, history collection, span of control, diagnostics, and security. Advanced features such as configuration audit trail and authorization can also be provided.

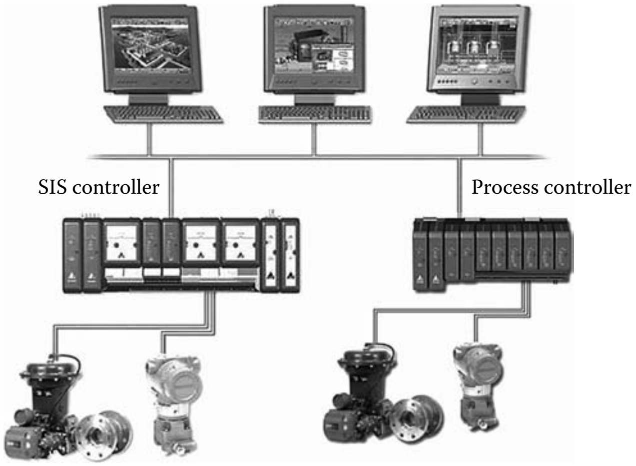


FIG. 4.10f
The system configuration if SIS controller support is available in the process control system.

CONCLUSIONS

Modern process control systems support a variety of communication interfaces. These may be used to integrate information from intelligent external devices, subsystems, and existing DCS systems into the total process control system of the plant. Where the products of several manufacturers are used, it may be necessary to implement proxies in the controller, which can be used to map the information from these "other" suppliers to its equivalent representation in the total control system of the plant.

Reference

1. Berge, J., *Instrument Engineers' Handbook,* Vol. 3, Section 5.4, Joint publication of CRC Press and ISA, 2002.

Bibliography

Fan, X.-G., Wang, Z., Sun, Y.-X., and Wang, L., "OPC Based Ethernet Control and Management," *Information and Control,* Vol. 32(2), pp. 113–117, 2003.

Hoon, P. S., and Huan, Y. R., "FOUNDATION Fieldbus Systems Integration Using OPC," *Technical Papers—ISA,* Vol. 397, pp. 143–152, 1999.

Janke, M., "OPC—Plug and Play Integration To Legacy Systems," IEEE Conference Record Annual Pulp and Paper Industry Technical Conference, pp. 68–72, 2000.

Kleines, H., Zwoll, K., and Sarkadi, J., "Integration of Industrial Automation Equipment in Experiment Control Systems via PROFIBUS—Developments and Experiences at Forschungszentrum Julich," *IEEE Transactions on Nuclear Science Ns,* Vol. 47(2p1), p. 229, 2000.

Luth, J., "OPC Brings XML-DA to the Factory Floor," *Control Solutions*, Vol. 75(7), pp. 34–36, 2002.

Macias-Hernandez, J. J., and Feliu, J. A., "Dynamic Simulation and OPC," *Hydrocarbon Engineering,* Vol. 8(9), pp. 53–58, 2003.

OPC DA 3.00 Specification, "Data Access Automation Interface Standard," www.opcfoundation.com, 2003.

PROFIBUS Organization, "Specification for PROFIBUS Device Description and Device Integration Volume 3: FDT V1.1," Karlsruhe, Germany, November 2000.

Spear, M., "Bridging the Data Gap," *Process Engineering,* Vol. 81(3), p. 25, 2000.

Wang, X. Z., Garcia-Flores, R., Hua, B., and Lu, M. L., "Advanced Information Systems for Process Control and Supply Chain Integration," *Computer Aided Chemical Engineering*, Vol. 15(A), pp. 636–641, 2003.

4.11 DCS: Management of Abnormal Conditions

B. A. FITZPATRICK (2005)

Professional Organization:	Abnormal Situation Management (ASM) Consortium® (www.asmconsortium. com)
Consulting Services:	Brad Adams Walker Architecture (www.bawarch.com) Human Centered Solutions (www.applyhcs.com) Mustang Engineering (www.mustangeng.com) TTS Performance Systems (www.myplantstraining.com) User Centered Design Services (www.mycontrolroom.com)
Alarm Management Software and Its Suppliers:	
Asset Integrity Management:	SILAlarm (www.assetintegrity.co.uk)
Control Arts:	Process Alarm Toolkit and Alarm History Analysis (www.controlartsinc.com)
Gensym:	G2-Expert Systems (www.gensym.com)
Honeywell:	Alarm and Events Analysis, Alarm Configuration Manager (ACM), and Alarm Scout (www.honeywell.com)
ICS Online:	IMAC (www.ics-ltd.co.uk)
Matrikon:	Process Guard (www.matrikon.com)
Nexus Engineering:	Real-Time Operations Excellence (rtOpx™) (www.nexusengineering.com)
PAS:	Plant State Suite (www.pas.com)
ProSys:	Special Alarm Management (SAM) (www.prosysinc.com)
TIPS:	Logmate/AMS (www.tipsweb.com)
Yokogawa:	Advanced Alarm Administrator (AAA Suite), Event Analysis Package (Expalog), and Operation Efficiency Improvement Package (Exapilot) (www.yokogawa.com/us)

INTRODUCTION

From an industrial process control standpoint, an abnormal condition develops when the process deviates significantly from the “normal” and acceptable operating range. In pragmatic terms, the “abnormal” that one must manage is a deviation that will lead to financial losses. The challenge of abnormal condition management is to provide the operating team with tools that will enable them to avoid or minimize the impact of abnormal conditions. This is not an insignificant undertaking, but it is one that stabilizes operations and has dramatic financial rewards.

This section will provide a conceptual overview of abnormal condition management and will highlight key considerations in a formal abnormal condition management program focusing on design elements for control room layout, operator training systems, alarm systems, and the graphical operator’s interface to the control system.

ABNORMAL CONDITION MANAGEMENT

To manage anything, one must first understand its basic nature. From an industrial process control standpoint, an abnormal condition can be defined as the condition that develops when the process deviates significantly from the “normal” acceptable mode of operation. In pragmatic terms, the “abnormal” that one must manage is a deviation that will lead to unsafe

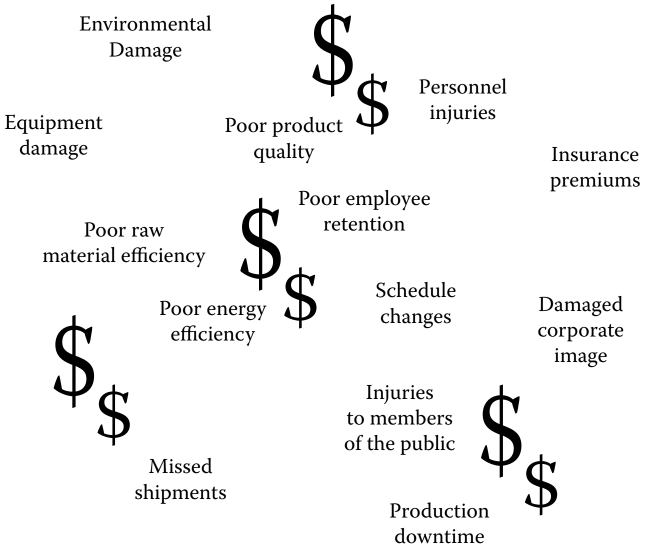


FIG. 4.11a
Types of losses that can be caused by abnormal conditions.

conditions or losses. Figure 4.11a shows a sampling of losses that can be caused by abnormal conditions, including losses in efficiency, quality, production, or equipment or environmental damage, personnel injuries, and other softer issues.

The preventable losses to the U.S. economy are estimated at several billion dollars annually. Recognition of this led to the formation of the Abnormal Situation Management (ASM) Consortium^{TM}, a group managed by Honeywell and composed of several manufacturing companies, universities, and other organizations performing research and development focused on developing technologies to aid in abnormal management. The consortium has published several guidelines and documents for use by member companies. Additional information can be found at www.asmconsortium.com.

Types of Control

In most industrial processes, the operator no longer initiates the majority of the steps of the operation. The days of (local- and panel-mounted) single-loop controllers whose performance was orchestrated by human operators is largely gone. Distributed control systems (DCSs) are generally financially attractive to implement and have become quite common.

Control schemes have become quite robust and generally require little operator intervention. The trend in staff downsizing has resulted in ever-fewer operators managing an increasing number of loops. In the days of the panel-mounted single-loop controller, the entire span of control for an operator was generally mounted within his or her sight. Alarms and annunciators were mounted on the control panel. The action required to intervene was generally visually apparent and in close proximity.

With the advent of the DCS and optimizing control systems, the interactions between control loops became less apparent in the new operating graphics. Today, a single line of text in the alarm summary can be all the information that is provided to the operator. The degree to which the operator understands the process provides the context that leads to the decisions about intervention. In case of a significant upset it is common that the operator will be overwhelmed or "flooded" with hundreds or even thousands of alarms, signaling a variety of abnormal conditions. This sudden flood of information is hard to handle.

The newer types of controllers tend to also diminish the likelihood that the operator will understand all the details of the control algorithm and its interactions. As control systems are becoming more complicated and more robust, an unfortunate result is that abnormal conditions that occur are increasingly difficult for the operator to recognize and respond to in a manual mode.

Figure 4.11b shows a pyramid of a typical control system. Moving up the pyramid, profitability and on-stream reliability increases, but the complexity of the control strategy rises and therefore the role and the understanding of the operator can diminish. When the control strategy is very complex, no matter how well the operators are trained, they cannot fully understand all the possible interactions and consequences. One of the challenges of abnormal conditions management is to provide tools to the operations team to bridge this gap in understanding.

Need for Operator Intervention

With all of the capabilities of current technology, one might question why operator intervention is needed at all. The reason is Murphy's Law—anything that can happen, will. For this reason, human operators are required to reason and adapt their knowledge of the process and develop a response plan to any situation. While adaptive controls and optimization technologies are emerging and are useful in obtaining ever-higher levels of automation, it is still probable that human operators will be needed when systems fail.

An analysis of the root causes of abnormal conditions and incidents suggests that in some 40% of the cases, operators have played a role in causing the accidents. Figure 4.11c recaps the general findings of varied studies[1] performed to determine the root causes of incidents that have prompted significant losses.

PSYCHOLOGICAL BASIS FOR INTERVENTION

To work out how to best enable the operating team in the management of abnormal conditions, it is useful to understand the basic steps of cognitive processing. There are numerous models for the psychology of operations, notably that of Swain and Guttman, who suggest a three-step model of Orienting, Evaluating, and Acting. These steps are key elements in the Human Intervention Framework developed by the Abnormal Situation Management Consortium. Figure 4.11d

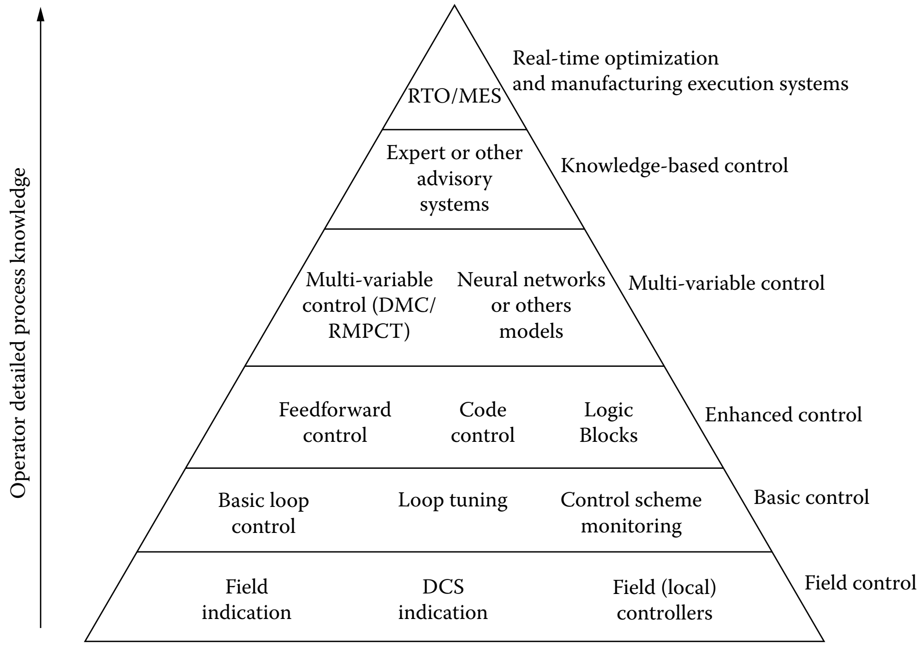


FIG. 4.11b
An approximate pyramid of the elements of an advanced control system.

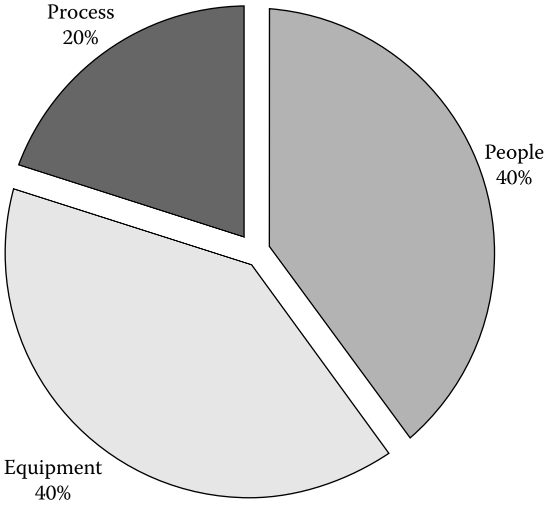


FIG. 4.11c
Root causes of industrial upsets and accidents.

proposes a simple sequential model of Detect, Sort/Select, Plan/Act, and Monitor.

Detect Phase

Tools for the Detect Phase of operator intervention should be focused on ensuring that the operator is aware of conditions in his or her span of control. There are several elements that are critical to successfully managing the operator's awareness of process conditions, including the design of the control room environment, the alarm system, and the graphical user interface (GUI).

It is very important to remember that operators have human physical and mental limits that can be exceeded in abnormal conditions. Thus, it is important to factor these limits into the design of critical systems. Also, advanced detection tools exist that can aid the operator by helping to detect abnormal conditions early and give the operations team additional time to avert an upset or loss.

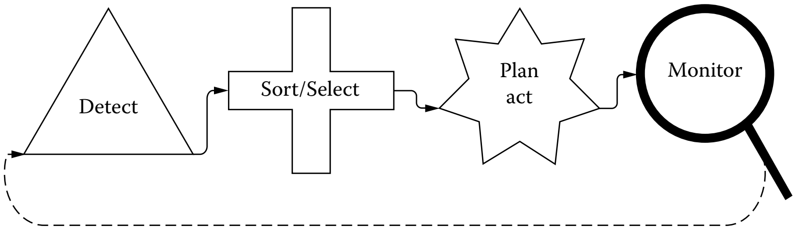


FIG. 4.11d
Sequential model for operator intervention in case of an upset.

Sort/Select and Monitor Phases

Tools for the Sort/Select and Monitor phases of operator intervention should be focused on ensuring that the operator will draw the correct conclusions regarding the state of the process. Key elements include Operator Training Systems, Advanced GUI Design, and Advanced Alarm System handling. Fundamentally, any clues regarding the root causes of upsets and any prevailing interactions should be presented and managed in these systems. One other important consideration is Additional Operator Information Needs. Systems should be assessed to ensure that any information relevant to diagnosing abnormal conditions should be presented to the operator if feasible.

Plan/Act Phase

Tools for the Plan/Act phase of operator intervention should focus on ensuring that whatever action taken is the correct one. Key elements here include:

- Graphics showing the entire span of control
- Contextual information displayed on key graphics (including advice on response)
- Appropriate update rates that show the impact of operator moves

Several other factors can impact the ability of people to perform effectively in a plant environment. These include personal factors such as knowledge, skills, motivation, and personality, and group factors such as the working environment, organizational structure, and work hours. Psychological and work process factors like availability of procedures, communications within the organization, work methods, control display relationships, and task criticality also have a profound effect on an operator's performance.[1]

Response Time

The time available for operators to respond is a primary requirement. Industrial processes are dynamic, and the pace of change during an abnormal condition, particularly once it has been detected, can be quite rapid. Varied studies have repeatedly suggested that the less the time available to respond, the more likely that the correct response will not be found.[2]

TABLE 4.11e
Probability of Failure to Respond[2]

Time Available (Minutes)	*Probability of Failure*
1	~1
10	0.5
20	0.1
30	0.01
60	0.001

In the stress of the moment, it is likely that the operator will not process all available information, but will act when a theory of the state of the process makes sense. Operators generally look for patterns in the behavior of the process. Once a pattern is recognized, then it becomes "reality" until an accumulation of data might prove it otherwise.

In fact, as soon as a pattern is selected, the operator focus will shift from understanding the upset to monitoring intervention activities. The focus will not shift back to a search for general system understanding until it becomes clear that the intervention is not improving the situation or *several* new alarms do not match the pattern selected. Thus it becomes critical to provide support in identifying the root cause of a given condition.

Planning the Intervention

Once the problem has been identified, the key issue becomes one of planning and executing the appropriate intervention. Any known responses should be designed into the automatic control system. If the response cannot be automated, then the operator should be instructed to make the required changes. Any advice that can be concisely presented may speed up the intervention process. If there is no pattern that can be diagnosed automatically, then the interface to the control system should be designed so that the process state and critical interactions are visually apparent.

It is important to remember that the abnormal conditions will generally be complex and will result from a series of consequences rather than one single event. Belke[3] performed a review of chemical industry accidents looking for common causes and found that major disasters are often preceded by a series of smaller accidents, near-misses, and other known precursors.

Plants are routinely run "blind" with failed instruments. And, at times, "without a safety net" because shutdown devices are bypassed. The decision to continue operating under these conditions is a critical one because a new event or failure combined with the existing failures is often simply too much for the operators to overcome in the time available to intervene.

Types of Operations

Finally, it is also important to note that the type and quantity of operator interventions needed are functions of the type of operation. Figure 4.11f shows a proposed model for the types or modes of operation. The first type of operation is under Optimal Operations, where the process is operating at peak efficiency and quality.

Optimal Operations is considered to be a subset of Normal Operations. Normal Operations can have some efficiency and quality losses, recognizing a range of operations that is considered acceptable. As the magnitude of losses increases, the type of operation becomes Upset Operations. This mode of operation might incur some minor equipment damage.

As the problem escalates, the mode of operation can shift into Emergency Operations. Emergency Operations has more

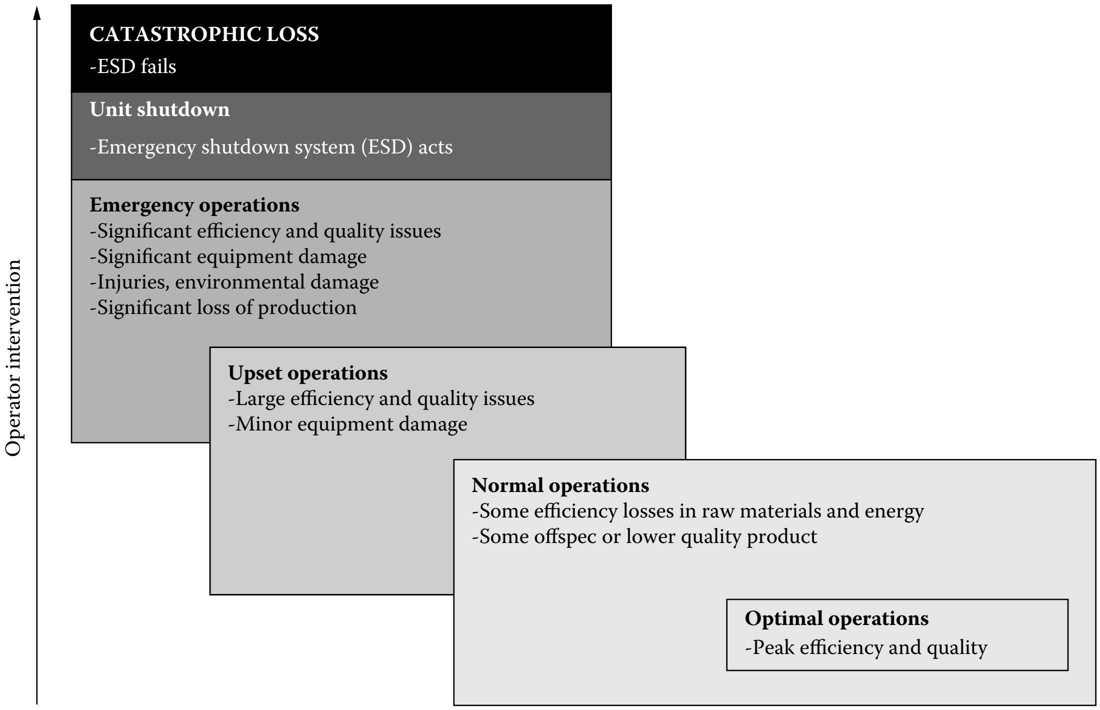


FIG. 4.11f
Model for defining the types of prevailing states of operation.

significant losses in efficiency, quality, and equipment damage. Also possible are injuries, environmental damage, and loss of production. The model also includes an escalation into Unit Shutdown and, with a failure of the Emergency Shutdown Systems, an escalation into a Catastrophic Loss.

As the type of operation changes, the operator also has different primary active goals. While safety is generally the primary goal of all types of operations, it is not the most active goal during, for example, Optimal Operations. During Normal Operations, operators can, in theory and in fact, undertake a variety of tasks that are not directly related to "operating" the plant.

It is common for operators to undertake personal training or to support projects during these relatively quiet times. However, as an abnormal condition develops, the operator becomes fully engaged in operating. The pace and magnitude of intervention generally accelerate. Table 4.11g summarizes the most common active goals and types of operator intervention during the different types of operations.

So, as the time available to respond drops and the likelihood of successful intervention diminishes, the need for operator intervention increases, as does the magnitude of the likely losses if the intervention is not successful.

MANAGING ABNORMAL CONDITIONS

There are certainly as many approaches to managing abnormal conditions as there are causes for the upsets themselves. Perhaps the most prominent and effective lines of defense

TABLE 4.11g
Overview of Operator Intervention and Goals

Type of Operation	*Operation's Active Goals*	*Major Type of Intervention*
Optimal	Exceed targets for production, efficiency and quality	Process monitoring (minor SP and OP changes)
Normal	Move to Optimal Operations	Process optimization (small SP and OP changes)
Upset	Return to Normal Operations	Process stability (large SP and OP changes)
Emergency	Ensure safety, stabilize unit (if possible)	Process safety (major SP and OP changes)
Shutdown	Secure unit, prepare for restart (if possible)	Process safety (major SP and OP changes); interact with ESD system
Catastrophic Loss	Secure unit for safety, minimize losses (if possible)	Emergency containment, field isolation

SP: set point; OP: output.

are in the control schemes and in the design of the emergency shutdown systems. Care should be and is taken in the design of these systems. However, the following ancillary systems can also be useful in the management of abnormal conditions and care should be taken in their design:

- Control rooms
- Operator training programs
- Alarm systems
- Graphical user interfaces (GUI)
- Advanced condition detection and advice systems

Control Room Design

Early control rooms were not much more than glorified umbrellas designed to do little more than provide shelter for equipment and operators. The early equipment was reasonably impervious to the elements, so the designs were not complicated and operators got little extra consideration. As the control equipment became more advanced and environmental requirements were added, the control rooms generally became actual buildings requiring architectural design support.

Over time, safety considerations have became more prominent, so it is relatively common today to have blast-proof buildings with advanced environmental controls. In fact, a specialty within the field of architecture eventually emerged, focused specifically on control center design. Additionally, reliance on video display terminals for process control system interactions with operators makes the ergonomics of workstation and console design critical. A detailed discussion of the design and upgrading of control centers is provided in Sections 4.1 and 4.2 in this chapter.

It is important to remember, however, the needs of the operating team during abnormal conditions and to factor those needs into the final control center design. The control room is the communications center for operations. Varied types of people need to interact with the workstation operators. But these groups of people also need to interact among themselves, as is illustrated in Figure 4.11h.

Thus, it is important to provide meeting facilities and entryways in such a way that they do not disturb the workstation operators. In abnormal conditions, the traditional control center design with its direct entry into the console area and its limited space often hinders operator performance.

The following key design specifications have been proven effective in designing for the management of abnormal conditions:[4]

1. Control center arrangements (focused on different types of rooms)
2. Control room layout (including usable space, furniture, maintenance access, storage, entrance and exits)
3. Workstation layout and dimensions (including communication systems)
4. Displays and controls design

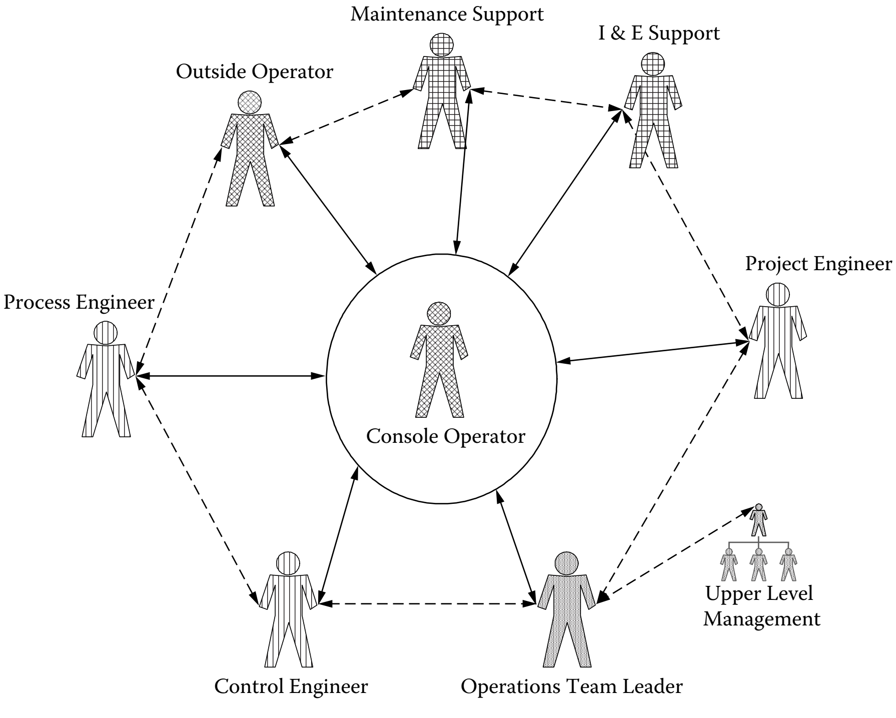


FIG. 4.11h
The required communication interactions in a control center.

5. Environmental design (including air quality, lighting, noise, static, electromagnetic field, hearing clarity)
6. Operational and managerial requirements (practices and organizational policies)

Summarizing, care should be taken in the design and retrofit of existing control rooms to address the communication and ergonomics needs of the operating team. Efficient designs can be effective in dramatically improving operator response to abnormal conditions.

Operator Training

Formal operator training programs, for the most part, evolved from programs designed to meet the first governmental training regulations. The early programs were largely on-the-job training (OJT) systems that were largely modeled after the traditional apprenticeship system. Programs today commonly have a mix of OJT, classroom lectures, and computer-based training (CBT) systems (some of which are programmed to adapt to the student's pace and success).

Increasingly, training simulators have also been developed to simulate the operation of varied industrial processes, with varied degrees of process model fidelity. The simulators are generally either a generic training module at moderate cost or a high-fidelity model of that actual process unit at generally very high cost. These high-fidelity simulators allow operators to operate the plant offline in simulation mode, but with the same look and feel of the actual plant. The simulators can be used for operator training, as well as to verify process engineering and process control designs.

Progressively more companies that are focused on achieving manufacturing or operational excellence have acknowledged the need for structured training in preparing for emergency operations. This is because there is no time to review emergency operating procedures or to call for engineering support during an emergency. By the time an emergency arises, the operators must already have a series of possible response plans internalized or "unless special care is given to developing and maintaining the operator's abnormal operations skills, such as through simulator-based training, he will have difficulty fulfilling his role effectively once an abnormal event occurs."[5]

New regulations have also spurred an emerging approach to training systems based on operator competency, where the specific skills and knowledge for a given job are documented and the operator is assessed against these requirements. Many companies have augmented this approach with a tiered training curriculum as seen in Figure 4.11i.

The number and specifics of the tiers vary from company to company, depending on the work process design at each location. The example in Figure 4.11i shows that all site field operators complete the first three tiers, while all console operators complete Tier 4 and, for the unit process they operate, would complete two additional tiers that would be unique to their specific jobs. From an abnormal condition management perspective, it is useful to consider possible abnormal conditions at each of the tiers, so that the operator would know how to respond to major site risks and sitewide upsets, as well as the specific risks within the process unit and his or her specific job tasks.

One effective way to document much of this is through periodic process hazards analysis reviews. Fundamentally, it is important to train operators in the response to abnormal conditions or upsets either through OJT discussions, classroom training, CBT reviews of past events, or the use of process simulators. This training, periodically refreshed, is a first line of knowledge that can be critical in avoiding or managing abnormal conditions.

Alarm System Design

As was discussed earlier, the financial losses due to upsets are varied and many. More functional alarm systems can help

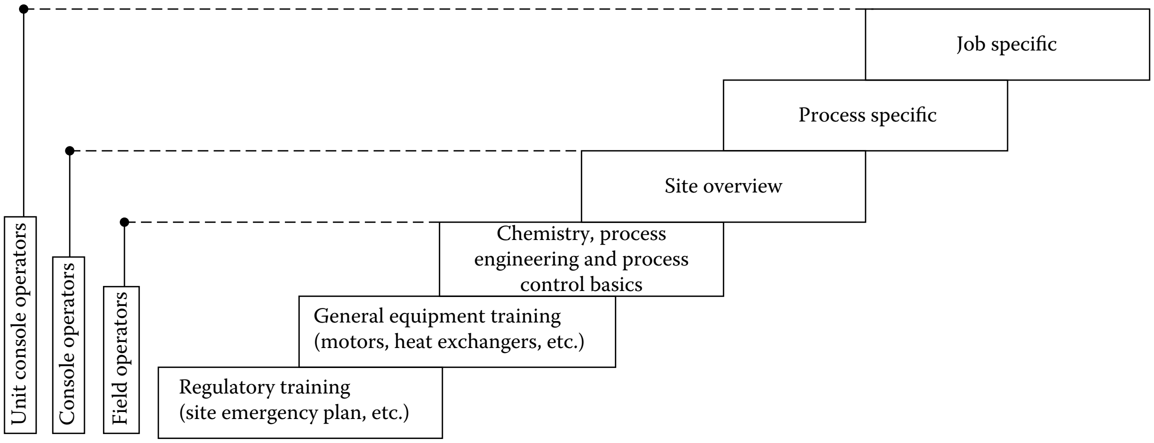


FIG. 4.11i
Illustration of a tiered approached to operator training.

minimize losses in several ways by:

1. Avoiding overloads in upsets
2. Decreasing trips and downtime
3. Decreasing losses in efficiency
4. Identifying areas for improvements in instrumentation, operations, and/or control

It has been observed that severe upsets can corrupt DCS database files, requiring involved rebuilding of these databases. A corrupt database can result in inaccurate alarm system displays, providing the operator with an inaccurate image of the process. The possibility of inaccurate alarming could justify an alarm management evaluation project.

Perhaps the most commonly referenced guideline for the design and use of alarm systems is the Engineering Equipment and Materials Users Association (EEMUA) guideline entitled "Alarm Systems, a Guide to Design, Management and Procurement." The guideline provides a comprehensive look at the design and procurement of alarm systems and the assessment of alarm system performance.

Additionally, the ISA SP18 committee started work in 2004 on a new ISA standard (ISA-18.00.02), tentatively titled "Alarm Systems Management and Design Guide." The Abnormal Situation Management Consortium published a guideline titled "Effective Alarm Management Practices" in 2004.

The following ten-step process, as depicted in Figure 4.11j, is recommended for managing an alarm system performance improvement project:[6]

1. Set alarm system objectives
2. Set an alarm philosophy
3. Set an alarm interface standard
4. Set up the standard approaches
5. Set up the collection infrastructure
6. Set goals and measures
7. Perform alarm analysis
8. Perform alarm rationalization (as needed)
9. Implement improvements
10. Repeat the process

Alarm System Objectives As with any course of action, setting clear objectives is critical to success. Example objectives include the elimination or optimization of:

- Chattering alarms, i.e., alarms that rapidly cycle in and out of the alarm state, causing a large load for both the DCS system and the operators.
- Nuisance event-based alarms, i.e., alarms that are just expected steps in routine procedures or sequence of events.
- Duplicate alarms, i.e., alarms that detect and annunciate both a high and a high–high process variable condition. These alarms eclipse one another and essentially are duplicated entries in the alarm system.
- Stale alarms, i.e., alarms that have been in their alarm condition for a protracted period and do not need to remain in the active alarm summary.
- Disabled or inhibited alarms, i.e., alarms whose annunciation is not enabled, indicating a potential problem in the alarm design.
- Operator alarm changes. Frequent changes by the operator indicate a potential problem in the alarm design.
- Other operator actions, e.g., changes to outputs or set points, that might indicate areas for improvement in instrumentation, process design, or process control.

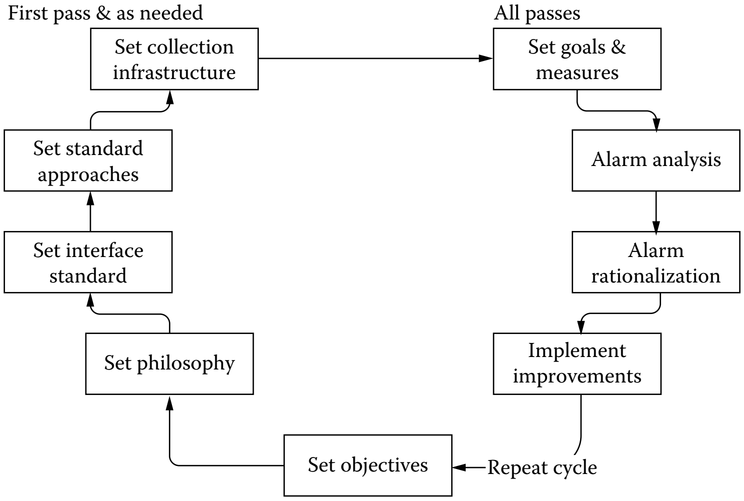


FIG. 4.11j
Steps required in an alarm system performance improvement project.[6]

- Controls not in Normal Mode, if the process is not controlled in the normal mode, that might also indicate potential areas for improvement in instrumentation, process design, or process control.

Alarm Philosophy Setting an alarm philosophy can be difficult, since the varied stakeholders have disparate needs and generally strong opinions. The needs of most of the alarm systems tend to meet a variety of different needs. From a process safety standpoint, it is vital to start and end with operators. The alarm system should primarily be an Operations tool.

It is also important to ensure that Operations is aware of the limitations of the DCS, such as the bandwidth limits that cannot be exceeded, because systems can be overburdened to the point of being unreliable in an upset. The alarm practices could easily be different for every DCS system at a given site, though some level of common ground is useful. Examples of common philosophy include:

- Stipulate that annunciated alarms *require* an operator action. If no action is needed, then it is *not* an alarm.
  - While it can be useful to alarm a variety of items for operator information (e.g., alarms to alert the operator that a sequence program has moved from one step to another), it is important to remember that the primary goal of the alarm system is process safety. Overloading the system with message-type information can make it fail to perform during emergency operations.
- Reserve emergency alarms for true emergencies. (*Note:* Emergency is defined as the highest DCS alarm priority, recognizing that different DCS vendors have different naming conventions.)
  - It is very useful to have a sound and color reserved for true emergencies that require critical responses from the Operations team. Deciding which alarms warrant emergency could be based on, e.g., time to respond or potential impact severity.
  - It is vital to ensure that the operator response is not diluted with too frequent emergency alarms. True emergencies should be rare occurrences.
- All alarm types need a defined rationale and response.
  - A good first step at improvement can be setting the rationale for the emergency alarms in the system. However, every alarm type should have a defined meaning and expected operator response. It is important that the operations team understand and internalize the meaning of the different types of alarms. Recommendations on how to proceed with defining the rationale will be discussed under standard approaches below.
  - In general, three levels of alarms are considered sufficient. A proliferation of alarm types makes color and sound coding difficult and success at internalizing operator understanding unlikely.
  - Several publications recommend certain ratios of alarm types by priority, including the EEMUA guideline. Such ratios are illustrated in Figure 4.11k:
    - — *Emergency* alarms should represent true emergencies requiring immediate operator intervention and be less than 3 to 5% of the total configured alarms in the system.
    - — The *High* alarm priority should represent upsets that are likely to move the process into emergency operations, requiring timely operator intervention. This category should be no more than a third of the remaining alarms or a maximum of roughly 30%.

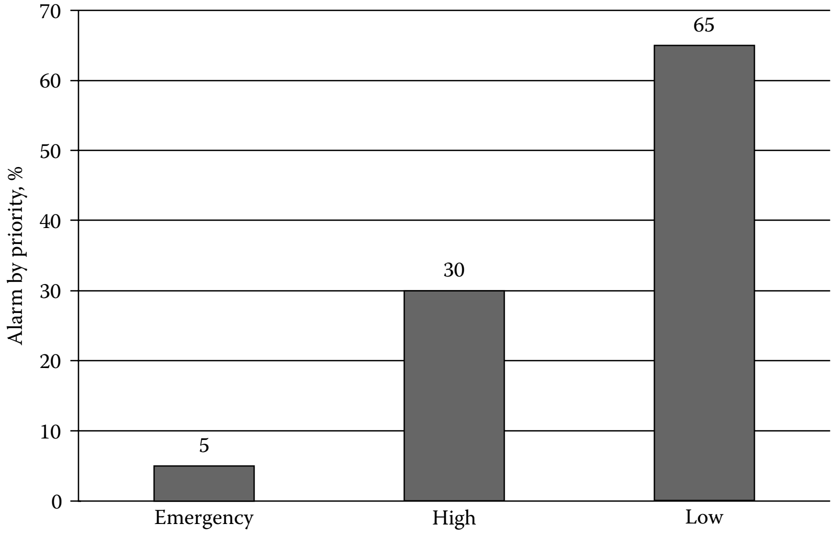


FIG. 4.11k
Recommended alarm priority ratios of low, high, and emergency alarms.

    — The *Low* alarm priority represents the balance of the annunciated alarms and should sound as the operation is moving from normal into upset conditions, requiring operator action, but generally not with great urgency.
- Operator training operations will be provided for all deleted alarms.
    - At first glance, this might seem more of a step in the project process than a true philosophy element. However, it is important to understand that operators have an internal mental model of the process and that a missing alarm will be assumed not active and actions will be taken consistent with that assumption. Thus, removing alarms can have serious consequences if the operators are not adequately trained.

Interface Standards Rapid visual identification of alarm conditions can be an important part of an effective alarm system. Thus it is important that alarms have a standard visual and audible presentation format.

Some example elements of an interface standard include:

- Alarm presentation is consistent
- Alarm presentation is redundant (e.g., by color and symbol to ensure that it is prominent and insensitive to loss of color in the display or to colorblind operators).
- Red and yellow are reserved for alarms *only.*
    - This can be hard to initially sell to the operations team that is accustomed to red used for tripped equipment, but generally once implemented, operators grow to appreciate the value of reserving the color for alarms. One effective means of differentiating tripped versus running equipment is the use of hollow symbols for tripped and filled for running.
- Ensure that the alarm enable state status is redundantly apparent on the graphic, in color and with a coded symbol, e.g., INH for inhibited alarms or DIS for disabled.
- Ensure that the operator is aware of critical code execution status on affected controls (e.g., if code acts on a given controller and it is not functioning, ensure that a visual message or symbol is present on the operating graphic).

Standard Approaches Alarm management projects can be daunting to undertake. There are, however, four things that can make the projects more likely to be successful:

1. Get strong management support and charter a cross-functional team that includes representation of all areas of the operations team.
2. Start with Basic Alarms Cleanup, which will generally show initial improvement.
3. Consider the use of commercial software, which will save time in setup and ensure that someone else will manage any needed updates to keep in synch with the DCS database structures.
    a. Most DCS vendors have some form of alarm management software.
    b. Other commercial packages that are available were listed at the beginning of this section.
    c. If commercial packages are not feasible, consider automation of key reports to ensure that the analysis is ready whenever time is available to work on improvement.
4. Find a specific area of concern to the stakeholders and work on that in parallel with the rest of the general project. Success at resolving a key stakeholder worry will help build enthusiasm.

Improvement projects can be most effectively managed when the engineering decisions are made at the beginning and then used as templates for specific applications. The most common solutions are:

- Database management
- Group alarming
- Alarm "silencing"
- Alarm analysis
- Alarm rationalization

Database management should be designed to run in an automated manner, providing a periodic audit of the DCS, including:

- Disabled or inhibited alarms
- Changes since last audit

Group alarming can be used where a process upset results in numerous alarm conditions, which the operator does not need to be aware of:

- For example, eliminate alarms when a given section of the plant is down and secured. Include logic to test for down and secure. Perform a safety review indicating which alarms are not needed when down and secure. Visual indication of the true state of all alarms should be available at the user interface.
- A second example of alarm grouping is to take all of the noncritical alarms in one area of the process into a single alarm. This can dramatically improve the clutter of stale or standing alarms (defined as the alarms that remain on and standing in the alarm summary). Standing alarms can be a significant issue during upsets as an additional information load that the operator has to sort through to understand the state of the process.

Alarm "silencing" can be an effective temporary tool. Consider the automation of alarm "silencing," where the operator

TABLE 4.11l
Types of Alarm System Analyses

Alarms per time period (minutes, hours, days, shift, etc.)
Time in alarm (both event and cumulative over the period)
Time between new alarms
Duplicate alarms (query for alarms with *xxx)
Alarm annunciation not enabled
Operator changes to: alarm parameters, controller mode, etc.
Time not in normal mode
Time to acknowledge
Time in alarm (time to return from alarm)
Sorting: by area, unit, priority, unit and priority, operating shift, day or night, etc.

is allowed to inhibit the audible annunciation of alarms until the next shift change. This can eliminate the stress of an alarm that is malfunctioning and chattering on and off with high frequency or any other temporary condition. It can also help manage standing alarms due to instrumentation problems.

Such an automation step would reintroduce the alarms after shift change (checking for a lack of abnormal condition) or on operator demand. This would eliminate the possibility that the alarms could be left in their inhibited state for a long period of time.

Alarm analysis is the bulk of the effort needed in an alarm management program. A variety of analyses can be useful, as listed in Table 4.11l. The most important data are the number of alarms per operator per time period, the number of chattering alarms, and the number of standing or stale alarms.[6]

Alarm rationalization is the formal process for setting alarm priorities. The alarm philosophy will define the meaning and calculation methods for alarm priorities. The most common approaches involve the consideration of the time available to intervene and correct the situation and the potential magnitude of impact (type of injuries and equipment or environmental damage). There are varied commercial software packages available to assist in the documentation and calculation of priorities.

Rationalization is the process for applying the alarm priority to individual alarms. It is perhaps simplest to do this rationalization as part of the periodic process hazard analysis (PHA). Alarm rationalization can be a time-consuming process, though detailed review is likely to be only needed on such parts of the process that can lead to emergency operations and should be identified as part of the PHA cycle.

Alarm Collection Infrastructure Most DCS systems are designed to manage a rolling buffer of alarm information and have no standard means to store a continuous stream of alarm information. Alarm understanding is generally improved with a review of operator changes, which is also generally stored on a rolling buffer. A wide variety of options are available for the collection of alarm information, from custom to commercial applications:

- DCS vendors are starting to offer optional packages.
- Data historian companies generally offer optional tools.
- Most DCS systems can stream this information to a “system printer,” which can be directed to a virtual file system.
- The system printer can also be replaced with a connected personal computer, running terminal/modem software managing the data or directly to varied third-party applications.

The most critical issue is to make the data readily available to all users and to automate as much as possible. The most effective improvement programs make the alarm information available to the entire operations team on their desktops so that access to data is not a roadblock to improvement.

Goals and Measures Alarm Management is never complete; monitoring the performance of the system is a never-ending task. Thus it is important to set intermediate improvement goals and minimum performance limits on key performance targets. Management commitment to the goals will help make the process of Alarm Management become a way of life.

Common problems beyond simple redesign of alarm systems include:

- Noncritical digital indicators failing/failed
- Field switches that need new set points due to process changes
- Noncritical loops with chronic performance issues
- Critical loops in need of maintenance
- Different modes of plant operation needing different alarm trip points

Graphical User Interface

The graphical user interface can be designed in several ways to support the operators during upsets. Key issues to consider include:

- View of span of control
- Navigation design
- Dedicated displays for abnormal conditions
- Dedicated displays for task support
- Basic display design guidelines

View of Span of Control In modern digital control systems, the full span of control for an operator is often not shown on a single screen. In an abnormal operating condition, this lack of the “full picture” can lead to errors in judgment. Thus it is recommended that an operating graphic be designed that clearly shows the entire operator span of control. In order to manage graphic throughput constraints, alarm grouping techniques or alarm counts by area can be used to manage the information needs for this process overview display.

Navigation Design Navigation design can be critical because it directly impacts the time to understanding and ultimate action. Key considerations include:

- Display every loop in the unit on at least one operator graphic.
- Keep the number of operator graphics to a minimum.
- Keep the navigation scheme fairly simple and flat.
  - Most companies have display design guidelines that limit the depth and breadth of the graphics hierarchy. There has been significant research into "menuing" and navigation design. A simple rule of thumb is that every loop should be visible within no more than three navigation moves. Graphics for critical loops should be accessible with one move.

Dedicated Displays for Abnormal Conditions Dedicated displays for special tasks and abnormal conditions can be very effective. However, care must be taken to either use the same type of layout as used in the normal operating graphic or have periodic training sessions that include the actual use of the dedicated displays. It can be very effective to display context-sensitive information. Examples include:

- Information about interlocks; include details on permissive and bypasses that are available as required.
- Conditional advice embedded in the graphics that displays only when logic tests are true (or only at high intensity when true).
- Sections of the graphic dedicated for the display of information relevant to the actions of the operator.

Dedicated Displays for Task Support Where there are specific tasks that are more prone to generate abnormal conditions, it can be valuable to perform a process mapping of the task and to then generate specific operating graphics to support the effort. Additional automation ideas are likely to be generated as part of this process mapping. There is an emerging field of specialty in procedure automation and verification, partially driven by OSHA 1910.119 and partially by initiatives like abnormal situation management.

Basic Display Design There are some basic design guidelines that can be effective in designing operating graphics that perform well under abnormal conditions. Perhaps the most important considerations are those concerned with the basic ergonomics of the operator, ensuring that the pace of information delivery matches reasonable human capabilities. Key considerations include:

- The display refresh rate is appropriate. Just because the DCS can refresh the graphic every 1/10 of a second does not mean that this speed is needed. The speed of refreshing the graphic should be based on the actual response time of the process, thus reducing the information load on the operator and on the system.
- Error avoidance techniques should be implemented.
  - Require a double confirmation on critical steps.
  - Redisplay the requested action in the verification, if appropriate.
  - Enforce high and low limits of change on typed entries.
- Design the graphic to ensure that the equipment and piping fade into the background. While the graphic represents the process of, for example, a distillation tower, the valves, pumps, and instruments are more important than the tower or its trays.
- Detailed information should be available, but only on operator demand. Do not clutter the operating graphic with all the information that might ever be needed, but the tag numbers and shutdown limits provide methods to display, etc.
- The direction of flow and the overall layout should be consistent with the actual process.
- Reserve red and yellow for alarms only. It is also prudent to keep the number of other colors to a minimum and apply the meaning of these colors consistently. It can be very effective to make all process equipment gray and all text some other dull color. This makes the red and yellow for alarms very prominent.
- Consider colorblindness and contact in the design of the graphics.

Advanced Detection and Advice A key goal of abnormal condition management is to ensure that the operator gets the correct information at the correct time to be able to make the required corrective action.

Advanced detection and advice systems strive to provide real-time alerts and dynamic advice to the operator. Technologies that support this include knowledge-based systems (e.g., expert systems), model-based systems, data-driven approaches (e.g., statistical process control, Six Sigma, and neural networks), and engineering process control.[7]

Research into abnormal condition management, especially with the ASM Consortium, is currently striving to create a collaborative technology that will further enable the generation and delivery of this advice. Much of this is presented in contextual-based graphics within the DCS or in a technology-specific interface on a non-DCS computing platform. One useful presentation aid is that of the polar star, which provides a multi-dimensional image of "normal" operating conditions for key controllers or properties and a real-time view of the process in the same format.

Providing more time to orient the operator with early detection and dynamic real-time advice can be a very effective means of averting abnormal conditions or of minimizing their impact.

CONCLUSIONS

Abnormal conditions cost the U.S. economy billions of dollars every year. The keys to the management of abnormal conditions are the use of complex control and emergency shutdown systems, as well as early detection systems, comprehensive operator training, performant alarm systems, dynamic well-designed user interfaces, and advanced operator advice systems.

No one of these elements is the magic bullet to solve the problem; rather, they form a basis for context and understanding for the operations team. In reality, the problem is never solved; it is a continuous work process for operations, requiring vigilance and unremitting improvement.

ACKNOWLEDGMENTS

The author would like to recognize the influence of the ASM Consortium in this material and the critical review of Byron Lemmond and Ed Quesada.

References

1. Lorenzo, D., *A Manager's Guide to Reducing Human Error,* Washington, DC: Chemical Manufacturers Association, 1990.
2. Swain, A., and Guttman, H., *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications* (Final Report), Washington, DC: United States Nuclear Regulatory Commission, 1983.
3. Belke, J., "Recurring Causes of Recent Chemical Accidents," Paper Presented at MAPP, 1999.
4. Errington, J., and Bullemer, P., "Designing for Abnormal Situation Management," Proceedings of the 1998 AIChE Conference on Process Plant Safety, Houston, TX, 1998.
5. Emigholz, K., "Improving the Operator's Capabilities During Abnormal Operations: Observations from the Control House," Paper Presented at AIChE Loss Symposium: Human Factors in Plant Operations Session, Boston, MA, 1995.
6. Fitzpatrick, B., "Mustang Alarm Management," Houston, TX: Mustang Engineering, 2005.
7. Cochran, E., and Rowan, D., "Human Supervisory Control and Decision Support: State of the Art," Paper Presented at ISPE Conference, 1995.

Bibliography

Anderson, N., and Vamsikrishna, P., "Ergonomic Best Practices for Information Presentation to Operators," Paper Presented at AIChE Meeting, 2000.

Attwood, D., and Fennell, D., "Cost-Effective Human Factors Techniques for Process Safety," Paper Presented at CCPS International Conference and Workshop, Toronto, 2001.

Brown, S., *Advances in Computer-Based, Multimedia Training Provide Significant Opportunities to Improve Results*, Tucson, AZ: Performance Associates International, Inc., 1999.

Bullemer, P., "Managing Abnormal Conditions: A New Operations Paradigm," Paper Presented at Honeywell North American Users Group Meeting, Phoenix, 2001.

Bullemer, P., Cochran, T., Harp, S., and Miller, C., "Managing Abnormal Conditions II: Collaborative Decision Support for Operations Personnel," Paper Presented at ISA INTERKAMMA, 1999.

Bullemer, P., and Nimmo, I., "A New Training Strategy: Design the Work Environment for Continuous Learning," Paper Presented at the Mary Kay O'Connor Process Safety Center Annual Symposium, College Station, TX, 1998.

Campbell Brown, D., and O'Donnell, M, "Too Much of a Good Thing? — Alarm Management Experience in BP Oil, Part 1: Generic Problems with DCS Alarm Systems," Paper Presented IEE Colloquium on Stemming the Alarm Flood, London, 1997.

Cochran, E., and Bullemer, P., "Abnormal Conditions Management: Not by New Technology Alone," Proceedings of the 1996 AIChE Safety Conference, Houston, TX, 1996.

Cochran, E., Bullemer, P., and Millner, P., "Effective Control Center Design for a Better Operating Environment," Paper Presented at NPRA Computer Conference, 1999.

Cochran, E., Miller, C., and Bullemer, P., "Abnormal Conditions Management in Petrochemical Plants: Can a Pilot's Associate Crack Crude?" Proceedings of the 1996 NAECON Conference, 1996.

Errington, J., and Nimmo, I., "Designing for an Ethylene Plant Control Room and Operator User Interface Using Best Practices," Paper Presented at AIChE National Meeting, 2001.

Gerold, J., "Managing Abnormal Conditions Pays Process Dividends," *Automation World,* September 2003.

Ghosh, A., *Abnormal Situation Management is Your First Life of Defense*, Dedham, MA: ARC Advisory Group, 2000.

Hendershot, D., and Post, R., "Inherent Safety and Reliability in Plant Design," Proceedings of the Mary Kay O'Connor Process Safety Center Annual Symposium, College Station, TX, 2000.

Jamieson, G., and Vicente, K., "Modeling Techniques to Support Abnormal Situation Management in the Petrochemical Processing Industry," Proceedings of the Symposium of Industrial Engineering and Management, Toronto, 2000.

Mattiasson, C., "The Alarm System from the Operator's Perspective," Paper Presented at IEE People in Control Meeting, Bath, UK, 1999.

Metzger, D., and Crowe, R., "Technology Enables New Alarm Management Approaches," Paper Presented at ISA Technical Conference, Houston, TX, 2001.

Mostia, B., "How to Perform Alarm Rationalization," *Control,* August 2003.

Mylaraswamy, D., Bullemer, P., and Emigholz, K., "Fielding a Multiple State Estimator Platform," Paper Presented at NPRA Meeting, Chicago, 2000.

Nimmo, I., "Abnormal Situation Management," Paper Presented at NPRA Conference, 1995.

Nimmo, I., "Adequately Addressing Abnormal Situation Operations," *Chemical Engineering Progress,* September 1995.

Nimmo, I., "Ergonomic Design of Control Centers," Paper Presented at Honeywell Users Group Meeting, 2000.

Nimmo, I., "The Importance of Alarm Management Improvement Project," Paper Presented at ISA INTERKAMMA, 1999.

Nimmo, I., and Cochran, E., "Future of Supervisory Systems in Process Industries: Lessons for Discrete Manufacturing," Proceedings of the MVMT Workshop, Ann Arbor, MI, 1997.

Nochur, A., Vedam, H., and Koene, J., *Alarm Performance Metrics,* Singapore: Honeywell Singapore Laboratory, 2001.

O'Donnell, M., and Campbell Brown, D., "Too Much of a Good Thing? — Alarm Management Experience in BP Oil, Part 2: Implementation of Alarm Management at Grangemouth Refinery," Paper Presented IEE Colloquium on "Stemming the Alarm Flood," London, 1997.

Pankoff, J., "Look Beyond Traditional Organizational Structure: Production-Centered Design Improves, Sustains Total Plant Performance," *Hydrocarbon Processing,* January 2003.

PAS (Plant Automation Services), *The Cost/Benefit of Alarm Management,* Houston, TX: Plant Automation Services, 2000.

PAS (Plant Automation Services), *White Paper: Alarm Management Optimization,* Houston, TX: Plant Automation Services, 1998.

Smith, W., Howard, C., and Foord, A., "Alarm Management — Priority, Floods, Tears or Gain?" www.4-sightconsulting.co.uk: 4-sight Consulting, 2003.

Walker, B., Smith, K., and Lenhart, J., "Optimize Control Room Communications," *Chemical Engineering Progress,* October 2001.

4.12 DCS: Modern Control Graphics

T. L. BLEVINS, M. NIXON (2005)

Industry Control Graphic Standards:	A. Fieldbus Foundation (www.fieldbus.org) B. PROFIBUS (www.profibus.com) C. IEC (www.iec.org) D. World Batch Forum (www.wbf.org)
Partial List of Software Suppliers:	ABB (B,C) (www.abb.com) Allen-Bradley (C,D) (www.ab.com) Emerson (A,C,D) (www.EasyDeltaV.com) Honeywell (A,C,D) (www.honeywell.com) Invensys (A,C,D) (www.invensys.com) Schneider (B,C) (www.schneider.com) Siemens (B,C) (www.siemens.com) Yokogawa (A,C) (www.yokogawa.com/us)

INTRODUCTION

Graphical standards for defining control and calculation were in common use long before the introduction of the DCS. However, early digital control systems were limited by computer, software, and interface technology in supporting graphic control languages and in implementing process control and simulation graphics. With the introduction of the IEC 61131-3 standard in the early 90s, most manufacturers began to provide general support for control graphics.

The IEC 61131-3 standard is in its third edition and defines four control languages for process control and manufacturing automation. Three of these languages use graphical representation of the control, as follows:

- Function block—Allows both continuous and discrete control to be represented by reusable blocks of functionality. The information flow between blocks is by wiring between the function block input and output connections.
- Sequential function chart—Defines calculation and control where the logic evaluation is done in a sequential manner and may follow different paths depending on the operating conditions. Such capability is often required in the sequential controls of batch processes.
- Ladder diagram—Allows discrete logic to be implemented as rungs of contact, coils, and function blocks. It has been utilized in discrete control and in motor control interlock and permissive logic applications.

Major equipment manufacturers and organizations of Foundation Fieldbus and Profibus have supported the use of function blocks for control in the process industry. Many of the fieldbus devices utilized in the process industry today are based on the function block specifications defined by these industry sponsored organizations.

The IEC 61804 international standard for process industry function blocks has adopted much of the function block architecture of Foundation Fieldbus and Profibus. Modern control systems are based on this standard for function block representation. The IEC 61804 standard also defines a migration path that allows older systems and field devices to be used in a consistent fashion within these systems.

The release of batch control standard ANSI/ISA-88.01-1995, Batch Control Part 1: Models and Terminology, led to revolutionary advances in the way factories design, implement, and integrate flexible, modular batch processes. The standard made it possible to design better processes by combining recipes with equipment control in a standardized manner.

Safety logic functions are also being integrated into the DCS systems. Doing so will greatly simplify the configuration and support for safety instrumented control strategies. Improved graphic displays contribute to better control of both continuous and batch processes.

FUNCTION BLOCK REPRESENTATION

Control definitions based on function block representations are very common, providing graphic displays for continuous and batch-type processes. Much of this acceptance is based on the fact that common components that have been traditionally used for process control such as feedback controllers, computing elements, and interface stations can each be represented by function blocks. For example, the analog PID controllers and their interfaces on the central control panel can, along with their associated field input and outputs, be represented as function blocks.

The IEC 61131-3 made a significant contribution in standardizing the graphical representations of measurement, control, and calculations by function blocks. This standard specifies that any independent function which, when executed, yields one or more values may be represented as a block. Multiple, named instances of a function block can be created. Each instance must have an associated identifier (instance name) and defined input and output data structure. All the values of the output variables and the necessary internal variables must persist from one execution of the function block to the next.

A function block instance is graphically represented as a block with the function block name inside the block and the instance name above the block. Any values that are required for the associated calculation must be shown as input connections on the left-hand side of the block. The size of the block may vary depending on the number of inputs and outputs. Formal input and output parameter names must be shown respectively at the inside left and right side of the block.

Measurement and actuator access associated with Input and Output function blocks may not be explicitly shown since they are accessed in a proprietary fashion. For example, the graphical function block representation of the activation of a solenoid (DO function block) based on the logic OR of the contact status of two limit switches (using two DI function blocks) is shown in Figure 4.12a.

Standard for Process Control

The foundation for function block implementation has been further extended by the IEC 61804-2 Function Block standard for Process Control. In particular, this standard requires that function block implementations conform to the following:

- Function block inputs and outputs include both a value and status attribute. The quality associated with a function block input value is indicated by its status. The status of calculated function block outputs is based on the status of the block's inputs and on the block mode.
- All function blocks associated with a control function will support a mode parameter. The mode parameter allows the source of the block setpoint and outputs to be explicitly selected by a plant operator.
- Configuration values used in the block algorithm, operating targets, and calculation results may be defined as contained parameter and do not need to be shown as input connections—as defined in IEC 61131-3. The majority of parameters associated with process industry function blocks fall into this category.

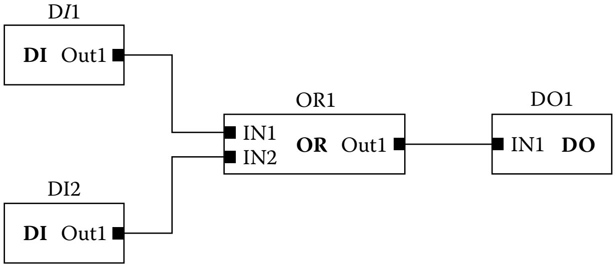


FIG. 4.12a

Block diagram representation of "or" logic, utilizing discrete I/O: The status of two digital inputs (DI) from two limit switches is evaluated by the logic block OR, which sets a digital output (DO) to a solenoid, which will be energized if either of the two limit switches is closed.

Function Block Modes

The IEC 61804 standard defines the mode parameter used in control blocks. The allowed target mode values of a block are defined by the permitted mode attributes of the mode parameter. Thus, the modes available for controlling a block may vary with each block instance. To help an operator select the normal mode of operation from those available, a normal mode attribute is provide in the mode variable. This attribute may be used by a human interface application to help guide the operator in setting the operation mode of a block.

The values assigned to the permitted mode attribute are selected from those defined by the block designer. They are assigned during block configuration to meet the specific requirements of a function block application. Interface devices and temporary devices may use the permitted mode to limit operator changes in target mode. Field devices limit changes in target mode to those allowed by the permitted mode attribute. Once the actual mode is determined, the block execution progresses and the outputs are generated.

Figure 4.12b illustrates a block that contains a setpoint variable and an output variable whose values are determined by the mode variable. The target operating point of a control is determined by the setpoint variable. In addition to indicating how the algorithm is to execute, in this case, the mode also indicates the source for the block's setpoint variable and the destination for its output variable.

As shown in Figure 4.7b, the source of the setpoint or destination of the output variable value may be an input variable, the algorithm itself, a control application or a human interface application in an interface or temporary device.

The effect of the mode on the operation of the function block is as follows:

Out of service (O/S) mode—The block is not being evaluated. The output and set point are maintained at their last values.

Initialization manual (IMan) mode—The block output is being set in response to the status of the "back-calculation input variable."

Manual (Man) mode—The block output is not being calculated, although it may be limited. It is directly set by the operator through an interface device (*A*).

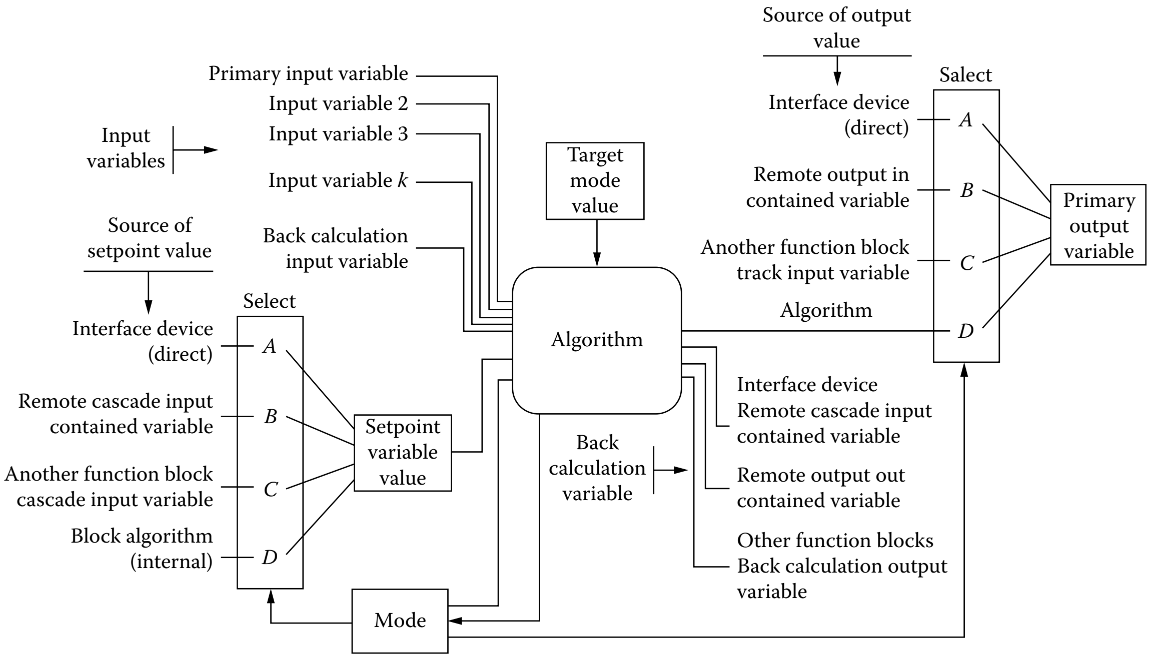


FIG. 4.12b
The internal structure of an algorithm block with mode-controlled set point and output parameters.

Local override (LO) mode—In the local override mode, the block output is being set to track the value of the tracked input parameter.

Automatic (Auto) mode—A local set point value is used by the normal block algorithm in determining the primary output value. The local setpoint value may be written to by an operator through an interface device.

Cascade (Cas) mode — A set point value supplied by another function block through the cascade input parameter (*C*) is used by the normal block algorithm in determining the primary output value (*D*).

Remote-Cascade (RCas) mode — The block setpoint is being set by a control application running on an interface device through the remote-cascade input variable (*B*). Based on this setpoint, the normal block algorithm determines the primary output value (*D*).

Remote-Output (ROut) mode—The block output is being set by a control application running on an interface device through the remote output variable.

Function Block Types

The architecture and basic parameters of function blocks are defined for measurement, calculation, and control. The standard allows DCS manufacturers to identify specific function blocks and to define how the requirement for status attribute will be achieved. Providers of control equipment may contribute by specifying additional features (named profiles) to the standard. The standard includes profiles for Foundation Fieldbus, Hart, and Profibus Organization.

The Foundation Fieldbus profile is used by many manufacturers of fieldbus devices. Many control systems have adopted the Foundation Fieldbus function block architecture and specific function blocks. In this architecture, 10 basic blocks are available to specify the basic control and measurement requirements:

- Discrete input (DI)
- Discrete output (DO)
- Analog input (AI)
- Analog output (AO)
- PID, PI, I control (PID)
- P, PD control (PD)
- Control selector (CS)
- Manual loader (ML)
- Bias/Gain station (BG)
- Ratio station (RA)
- PID

In addition, 19 advanced function blocks have been defined by the function block specification that provide pulse count input and pulse duration output, dynamic compensation of feedforward, characterization, integration, and discrete device control for motors and valve manifolds. Most control applications may be addressed using these blocks.

As a part of input block processing, the Foundation Fieldbus specification requires that checks be performed on associated hardware and software. The status of an output parameter is calculated by the block to give an explicit indication of the quality of the value — *Good, Uncertain,* or *Bad* — as shown in Figure 4.12c.

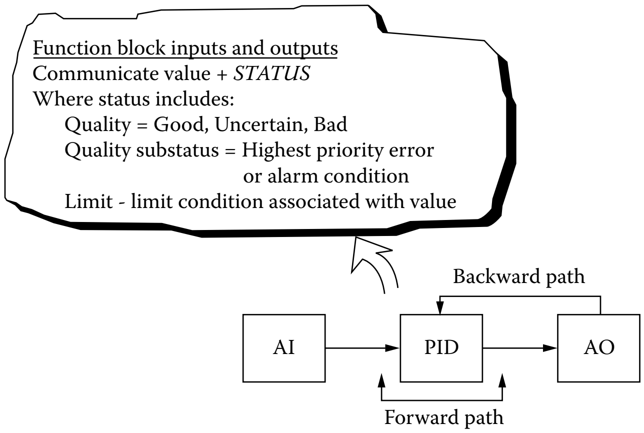


FIG. 4.12c

When a function block calculates an output value, status and substatus statements are also generated. The status statement indicates whether the quality of the output value is good, uncertain, or bad. The substatus statement indicates whether the value has reached a limit or alarm setting.

A substatus attribute indicates the primary condition, which determines the quality. Also, there is an indication in the status attribute whether the value has reached its high or low limit. When a block input or output parameter is accessed for viewing at an operator console, historian, or diagnostic tool, both the value and its associated status can be communicated. The quality information and its substatus may be useful in diagnosing the cause of a Bad or Uncertain measurement.

Quality is indicated by a substatus attribute. If the substatus indicates that the quality is Good, it can also provide additional information that may be needed for control or monitoring purposes. For example, if the substatus of an input block is Good, it will also give information on the status of alarms that are defined for that block.

As part of block definition, the processing and propagation of status is defined. To facilitate the use of status information, the representation and meaning of quality and its associated substatus have been defined in the Foundation Fieldbus specification.

SEQUENTIAL FUNCTION CHART (SFC)

The IEC 61131-3 standard defines the capabilities of the sequential function chart (SFC) that is used to determine the execution of control functions. The set of SFC execution control elements as defined by IEC 61131-3 was originally designed as a program organization unit for a PLC.

Many process control systems have adopted this graphical language as the primary means of sequential logic specification. Sequential evaluation of function blocks may be structured as sets of steps and transitions that are interconnected by directed links that use the SFC graphic language.

At any time the state of logic evaluation is defined by a set of active steps and by the values of the internal and external output parameters. The definition of these elements according to IEC 61131-3 is as follows:

Step—A situation in which the behavior of a program organization unit with respect to its inputs and outputs follows a set of rules defined by the associated actions of the step.

Transition—Represents the condition when control passes from one or more steps preceding the transition to one or more successor steps along the corresponding directed link. Each transition has an associated transition condition, which is the result of the evaluation of either a Boolean expression, a ladder diagram, or a network in the function block diagram language. When the transition condition is evaluated as TRUE, the transition is triggered and associated steps are enabled or disabled.

Action—Can be defined as a collection of networks in the function block diagram or by various other means such as a collection of rungs in the ladder diagram language. It is used to assign the result of an expression to a database item, which can be an SFC step.

A specific task may be represented graphically using steps, transitions, and actions. A sequence in an SFC is drawn as a series of steps and transitions. Steps are represented by boxes and transitions by vertical lines with crosses attached, as illustrated in Figure 4.12d.

Each step contains a set of actions that affect the process. At any given time, one or more of the steps and transitions can be active. Each time the SFC scans, the active steps and transitions are evaluated. When a transition evaluates as TRUE (for example, when a transition condition is met), the steps prior to the transition are made inactive and the step(s) following the transition become active. This way, the SFC can sequence through the various control states as defined by the SFC diagram.

Transitions allow single-stream or parallel execution of logic within the SFC. By using a sequence select divergence, divergent paths may be used to enter alternative sequences. To converge paths again, a sequence select convergence can

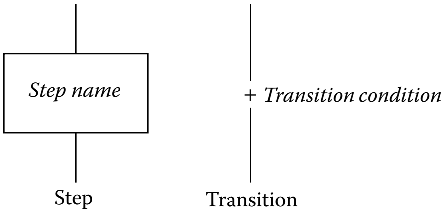


FIG. 4.12d

Representation of a step and of a transition, which are used in a sequential function chart (SFC) as sequentially numbered blocks.

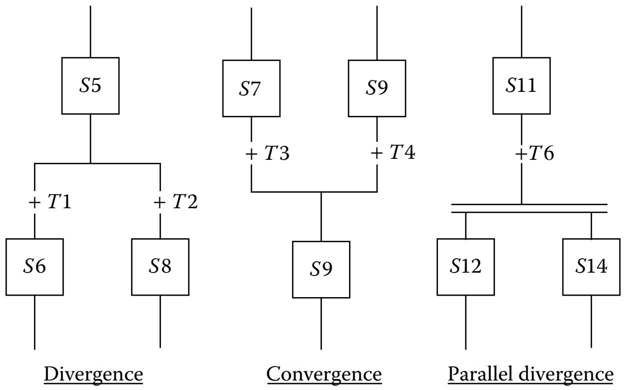


FIG. 4.12e
Examples of divergence and convergence paths in an SFC chart presentation.

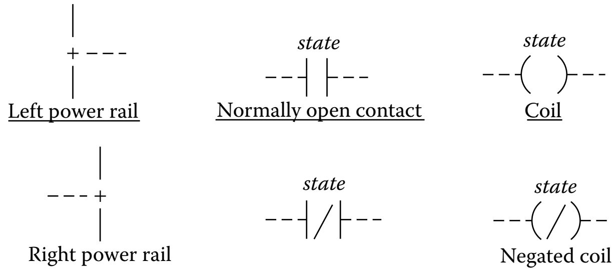


FIG. 4.12f
The component element symbols used in a ladder diagram include the left and right power rail and the contact and coil symbols.

also be graphically defined. Simultaneous or parallel sequences can be specified using parallel divergence. The graphic representation of divergent and convergent paths is illustrated in Figure 4.12e.

The logic associated with an action or the logic for a transition may be defined as an expression. Within IEC 61131-3, the semantics and syntax of structured text language (ST) may be used to define an expression. Expressions may also be used to provide information for process operators so that they can make control decisions.

LADDER DIAGRAMS

The ladder diagram as defined by IEC 61131-3 is utilized by many manufacturers of programmable controllers as their primary means of implementing discrete logic. By using this language, the PLC may test and modify data by means of standardized graphic symbols. These symbols are laid out in networks in a manner similar to "Rungs" of a relay ladder logic diagram.

As defined in IEC 61131-3, the ladder diagrams are bounded on the left and right by power rails. The state of the left rail is considered ON (powered) and no state is defined for the right rail. A horizontal link element is represented by a horizontal line and transmits the logical state (ON or OFF) of the element on its immediate left to the element on its immediate right. A vertical link element consists of a vertical line intersection of one or more horizontal link elements on each side. The state of the vertical element represents the inclusive OR of the ON states of the horizontal links on its left side. The state of the vertical link is copied to all of the attached horizontal links on its right.

Three basic elements that are used within the ladder diagram to express the implemented logic and calculations are:

Contact—Imparts the state of the horizontal link on its right side that is equal to the Boolean AND of the state of the horizontal link at its left side with the associated Boolean input, output, or variable. Four basic types of contacts are defined: normally open, normally closed, transition-sensing, and negative transition.

Coil—Copies the state of the link on its left to the link on its right without modification and stores an appropriate function of the state or transition of the left link into the associated Boolean variable. In addition to the basic coil and negated coil, various types of latched, retentive, and transition-sensing coils may also be used in the ladder diagram.

Function Block—At least one Boolean input and one Boolean output must be shown to allow for power flow through the block.

Examples of the graphical representation of a power rail, contact, and coil are illustrated in Figure 4.12f.

Within the ladder diagram, the network is evaluated in a top-to-bottom order. Figure 4.12g provides an example of how these basic elements may be combined to serve a motor control application.

Even though logic applications can be implemented by using ladder logic, many applications require too many elements to fully address the control and interface requirements of the system. For example, a motor control that includes features such as capturing first-out trip indication and includes

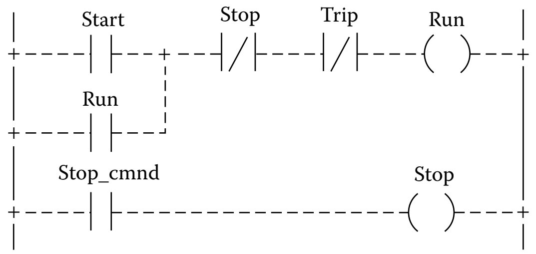
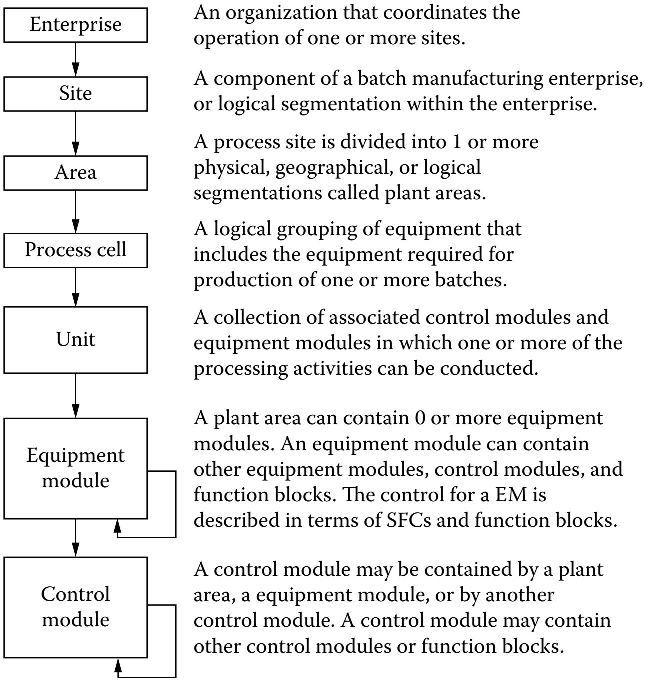


FIG. 4.12g
This motor control application illustrates how the component elements of a ladder diagram can be combined to serve a particular logic sequence.



An organization that coordinates the operation of one or more sites.

A component of a batch manufacturing enterprise, or logical segmentation within the enterprise.

A process site is divided into 1 or more physical, geographical, or logical segmentations called plant areas.

A logical grouping of equipment that includes the equipment required for production of one or more batches.

A collection of associated control modules and equipment modules in which one or more of the processing activities can be conducted.

A plant area can contain 0 or more equipment modules. An equipment module can contain other equipment modules, control modules, and function blocks. The control for a EM is described in terms of SFCs and function blocks.

A control module may be contained by a plant area, a equipment module, or by another control module. A control module may contain other control modules or function blocks.

FIG. 4.12h
ISA Standard S88 provides a basic structure for the hierarchical definition of batch controls.

permissive as well as interlocks in the control logic may require 50 or more elements and many rungs of logic.

This complicates the implementation and troubleshooting of logic interlocks in the field. Since a single function block can be designed to perform the same function in a standardized manner, many control system manufacturers provide a set of function blocks designed to address discrete applications in lieu of supporting the ladder diagram language.

BATCH S88

The ISA SP88 developed the S88 Standard to define batch control as a way to standardize recipe handling, equipment control, and the configuration of batch control systems so they fit together. The S88 Standard has served as the basis for almost all of the newer batch-oriented control systems on the market.

The S88 standard defined a hierarchy of modular control that supports reusable control components. The standard also provides for flexibility by allowing a recipe to determine the product-related characteristics of a process, while equipment control functions are embedded in the reusable modules.

Further, it defined and gave names to the modules that are in continuous and discrete control—referred to as "basic control" in the S88 standard. The standard also defines a layer of procedural controls that allows sequences of control actions to step a process through a recipe-defined procedure. Moreover, it defines the coordination controls, which keep all the other pieces of the system properly sorted.

Overall, the S88 Standard has provided an efficient means for collecting data in an understandable way from a process that, by its nature, is in a constantly changing state. S88 was adopted as an ANSII standard and has become an international standard (IEC61512-01).

The basic structure of the S88 standard is illustrated in Figure 4.12h.

SAFETY LOGIC

Safety systems and safety interlocks must be integrated into the overall process controls of the plant. Safety interlocks serve the protection and safety of operators, equipment, and the environment. Two major types of interlocks are used:

Failure interlocks—which are continuously active and usually serve to initiate equipment shutdowns
Permissive interlocks—which evaluate conditions in order to allow the starting or the continuation of actions

Several control languages have been used to implement safety instrumented control strategies. The most often used languages used are ladder logic, SFCs, and function blocks.

More recently there has been a move to embed safety features inside the DCS. Doing so leverages the configuration, diagnostics, alarming, history collection, and control capabilities of the DCS.

CONCLUSIONS

The IEC 61131-3 and IEC 61804-2 are graphical control language standards used in many modern control systems. The function block-type graphic language defined by IEC 61804-2 can be used in many measurement, calculation, control, and discrete logic applications.

For logic and interlock applications, the SFC-type control language defined by IEC 61131-3 may be effectively utilized. The ladder diagrams graphic language defined by IEC 61131-3 is supported by many PLC manufacturers for the graphic definition of their discrete logic. However, because it requires a large number of rungs for such applications as motor control, it can complicate the troubleshooting of the system. As a result, many control systems provide function blocks, which are designed to fully address such applications.

S88 has been widely adopted by the batch processing industries and is now part of most modern DCS systems. Similarly, the SIS safety control strategies are also embedded into some of the DCS systems.

Bibliography

Bruccoleri, M., La Diega, S. N., and Perrone, G., "An Object-Oriented Approach for Flexible Manufacturing Control Systems Analysis and Design Using the Unified Modeling Language," *International Journal of Flexible Manufacturing Systems,* Vol. 15(3), pp. 195–216, 2003.

Diedrich, C., Russo, F., Winkel, L., and Blevins, T., "Function Block Applications in Control Systems Based on IEC 61804," ISA Conference, September 10–13, Houston, TX, 2001.

IEC 61131-3 — Programmable Controller Languages, 1993.

IEC 61804-2 — Function Blocks for Process Control, 2004.

John, K-H., and Tiegelkamp, M., *IEC 61131-3: Programming Industrial Automation Systems,* New York, Springer-Verlag.

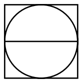
Lewis, R., "Design of Distributed Control Systems in the Next Millennium," *Computing & Control Engineering Journal,* Vol. 8(4), p. 148, 1997.

Lewis, R. W., *Programming Industrial Control Systems Using IEC 1131-3,* IEE Publishing, October 1998.

Mello, R. W., "Flexible System Control and Special-Purpose Languages," *Dr. Dobb's Journal,* Vol. 23(2), p. 68, 1998.

Vogel-Heuser, B., and Biermann, C., "Requirements of Process Control Description Language for Distributed Control Systems (DCS) in Process Industry," *Iecon,* Vol. 3, pp. 2272–2277, 2002.

4.13 DCS: Operator's Graphics



Distributed control display
for an indicator, controller,
recorder or alarm point,
usually for a video display.

Flow sheet symbol

T. L. BLEVINS, M. NIXON (2005)

Types of Software Products:
A. Graphic Symbol Library
B. Graphic Display Application

Partial List of Suppliers:
ABB (A,B) (www.abb.com)
Emerson (A,B) (www.EasyDeltav.com)
ESA Technology (B) (www.esa.int)
GE Fanuc Automation (B) (www.faunc.com)
Honeywell (A,B) (www.honeywell.com)
ICONICS (A,B) (www.iconics.com)
Invensys (A,B) (www.invensys.com)
Kessler-Ellis Products (B) (www.kep.com)
National Instruments (A,B) (www.ni.com)
Nematron (A,B) (www.Nematron.com)
OMRON Electronics (B) (www.omron.com)
Reichard Software (A) (www.reichard.com)
Rockwell (A,B) (www.rockwell.com)
Schneider Electric (B) (www.Schneider.fr)
Siemens (A,B) (www.siemens.com)
XYCOM Automation (B) (www.xycom.com)
Yokogawa (A,B) (www.yokogawa.com/us)

INTRODUCTION

The operator's workstation is the area where operators follow the process and, if assisted by fast and accurate translation of raw data into useful animations, trends, and patterns, can make the required decisions and take the correct actions. As was shown in Figures 4.2d and 4.2e in Section 4.2, one continuously attended operator console is usually dedicated to each section of the plant.

The operator's console includes one or more cathode ray tube (CRT)/liquid crystal display (LCD) monitors (described in Section 4.5), a standard or custom keyboard, a sound card, speakers, and a pointing device such as a mouse, trackball, or touch screen. This operator interface either replaces, or in case of upgrading of existing plants (Figure 4.3f in Section 4.3) supplements, the control room panel, which was/is filled with single-case analog controllers, meters, and digital indicators.

In addition, the operator interface also replaces the previously panel-mounted start/stop pushbuttons and status indications, chart recorders, annunciators, and subsystem interfaces. In the early DCS systems, because of processor and memory limitations and because of lack of software, the operator interface displays consisted only of preformatted faceplates that provided measurement, alarm, and control. In modern DCS the faceplate-type displays have been replaced by graphic displays.

From the console the operators respond to alarms, adjust the operation of the process by changing set points or other parameters, "zoom in" on particular portions of the process for further details, and utilize specialized batch, advanced control, or business applications.

Here, the hardware and functions of the operator console are discussed first, followed by a more detailed description of the operator's graphics including both static and dynamic ones.

OPERATOR CONSOLE EQUIPMENT

The operator console is usually an Intel processor-based workstation with single or dual 19-in. (475-mm) or 21-in. (525-mm) monitors. Memory requirements vary from 256 Mbyte to 1 Gbyte or greater. Hard disks can be either SCSI or IDE and range from 10 to hundreds of Gbytes.

Peripherals include the two-button mouse, track ball, or touch screen with full duplex audio built into the motherboard. The allowable environmental operating ranges for the consoles are temperature 10 to 35°C (50 to 95°F), relative humidity 8 to 80% (noncondensing), vibration 0.25 G at 3 to 300 Hz for

FIG. 4.13a
Standard workstation hardware includes a monitor, CPU, keyboard, mouse, and speakers.

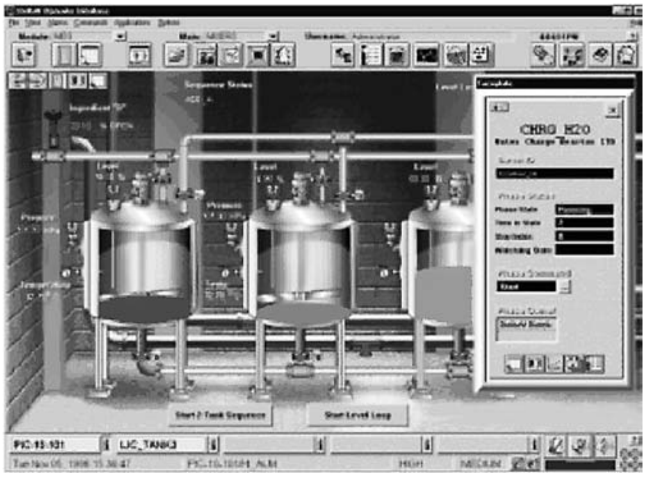


FIG. 4.13b
Graphic display providing an overview of a unit operation in the plant.

a maximum of 15 minutes. A typical operator station is illustrated in Figure 4.13a.

Video Display

The process monitor, sometimes referred to as the operator's window into the process, is usually a 19-in. (475-mm) or 21-in. (525-mm) diagonal, multicolor CRT or LCD display. Some DCS systems offer smaller, mono-color screens for remote or dedicated functions. A few suppliers offer 25-in. (625-mm) diagonal screens for wall-mounted display, which are observable from the working area of the control room. The more common 19-in. (475-mm) and 21-in. (525-mm) monitors are for tabletop or console mounting.

Most suppliers depend on visible cursor control to define screen areas; some use light pens. Touch-sensitive screens (Section 4.22) are quite common. They give the operator the ability to call up, for example, a detail from a larger graphic display by touching the portion of the screen where the corresponding process is located.

One type of touch-sensitive screen uses a grid of conductive material that changes circuit capacitance when a crossing pair is touched. The material is embedded in a sandwich of transparent plastic. Another uses a grid of infrared light beams. Touching the screen breaks two crossing beams and triggers an appropriate response.

Flat tubes are seldom used because of cost and resolution limitations. Similarly, standard distributed control systems do not accept voice commands or talk back to the operator. The technology exists for both, and at the present rate of technological change, these features will be soon available.

The most difficult task at the console-based operations center is to effectively display on a limited screen area a complex plant operation, monitored by hundreds if not thousands of sensors. This task is made more difficult by the fact that a human being is capable of assimilating only a limited amount of information at one time. Because of these considerations there is a need for both "bird's-eye view" type overview displays, and for more details as contained in equipment graphic displays (Figure 4.13b). These displays tend to emulate the graphic panel of the traditional control room.

If the operator needs additional information concerning the operation of a particular control loop, detailed faceplate displays are also available (Figure 4.13h). Other displays provide graphs, trends, and information on dynamic responses. Most importantly, such features as hot links allow the operator to "zoom in" on specific problem areas by drilling into detailed information that concerns only that area.

Although providing useful displays for operators is a common goal of all DCS manufacturers, the solutions provided vary. Some suppliers give more emphasis to the keyboard in moving among the displays, effected by the use of dedicated function keys, addresses based on instrument tag numbers, addressing by group numbers, and function keys that are display dependent.

Display-dependent function keys relate directly to the display on the video screen: depressing the key selects the associated screen segment and expands its information content in the next display. Other techniques used in obtaining this kind of telescoping effect can include the use of auxiliary cursor positioning (mouse), touch-sensitive screens, light pens, pan-and-zoom joysticks, and some other more advanced and less developed techniques, such as voice actuation.

Keyboards

Both standard and custom-made keyboards are used. Standard keyboards tend to sit on a table top. Custom keyboards are often incorporated into the monitor housing or connected to the operator station by cable. The keys may be movable pushbuttons or they may be printed squares on a flexible membrane. The membrane-type keyboard is familiar to users of microwave ovens and pocket calculators. It has a cost advantage, besides permitting the design of imaginative keyboard layouts.

The switches operated by the keys may be sealed reed switches, Hall-effect switches that are operated by actuating a magnetically energized semiconductor, or capacitative switches that are operated by the motion of a plate at the end of a plunger, which increases the capacitative coupling between two other plates.

The membrane-type switch has a flexible, hermetically sealed covering. When the pattern of a key printed on the membrane is pressed, a conductive elastic sheet is pushed through an opening under the key picture, making contact with another conductive sheet beneath it.

Some custom keyboards are implemented as a touch screen/monitor combination. This approach provides considerable flexibility in arranging the key layout, choosing which keys to automate or script, and support. Usually a fourth display is dedicated for this style of usage.

Some keyboards use special function keys, which are programmable by the user. Others use a conventional typewriter keyboard combined with blocks of special-function keys whose purpose is predefined. Table 4.13c lists the functions that are commonly provided by keyboard keys.

TABLE 4.13c
Some of the Typical Key Functions Provided on DCS Keyboards

Standard typewriter keyboard:

- Numberpad: 0–9, + and –
- Cursor control: up, down, left, right, home
- Cluster
  - Alarm
    - Acknowledge
    - Defeat
    - Restore
- Trending
  - Trend time space
    - Increase
    - Decrease
    - Range limit adjust
      - Increase
      - Decrease
- Mode select
  - Automatic
  - Manual
  - Cascade
  - Computer
- Direct input keys (input number keyed from keypad)
  - Set point
    - Output
    - Bias
    - Ratio
  - Set-point trim—slow and fast entry
    - Raise
    - Lower
  - Output trim—slow and fast entry
    - Raise
    - Lower
  - Logic
    - On, Raise, Start, Reset
    - Off, Lower, Stop, Reverse
- Overview select
- Detail select
- Special function keys

Peripheral Devices

Operator stations can be customized by adding extra memory, disk drives, networking cards, monitors, printers, and in some cases, backup devices. Whereas in the past printers were standard accessories, today they also store events in structured or relational databases to provide better reports, responses to queries, and analysis. Some government agencies, such the Food and Drug Administration (FDA), require the storing and inclusion of all operator actions in the batch end report.

REMOTE AND WEB-BASED STATIONS

A number of vendors now offer remote access into their process as well as access to remote systems. The approaches seem to have been standardized; the two prominent ones are referred to by some suppliers as remote clients and Web pages.

Remote Clients

Remote Client access supports terminal services or remote desktop functionality. This capability allows full function operator workstations to be located remotely from the DCS. If it is used by engineers, they can operate and troubleshoot the DCS from their desktops with a direct local area network (LAN) connection or from remote locations using any of a variety of communications methods, including dial-up modems and virtual private network (VPN) connections. This capability is illustrated in Figure 4.13d.

The remote client is made up of two hardware components—a client computer and a server computer. The server computer provides information to the remote clients. The client computer can be any Windows-based computer or thin-client hardware capable of running the remote desktop

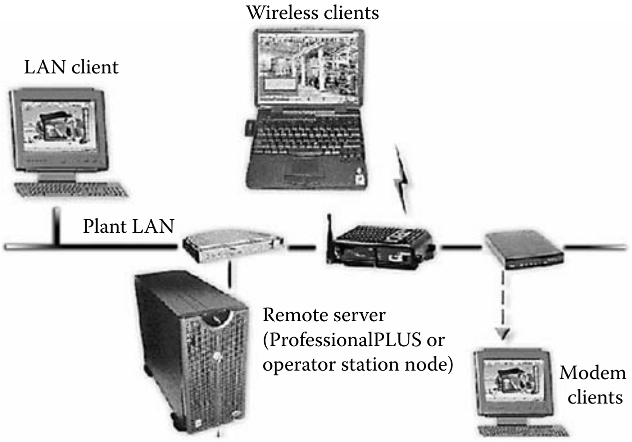


FIG. 4.13d
The control and access capability of the operator workstation can be made available remotely by wired or wireless means.

FIG. 4.13e
Example of a Web-style interface, the use of which is supported by some DCS suppliers. These displays can monitor process conditions and alarms and can also provide multivariable trend recordings.

connection software. Operator station software is not installed in this computer.

Web Pages

The widespread use and acceptance of the Web has led a number of vendors to fully integrate and support Web-style interfaces. Although security remains the key issue in this approach, at least one of the vendors, Honeywell, appears to have gone to great lengths to secure its Web services infrastructure.

Web pages and Web services offer a number of advantages. They include the ability to integrate more enterprise application content and make that content available to many users on their local or remote desktops. Examples include monitoring and analysis of process conditions, performance indicators, and alerts from devices and equipment. The use of Web pages is illustrated in Figure 4.13e.

OPERATOR GRAPHICS

A graphic operator interface shows the piping, instrumentation, and other equipment that is part of a unit operation of the process. The primary advantage of such an approach is that more information can be provided to the operator. However, the effectiveness of such interfaces depends a great deal on the accuracy and completeness of the graphic displays. Standardized display packages cannot provide that accuracy, while custom graphics require substantial investments of time from both the user and the vendor.

Modern distributed control systems include a variety of standard tools that can be used to create graphic operator displays. However, since the process equipment and associated field devices are unique to each plant and because most plants develop their own standard symbols and color schemes, the development of the graphic interface tends to be unique for each plant. The control system manufacturer does provide some tools and some guidance in the development of the graphic displays, but the actual work is usually done by the design engineering firm or the user.

The characteristics of the graphics systems vary with the needs of the users, but they all have to meet some basic requirements:

- The operator graphics must be highly reliable because process control operators cannot control the "process" without them. Their characteristics include:
  - *Must always be available.* In many cases, the application must be available for long periods of time (e.g., weeks, months, even years) without unplanned disruptions.
  - *Must be interacting.* Operator must always have the opportunity to "move on" to other displays and to get predictable responses to new directives even when previous requests take "too long" to complete or "everything is changing" due to a plant upset.
  - *Must be mistake-resistant.* Repeatable behaviors and user interaction features have to be provided to help reduce the possibility of high-cost "human errors," especially in high-stress situations.
- The graphics should provide extremely responsive interactions and fast call-up of process and historical information. This means that even large user-configured displays should be updated with live data from the DCS system within 1 second.

- The updating of data points, alarms, history, and dynamic graphic objects on the displays should occur in real time.
- The graphics should be suited for a dedicated (kiosk-style) operator environment and be resistant to programs or data on the workstation being damaged by users or people gaining access to unintended applications. The degree of restriction is usually user configurable so that the user can decide on access to the workstation.
- The graphics should interface with a wide range of user interface hardware features:
  - Two-, three-, or four-monitor video configurations
  - 4:3 (PC format) and 16:9 (HDTV format) aspect ratio monitors
  - Mouse + keyboard operation
  - Touch-screen operation
  - In newer systems, Tablet PC and PDA form factors and user input techniques
- The display graphics should provide the means for integration of operator tools/applications (e.g., batch, diagnostics, history viewing) with user-configurable displays (and user display logic) and also easy access to information related to display elements. These systems usually also provide easy access to auxiliary information sources (e.g., video feeds) and collaboration technologies (e.g., e-mail and instant messaging).
- The graphics should minimize the requirement for the operator to spend time in managing the windows and should allow the use of data sources not specifically engineered for use by the particular display.
- The display graphics must be provided with integrated user security and authentication facilities to support fast user switching and multi-user authentication for electronic signature requirements.
- The displays should provide a platform for improved information capture and analysis for operations staff. This usually includes features such as snapshot logs and the analyzing of alarm floods and history.

TYPES OF DISPLAYS

The operator station in the central control room functions somewhat like the cockpit of an airplane, where the pilot sits and observes all the information relevant to the performance of the plane. The operator of a DCS system depends on the video screen of the CRT/LCD for plant information. Because the preparation of the various graphic and faceplate displays for a particular DCS application is both expensive and time consuming, it is important that the responsibility for their preparation be clearly resolved before the purchase order is issued.

Each supplier has its own complement of lists, menus, and libraries (including measurement units and messages), which might include:

- Graphic displays
- Detail displays
- Faceplates
- Alarm summary displays
- Event summary display
- Trend displays
- Loop tuning displays
- Diagnostic displays
- Standard reports

The operator interface most often consists of overview, detail, faceplate, and trend displays. These will be discussed next.

Overview Graphic Displays

The graphic display of the whole plant provides a graphical and logical representation of the process, which makes it easier for operators to visualize what is happening. Overview displays are graphical displays with a lot of information to provide the operator with a "bird's eye view" of the process.

Such an overview display, as in Figure 4.13b, shows a section of the plant, with abnormal conditions animated (e.g., blinking) and/or colored (red). Additional animations are often provided to make the display more representative of the process. For example, a tank may be in the process of being filled, the agitator might be operating, etc.

The operator, looking at an overview display, can see at a glance the condition of all the loops and can quickly spot a process variable or piece of equipment that is out of control. The operator can also note trends or recognize patterns, even prior to the full development of their consequences.

On/off status can also be shown on an overview display. Discrete conditions (an open or closed switch, for example) can be shown as the presence or absence of a bar rising from the reference line. Sequential events can be displayed by displaying messages that change with the advance of the sequence.

If the overview display indicates an alarm condition the operator can, with a single keyboard stroke, easily call up a more detailed graphic display which shows the active alarm. If the operator wants still more detailed information, another keystroke can often be used to call up a faceplate and detailed display of the loop that generated the alarm.

Graphic Displays

Graphics are valuable training tools and help the operator in following plant conditions when a number of variables are changing. The graphic display capability of a DCS system allows for the creation of both overview and more detailed graphic displays.

Figure 4.13b illustrates the graphic display of a process that consists of three tanks. The graphic contains both process and control information, which is continuously updated in real time. On such a graphic, a pipeline downstream of a valve, for example, can become filled with color when the valve is opened or the symbol of the value can change color, say from black, when closed to green, which can indicate an open condition.

Some graphic displays are also capable of showing movement. For example, when liquid is flowing in a pipeline, when solids are traveling on a belt, when agitators are turning, or when fuel is burning in a combustion process, the associated movements are dynamically displayed. Obviously, this capability adds to both the cost and to the information content of the display.

Faceplate with Detailed Display

The detailed display is usually specific to a single loop or control function. Figure 4.13h illustrates a typical controller faceplate display. The bar graph in the faceplate display can show the controlled process variable, the set point, and the output. The transmitters or other sources from which these signals are received are also listed on the screen.

The display is identified with the tag number of the associated instrument; this tag number is also used in the DCS software to define the function. The tag number is added, using the keyboard like a typewriter to "fill in the blanks." The configuration of a function is done by moving a cursor to locations on the screen where values are to be entered and typing in the values on the keyboard.

In order to avoid errors or startup delays, it is advisable to make the preparation of the various graphic displays part of the DCS system supplier's scope.

Trend Displays

Analog strip chart recorders have been replaced by trend displays in DCS systems. They display the past values of the recorded process variable over a period of time. Some detailed graphic displays (see Figure 4.13e) include a real-time trend graph of the process variable covering the selectable period of the last 90 seconds, 1 hour, or 24 hours. In some displays, several trend graphs can be displayed at once, allowing comparison of the history of several variables. One supplier can accommodate up to four trends graphs in a single display.

Figure 4.13e also shows the trend display of several variables. This information is valuable to, for example, a foreman coming on shift who, by looking at these trends, can observe the recent patterns of operating history. It is also valuable to an operator after an upset has occurred because such trend recordings can help determine which condition was the cause and which were the consequences of an upset.

Trends can also display historical data records. This allows the operator to go back in time to find patterns or conditions that might have caused similar previous upsets.

STATIC GRAPHIC COMPONENTS

ANSI/ISA-S5.5-1985 is an American National Standard for Graphic Symbols for Process Displays. The stated purpose of this standard is to establish a system of graphic symbols for displays that are used by plant operators in the area of process measurement and control.

The goal of such standardization is to help operators comprehend the information that is conveyed through the displays and to provide uniformity throughout the process industries. The benefits of such standardization are a reduction in operator errors, faster training, and more accurate presentation of control systems. This standard is followed by many companies and is suitable for use in numerous industries.

However, the standard's two-dimensional line drawings are a mismatch with current commercial workstation technology, where some of the graphic displays are three-dimensional. Faster computers with graphics card accelerators and vector graphics are supported by most DCS vendors. High-definition visual display units, VDUs, are commonly available based on CRT or LCD technology that support 800 × 600 to 2048 × 1536 resolution. Object-oriented libraries or GDI software allow more complete three-dimensional representations of the process equipment.

Modern DCS systems include a graphic display editor and a comprehensive library of prebuilt display objects with professionally drawn three-dimensional representations of process equipment including pumps, valves, meters, piping, tanks, and other graphic objects.

Also, some systems may offer the capability to import graphics from a number of different sources including AutoCAD and Windows metafiles such as Visio vector drawings, as well as Windows Bitmap (.bmp) and Joint Photographic Experts Group (.jpg) image formats. These components may be used to create a realistic representation of the process with added dynamic information. An operator display created by using a library of three-dimensional components is illustrated in Figure 4.13f.

DYNAMIC ELEMENTS

If users decides to prepare the graphics on their own, detailed familiarity is required not only with the capabilities of the various DCS suppliers, but also with their software and the language they use, which might include such terms as "dynamos," "aliases," and "animation." Some of these terms will be discussed below.

Predefined objects such as knobs, dials, slide bars, and buttons are standard parts of any DCS graphic tool set and can be dynamically linked to information within the control system. Programmers call the workhorse of this linking process the dynamo (which has nothing to do with its "noncomputerize" meaning, the electric generator). The dynamo allows a particular field on the faceplate to be linked with a real-time reading and/or color that describes the state of the parameter associated with that field. In other words, "dynamo" in the language of the noncomputerese means the tool that is used to convert a field on a faceplate into a dynamic one.

A library of custom dynamos is typically provided by the DCS supplier; it can be used to access the actual real-time values of measurements, calculations, analog and control components, etc. Often these preengineered dynamos can be

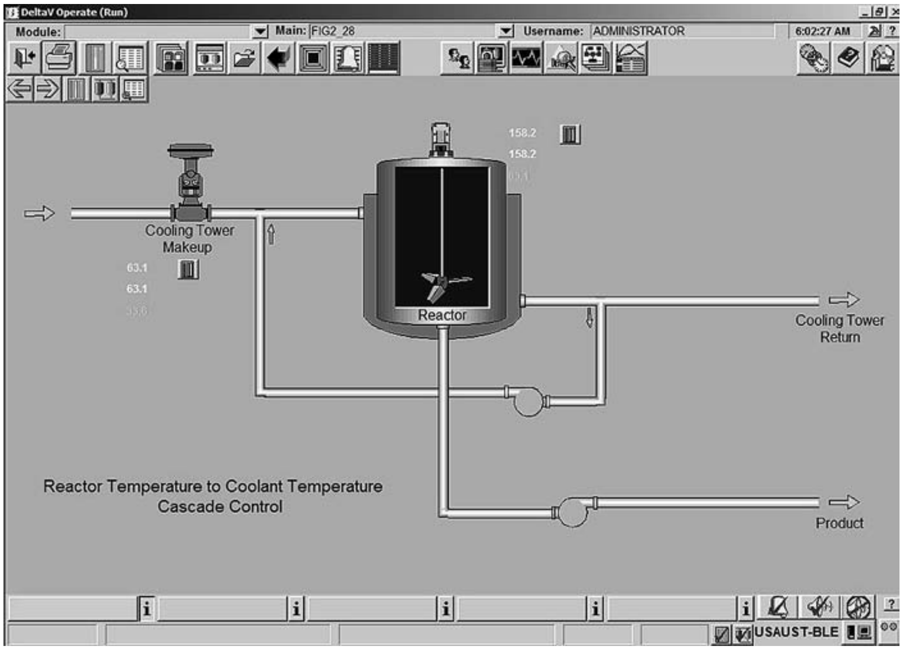


FIG. 4.13f
Some DCS systems offer the capability to import graphics, from which the experienced user can build the type of graphic display shown here.

customized to fit the specific needs and standard practices of the particular plant.

Dynamos

Standard dynamos typically include the key parameters associated with measurement and control functions. For example, a dynamo used to display a control loop may only display the controlled process variable, the set point, and the controller output. Therefore, the engineering units, for example, have to be added by the help of dynamos.

Process alarms associated with control loops can be shown in the dynamo by color change, such as changing the background color behind a number representing the value of a process variable.

Another use of color change can be to indicate that a loop is not in its designed/normal mode of operation. The ANSI/ISA-S5.5-1985 standard provides general guidelines for the use of colors in displays. Red, for example, is reserved to indicate an emergency situation such as the stopping of a motor or the actuation of a highest priority alarm. The predefined dynamos provided with a control system conform to this standard.

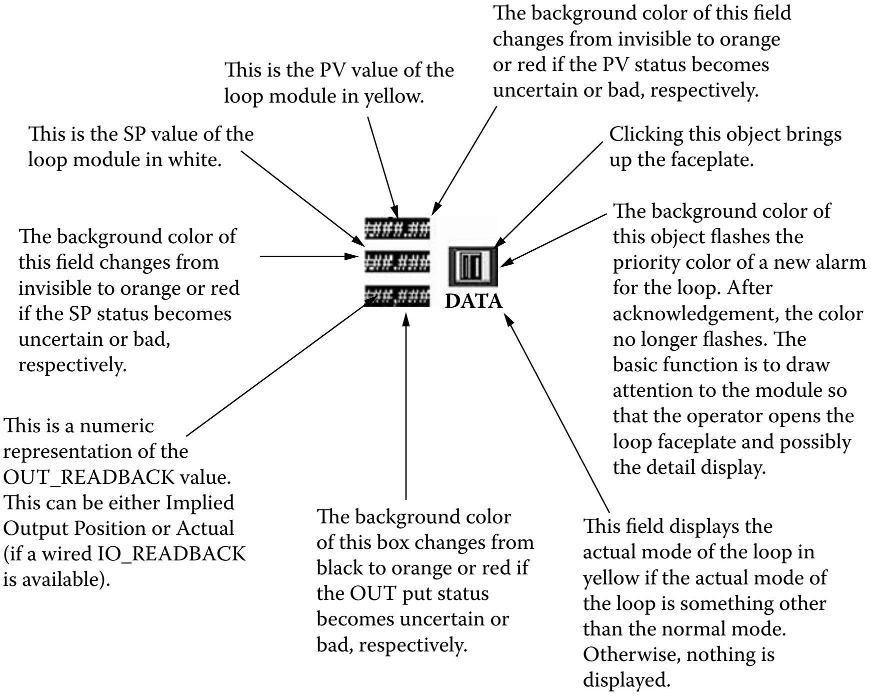
However, these guidelines may not match the traditions and past practices of the plant. In such cases, the colors used in dynamo templates can be modified to conform to the plant's standard. An example of a dynamo is illustrated in Figure 4.13g.

When the operator accesses a dynamo, a faceplate may be automatically presented providing further information on the parameter shown. For example, when a dynamo associated with a control loop is accessed, the faceplate presented to the operator may be used to modify the mode or change a set point or output. In addition, the capability may be provided to access further information, such as loop tuning, by selecting a detail display reference.

Parameters in a detail display, which can be changed by the operator, are determined by the system security. As shown in Figure 4.13g, in some designs, not all the parameter values are displayed all the time; instead, it is necessary for the operator to sequentially obtain them (by clicking) on a parameter-by-parameter basis. An example of a faceplate display with the associated detail display generated by accessing a control loop dynamo is shown in Figure 4.13h.

Just about any property in a display can be converted to a dynamic element and animated. A list of the more commonly used dynamic behaviors are:

- Text
- Visibility
- Scale
- Position
- Rotation
- Skew
- Enabled, disabled
- Stroke (pen size)
- Fill color
- Stroke color
- Gradient (both linear and radial)
- Font
- Font size
- Filter
- Images



This is the PV value of the loop module in yellow.

The background color of this field changes from invisible to orange or red if the PV status becomes uncertain or bad, respectively.

This is the SP value of the loop module in white.

Clicking this object brings up the faceplate.

The background color of this field changes from invisible to orange or red if the SP status becomes uncertain or bad, respectively.

The background color of this object flashes the priority color of a new alarm for the loop. After acknowledgement, the color no longer flashes. The basic function is to draw attention to the module so that the operator opens the loop faceplate and possibly the detail display.

This is a numeric representation of the OUT_READBACK value. This can be either Implied Output Position or Actual (if a wired IO_READBACK is available).

The background color of this box changes from black to orange or red if the OUT put status becomes uncertain or bad, respectively.

This field displays the actual mode of the loop in yellow if the actual mode of the loop is something other than the normal mode. Otherwise, nothing is displayed.

FIG. 4.13g
As can be seen from this example, the "dynamo" allows the programmer or the operator to link a particular field in the faceplate with a real-time reading and/or color that describes the state of the variable or parameter that is associated with that field.

Aliases

In order to speed up the tedious work involved in designing displays, some DCS manufacturers use "aliases." These software tools allow the designer to reuse previously designed displays or display segments and thereby eliminate the need for complete rebuilding of the displays for applications that are similar but are connected to different I/Os.

If the DCS system supports the use of aliases in the definition of similar pieces of equipment, the dynamic display components are usually designed to support dynamic referencing based on the piece of equipment selected for display. In such cases, preconfigured aliases and attributes are used in place of object tag or graphical attributes normally defined as part of the display object.

Thus, aliasing improves flexibility and reusability because display objects can be connected to different I/O points and can represent different graphical attributes, appearances, and security. Such capability can eliminate the need to completely

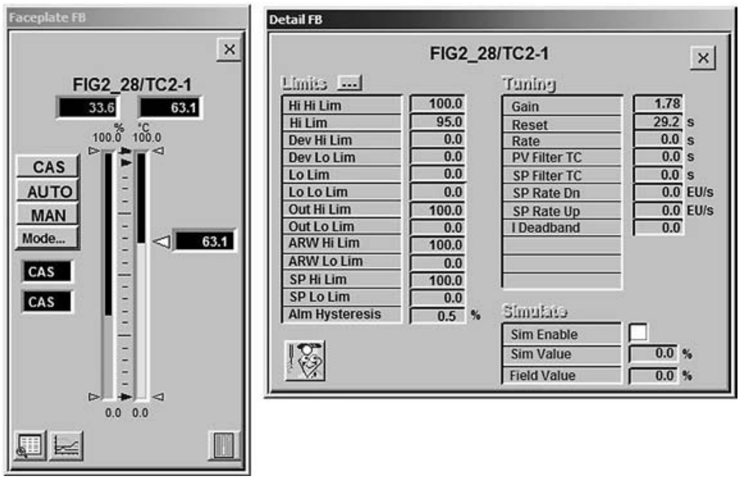


FIG2_28/TC2-1

33.6 63.1

100.0 100.0

63.1

CAS

CAS

0.0 0.0

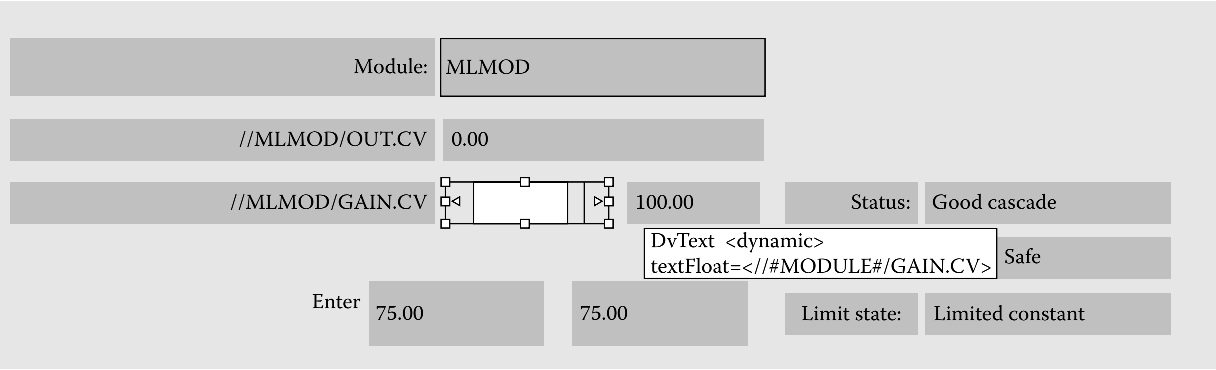
FIG2_28/TC2-1

Limits	
Hi Hi Lim	100.0
Hi Lim	95.0
Dev Hi Lim	0.0
Dev Lo Lim	0.0
Lo Lim	0.0
Lo Lo Lim	0.0
Out Hi Lim	100.0
Out Lo Lim	0.0
ARW Hi Lim	100.0
ARW Lo Lim	0.0
SP Hi Lim	100.0
SP Lo Lim	0.0
Alm Hysteresis	0.5 %

Tuning	
Gain	1.78
Reset	29.2 s
Rate	0.0 s
PV Filter TC	0.0 s
SP Filter TC	0.0 s
SP Rate Dn	0.0 EU/s
SP Rate Up	0.0 EU/s
I Deadband	0.0

Simulate	
Sim Enable	☐
Sim Value	0.0 %
Field Value	0.0 %

FIG. 4.13h
This example illustrates a faceplate that has been accessed (obtained) from a control dynamo.



Module: MLMOD

//MLMOD/OUT.CV 0.00

//MLMOD/GAIN.CV 100.00 Status: Good cascade

DvText <dynamic>
textFloat=<//#MODULE#/GAIN.CV> Safe

Enter 75.00 75.00 Limit state: Limited constant

FIG. 4.13i
An example of aliases used for a graphic element.

rebuild similar display objects. An example aliased substitution is shown in Figure 4.13i.

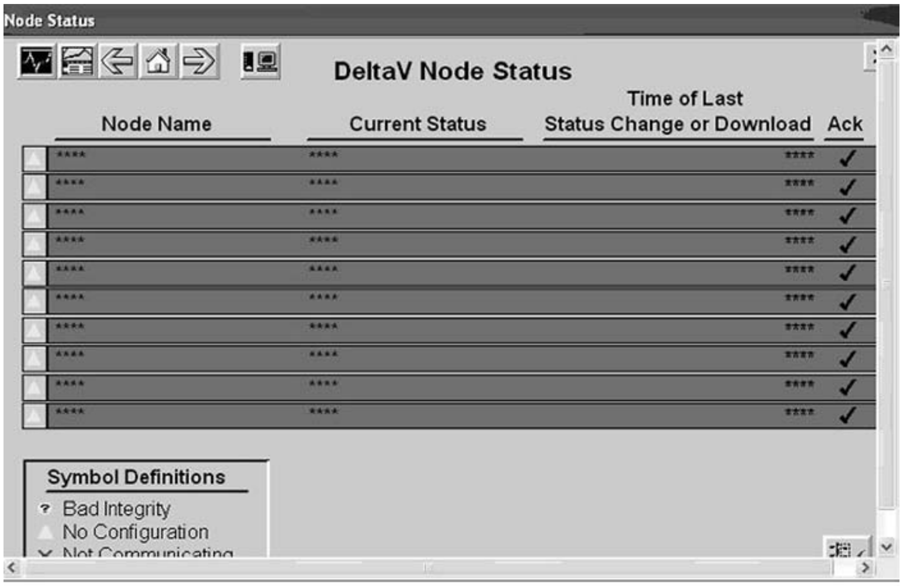
Another way to improve the reusability of displays, faceplates, dynamos, and similar equipment is through the use of class-based displays. With this technique the display designer can develop displays with dynamic or aliased references. For example, the user could base the references in a display against library- or class-based items (e.g., a Phase Class in a batch configuration). When the display is instantiated during runtime, the graphics engine would resolve references based on the object that the display is being opened against.

Display Access

The increased use of a mouse in the operator interface has changed the way operators accesses information within a graphic display. Displays are typically designed to support both horizontal (at the bottom of the display) and vertical (to the right of the display) toolbars. The control system typically provides default toolbars to support the following functions:

- Time and date display
- Alarm list with direct access to the display required to acknowledge the alarm
- Navigation to alarm summary, main menu, system status, last display, page back and forward, and print
- Alarm acknowledge and silencing

Figure 4.13j illustrates the operator display of one DCS supplier, showing scroll bars and navigational items that can be used by the operator.



DeltaV Node Status

Node Name	Current Status	Time of Last Status Change or Download	Ack
****	****	****	✓
****	****	****	✓
****	****	****	✓
****	****	****	✓
****	****	****	✓
****	****	****	✓
****	****	****	✓
****	****	****	✓
****	****	****	✓
****	****	****	✓

Symbol Definitions
? Bad Integrity
No Configuration
✕ Not Communicating

FIG. 4.13j
Navigational items in operator displays.

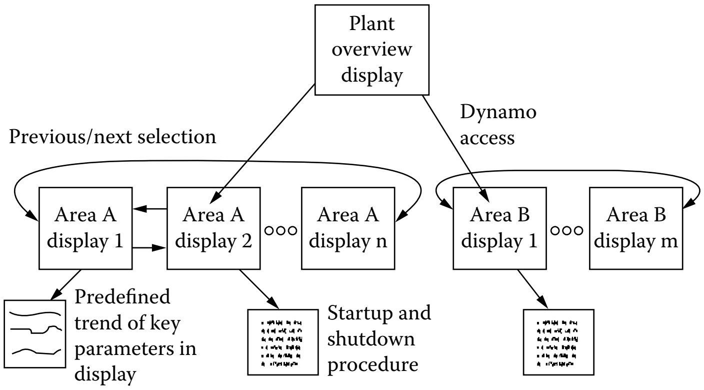


FIG. 4.13k
Example of the hierarchy for accessing information and trend for a display.

Using the page forward and page back feature of the toolbar, it is possible for an operator to access displays that contain information upstream or downstream of the displayed process. In addition, dynamos may be added for accessing another display. Through the use of these tools, it is possible to create a display hierarchy that allows, from an overview display, access to the key display in each process area.

Within the process area, it is possible to pan through the displays using a page forward/backward toolbar. Also, using a dynamo for display access and trend window access, it is possible to provide the operator with startup/shutdown information and trends of key parameters in a display, as well as in-context documentation, as illustrated in Figure 4.13k.

The graphics editor supported by the control system allows incorporation of animation into a graphic display. Using this capability, the static graphic component for process equipment may be modified to indicate the status of the equipment for example motor on or tripped. Also, animation may be used to represent dynamic data associated with the equipment such as showing the level of a tank using a filling technique or showing the status of an agitator by rotating the agitator to indicate that the agitator is on. In newer products, adding dynamics to the properties of controls (send-to-front, -back, send-forward, -backward) are used to show different information in the display based on the state of the process operation.

In some cases it is necessary to create a specialized display complete with user interface (UI) controls such as buttons, check boxes, or sliders. The user would also add behaviors to these specialized displays—the behaviors could be driven from parameters internal to the special display or from parameters in another display, the DCS, or some external source such as OPC. A designer creating such a display would in general keep all of the references unbound so that the specialized display could be activated in context from some other graphic item at run time.

PROCESS PERFORMANCE MONITORING

The graphic display capability may be used to create special displays, allowing easily monitoring of the status of critical equipment. Some examples of these types of applications are:

- First-out indication on a process shutdown
- Vibration monitoring
- Burner management
- Sootblower
- Safety system status

The associated displays are structured to summarize the information. In cases where moving equipment is involved, such as a sootblower, animation may be effective, allowing an operator to quickly access the operation.

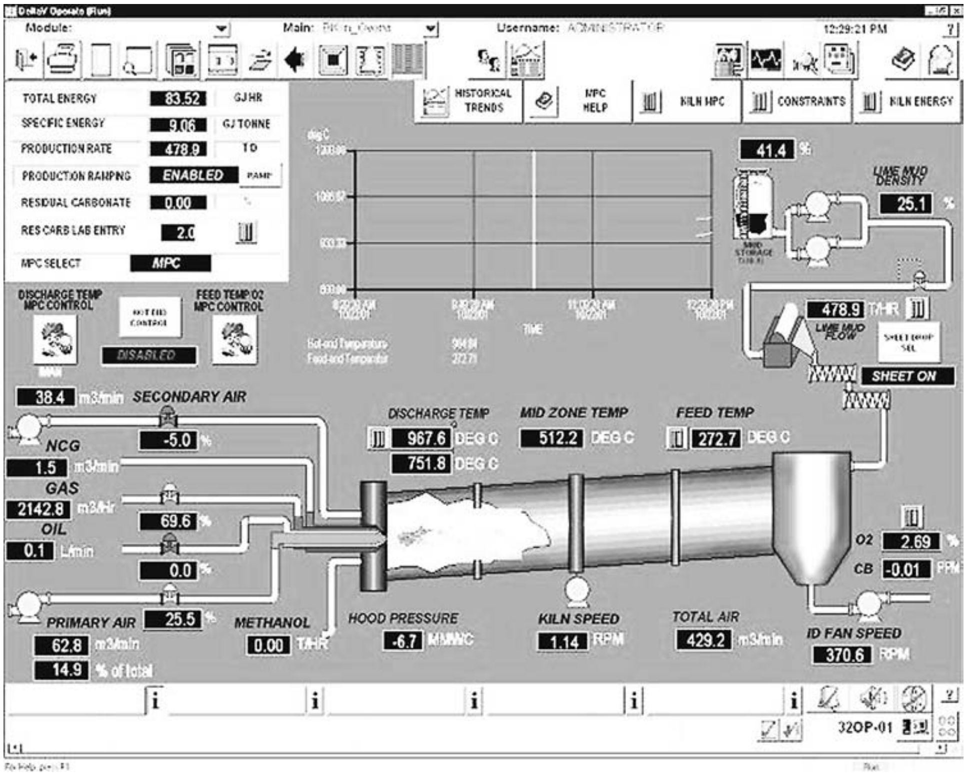
The calculation capability of most control systems is used to implement online calculation of operation cost, efficiency, and other performance indicators. This type of information may be easily incorporated into the operators' graphic display so that they can use this information to improve the process operation. An example of such types of calculations for a lime kiln operation is shown in Figure 4.13l.

PROCESS GRAPHIC DATA INTERFACES

A variety of data sources may be integrated into the process display graphics. The data may be sourced internal to the DCS from the runtime, historians (continuous, batch, event), and alarm services. The data may also be sourced externally from OPC, XML files, or even a hypertext transfer protocol (HTTP) browser. Supporting both internal and external data sources allows the operator displays to be used for a wider range of functions.

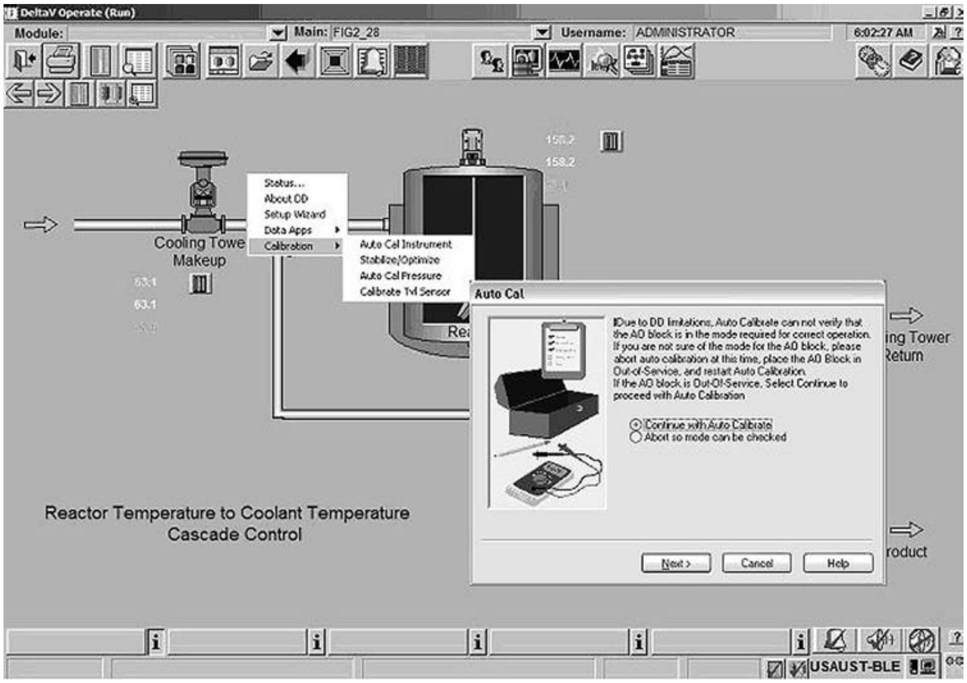
The ability to activate interfaces external to the operator graphics is a very powerful feature. Examples of this include calling up menus in context related to applications external to the graphics—for example, bringing up the menu to calibrate a device from the graphics system. This is illustrated in Figure 4.13m.

In special cases, the information may need to be displayed in a manner that is not supported by the control system. For example, some control systems do not support three-dimensional plotting of matrix values as is required



TOTAL ENERGY 83.52 GJ HR
SPECIFIC ENERGY 9.06 GJ TONNE
PRODUCTION RATE 478.9 TD
PRODUCTION RAMPING ENABLED
RESIDUAL CARBONATE 0.00
RES CARB LAB ENTRY 2.0
MPC SELECT MPC

FIG 4.13l
Process performance indicators incorporated into the operator's graphic.



Due to DD limitations, Auto Calibrate can not verify that the AO block is in the mode required for correct operation. If you are not sure of the mode for the AO block, please abort auto calibration at this time, place the AO Block in Out-of-Service, and restart Auto Calibration.
If the AO block is Out-Of-Service, Select Continue to proceed with Auto Calibration

● Continue with Auto Calibrate
○ Abort so mode can be checked

FIG. 4.13m
An example of performing operations on the process shown in Figure 4.13f by activating various applications from the operator graphics.

to show sheet gauging information. To address these special requirements, the graphics display may support terminal server client/server interface for accessing the subsystem.

CONCLUSION

The graphics capability of modern control systems allows construction of an effective operator interface. The graphic editors provided with the control systems are used to create custom graphic displays that represent the process equipment and the associated field measurements. Preengineered graphic components and dynamos are used to easily create the operator graphics. The display features may be used to create a display hierarchy that allows the operator to quickly access information in any part of the plant. Information from performance monitoring and from subsystems may be directly included in operator graphics or in some cases incorporated using terminal server technology.

Bibliography

Ancilotti, P., Buttazzo, G., Natale, M. D., and Spuri, M., "Design and Programming Tools for Time-Critical Applications," *Real-Time Systems,* Vol. 14(3), pp. 251–267, 1998.

Barnard, H., "Engineering and Ergonomics in Control Room Design," *Measurement & Control News,* Vol. 33(4), p. 121, 1999.

Fransen, P., Borthwick, K., and Thind, P., "Implementation of a PC-Based Operator Interface for a Paper Machine Drive," *Tappi Journal,* Vol. 82(3), p. 209, 1999.

Goonetilleke, R. S., Martins Shih, H., Kai On, H., and Fritsch, J., "Effects of Training and Representational Characteristics in Icon Design," *International Journal of Human-Computer Studies,* Vol. 55(5), pp. 741–760, 2001.

Humbel, C., and Tasset, D., "Advanced Human-System Interfaces: Insight of a Validation Process in a Control Room of a Swiss Nuclear Power Plant," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* Vol. 3, pp. 185–188, 2000.

Hwang, F. H., and Hwang, S. L., "Design and Evaluation of Computerized Operating Procedures in Nuclear Power Plants," *Ergonomics,* Vol. 46(1–3), pp. 271–284, 2003.

ISA-S5.3 Graphic Symbols for Distributed Control/Shared Display Instrumentation, Logic, and Computer Systems, 1983.

Medjdoub, B., Richens, P., and Barnard, N., "Generation of Variational Standard Plant Room Solutions," *Automation in Construction,* Vol. 12(2), pp. 155–166, 2003.

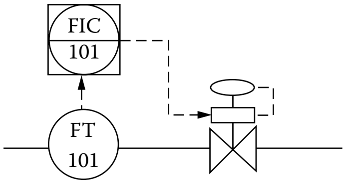
Tollmar, K., and Sundblad, Y., "The Design and Building of the Graphic User Interface for the Collaborative Desktop," *Computers & Graphics,* Vol. 19(2), pp. 179–188, 1995.

Vicente, K. J., "Operator Adaptation in Process Control: A Three-Year Research Program," *Control Engineering Practice,* Vol. 5(3), pp. 407–416, 1997.

4.14 DCS: System Architecture

J. A. MOORE (1985) **B. G. LIPTÁK** (1994)

T. L. BLEVINS, M. NIXON (2005)



Shared DCS flow control loop and valve with digital valve positioner

Flow sheet symbol

Analog Inputs:	Single-ended (reference to signal common) or isolated, channel to ground — 600 V RMS Intrinsically safe (field circuit Class 1 Division 1, Zone 1, Zone 0) and redundancy options
Analog Input Ranges:	4 to 20 mA, 1 to 5 volts with optional digital communications (HART) Thermocouple, RTD Input resolution 12 bits (1 part in 4000) to 16 bits (1 part in 64,000)
Analog Outputs:	4 to 20 mA into a 700 ohm load Intrinsically safe (field circuit Class 1 Division 1, Zone 1, Zone 0) and redundancy options Resolution: 10 to 12 bits
Digital Inputs and Outputs:	24 or 48 volts DC; 120 or 230 volts AC; optically isolated Simplex or redundant
Fieldbus:	AS-Interface, DeviceNet, Foundation Fieldbus, Profibus
Other I/O Types:	RS232, RS422/485 half duplex, RS 422/485 full duplex serial communication ports, binary coded decimal (BCD), pulse count input, pulse duration output, sequence of events
Number of Loops of Control:	8 to 30,000
Display Interface:	CRT or LCD monitor, 19 in diagonal measure
Normal Operating Limits:	Temperature: 0 to 60°C Relative humidity: 5 to 95%, noncondensing Power supply: 115/220 VAC
Network:	IEEE 802.3 (Ethernet) 10BaseT or dual-speed 10/100BaseT network, depending on the hubs or switches; shielded twisted-pair cable connects each node to the hub or switch; maximum cable length from the hub/switch to a node is 100 m. For longer distances, use a fiber-optic solution.
Cycle Times:	Sample interval cycle times range from 0.02 to 0.2 sec/cycle for dedicated loop controllers and from 0.1 to 1 sec/cycle for multiloop unit operations controllers. Scan periods can be fixed or individually specified for each loop.
Partial List of DCS Suppliers:	ABB (www.abb.com) Emerson (www.EasyDeltaV.com) Honeywell (www.honeywell.com) Invensys (www.invensys.com) Siemens (www.siemens.com) Yokogawa (www.yokogawa.com/us)

INTRODUCTION

The instrumentation used to implement automatic process control has gone through an evolutionary process and is still evolving today. In the beginning, plants used local, large-case pneumatic controllers; these later became miniaturized and centralized onto control panels and consoles. Their appearance changed very little when analog electronic instruments were introduced. The first applications of process control computers resulted in a mix of the traditional analog and the newer direct digital control (DDC) equipment located in the same control room. This mix of equipment was not only cumbersome but also rather inflexible because the changing of control configurations necessitated changes in the routing of wires. This arrangement gave way in the 1970s to the distributed control system (DCS).

The DCS offered many advantages over its predecessors. For starters, the DCS distributed major control functions, such as controllers, I/O, operator stations, historians, and configuration stations onto different boxes. The key system functions were designed to be redundant. As such the DCS tended to support redundant data highways, redundant controllers, redundant I/O and I/O networks, and in some cases redundant fault-tolerant workstations. In such configurations, if any part of the DCS fails the plant can continue to operate.

Much of this change has been driven by the ever-increasing performance/price ratio of the associated hardware. The evolution of communication technology and of the supporting components has dramatically altered the fundamental structure of the control system. Communication technology such as Ethernet and TCP/UDP/IP combined with standards such as OPC allowed third-party applications to be integrated into the control system. Also, the general acceptance of object-oriented design, software component design, and supporting tools for implementation has facilitated the development of better user interfaces and the implementation of reusable software.

Major DCS suppliers introduced a new generation of process control systems based on these developments. These systems incorporate commercially available hardware, software, and communications. They fully integrate I/O Bus technology such as Fieldbus and Profibus into the system. Batch technology, advanced control, and safety system-oriented software packages are also being included as embedded technologies within the DCS system, although some suppliers might charge extra for some of that software.

These new systems are the foundation for the instrumentation and control implemented in new grass-roots plants. Also, because of the significant operational improvements, including such advanced features as abnormal situation prevention, which may be achieved with such systems, they are quickly replacing the early DCS systems.

More recently, some of the control functions have begun to move into the field. This move to further distribute control and functionality has opened the door to hybrid controllers available from most of the DCS suppliers.

These newer controllers are also being used as linking devices—interfacing with and integrating multiple I/O buses such as Fieldbus, DeviceNet, AS-Interface, HART, and conventional I/O into a single system, as is illustrated in Figure 4.14a.

DCS control, which has dominated the process control industry for years, has improved its performance and reliability. Over the years the DCS design has become more modular and, due to the reduction in the cost of its hardware, has penetrated even some of the smaller installations—especially where advanced capabilities such as alarm management, batch, and advanced control are needed. Installations consisting of only a single controller, one workstation, a bus arrangement such as AS-Interface and Fieldbus, and a small number of I/O are not uncommon.

A review of the nature of analog and DDC is useful in understanding why the DCS has been so successful.

Analog Control

Operators often receive large amounts of information concurrently. This information arrives in the form of alarms and value updates. The operator needs to be able to focus on critical information when an issue arrives, but at the same time, operators should be able to freely move about displays displaying the key aspects of their process. For these reasons, panel arrangements in the past proved to be unsatisfactory, particularly when handling a plant upset or emergency.

The DCS addressed these problems by assigning priorities to issues. Alarming features such as alarm prioritization, acknowledgement, suppression, and filtering play an important part in plant operations. Alarms are associated with control items so that a user in a single click can move from the alarm directly to the appropriate display or faceplate.

Analog instrument panel designs of the past consumed too much expensive space. Issues included recording instruments whose charts were filed in equally expensive storage locations, fixed instrument arrangements on the panel, and the fact that the panels do not easily permit the relocation of associated devices so that they can be observed as a unit. The requirement to run individual electric wiring from every transmitter and final control device to the control room also increased the cost of a project.

Other costs include the cost of real estate for the control room, the cost of the instrument panel, and the cost of utilities necessary to provide a clean and comfortable environment for the instruments. The DCS addressed this need by including trending directly into the operator workstations. Specific process conditions to be monitored were added directly into the displays. Other related items can be placed on the same trend or on a graphic display along with the trend.

More recently the cost of running individual pairs of wires from transmitters and to valves has been reduced by substituting the use of networks and I/O buses such as Fieldbus and Profibus.

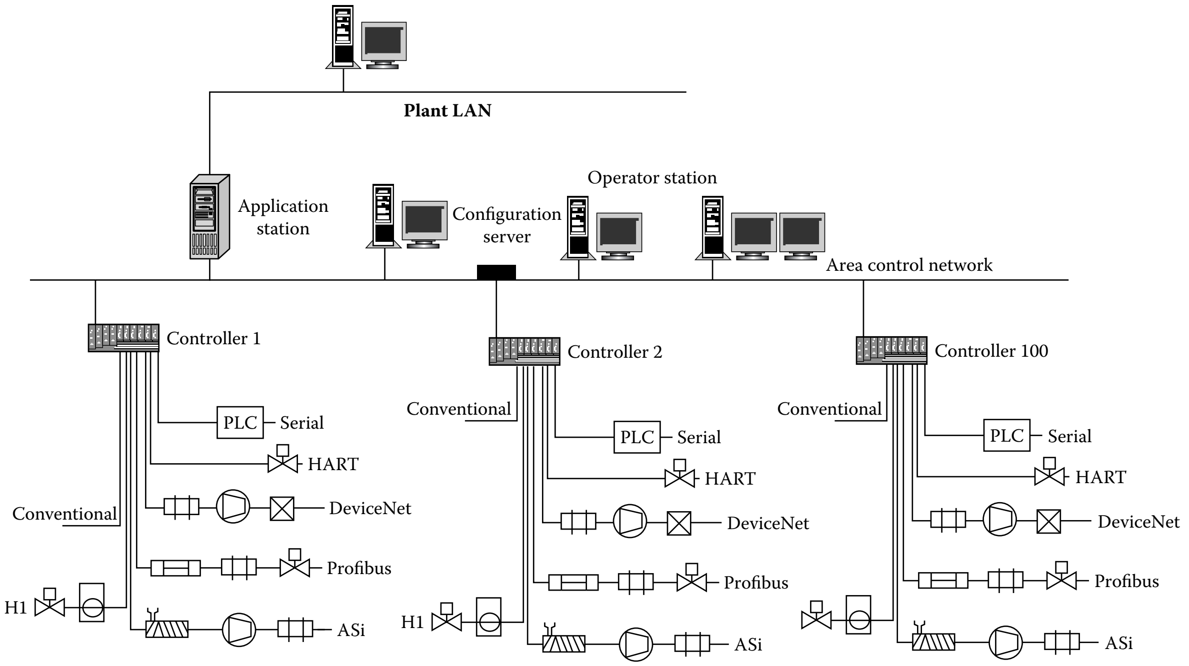


FIG. 4.14a
Current DCS topology integrating multiple I/O buses such as Fieldbus, DeviceNet, AS-Interface, HART, and conventional I/O into a single system.

DIRECT DIGITAL CONTROL

Direct digital control (DDC) and supervisory control of analog systems by computers are two other control options. In DDC a digital computer develops control signals that directly operate the control devices (DDC is rarely used in current DCS installations). In supervisory control, a digital computer generates signals used as reference (set-point) values for conventional analog controllers. Both options were offered by earlier generations of the DCS. Even with DCS systems this is still a problem to be reckoned with because each installation is different and requires a separate programming effort; however, the availability of standardized and tested DCS software packages for the more routine functions reduces this problem.

DISTRIBUTED CONTROL SYSTEM

The major components that make up a DCS process control system are illustrated in Figure 4.14b. The operator interface to the process is typically made up of standard off-the-shelf personal computers (PCs), standard keyboards, mice, and CRT or LCD monitors.

The use of custom keyboards and furniture for the operator interface is often not a viable option because of the associated cost and restrictions on operation. Initial fears that operators would not accept a standard keyboard and mouse have been proven wrong by successful installations on a variety of processes. To provide a wider view and range of control, dual or quad monitor arrangements are often included as part of the operator station.

Similarly, the speed, memory, and disk capacity of personal computers have proven sufficient to address the requirements of the engineering stations that are utilized for system configuration and diagnostics. Also, the price-performance of PCs has driven their adaptation as application stations for the integration of third-party software into the control system. Standard operating systems such as Windows NT and XP are often preferred because of the broad support available to manufacturers.

Equipment that the operator uses to monitor process conditions and to manipulate the set points of the process operation is located in a central control room or distributed on the plant floor close to the equipment. From these locations the operator can (1) view information transmitted from the processing area and displayed an operator display and (2) change control conditions from an input device. The controlling portions of the system, which are distributed at various locations throughout the process area, perform two functions at each location: the measurement of analog variable and discrete inputs and the generation of output signals to actuators that can change process conditions.

Input and output signals can be both analog and discrete. By means of electrical transmission, information is communicated between the central location and the remotely located controller locations. The communication path is either a cable from each remote location to the central station or a single

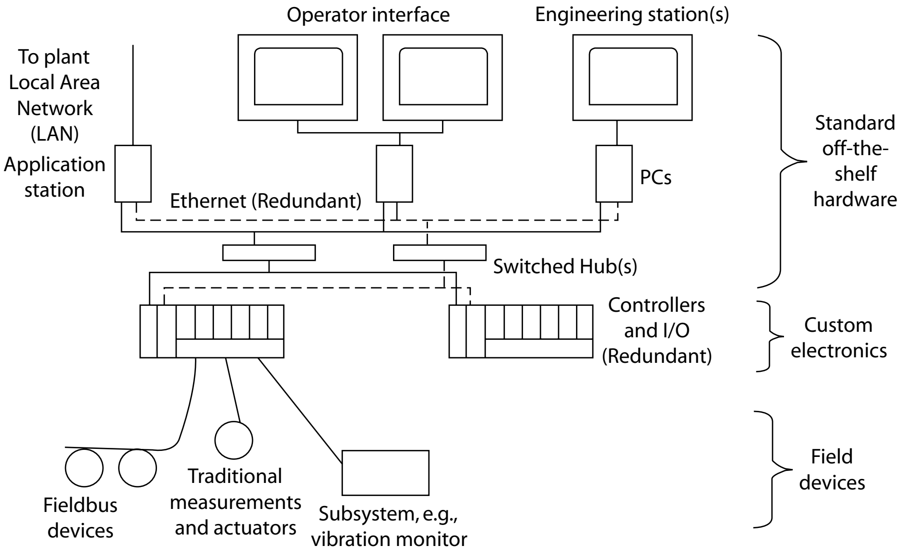


FIG. 4.14b
The components of a DCS system, shown from the perspective of its physical structure.

cable data highway interfacing all the remote stations. The cable in some cases can be a wireless connection via radio, microwave, or satellite.

Functional Components

Distributed control systems are made of several components including workstations, controllers, I/O cards, I/O buses, a control network, control technology, and software. The controllers are connected to field devices via analog, digital, or combined analog/digital buses. The field devices, such as valves, valve positioners, switches, and transmitters or direct sensors, are located in the field.

Smart field devices, if they are designed to conform the bus protocols for the I/O, can communicate on the buses while locally performing control calculations, alarming functions, and other control functions. Control strategies that reside in field-mounted controllers send their signals over the communication lines to the final control elements, which they control. The main components of a process control system are illustrated in Figure 4.14c.

Information from the field devices and the controller can be made available over a control network to the operator workstations, data historians, report generators, centralized databases, etc. These nodes run applications that enable the operator to perform a variety of operations. An operator may change settings, modify control modules within the controller or within the field devices, view the current state of the process, or access alarms that are generated by field devices and controllers. The system may support process simulation for the purpose of training personnel or testing the process control software, plus keep and update a configuration database.

DCS control systems are always sold as packages. Suppliers do not sell only the remote portions or only the centrally located portion. This is because the parts function together as a system; they must be completely integrated and tested as a system. Because the components of the system communicate over a shared data highway, no change is required to the wiring when the process and its control system are modified (Figure 4.14d).

DCS Control Network

The communication link from the controller and from the PCs supports the workstations for the operator, engineering, and applications of the DCS system. For process alarms and values needed by the operation, earlier DCS systems utilized customer communication interfaces. These have been replaced in most process control systems by less expensive Ethernet interfaces operating at communication rates of 10 or 100 Mbit.

With twisted pair wiring, the maximum distance between hub and workstations is limited, but this distance limitation can be overcome by using fiber-optic cables. For example, at a rate of 100 Mbit per second, the distance limitation for twisted pair of wiring is 300 ft (100 m), while with fiber-optic cable it is 6000 ft (2000 m). For the distance capabilities and relative features of various cable designs refer to Figure 4.14e and Table 4.14f.

Uncertainty in communications due to packet collisions can be eliminated by using full-duplex switches rather than hubs because each interface has its own channel on which to transmit. By designing the PC communication interface to utilize two Ethernet interface cards, it is possible to provide fully redundant communications.

As was shown in Figures 4.14b and 4.14c, the operators' console in the control room can be connected through a shared communications facility (e.g., area control network consisting of Ethernet hubs, switches, and CAT5 cable) to

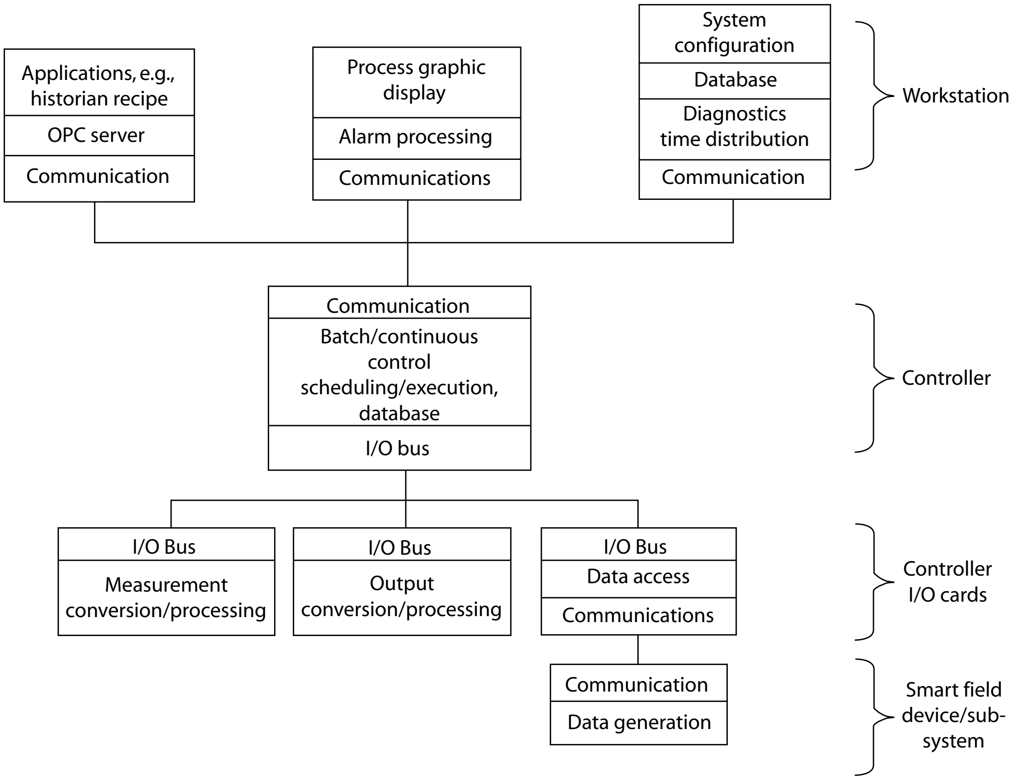


FIG. 4.14c
The main components of a DCS system, shown from the perspective of its main functional components.

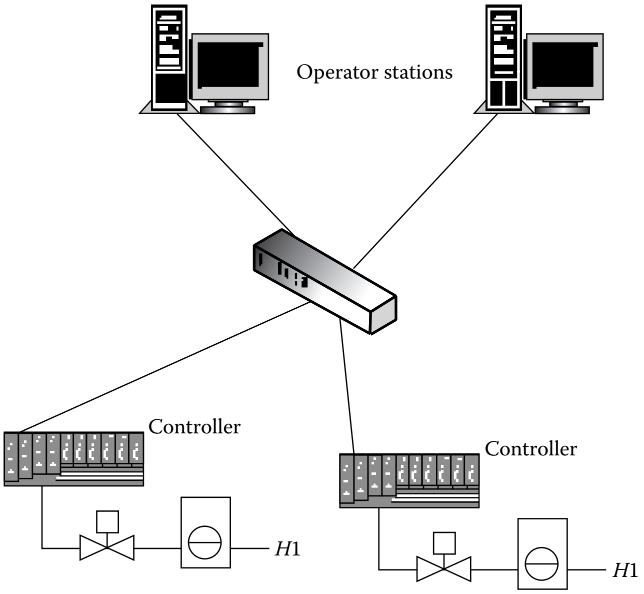


FIG. 4.14d
DCS communications are over area control network. In some cases switches may be used to manage the flow of traffic on specific network segments.

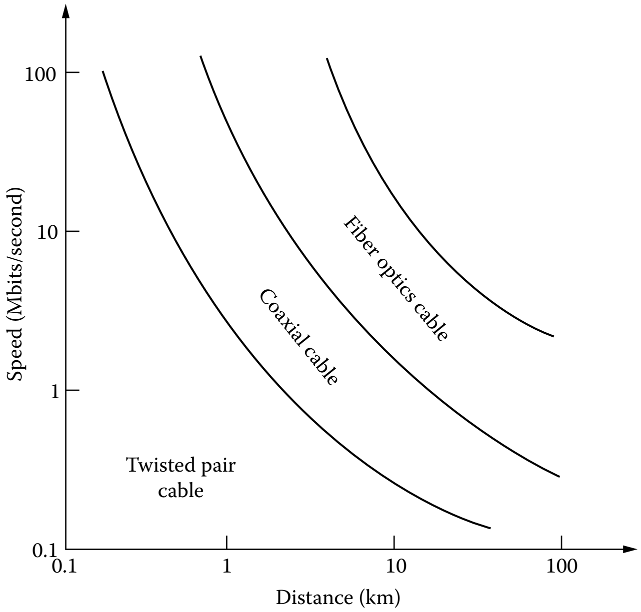


FIG. 4.14e
The capabilities of data highways made of twisted pairs, coaxial cables, or fiber-optic cables. (From M. P. Lukas, Distributed Control Systems, Van Nostrand Reinhold, 1986.)

TABLE 4.14f
*Relative Features of Different Data Highway Cables**

Feature	*Twisted Pair Cable*	*Coaxial Cable*	*Fiber-Optic Cable*
Relative cost of cable	Low	Higher than twisted pair	Multimode fiber cable comparable with twisted pair
Cost of connectors and supporting electronics	Low due to standardization	Low due to CATV standardization	Relatively high—offset by high performance
Noise immunity	Good if external shield used	Very good	Excellent—not susceptible to and does not generate electromagnetic interference
Standardization of components	High—with multiple sources	Promoted by CATV influences	Very little standardization or second sourcing
Ease of installation	Simple due to two-wire connection	Can be complicated when rigid cable type is used	Simple because of light weight and small size
Field repair	Requires simple solder repair only	Requires special splice fixture	Requires special skills and fixturing
Network types supported	Primarily ring networks	Either bus or ring networks	Almost solely ring networks
Suitability for rugged environments	Good, with reasonable cable construction	Good, but must protect alumium conductor from water or corrosive environment	Excellent—can survive high temperatures and other extreme environments

*From M. P. Lukas, *Distributed Control Systems,* Van Nostrand Reinhold, 1986. CATV: cable TV.

several distributed system components. These components can be located either in rooms adjacent to the control room or out in the field.

These distributed control units, which can be remotely located controllers, intelligent fieldbus devices, or remote I/O modules, can in some cases also provide a limited amount of display capability (low-level operator's interface, LLOI). An example of this is a local panel connected to a serial device using a MODBUS protocol.

A specific DCS for a particular plant is configured from standard building blocks marketed by most DCS suppliers. Figure 4.14b illustrates the categories of components that are available to configure various DCS systems. These components include the operator consoles in the central control room, controllers, I/O cards, communications components, and serial cards serving the interconnections with other digital systems such as PLCs and supervisory computers. The components also include bus cards interfacing to Fieldbus, Profibus, DeviceNet, AS-Interface, and other buses.

The process I/O signals are connected to I/O cards, which can fail. For all control loops that should continue functioning if the central processor or the data highway fails, the I/O should be directly connected to the location where the control is executing.

The preferred approach of control system layout is to keep all the I/O and all the associated controllers for a particular unit operation of the process (chemical reactor, distillation column, etc.) in the same physical controller. If this approach is implemented, the process will remain under control as long as the controller is functioning.

In critical applications, the controller and I/O modules can be made redundant.

Operator Console

The viewing applications, which may run on one or more operator workstations, receive data from the controller application via the control network and display this data to process control engineers, operators, or users of user interfaces, and may provide any of a number of different views, such as an operator's view, an engineer's view, and a technician's view.

Operator display applications are typically implemented on a systemwide basis in one or more of the workstations and provide preconfigured displays to the operator or maintenance people regarding the operating state of the control system or the devices within the plant. Alarm displays receive alarm signals generated by controllers or other devices within the process plant, control displays indicating the operating state of the controllers and other devices within the process plant, maintenance displays indicating the operating state of the devices within the process plant, etc.

These displays are generally preconfigured to display, in known manners, information or data received from the control modules or the devices within the process plant. Often displays are created through the use of objects that have a graphic that is associated with a physical or logical element and that is tied to the physical or logical element to receive data about the element. The object may animate the graphic on the display screen based on the received data to illustrate, for example, that a tank is half full or to illustrate the flow measured by a flow sensor.

Although the information needed for the displays is sent from the devices or configuration database within the process plant, that information is used only to provide a display to the user containing that information. As a result, all information and programming that is used to generate alarms, detect problems within the plant, etc., must be generated by and configured within the different devices associated with the plant, such as controllers and field devices, during configuration of the process plant control system.

Although error detection and other programming are useful for detecting conditions, errors, and alarms associated with control loops running on different controllers and problems within the individual devices, it is difficult to program the process plant to recognize system-level conditions or errors that must be detected by analyzing data from different devices within the process plant.

New control systems provide various levels of support for alarm management. In some cases control logic must be built into the control strategies. In other cases smart objects or agents are configured on top of the control strategy to detect abnormal conditions.

Core Architectural Components

The core architectural components of the DCS are:

- System configuration
- Communications
- Control
- Alarms and events
- Diagnostics
- Redundancy
- Historical data
- Security
- Integration

System Configuration Like any computer, distributed control equipment must be told what to do. Programming the process control system instructions is called configuring. There are several aspects to the configuration—the physical configuration and the control strategy configuration. These two activities are generally run in parallel and brought together as the project is engineered.

The configuration database enables users to create and modify control strategies and download these strategies via the control network to distributed controllers, consoles, and devices. Typically, the control strategies are made up of interconnected function blocks, sequential function charts (SFC), and equipment and units representations, which perform functions within the control scheme based on inputs and which provide outputs to other function blocks and/or I/O within the control scheme.

The configuration application also allows a designer to create or change operator interfaces, which are used by a viewing application to display data to an operator and to enable the operator to change settings within the process control system. Each controller and, in some cases, field devices too, stores and executes controller applications that run the control modules assigned and downloaded to implement actual process control functionality. The general configuration items are illustrated in Figure 4.14g.

For regulated and highly critical applications, such as those requiring Food and Drug Administration (FDA) certification, a record can be kept of any changes that are made to the control system configuration. Such an “audit trail” records all changes that were made, the names of people who made the changes, and the time and date when the changes were made. Provisions are also provided to automatically or manually undo any changes that are made in the control system.

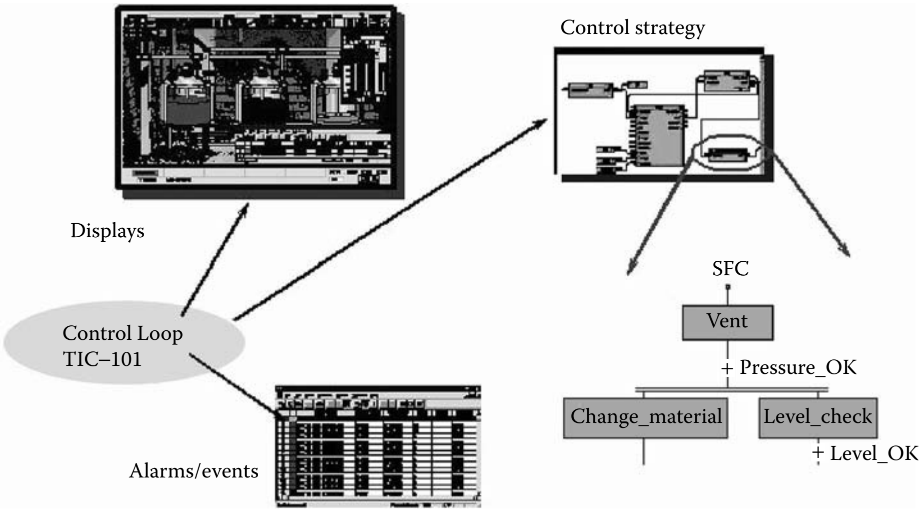


FIG. 4.14g
DCS system function configuration, showing the process display and a listing of alarms and other key events.

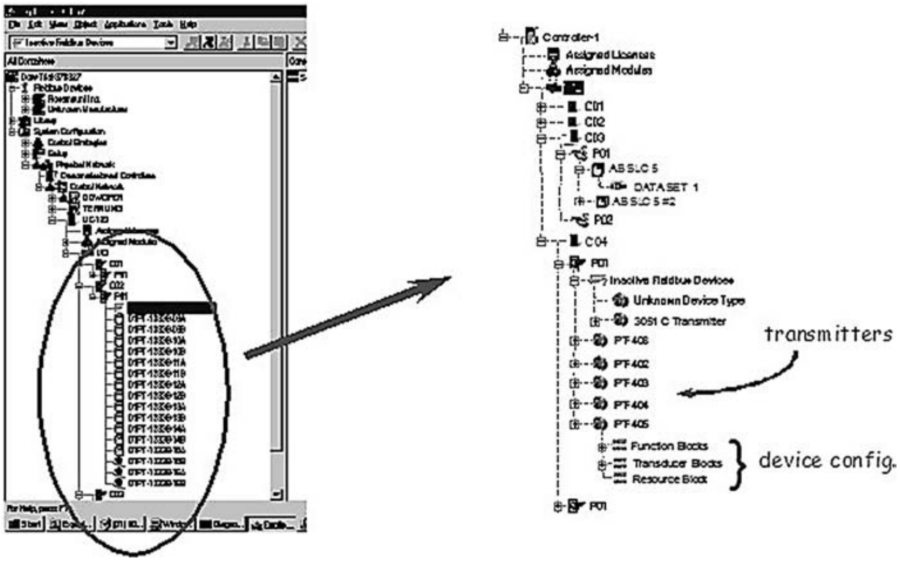


FIG. 4.14h
Physical configuration of nodes, cards, and other devices in the DCS system.

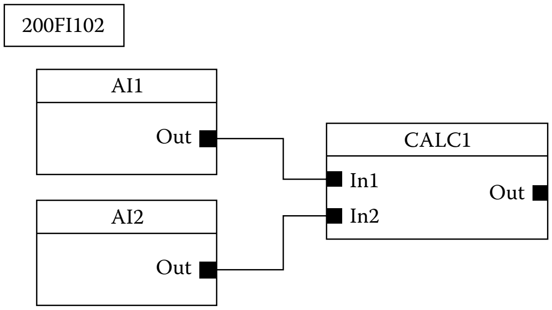


FIG. 4.14i
The configuration of a flow indicator module, with a tag number of 200FI102, is shown here, where two inputs (AI1 and AI2) are sent to calculation CALC1, which generates an output .

Physical Configuration The physical configuration requires configuring the nodes, cards, and devices. In many systems this activity is greatly simplified using auto-sense capabilities. The physical configuration of part of the system is illustrated in Figure 4.14h.

Logical or Control Strategies A distributed control system must have a consistent means of representing and referencing information. Ideally, such reference can be made independent of the physical device that holds this information. A common way to divide data within the control system is according to identifier tag numbers.

The S88 batch standard defines such logical grouping of measurement, calculation, or control as a module. When a control system follows this convention, then each module is assigned a tag that is unique within the control system. Based on this tag and the structure of the components in the module, it is possible for applications in the control system to reference any piece of information. For example, consider the module of an instrument with the tag number 200FI102 (which represents flow indicator No. 102 in area 200 of the plant). In the module a calculation identified as CALC1 is made using inputs AI1 and AI2 and generating an output as shown in Figure 4.14i.

Based on the module tag number and unique function block names within the module, the output of the calculation block is identified as 200FI102/CALC1/OUT.

DCS Systems support multiple control languages. The control languages include function block diagrams, sequential function charts, and structured text and may also include ladder diagrams and instruction lists. Some systems may be IEC 61131-3 compliant. Most control systems also include interlocking and batch capability — in most cases supporting S88. Some systems also support embedded advanced control and safety functions. The control strategies can often be mixed. Strategies can reference I/O as well as local and remote parameters.

An important feature of DCS systems is their ability under certain conditions to be upgraded online. In case of failure, most systems have extensive support for holding the last value, using a default value, or moving to some known state.

One of the features that make distributed control systems so powerful is their function library, which is available and can be used just by calling for it. This availability simplifies the task of the process control engineer, if he or she is familiar with the particular vendor's practices. Such a function library list of one supplier is shown in Table 4.14j.

What distinguishes some DCS suppliers from most PLC manufacturers and what distinguishes the various DCS suppliers from each other is the size and quality of the algorithm library that is embedded and freely available with their basic

TABLE 4.14j
Distributed Control Functions That Can Be Entered by Configuration

1. Highway definition — Assign names and addresses to the workstations and controllers that make up the DCS.
2. Configure reusable configuration components and store them in a configuration library.
3. Configure system level items such as enumeration sets, engineering units, and alarm priorities.
4. Configure loops, equipment, units, process cells, and areas. Make use of library as much as possible.
5. Bind inputs and outputs in control strategy to actual IO and devices in physical hierarchy.
6. Configure additional alarming.
7. Assign control strategies to controllers.
8. Configure historical values.
9. Download configuration into controllers, workstations, IO cards, and devices.
10. Tune parameters (gain, reset, sensitivity, ratio, etc.) and limits (alarm limits, output rate, etc.) for each control loop.
11. Check out sequences.
12. Check out first level of control strategy, displays, alarming.
13. Run water batches checking out critical loops. If this is a batch system, begin checking out phases.
14. Check out first level of control strategy, displays, alarming.

packages. When it comes to implementing some of the more advanced control strategies, it makes all the difference whether the algorithm library includes the algorithms that the particular project requires.

Some of these are analog inputs, sample and hold for dead time processes, dead band control for fast processes, set-point filtering, error squared, integral squared, dead time PID, set-point and process variable characterizers for nonlinear processes, decoupling of interactions, linear dynamic compensation of lead/lag blocks for feedforward, external feedback for antireset windup, self-tuning, and nonlinear adaptive controls, not to mention the more demanding algorithms for optimization, statistical process control, fuzzy logic, model-based optimization, sampled data control, sliding mode control, neural networks, and state-space controls.

Configuration of the control strategy often makes use of libraries of prebuilt control logic. The prebuilt control logic can be linked into final control strategies, in which case changes to the library can be automatically propagated to each control item. Alternatively, the prebuilt control logic can be embedded or unlinked so that individual control strategies are unaffected by changes in the library.

Batch and larger continuous projects, following the suggestions of S88, define strategies as a set of class-based items that can come together as a complete class-based strategy. If the required elements are offered by the DCS supplier, creating a control strategy from a class-based library can result in a significant set of configurations. Binding these configurations to actual I/O, loops, and equipment can become a “fill in the table” exercise.

Input/Output The requirements for redundancy and interfacing for I/O processing dictate that the process controller be of custom hardware design. Multiple processors are often used to address the communication and I/O processing and control execution. Also, a real-time operating system for embedded applications is often used to provide deterministic scheduling and execution of control. A large variety of I/O cards are normally provided to address a variety of field measurements and actuators:

Analog Input (isolated) 1 to 5 volt DC, 4 to 20 mA
Analog Output 4 to 20 mA
Isolated RTD input (2, 3, 4 wire) and Thermocouple Input (B, E, J, K, N, R, S, T)
Discrete Input 24 VDC, 120/230 VAC
Discrete Output 24 VDC, 120/230VAC
Pulse Count Input
Pulse Duration Output

Since digital transmitters and actuators that utilize a variety of communication protocols and physical interfaces are available, many manufacturers offer interfaces to the most common buses. Also, serial interface cards are often supported for interfacing to supporting systems. Examples of these communication interface cards are:

HART AI-Card, 4 to 20 mA
HART AO Card, 4 to 20 mA Series
DeviceNet (Baud rate 125 250 500 kbit/sec)
FOUNDATION Fieldbus
AS-Interface
Profibus DP Baud rate (9.6 19.2 93.75 187.5 500 1500 kbit/sec)
Serial Interface (MODBUS or Allen-Bradley’s Data Highway Plus protocol)

In addition, some manufacturers may offer I/O cards to meet special requirements. For example, sequence of events (SOE) input cards are used to capture process-upset events coming directly from devices in the field. Because events are captured and temporarily stored locally—on the SOE input card itself—faster recording for each of the channels on the card is possible. For example, events captured by an SOE input card are time stamped using a 1/4-millisecond resolution.

Input and output terminations are made at terminals that are either part of the electronic mounting frames or on separate terminal boards. In the latter case there will usually be a cable connection between the terminal board and the electronic controller file. Connections are usually made from the front of the cabinet. An alternate method is to use a separate termination cabinet, filled with rows of terminal strips. This requires extra wiring from the termination cabinet over to the terminals in the remote controller cabinet, but it has the advantage that field wiring can be completed before the distributed control housings are delivered and installed.

Conventional I/O Analog input and output signals will usually be carried on shielded, twisted pairs of copper wire. Digital inputs and outputs, either 120-volt AC or 24-volt DC, can be carried on twisted pairs, which do not, however, have to be shielded. Analog signals should never be run in proximity to alternating current wiring. The controller files operate almost universally on 1- to 5-volt signals, so the most common input is a 4- to 20-mA current signal, developing a 1- to 5-volt input across a 250-ohm resistor mounted on the input terminal board. Most distributed control systems can accept low-level signals from RTDs and thermocouples, performing the signal amplification in their input electronic circuitry. A few systems can accept pulse input with frequencies sufficiently high to allow signals from turbine flow meters to be used directly.

Most suppliers offer some conditioning of signals. Providing square root extraction, linearizing thermocouples, and resistance thermometers and dampening noisy inputs can be selected by configuration. Some input/output boards provide terminals with fused 24-volt DC power that can be used to supply a positive voltage to two-wire transmitters.

Separate terminal boards may also be supplied for digital input and output signals. Usually, optical isolation is provided. A DC input signal (or a rectified AC input signal) causes a light emitting diode (LED) in the isolating relay to be energized. A photoelectric device energized from the LED actuates a transistor in transistor–transistor logic (TTL) input circuitry to signal a digital input. A digital output signal is similarly isolated to actuate a transistor driver circuit for DC outputs or a triac for AC outputs. The solid-state relay from which the output is generated functions like a dry contact, and the output must be powered from a separate power source.

Diagnostics Integrated diagnostics is an important feature of the DCS. The diagnostics cover the hardware, redundancy, communications, control, and to some extent, the software that makes up the DCS.

Redundancy Redundancy is an important requirement for any critical process control application using DSC systems. These systems must have redundant communications, redundant controllers, redundant I/O cards, and redundant I/O communications. It is also possible to take redundant or preferably two out of three voting measurements and discard the defective or inaccurate one during control execution.

One advantage of redundancy is the ability to upgrade components online in the control system, but on critical processes this has to be very carefully planned so that safety will not be compromised.

Historical Data The DCS usually includes the ability to collect batch, continuous, and event data. A centrally defined history database is available for the storage of historical data. The value of any attribute, alarm or any control strategy, alert, or process condition can be recorded in the history database along with its status. In modern control systems the data values are collected as an integrated feature of the system. Events are collected and time-stamped at their source—in some cases down to a few millisecond resolution.

Users and layered applications can retrieve the batch, continuous, and event data in a time-ordered fashion. For security reasons values cannot be edited without leaving behind an audit trail.

Security Security is essential in process control. The DCS system must be able to limit the access to the various parts of the control system to authorized people only. This is done by user, plant area, and workstation. Layered applications have to form a session before they are allowed access into the system. There are several aspects to security as summarized below:

- *Authentication.* Access to the DCS for human users and layered applications users will be controlled through the use of password-protected user accounts.
- *User.* A human user of the DCS must have a user account on the system in order to gain access. All user accounts are named. User accounts have unique names within the scope of a site. All user accounts have a password, which must be provided in conjunction with the account name in order to start a DCS session.
- *Plant area security.* A user account can be permitted or denied access to make changes within zero or more plant areas within a site.

For each plant area where access is permitted, access can be restricted at runtime according to the classification of the runtime attribute data. For each plant area where access is permitted, the ability to make configuration changes can be restricted.

A user account can be permitted or denied access to view or modify user account and privilege information.

In some systems it is also possible to enable authorization. In these cases a user, or in some cases several users, will need to confirm by password the changing of certain parameters, starting/stopping a batch, etc.

Integration When a new plant area is added or expanded, the operators of the new area may need some information about the existing plant to provide a coordinated operation. Similarly, the operators of the existing plant may need to have feedback from the new process area in making decisions on how best to run the balance of the plant. In most cases, only a small fraction of the information in either system must be communicated to support such coordination between these areas. Several techniques are used to integrate systems.

The OPC Foundation has defined an industry standard for accessing information within a control system. Thus, many control systems provide OPC server capability in workstations designed for interfacing to the plant local area network (LAN).

OPC client applications in this station or on the network may access information using the path convention supported by the control system.

International Fieldbus Standards

The adoption of the IEC1158-2 Fieldbus standard by the major DCS manufacturers has ushered in the next generation of control and automation products and systems. Based on this standard, fieldbus capability may be integrated into a DCS system to provide:

- Advanced function added to field instruments
- Expanded view for the operator
- Reduced wiring and installation costs
- Reduced I/O equipment by one half or more
- Increased information flow to enable automation of engineering, maintenance, and support functions

Similarly, the Ethernet and OPC industry standards have provided DCS manufacturers a defined means for other applications to access information in a process control system. Through the active use of control system data in a plant information system, the operation benefits that may be achieved are:

- Improvement in production scheduling
- Better inventory control
- Consolidation of maintenance and operation information from multiple sites

DATA HIGHWAY DESIGNS

DCS systems today range in size from a single stand-alone laptop for the purpose of plant design and simulation to full-scale systems covering a whole plant. The systems come complete with integrated Web services for plant integration — often supporting a variety of open standards, such as OPC, for communicating with outside data sources.

Control Network

The communication infrastructure on the area control network supports the following:

- *Connections.* Connection between nodes in the system.
- *Unsolicited communications.* Transferring real-time information as data changes.
- *Synchronous and asynchronous read/writes.* Reads and writes block/don't block during the transaction.
- *Passthrough messages.* Transfer of hosted messages across control network to device, serial, or other network.
- *Configuration downloads.* Transfer of configuration from engineering node to controller and devices.
- *Auto-sensing.* Automatic detection of controllers, workstations, I/O cards, devices.
- *Diagnostics.* Diagnosis of system components, control strategies, etc.
- *Debugging.* Debugging of control strategies.
- *Directory services.* Location services to find nodes, cards, control strategies, devices, and other items in the system.
- *Online upgrades.* Upgrading a system that is in operation.
- *Hot/warm /cold restart.* Restarting a control strategy from backup.
- *Secure and unsecured access.* Security required to access information in the system.
- *Alarms and events.* Alarms and events are generated by the control strategy and system.
- *Device alerts.* Device alerts are generated by devices and equipment in the control system.
- *Time synchronization.* Time synchronized across nodes, devices, and sometimes I/O (e.g., sequence of events recording).

The data highway is the communication device that allows a distributed control system to live up to its name, permitting distribution of the controlling function throughout a large plant area. Data highways vary in length as a function of traffic capability and speed of transmission.

The most popular medium is Ethernet CAT5 cable. Several suppliers still offer twisted and shielded coax cables. In situations where noise is extensive, either fiber optics or wireless may be used—with fiberglass cable being the most prevalent. This is used most commonly for point-to-point connection between switches and hubs. Fiber optics is attractive for use as a data highway because it eliminates problems of electromagnetic and radio frequency interference, ground loops, and common mode voltages; it is safe in explosive or flammable environments; it can carry more information than copper conductors; it is inert to most chemicals; and it is lighter and easier to handle than coaxial cable. For more details on fiber-optic transmission refer to Section 3.7.

Ethernet Configuration

Hubs and switches can be used for 10 or 100 Mbit and even 1 Gbit per second networks. With Ethernet switches, each port auto-senses the speed of the attached workstation or controller and then operates at that speed. Standard DCS architectures operate at 100 Mbit per second speed.

Figure 4.14k illustrates a Class II repeater network where the maximum distance between workstation A and workstation B (the furthest points) can be up to 672.6 feet (205 meters) when up to two of the dual speed hubs are daisy-chained with twisted pair link segments. Each port on the switch can support up to 328 feet (100 meters) of twisted pair cable; however, only 672.6 feet (205 meters) is allowed end to end. This leaves 16.4 feet (5 meters) for the twisted pair link segment between the hubs.

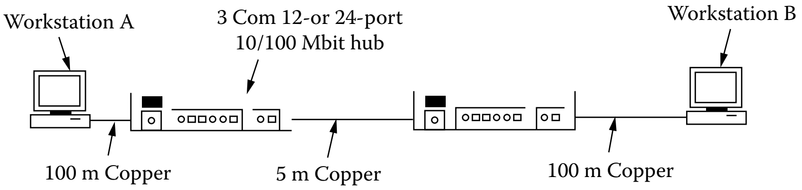


FIG. 4.14k

Ethernet-based control network switch configuration, where the maximum distance between workstations is up to 672.6 feet (205 meters). Each port on the switch can support up to 328 feet (100 meters), and 16.4 feet (5 meters) are allowed for the twisted pair link segment between the hubs.

An alternative is to use shorter-length cables (less than 328 feet [100 meters]) from workstation or controller to hub and allow a longer cable between hubs. Any combination will work as long as the total of all cable lengths does not exceed 672.6 feet (205 meters). It is a good practice to use a 16.4-foot (5-meter) or shorter cable between the hubs to avoid problems later if a new cable needs to be attached to the hub. This is to avoid a situation in which the 672.6-foot (205-meter) maximum length is exceeded because the length of the intra-repeater link is not known and an assumption is made that all ports can have a 328-foot (100-meter) cable.

ALARM MANAGEMENT

A critical part of the DCS is its integrated alarms and events system. The system provides configuration, monitoring, and notification of significant system states, acknowledgments, and priority calculations. Events represent significant changes in state for which some action is potentially required. An active state indicates that the condition that caused the event still exists. The acknowledge state of an event indicates whether an operator has provided acknowledgement that an event has occurred. In most systems event types can also be defined.

The event type specifies the message to be displayed to an operator for the various alarm states and the associated attributes whole value should be captured when an event of this type occurs. Event priorities can also be defined. An event priority type defines the priority of an event for each of its possible states.

Many DCS systems now also support device and equipment alerts. Like process alarms, alerts can have priority assigned to them, can be acknowledged, and convey information related to the condition that caused them. Unlike process alarms, however, these alerts are generated by the DCS hardware or devices and equipment external to the DCS.

Alarms and alerts are presented to the operators in alarm banners and summaries. Using these specialized interfaces, operators can quickly observe and respond to conditions. They typically use these specialized displays to navigate to a display where they can view additional details and take action as appropriate.

Operators can also suppress and filter alarms. Alarm suppression is typically used to temporarily remove alarms from the system for which some condition exists that the operator knows about (e.g., a piece of equipment has been shut down). Alarm filtering provides a way for the operator to view collections of alarms.

The following paragraphs describe alarm processing and alarm management and conclude with a discussion of higher-level applications built on top of the overall alarm processing and management system.

Alert Processing

In 1996 the SP50 committee finally finished its standards, including field-testing, and made the standards available in the form of a commercial product.

As part of SP50 alarms and events, collectively known as alerts, were included as part of the function block efforts of Foundation Fieldbus and IEC 61804. In these standards, alarms and alerts represent state changes within function block applications. Resources each have an alert notifier responsible for reporting their alert occurrences. Alerts objects are used to communicate the event to other devices.

An alert notifier examines the results of resource, transducer, and function block executions to determine whether any of a defined set of alert states has been entered. For alarms, both entering and exiting alarm conditions are defined as alert states. When an alert occurrence has been detected, the alert object builds a report message, referred to as an event notification, and publishes it onto the network. The time at which the alert state was detected is included as a time stamp in the alert message. The reporting of alerts may be individually suppressed.

A reply is required that confirms receipt of the notification. If the reply is not received within a time-out period, the alert may be retransmitted.

Alerts may also be acknowledged. Acknowledgment indicates that the alert has been processed by an interface device to satisfy operational interface requirements.

An alert notifier examines the results of function block executions to determine whether any of a defined set of alert states has been entered. When an alert occurrence has been detected, the alert notifier builds a report message, referred to as an event notification, and publishes it onto the network through the alert object.

Based on the type of alarm and event information, which may be reported by blocks contained in a resource, up to three classes of alerts may be defined in the resource:

1. *Analog alert.* Alert used to report alarms or events whose associated value is floating point
2. *Discrete alert.* Alert used to report alarms or events whose associated value is discrete
3. *Update alert.* Alert used to report a change in the static data of the block

A reply is required from one interface device that confirms receipt of the notification. If the reply is not received within a time-out period, the alert notifier will retransmit the notification. This method ensures that alert messages are not lost.

The following alarms are supported by Foundation Fieldbus devices:

0 = Discrete alarm
1 = High high alarm
2 = High alarm
3 = Low low alarm
4 = Low alarm
5 = Deviation high alarm
6 = Deviation low alarm
7 = Block alarm

Associated with each alarm is a time stamp that indicates the time when evaluation of the function block was started and a change in alarm state was detected that is unreported. The time-stamp value will be maintained constant until alert confirmation has been received, even if another change of state occurs. Also, the value of the associated parameter at the time the alert was detected is reported.

A function block must detect the alarm condition. The alarm must be transported to the responsible entity, e.g., interface device supporting human interface. The entity must confirm that the transport was successful. The alarm may require that a plant operator acknowledge that the alarm has been noticed, even if the condition has cleared.

Every occurrence of an alarm must be balanced by a notification that the alarm has cleared or that the same alarm has occurred again before the clear could be reported. An alarm will also be cleared in a device when 1) an alarm that is active is disabled or 2) a block containing an active alarm is placed in out-of-service mode. In these cases, specific alarm-clear messages should be generated, to allow remote alarm summaries to clear the alarm information for this block.

Each alarm and event parameter may have an associated priority parameter. The alert priority enumeration value is:

0 = The associated alert may clear when the priority is changed to 0, but it will never occur.
1 = The associated alert is not sent as a notification. If the priority is above 1, then the alert must be reported.
2 = Reserved for alerts that do not require the attention of a plant operator, e.g., diagnostic and system alerts. Block alarm, error alarm, and update events have a fixed priority of 2.
3 to 7 = Increasing higher priorities; advisory alarms.
8 to 15 = Increasing higher priority; critical alarms.

The alert object allows block alarms and events to be reported to a device responsible for alarm management.

The alert object contains information from an alarm or update event object, which is to be sent in the notification message. The alert object will be invoked by the alert notification task. If multiple alarms or event parameters are unreported, then the one with the highest priority or the oldest of equal priority will be selected by the alert notification task. The selected alert object is sent in a message at the first opportunity—less than the alert confirm time. If a confirmation from an interface device is not received by the alarm notification routine in the field device within a time determined by the resource block confirm time parameter, then the alert will be considered unreported so it may be considered for selection.

Figure 4.14l illustrates the transport of an alert.

Alarm Management Operator consoles maintain a list of active alarms in the system:

- Consoles register an interest in alarms/events (by plant area) in all other nodes in the system.
- The alarm state change events are routed to the software that maintains the active alarm list.
- Active alarms may be regenerated in order to build the active alarm list in workstations starting up, or when additional plant areas are required.
- In the background, alarms are resynchronized to keep the active alarm lists accurate (in particular, remove “dead” active alarm list items for alarm parameters/modules/nodes that are no longer out there and communicating (and did not get a message out before they went away).

Workstations select which alarms are eligible for inclusion in the workstation alarm list:

- Must be in workstation’s “alarm scope” (set of plant areas).
- Must be in current user’s “alarm scope” (set of plant areas that user has one or more security keys for).
- Area level alarm filtering (per workstation) must be set to enable alarms from that area.

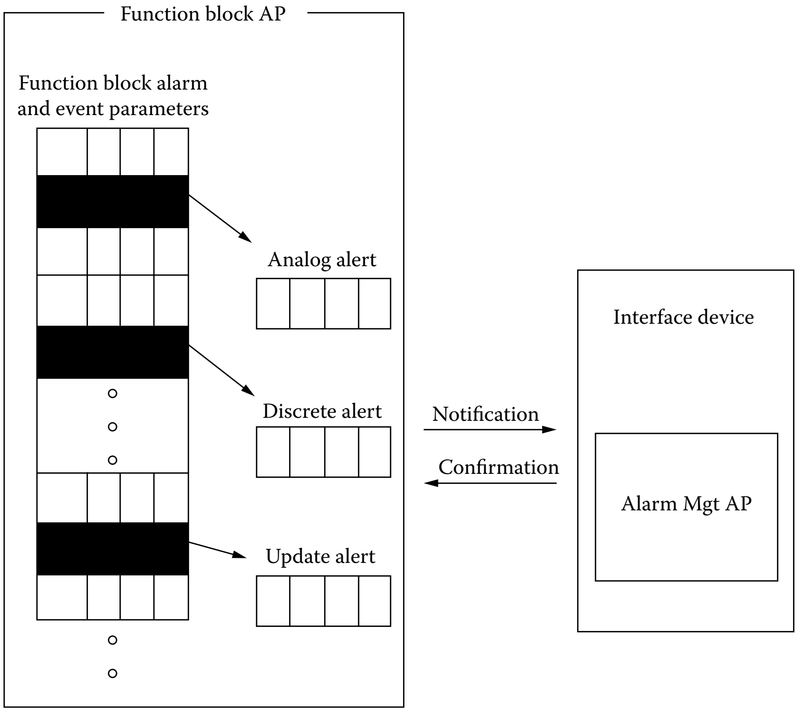


FIG. 4.14l
Alert event notification and confirmation.

- Unit level alarm filtering (per workstation) must be set so if the alarm is associated with a unit, that unit does not have the alarm turned off.

The workstation alarm list maintains the list in order of importance:

- Unack ahead of acked
  - Then: Higher priority ahead of lower priority
    - —Then: Condition still active ahead of condition cleared (latched)
      - —Then: more recent "went active" time ahead of older alarms

When alarms and alerts arrive at the operator console they are first classified. Alarms and alerts can be classified into process alarms, device/equipment alarms, and hardware alarms.

Process alarms cover "traditional" process alarms: HI, LO, HI-HI, LO-LO, DEV, etc. They are highly configurable:

- User-configurable alarm (parameter) names
- User-configurable alarm types (alarm words, category, message content)
- User-configurable alarm condition (with arbitrarily complex logic to arrive at the alarm condition)
- User-configurable priority
- Unlimited number per containing module

Device alarms are from instruments, transmitters, valves, and equipment attached to the DCS. Relatively little configuration is needed:

- Limited/fixed number of distinct alarms:
  - FAILED
  - MAINT
  - ADVISE
- Alarm conditions have "fixed" mapping to the alarms.
- User-configurable priority, and whether or not alarm should be enabled.

Hardware alarms are triggered by the hardware components of the DCS (controllers, I/O cards, remote I/O communication links, redundant hardware, etc.). Relatively little configuration is required.

Operators usually interact with the alarm system through alarm banners and alarm summary displays.

DCS ATTRIBUTES

The DCS has a number of advantages over other control approaches. If the client and design engineering firm are highly experienced, distributed control can reduce overall installation, configuration, and commissioning costs. Less wiring is required when information is transmitted over bus networks.

From the point of view of the operator, the interface with the process is improved. Integrated alarming and diagnostics substantially improve the operator's ability to identify the causes of upsets and to bring the process back under control.

Configuration capabilities allow for prebuilding portions of the control system configuration, which can be reused on other projects. There is great potential for extensive standardization of control systems.

The distributed control system is flexible and relatively easy to expand.

Reliability

Digital computers are more reliable today than when they were first introduced, but the possibility of failure of a single piece of electronic equipment causing the shutdown of an entire production facility still raises concerns that cannot be ignored. How does distributed control satisfy the requirement for continuous production?

Most suppliers subject their equipment to extensive periods of cycling at temperatures exceeding the extremes listed in equipment specifications. This weeds out the components most likely to fail. Failures of marginally operational parts will usually show up in the first few weeks of operation. Once past that period, electronic equipment seems to operate indefinitely, so long as limits of temperature and atmospheric cleanliness are observed. The advent of large-scale integration (LSI) has fostered this reliability. As size has been reduced, so has heat generation.

Nevertheless, failures will inevitably occur. Consequently, suppliers provide redundancy in their design, as well as backup. Some suppliers simply build two of everything and supply it as standard. Others offer redundancy on an optional basis. Power supplies, data highways, traffic directors, and remote controller electronics are important links in the communications chain and should be considered as candidates for redundancy. The operator station itself, with its video terminal, will not shut the system down if it fails, but it will leave the operator blind to the condition of the process, and so it is another candidate for redundancy.

There should be automatic transfer between redundant parts, so that if one fails the other takes over with no disturbance of the operation or output. At the same time, there must be some sort of alarm to alert the operator to the fact that a failure has occurred. How much redundancy is provided must depend on how much loss of production will cost. If continuous production is absolutely necessary, no expense cutting is justified.

Another form of redundancy, available from most suppliers, is controller backup. A complete modular file is mounted in the same remote location as others that are considered important enough to require backup protection. Some suppliers back up one complete set of electronics with another complete set, updating the database of the backup unit as that of the primary is changed. The two are cabled together, and on failure of the primary, the secondary takes over automatically.

Another supplier backs up eight controller files with a single backup file, and if any of the eight fails, its place will be taken by the backup unit. Still another supplier, in addition to supplying one-for-one backup, allows portions of the backup file to back up portions of the primary controller file, freeing the unused portions of the backup file for additional control tasks.

Mean Time between Failure

High availability is as important as reliability. Defining availability as the ratio of mean time between failure (MTBF) to mean time between failure plus mean time to repair (MTBF + MTTR), it is clear that a system will be most available when it is very reliable (high MTBF) and can be quickly repaired (low MTTR). Since distributed control equipment is highly modular and contains many printed circuit cards, time to repair can be very short if sufficient spare parts are available.

Most systems make good use of diagnostics; internal failures are reported on the CRT screen, indicating where in the system a failure has been detected and providing some clue to the cause. Substitution of printed circuit cards can often restore operation, and this can be done quickly by a service technician who knows where to look for the trouble and has spare parts for making the substitution. The failed card can then be returned to the supplier for replacement.

High-availability process control systems are essential in case of large continuous processes such as a cracking unit in a petroleum refinery. Such units can be designed to run with no shutdown for 5 to 6 years while running 24 hours a day 7 days a week. Some parts of the system can be made redundant so that automatic switchover might occur in case of a failure. Other parts of the system can be provided with automatic fault diagnosis and identification for rapid repair.

For such applications obviously one would hope and aim to design a system that allows the failed unit to be replaced while the process continues without interruption, but in the real world Murphy's Law often comes true.

The high diagnostic coverage and rapid replacement capabilities of the controller and I/O assemblies allow the user to decide how much redundancy is really needed for an application. A customer may choose to use simplex I/O and possibly even simplex controllers, where failure events are not significant as long as the failures are automatically diagnosed and reported and the repair steps are simple.

To address this, built-in diagnostics that identify the majority of assembly failures are provided by some suppliers. In addition, all of these assemblies are "hot pluggable" (a "computerize" buzzword for testing while in operation) so that an I/O assembly can be diagnosed and repaired efficiently. Some controllers are also able to check whether a new card is the proper type when replacement of a failed card is needed.

TABLE 4.14m
Sample Price List for DCS Components

	Description	*Price Range ($US)*	*Typical ($US)*
1.0	HARDWARE W/SOFTWARE LICENSE		
1.1	Control unit, w/carrier, power supply	5300 to 6600	5500
1.2	Redundant control units	11,700 to 14,300	13,000
1.2	Operator's console w/21" monitor (1200 I/O)	9,800 to 10,300	10,000
1.3	Engineer's station w/21" monitor (1200 I/O)	20,000 to 22,000	21,000
1.4	Analog input (AI) card, 8 channel w/HART	1400 to 1800	1500
1.5	Analog input (AI) card, 8 channel w/HART redundant	2500 to 3000	2800
1.6	Analog output, 4 to 20 mA, 8 channel w/HART	1700 to 1900	1800
1.7	Analog output, 4 to 20 mA, 8 channel w/HART, redundant	2700 to 3300	3000
1.7	Foundation Fieldbus interface card w/2 segments	2200 to 2900	2500
1.8	Discrete input, 24 Vdc, 8 channel	520 to 630	550
1.9	Discrete output, 24 Vdc, 8 channel	1000 to 1150	1100
1.10	Discrete input, 24 Vdc, 8 channel as fieldbus I/O	820 to 930	850
1.11	Discrete output, 24 Vdc, 8 channel as fieldbus I/O	1000 to 1150	1000
1.12	Discrete input, 24 Vdc, dry contact, 8 channel, redundant	1000 to 1200	1100
1.13	Discrete output, 24 Vdc, hi-side, 8 channel, redundant	1500 to 1700	1600
1.14	Carrier for 8 cards	600 to 800	700
1.15	Bulk power supply	900 to 1200	1000

Some manufacturers also have extensive diagnostics on their bus cards such as Fieldbus Foundation H1 Fieldbus IO cards and HART analog input and analog output cards. Some conventional cards, such as discrete input and output cards, also support automatic short-circuit and open-circuit detection on discrete sensors and actuators.

PRICING

Utilizing PCs and standard monitors and keyboards for operator and engineering stations and the use of standard Ethernet for communication have reduced the cost of DCS system hardware. Also, the introduction of fieldbus devices has dramatically impacted the cost and space required for controller I/O.

In the cost information provided here, the cost of services associated with system configuration, factory acceptance, installation, simulation, testing, commissioning and startup are not included. Similarly excluded is any software not embedded in the DCS supplier's standard package. This usually excludes control and simulation or modeling algorithms that are unique to the particular process, interfacing with software or hardware of other suppliers that the DCS vendor does not support with embedded software, and the configuration of most graphics that are specific to the particular project.

A sample price list for DCS hardware components is shown in Table 4.14m. This price list basically covers the cost of DCS control system hardware. Most of the software costs are not included in the list.

As an example, Table 4.14m lists the cost of the DCS hardware (and the licenses for some basic software) for a project that requires 256 analog and 961 on/off I/O as follows:

- 25% Fieldbus Analog (64 I/O—12 fieldbus devices per segment)
- 75% Non-Fieldbus I/O (192 I/O, redundancy for 5% of analog I/O)
- 10% Fieldbus Discrete (96 I/O—8 discrete fieldbus devices)
- 90% Non-Fieldbus Discrete (865 I/O, redundancy for 10% of discrete I/O)

Based on the above scope definition, the cost of the package has been estimated using typical list prices for the components, as shown in Table 4.14n.

After the hardware components of the DCS have been manufactured, the system must be integrated and made fully operational before it can be tested. If the user wants to have a complete estimate for the whole DCS project, including factory acceptance test and startup, it is advisable to include *all* the software development, integration, and training tasks in the initial specifications for the DCS bid package, when it is sent out for competitive bidding. Following system integration, a factory acceptance test (FAT) is normally conducted.

TABLE 4.14n
Cost Estimate for the System Described in the Text

	Description	*Quantity*	*Unit Price ($US)*	*Total ($US)*
1.	Redundant control units	3	13,000	39,000
2.	Operator's console	2	10,000	20,000
3.	Engineer's station	1	21,000	21,000
4.	Analog input (AI) card	12	1,500	18,000
5.	Analog input (AI) card, redundant	1	2,800	2,800
6.	Analog output, 4 to 20 mA	12	1,800	21,600
7.	Analog output, 4 to 20 mA, redundant	1	3,000	3,000
8.	Fieldbus interface	3	2,500	7,500
9.	Discrete input, 24 Vdc	65	550	35,750
10.	Discrete output, 24 Vdc	34	1,100	37,400
11.	Discrete input, 24 Vdc, fieldbus I/O	8	850	6,800
12.	Discrete output, 24 Vdc, fieldbus I/O	5	1,000	5,000
13.	Discrete input, 24 Vdc, redundant	8	1,100	8,800
14.	Discrete output, 24 Vdc, redundant	4	1,600	6,400
15.	Carrier for 8 cards	20	700	14,000
16.	Bulk power supply	1	1,000	1,000
				Total: 248,050

If an engineering firm or system integrator performs some of the software development, simulation, training, and commissioning tasks, the initial bid package should clearly define the areas of responsibilities that are to be met by the DCS supplier.

Bibliography

Allen, B. S., "Data Highway Links Control Equipment of Any Number of Different Manufacturers," *Control Engineering,* July 1981.

Allen, R., "Local Networks," *Electronic Design,* April 16, 1981.

Archibald, W., "Remote Multiplexing," Burr-Brown Application Note AN-80, January 1976.

Aronson, R. L., "Line-Sharing Systems for Plant Monitoring and Control," *Control Engineering,* January 1971.

Atif, Y., "System Software Support for Distributed Real-Time Systems," *Journal of Systems and Software,* Vol. 53(3), pp. 245–264, 2000.

Balph, T., Schreyer, A., and Tonn, J., "The Impact of Semiconductors on Industrial and Process Control," *Instruments and Control Systems,* September 1981.

Barnes, G. F., "Single Loop Microprocessor Controllers," *Instrumentation Technology,* December 1977.

Bibbero, R. J., *Microprocessors in Instruments and Control,* New York: John Wiley & Sons, 1977.

Brook, R. C., "Use of Microprocessors for Process Control," *Food Technology,* October 1981.

Buckley, P. S., "Distillation Column Design Using Multivariable Control," *Instrumentation Technology,* September and October 1978.

Campelo, J. C., Rodrguez, F., Rubio, A., Ors, R., Gil, P. J., Lemus, L., Busquets, J. V., Albaladejo, J., and Serrano, J. J., "Distributed Industrial Control Systems: A Fault-Tolerant Architecture," *Microprocessors and Microsystems,* Vol. 23(2), pp. 103–112, 1999.

Carlo-Stella, G., "Distributed Control for Batch Systems," *InTech,* March 1982.

Crutchley, W., "Software—A Guide Through the Maze," *Instruments and Control Systems,* February 1979.

Davis, H., "Introduction: Pascal," *Instruments and Control Systems,* June 1979.

Dobrowolski, M., "Guide to Selecting Distributed Control System," *InTech,* June 1981.

Emigh, J., "Struggle Heats Up Over Fieldbus Standardization," *InTech,* July 1994.

Fadum, O., "Plantwide Automation," *InTech,* September 1983.

Fan, C. K., and Wong, T. N. "Agent-Based Architecture For Manufacturing System Control," *Integrated Manufacturing Systems,* Vol. 14(7), pp. 599–609, 2003.

Garrett, L. T., and McHenry, J. M., "Analyzing Costs of Digital and Analog Control Systems," *Hydrocarbon Processing,* December 1981.

Gianamore, D., "UNIX Bridges the DCS Interface," *InTech,* June 1993.

Glanzer, D. A., "Take Process Control to a New Level," *Chemical Engineering Progress,* Vol. 94(10), 1998.

Gooze, M., and Nelson, G., "What Do You Have to Know About Serial Data Communication?" *InTech,* June 1981.

Griem, P. D., "Security Functions in Distributed Control," *InTech,* March 1983.

Groves, B., "Microprocessors," *Instruments and Control Systems,* March 1975.

"Guide to Selecting Temperature Multiplexers," *Instrumentation Technology,* February 1980.

Hackmeister, D., "Focus on Printers: The Application Determines the Type and Technology to Choose," *Electronic Design,* June 21, 1979.

Herb, S. M., and Moore, J. A., "Understanding Distributed Process Control," ISA, 1987.

Higham, J., Kendall, B., and Gerdts, M., "The Data Freeway Network: A Coaxial Multidrop Management Control Bus," *Control Engineering,* September 1981.

Hollister, A. L., "A Primer on Microprocessor Software," *Instrumentation Technology,* October 1977.

Hu, S. C., "Microprocessors—Characteristics and Role in Process Control," 1976 ISA Conference, Preprint No. 76-552.

Ipock, I., and Carroll, G., "DCS or PLC: A User Decides," *InTech,* January 1991.

Kaiser, V. A., "Changing Patterns of Computer Control," *Instrumentation Technology,* February 1975.

Kawai, K., Takizawa, Y., and Watanabe, S., "Advanced Automation for Power-Generation Plants—Past, Present and Future," *Control Engineering Practice,* Vol. 7(11), pp. 1405–1411, 1999.

Kovalcik, E. J., "Understanding Small Computers," *Instruments and Control Systems,* January 1976.

Kwok, T., "Advanced Workstations as DCS Operator Consoles," 1992 ISA Conference, Houston, October 1992.

Lagana, T., "Digital Control System Interfaces: The Same but Different," *InTech,* November 1981.

Larsen, G. R., "A Distributed Programmable Controller for Batch Control," *InTech,* March 1983.

Larson, K., "The Great DCS versus PLC Debate," *Control,* January, February, and March 1993.

Laspe, C. G., "Personal Microcomputers," *InTech,* November 1982.

Losavio, F. and Matteo, A., "Multiagent Models for Designing Object-Oriented Distributed Systems," *Journal of Object Oriented Programming,* Vol. 13(3), p. 8, 2000.

Lukas, M. P., *Distributed Control Systems,* New York: Van Nostrand Reinhold Co., 1986.

Lynn, R. L., "Guidelines for Specifying and Procuring Distributed Control Systems," *InTech,* September 1980.

Maczka, W. S., "DCS Users Want It All," *InTech,* April 1993.

Mahalik, N. G., and Lee, S. K., "Client Server-Based Distributed Architecture for Concurrent Design of DCS Networks: A Case Study," *Integrated Manufacturing Systems,* Vol. 13(1), pp. 47–57, 2002.

Merritt, R., "Large Control Systems," *Instruments and Control Systems,* December 1981.

Merritt, R., "Distributed Controls: A Technology Update," *Instruments and Control Systems,* September 1983.

Myron, T. J., "Digital Technology in Process Control," *Computer Design,* November 1981.

Ogdin, C. A., "The Highs and Lows of Microcomputer Programming Languages—Part I," *Instruments and Control Systems,* May 1978.

Ogdin, C. A., "The Highs and Lows of Microcomputer Programming Languages—Part II," *Instruments and Control Systems,* June 1978.

Park, J., Reveliotis, S. A., Bodner, D. A., and McGinnis, L. F., "A Distributed, Event-Driven Control Architecture for Flexibly Automated Manufacturing Systems," *International Journal of Computer Integrated Manufacturing,* Vol. 15(2), pp. 109–126, 2002.

Phinney, T., "Fieldbus—The Bottom-Up Approach," *Control Engineering,* March 1991.

Pracht, C. P., "A Distributed Microprocessor-Based Control System," 1976 ISA Conference, Preprint No. 76-515.

Prickett, M. J., "Microprocessors for Control and Instrumentation," *InTech,* March 1983.

Ritz, G., "Keys to Integrating PLC and DCS Architectures," *Control,* March 1994.

Ritz, G., "DCS Looks to Satisfy Users' Cravings," *Control,* April 1994.

Shaw, W. T., "Using Distributed Control Systems for Process Simulation," *Instruments and Control Systems,* December 1981.

Shepherd, B., "Watch Those Other Factors in Your Bus System Design," *Instruments and Control Systems,* June 1983.

Sheridan, T. B., "Interface Needs for Coming Industrial Controls," *InTech,* February 1979.

Stockdale, R., "Smart Transmitter Users Speak Out for Global Standardization," *Control Engineering,* September 1991.

Stoffel, J., "DCS Melds Analog, Digital I/O," *Control Engineering,* October 1991.

Studebaker, P. S., "The State of the Architecture," *Control,* December 1993.

Sulzer, E., and Bertsch, M., "Design and Engineering of a Modern Process Control System," *World Cement,* Vol. 28(10), p. 72, 1997.

Tebbett, G., "Putting the System Together," *Instruments and Control Systems,* May 1983.

Thaler, R. M., "Things to Consider When Selecting a Minicomputer System," *Instruments and Control Systems,* January 1981.

Tom, T. H., "DCS Selection for Flexible Patch Automation," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Troutman, P., "Distributed Digital Process Control at the Control Valve," 1978 ISA Conference, Preprint No. 78-543A.

Uyetani, A., "Multiloop Process Controller," 1976 ISA Conference, Preprint No. 76-823.

Wade, H. L., *Distributed Control Systems Manual,* Applied Digital Research Inc., 1991.

Warren, C., "Disk Drives," *Engineering Design News,* August 19, 1981.

Washburn, J., "Communications Interface Primer," *Instruments and Control Systems,* Part I, March 1978 and Part II, April 1978.

Williams, T. J., "Hierarchical Distributed Control," *InTech,* March 1983.

Yook, J. K., Tilbury, D. M., and Soparkar, N. R., "A Design Methodology for Distributed Control Systems to Optimize Performance in the Presence of time Delays," *International Journal of Control,* Vol. 74(1), pp. 58–76, 2001.

Zama, E., Chaillet-Subias, A., and Combacau, M., "An Architecture for Control and Monitoring of Discrete Events Systems," *Computers in Industry,* Vol. 36(1), pp. 95–100, 1998.

Zielinski, M., "Microprocessor-Based Controller Troubleshooting," *Instrumentation Technology,* June 1977.

4.15 Digital Readouts and Graphic Displays

J. VENCZEL (1985) **B. G. LIPTÁK** (1995) **P. M. B. SILVA GIRÃO** (2005)

Types:	A. Mechanical and electrical counters B. Gaseous discharge displays C. Cathode ray tube displays (CRTs) D. Rear projection displays E. Light-emitting diode (LED) displays F. Liquid crystal displays (LCD) G. Vacuum fluorescent displays (VFD) H. Digital panel meters
Note:	In this feature summary, the letters A through H refer to the listed alphanumeric readout types.
Display Application:	A and H. Numeric B and G. Alphanumeric C, D, E, and F. Alphanumeric and symbols
Character Sizes:	A. 0.1 to 0.5 in. (2.5 to 12.5 mm) E and G. 0.1 to 4 in. (2.5 to 100 mm) and larger H. 0.1 to 6 in. (2.5 to 150 mm) and larger B and C. 0.3 to 2 in. (7.5 to 50 mm) D. 0.12 to 3.38 in. (3 to 84.5 mm) F. 0.3 to 6 in. (7.5 to 150 mm) and larger
Maximum Viewing Distance:	A. 30 feet (9.12 m) G and H. 50 feet (15 m) B and C. 100 feet (30 m) D. 150 feet (45 m) E and F. 200 feet (60 m) or more
Viewing Angle:	A, C, E, and F. 90° D. 100° B, G, H. 120°
Brightness in Foot-Lamberts:	A. A function of ambient lighting F. A function of ambient lighting or 75 (255 cd/m^2) H. A function of ambient lighting or 75 (255 cd/m^2) or 200 (680 cd/m^2) D. 75 (255 cd/m^2) C. 100 (340 cd/m^2) B. 150 (510 cd/m^2) E. 200 (680 cd/m^2) G. 300 (1020 cd/m^2)
Life:	D. 10,000 hours C. 30,000 hours A and B. 100,000 hours G. 300,000 hours E, F, and H. 500,000 hours
Operating Voltage:	E. 3 to 5 V DC D and F. 5 to 28 V DC H. 5 to 30 V DC or 45 to 270 V AC A. 0 to 115 V AC or V DC B. 100 to 300 V DC G. 15 to 40 V DC C. 3,000 to 15,000 V DC

Costs: Type A from $10 to $30; type B from $20 to $40; type C from $75 to $500; type D from $30 to $60; type E from $1 to $100, type F from $4 to $40; and type G from $10 to $30. When the display is built into digital panel meters, voltmeters, multimeters the cost ranges from $100 to $500 or more. Large-size, high-resolution state-of-the-art CRT, rear projection, LED, and LCD displays may reach the tens of thousands of dollars.

Partial List of Suppliers:
Acculex Inc. (A, E, F, H) (www.acculex.com)
Agilent Technologies (E) (www.agilent.com)
American Opto Plus LED Corp. (E) (www.aopinc.net)
Amperex Electronic Corp. (B) (www.rell.com)
Automatic Timing and Controls (A, E, F, H) (www.automatictiming.com)
AZ Displays Inc. (E, F) (www.azdisplays.com)
DCI Inc. (A, F) (www.dciincorporated.com)
Dl International Inc. (E, F) (www.d1international.com)
Dialight Corp. (E) (www.dialight.com)
Elcon Instruments (H) (www.elconinst.com)
Electro-Numerica Inc. (E, H) (www.electronumerics.com)
Endress+Hauser (H) (www.us.endress.com)
Fairchild Semiconductor (E, H) (www.fairchildsemi.com)
Futaba Corp. (G) (www.futaba.com)
Hengstler GmbH (A) (www.hengstler.de)
Hoyt Electrical Instrument Works Inc. (H) (www.hoytmeter.com)
Industrial Electronics Engineers Inc. (F, G) (www.ieeinc.com)
Jewell Instruments LLC (H) (www.jewellinstruments.com)
Laurel Electronics Inc. (A, H) (www.laurels.com)
London Electronics Ltd. (E) (www.london-electronics.com)
Lumex Inc. (B, E, F) (www.lumex.com)
Martel Instruments (E, F, H) (www.martelinstruments.com)
Nec Corp. (F) (www.nec.com)
Newport (E, H) (www.newportus.com)
Noritake Co., Inc. (G) (www.noritake-elec.com)
Omron Electronics LLC (A, E, F, H) (oeiweb.omron.com)
Optrex (F) (www.optrex.com)
Precision Digital (E, F, H) (www.predig.com)
Red Lion Controls (E, F, H) (www.redlion-controls.com)
Rockwell Automation (C, E, F, G) (www.ab.com)
S-Products Inc. (E, F, H) (www.s-products.com)
Selco Products Co. (H) (www.selcoproducts.com)
Samsung Electronics Co. (C, F, H) (www.samsung.com)
Sanyo Semiconductor Co. (E, F) (www.semic.sanyo.co.jp/index_e.htm)
Sharp Electronics Corp. (F) (www.sharp-usa.com/SharpHome)
SunLED Corp. (E) (www.sunled.com)
Toshiba América Inc. (F) (www.toshiba.com/tai-new)
VarTech Systems Inc. (C, F) (www.vartechsystems.com)
Vishay Intertechnology Inc. (E, F) (www.vishay.com)
Weschler Instruments (A, H) (www.weschler.com)

INTRODUCTION

The development of digital technology in the 1970s was followed by the increased use of digital or alphanumeric (a/n) displays. As pocket calculators became commonplace, light-emitting diode (LED) displays and liquid crystal displays (LCD) were produced in very large quantities. This led to increased research and development efforts and greatly reduced prices. The evolution of cheaper monolithic analog-to-digital (A/D) converters paved the way for the widespread use of digital voltmeters. This again opened up new markets for alphanumeric displays. The use of microprocessors only added to the popularity of digital readouts, and so the digital era has led to the obsolescence of analog art.

As their name implies, a/n readouts furnish alphabetical (letters or legends) and numerical (digital or numbers) information. Common examples are the television cathode ray tube and the automobile odometer. These readouts provide accurate, easily understood displays and require no operator interpretation because they present clear, concise legends or exact numerical values.

Versatility is the primary advantage of a/n readouts because they have the capability to display many different types of information. Some readouts display only numbers, whereas others display numbers, letters, and symbols. The contents of the display can be tailored to the type of variable being displayed. For example, the readout of a digital voltmeter or electronics calculator may accommodate the maximum value and the decimal point location. In a process control display the readout may include the loop name and tag number with the value of the variable being measured.

Automated process control, in particular of complex distributed processes, is increasingly computer based. This means that from the human–machine interface point of view, significant changes are taking place and will continue to do so. The amount of information that can be easily presented to an operator is extremely large and diversified, recommending the use of displays with graphics and image capabilities of the same types used in computer monitors (cathode-ray tubes, liquid crystal displays, and even plasma-based monitors). The information conveyed by such displays, although eventually appealing and complete is very demanding for the operator since (a) to obtain the required information on a process several levels of pull-down menus must be navigated, and (b) the quantity of information for the operator to process can harm an effective and prompt answer to abnormal situations.

In the context of process control, and as far as the human–machine interface is concerned, two distinct situations may occur:

1. Simple processes (e.g., single loop control, single or small number of process variables) requiring simple interfaces (small number of displayed quantities and alarms). In this instance, LED, LCD, and VFC displays or panel meters are used.
2. More complex processes, requiring a heavy interface with a large or very large quantity of information to be analyzed. The display is then assured by either CRTs or flat-panel displays, mainly of the LCD type.

In the present section, different types of digital readouts are discussed. Some of them, namely type B, are clearly of the past or marginally used but, at least for historical reasons, still justify some attention. Because the interface between an operator and the process(es) control system (human or man–machine interface) is being constantly enriched with information technology-based solutions, some references are made to the pertinent technologies.

HUMAN FACTORS

The design of a human–machine interface (including display and means of obtaining the presented information such as pushbuttons, up–down buttons, or touch screen controls), as detailed in Section 4.17, "Human-Machine Interface Evolution," must be carefully designed to reduce operator stress and to quickly provide the information required. Some important aspects to consider are:

1. Organization of the information on the display, including its consistency (where it is shown and how it is presented)
2. Type of data presentation (alphanumeric, graphics, etc.)
3. Colors and character size

The design is particularly critical when the information requires large displays, several displays or, as it is more common, a single monitor but several menus and windows. In this last case, the performance of the software used for data presentation is extremely important.

To transmit information efficiently, readouts must have sharp character resolution, which is a measure of the character image sharpness or clarity and relates directly to readability. Readability is the quality that allows an observer to perceive information with speed and accuracy and is a function of character style, proportions, height, contrast, color, and refresh rate.

Characters that are pleasing to the eye are generated by simple continuous lines. To aid in character recognition, critical details should be prominent, and special features such as openings or breaks should be readily apparent. Many a/n readouts use a matrix of dots or bar segments for character generation. Closely spaced dots are more legible and natural-looking, whereas bar segments usually form boxlike characters containing noticeable inter-segmented spaces. Character proportions should be predicated on a height-to-width ratio of roughly three to two. Line width should be approximately one-seventh of character height. Minimum spacing between characters should be two line widths, with approximately six line widths between words.

SIZE AND CONTRAST

A guideline for determining character height based on viewing distance is given in Table 4.15a. The heights can be modified slightly for high ambient illumination or high-brightness displays. A display mockup is recommended if a particular viewing distance is critical.

Contrast is the ratio between character image brightness and its background brightness when measured under normal ambient illumination. Acceptable ratios for a/n readouts are about five or ten to one. Assuming the background to be a typical control console surface with a brightness of 20 to 50 foot-lambert (ft-L), or 68 to 170 candela per square meter (cd/m^2), a nominal brightness range for readouts should be 100 to 500 ft-L (340 to 1700 cd/m^2). Contrast depends on ambient illumination. Consequently, locations near an outside window or directly under a lighting fixture may require a filter to prevent image washout or to reduce objectionable

TABLE 4.15a
Character Heights Based on Viewing Distance

Required Viewing Distance in Feet (m)	*Nominal Character Height in Inches (mm)*
2.3 (28 in. or 0.7 m)	0.12 (3)
5 (1.5)	0.18 (4.5)
10 (3)	0.25 (6.25)
15 (4.5)	0.31 (7.75)
20 (6)	0.38 (9.5)
30 (9)	0.50 (12.5)
40 (12)	0.68 (17)
50 (15)	0.75 (18.75)
65 (19.5)	1.38 (34.5)
100 (30)	2.00 (50)

reflections. An antireflective filter reduces reflections and improves contrast by passing proportionately more self-generated image light than reflected image background light.

The optimum readout colors are green and yellow because the eye is most sensitive to them (Figure 4.15b). Amber, red, and orange are the next best choices. A filter can be used to obtain a desired color. A red filter, for example, will pass only the red light of an incandescent readout while it absorbs all other colors. Several types of a/n readouts generate character images by addressing time-displaced current pulses to dots or segments. The number of times (per second) that the image is generated is called the refresh rate. Low refresh rates can result in lowered brightness and occasionally cause flicker.

Two additional human factors involve viewing angle and change or update rate. If the readout must be seen from several operator positions, it should have a viewing angle of approximately 120 degrees. Recessed characters are satisfactory only for direct viewing. The maximum readout update rate is twice per second if the operator is expected to read consecutive values. Numerical readouts are not recommended for determining rate of change, direction, or tracking.

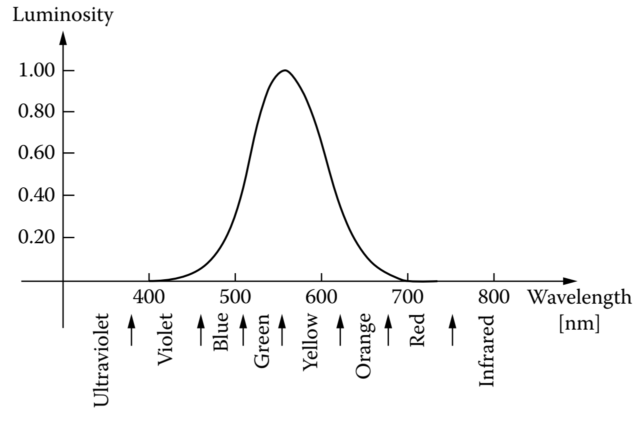


FIG. 4.15b
Luminosity as a function of wavelength. The National Bureau of Standards developed the above spectral-luminosity curve based on the findings of 52 experienced observers. The curve indicates that the eye responds to light between 380 and 760 nanometers, reaching maximum sensitivity at 555 nm in the yellow–green region.

APPLICATION NOTES

For critical readouts there should be a test function or switch to locate burned-out lamps or tubes or lost character portions. The loss of the horizontal center section of some matrix readouts, for example, can cause an operator to misread an integer as zero when eight is really intended. The readout bezel should be aesthetically pleasing and in contrast with the panel mounting surface.

Most a/n readouts require an electrical power supply to provide memory and to keep the readout continuously lighted. The readouts are updated or changed by a decoder driver that accepts binary coded decimal inputs. Readout modules containing standard decoder drivers with memory are preferred. Characteristics of common a/n readouts are listed in Table 4.15c, and a description of each type follows.

MECHANICAL AND ELECTRICAL COUNTERS

Mechanical and electrical counters (Figures 4.15d and 4.15e) are relatively simple and inexpensive devices that provide a cumulative count of sequential events. Usual configurations consist of a series of numbered wheels mounted on a common input shaft rotated by mechanical or electrical pulses. The numerical readout is visible through a window placed over the foremost row of digits. Digits should be as close as possible to the window to provide the maximum viewing angle and to eliminate shadows from ambient lighting. Internal lamps should be supplied if these counters are to be used in darkened areas.

The driver output and the counter input should be compatible. Mechanical counters are driven by mechanical rotation or oscillation of the input shaft. Rotation changes the readout through internal gearing, and oscillations use ratchet-pawl mechanisms. Electrical counters work similarly, except that a solenoid or an electromagnet actuates the input shaft. Counters operating with mechanically rotating inputs usually provide one or ten counts per complete rotation of the input shaft. Inputs can be either clockwise or counterclockwise; consequently, the count adds in one direction and subtracts in the other. Oscillating inputs provide one count per oscillation, and counting is in one direction only.

Electrical counters ordinarily supply one count per pulse and are available for AC or DC operation with a wide selection of voltage coils. Counters with two coils will add one count for a pulse through one coil and will subtract one count for a pulse through the other coil. Zero reset counters should be used for applications requiring a cumulative count

TABLE 4.15c
Alphanumeric Readout Characteristics

Type/Feature	*Method of Operation*	*Operating Characteristics*	*Typical Application*
Mechanical and electrical counters	Numbered wheels are rotated behind a viewing window.	Contain inherent memory; power required only to change display; usually illuminated by ambient light.	Digital clock; water meter; gasoline pump
Gas discharge displays	Shaped electrodes, in form of characters, ionize surrounding neon gas.	Require external memory and operate on high voltages (170 to 300 V DC).	Electronic calculators; electrical meters
Cathode ray tubes	Shaped electron beam is projected on phosphor screen.	Require character generation and memory circuitry; operate on very high voltages (3,000 to 15,000 V DC).	Computer-controlled displays
Rear projection displays	Miniature optical projectors containing incandescent lamps display images on viewing screen.	Require external memory; models are available for wide range of operating voltages.	Control-console displays, television monitors
Light-emitting diode displays	Semiconductor junction emits light when DC current is passed through it. Individual diodes make up dot or bar matrix.	External memory and drivers are required. Low-voltage, high-current operation (2 V, 15 mA per segment).	Electronic calculators, clocks, panel meters, computer-controlled displays
Liquid crystal displays	Transmissive technology. LCD does not generate light. It can be made to pass or block light by application of AC voltage.	External memory and AC drivers are required. Low-voltage, low-current operation (5 V, 1 $\mu A/cm^2$).	Digital watches, portable instrument displays, panel meters, computer monitors, computer-controlled displays
Vacuum fluorescent displays	A triode vacuum tube with phosphor-coated anodes. Variously shaped segments form matrix.	Medium-voltage (20 to 40 V), low-power device. External memory and drivers are required.	Clocks, calculators, and alphanumeric displays on instruments

between random times or events. A single knob depression for mechanical counters or pulse for electrical counters will set all wheels to zero, and a new count can be established.

GAS DISCHARGE DISPLAYS

The forerunners of this outdated group of displays were the so-called Nixie tubes (Figure 4.15f). The operation of these tubes is similar to that of neon lamps. A sealed glass tube contains a common anode and a stack of ten independent cathodes shaped like numbers. When a large negative voltage is applied to a selected cathode, the gas around it glows from being ionized and emits a yellow–orange light.

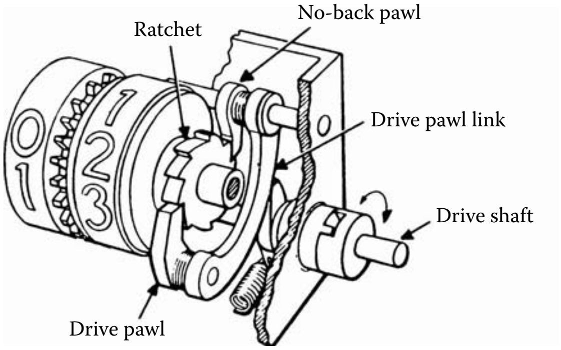


FIG. 4.15d
Mechanical counter.

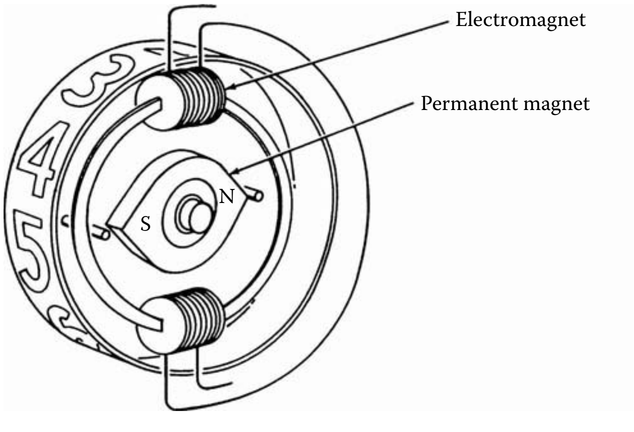


FIG. 4.15e
Electrical counter.

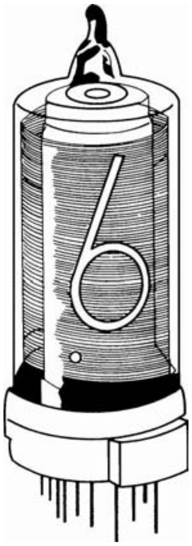


FIG. 4.15f
Gaseous discharge tube.

The gas in the tube is usually a mixture of neon and argon. A small amount of radioactive krypton and liquid mercury is added to the mixture to improve starting and to reduce cathode sputtering. These tubes were used up to the 1970s, when the planar-type matrix displays took over. These displays involve the same physical phenomenon as the Nixie tubes, except that the characters are made up of selectively lighted segments. The most common version is the seven-segment display, in which by using variable combinations of seven bar-shaped segments one can write all ten numbers and some other characters as well (Figure 4.15g).

The characters may be straight or slanted. The segments can be rectangular, mitered, or a combination of both with rounded corners (Figure 4.15h). For construction details, refer to Figure 4.15i. The whole assembly is placed in a flat, sealed glass container. The character segments are the cathodes and are located on the bottom of the glass enclosure. The anode is a fine metal mesh or a transparent tin oxide or indium coating on the inside surface of the cover glass. There are generally two or three characters to an envelope, with the connecting pins coming out on the rear side. A seal-off tube is located on the rear side for pumping the air out and inserting the gas mixture.

These devices are presently hard to find. They provide high brightness, good viewing angle, and 10,000 to 30,000 hours of operation and are moderately priced ($3 to $5 per digit). Among the disadvantages are the need for a high-voltage power supply (~180 V DC) and a high-voltage driving circuit.

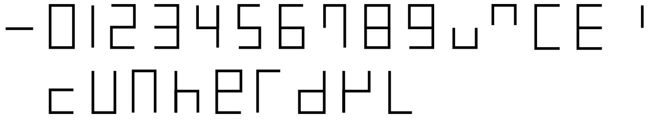


FIG. 4.15g
Some symbols available from the seven-segment character.

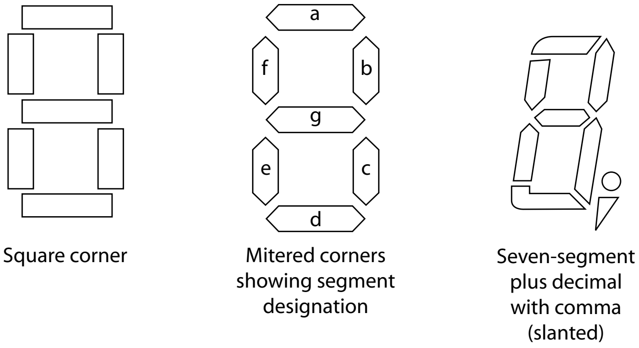


FIG. 4.15h
Typical character shape of seven-segment displays.

CATHODE RAY TUBE DISPLAYS

Cathode ray tube displays, commonly called CRTs, are discussed in detail in Section 4.5. CRT displays are versatile alphanumeric readouts. They are based on a cathode ray tube and from the outside they look like ordinary television sets. In effect, the operation is very similar to that of the display section of a television set. The characters may be drawn by a line raster that blanks and unblanks the beam or by X–Y writing, also known as stroke writing. A CRT is shown schematically in Figure 4.15j.

This type of display requires the most complicated support circuits. It requires a high-voltage power supply (3 to 15 kV), horizontal and vertical deflection circuits for raster generation, and a character generator. The raster generally consists of 525 horizontal lines or 1200 lines in high-resolution displays. Digital character generators can produce any types of characters previously defined and stored in semiconductor memories. CRT displays are complicated but can display a large amount of information. Typical uses are in computer terminals and airport departure–arrival displays.

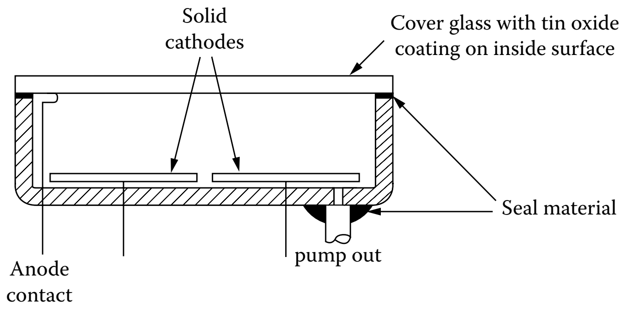


FIG. 4.15i
Cross-sectional view of a popular construction technique for the gas discharge display.

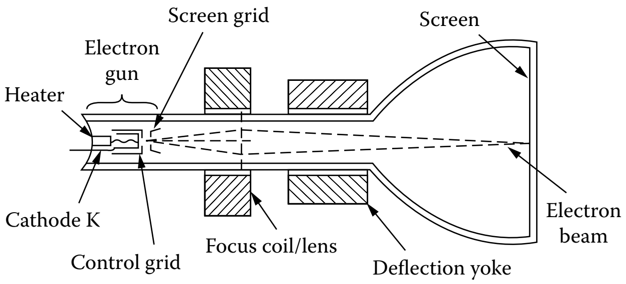


FIG. 4.15j
Basic CRT. The electron gun projects an electron beam on the screen that becomes luminous at the point of impact. The electromagnetic focus lens causes beam convergence, reducing spot size to usable dimensions.

The CRT, based on well-established principles and using commonly available materials, provides inexpensive, high-performance, true-color, stable, and high-resolution displays. However, CRT-based displays also have the drawbacks of high power consumption, divergence, and color variations across the viewing area. Their size and weight are also limitations (Figure 4.15k). For these reasons, the CRT is slowly but progressively being replaced by other technologies (e.g., LCD and plasma), particularly for mobile and/or large-size displays.

REAR PROJECTION DISPLAYS

Rear projection displays (Figure 4.15l) are miniature optical projectors stacked in cordwood fashion. They can display anything that can be put on film, including symbols, words, and colors. Twelve incandescent lamps at the rear of the display ordinarily illuminate the corresponding filmed messages, which are focused through a lens system and projected onto a single-plane, nonglare viewing screen. Rear projection displays exhibit excellent character resolution and readability and produce a very natural a/n readout appearance. They can be operated directly from electrical relays or from self-contained decoder drivers.

FIG. 4.15k
Twenty-inch Fast Scan Console Mount Monitor with touch screen capabilities. Dimensions: 17.62 in. (447.55 mm)(w) × 13.72 in. (348.48 mm) (h) × 20.35 in. (516.89 mm)(d); Weight: 50 lb (22.6 kg). (Courtesy of VarTech Systems Inc.)

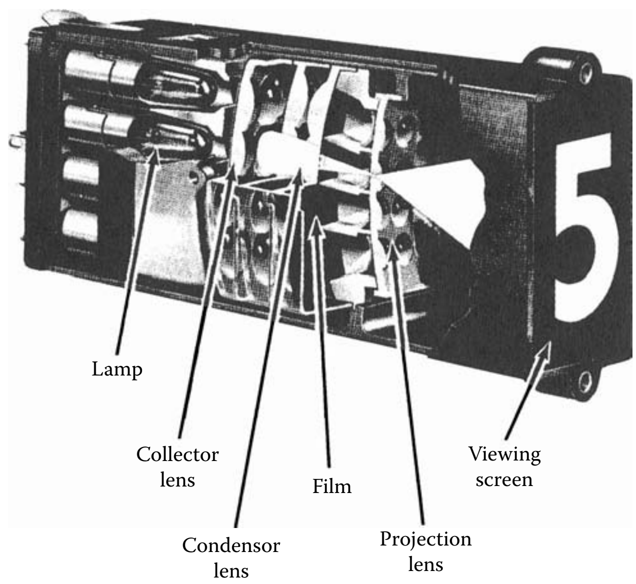


FIG. 4.15l
Rear projection readout.

Rear projection has recently experienced a strong technical development toward the implementation of performing rear projection monitor.[1] Nevertheless, the use of such monitors in process control rooms is insignificant because they cannot compete with CRT and LCD-based displays.

LIGHT-EMITTING DIODE DISPLAYS

Light-emitting diode (LED) displays have a wide variety of uses. They were developed in the 1970s, and by the end of that decade they could be found in all types of products, from watches and calculators to instruments, appliances, and point-of-sale advertising. Among their advantages are that they are inexpensive and easy to drive and have reasonable power requirements. The disadvantages include a reduced viewing angle for displays with magnifiers, and possible nonuniform segments for larger displays. They can be used as solid-state lamps, bar-segment displays, and dot-array displays.

Three colors are available for LED displays: red, yellow, and green. The red displays have the highest brightness and are the most widely used. The yellow is acceptable, and the green is the least legible. The sizes range from 0.1 to 0.8 in. (2.5 to 20 mm). As an example, Figure 4.15m shows a seven-segment digital LED display manufactured by Hewlett-Packard.

The most recent developments in LED technology are the introduction of the organic light-emitting diode (OLED) and the polymeric light-emitting diode (PLED) displays. Both types of displays are organized as matrices of pixels of OLEDs and PLEDs, respectively, formed on a thin film transistor (TFT) array.

An OLED is a device that sandwiches carbon-based films between two charged electrodes, one a metallic cathode and

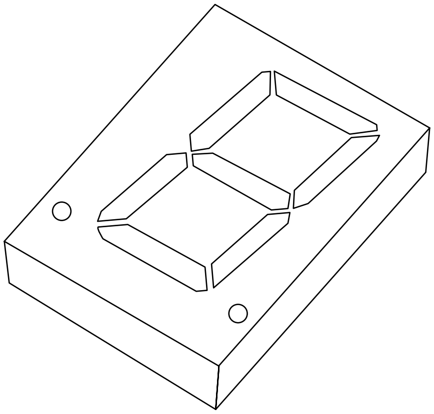


FIG. 4.15m

The HDSP-3400 Series are very large 20.32 mm (0.8 in.) GaAsP LED seven-segment displays. Designed for viewing distances up to 10 m (33 ft), these single-digit displays provide excellent readability in bright locations.

one a transparent anode, usually glass. The organic films consist of a hole-injection layer, a hole-transport layer, an emissive layer, and an electron-transport layer. When voltage is applied to the OLED cell, the injected positive and negative charges recombine in the emissive layer and create electroluminescent light. Invented by Eastman Kodak in the 1980s, the technology of OLED is brighter, thinner, faster, and lighter than LCDs, uses less power, offers higher contrast, and is cheaper to manufacture.

A PLED is created by sandwiching an undoped conjugated polymer between two proper electrodes at a short distance. The polymer emits light when exposed to electricity. It is a possible alternative to liquid crystal-based displays with the advantage of not requiring backlight. PLEDs enable full-spectrum color displays, are relatively inexpensive compared to other display technologies such as LCD or OLED, and require little power to emit a substantial amount of light.

LIQUID CRYSTAL DISPLAYS

The liquid crystal display (LCD) is an electronic readout that uses the unique characteristics of liquid crystals. This display technology is unique in that electrical energy is not converted to visible light, as with LEDs, gas discharge, and vacuum fluorescents; rather, the display modifies light. To create substantial contrast, the other display technologies require either considerable power (milliwatts per display element) or high voltage.

By taking advantage of unique characteristics of the liquid crystal materials, one can design LCDs to pass or block light. Only the orientation of the liquid crystal molecule must be altered in order for this effect to be created. This takes only microwatts of power per display element.

To understand the operation of an LCD and a twisted nematic liquid crystal display (TN LCD), refer to Figures 4.15n and 4.15o. The molecules in a liquid crystal display form a helix when viewed from the front to the back of the display cell. That is, the cigar-shaped molecules at the front of the cell may be oriented vertically and those at the rear horizontally. Those in the center are somewhere in between, forming a very orderly 90-degree helix. This 90-degree angle can be controlled exactly during construction.

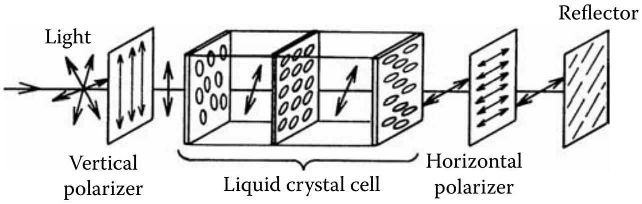


FIG. 4.15n

Field effect display unenergized between segments. (Courtesy of Fairchild Corp.)

The key to the operation of a liquid crystal display is polarized light. When scattered light is polarized (see Figure 4.15n) in the vertical plane and passed through the liquid crystal helix described previously, light is twisted, following the helical arrangement of the molecules. The light leaving the back of the display will be polarized in the horizontal direction. If a reflector is placed behind the display, the polarized light will return along its entry path, being twisted as it passes through the helix in the opposite direction.

If the helix of molecules is placed in an electric field, the molecules will align with the electric field, destroying the molecular helix. The polarized light entering the front of the liquid crystal cell will not be twisted but will emerge from the rear of the display polarized in the vertical plane. When vertically polarized light encounters the horizontally oriented rear polarizer, the light will be attenuated (see Figure 4.15o). If a reflector is present, no light will be reflected; the display would appear dark.

The construction of the LCD involves simple implementation of the operating principles (Figure 4.15p). The liquid crystal material is encapsulated in a glass "sandwich." The glass is coated with a transparent conductive material, usually indium-tin oxide, which is etched back to the preferred pattern.

Once the electrode has been formed, the inside surface must be treated to give the molecules a 90-degree twist, After this, the package is sealed with a small gap left between the

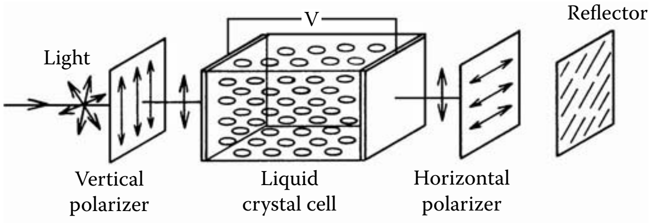


FIG. 4.15o

Field effect display energized between segments and backplane. (Courtesy of Fairchild Corp.)

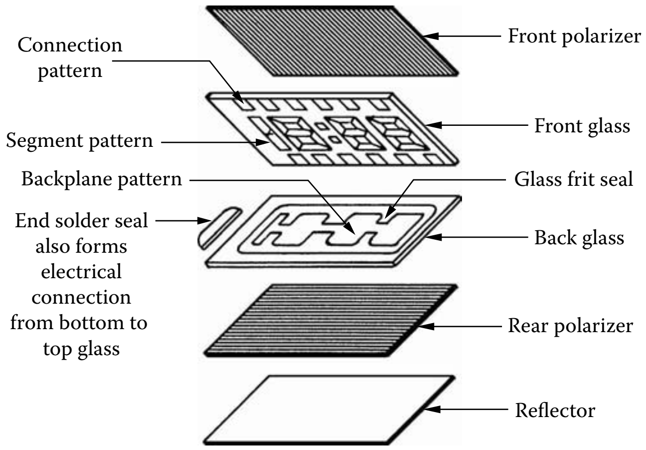


FIG. 4.15p
Liquid crystal cell structure. (Courtesy of Fairchild Corp.)

plates—generally 10 microns or less—for the liquid crystal material.

Liquid crystal displays are available in several configurations; reflective, transmissive, and transreflective. The choice of configuration depends on the lighting to be used. The reflective option includes a small reflector laminated on the rear of the display. This option gives the best performance and appearance in bright light. In a dark environment, however, the display has to be lighted from the front.

Liquid crystal technology was first applied to monochromatic numerical and alphanumeric displays. Its development was boosted by the television and computer monitor markets. The research involving materials and display construction is very active today. The available LCD panels can be divided into picture elements (pixels), which are arranged as a matrix. They fall into the categories of either passive matrix (PMLCD) or active matrix (AMLCD).

Passive Matrix Liquid Crystal Displays

One of the more common PMLCD designs is dual-scan twisted nematic (DSTN). The design includes several layers (Figure 4.15q):

1. A sheet of glass coated with metal oxide that operates as a grid of row and column electrodes, which pass the current needed to activate the screen elements.
2. An alignment layer made of a polymer of parallel grooves running across it to align the liquid crystal molecules in the appropriate direction and to provide a base on which the molecules are attached. This layer is repeated on another glass plate that also carries a number of spacer beads, which maintain a uniform distance between the two sheets of glass when they are placed together.
3. Polarizing layers applied to the outermost surfaces of each glass sheet to match the orientation of the alignment layers.

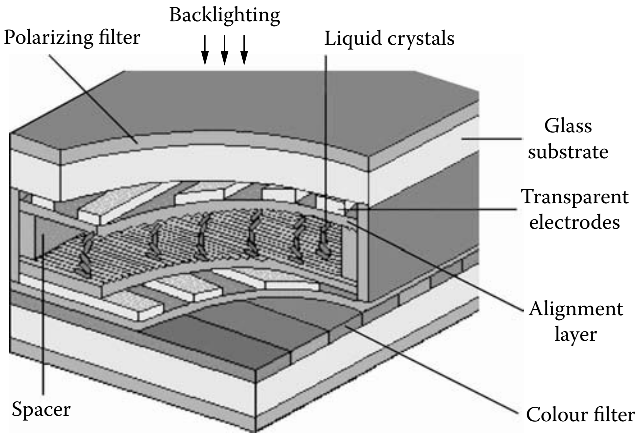


FIG. 4.15q
Passive matrix liquid crystal display (PMLCD) structure.

Backlighting is added, typically in the form of cold-cathode fluorescent tubes mounted along the top and bottom edges of the panel; the light from these is distributed across the panel using a plastic light guide or prism.

The image on the screen is created by this light as it passes through the layers of the panel. With no power applied across the LCD panel, light from the backlight is vertically polarized by the rear filter and refracted by the molecular chains in the liquid crystal so that it emerges from the horizontally polarized filter at the front. Applying a voltage realigns the crystals so that light cannot pass, producing a dark pixel. Color LCDs simply use additional red, green, and blue colored filters over three separate LCD elements to create a single multicolored pixel.

PMLCDs respond slowly, which causes smeared images and ghosting. Ghosting is when an area populated by pixels that are “on” causes a shadow on “off” pixels in the same rows and columns. Several solutions have been developed to overcome such problems, including the use of hybrid passive displays (HPD) and multiline addressing.

Full-color display requires intermediate levels of brightness between all-light and no-light passing through the crystals. The varying levels of brightness are achieved by changing the strength of the voltage applied to the crystals since liquid crystals untwist at a speed directly proportional to the strength of the voltage.

In contrast to full-color CRT displays that can provide 256 brightness levels (8-bit), PMLCDs offer only 64 different shades per element (6-bit), which means that using three elements per pixel, color PMLCDs deliver a maximum of 262,144 colors (18-bit). This compares with 16,777,216 colors (24-bit) of true-color CRT monitors. Figure 4.15r shows a small-size, low-cost, 64,000-color PMLCD, whose viewing angles are of 60° x-axis and 35° y-axis.

Active Matrix Liquid Crystal Displays

Thin film transistor (TFT) technology is presently used both in high-image-quality computer and process control monitors.

FIG. 4.15r
Seven and one-half inch color scan twisted nematic (STN) passive matrix LCD Module 640 × 480. Outline dimensions: 7.76 × 5.7 in. (197 × 145 mm); viewing area: 6.22 × 4.48 in. (158.00 × 113.75 mm). (Courtesy of Lumex, Inc.)

In a TFT screen, an extra matrix of transistors is connected to the LCD panel—one transistor for each color (RGB) of each pixel (Figure 4.15s). Transistor polarization controls the voltage applied across the liquid crystal element and thus the degree of twist and hence the intensity of the red, green, and blue elements of each pixel forming the image on the display. Thereby both slow response speed and ghosting of PMLCDs is eliminated.

A 640 × 480 color display AMLCD, such as the one shown in Figure 4.15r, would require 921,600 perfect transistors on a silicon wafer. In addition it would require an external driver circuitry, which justifies the higher prices of TFT LCDs compared to the passive LCDs. Similarly to PMLCDs, AMLCDs have also gone through several cycles of development to improve their viewing angles, response times, and contrast ratios. These developments involved in-plane switching (IPS), vertical-aligned (VA) systems, and multi-domain vertical-aligned (MVA) systems.

Table 4.15t shows the main characteristics of a 13.5-in. PMLCD and AMLCD and of a 15-in. CRT monitor. LCDs have advantages in terms of bulk, power consumption and flicker, but they are much more expensive, have poorer viewing angles, and are less accurate in their color reproduction.

The lower viewing angles of LCDs result from the fact they are transmissive systems that work by modulating the light that passes through the display, while CRTs are emissive. With emissive displays, there is a material that emits light at the front of the display, which is easily viewed from greater angles. In an LCD, in addition to the light, which is passing through the intended pixel, obliquely emitted light also passes through the adjacent pixels, causing color distortion.

Contrast ratio is a measure of how much brighter a pure white output is compared to a pure black output. The higher the contrast, the sharper the image and the purer the white will be. The contrast ratio of LCDs is lower than that of CRTs. Nevertheless, ratio values of 500:1 are currently available both in PMLCD and AMLCD displays.

Response time is the time it takes for each pixel to respond to the command it receives. Response time is a feature of LCDs because of the way they send their signal. Response time does not apply to CRTs because of the way they handle the display of information (an electron beam exciting phosphors). An AMLCD has a much faster response time than a PMLCD.

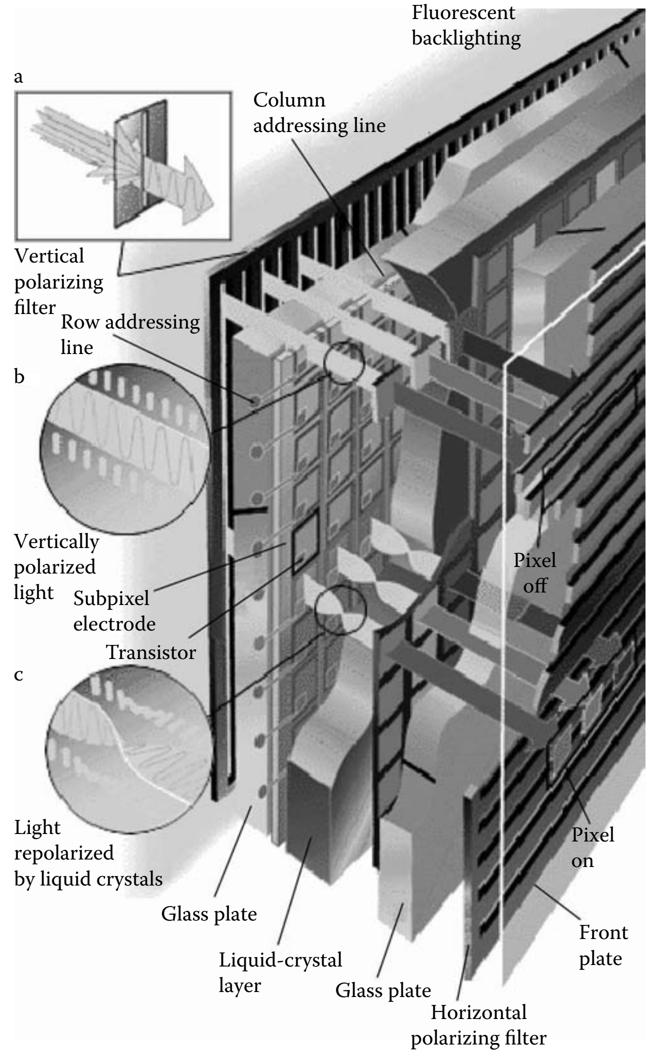


FIG. 4.15s
TFT active matrix liquid crystal display (TFT AMLCD) structure.

Brightness can be measured in different ways, but the higher the brightness number in Table 4.15t, the brighter the white display is.

Life-span value of an LCD is taken as the mean time before failure for the flat panel, which means that if it runs continuously it will have an average life of 60,000 hours before the light burns out. CRTs can last much longer, but while LCDs simply burn out, CRTs get dimmer as they age, and in practice, they do not have the ability to produce a standard luminance after around 30,000 hours of use.

While CRTs are capable of displaying a range of resolutions and can be scaled to fit the screen, an LCD panel has a fixed number of liquid crystal cells and can display only one resolution at full-screen size using one cell per pixel. Lower resolutions can be displayed by using only a portion of the screen or by using complex rescaling techniques. On

TABLE 4.15t
Comparison between a 13.5-in. Passive Matrix LCD (PMLCD) and Active Matrix LCD (AMLCD) and a 15-in. CRT Monitor

Display Type	*Viewing Angle (degrees)*	*Contrast Ratio*	*Response Speed (ms)*	*Brightness (cd/m^2)*	*Power Consumption (watts)*	*Life*
PMLCD	49 to 100	40:1	300	240 to 300	45	60 K hours
AMLCD	>140	140:1	25	240 to 300	50	60 K hours
CRT	>160	300:1	n/a	220 to 270	180	4–5 Years

the other hand, the ability to pivot the screen from a landscape to a portrait orientation is a feature of LCDs that is particularly suited to flat panels.

VACUUM FLUORESCENT DISPLAYS

The vacuum fluorescent (VF) display is based on the electronic tuning eye used in radio receivers up to the 1960s. The VF display consists of a vacuum tube triode with phosphor-coated anodes arranged in a planar configuration. It is widely used in digital alarm clocks and radios and as a digital display in automobiles. The typical color is a bright bluish green. The construction of the device is shown in Figures 4.15u and 4.15v, and a typical alphanumeric display is shown in Figure 4.15w.

The display is sealed in a flat glass enclosure. The anode segments, which represent the display elements, are located on the bottom. Each segment is connected to outside terminals sealed in the glass envelope. The anode segments are coated with phosphor material that emits visible light under electron bombardment. The grid is a fine mesh located between the filament (cathode) and the anode segments. The device requires a separate filament voltage, which may be AC or DC and ranges from 1.5 to 5 V. The anode voltage is typically between 12 and 40 V.

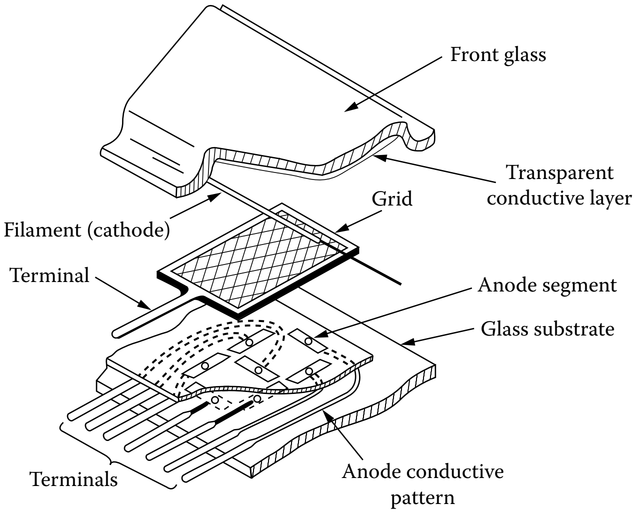


FIG. 4.15u
Construction of a vacuum fluorescent display. (Courtesy of Futaba Corp.)

The grid is used for turning on or blanking a whole digit for multiplexing purposes, whereas the combination of anode segments defines the character. One glass envelope generally contains several characters. The number of characters ranges from four to 40 or more. Displays with reduced graphic capabilities are available. Some modules have the same mounting hole location and have compatible interface with LCD devices, which means that such VF displays can replace LCDs with the advantages of lower power consumption, wider viewing angles, and higher brightness.

There are various numbers given in the literature for life expectancy or mean time between failures (MTBF), of a VF display. These numbers also depend on the definition of failure. Filament breakage is not a serious problem since it rarely happens. The phosphor coating will lose its effectiveness in time and this will result in lower brightness. Some manufacturers report VF devices with life spans of 300,000 hours.

In summary, the VF displays are adequately illuminated, have a wide viewing angle, and involve a medium-voltage, low-power operation. Their disadvantage is the low contrast at normal ambient lighting. This can be improved with green or blue filters, which also reduce distracting glare. These devices can be quite inexpensive, especially when purchased in large quantities.

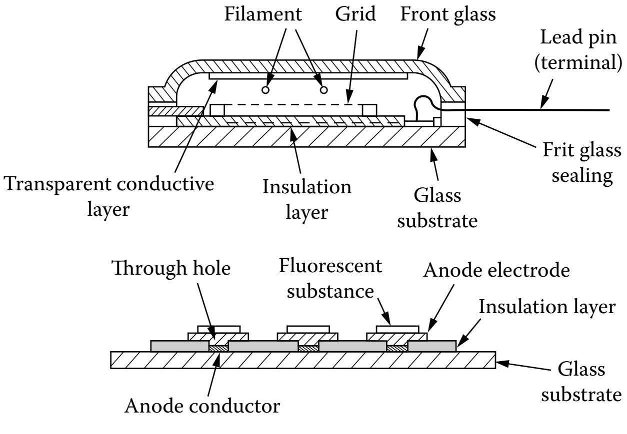


FIG. 4.15v
Cross section of a VF display. (Courtesy of Futaba Corp.)

FIG. 4.15w
Alphanumeric VF display. (Courtesy of Futaba Corp.)

NEW TRENDS IN GRAPHIC DISPLAYS

Several new static and dynamic display technologies are under development. These include organic LEDs (OLEDs); polymeric LEDs (PLEDs), which are also known as light-emitting polymers-based displays (LEPDs); plasma display panels (PDPs); field emission displays (FEDs); electroluminescent displays (ELDs); and digital light processor-based displays (DLPDs) for two-dimensional displays and holographic autostereoscopic display (HAD) for three-dimensional displays.

The required performance of the displaying devices depends on the applications. The more significant markets are television, computer monitors, and displays for small portable devices such as mobile phones. Other domains, such as medical imaging, are likely to benefit from the outcome of the ongoing research. A short reference to some of those technologies follows, even if they are not yet used in process control.

Plasma Displays

Plasma display panels (PDPs) are a fast-growing and very popular new generation of large-size displays.

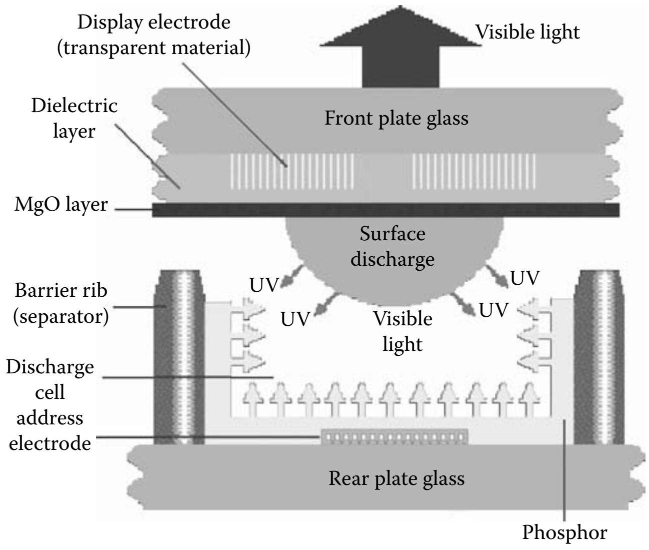


FIG. 4.15x
Plasma display panel structure and operating principle.

PDPs are based on multiple micro-discharges working on the same basic principles as fluorescent tubes. Plasma panels (Figure 4.15x) are arrays of cells, known as pixels, which are composed of three sub-pixels, corresponding to the colors red, green, and blue. Each cell has gas confined in the plasma state and electrodes that enable gas ionization under an electric field.

The UV rays that are created from this plasma gas excite the phosphors in each sub-pixel to produce colored light (red, green, or blue). Each sub-pixel is individually controlled by advanced electronics to produce over 16 million different colors. PDP technology is able to provide direct view displays with diagonals in the range 30 to 65 in., different resolutions, and viewing angles of more than 160°. Unlike projectors, they can be viewed under room-light or sunlight conditions. The main applications are public information displays and home cinema.

Table 4.15y summarizes the specifications of a 25-in. SXGA PDP released by a manufacturer that developed one of the several techniques presently available, an interlaced technique called alternate lighting of surfaces (ALiS).

Field Emission Displays

The field emission display (FED) (Figure 4.15z) is a promising flat panel display design that could replace both CRTs and LCDs. Similarly to CRTs, its operation is based on phosphors excited by electrons traveling in vacuum. The difference is in the electron emitters: instead of one electron gun, FEDs use millions of microscopically small electron-emitting cathodes that are matrix addressed. FEDs feature high brightness and

TABLE 4.15y
Specifications of a 2-in. SXGA PDP

Display pixels	1280 × 1024
Pixel pitch (mm)	0.39 × 0.39
Effective display size (mm)	499 × 399
Number of colors	260,000
Luminance (cd/m^2)	150 (white peak)
Contrast ratio (in dark room)	80:1
Weight (kg)	10
Viewing angle	Greater than 160°

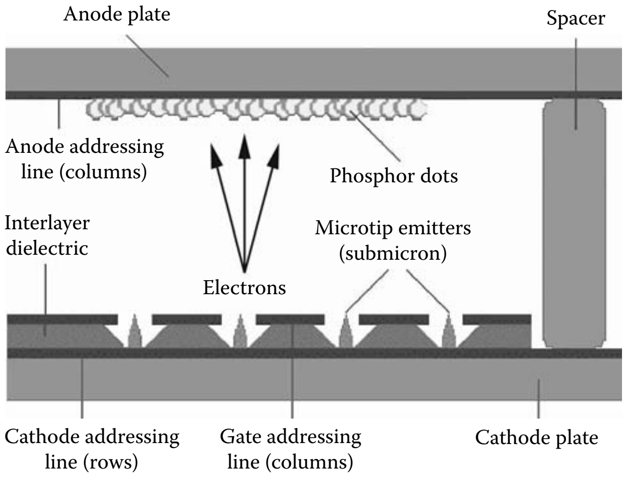


FIG. 4.15z
Field emission display structure and operating principle.

high efficiency, with a wide viewing angle, fast response time for video viewing, and perfect color quality. Actually, they are flat, thin, low-power CRTs.

Electroluminescent Displays

The operation of electroluminescent displays (ELDs) is based on the nonthermal conversion of electrical energy into luminous energy, the same principle as that of LEDs.[2] However, in ELDs light is generated in special materials when high-energy electrons that have been accelerated by an electric field impact on light-emitting centers (called activators).

Figure 4.15aa shows the schematic diagram of a thin-film electroluminescent display (TFELD) structure. The central layer is the thin film phosphor that emits light when a large electric field (on the order of some 1.5 MV/cm) is applied across it. Current limiting layers (the insulators) are needed on either side of the phosphor layer to protect it from the huge amount of dissipated energy if the phosphor were connected to the electrodes and a short circuit due to imperfection in the thin film would occur.

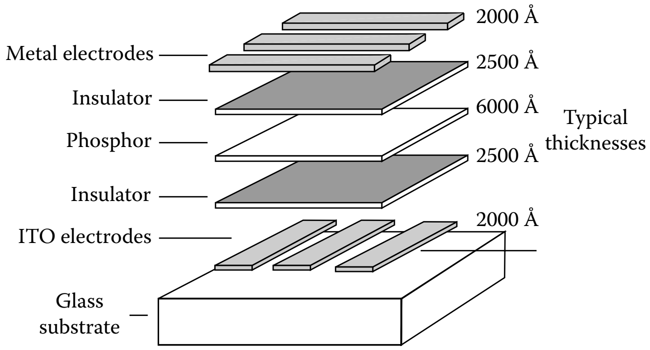


FIG. 4.15aa
Schematic diagram of a conventional thin film electroluminescent display (TFELD) structure.

The electrodes on the top and bottom of the device complete a basic capacitive structure. At least one set of these electrodes should be transparent, to permit viewing of the emitted light. Color displays have been demonstrated either using white or broadband phosphors and filters to produce a red–green–blue display or red, green and blue electroluminescent phosphors.

TFELDs have short response time and wide viewing angles (>160°) and are quite rugged, which makes them particularly attractive to produce high-information-content flat panel displays with the image quality of the CRT in portable applications requiring low weight and compact size.

References

1. Bush, R. C., “Rear Projection Screens for Different Applications,” Technical paper, Fresnel Optics, Inc., Rochester, NY, 2000.
2. King, C. N., “Electroluminescent Displays,” white paper, Planar Systems, Inc., Beaverton, OR, 2001.

Bibliography

Chigrinov, V. G., *Liquid Crystal Devices: Physics and Applications,* Boston: Artech House, 1999.

Compton, K., *Image Performance in CRT Displays,* Bellingham, WA: SPIE Press, 2003.

Gadberry, B. E., “Designing Integrated Control System Displays,” Paper #91-0354, 1991 ISA Conference, Anaheim, October 1991.

Guerlain, S., and Bullemer, P., “User-Initiated Notification: A Concept for Aiding the Monitoring Activities of Process Control Operators,” Honeywell Technology Center, 1996.

“Guide on Ergonomics of Man-System Communication,” SIREP-WIBEX-ERA Report S3630X94, 1994.

Keller, P. A., *Electronic Display Measurement: Concepts, Techniques, and Instrumentation,* New York: John Wiley & Sons, 1997.

Lanier, G. W., “Developing CRT Graphics Displays,” 1992 ISA Conference, Houston, October 1992.

Lueder, E., *Liquid Crystal Displays: Addressing Schemes and Electro-Optical Effects,* New York: John Wiley & Sons, 2001.

MacDonald, L. W., and Lowe, A. C., *Display Systems: Design and Applications,* New York: John Wiley & Sons, 1997.

Ono, Y. A., *Electroluminescent Displays,* Singapore: World Scientific, 1995.

Reinhart, J., “Are Your Central Displays User Friendly?” ISA/93 Technical Conference, Chicago, 1993.

Sherr, S., *Electronic Displays,* New York: John Wiley & Sons, 1993.

Shirley, R. S., et al., “What’s Needed at the Man/Machine Interface?” *InTech,* March 1981.

Stupp, E. H., and Brennesholtz, M. S., *Projection Displays,* New York: John Wiley & Sons, 1999.

Thumm, M., et al., “Displays and Vacuum Electronics,” April 29–30, Berlin: Garmisch-Partenkirchen, VDE-Verlag, 1998.

4.16 Fieldbuses and Network Protocols

P. G. BERRIE (1995, 2005) **B. GUT** (1995) **K-P. LINDNER** (1995, 2005)

Partial List of Suppliers: (Most fieldbuses or control system communication systems that have reached the stage of international or industrial standard are supported by user groups who maintain Internet sites. These sites contain lists of products and suppliers. For the systems covered in this section these are as follows:)

AS-i (www.as-interface.com)
HART (www.hartcomm.org)
PROFIBUS (www.profibus.org)
FOUNDATION Fieldbus (www.fieldbus.org)
ControlNet (www.controlnet.com)
MODBUS (www.modbus.org)
Ethernet TCP/IP (www.industrialethernet.com)

INTRODUCTION

In the ten years since the publication of the third edition of the *Instrument Engineers' Handbook*, fieldbuses and network protocols have become a major point of interest in process automation. While it is true that conventional plants, i.e., plants using 4 to 20 mA or pneumatic equipment for the collection and distribution of control signals, still represent 90% of production capacity, the majority of new and extension projects are now utilizing fieldbus technology.[1] The last decade has also seen HART technology achieve the status of an industrial standard. Thanks to a comprehensive range of auxiliary equipment and its ability to provide communication between conventional 4 to 20 mA signals and DCS systems, it has become the major method of migration toward fieldbus-like networks.

There have also been considerable changes at higher levels, where the success of Ethernet office solutions has broken down the monopoly of proprietary systems in plant operation and management. At the moment there is no agreement between manufacturers on a common protocol—FOUNDATION Fieldbus HSE, PROFInet, and MODBUS TCP (also known as MODBUS TCP/IP) are just three examples—but all use the same Ethernet physical layer and Internet TCP/IP protocols, and so can run on the same network. In this respect MODBUS TCP is assuming the same role as HART at the field level as a practical means of linking controllers or remote I/Os interfacing to networks operating with a different protocol.

Such technological advances mean that information technology (IT) now plays a major role in the design and operation of plant networks. As a consequence, a number of computer models for the structuring, processing, or exchange of data in a plant network are in the process of becoming international standards.

At the other extreme, many users are uncertain of how to implement the new technology, and a quick look at one of the many fieldbus user forums reveals lively discussions on, e.g., how best to ground a fieldbus segment. It is beyond the scope of this section to go into depth on either of these subjects; however, it is the authors' intention to provide interested readers sufficient insight into the basics of networks and fieldbus protocols that they can pursue the various aspects in more depth by referring to the sources in the References and Bibliography at the end of this section.

COMMUNICATIONS HIERARCHY

Before discussing the basics of digital communication, it is useful to consider how a plant network is structured and what activities occur at the various levels. Figure 4.16a shows a so-called automation pyramid. In this case, it comprises four levels: field, control, operations, and enterprise, each of which is discussed below. The reader may be familiar with other versions of the pyramid with a different number of levels, but the basic hierarchy remains the same. It must be said that nowadays the division into "levels" is more a matter of convenience than reality.

Some of today's fieldbus devices are equipped with the capability to drive control loops, so that the difference between control and field level is becoming blurred. In addition, the term "pyramid" is also misleading. There might be more equipment at field level, but the enterprise level with its

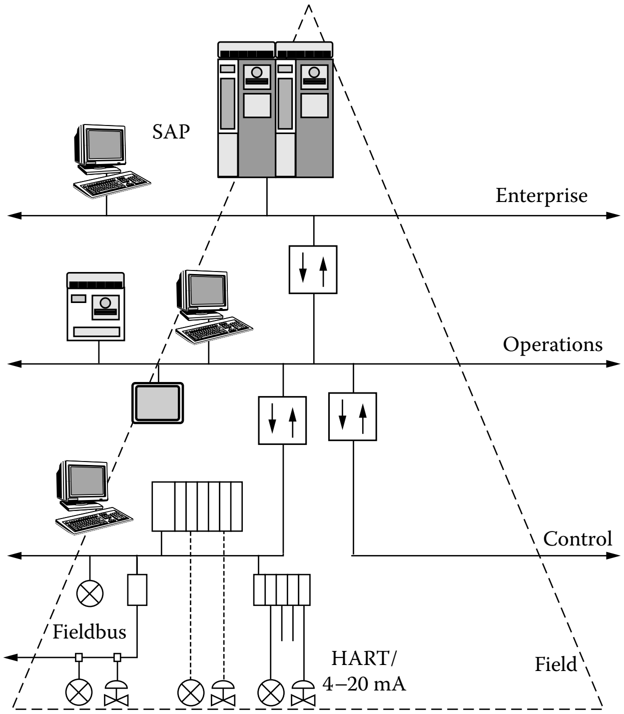


FIG. 4.16a
Automation pyramid.

business management programs is just as extensive, and the control level requires more programming effort. Nevertheless, the pyramid provides a simple overview of what should be attained by process automation, even if the realization may be somewhat different.

Field Level

The field level comprises the sensors and actuators that are required for the control of the processing plant. The sensors measure their process variables, i.e., temperature, pressure, level, flow, etc., and pass the signals up to the control level. There are two basic types of sensor:

- Limit switches signal one of two states: They detect whether the process variable that they are monitoring is in or out of limit. Conventional switches require the appropriate I/O card in the control system: relay, NPN, or PNP transistor circuit pulse frequency, NAMUR are examples. Fieldbus switches, e.g., PROFIBUS or AS-i, transmit a discrete digital signal, i.e., "0" or "1," directly over the fieldbus segment.
- Continuous measuring devices deliver a signal that is proportional to the process variable that they are monitoring. Conventional instruments deliver a standard current or voltage signal, often 4 to 20 mA, to an analog I/O card. Fieldbus sensors transmit an "analog output signal" — in reality a digital signal representing the current measured value — directly over the fieldbus segment.

The actuators receive signals from the controller in much the same way and perform a function, e.g., they start a pump when the signal changes state or close a valve in proportion to the magnitude of the control signal.

Control Level

At the control level, the signals from the sensors in the field are processed and the commands to the actuators are generated. The usual agent is a programmable logic controller (PLC), a process control system (PCS), or a distributed control system (DCS):

- A PLC uses simple ladder logic and provides real-time control. It executes its tasks in strict sequence by scanning the status of the analog and discrete inputs and generating appropriate output commands. The controller is programmed either directly through a console or from a remote engineering console — a personal computer — using the manufacturer's operating program.
- Process control systems are based either on mainframe computers or increasingly on personal computers or workstations. They differ from PLCs in that they are not bound to strict sequences and are capable of executing several tasks at once. A PC-based system cannot handle conventional signals, so these must be fed in via multiplexers or gateways on the control level bus. Since the information is already in the computer, they offer more possibilities for displaying the incoming signals.
- Distributed control systems: Modern process control systems are based on distributed intelligence. Here the control of the various parts of the plant is in the hands of several controllers or control systems. Each is responsible for its particular part and works autonomously. Each reports back to a central station based in the control room. Depending upon the network structure, the controllers may communicate over the control level network or be directly connected to the plant management level via Ethernet.

Operations Level

At the operations level we are no longer dealing with individual signals, but with the collection of plant data. It is also the last level at which intervention in the process is possible. Data are acquired directly from the field or via the controller by so-called SCADA programs (SCADA = supervisory control and data acquisition), giving the operator a schematic plant view on the monitor. Here, the operator can check all process values and take appropriate action when a value is out of limit. The signals are registered and stored in a daily log, alarm histories are archived, and trends are visualized. The performance of the plant is continuously monitored, analyzed, and optimized by specially adapted software. When necessary, recipes are changed.

This level also accommodates so-called asset management programs. These may operate independently of the control system, collecting fieldbus device process and status data at specific intervals rather than continuously. These programs usually allow configuration of the devices and may be linked to plant documentation, maintenance, and calibration scheduling programs. Asset management is a fast-developing area at the present moment, and the trend will be to marry both control and asset management into a single entity.

Enterprise Level

The enterprise level is where the information gained from the operations level flows into the business administration world, e.g., for production planning or ordering and billing of material. The enterprise level of a company network uses standard office networks and programs. A field device may signal the amount of liquid pumped in a day, but it cannot supply this information directly to a program at the enterprise level. Its data must first be processed and analyzed at the operations level, before being passed on through special software interfaces to the enterprise system.

Data Models

The communications hierarchy just described also forms a broad basis for two important standards that are concerned with the computer modeling of process industry networks. These are primarily concerned with the exchange of data throughout all levels of the factory, but are sufficiently well accepted that most major vendors base their engineering and enterprise tools on them. The oldest, the ISA S88 standard for batch control,[2] describes a more refined control hierarchy that splits the plant into operational units or parts thereof. This has been adopted by the International Electrotechnical Commission (IEC) in their IEC 61512 standard series,[3] and extended to include continuous processes, discrete processes, and inventory management. The structure is shown in Figure 4.16b.

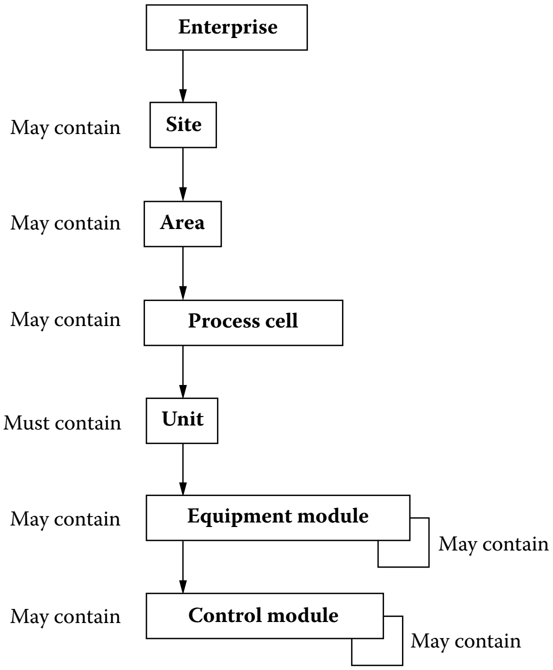


FIG. 4.16b
ISA S88/IEC 67512 model for process control.

Processes at the enterprise level and the exchange of data between it and the operations level are the subject of the ISA S95 standard.[4] Again, this is being adopted by the IEC as the IEC 62264 standard series.[5]

NETWORK BASICS

Digital communication between components at the top levels of control systems first became a serious proposition in the mid-1970s. System suppliers were quick to develop their own systems, a situation that repeated itself when field devices also began to talk a decade later. What was a fairly profitable operation for the vendor was a particularly expensive adventure for the user, who when wanting to use the best available components was faced with enormous programming costs for the integration. Not only did two systems speak different languages, there was also no common model for handling communication. Faced with this proliferation of proprietary solutions, the International Standards Organisation (ISO) took corrective action when it published the central pillar of all network standards, the so-called ISO-OSI reference model.

OSI Reference Model

The Open Systems Interconnection Model[6] was published in 1978 and aimed to bring uniformity and transparency into communication network standards. It defines a communication model that is applicable to all network devices, from mainframe computers operating at enterprise level to simple actuators in the field. It makes no physical specifications: rather, it provides a framework in which communications standards can be placed.

The OSI model splits the communication process into seven functional levels or "layers," see Figure 4.16c and Table 4.16d, each of which performs a clearly defined task. By defining the interfaces between contiguous layers, the model ensures the compatibility of all devices using a particular network protocol. It also simplifies the exchange of data between networks operating with different protocols. The application process that requests or receives transmissions is above Layer 7. This may be an enterprise system, an office program, a controller, a database, a SCADA program, a workstation, a sensor or actuator, etc. This so-called Layer 8 or "User Layer" is not part of the OSI standard.

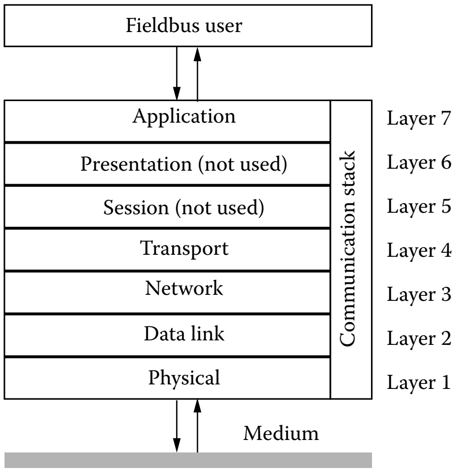


FIG. 4.16c
ISO-OSI communications model.

The seven layers can be divided into two functional groups:

- Layers 1 to 4 serve the network and are set the task of transporting data from one part of the network to another.
- Layers 5 to 7 serve the application and must present the data transported in an understandable form to the network user.

Since fieldbuses and control networks are much simpler than, say, the Internet, not all layers are required. Fieldbuses normally operate with Layers 1, 2, and 7 only; the devices on the bus receive their data from predefined controllers over so-called communication relationships. Since "Ethernet TCP/IP" is now found at control level, Layers 3 and 4 have also become relevant to process control networks. It can be seen from Table 4.16d that quite a number of standards of relevance to process control are already in place within the OSI framework. These together with other aspects of the physical network are discussed briefly in the next sections.

Physical Layer

The physical layer is concerned with how signals are placed on and carried over the network. The extent of standardization depends on the protocol used in the network. In some cases, a transmission standard only is specified; in others, the physical layer is completely defined including topology, transmission rate, connection pieces, cable length, cable type, etc. From a practical point of view, the most important aspects of the physical layer are the cable, the topology of the network, and the electronics residing in the fieldbus devices or control equipment that put the signal on the line.

Cabling The backbone of all networks is the transmission medium, i.e., the "cabling" that connects the devices. At field level this is usually copper cable, of which there are three main types, see Figure 4.16e. At higher levels it might be fiber optics or even "wireless":

- Unshielded Twisted Pairs (UTP). These comprise one or more twisted pairs of copper conductors within a protective jacket. UTP is used primarily in Ethernet networks. The cables are rated according to the

TABLE 4.16d
The Layers of the OSI Model

Layer	*Function*	*Task*	*Examples: Standards/Realizations for Fieldbuses*
Layer 7	Application	Provides the user with specific network commands and functions	Dynamic host configuration protocol (DHCP) Simple network time protocol (SNTP) Simple network management protocol (SNMP)
Layer 6	Presentation	Encodes the application layer data before forwarding them for transmission and decodes incoming data before forwarding them to the application layer	Not relevant to fieldbuses
Layer 5	Session	Synchronizes communication sessions between two applications	Not relevant to fieldbuses
Layer 4	Transport	Prepares the data string for transmission and ensures that it is reliably exchanged	Transmission control protocol (TCP) User datagram protocol (UDP)
Layer 3	Network	Selects the data route and ensures that the network is not overloaded	Internet protocol (IP)
Layer 2	Data link	Establishes and maintains connection between two participants	IEEE 802.2, Token passing IEEE 803.3, CMSA/CD IEC 61158/PROFIBUS PA: Master–slave IEC 61158/FF: Bus arbitration (LAS)
Layer 1	Physical	Puts the data on the physical medium and takes it off at its destination	EIA RS-232, EIA RS-422, EIA RS-485 IEC 61158-2 100BaseT (IEC 802.3u) 100BaseFX (ISO/IEC 9314-3)

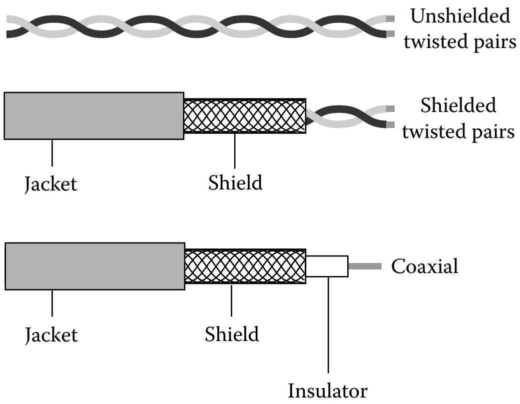


FIG. 4.16e
Copper cable types met in control applications.

EIA/TIA-568 standard,[7] whereby a Category 5 cable is recommended for transmission rates of 100 Mbit/s

- Shielded Twisted Pairs (STP). These are similar UTP cables, but the pairs are either shielded individually or share a common shield. The shield protects against electromagnetic interference. STP cables are recommended for fieldbuses: they are classed as Type A cables in the IEC 61158-2 physical layer standard.[8]
- Coaxial cable (Coax). This has a solid central conductor separated from an outer, tubular braided or foil conductor by an insulating medium, all within a protective jacket. The shield reduces electromagnetic interference and acts as the signal reference conductor. Coaxial cables are graded according to U.S. Military Standard MIL-C-17 with a RG (Radio Guide) number. In view of their expense, they are seldom found in the field, but might be used in computer systems operating at higher levels.
- Fiber-optic cables are specified as an alternative to copper cables in Ethernet (100BASE-FL), PROFIBUS DP, and FOUNDATION Fieldbus HSE. They are able to transmit at much higher rates and over much longer distances. Moreover, they are completely unaffected by electromagnetic interference and can be run into explosion hazardous areas without the need for barriers.
- Wireless at field level is still more of a dream than reality at the time of writing, but it has established itself within Internet applications as a means of connecting to remote networks.

For an overview of the connectors used in control systems and other aspects of the physical layer, the reader is referred to Barton.[9]

Topology The network topology describes the way in which the devices in the network are connected together. The basic topologies are star, ring, and bus: all other topologies are built from these basic elements. Figure 4.16f shows all three together with a so-called “tree” structure:

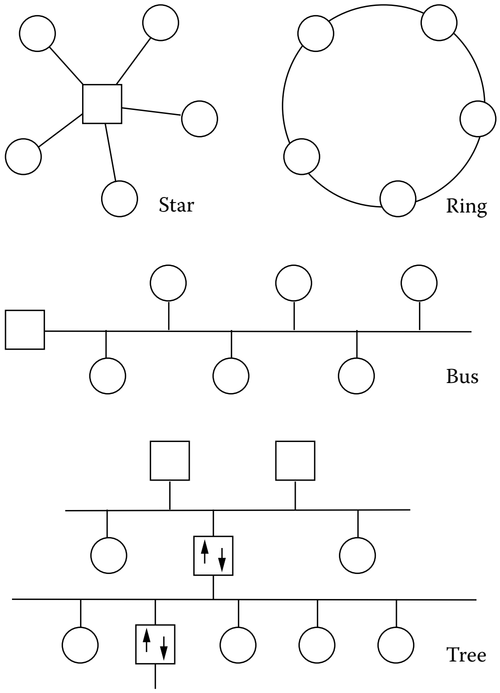


FIG. 4.16f
Network topologies.

- A star topology comprises a central device with point-to-point connections to all peripherals. The central device controls access to the network, and any data exchanged between peripherals must pass through it. If the central device fails, the network fails. The failure of individual lines is uncritical. The extent of the network is limited by the number of connections to the central controller. Shielded cable should be used. As transmission interface RS-232C is sufficient, however, in the classic star application, a computer system, this is now being displaced by the much faster universal serial bus (USB) interface.
- In a ring topology all devices are connected together in a physical ring. Decentralized control is required and access rights are passed from device to device. If a line fails the network fails, so bypass switches are required to avoid this. The number of participants is unlimited; however, the time it takes to circuit the ring sets a practical limit since this governs the response time of the system. The line must be interference free, i.e., coaxial or other reliable cable is required.
- In a bus topology all devices are connected either directly or via spurs to a single physical line. Both centralized and decentralized control is possible. Depending on control mode, line failure has the same consequences as in a star or ring. Failure of individual devices, however, does not affect the network function.

The number of participants is unlimited; however, the time taken to poll all devices sets a practical limit. The line must be interference free, i.e., coaxial or other reliable cable is required.

- A tree topology is basically a bus topology with branches or so-called "chicken feet," i.e., points where several devices are connected to one junction box. It is frequently met in PROFIBUS and FOUNDATION Fieldbus applications where a fieldbus operating at 31.75 kbit/s is connected to a high-speed bus via a link or coupling element.

Although, e.g., a bus or ring topology allows an unlimited number of participants, in practice, the associated communication protocols and transmission interfaces impose limits. The design of a network must also take into consideration other factors such as maximum cable length, grounding and, if power is taken from the cable, maximum allowable load. More information on designing FF and PROFIBUS segments can be found in Volume 3 of the *Instrument Engineers' Handbook*.[10]

Transmission Interfaces The digital signals produced by the devices are launched onto the network via a transmission interface. In order to communicate with each other, all the devices must use the same interface, and as such it will be part of any system or fieldbus specification. The transmission interfaces of interest to control networks are listed together with their attributes for copper cables in Table 4.16g.

It is beyond the scope of this section to go into detail about how the various interfaces work; however, it can be seen that the transmission standard itself often imposes conditions on the topology and protocol. Thus, RS-232 can only be used in a star topology with one device driving the communication. It should also be noted that the line lengths quoted are the maximum for the standard. The 1900 m quoted for IEC 61158-2 becomes 1000 m with an intrinsically safe bus, and the 1200 m copper for RS-485 shrinks to 100 m when the maximum rate of 12 Mbit/s is required. The other attributes in Table 4.16g together with a number of terms frequently met when transmission interfaces are discussed are found in Table 4.16h. Readers requiring detailed information are referred to Thompson.[11]

Data Link Layer

The data link layer provides the tools necessary for two or more devices to establish and remain in communication. It is responsible for fault-free transmission of bit streams over the network, calculating various cross-sums, and comparing them to so-called check bits. When a data packet is lost or faulty it may also have the responsibility for retransmitting the message. Another function of the data link layer is to add extra data frames to the bit stream that contain information about the transmitter and receiver, the priority of the message, and whether or not its receipt is to be acknowledged.

The data link layer is also responsible for media access, i.e., the controlled use and sharing of the network by the various devices connected to it. There are two distinct methods to regulate access:

- Centralized bus control (fixed master): the central bus master controls the access to the bus by the devices. Examples are the master–slave and the bus arbitrator models.
- Decentralized bus control (flying master): thanks to its intelligence, each device is in a position to control the bus itself when it wants to communicate. Two examples of decentralized bus control are the Token Passing model as per IEEE 802.2[12] and the CMSA/CD model as per IEEE 802.3.[13]

TABLE 4.16g
Characteristics of Transmission Interface Standards of Interest to Control Systems

Attributes	RS-232C	RS-422	RS-485	IEC 61158-2	100BaseTX
Max. devices	1 transmitter 1 receiver	1 transmitter 16 receivers	32 transmitters/ receivers	32 transmitters/ receivers	Limited only by address range[1]
Signal coding	±12 V pulses	Differential	Differential	Manchester II	8B6T
Topology	Star	Bus	Bus	Bus	Star
Max. line length[2]	15 m	1200 m	1200 m	1900 m	100 m
No. of lines	min. 3	4	2	2	4
Max. rate of transmission	19.2 kbit/s	10 Mbit/s	10 Mbit/s	37.5 kbit/s	100 Mbit/s
Communication mode	Duplex	Duplex	Half-duplex	Half-duplex	Duplex
Transmission type	Asynchronous	Asynchronous	Synchronous	Synchronous	Synchronous
Connectors	9- or 25-pin connectors	Not specified	Not specified	Not specified	RJ45 connectors

[1] A practical limit to the address range is given by the response time.

[2] Maximum possible allowed by the standard. Line length is dependent upon transmission rate and application.

TABLE 4.16h
Terms and Definitions Associated with a Transmission Interface

Attribute	Alternatives	Description
Transmission mode	Parallel transmission	Byte-by-byte transmission over a minimum of eight parallel lines. High transmission speeds, but expensive cables. Used by some buses for automated manufacturing.
	Serial transmission	Bit-by-bit transmission over a minimum of one line. The transmission time is dependent upon the length of the data string. Often used for fieldbuses.
Carrier mode	Baseband	Direct transmission of data without prior modulation.
	Broadband	Transmission of data by modulating the data on a high-frequency carrier signal. The data can then be simultaneously transmitted with other data modulated on a different carrier signal on the same transmission medium.
	Carrier signal	Frequency in a communications channel modulated to carry analog or digital signal information.
Signal transmission	Amplitude modulation	The amplitude of the data signal (“0” = low DC voltage or current, “1” = high DC voltage or current) causes a change in amplitude of a carrier signal, usually an alternating current.
	Frequency and/or phase modulation	The amplitude of the data signal causes a change in frequency or phase of a carrier signal, usually an alternating current.
Transmission type	Asynchronous transmission	Spontaneous transmission of data, requiring agreements on set sequences and types of information (protocol). Suitable for smaller blocks of data.
	Synchronous transmission	Transmission requiring that system clocks in both transmitter and receiver are in phase. It allows long data blocks to be transmitted efficiently and safely with a high proportion of useful data.
Communication mode	Half-duplex	In half-duplex communication, information flows in both directions. First one device transmits: after it has finished, the partner answers.
	Duplex	In duplex communication information can be simultaneously transmitted and received.
Transmission rate	1200 bit/s in steps to 100 MB/s	Indicates how many bits per second can be transmitted between the transmitter and receiver. All devices on a network must operate at the same transmission rate. The maximum rate is limited by the type of interface and transmission medium used.

At field and control level, it is often the case that two methods are used together, e.g., when several masters share the same backbone but communicate in a master–slave relationship with their field devices and other control equipment.

Communication Modes Communication modes include:

- *Master–slave.* In this method, one bus device is the master. The master alone possesses a continuous right to transmit. It polls (or addresses) each bus participant in turn, supplies it with data, and/or asks it to transmit its data, e.g., its measured value and status. To increase the efficiency, i.e., the amount of useful data to overheads, the protocol is kept as simple as possible. The security of the data depends upon the protocol structure and the error checking method used. MODBUS, PROFIBUS DP/PA, and AS-i are all protocols that use the master–slave method.
- *Bus arbitration.* In this method, every device on the network may transmit and receive data. The right to transmit is organized by a central controller, the so-called bus arbitrator. The arbitrator maintains a list containing the order in which and the length of time devices are allowed to transmit. The arbitrator works through the list, passing the right to transmit to each device in turn. If different scanning frequencies are required, the device can appear more than once in the table. It should be noted that although this is classified as centralized control, token passing is a central element. The method is used in the WorldFIP and FOUNDATION Fieldbus H1 protocols.
- *Token passing (token ring).* In this access method, the token, i.e., the right to transmit, is passed from device to device. This may be done on a master–slave or peer-to-peer basis. In the event of peer-to-peer passing, a so-called token ring is built, which may be real or logical. The passing sequence depends upon the application and is defined during the planning of the system. This method allows each device a fair share of the bus, since each is allowed to transmit within a preset period of time. The time taken to pass the token around the system determines the frequency of polling of the individual members. Token passing is used by FOUNDATION Fieldbus to ensure deterministic communication on the H1 layer.
- *Stochastic or random bus access.* There are several random access methods of which CSMA/CD is the most well known. CSMA/CD stands for Carrier Sense

Multiple Access with Collision Detection. All devices on the bus have the right to transmit. Each continually listens to (or senses) the bus. If this is free, then any of the devices can transmit its data. If several devices want to transmit simultaneously, a collision is detected and all withdraw. A random generator in each of the devices then determines the time interval that must elapse before it can attempt to transmit again. The CSMA/CD bus access method is specified in the Ethernet protocol and is used in office networks and the higher levels of automation systems. At field and control levels it has to be supported by a centralized access management method because there is no strict periodicity of scanning.

Communication Relationships Several types of communication relationships between devices may be supported by the bus protocol:

- *Client–server.* The client transmits a request to its server communication partner. When the requested action is executed and its confirmation is necessary, the server issues a response to the client. If the client is also a server to its communication partner, the relationship is said to be peer-to-peer. Otherwise it can be considered to be master–slave.
- *Publisher–subscriber.* All the devices sense the traffic on the bus continuously. The publisher broadcasts its data when it has the right to transmit. The subscribers to the data recognize that they are of interest to them and update their local copies. There is no direct confirmation of receipt. When communication is controlled by bus arbitrator, transfers can be scheduled on a precisely periodic basis. Publisher–subscriber is also known as the producer–consumer method.
- *Source–sink.* A source transmits when it detects a necessity to report, e.g., an event and has the right to do so. One or more sinks receive the (multicast) message, ensuring that the event is detected. One of the sinks is responsible for acknowledging the event via a separate client–server relationship.

Network Layer

The network layer is required neither by "H1" fieldbuses nor by Ethernet itself, which defines only the physical and data link layers. It is used by so-called "Industrial Ethernet" or "Ethernet TCP/IP" and by other industrial standards such as MODBUS TCP and FOUNDATION Fieldbus HSE.

The network layer is responsible for data routing. When the transport layer demands a connection between A and B with a specific capacity, it is the network layer that finds a suitable route. It must also make sure that the network is not overloaded. Where intermediate networks are to be used, the data must be stored before it is transferred from one network to another. The Internet protocol is of interest to process control:

- *Internet protocol* (*IP*). IP is the Internet's most basic protocol and is used to forward data packets. It is a datagram-oriented protocol that treats each data packet independently. This means that each packet contains complete addressing information. Packets can also be assigned priorities via a "type of service" field. IP has, however, no mechanisms to determine whether packets reach their destination. Furthermore, each packet has a "time to live" (TTL) field: when the TTL value reaches zero, the packet is discarded. The contents of a packet are not checked for transmission errors; only the IP header is checked. These functions are provided by such transport layer protocols as UDP (to a limited extent) and TCP. IP also provides several optional features, allowing a packet's sender to set requirements on the path it takes through the network (source routing), to trace the route a packet takes (record route), and to label packets with security features.

Transport Layer

The transport layer is responsible for the interference-free transport of data via the various routes and networks connecting any two communicating units. According to requirement, the transport layer sets up several logical channels and chops long information strings into small packets. On the receiving side, it reassembles the packets into the original string or, if the data are not received in sequence, attempts to place them in correct order. The transport layer also checks that the data packets have not been falsified during transport. Two protocols are of interest to process control:

- *Transmission control protocol (TCP).* The transmission control protocol runs on top of the Internet protocol, see above, and enables two hosts to establish a connection and exchange streams of data. TCP guarantees delivery of the data and also guarantees that the data packets are delivered in the same order as they were sent.

- *User datagram protocol (UDP).* The user datagram protocol also runs on top of IP. Unlike TCP it provides very few error recovery services. Instead it offers a direct means of sending data telegrams over an IP network. It is used primarily for broadcasting messages.

Application Layer

The application layer defines the services supported by the network, for instance, read and write commands, program management functions, network synchronization, addressing, up- and downloading of data, and virtual device images. It is not concerned with the actual carrying out of services, but rather with their presentation to the user so that the user can program the network. The procedures are then mapped onto the lower layers. In the full model, the application layer is connected to the lower layers through the presentation and session layers. In both fieldbuses and in Ethernet, these layers are bridged by special communication programs. Of the various protocols available to the application layer at least three find use in control applications:

- *Dynamic host configuration protocol (DHCP).* A protocol that automatically assigns IP addresses to devices on a network every time they connect to it. This means that a new device can be added to a network without the need to manually assign it a unique IP address.
- *Simple network time protocol (SNTP).* Simplified version of the Network Time Protocol (NTP), an Internet standard protocol that assures accurate synchronization of device clock times in a network. Running as a background client program on a computer, NTP sends periodic time requests to servers, obtaining server time stamps and using them to adjust the client's clock.
- *Simple network management protocol (SNMP).* The Internet standard protocol for network management software. It monitors devices on the network and gathers device performance data for management information databases.

FIELDBUS PROTOCOLS

Readers familiar with the history of digital communication in process control will know that work on standardizing an international fieldbus for the process industries was started in the late 1980s under the auspices of the ISO/ISA SP50 committee. Its feasibility was demonstrated by a multi-vendor demonstration at both ISA and Interkama exhibitions in 1989. Before the events occurred that caused to committee to become the scene of the so-called "fieldbus wars," it had been decided that a two-tier structure was best suited to the application. At fieldbus level, a bus with moderate transmission rates was to offer device power and be suitable for use in hazardous locations. This became known as H1—"Hunk 1," referring to the work package. "Hunk 2" was the control level requiring faster transmission rates, no bus power, and no requirement for use in hazardous locations.

The committee standardized the H1 physical layer in IEC 61158-2,[8] but there was no agreement on the data link and application layers or on the H2 bus. Company politics and various other factors caused the standardization to grind to a halt. Without going into further details, the result of this stand-off was emergence of the PROFIBUS standards and the setting up of the FOUNDATION Fieldbus. Enmity continued for several years until the IEC decided to publish both standards, plus a variety of proprietary systems that were prepared to open their specifications to the public, in the IEC 61158 set of standards,[14] see Table 4.16i.

At the same time, the FF standard was published as a European (EN) standard.[15] The decision was much criticized, but it has basically set up a situation where individual users can decide for themselves which fieldbus "flavor" they prefer.

TABLE 4.16i
Fieldbus Types Standardardized in IEC 61158: 2000–01

Type	*Protocol*
IEC 61158, Type 1	FOUNDATION Fieldbus H1
IEC 61158, Type 2	ControlNet
IEC 61158, Type 3	PROFIBUS DP/PA
IEC 61158, Type 4	P-Net
IEC 61158, Type 5	FOUNDATION Fieldbus HSE
IEC 61158, Type 6	SwiftNet
IEC 61158, Type 7	WorldFIP
IEC 61158, Type 8	Interbus

Note: The IEC 61158 standard itself currently comprises six parts that specify and define services for the physical, data link, and application layers. The types reference the parts that are relevant to the particular standard.

Although the idea of a single fieldbus is laudable, in practice there are also different demands at field level. In a typical production facility, there is a mixture of factory automation (e.g., filling and packaging) and process automation (e.g., batch and continuous control), which justifies the use of different protocols. In the following, those control and fieldbus protocols of interest to process control will be discussed. The various types of components used to build the networks are listed with a short description in Table 4.16j.

AS-i

AS-i (Actuator-Sensor interface) is a primarily a factory automation bus for simple binary sensor and actuators operating in safe areas. An overview of the principal technical data is given in Table 4.16k. AS-i is supported by a consortium of users and manufacturers, which is responsible for development and conformance testing.[16] Currently over 100 vendors offer AS-i devices or equipment.

Figure 4.16l shows a typical architecture. The AS-i bus is normally connected to a PLC or PC by using an appropriate interface card. It can also be connected to a control network by means of a gateway and multiplexer acting as a master. Power is supplied over the bus, and each participant can draw up to 100 mA at 24 VDC. A maximum of 8 A can be drawn.

AS-i is a cyclically operating system with a master–slave structure. In the standard configuration, the master interrogates all connected slaves in strict rotation and receives telegrams from them. The data structure is extremely simple: AS-i telegrams have four output bits that are used to control connected devices—to open a valve or close a switch. A slave answers immediately, returning four bits indicating that the function has been performed. A sensor, e.g., a level limit switch returns its current status. Alarm interrupts or similar event-controlled messages are not provided. The

TABLE 4.16j
Network Components

Component	*Standard*	*Description*
Bridge	Ethernet	Device that connects two local area networks (LANs) or two segments of the same LAN that use the same protocol.
Field device	General	Sensor or actuator attached to the process under control.
Frequency converter	General	Device on a network that serves to control motor drives, etc.
Gateway	General	Device on a network that serves as an entrance to another network. In addition to interfacing physical layer and protocols, a gateway may also scan and assemble data for forwarding as a single packet.
Host	Ethernet	Computer that is connected to a TCP/IP network, including the Internet. Each host has a unique IP address.
Hub	Ethernet	Common connection point for devices in a network with several ports. When a data packet arrives at one port, it is copied to the others so that all segments can see it.
I/O	General	Short for input/output. Device that transfers data, usually in the form of analog or discrete signals, to and from a controller.
Interface	General	Boundary across which two independent systems meet and act on or communicate with each other.
Link (PROFIBUS)	PROFIBUS	Device comprising a gateway and one or more segment couplers.
Linking device	FOUNDATION Fieldbus	Device serving as the interface between FOUNDATION Fieldbus H1 and HSE segments.
Modem	General	Device to enable transfer of digital data over analog telephone lines or cell telephone networks.
Multiplexer	General	Device that combines (multiplexes) several signals for transmission over a single medium. Often called a mux.
Port	General	Interface on a computer to which a device or network can be connected.
Programmable logic controller		Device using ladder or similar simple logic to provide real-time control.
Repeater	General	Device used to regenerate or replicate a signal.
Router	Ethernet	Device that forwards data packets along networks. Routers are located at gateways, i.e., the places where two or more networks connect.
Remote I/O	General	I/O device that is not an integral part of the controller but an independent entity residing within the network.
Segment coupler	PROFIBUS	Device comprising a signal coupler, a bus power unit, and a PROFIBUS DP/PA interface.
Switch	Ethernet	Device that filters and forwards data packets between LAN segments.
Switching hub	Ethernet	Short for port-switching hub, a special type of hub that forwards data packets to the appropriate port based on their address.

error detection is also much simplified and provides only moderate security.

Originally AS-i supported 31 slaves, whereby a slave could be an intelligent device or a so-called user module with 4 inputs and 4 outputs. The recently introduced Version 2.1

TABLE 4.16k
Principal Technical Data of AS-i

Property	*AS-i*
Standard	IEC 62026-2; EN 50295-2
Physical layer	AS-i specific, UTP
Length	Max. 100 m
Transmission rate	167 kbit/s
Bus access method	Central master slave
Participants	1 master, max. 31 slaves

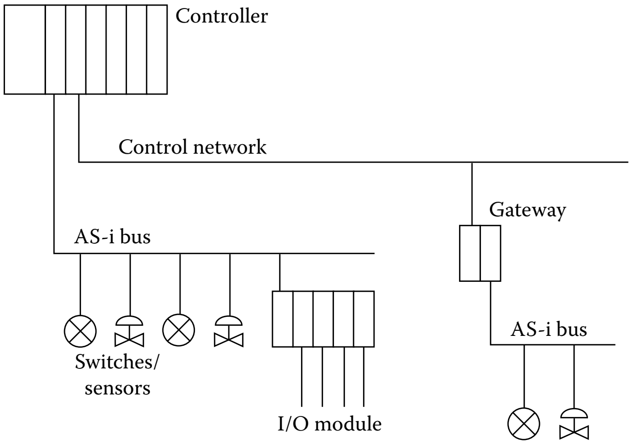


FIG. 4.16l
Typical AS-i architecture.

specification doubles the number of slaves to 62, while providing backwards compatibility with earlier devices and networks. The maximum cycle time of a fully loaded network is around 5 ms (double for Version 2.1). Version 2.1 also offers better diagnostics and a profile for analog signals. It is now possible to transmit analog signals such as temperature over 5 cycles for PID control.

HART

HART (Highway Addressable Remote Transmitter) is primarily a smart protocol; its information is transferred point-to-point by superimposing a digital signal on a standard 4 to 20 mA output. The protocol also allows for a multi-drop bus structure, however, at the expense of the conventional output. An overview of the principal technical data is given in Table 4.16m. More detailed information on HART is available in Volume 3 of this handbook.[17] HART is supported by the HART Communication Foundation,[18] an independent body of users and vendors who are responsible for development and conformity testing.

HART is a single master protocol, which for multi-drop applications allows a secondary master. This is the case when a master reading process data cyclically from the field devices and a terminal for device configuration (handheld or laptop) are connected to the same segment. As far as the coding of the data frames is concerned, HART uses a mixture of Layer 7 and Layer 2. Every data frame contains a command byte that determines the service to be performed (Layer 7) as well as the significance of the data for the data link layer (Layer 2). The command codes are divided into three classes: Universal, Common Practice, and Device Specific.

TABLE 4.16m
Principal Technical Data of HART

Property	*HART Point-to-Point*	*HART Multi-drop*
Standard	Industrial standard maintained by HART Communication Foundation	
Physical layer	Bell 202 FSK (Frequency Shift Keying) "0": 2200 Hz, "1": 1200 Hz	
Length	As for analog signals on UTP cable	Max. 1500 m with multi-core STP cable Max. 3000 m with two-core STP cable
Transmission rate	Digital information at 1200 bit/s with simultaneous analog transmission	1200 bit/s
Bus access method	For digital communication, master–slave with secondary master	Master–slave with secondary master
Participants	1 slave per connection	Max. 15 slaves with central power supply

All HART instruments must be able to process commands of the type universal. Common practice commands are not mandatory, but when they are implemented, the meaning of the commands must conform with the specification. In the case of device-specific commands, the same designations may have different meanings in different instruments. The data security in HART is ensured by the building of cross-sums, parity building in each frame, and an additional cross-summing of the parity bits.

In view of the fact that the protocol allows for device-specific commands, manufacturers must supply device description files (DDs) if their products are to be fully operable by a HART application. If no DD is available, operation is restricted to the basic parameters encompassed by the universal commands such as the writing of zero, span, and damping and the reading of process value and status.

Figure 4.16n shows typical ways in which HART devices are integrated into control system architectures:

- *Conventional.* Where digital communication is not of interest (except perhaps for device configuration via handheld terminal), HART devices are connected directly to analog or binary PLC cards via a point-to-point connection. Where the 4 to 20 mA device is loop-powered, the PLC provides the power.
- *Multiplexer.* If the HART digital signal is required, e.g., to configure the devices or monitor their readings from a workstation in the control network, a multiplexer may be used. This is often operated parallel to conventional point-to-point wiring. The multiplexer acts as HART master and scans each of the transmitters in turn, storing the measured values (and status information) in a buffer. The controller now accesses the buffer over the control network and reads the information it requires. The data

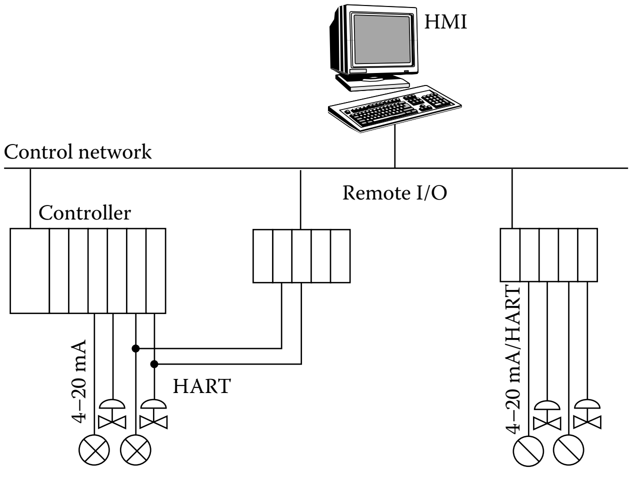


FIG. 4.16n
Typical HART architectures (not including multi-drop bus).

can also be accessed by a SCADA program, which is running in parallel to the controller on a workstation.

- *Remote I/O.* Remote I/Os are wired point to point and may provide loop power for 4 to 20 mA devices. Intelligent I/Os may also contain processing abilities, e.g., the ability to link signals from several sensors to produce a secondary process variable, e.g., density or mass, or be able to perform simple control actions such as closing a valve. Device information is collected by the controller or application software via the control network. Remote I/Os are often transparent, allowing direct configuration of the devices by an appropriate engineering tool.
- *Direct integration.* A recent development on the market is the introduction of controller I/O cards which also process the HART digital signal. This allows the controller to directly access the additional information provided by the devices.

It was mentioned at the start of the section that HART is extremely popular. It has sustained its popularity by successively enhancing its protocol with features that the users require. Where it has not been successful is to free itself from its origins as a protocol for pressure transmitters. This means that more complex devices such as flow transmitters and analyzers rely heavily on device-specific parameters for configuration to the detriment of interchangeability. In addition, the low transmission rate means that the time required for configuration is some 20-fold that of a standard fieldbus protocol.

PROFIBUS DP/PROFIBUS PA

PROFIBUS is an open fieldbus standard. It was developed in Germany and was published as national standard in 1992. European Standard EN 50170 followed roughly a year later.[19] The application profiles PROFIBUS DP (Decentralized Periphery) and PROFIBUS PA (Process Automation), which are used in process automation, are incorporated into the IEC 61158 standard published in 2000.[14]

An overview of the technical data is given in Table 4.16o. PROFIBUS is supported by an international network of PROFIBUS User Organizations, which are also responsible for the maintenance of the standards and certification and testing of the devices.[20] Readers requiring more detailed information on PROFIBUS are referred to Sections 4.13 and 4.14 in Volume 3 of this handbook.[10,21,22] Here they will find a full description of PROFIBUS DP and PROFIBUS PA, as well as information on how to design IEC 61158-2 segments (PROFIBUS PA).

TABLE 4.16o
Principal Technical Data of PROFIBUS PA and PROFIBUS DP Version DPV1

Property	*PROFIBUS PA*	*PROFIBUS DPV1*
Standard	DIN 19245 Part 4; EN 50170, Part 2; IEC–61158 Type 3	DIN 19245 Parts 1 to 3, Version DPV1; EN 50170, Part 2; IEC 61158 Type 3
Protocol	PROFIBUS DP	PROFIBUS DP
Physical layer	IEC 61158-2	RS-485 and/or fiber optics
Length	Max. 1900 m for safe and EEx ib areas Max. 1000 m for EEx ia	Up to 1200 m, depending upon transmission rate
Transmission rate	31.25 kbit/s	9600 bit/s to 12 Mbit/s
Bus access method	Handled from PROFIBUS-DP side	Master–slave with token passing
Participants	32 in safe areas, approx. 24 in EEx ib and approx. 10 in EEx ia	Per segment: max 32 Logical: max 126 (using repeaters), including max. 32 as masters

Figure 4.16p shows a typical PROFIBUS process automation network. It can be seen that PROFIBUS DP (Version DPV1) provides the H2 control network backbone and PROFIBUS PA the H1 fieldbus. Many systems comprise a programmable logic controller (PLC) and a personal or laptop computer equipped with a PROFIBUS DP card, on which the engineering tools are installed. All other devices, e.g., drives, valve positioners, and sensors, are slaves. They may be PROFIBUS DP or PROFIBUS PA devices:

- The PLC acts as a Class 1 master and communicates cyclically or acyclically with the slaves.
- The engineering tool acts as a Class 2 master and communicates acyclically with any device, either

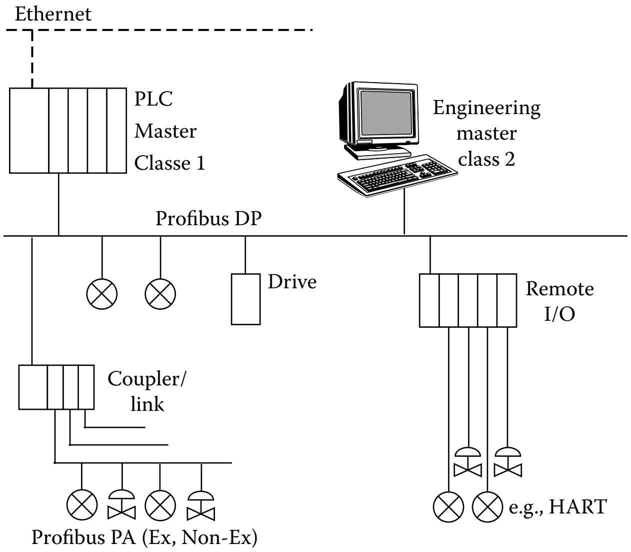


FIG. 4.16p
Typical PROFIBUS architecture.

according to its own schedule or at the request of the device.

- The medium is shared between the two masters by token passing.
- When a master has the right to transmit, it talks with its slaves in a master–slave relationship

The field devices, i.e., sensors and actuators, on the PROFIBUS PA segment communicate with a PROFIBUS DP master via a segment coupler or link:

- A segment coupler interfaces the RS-485 (PROFIBUS DP) and IEC 61158-2 (PROFIBUS PA) physical layers and provides (intrinsically safe) bus power: the same protocol is used throughout the network.
- A link comprises two or more segment couplers and acts as a slave on the PROFIBUS DP side and a master on the PROFIBUS PA side. It collects data cyclically from its slaves on the PROFIBUS PA segments and parcels them into a single telegram, which is sent to the PROFIBUS DP master. In acyclic communication a PROFIBUS DP master, it provides a direct connection with the PROFIBUS PA device addressed.

PROFIBUS PA supports the FISCO model (Fieldbus Intrinsically Safe COncept), so that for segments operating in explosion-hazardous areas, certification is required only for the individual components, not for the network as a whole.[23]

The PROFIBUS standard specifies that each device must contain a set of universal parameters. Mandatory parameters must always be present. Optional parameters are only present when required. Finally, manufacturer-specific parameters are used to realize device functions that are not in the standard profile. Certification ensures configuration compatibility between identical "device-types" manufactured by different manufacturers, e.g., pressure transmitters.

For devices that conform to the standard, these parameters are managed in block objects, e.g., resource, transducer, input, and output blocks. Manufacturers supply both device description files (DDs) for device configuration and device database files (GSMs) for system integration. Should these not be available, devices can still be operated, if not fully, by means of the profile parameters (mandatory and optional) for the particular device type.

Data are exchanged by means of standard telegrams. An "analog" value requires 5 bytes (4 bytes value, 1 byte status) and a discrete value 2 bytes (1 byte value, 1 byte status). A device may output more than one measured value; for instance, a Coriolis flow meter may offer up to 10 different outputs, i.e., a maximum of 50 bytes. When the segment is operated through a segment coupler, data exchange presents no restriction to segment size. For a link, however, there is an upper limit to the length of the telegram, often 244 bytes, but sometimes 122 bytes. This may reduce the number of devices connected to a segment to below the limits shown in Table 4.16o.

Another limiting factor is the current available to devices on the bus. FISCO devices typically draw less current than other intrinsically safe devices, but nevertheless, eight devices is often the maximum per segment for applications in explosion-hazardous areas.

Cycle times vary according to the number of devices in the network and the number of outputs being scanned. An estimate can be made as follows:

Total cycle time = Sum of the cycle times of the field devices + internal PLC cycle time + PROFIBUS DP system reaction time

The cycle time of the PLC is typically 100 ms, that of the devices typically 10 ms for one measured value and 1.5 ms for each additional measured value. The PROFIBUS DP system reaction time is dependent upon the number of slaves and the retry setting. Both are dependent upon the baud rate on the PROFIBUS DP side, which may range from 45.45 kbit/s to 12 Mbit/s depending upon the components used.

PROFIBUS has a large installed base covering a variety of applications in both factory and process automation all over the world. Although it is true to say that its stronghold is in Europe, it would be foolish to dismiss it as a European phenomenon. There are many important users to be found in North America, Southeast Asia, and South Africa. Its strengths lie in the amount of vendor support it has, its ability to use one protocol for factory and process automation, and last but not least, the ease with which it fits into PLC architectures.

FOUNDATION FIELDBUS

FOUNDATION Fieldbus is an open fieldbus standard to IEC 61158[14] and European Standard EN 50170[15] that was developed and is supported by the Fieldbus Foundation.[24] It has been designed to solve the measurement and control tasks associated with process automation, aiming to provide optimum communication between programmable logic controllers (PLCs) or process control systems (PCS) and the plant equipment operating on the field level. A detailed description is to be found in Volume 3 of this handbook.[25] The standard specifies two communication levels:

- HSE, on which traffic between controllers, computers, frequency converters, and other control equipment is handled
- H1, on which traffic between the process sensors and actuators is handled

HSE (High-Speed Ethernet) is the high-speed communication specification for the FOUNDATION Fieldbus. It provides a cost-effective, high-speed backbone by using standard Ethernet technology running at a fixed rate of 100 Mbit/second. FOUNDATION Fieldbus H1 provides low-speed communication for field instrumentation. The FOUNDATION

TABLE 4.16q
Overview of FOUNDATION Fieldbus H1 and HSE

Property	*FOUNDATION Fieldbus H1*	*FOUNDATION Fieldbus HSE*
Standard	EN 50170, Part 4; IEC 61158 Type 1	EN 50170, Part 4; IEC 61158 Type 5
Protocol	FOUNDATION Fieldbus	FOUNDATION Fieldbus
Physical layer	IEC 61158-2	100BaseTX, 100Base FX
Length	Max. 1900 m for safe and Ex ib areas Max. 1000 m for Ex ia	Up to to 100 m (copper) Approx. 2 km (fiber optics)
Transmission rate	31.25 kbit/s	100 Mbit/s
Bus access method	Bus arbitrator (Token passing)	CSMA/CD
Participants	32 in safe areas, approx. 10 in Ex ia	Logical: limited by address range only Practical: limited by response time

Fieldbus protocol is used for both HSE and H1 levels. Table 4.16q gives an overview of the main technical data.

The H1 physical layer (wiring and data transfer mechanism) is based on the international standard IEC 61158-2,[8] i.e., it is exactly the same as PROFIBUS PA. Originally, safety in hazardous areas foresaw only an Ex i or Ex d concept, but the FISCO model[23] was added to the specification in 2002. More information on design of H1 buses can be taken from the corresponding section in Volume 3 of this handbook.[10]

Figure 4.16r shows a typical architecture as foreseen by the specification:

- The process is run by one or more controllers connected to the HSE backbone. These may be separate entities or reside in the linking device. The engineering tool may be part of the system or may run on a separate workstation. Similarly, there may be a SCADA system running in parallel.
- FOUNDATION Fieldbus HSE handles the communication at the control level. Drives, remote I/Os and linking devices, etc., are connected to this bus. At this level the devices are externally powered and HSE ensures that data are quickly exchanged.
- FOUNDATION Fieldbus H1 is used at the field level. The linking device serves as interface to the HSE system. In addition, a power unit and conditioner are required to power the field devices over the bus. Depending upon power conditioner type, the H1 segment can be installed in safe or hazardous areas.

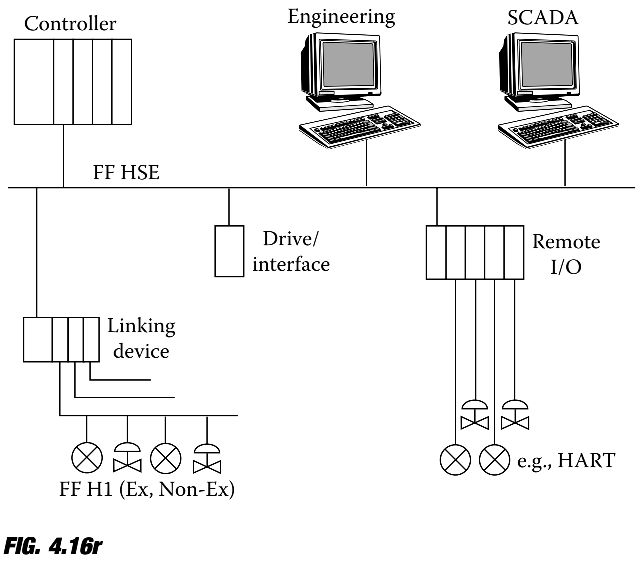


FIG. 4.16r
Typical FOUNDATION Fieldbus architecture.

In view of the fact that the HSE specification was published much later than the H1 specification, there are at least two alternatives to the above architecture:

- Connection via ControlNet: The field devices on the H1 bus are connected to the ControlNet network via a bridge (linking device), which in turn is connected to the ControlNet card of the controller.
- Connection via H1 card: The FOUNDATION Fieldbus H1 bus is connected directly to the controller or control system via a FOUNDATION Fieldbus H1 I/O card.

Access to the fieldbus is managed through a deterministic centralized bus scheduler called the link active scheduler (LAS). Fieldbus devices may be of two types: link masters or basic devices. A link master has LAS functionality; a basic device does not. Each fieldbus segment contains one LAS. It may also contain other link masters that are not active, but that can be called upon to become the LAS in the event of failure of the preferred device. All transmissions, whether scheduled or unscheduled, are instigated at the request of the LAS:

- Scheduled transmissions are used to cyclically transmit process values and control block I/O parameters. The LAS sends a compel data (CD) request at the scheduled time, and the device publishes its data on the bus.
- Unscheduled transmissions are made during the periods when the field devices are executing their function blocks or at the end of the CD schedule. The LAS temporarily gives device permission to transmit in a client–server or source–sink relationship, by sending a pass token message.

Table 4.16s gives more details of how the various communication methods are used.

Similar to PROFIBUS PA, FOUNDATION Fieldbus also manages device parameters in block objects, e.g., resource, transducer, input, and output blocks. It is unique, however, in

TABLE 4.16s
Types and Uses of Communication Relationships in FOUNDATION Fieldbus

Publisher/ Subscriber	*Client/Server*	*Report Distribution (Source–Sink)*
Scheduled	Unscheduled	Unscheduled
Used for publishing data	Used for operator messages	Used for event notification and trend reports
Sends parameters between blocks, e.g., the PV value to a PID control block and operator console	Set-point changes Mode changes Tuning changes Upload/download Alarm management Access display views Remote diagnostics	Sends process alarms to operator consoles Sends trend reports to data historians

TABLE 4.16t
Principal Technical Data for MODBUS

Property	*Serial MODBUS*	*MODBUS TCP*
Standard	Industrial standard	Industrial standard
Physical layer	Not specified, but usually RS-422, RS-485	10BaseTX, 100BaseTX
Length	1200 m at 19.2 kbit/s	100 m
Transmission rate	Max. 19.2 kbit/s	10/100 Mbit/s
Bus access method	Master–slave	CSMA/CD
Participants	1 master, max 247 slaves	Logical: limited by address range only Practical: limited by response time

specifying standard control blocks, e.g., PD, PID, Arithmetic, etc., that may be implemented in the field devices. Manufacturers supply both device description files (DDs) for device configuration and device database files (CFFs) for system integration. Unlike PROFIBUS, up to the time of writing FOUNDATION Fieldbus has not specified device profiles, so that generic operation of particular device types is not possible. Usually, the system engineering tool is used both to configure the network and set up the devices parameters.

The control cycle time depends upon the number of devices on the segment and the way in which the control loop has been configured. Although control block execution in a controller is quicker than in a field device, it involves a greater number of data transfers, each taking around 10 ms. In addition, unscheduled transmissions must be made to ensure the data flow within the system. On average, therefore, the full cycle (macrocycle) is 150 to 200% of the time it takes to execute the control itself. Typically this will be around 1 to 2 seconds. If the cycle time is unacceptable for the application at hand, it must be optimized by reducing the number of devices on the segment or changing the location of the function blocks.

FOUNDATION Fieldbus was later on the market than PROFIBUS and suffered from the late introduction of the HSE layer. As a result, support for high-level control equipment, i.e., drives, etc. is still weak. Initially, it suffered from interoperability problems because only devices and not host systems were tested for conformance. This has now been remedied, however, and the market has responded accordingly. FOUNDATION Fieldbus has strong support in North and South America as well as Southeast Asia and China. It is beginning to make an impact on Europe.

Its installed base is in traditional DCS industries such as oil and gas and pulp and paper, but it is also penetrating water and wastewater and chemicals. Although it offers control in the field, and many users see great potential in this functionality, most system vendors steer clear of such applications and use FF in a pure "master–slave" environment.[26]

MODBUS

MODBUS is a quasi-industrial standard developed some years ago by Gould-Modicon[26] and is a messaging service that may run on a variety of physical layers. Its two open versions are of interest here:

- Serial MODBUS, using RS-485 (or RS-422) as physical layer allows the connection of MODBUS devices to a PLC in a bus structure, whereby analog devices can be connected via gateways.
- MODBUS TCP (also known as MODBUS TCP/IP), using Ethernet as physical layer, allows the exchange of data between PLCs in different networks.

Table 4.16t summarizes the principal technical data. Figure 4.16u shows a typical system architecture.

The MODBUS protocol exchanges data in a master–slave relationship. Each slave has a unique address, and the data are identified by their location in the slave address register. Certain characteristics of the MODBUS protocol are fixed,

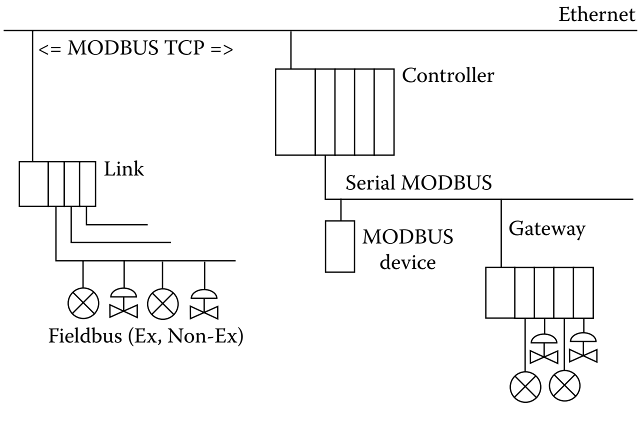


FIG. 4.16u
Typical MODBUS architecture.

such as the frame format, frame sequences, handling of communications errors, exception conditions, and the functions performed. Other characteristics are user selectable; these include transmission medium, baud rate, character parity, number of stop bits, and transmission modes (ASCII or RTU). The contents of the data carried by the protocol are also freely selectable, i.e., nothing is said about strings, integers, floating-point numbers, etc.

The MODBUS protocol controls the query and response cycle between master and slave devices. Only the master can initiate a transaction. A query and response may involve only a single slave, or it may be in the form of a broadcast, in which case the slaves do not answer. The query is contained in a frame that includes the address of the intended receiver, what this slave is to do, data needed to perform the action, and a means of checking for errors. The slave checks whether errors have occurred and performs the desired action. After the action is performed, the slave builds the response and returns it to the master. The master can send another message to any slave as soon as it receives a valid response or after a user-selected time interval. This "time-out" time has to be selected on the master device and depends on the slave response time.

The data can be exchanged in two transmission modes: ASCII (American Standard Code for Information Interchange) and RTU (Remote Terminal Unit). The major differences between them are the type of error check performed on the message and the number of characters used. MODBUS offers several read, write, and test functions, each identified by a code number. They are designed as control commands for sensors and actuators, e.g., coils, inputs, input registers, holding or output registers, diagnosis and test reports, programs, polling control, and reset. For MODBUS TCP the serial frame is simply inserted into the Ethernet data frame. In addition, not all codes are implemented (see Reference 27).

MODBUS has a large, worldwide, installed base. It remains popular because it is a very simple protocol that requires little programming effort to get it running. It enjoys extremely good support from system manufacturers both on the field and now on the control network side. The support for MODBUS TCP by manufacturers and users makes it a serious contender, alongside FOUNDATION Fieldbus HSE, to become the universal "Industrial Ethernet."

CONTROLNET

ControlNet is an open fieldbus standard to IEC 61158[14] that was developed as a proprietary protocol by Rockwell Automation. It is supported by the ControlNet User association.[29] Essentially, it is a control network that shares a common application protocol with DeviceNet[30] and Ethernet IP.[31] Its installed base is in factory automation, but like PROFIBUS DP it crosses the border to process automation. When used in the latter, FOUNDATION Fieldbus provides the H1 layer; the connection is made by a ControlNet/FF gateway that interfaces both physical layer and protocol. This solution may be short lived, however, since manufacturers are working on a full FOUNDATION Fieldbus solution for their controllers.

ControlNet is important because it has a large installed base in a broad range of applications in North America. Apart from isolated countries, however, it has no great significance for the rest of the world.

INDUSTRIAL ETHERNET

As mentioned in the introduction, Ethernet has become the dominant technology at the operations and in some cases the control network level. Detailed information and white papers are available at a number of company Web sites, e.g., References 31 and 32. A typical system architecture is shown in Figure 4.16v. Unlike PROFIBUS and FOUNDATION Fieldbus, it is not a designer solution for Process Automation; rather, it has been found to have a lot of characteristics required for the H2 layer: openness, inexpensive components, and support from major systems and software manufacturers.

Unfortunately, there is no "Industrial Ethernet" standard as such, although there is a supporting association.[33] Rather there is a collection of protocols that use the Ethernet physical layer, IP, and TCP. The most fully defined protocol is FOUNDATION Fieldbus HSE. MODBUS TCP is not a standard but is popular and in widespread use. Ethernet IP has some

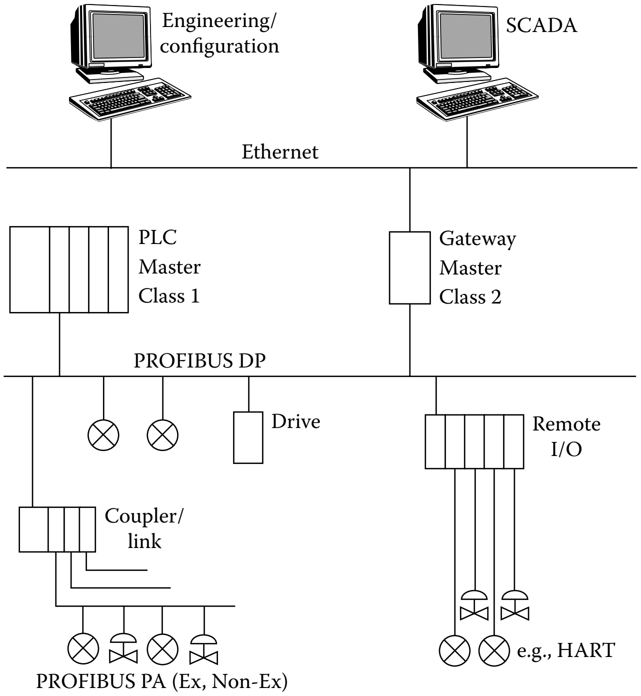


FIG. 4.16v
Typical Ethernet TCP/IP architecture using the example of PROFIBUS DP as control network.

support, and there are a number of proprietary systems with Ethernet TCP/IP backbones.

The protocols function by either packing their information in the Ethernet data frame or mapping proprietary applications on the TCP and IP protocols. Devices using different protocols cannot understand each other, but in some cases, e.g., with HSE and MODBUS TCP/IP, it is possible to use the same network to transport data. The matter is further complicated by the existence of two different Ethernet frames: devices receiving the wrong kind simply assume that a collision has occurred and reject the data. On the other hand, it is a fairly simple matter to map compatible systems, and for example, FF Function blocks for the interpretation of MODBUS data already exist, providing a excellent means of communication between FF and legacy systems.

Ethernet is accessed by means of the CSMA/CD method. This means that it is ideal for the transfer of non-time-critical information but requires an additional means of controlling medium access if it is to be used in deterministic control applications. Current equipment is usually capable of transmitting at 10 or 100 Mbit/s as required over 4-wire twisted pairs as well as fiber-optic cables. Commercial equipment is sufficiently rugged for a control room environment.

A frequently asked question is how long will it take Ethernet to penetrate the field. Four factors may have an influence:

- *Cost and availability of suitable chips.* This is probably a question of supply and demand; if there is a move toward Ethernet devices, the costs will decrease in inverse proportion to the numbers produced. In the long term, therefore, no hindrance will occur.
- *Information content and response times.* The standard TCP/IP frame has a minimum length of 72 bytes. Although the transmission of information across the network is fast, the protocol overhead increases the turnaround time and hence the response time of the system. At present, Ethernet is no match to a binary protocol such as AS-i, and in this case, would be definitely more expensive. The increase in transmission speed provides little advantage where process variables such as temperature and level are concerned, because the sensor response time places a lower limit on the polling interval.
- *(Intrinsic) powering of the devices.* Power supply over Ethernet is not an issue, but intrinsically safe power is. This is the main reason why IEC 61158-2 was developed for process automation. Any substitute must solve this problem before it can penetrate at field level. Since intrinsic power is not of interest to the home and office applications, it is unlikely that there will be much pressure in this direction.
- *Ruggedness.* Industrial connectors do exist, but other Ethernet equipment, such as cables and switches, is seldom designed for an industrial environment. An adaptation to, e.g., IP 67 or the equivalent NEMA rating would cut off the user from the benefits of a mass market and lead to a considerable increase in price.

Summing up, Ethernet at field level still seems far away, but devices operating in safe areas will be provided with Ethernet connections for integration at control level in the not-too-distant future. More information is contained in the recent ARC survey on the subject.[34]

NETWIDE DATA EXCHANGE

Where a process facility has a proprietary process control system installed, the problem of information exchange across the plant will have been solved by the system vendor. This is done, however, at the expense of freedom of choice regarding the sensors, actuators, and auxiliary equipment as well as flexibility in system design. Being locked into a particular system can also be expensive. Open systems, on the other hand, allow the user to choose the best devices and components or simply ensure that the best performance is being bought for the budget at hand.

For standard systems, data exchange between field devices and controllers is governed by the standard specification. There are many applications, e.g., a SCADA system or Microsoft Office program running on a personal computer connected to an Ethernet at the operations level, where information needs to be exchanged outside the bounds of the standard specification. This also applies when data from different systems operating with different protocols is to be gathered together at a single point.

Plantwide information exchange in an open system is often solved by the use of OPC servers. Since the support by Microsoft for XML, there is also considerable interest in this technology as a means of handling structured information in open networks. Finally, FDT/DTM, although still regarded skeptically by some parts of the I&C establishment, provides a protocol-independent technology for exchange of data for device configuration and asset management tasks. The paragraphs below describe these three technologies.

OPC Servers

OPC is a series of specifications based on basic PC standards and technology that provides open connectivity in industrial automation and enterprise systems. The standards are created, maintained, and adapted to new technologies by the OPC Foundation, a user group, which owns the technology and provides compliance testing.[35]

The original OPC standard, covering data access, was the result of collaboration between a consortium of system vendors and Microsoft. It defined a standard set of objects, interfaces, and methods that provided a simple means of ensuring interoperability of process and factory automation applications. This framework allowed a large number of software products to be developed that ensure seamless data exchange throughout the factory network.

Before the development of the OPC specification, a special driver had to be written if one application on a network, e.g., a controller, was to exchange data with another, e.g., a human-machine interface (HMI) application. Since no two controllers handled data in exactly the same way, an HMI supplier had to supply a driver for every controller on the market. In practice this did not happen, so users were restricted in their choice. By installing standardized OPC interfaces in both components, only one driver was required and data could be exchanged in a client (HMI)–server (controller) relationship. This technology is not restricted to controllers, however, and other control devices such as remote I/Os, gateways, FF linking devices, or even field devices can publish data in the same manner.

The benefit to users is flexibility and lower costs. Software components can be chosen on the basis of their features and performance rather than a need to connect to a particular device. Quality is improved since components are subject to independent testing. In addition, because a component is reused many times, it undergoes continuous improvement.

The original standard regulated the acquisition of process data, but it was quickly realized that other types of data could benefit from standardization. At the time of this writing, seven specifications have been completed or are in development, covering many aspects of data exchange, from data access, through alarm and event handling, to the exposure of XML data in field devices. Cooperation between the OPC and field device tool-joint interest group (FDT-JIG) in the field of device configuration and asset management standards has also been announced recently.

XML

XML stands for eXtensible Markup Language and is a text markup language that in functionality and simplicity lies somewhere between its parent SGML, which was standardized in the 1970s, and HTML. It is designed to carry structured textual information independent of style information. HTML differs, because style information, e.g., font and font size, is carried by predefined tags within the file. XML does not have predefined tags and uses a Document Type Definition (DTD) or an XML Schema to describe the data held in an XML file.

Readers familiar with DD technology will immediately recognize the similarities between the two technologies: a TXT file, DD or XML, carries the data and the Type Definition, Device or Document, respectively, interprets them.

Separating data content from the way in which it is displayed ensures that any change to one does not affect the other. This leads to several interesting applications in process automation:

- XML data can be published in HTML pages provided by a Web server and read by a local browser.
- XML data can be transmitted together with its DTD across the Internet or control network and read directly by, e.g., Microsoft Office programs such as Excel that support XML data exchange.
- An OPC server is capable of taking the XML data and storing it for use by a HMI or SCADA application.
- The simple nature of the file form means that it is easier to share data between different components within a system.
- The structured nature of the data means that it is easy to store them in files or databases. Storage and retrieval programs are easily written and generic applications can be used to display the data.

XML can also be used for exchange of data between control and enterprise systems as well as for business-to-business (B2B) applications. Thus, it is being considered as a data carrier in e-business applications and e-purchase based on the exchange of electronic data sheets. The latter foresees the automatic generation of a measuring point specification from an engineering tool and in response a corresponding device data sheet from an electronic catalogue. At each stage of the purchase process, data are added to the sheet, e.g., serial number, device attributes, so that on delivery of the physical device, its full electronic description can be loaded into the engineering or asset management system.

Finally, moves are afoot to bring XML into the field devices themselves. This would mean that eventually, all the parameters and information held within a fieldbus device would be structured in the same manner, with the same tags, independent of vendor and protocol. Only the communication across the network would be protocol dependent, and when the features of FDT technology are considered, even this will play a minor role in the future.

FDT/DTM

With digital communication, field devices must be integrated into the control system before they can be operated. To this end the manufacturer supplies a device description with each device, which tells the operating software which parameters the device contains and how incoming data are to be interpreted.

Unfortunately, device integration is seldom a matter of plug and play. Although device descriptions are standardized for HART, PROFIBUS, and FOUNDATION Fieldbus, it is often the case that additional data or proprietary formats are required for particular systems.

Quite detailed knowledge of the system is required if the device is to work properly the first time. Moreover, it is seldom that all device functionality can be integrated into the third-party system. Even when system and device vendors cooperate to optimize interoperability, updates to software can cause devices to fail because the updated device descriptions are not immediately available. In addition, although the device appears in a uniform environment, it may be totally different from the one described in the device operating manual.

This is essentially the same as the HMI/controller driver issue, which was solved by the OPC initiative. The prime mover in the case of FDT (field device tool) was the PROFIBUS User Organisation. The idea was to define a set of protocol-independent interface standards for the exchange of device configuration and later asset management data throughout the control network. Each intelligent device in the network is supplied with a so-called DTM, which contains both the parameters and the operating interface for full device configuration. The DTMs are programs that run in an FDT frame program. There are two types: Device DTMs for field devices and CommDTMs for communication devices, e.g., gateways or remote I/Os. Transparent communication devices requiring no configuration, e.g., couplers, require no DTM.

The standard ensures that all DTMs run in all FDT frame programs. Responsibility for the production of the DTM lies with the manufacturer, who together with the user has the guarantee that it will run in all systems that are FDT compliant. The DTM is generated from the existing HART, PROFIBUS, or FOUNDATION Fieldbus device descriptions (DD or EDD). This means that even if a manufacturer does not support the standard, DTMs can be obtained for its devices. An alternative, with restricted operability, is to use generic DTMs for a particular protocol. Since FDT essentially frees device operation from protocol-dependent factors, it also provides seamless data flow across networks in which different protocols are operating.

The year 2002 saw the establishment of the FDT-JIG,[36] which has now taken over the responsibility for developing, maintaining, and testing compliance to the standard. At the time of writing, over 30 companies worldwide are actively involved in the group, many with DTMs or FDT frame tools on the market. CommDTMs are available for Ethernet, PROFIBUS, and HART components; MODBUS may also be a possibility in the near future. Device DTMs are available for HART and PROFIBUS. The specification for FOUNDATION Fieldbus is complete; its feasibility was demonstrated at the ISA shows in 2002 and 2003, and products were shown at the Inter Kama 2005 show.

CONCLUSION

Having seen the progress made over the last 10 years in fieldbus technology, it would be appropriate at this point if the authors could look into the future. Unfortunately, this is not possible. There are, however, certain trends observable that will probably have a great influence on the control system of the future:

- Field devices are becoming increasingly important as a source of information for plant and enterprise asset management systems. The standards are already in position; vendors will adapt their systems to allow free flow of information from field to enterprise level.
- As plant and enterprise asset management grow in importance, the influence of information technology on control system design will increase. The databases behind the system will become standardized, and proprietary systems will move towards openness.
- As protocol-dependent factors are eliminated from the system database, the role of the protocol in control and asset management systems will be reduced. OPC, XML, and FDT are already eroding its influence; cooperation between them will eliminate it completely.
- Web technologies are beginning to appear in remote monitoring applications. The embedding of sequential or continuous control in “Plant Access Points” utilizing this technology will open a new era of “Microcontrol,” where size and control capability will be exactly matched to the task at hand.

There are, of course, a great number of other factors that can and will influence the development of both networks and the protocols that are used on them. Perhaps the future is different—but we will have to wait until the next edition of this handbook before we will know for certain.

References

1. O’Brien, L., “Process Fieldbus User Survey,” ARC Advisory Group, arcweb.com, 2002.
2. ISA S88.01 Series, “Batch Control,” Instrument Society of America, www.isa.org, 1995.
3. IEC 61512 Series, “Batch Control,” International Electrotechnical Commission, www.iec.ch, 1997.
4. ISA S95.01 Series, “Enterprise-Control System Integration,” Instrument Society of America, www.isa.org, 2000.
5. IEC 62264 Series, “Enterprise-Control System Integration,” International Electrotechnical Commission, www.iec.ch, 2003.
6. ISO/IEC 7498-1 “Information Technology—Open Systems Interconnection—Basic Reference Model: The Basic Model,” International Standard Organisation/International Electrotechnical Commission, www.iec.ch, 1994.
7. ANSI/EIA/TIA-568, “Commercial Building Telecommunications Wiring,” American National Standards Institute, www.ansi.org, 1991.
8. IEC 61158-2, “Digital Data Communications for Measurement and Control—Fieldbus for Use in Industrial Control Systems, Part 2: Physical Layer Specification and Service Definition,” Edition 3, 2003, International Electrotechnical Commission, www.iec.ch.
9. Barton, C. S.,“PLC Proprietary and Open Networks,” Section 4.2, *Instrument Engineers’ Handbook, Vol. 3, Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002.
10. Berrie, P.G., and Korsten, K., “Designing PROFIBUS PA and FOUNDATION Fieldbus Segments,” Section 4.16, *Instrument Engineers’ Handbook, Vol. 3, Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002.
11. Thompson, L. H., *Industrial Data Communications,* ISA Resources for Measurement and Control Series, Instrument Society of America, 1997.
12. IEEE 802.2/ISO 11802/2, “Bus Arbitration by CMSA/CD,” International Standards Organisation, iso.org.
13. IEEE 802.3/ISO 11802/3, “Bus Arbitration by Token Passing,” International Standards Organisation, iso.org.

14. IEC 61158 Parts 1 to 6, "Digital Data Communications For Measurement And Control—Fieldbus For Use In Industrial Control Systems," 2003, International Electrotechnical Commission, www.iec.ch.
15. EN 50170 Part 4, "General Purpose Field Communication System — FOUNDATION Fieldbus," CENELEC, www.cenelec.org.
16. AS-i Users Consortium, www.as-interface.com.
17. Pratt, W. A., Jr., "HART Networks," Section 4.11, *Instrument Engineers' Handbook, Vol. 3, Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002.
18. HART Communication Foundation: www.hartcomm.org.
19. EN 50170 Part 2, "General Purpose Field Communication System — PROFIBUS FMS, PROFIBUS DP, PROFIBUS PA, CENELEC," www.cenelec.org.
20. PROFIBUS User Organisation, Haid-und-Neu-Str. 7, 76131 Karlsruhe, Germany, www.profibus.org.
21. Vitturi, S., "Fieldbus Networks Catering to Specific Niches of Industry," Section 4.16, *Instrument Engineers' Handbook, Vol. 3, Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002.
22. Berrie, P. G., and Füchtler, L., "PROFIBUS PA," Section 4.13, *Instrument Engineers' Handbook, Vol. 3, Process Software and Digital Networks,* Boca Raton, FL: CRC Press, 2002.
23. Johannsmeyer, U., FISCO, PROFIBUS-PA User and Installation Guideline, Profibus User Organisation, www.profibus.org / PTB Report W-53, Investigation into the Intrinsic Safety of Field Bus Systems.
24. FOUNDATION Fieldbus: www.fieldbus.org.
25. Caro, R. H., "Foundation FIELDBUS Network," Section 4.12, *Instrument Engineers' Handbook, Vol. 3, Process Software and Digital Networks*, Boca Raton, FL: CRC Press, 2002.
26. Neto, E. de S., and Berrie, P. G., "Who's Afraid of Control in the Field?" FF End User Conference, Singapore, 2003.
27. MODBUS Protocol, Reference Guide M-MBUS 300, Gould-Modicon, www.modicon.com, see also user group Internet site below.
28. MODBUS User Organisation: www.modbus.org.
29. ControlNet International: www.controlnet.org.
30. Open DeviceNet Vendor Association: www.odva.org.
31. Ethernet IP Organisation: www.industrial-ip.org.
32. Ethernet Users Guide, Lantronix Tutorial: www.lantronix.com.
33. Industrial Ethernet Association: www.industrialethernet.com.
34. Ethernet at the Device Level Worldwide Outlook, ARC Advisory Group, 2003, www.arcweb.com.
35. OPC Foundation: www.opcfoundation.org.
36. FDT Joint Interest Group: www.fdt-jig.com.

Bibliography

Anderson, N. A., *Instrumentation for Process Automation and Control,* 3rd ed., Boca Raton, FL: CRC Press, 1997.

Berge, J., *Fieldbuses for Process Control: Engineering, Operation, and Maintenance,* Research Triangle Park, NC: ISA, 2001.

Boyer, S. A., *SCADA (Supervisory Control and Data Acquisition),* 3rd ed., Research Triangle Park, NC: ISA, 2004.

Caro, D., *Automation Network Selection,* Research Triangle Park, NC: ISA, 2004.

Goldfarb, C. F., *XML Handbook,* 5th ed., Englewood Cliffs, NJ: Prentice Hall, 2003.

The HART Book, Analog Services Inc. www.analogservices.com.

Marshall, P. S., and Rinaldi, J.S., *Industrial Ethernet: A Pocket Guide,* 2nd ed., Research Triangle Park, NC: ISA, 2005-(online publication).

Mitchell, R., *PROFIBUS: A Pocket Guide,* Research Triangle Park, NC: ISA, 2004.

Morris, A. S., *Measurement and Instrumentation Principles,* 3rd ed., Oxford: Elsevier, 2001.

Simon, R., et al., *Field Device Tool—FDT, Universal Field Device Integration,* Munich: Oldenbourg Industrieverlag, 2005.

Sterling, D. J., and Wissler, S. P., *The Industrial Ethernet Networking Guide*, Thomson Delmar Learning, 2002.

Thompson, L. H., *Industrial Data Communications,* 3rd ed., ISA Resources for Measurement and Control Series, Research Triangle Park, NC: ISA, 2002.

Verhappen, I., and Pereira, A., *FOUNDATION Fieldbus: A Pocket Guide,* Research Triangle Park, NC: ISA, 2002.

Weigmann, J., and Kilian, G., *Decentralisation with PROFIBUS DP: Structure, Configuration and Use of PROFIBUS DP with SIMATIC S7,* 2nd ed., Erlangen, Germany: Siemens, 2004.

ACRONYMS

CSMA/CD	Carrier sense medium access/collision detection
DCS	Distributed control system
DHCP	Dynamic host configuration protocol
DTM	Device type manager
FDT	Field device tool
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
IP	Internet protocol
ISA	Instrument Society of America
ISO	International Standards Organisation
MES	Management execution systems
MIS	Management information system
OSI	Open systems interconnection (model)
PCS	Process control system
PLC	Programmable logic controller
SCADA	Supervisory control and data acquisition
SNMP	Simple network management protocol
SNTP	Simple network time protocol
STEP	Standard for the exchange of product model data
TCP	Transmission control protocol
UDP	User datagram protocol

4.17 Human–Machine Interface Evolution

G. KEVIN TOTHEROW (2002, 2005)

Partial List of HMI Software Suppliers:	(Software is sold based on tag counts or I/O count. System sizes range from 64 tags to unlimited system tags, and most vendors have a variety of options within those sizes.) Adroit (www.adroit.co.za) Citect (www.citect.com) GE Automation Cimplicity (www.geindustrial.com/cwc/gefanuc/software.html) GE Automation Intellution (www.geindustrial.com/cwc/gefanuc/software.html) Iconics (www.iconics.com/) Rockwell Software (www.software.rockwell.com/navigation/products/index.cfm) Siemens (www.ad.siemens.de/hmi/html_76/products/software/wincc/index.htm) USData (www.usdata.com/usdata.html) Wonderware (www.wonderware.com)
Alarm Management Software Supplier:	TiPs (www.tipsweb.com/)

INTRODUCTION

The devices that operators use to control process equipment have changed dramatically since the beginning of the digital age. In some cases the operator is surrounded with multiple human–machine interfaces that more closely resemble NASA's mission control center than the bench boards and instrument panels of the past. In other cases the operator control room is becoming less significant as operators are using handheld devices and Web pages to access the process information and control systems. However, we still see much of the information presented on digital displays with images of the control stations, pushbuttons, and trend charts resembling bench-board equipment used 50 years ago. Further, the digital displays are often installed and configured by technicians and system integrators without regard to supporting the operator's entire job, making the modern controls less intuitive than the bench-board controls were (Figure 4.13h). Notwithstanding these issues, the digital display HMI has allowed a single operator to monitor and control thousands of process control loops from a single station (Figure 4.5c). The latest advances in HMI technology using inexpensive hardware, thin-clients, handheld computers, and cell phone technology have freed the operator from confinement to a particular location for control and information systems. Understanding the evolution of the operator interface and the changing role of the operator in a plant allows suppliers to build a better HMI product, plant engineers to pick the best technology, and integrators to make the operator more effective and efficient.

The modern HMI is at the evolutional convergence of two different development paths. One development path is that of the operator console of the distributed control system (DCS). The other path is the development of the HMI for the programmable logic controller (PLC). This makes a chronological discussion of HMI evolution somewhat difficult without getting the reader confused about which HMI is being discussed. Therefore, this author will discuss the basic functions of an HMI and then attempt to describe the evolution of DCS and PLC human–machine interfaces separately. Toward the end of this section, once again the discussion of control rooms, operators, and the future will apply to all HMI systems.

FUNCTIONS OF THE CONTROL SYSTEM HMI

Operators have always had HMIs, with some of the oldest being the handle of a valve or the faceplate of a pneumatic controller. Although the term HMI can refer to any type of device that allows a person to manipulate a machine or process, the term as used in process and manufacturing usually refers to the display, computer, and software that serve the specific role as the operator's interface to a controller or control system. The same display, computer, and software used for another purpose do not make an HMI. An analogy is a taxi. A taxi is an automobile, boat, coach, or rickshaw that serves a specific function of transporting a customer from one location to another. Use the same vehicle in another function and it is not a taxi.

This section uses the term HMI to describe the computer hardware and software that allow interaction between the operator and the controls.

Many HMI product suites provide functionality beyond HMI. These system functionalities will be discussed, but we will always try to use a term other than HMI to describe these devices. Further clarification will be provided as necessary to describe specific HMI hardware, software, or a specific type of HMI. Table 4.17a is a quick reference of the specific "HMI" terms used in this section.

The operator interface to the plant process, whether bench-board devices or HMI, has four primary functions. The first function is to provide visualization of process parameters. The second function is to enable interaction or manipulation of the process or machine. Third, the interface is to provide alarms and event notification to the operator that some area of the process should be monitored more closely or that some type of corrective action is required. Finally, the fourth function of the operator interface is to provide a method, such as a trend chart, to allow the operator to understand where the process is going and how fast. Other uses of the operator interface should be secondary to these.

TABLE 4.17a
Definitions of HMI Terminology

Term	*Definition*
HMI	Human–machine interface. Computer hardware and software that allows operators to interface with the control system.
CRT HMI	Cathode ray tube HMI. The term is specifically used where there is a transition from hand stations to the HMI.
DCS HMI	Proprietary HMI from DCS vendors that is optimized to work with the DCS.
GUI	Graphical user interface. General computer term for software that allows graphical interaction with the computer.
Open HMI	HMI built on PC hardware and Windows operating system for general use. Not specific to any hardware.
Operator's window	Open HMI that incorporates information systems and controls.
PLC HMI	Proprietary HMI from PLC vendors that is optimized to work with the PLC.

Visualization and Control

The first two responsibilities of the HMI are to provide the operator with a view of the process and a way to control the process. The panel board was the operator's window to the process through the various control stations and the analog instruments before the digital age of controls. The controllers themselves were mechanical devices that mounted to the bench board or the control panel.

Most of the controller was hidden from view, and the part of the controller that was above the bench board for the operator to see and use was the controller faceplate. This faceplate displayed the process variable and provided access to change the mode, set point, or output to the loop. The controllers were mounted on the panel in logical clusters based on the process group. Figure 4.17b shows a power

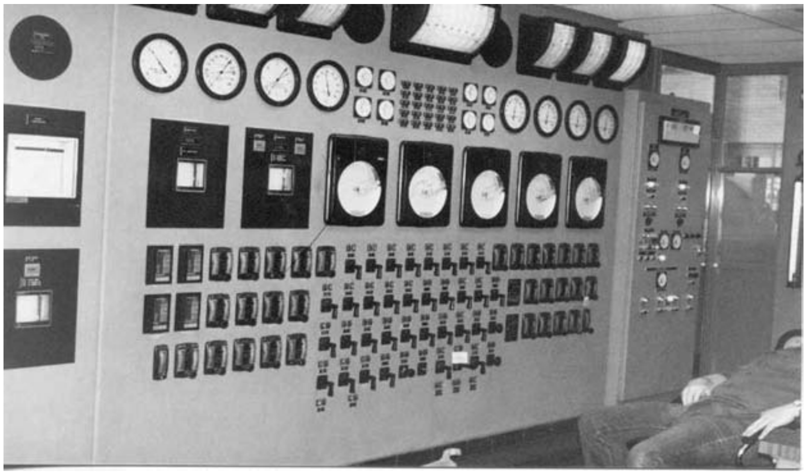


FIG. 4.17b
Power boiler control panel circa 1984, picture by Jim Mahoney. Nineteen hand stations are in the bottom center of the panel. Four are digital stations for a Bailey Net90 DCS, the other fifteen are pneumatic controllers.

FIG. 4.17c
Group display of hand-station faceplates on a Bailey Net90 HMI circa 1985. Photograph provided by Jim Mahoney.

boiler control panel with pneumatic controllers to the right and a few electronic trend units and DCS hand stations on the left.

The HMI provides the same functionality to monitor and control the process in a newer, highly customized, and constantly evolving package. The HMI makes it possible to increase the number of control loops and process information presented to the operator many times higher than could be mounted in a control panel. A configurable display also gives the ability to present the information to the operator in many and various ways. Figure 4.17c shows a very simplistic graphical operator interface with the controller hand stations implemented on the HMI.

Process Alarming

The third responsibility of the operator's HMI is generation, annunciation, and manipulation of process and system alarms. Prior to the digital control age, generation of the alarm was the sole function of a switch in the field to determine whether a process parameter such as a pressure or temperature exceeded specifications. A backlit light box with translucent window panels with the alarm written on the panel, called an annunciator, is the operator's indication of the alarm. The discrete alarms were wired from switches to the annunciator to provide a visual and audible alarm for serious process upsets or failures.

The annunciator allowed the operator to manipulate the alarm by silencing the audible portion or acknowledging the alarm. Alarming process parameters with an annunciator panel is very straightforward—install and calibrate the proper switch in the process, and wire the switch contact to a power source and to the annunciator panel. Of course, the alarms were expensive because of the required field installation of switches and wiring. Also, there was a finite amount of wall space in a control room to install annunciator panels, but on the positive side there were very few nuisance alarms.

The modern HMI generates process alarms from the analog process variable or a discrete input from the control system. The HMI displays, stores, and manages the alarms. Process switches are not needed to provide process alarms, and the number of alarms that can be generated is virtually unlimited. The modern HMI makes process alarms cheap and easy to implement, thus allowing engineers and managers to create nuisance alarms at an unprecedented rate.

Management of process alarms through silencing, acknowledging, and summarizing is now a primary function of the HMI. Advanced alarm management, organization, and analysis are distinguishing features among HMI products.

Trending

Long before the digital control system age, operators depended on pen recorders to tell then what the process was currently doing and also the trend of the process variable. Analog pen recorders were set up to record the process at variable rates to make the chart show the process trend for a certain number of previous hours. The resolution of the recorder was inversely proportional to the amount of time the recorder captured on paper. Getting the proper balance of time and resolution on the recorder was very important to helping the operator understand the direction and rate of change of the process. Recorder paper was changed as necessary and kept on file to enable process engineers and managers to look at the history of the control parameter.

The first digital display HMI immediately had provisions for keeping a history and the capability to provide the operator with trending to replace the chart recorder. The trending is very similar to the chart recorders—the data are collected at a rate that will balance resolution with the total trend time.

However, unlike with the chart recorder, the operator cannot change the paper and store an unlimited amount of process history.

Most of the control system and HMI companies abdicate the long-term historical systems to others, choosing instead to provide medium- to short-term trending for operators only. This decision is still reflected in many HMI products, where operator history and trending are an integral function of the HMI, but long-term history for management or regulatory compliance is not.

DCS CONSOLES

The first companies offering digital display operator consoles for process controls are the DCS companies, beginning with Honeywell in 1975 with the TDC 2000.[1] This was followed closely by several other major manufacturers of control products. The introduction of a graphical user interface (GUI) display as the operator HMI is a technology leap destined to change the way many industries operate. Yet many operators and some engineers feel more comfortable operating the process from a digital representation of their own older process controller faceplate than from a graphic of the process.

The faceplate is familiar to the operators, provides all of the detail about the controller and the control loop, and is easily implemented by engineers. The problem with using the faceplate from the HMI is that the operator often had a much better intuitive feel for finding the right controller and responding quickly to upsets with the bench-board controls as compared to navigating through various displays of screens to find the right controller.

Custom process graphics, or mimic displays, showing representations of pumps, motors, vessels, agitators, heaters, and other process equipment along with good overview displays and standards for navigation make controlling large processes from the HMI display much more intuitive for operators.

In 1983, Savannah Electric and Power Company bought a Bailey Net90 DCS to replace pneumatic boiler controls and, like many others, chose to only replace pneumatic controllers on the bench board with rows of digital hand stations connected to the DCS. Figure 4.17d shows digital hand stations for a Bailey Net90 mounted in a panel beside pneumatic controllers.

The Bailey OIU (operator interface unit) displaying process graphics and graphical representations of the hand station faceplates sits beside the operator bench board. The logic of the transition is that the operators are accustomed to controlling the process by controller faceplates, and process graphics are new and will take time to learn. The expense of redesigning the control room to make the HMI useful as the primary interface is also a large factor in a phased approach to introducing the HMI.

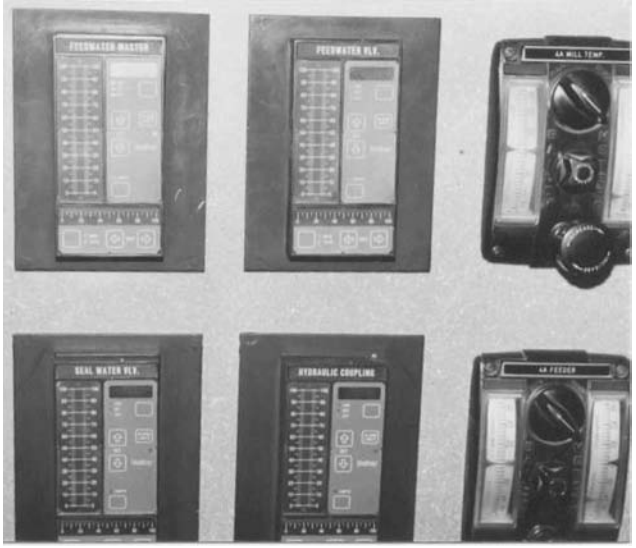


FIG. 4.17d
Four Bailey Net90 DCS hand stations mounted beside two pneumatic controllers, circa 1984. Photograph courtesy of Jim Mahoney.

At Savannah Electric, the HMI offers more information than the hand stations show, and the operator can sit in one position to control and monitor the boiler and auxiliaries. Eventually, the operators use the HMI as their primary control interface.

DCS Console Graphic Standards

The early DCS console was built specifically for a single control system. Controller faceplates were standard graphical objects that looked and functioned like bench-board hand stations. The most popular early displays were the faceplate, group, and area.[1] The faceplate is the graphical representation of the hand station with all of the process information and loop control. The group display is a display showing four to eight faceplates in a little less detail. The area display is an overview of many process variables that allows operators to quickly see whether the processes are stable. DCS mimic displays are the highly customized displays that operators like.

Mimic displays show graphical illustrations of the process with process information. Good mimic displays and good overview pages that show only the most pertinent information about large process areas make the operators more efficient. The HMI is so flexible and easy to change that companies quickly realized that they needed to adopt at least basic standards for building these custom displays to ensure consistency between engineers and projects.

Suddenly, industry and plant questions arose as to whether a red motor indicated the motor is energized or did it mean the motor is stopped? Does a green valve indicate open, closed, or deenergized? The standards that most companies adopted include object size, conditional colors, process colors, screen navigation, and screen density guidelines. Organizations such as the ISA developed standards for operator displays that include the items above and many more.[3]

The standard displays for most DCS operator consoles include:

- Process graphic displays
- Faceplates
- System diagnostic graphics
- Controller tuning pages
- A list or hierarchy of displays
- Alarm summary page
- Trending pages

Screen/Page Navigation and Item Selection Navigating between DCS displays on first- and second-generation systems and selecting the controllers on the display is highly customized with custom keyboards to make the operator efficient. Generally, the only "mouse" in the control room at this time (early to mid-1980s) was the real one that shorts out control loops while eating insulation off the cables. The common methods of screen navigation included:

- *Touch screen.* Touch the object to select it. This included paging objects.
- *Direct screen reference (DSR) numbers from one display to the next.* The DSR selected the display.
- *Direct screen reference (DSR) numbers from the display list.* Each display on the display list was assigned a DSR number. Selecting that DSR from the display list would take the operator to the display.
- *Configured forward and backward display keys.* Each display could have configured forward and backward displays to navigate logically.
- *Show detail keys.* Devices selected could bring up a detailed display of the object or controller.
- *Defined function keys.* Console keys that change to preconfigured display.
- *Trackballs.* The trackball was mounted in the custom keyboard and manipulated a pointing device on the display.

More vendors used a mouse or trackball and fewer use custom keyboards with later revisions of the DCS operator console, although the best operators swear they are more efficient with the older custom keyboards and DSRs on the display.

DCS HMI Redundancy

Many managers and engineers were worried about operating potentially dangerous processes through a computer when the DCS first arrived on the scene. DCS vendors worked hard on the technical issues of redundancy and selling the concept of the HMI to industries that liked the idea of direct manipulation of hand stations on controllers.

First-generation DCS consoles were built of a minicomputer and multiple workstations that were dumb terminals (thin-clients) of the computer. Installing at least two computers or sets of console electronics, and using two different power sources for the electronics and the operator terminals achieved a comfortable level of redundancy. See Figure 4.17e again for the "typical" DCS HMI arrangement. Notice in the diagram that the operator has two workstations from two different console electronics. If one power source fails, or one HMI computer fails, the operator loses only one workstation.

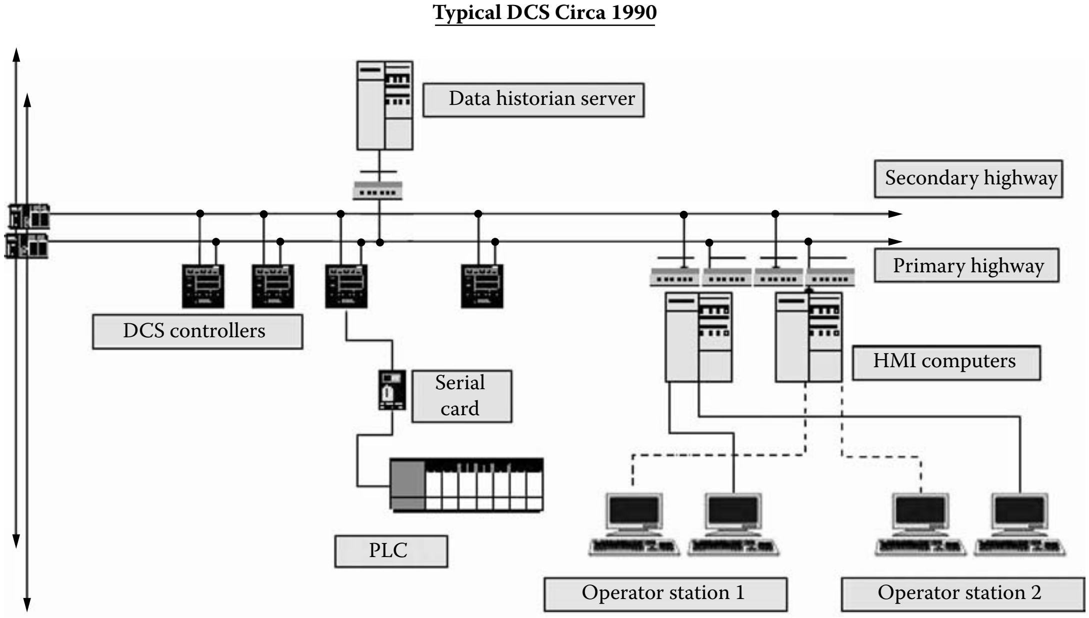


FIG. 4.17e
This is a typical DCS system from the early 1990s.

The first-generation consoles from vendors such as Fisher Controls and Bailey Controls could be configured with up to 2,500 points. Second-generation consoles could handle 10,000 points. The HMI had to provide operators with an efficient interface to the DCS with the number of displays and point count necessary for the application yet still refresh quickly. A common console specification demands a maximum screen update time of 5 seconds and a refresh rate of 3 seconds. The console is a very large part of the selection criteria for a DCS for many customers.

DCS HMI Chronological Evolution

1975–1985 The various HMI offerings use proprietary hardware and software and, like other systems of that generation, hardware and software are inseparable. Nowhere is this more evident than in the HMI keyboards. The keyboards contain rows and rows of special fixed keys to start, stop, open, and close the selected devices and acknowledge alarms and silence horns. The keyboards also have configurable keys for display navigation and special functions.

Finally, the HMI includes function keys that change function with each display to help make the operator more efficient. Each DCS vendor has different ideas and standards for their HMI, but all approach the operator interface as the total solution to replace the bench board. The DCS HMI includes cabinets with room for mounting push buttons or telephones, wedge-shape tabletops to mount between HMI consoles to curve the console, matching tabletops for writing surfaces, and even chairs.

1986–1995 Later generations of these proprietary HMI products show incredible integration with other components in the "system." Most DCS HMI products incorporate distributed alarming, console redundancy, and a common configuration database for controllers and consoles. System ease of use is a theme. One price for this tight integration of proprietary systems is that by the second or third generation of DCS offerings, communication problems exist between vintages. The first-generation HMI cannot always communicate with second- or third-generation controllers and third-generation HMI cannot always communicate with first- or second-generation controllers.

The proprietary HMI communicates only with the DCS controllers. A company that employees PLCs to do motor logic has to interface the PLC to the DCS at the controller level to have the DCS HMI stop and start the motor. Figure 4.17e shows a typical DCS interface to PLCs and historians. This adds greatly to the cost of the system and adds unnecessary load on the DCS highway just to give the operator access to the motor controls through the HMI.

1996–2002 This is generally the first generation of HMI from DCS vendors that is based on the Microsoft Windows™ operating system and uses commercial off-the-shelf hardware where possible to keep costs lower. Proprietary may not be the correct terminology to describe this generation of DCS HMI products, but they are very often highly optimized with their own proprietary DCS controllers and much less capable of working with other controllers.

This generation DCS HMI uses inexpensive PC equipment and operating systems familiar to their clients but is not generally marketed as an open system to other controllers. Many of the DCS companies have trouble building a console on the Windows operating system that is as functional or as reliable as their older consoles. DCS companies control access to competitors' controllers from their HMI, and open operator interfaces are not generally part of their strategy even when the DCS is built on open-system HMI software.

Emerson's Delta V hybrid DCS uses Intellution's Fix software as the basis for their DCS console. However, Emerson maintains a single DCS configuration database that requires a third-party controller to interface at the controller level rather than allowing the HMI to access the controller directly. Similarly, Moore Products APACS chooses to build their Windows NT console on Wonderware's InTouch HMI software

2003– Today, DCS vendors are on their second or third generation of Windows-based HMI consoles. The products are better, more reliable, and more open. The DCS consoles are competitive with open-system HMI software as far as features and interfacing capability. Thin-client capability, Web client operator stations, and handheld technology are all possible with many of the systems. The distinguishing factor for many of the DCS consoles is the same as with open-system HMI products and that is ease of configuration, reusable configuration objects, and migration of existing engineering work. The converging of features and capability between the DCS console and the independent open-system HMI that grew up from the PLC world is nearly complete.

THE OPEN HMI

Open HMI development was very different from that of the DCS HMI. Since the early PLC market was replacing relays and timers, it was utilized in many applications without an operator interface of any type or was connected to pushbuttons and lights. PLC systems were relatively inexpensive, so the HMI had to be inexpensive too. There was little market for an expensive operator interface, as in the DCS world. However, the major PLC manufacturers saw the immediate need for industrial operator interfaces that could replace the pushbuttons and lights and give more information to operators. But it was several small companies that developed HMI software for personal computer hardware that really vitalized open HMI development. A distinguishing factor in the development of the open HMI vs. the DCS console was that the open HMI is not an integral part of a proprietary control system.

Open HMI Display Standards

Open HMI software is extremely versatile, providing a cost-effective operator HMI to a stand-alone machine or a networked series of many operator stations for a large system of many controllers. That versatility works against the open HMI in terms of configuration standards. The open HMI software companies cannot foresee how the HMI will be used by the purchaser. The open HMI companies follow the Microsoft business model of supplying shrink-wrapped software for their customers to build into a functional HMI for their particular installation. The organization of process information and the configuration standards for the open HMI falls entirely to system integrators, engineers, and end users—with mixed results.

Information on the ergonomics of human factors such as position of the chair, screen glare, and others is abundant. There is less information readily available on specific guidelines to help engineers organize the information and build an effective HMI. The ISA has a document, ISA-TR77.60.04-1996, for fossil fuel plants that provides some guidelines that are common to all process industries. The problem is that many companies implementing medium to small projects never dedicate resources to human-factors engineering and establishing HMI standards.

Documentation

Documentation is defined in this section to include installation, configuration, backup, and change control documentation. Documentation is a weak point for the open HMI when compared to any proprietary system. The first weakness is that the hardware, software, network, and configuration are individual components from different companies. The engineers who install and configure the open HMI must provide the documentation for the HMI system. The open HMI software companies provide software documentation, configuration training, and configuration examples and tips to their users. The software company cannot know the system hardware or the configuration application, leaving the end user and engineer responsible for documentation of the system.

The second area of documentation weakness for the open HMI is the configuration backup and change control documentation. The DCS prescribed database backup procedures, and most of the configuration systems automatically documented database changes. The configuration backup of a PC file such as an HMI application is intuitively easy, but change control documentation is not. Third-party software products such as MDT's AutoSave and GE's Cimplicity Manager fill this void with software that limits access to the HMI or PLC configuration and maintains a log of the changes that are made to the HMI and PLC software packages.

Open HMI Evolution Chronology

1980s Early open or independent HMI products are written to operate on all of the various computers and workstations of the day on all of the various operating systems. It is the office PC, however, that provides the basic HMI capability in an inexpensive package. Microsoft releases the Windows 3.1 operating system for the PC in the 1980s, and with each successive release it is more apparent that Windows is the platform of choice for the open HMI. This is the time when many of the first computer-based HMI products such as Intellution, USData, and Wonderware began to fill a void in manufacturing control systems.

1990–1997 The PLC is now a very powerful logic controller capable of PID control and processor redundancy. The low price of the PLC and open HMI make them competitive in the process industry over most of the "mini-DCS" platforms that are available. Technology increases and low cost continue to bring the PLC and open HMI into larger service in the process industries and particularly in supervisory control and data acquisition (SCADA) applications. The open HMI continues increasing in functionality compared to the DCS HMI, although many of the major DCS players scoff at suggestions of an HMI based on an Intel PC running Windows.

Microsoft releases Windows NT 3.5 in 1994, Windows NT 3.51 in 1995, and Windows NT 4.0 in 1996.[4] Most of the open HMI companies upgrade their software to the current releases from Microsoft, and with each release come new features and arguably better stability. PLC companies release open HMI software products at this time to capitalize on the burgeoning market for HMI products.

Major DCS vendors stop scoffing at the suggestion of utilizing PC hardware and the Windows operating system toward the end of this time period and begin projects to create PC-based operator stations.

1997–2001 The major open HMI companies such as Intellution, Rockwell, and Wonderware continue to expand HMI functionality as computer hardware becomes more capable of providing statistical process control (SPC), recipe management, soft control capability, and a host of other capabilities.

The open HMI begins to look even more impressive for process systems when several of the leading HMI companies introduce software supporting client–server architecture and distributed alarming. The open HMI can act as a common, integrated HMI for all control systems at a plant. Figure 4.17f shows the open HMI accessing data from multiple tag servers.

Through acquisition and partnering the open HMI, companies also begin the drive to integrate their HMI with a suite of software products to provide better information to the entire corporation. Process data historians, execution systems, Web-enabled clients, downtime tracking systems,

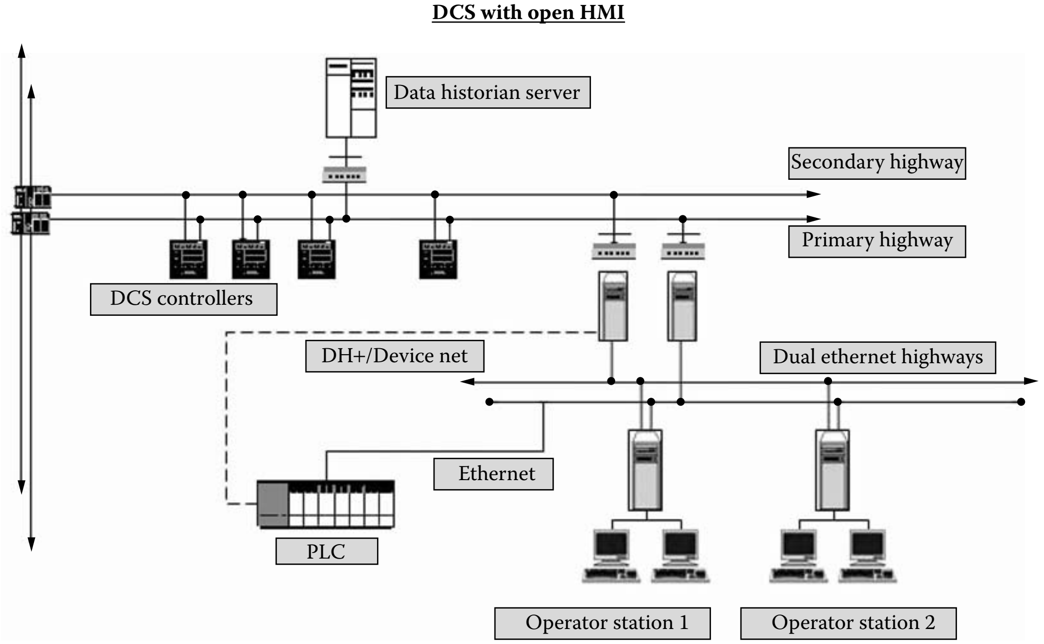


FIG. 4.17f
A DCS using Open HMIs. Note that the PLC has redundant communication directly to the HMI.

statistical quality control (SQC), and other systems are bundled with the large-system HMI providing a coordinated mid-level information system to just about any control or process information system. Rather than the HMI being an integral part of a control system like the DCS console, the open HMI is now part of a layer of plant-floor information systems above the control systems.

Larger automation companies purchase many of the independent HMI software companies during this time period.

2002 to Present Since about 2002, the development and capability of the DCS console and the open system HMI have virtually converged. The larger hardware-independent HMI products are as capable as the DCS consoles with functionality and are generally considered easier to use. DCS consoles are capable of interfacing to other controllers like open-system HMIs. The class distinction for new HMI products now appears to be between feature-rich HMI systems capable of running plants and those HMI products that are primarily designed to interface with a single controller.

Controller and PC hardware vendors are deploying Web servers in their components. The Quantum PLC from Schneider Electric and the WebOIT industrial CE computer from Advantech are examples of devices enabled with Web server capability. Any PC connected to the device via IP protocol can access the Web server using only a Web browser for device configuration of the PLC or even control in the process with the operator interface. Hundreds of HMI vendors, machine builders, and building automation companies are making low-cost products that are a single-window operator station into a single device.

The biggest and most sophisticated HMI systems also employ thin-client computing for operator interfaces. Full-function process control operator stations might use terminal services and thin-clients for operator stations while using a Web server providing an HMI for auxiliary operating or information-only stations. The future of HMI will include continued use of thin-client stations to control the cost of hardware and engineering time on configuration of the software.

Finally, vendors are working hard to make system configuration easy and highly reusable because these companies recognize that 60% of the cost of a process control or information system is in the engineering. The large open-system HMI companies are making their configuration systems object oriented. Configuration takes place in a separate environment where all devices within a plant, or even a company, are defined as objects within a common namespace.

Objects built within the configuration system are available to be copied and modified. The device object can then be attached to system objects and then deployed on a physical system component. The configuration system makes the appropriate connections at the physical device level. Adding, subtracting, or changing physical devices is now an easy configuration task by dropping the object into a different device.

EVOLUTION OF HMI ARCHITECTURE

The open HMI began as a single operator interface to a single PLC controller. In such a system, the HMI software contains the tag database and the graphical displays. In a system of multiple HMI machines, each one runs the controller I/O driver. All of the HMI machines initiate communication conversations with the controller. This is a fine arrangement for a small system but cannot support systems with many controllers

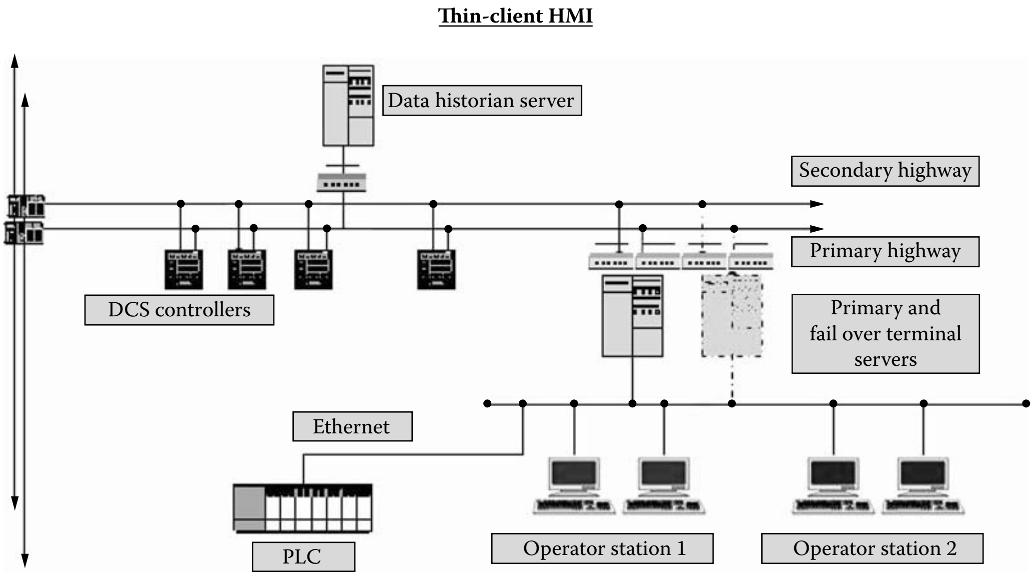


FIG. 4.17g
The architecture of thin-client open HMI on a traditional DCS.

and HMI machines. The answer is a client–server architecture and remote tag access in the HMI.

The client–server architecture for the open HMI was introduced in the mid to late 1990s. The server machine connects to the controller device or system. Multiple HMI machines act as clients accessing the tags in the one or more servers. The client–server architecture improved the ability of the open HMI to interface to multiple controllers and greatly opened the door for the HMI to become the operator's window to information from any source.

The cost of buying and maintaining the HMI hardware and software assets is high for large industrial plants. The client–server architecture described above still requires PC hardware and HMI software at both the client and server locations. The next major development in HMI architectural design is to reduce these costs by making the operator HMI a thin client. This architecture also requires upkeep of client software. The definition of a thin client is a client machine that displays the operator HMI displays without application software on the client.

The advantage of this architecture is the decreased cost of system management, increased system maintainability, and longer component life cycle since the thin client does little work. There are two types of thin client systems: thin clients that display and interface to an instance of an application on terminal server and thin clients that display information through a browser.

A terminal services thin client displays an HMI application session that is running in a Microsoft Terminal Server. The client box views and interfaces to the session through an Ethernet TCP/IP network. This HMI architecture, shown in Figure 4.17g, enables every thin client to have a full-functioned HMI that is very fast using either Microsoft Remote Desktop Protocol (RDP) or Citrix Independent Computing Architecture (ICA) protocol. This is an excellent way to enable wireless handheld computers.

Citrix MetaFrame software further enhances the architecture by allowing server farms and server load balancing. Most DCS operator stations of the 1970s, 1980s, and early 1990s were thin-client terminals displaying an instance of the program on the operator console computer. The operating system was likely UNIX or VME, and the client was called a dumb terminal, but the concept of the thin client is the same.

The second type of thin client is one that uses only an Internet browser to view the operator HMI from a Web server. Tim Donaldson, Marketing Manager for Iconics, an industrial automation software company, describes how their WebHMI functions (see Figure 4.17h):

> The Web publishing wizard publishes the graphics, trends, reports, and alarms to the Web server running WebHMI, Genesis32 software and Microsoft Personal Web Server or Microsoft Internet Information Server (IIS). WebHMI uses communication technology that supports routers, switches, firewalls, virtual private networks (VPN). Any client PC with an Internet browser can point to the desired Web page on the server and the HTML page will load all necessary components in the background providing a real-time OPC-enabled HMI.

The last architecture discussed in this section is the HMI application server. An HMI application server is built similar to the software application servers from companies such as Sun Microsystems, IBM, and Oracle. By the time of this printing, Wonderware, Adroit, and possibly other companies

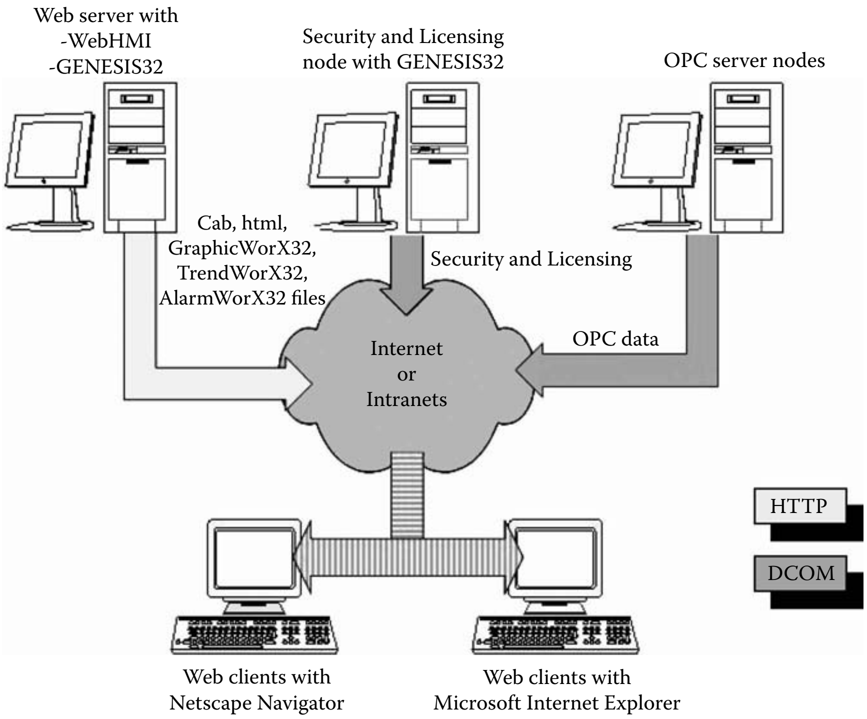


FIG. 4.17h
This diagram shows how the Web clients get information from the Web server and from the OPC server nodes.

have built their systems on this architecture, which once again radically will change the concept of the HMI.

The application server systems of which the author is aware are being built upon a Microsoft .NET framework and take full advantage of object configuration. Once an object is defined with all of its scripting, alarming, trending, and so on, it is then deployed in an application server creating an HMI engine. Application servers can be redundant. The distributed network of application servers perform all of the HMI work for gather control system information, scripting, alarming, trending, and the like. Since the full features of the HMI are being accomplished in the application server, the actual operator visualization is not doing much work. In many cases, the operator visualization may be accomplished in a .NET portal.

This architecture offers many advantages:

- In large systems, maintaining several application servers is much easier than maintaining hundreds of computers running HMI software.
- Implementing change is easier.
- Reducing licenses at every computer may reduce cost.
- HMI redundancy will be standardized.

Perhaps the best advantage of the application server architecture is that using a portal as the operator interface will further help to open the information that is available to present to the operator. The HMI engine will be only one of many application engines running in the plant. Other engines will include manufacturing execution system (MES) calculations such as downtime and overall equipment effectiveness (OEE), tracking, genealogy, statistical process control (SPC), and more. The HMI application server will sit on the plant network along side these other application servers all supplying real-time information to the operator interface.

EVOLUTION OF PLANT NETWORKING

One of the major technologies fueling HMI evolution is the plant information system networks that tie process control systems together. The DCS can be credited with introducing the plant communication network in many process companies. At the outset the DCS was meant to communicate with peer controllers and the HMI over a network highway, allowing the system to excel at SCADA, as well as grow with expanding system needs.

The early DCS network was easy (if expensive) to buy and easy to install because the DCS vendor sold all of the network components and prescribes every phase of the installation. The network was part of the DCS. Whether completely proprietary or an early adoption of IEEE standards, the DCS network provided the backbone for communication over long distances and a way to incrementally add to the existing control systems. The problem of the DCS network was that

FIG. 4.17i
Control room with pneumatic controllers, alarm annunciators, and pushbuttons, circa 1984. Photo provided by Jim Mahoney.

early DCS vendors did not foresee the vast amount of data that customers would pull across those early highways to higher-level systems.

PLCs commonly communicate on proprietary networks and fieldbus networks. PLC vendors and companies such as Woodhead SST build interface cards for the HMI to communicate on these various fieldbus highways. In the 1990s many process industries begin installing TCP/IP network systems for communication of personal computers. The inexpensive network equipment and IT support make them the *de facto* standard for all communication networks at the HMI level and above.

Development of industrial Ethernet networking equipment and widespread acceptance by equipment vendors is quickly making IP communication the standard down to the I/O level. This common and easy networking in process plants is a key factor in implementing large systems of open HMI operator stations.

EVOLUTION OF CONTROL ROOMS

The control rooms that predate the digital HMI have individual process controller devices for valves, vanes, dampers, and other process control equipment that allow operators to manipulate the controller output and set point. Push buttons and lights are the operator interfaces for discrete relay logic. Individual instrument gauges, thermometers, strip chart recorders, and sight glasses monitor process variables in the control room and in the field. Annunciator panels provide alarming and first-out indications.

The old operator panels of control devices, annunciator systems, telephones, and the operator's desk in the control room were part of an ergonomic design for the operator. See Figure 4.17i. Engineers designed the control panels for functionality during routine and emergency situations, as well as startup and shutdowns. Operators easily adapted the controls to their own needs by attaching notes to the instruments and using markers to indicate "sweet spots" of control or danger areas beside the controller. In a glance the operator could survey the control panel and easily recognize and respond to problems.

Unfortunately, some companies never duplicate the ergonomic functionality of the old control room when they install HMIs. An HMI designed by a system integrator sitting on a spare desk or mounted into a convenient panel board is not optimizing the operator of the process. Some managers believe that manufacturing processes run better when the operator is surrounded by human–machine interfaces that present a plethora of information and all controls at the operator's fingertips.

Capstone Technologies of Camas, Washington, designs displays that incorporate hundreds of process trends that surround the operator HMI. The concept is that operators do not need to understand each of the process variables trended, but they will begin to recognize anomalies in the trends and get others involved in diagnosing process issues before they create problems. Concepts and tools such as Capstone and others

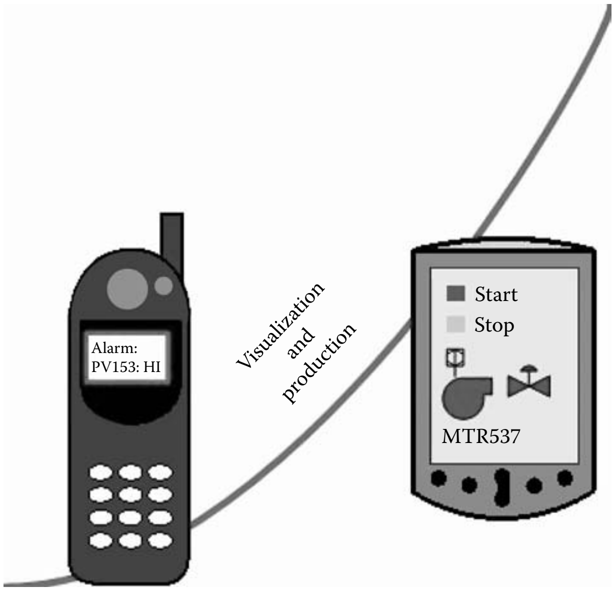


FIG. 4.17j
Mobile wireless communication devices provide interface capability for roving operators and maintenance personnel.

deliver require "maximizing operator butt time" which in turn requires that the HMI and auxiliary tools that the operator needs be designed to keep operators in front of their command stations.

Other managers believe that their industry or process is not conducive to bringing all information to the operator in a control room. Their operators are moving about the operating floor between several stations. Iconics is an industrial automation company that is at the forefront of providing flexibility of operator interfaces from control room to Web-based operator stations to handheld operator interfaces. The modern HMI can enable operators with a combination of handheld wireless devices, inexpensive Web-based HMI computers, and control room systems. See Figure 4.17j.

EVOLUTION OF THE PROCESS OPERATOR

The operator's job is more than to merely operate the process. The operator makes visual inspections of equipment, performs or coordinates chemical testing, and performs shutdown, lockout, and tagging of equipment. The operator implements manufacturing schedule changes and generates production reports and log sheets that describe operating conditions and show how the product was made. The operator reports productivity, dispatches support personnel, and coordinates problem resolution and maintenance activities. The operator makes thousands of individual decisions every day in reaction to material, system, and process equipment problems. The process operator is the conductor of the orchestra of people and systems that produce the product. The control system HMI must help the operator do this job more easily.

Operators are controlling and monitoring an ever-increasing number of process loops because of the capabilities of the modern HMI. Simultaneously with the increased process control responsibilities, many operators are also asked to perform tasks once accomplished by foremen such as coordinating product transitions, making shutdown and maintenance decisions about equipment, and monitoring product quality.

Many control rooms are an eclectic mess of HMI consoles, various information consoles, written logs and books, video monitors, chalkboards, and written notes because of the roles of the operator. The HMI must become a picture frame for the operator with a collage of control system objects, informational objects, and financial objects from many different systems arranged sensibly. The HMI will be the operator's window into the process, the production and maintenance schedules, and likely the customer's order.

The HMI will require increasing organization of the display of information to convey the real-time data, lab test data, historical information, maintenance information, tracking, and integration of video cameras. Pierce Rumph of Orion, CEM of Atlanta, has a hobby of studying the operator tasks and the HMI. Orion CEM, part of Emerson Process Management, is a leader in developing an HMI style guide that helps Pulp and Paper operators become process managers.[5]

Orion's CyberBAR style guide is a navigation bar that separates information above the bar from process interaction below the bar. The bar moves up or down depending on the information contained on the display. The CyberBAR is a navigational tool that uses icons for paging navigation and alarm notification. The CyberBAR also serves as an information bar calling up items such as diagnostics, trends, and SOPs. Figure 4.17k shows a display using the CyberBAR style guide.

The CyberBAR near the top of the display is showing navigation to information pages. The arrow shows where the operator has selected a flow loop on the display and the change zone is populated and ready to accept control inputs. Orion is also a leader in the use of audible messages from the HMI. Voice messages that notify the operator of alarms and console-activated sequences and audible feedback for motor starting and stopping keep the operator engaged with the process.

Unfortunately, navigational aids and intelligent audible signals are not sufficient for the process manager trying to manage a major crisis in a process area of thousands of points. Dynamic alarm management and intelligent alarming methods to enable better response to emergencies are necessary if the operator's process area continues to increase. Process alarms must be written smart to take advantage of alarm hierarchies to suppress minor alarms when a higher-level alarm for the same process group is active. Also, the operator's

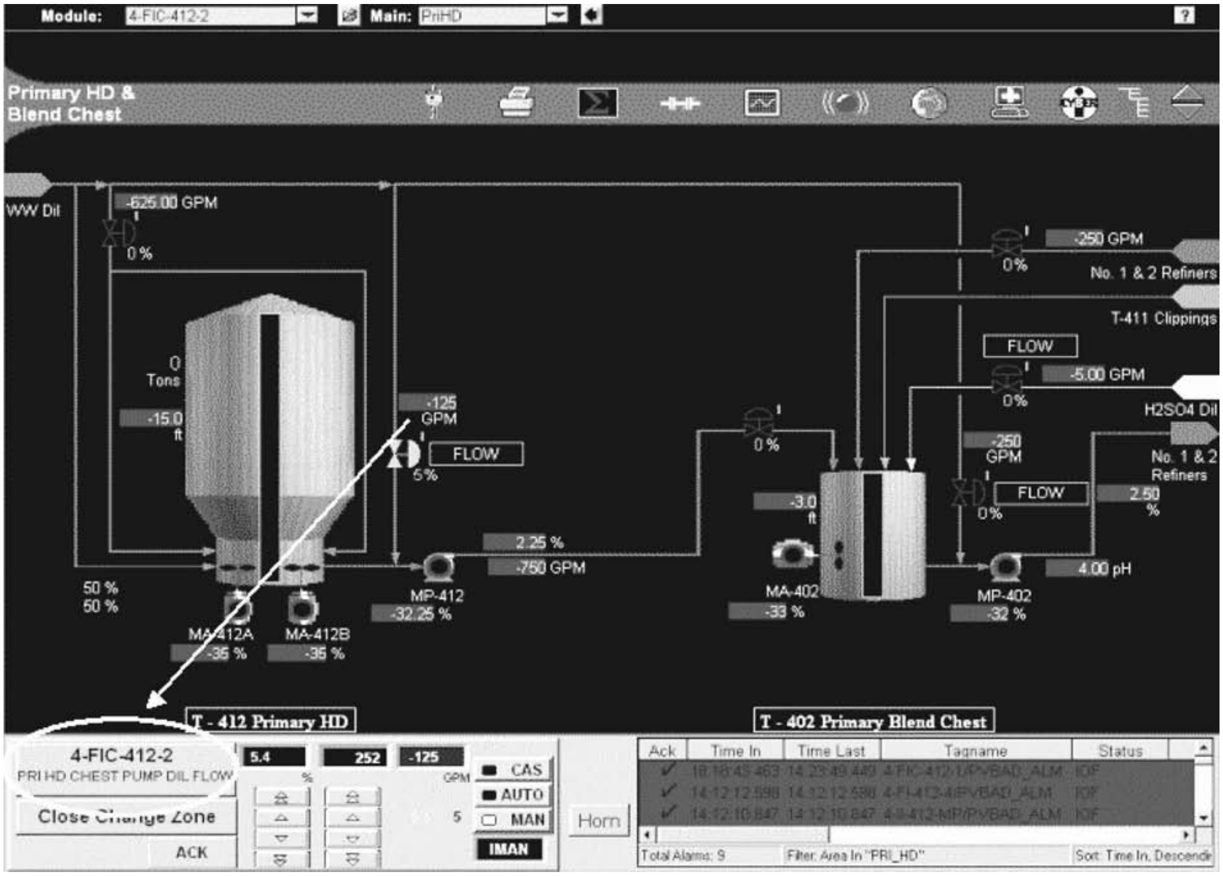


FIG. 4.17k
HMI graphic using Orion's CyberBAR display style guide. Selecting the flow on the graphic populates the change zone for operator entry. The CyberBAR near the top of the display is selected to navigate to information pages.

window must offer alarm disabling, silencing, and dynamic alarm level changes for various modes of operation for large highly sophisticated process industries. TiPS Incorporated of Georgetown, Texas has several products such as LogMate and ACE to help capture and analyze process alarms to reduce confusion during system upsets.

2005 AND BEYOND

Process control, the HMI, and the operator's job function will continue to evolve. The physical choices for providing the operator with a window into the process will continue to increase and will include HMI stations distributed around the plant, consolidated control rooms, personal human–machine interface devices such as handheld computers or cell phones, thin-client operator stations, and off-site human–machine interfaces.

Process industries such as pulp and paper, foundries, utilities, chemical production, and others are very capital intensive. Expensive process control equipment will not be replaced without a great return on investment, so existing controls will continue to be utilized for years, with new systems integrating with old components and plant support systems.

Three initiatives that may have a large impact on the process operator HMI through 2010 will be integration of data from any system into the HMI, widely distributed and mobile operator windows, and the remote-from-site HMI. These technologies can integrate into existing plant systems and utilize existing legacy systems, and they have the capacity to fundamentally change the way process systems are operated in the future.

Distributed and Mobile Control

Is a control room the best place for the operator to control the process? Or should the operator move about the process area?

The control room is a relatively new invention in manufacturing. Prior to the DCS, many processes were run from local panel boards. Coordination of the process control with the panel boards was very difficult, and the central control room increased the plant control quality by allowing for better coordination. The control systems are advanced enough that the control room is not needed to coordinate the control in some cases.

HMI technology is capable of distributing inexpensive operator stations throughout a process area so that an operator can easily walk around the process and still be very close to an HMI. The operator can also use wireless handheld computer technology or even cell phone technology to manage the process.

Remote Operation of the Plant

If your main operator is in a central control room 100 yards from the process today, why couldn't the operator be 100 miles from the process? Why not 1000 miles?

Control systems must be located within the propagation limits of the signal wiring and process lines without extraordinary expenses. However, there are virtually no longer limitations on distance away from the process for the HMI.

Thin-client HMIs and the Internet now enable an operator, or a backup "expert" operator, to be located anywhere in the world. The few technical limitations associated with remote operation of the plant today are dissolving, and future discussions about the location of operators will be philosophical.

The Future

Joey Rodems of InSource Solutions is a 10-year veteran sales consultant of HMI software. In that time the HMI has changed dramatically, but Mr. Rodems expects more and faster change.

> The need for mobile production workers to stay connected, combined with the explosion of new wireless digital personal communications devices, signals the eventual shift of traditional fixed location HMI terminals in many production environments to the production worker's secondary means of collaboration. Growing support for Web browsers, thin clients, smart phones, PDAs and tablet PCs is just the tip of the iceberg. Over the next 5-10 years look for SCADA's reach to extend beyond the HMI toward increasingly numerous and specialized personal communication devices. While these devices will provide a low-cost, natural and effective means of event-based communications, strategically positioned fixed location HMI terminals will provide rich analytical capabilities.

Human–machine interface is the subject of many research projects in private industry, governments, and academia. The work in virtual imaging is fascinating, but will it be relevant in process control? The work by NASA on changing HMI functionality associated with modes of operation[6] could have industrial applications one day. Audio-based interaction between humans and machines might prove to the greatest benefits for the industrial HMI in the near future.

Auditory messages can be sent in parallel with visual signals and can be spoken and received in hands-free operations providing unmatched flexibility by any other I/O method.[7] Engineers can implement auditory HMI in small phases to augment conventional control. The technology of audio HMI is only beginning and like most technology, it will be tested and refined in commercial applications long before it is introduced into industry.

However, the biggest gains in operational success associated with HMI may come not from the technology, but from the management and customization of HMI for the particular industry, process, and individual operator. Distributed HMI devices throughout the plant, personal HMI devices, ergonomically designed central control rooms, and perhaps auditory control can all be used with great success. Companies that implement or ignore the technology without a plan will probably achieve much less than will companies that have a plan for operating the plant and provide the right HMI tools to make it happen.

Increased priority on display design, information presentation, and operator interaction will provide a good return on investment. The process industries have more similarities in their needs than differences. This creates opportunities for toolkits from outside organizations and open HMI companies to increase the efficiency and standards of HMI displays.

Jim Mahoney, a Project Specialist with ABB Automation, has over 20 years' experience working with operator control system interfaces. He has migrated control rooms from pneumatic to digital, from first-generation HMI to the newest PC-based HMI. In a conversation about operator interfaces Jim said,

> The key to a good console is well-laid-out graphics that properly mimic the process, and ease in getting around the multitude of screens. A state-of-the-art console with lousy graphics does not help operators do their job and can actually be a major detriment.

Few experienced people will disagree with Mr. Mahoney, yet few realize that an HMI with mediocre displays helps make mediocre operators and an HMI with great displays helps make great operators.

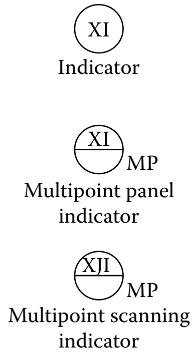
Finally, remember the four fundamental functions of the HMI software—visualization, control, alarming, and trending—discussed at the beginning of this section? These functions were lumped together in software in the 1970s because of the available technology to produce an HMI at that time. A future generation of HMI could see each of these functions performed by different software or application servers and presented through a very thin operator's window that is also looking at business systems, maintenance systems, and the company stock price.

References

1. Freely, J., Merrit, K., Ogden, T., Studebaker, P., and Waterbury, B., "The Rise of Distributed Control," *Control Magazine*, December 1999.
2. Galitz, W. *The Essential Guide to User Interface Design*, New York: John Wiley & Sons, 1997.
3. ISA-TR77.60.04-1996, "Fossil Fuel Power Plant Human–Machine Interface—CRT Displays," approved May 24, 1996.
4. *Business Week Online*, February 22, 1999.
5. Rumph, H.P., "Improved Information Interfacing Turns Operators into Effective Process Managers," *Pulp & Paper,* August 1998.
6. Degani, A., Shafto, M., and Kirlik, A., "Modes in Human—Machine Systems: Review, Classification, and Application," Human Factors Research and Technology Web page, NASA, http://human-factors.arc.nasa.gov, 2001.
7. Stephanidis, C., and Sfyrakis, M., "Current Trends in Man–Machine Interfaces: Potential Impact on People with Special Needs," In Patric, R., and Roe, W. (Eds.), *Telecommunications for All,* Brussels: ECSC-EC-EAEC, 1995.

4.18 Indicators, Analog Displays

F. F. MARTON (1970, 1985) **B. G. LIPTÁK** (1995) **G. I. GYOROK** (2005)



Flow sheet symbols

Types of Movements:	A. Dynamometer B. D'Arsonval C. Polarized iron D. Moving iron E. Electrostatic F. Optical G. Stroboscope H. Acoustic
Costs:	From under $100 to over $500, depending on size.
Partial List of Suppliers:	ABB Inc. (www.abb.com) Action Instruments (www.actionio.com) Ametek Process Instruments (www.ametekpi.com) Dresser Instrument (www.dresserinstruments.com) Fisher Controls International (www.fisher.com) Foxboro-Invensys (www.foxboro.com) Fuji Electric Corp. (www.fujielectric.com) Honeywell Industry Solutions (www.iac.honeywell.com) Jordan Controls Inc. (www.jordancontrols.com) Powers Process Control (www.powerscontrols.com) Rosemount Inc. (www.rosemount.com) Selco Products Co. (www.selcoproducts.com) Thermo Electric (www.thermo-electric-direct.com) Triplett Corp. (www.triplett.com) Westinghouse Process Control (www.westinghouseepc.com) Yokogawa (www.yokogawa.com/us)

This section is devoted to the description of analog indicators. Digital displays are covered in Section 4.15, CRTs in Section 4.5, and human–machine interfaces of other types in Section 4.17. Analog indication can be based on the expansion of liquids in thermometers, on the movement of liquids in manometers (Figure 4.18a), on the motion of flexible elements used for the detection of pressure (Figure 4.18b), on the movement of an electric sensing mechanism, and on a variety of force and motion detection principles.

TERMINOLOGY

Indication implies a representation from which either an individually distinct state or a quantity can be inferred. In instrumentation, indication usually involves a measurement of a quantity to which a numerical value is attached. The magnitude of measurement is displayed on a graduated scale. A movable pointer indicates the magnitude on the scale.

Indicators can also display the existence of a condition without necessarily attaching a quantitative value to it. The same basic sensor elements can be used for such on/off indications or for recorders.

The term *gauge* is used for an instrument that only indicates. The suffix *-scope* refers to an instrument used for viewing only. An oscilloscope is an indicator; an oscillograph is a recorder. The oscillogram is the recording itself.

All of the aforementioned indicators are analog instruments. They vary one physical phenomenon to indicate another

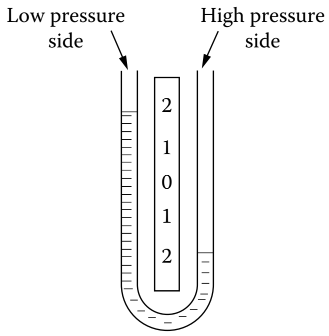


FIG. 4.18a
Differential pressure indicator utilizing a U-tube manometer.

one by analogy. One of the most widely used analog indicators is the liquid-in-glass thermometer, which uses the property of metallic media (such as mercury, mercury- thallium, or gallium) or of various organic liquids to expand under the influence of heat to indicate the temperature of another medium.

Other analog measuring methods include the following:

- One type of viscosity indicator uses the time required for a ball to sink from the surface of a liquid medium in a cylinder to the bottom as an analog measure of the viscosity of the liquid.
- The vapor-pressure thermometer indicates temperature by detecting the pressure of a filling liquid that is in equilibrium with its own vapor.
- The frequency of an alternating current can be indicated by the vibration of a reed in an instrument that contains several reeds, each tuned to a different frequency. The particular reed whose resonant frequency is closest to the current frequency swings with the greatest amplitude and indicates the frequency on an adjacent scale.

ELECTRICAL MOVEMENTS

Probably the most fundamental electrical movement is that of a dynamometer, consisting of two coils: one moving, the other stationary. When current flows through the coils, the resulting coil movement drives a pointer. This indicator can detect DC or AC (up to 200 Hz) current, voltage, or power, and the reading error can be as low as 0.1%.

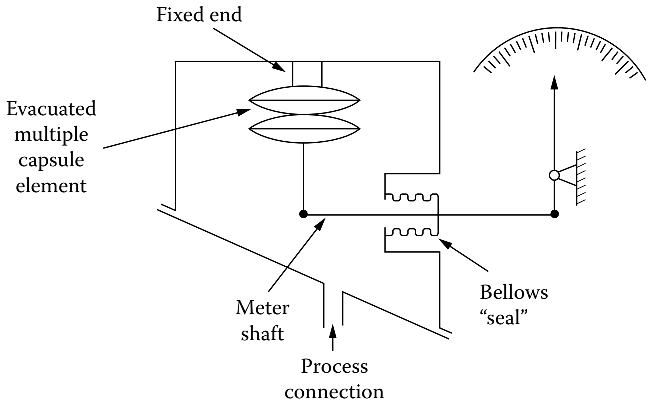


FIG. 4.18b
Motion balance absolute pressure indicator.

The D'Arsonval movement is the most sensitive of all movements. It consists of a fixed magnet and a moving coil, which moves in response to the field created by the magnet. This movement is inexpensive; it can detect DC or, if a rectifier is provided, AC (up to 10 kHz); and it requires only microwatts for operation. Its response is linear, and its error can be as low as 0.1%.

In the polarized iron movement the roles are reversed: the coil is fixed and the magnet is moving due to the field in the coil. Historically, this was the first inexpensive DC ammeter used, for example, in automobiles. The moving iron vane movement also consists of a fixed coil. Here the field generated by the fixed coil moves an iron vane. This is an inexpensive DC or AC (up to 125 Hz) indicator that requires a square-law scale, which is compressed at the low end.

For high AC or DC (over 10 volts) voltages and for applications where inaccuracies of 0.5 to 1% are acceptable, the electrostatic movement can be used. The response of this readout follows the square law, and the operation of the movement is based on the repulsion or attraction of charged electrodes. For a detailed discussion of all electrical sensors refer to Section 7.3 in Volume 1 of this handbook.

INDICATION OF MEASUREMENTS

Indications on a graduated scale require the relative motion of two elements. One of the two is a fixed reference. In most indicators, the scale is stationary and the indicating pointer moves. The indicating element may be a liquid column (Figure 4.18a), a float, a pointer, or a beam of light. The scale may be laid out on a straight line (Figure 4.18c) or arc (Figure 4.18d) or on a circular arc (Figure 4.18e).

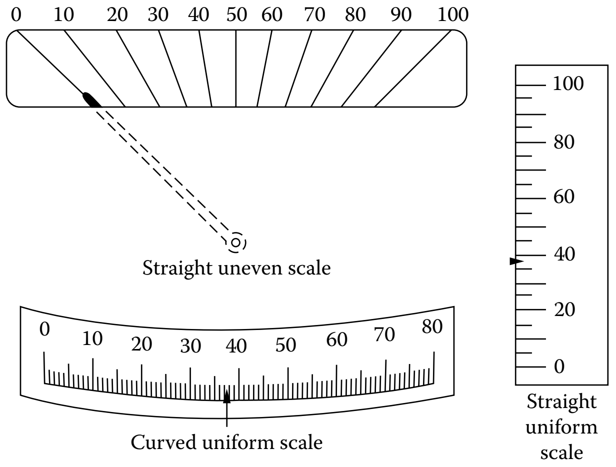
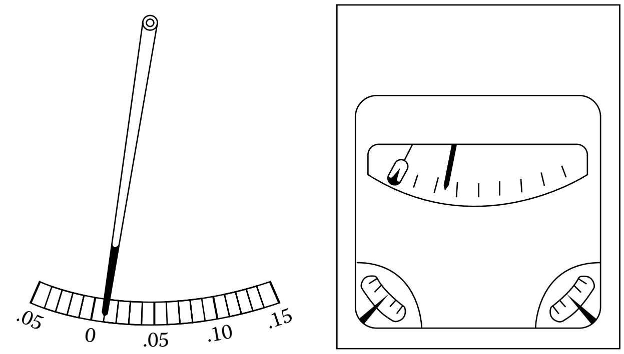


FIG. 4.18c
Straight and curved indicator scales.



Segmental uniform scale

Controller with segmental scale

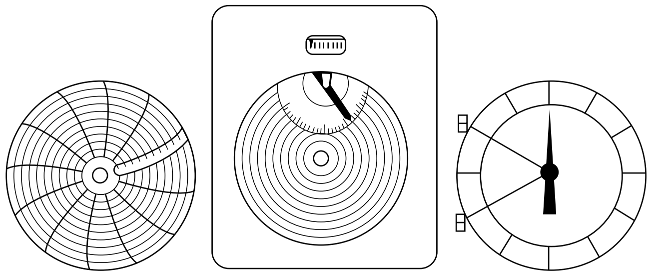
FIG. 4.18d
Segmental scale indicators.

The larger the radius, the better the readability. A scale of more than 180 degrees laid out on a circular disk is usually referred to as a dial. A scale is called uniform if the graduations on it are equidistant. The increments may also be different, reflecting other than a linear relationship. Flow meters utilizing orifice-type sensors often have square root scales; vapor-pressure thermometers have an uneven scale that becomes progressively wider. They are calibrated according to the individual filling medium.

Fixed-Scale Indicators

A liquid column moving in a transparent, usually cylindrical tube with a stationary scale on the tube forms the simplest fixed-scale indicator. Liquid-in-glass thermometers and manometers are examples of this type of indicator. Manometer and thermometer scales are often adjustable to allow recalibration, and the tubes are usually vertical.

One manometer type uses two tubes connected in the shape of a “u” (Figure 4.18a). If one leg of the “u” is bent on an angle, the device is called an inclined-tube manometer. For very low pressure difference measurements it offers the advantage of a longer, more accurately readable scale, stretched out by the reciprocal of the sine of the angle that the inclined leg forms with a horizontal line (Figure 4.18f). In some designs the inclined leg is provided with a logarithmic scale to stretch readings in the low range farther out.



Segmental scale on circular recorder

Circular recorder with concentric indication

Circular scale on circular recorder

FIG. 4.18e
Scales used on circular recorders.

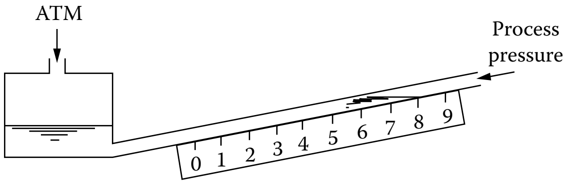


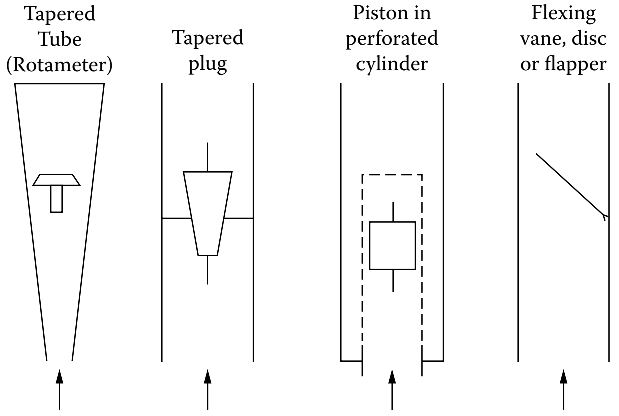
FIG. 4.18f
Inclined tube-type manometer indicator.

Readings in liquid columns are taken at the crest of the meniscus formed as a result of capillary attraction. Mercury forms a convex meniscus; most other liquids form a concave meniscus.

A plummet (or plumb bob) floating in a calibrated tapered tube is used in the variable-area-type flowmeter to indicate by its position the rate of flow through the tube (Figure 4.18g). The position of a plummet in a liquid inside a graduated transparent tube can be an indication of fluid density.

The pointer is the most commonly used moving reference for indication. The color of the pointer usually contrasts with the background for better readability. Pointers are frequently covered with luminescent materials to make readings in darkness possible. Transparent plastics are used if the pointer would otherwise obscure other indicators under it, e.g., when a totalizing counter is on the same dial.

Incandescent, fluorescent, or neon-edge lighting is also used to facilitate observation. To minimize the parallax or apparent displacement effects caused by the lens, knife-edge pointers are used. Also, mirrors are used in the same plane as the dial, and the readings are taken at an angle of observation where the pointer and mirror image coincide. Sometimes the dial is raised to pointer level. In another meter design, either a part of the rotatable scale or the pointer is optically projected onto a coated window, whereby parallax is completely eliminated.



Tapered Tube (Rotameter)

Tapered plug

Piston in perforated cylinder

Flexing vane, disc or flapper

FIG. 4.18g
Design variations of the variable-area-type glass tube flow indicator.

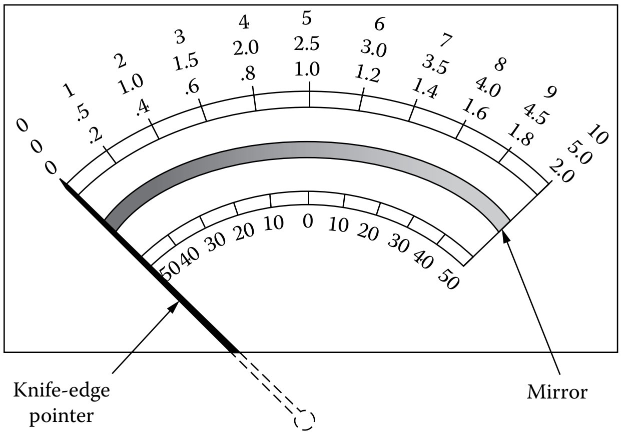


FIG. 4.18h
Multiple-scale indicator.

Indicating instruments usually read from left to right or clockwise. If negative values have to be read, as in vacuum indications, graduations progress from right to left. Compound gauges indicating pressure and vacuum have the zero point near the middle of the scale. When readings near zero are of lesser importance, suppressed-zero ranges are used with narrower graduations at the beginning of the scale followed by wider increments. There are scales with extended sections and scales with condensed sections. A multiple scale is illustrated in Figure 4.18h.

Thermometers are often provided with dual scales that indicate in both Fahrenheit and Celsius degrees. Pressure gauges usually have a circular scale of 270 angular degrees (Figure 4.18i). Precision gauges are made with 350-degree scales. Extending the pointer rotation to two turns covers a total of 660 degrees, resulting in a scale of 80 in. (2 m) in a gauge 16 in. (400 mm) in diameter. A light beam can be used as the movable reference, throwing a light spot on a scale. Readings in this case are taken at the center of the light spot.

Multiple-scale meters show different ranges of the same variable on concentric scales. A selector switch changes the range in such electrical indicators and shows the range of the scale on which the reading is to be taken.

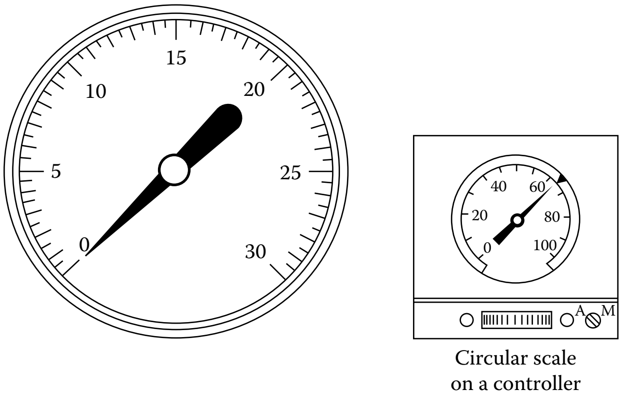


FIG. 4.18i
Circular gauge dials.

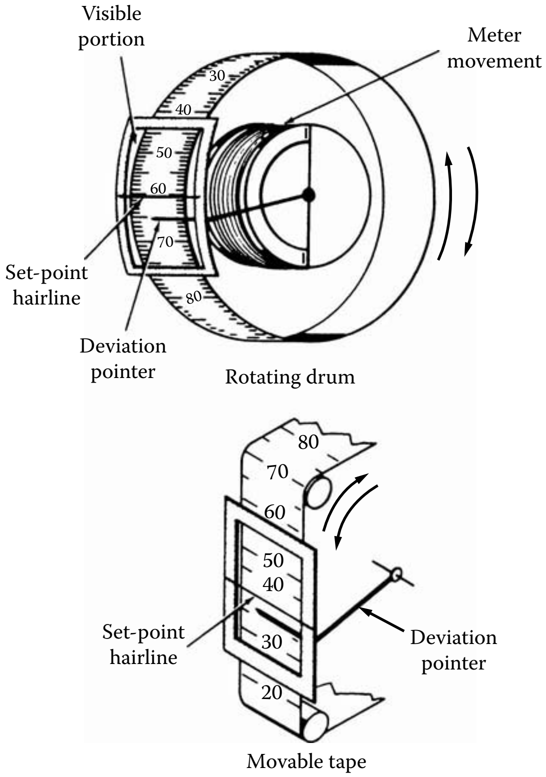


FIG. 4.18j
Movable scale indicator design.

Movable-Scale Indicators

The alternative to the moving pointer is the moving scale and fixed reference or index (Figure 4.18j). A dipstick with a scale engraved on it used for immersion measurements in a vessel is the simplest indicator in this category.

A large circular disk with graduations on its entire periphery, read at an index in 12 o'clock position, affords great precision in reading fine subdivisions (Figure 4.18k). The hydrometer is also in this category. The floating vertical scale projecting through the surface of a liquid shows the specific

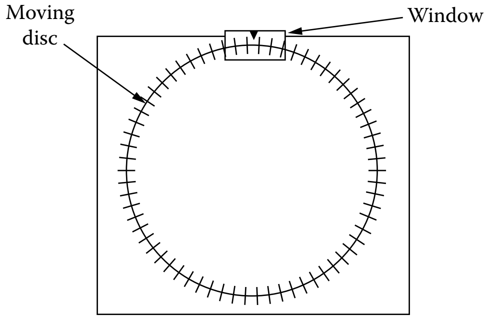


FIG. 4.18k
Rotating disk-type movable scale indicator.

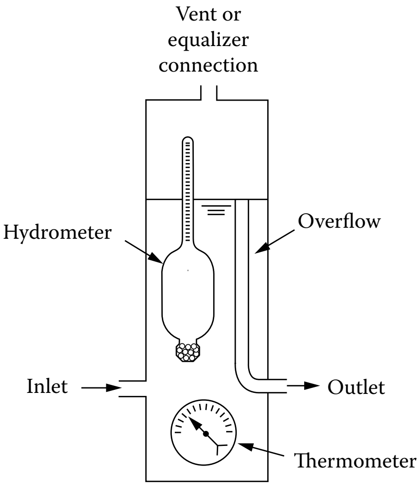


FIG. 4.18l
In-line hydrometer in rotameter housing.

gravity or density of the liquid as indicated by the surface of the liquid on the scale (Figure 4.18l).

One of the more special sorts of moving scale devices is the stroboscope-disk indicator. In these devices the exact value of the frequency to be measured is displayed in standing lanes resulting in a virtual image given by the coincidence of the velocity of the black and white lanes on the disc with the frequency of the impulse-type light lamp. If the frequency does not match the reference frequency, the displayed virtual image rotates to the right or left.

In one of the designs, the motor-driven disc is controlled by an AC reference frequency, which can be the power from the grid if this simple reference is acceptable. For higher precision the frequency of a quartz oscillator is used (Figure 4.18m).

The impulse-type light lamp (Glimm lamp, flash light, LED) is controlled by an amplifier from the power of the frequency to be measured. Slowly, the variable signals to be measured (or indicated) could function conversely.

PARAMETRIC INDICATION

A parametric indication tells only that a state exists, e.g., whether a fluid is flowing or static. For example, the precise level of water in a boiler may be less important than the knowledge of whether it is an adequate quantity.

Examples of indicators of states or conditions include the sight flow meter or the vane- or paddlewheel-type flow sensor, which allow the observer to see whether or not a fluid is moving. Similarly, a bubbler in which an air stream is passing through a narrow glass tube does not measure the air flow but only indicates that it is occurring.

In an electrical system, an indicator light is often used to show that an electric potential has reached a certain point. Similarly, a Geiger counter indicates radiation by clicking or blinking. Various other nonmetering analog indicators use color, which is easily distinguished by the human eye. Color changes are also used in instrumentation to indicate changes of state. Chemical conditions and reactions are often indicated by color. This is exemplified by the pH meter and by litmus paper.

The change that take place in the physical state of a compound can monitor approximate temperature. Solid waxes with various admixtures show the amount of heat absorbed by drooping in an oven. Color changes brought about by heat in certain materials are likewise used for approximate temperature indications.

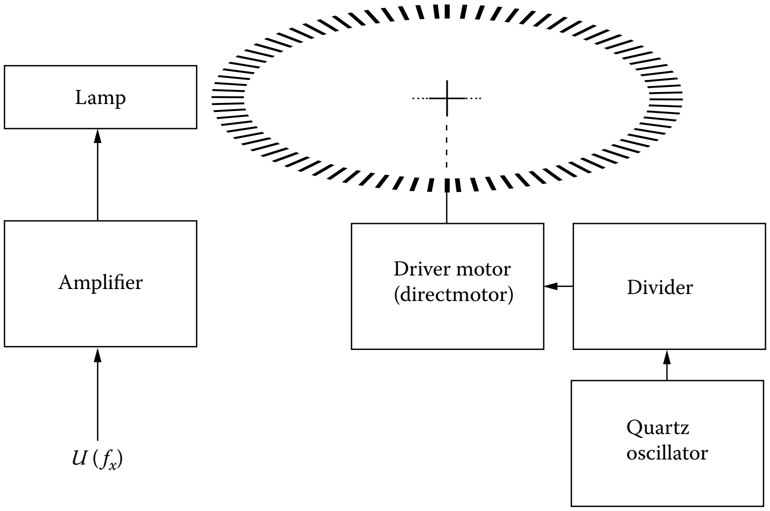


FIG. 4.18m
Frequency indicator with stroboscope wheel.

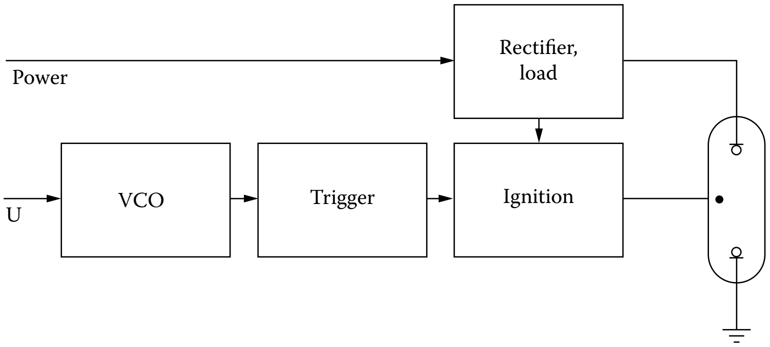


FIG. 4.18n
The variable frequency stroboscope-lamp.

DIGITAL INDICATORS

Digital indicators (Section 4.15) present the readout in numerical form. They are preferred, because fewer mistakes are made in reading a number than in deciphering an indication on a scale. Numerical readouts are also called for when a quantity is to be counted. Thus, counters usually present the result in figures even though an analog indication of the instantaneous value of a variable offers other advantages.

A speedometer is an analog indicator of an instantaneous rate of speed, whereas the mileage traveled is digitally shown on a counter.

In the case of flow meters, the flow rate is usually provided by analog indication (i.e., in units of the quantity flowing at a given instant per unit of time), but the total quantity is usually indicated numerically on a separate counter. To accomplish this, a somewhat complicated conversion by an integrator is required. A positive displacement meter indicates total flow and is generally provided with a digital readout.

Digital clocks and thermometers are easier to read than dial clocks and scale thermometers. The necessary analog-to-digital conversion is even simpler in electrical measurements. Digital meters are made for the readout of almost any variable. Voltmeters (including panel meters with gaseous numerical display devices); thermocouple thermometers; indicators of pressure, load, strain or torque; pH meters; oscilloscopes; and stroboscopes all have become available with digital displays. Digital displays are discussed in detail in Section 4.15.

Alarm switches can also be considered as a special group of the digital indicators. Their task is to detect faulty or dangerous conditions and draw attention to them. The simplest alarm device is a light bulb that signals the state or condition of a process. In other cases the frequency of a flashing light is used as an alarm indication, where higher frequencies can signal higher levels of danger.

The frequency of a stroboscope indicates voltage, as illustrated in Figure 4.18n. The input to the stroboscope enters a voltage-controlled oscillator (VCO), the frequency of which can be varied. The output of the VCO controls the flash tube through an ignition circuit.

ACOUSTIC INDICATORS

Acoustics is often used to indicate or signal to the operator the existence of some dangerous or undesirable condition. Acoustic information is usually provided in addition to visual and is in the form of ringing, beeping, or buzzing. Such combinations are used in annunciators (Section 4.1) and in other systems that manage abnormal conditions (Section 4.11).

These sounds can be produced electronically and can be made to be relaxing, harmonious, or polyphonic. Both the volume and the sound frequency can be used to provide

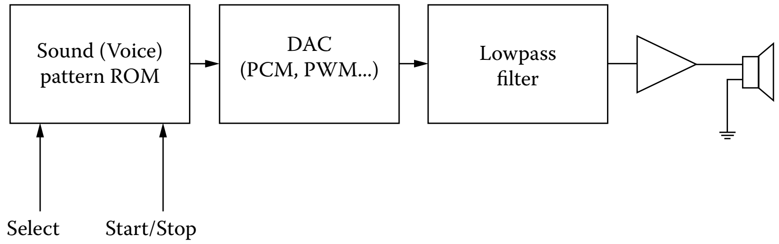


FIG 4.18o
The generation of machine sound.

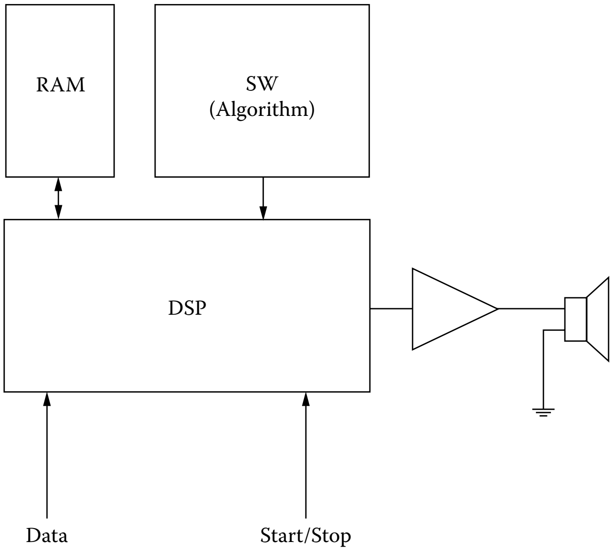


FIG. 4.18p
The synthetic generation of artificial sound.

information about the degree of danger. The choice of these sounds must be made very carefully, because if they are too monotonous or high pitched, they can irritate or stress the operators.

In addition to using sound to signal certain conditions, the human voice can also be used to indicate certain developments, state measured values, or describe the desired corrective actions. Such speech generators record a spoken text, then digitalize the recorded sound-pattern, breaking it down to units and writing it into a semiconductor memory (EEPROM, FLASH, NVRAM, ROM).

When generating the sound, first the digital units of the sample speech are read from the memory, then, by the help of an analog-to-digital converter (ADC) it is converted into a continuous signal. A low-pass filter can be used to reduce the quantity of quantum noises. Through this method one can obtain a device that can "utter" a few words or even a sentence. Such devices are available and are inexpensive in integrated circuit designs. (Figure 4.18o).

When the text is more complicated, speech synthesizers can be used, which can handle some basic grammatical rules, accents, and intonation of the language (Figure 4.18p). Such speech can be produced by a variety of algorithms in signal processors (DSP) or custom-made IC circuits.

The speech generators can be remotely controlled or supervised. In this case the speech generator can be connected to a regular or a cell phone. The input of the control unit is a driven order instruction that, with the meaning "what" and "whom," is transmitted to a particular phone with the desired message (Figure 4.18q). This way a particular operator can receive the message through the phone, in the form of a human voice (Section 1.7). There is also the possibility of interactive message reinforcement or message forwarding orders to other operators.

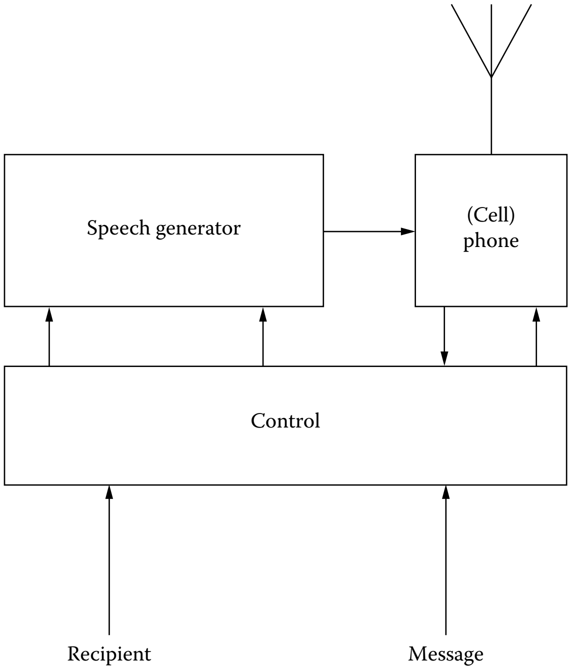


FIG. 4.18q
The generation of telephone messages.

Bibliography

"Analog Panel Meters," *Measurements and Control Systems*, February 1994.

Bailey, J., "Non-Video Display Rounds Out Vital Man–Machine Interfaces," *Control Engineering*, May 1980.

Bylander, F. G.. *Electronic Displays,* New York: McGraw-Hill Book Co., 1979.

Chigrinov, V.G., *Liquid Crystal Devices: Physics and Applications,* Boston: Artech House, 1999.

Degani, A., Shafto, M., and Kirlik, A., "Modes in Human–Machine Systems: Review, Classification, and Application," Human Factors Research and Technology Web page, NASA, http://human-factors.arc.nasa.gov, 2001.

Guerlain, S., and Bullemer, P., "User-Initiated Notification: A Concept for Aiding the Monitoring Activities of Process Control Operators," Honeywell Technology Center, 1996. http://www.winbond-usa.com/products/tts/productbriefs/wts701_prod_bul.pdf

Keller, P. A., *Electronic Display Measurement: Concepts, Techniques, and Instrumentation,* New York: John Wiley & Sons, 1997.

Lueder, E., *Liquid Crystal Displays: Addressing Schemes and Electro-Optical Effects,* New York: John Wiley & Sons, 2001.

MacDonald, L. W., and Lowe, A. C., *Display Systems: Design and Applications,* New York: John Wiley & Sons, 1997.

Ono, Y. A., *Electroluminescent Displays,* Singapore: World Scientific, 1995.

Reinhart, J., "Are Your Central Displays User Friendly?" ISA/93 Technical Conference, Chicago, 1993.

Sherr, S., *Electronic Displays,* New York: John Wiley & Sons, 1993.

Shirley, R. S., et al., "What's Needed at the Man/Machine Interface?" *InTech,* March 1981.

Stupp, E. H., and Brennesholtz, M. S., *Projection Displays,* New York: John Wiley & Sons, 1999.

Thumm, M., et al., "Displays and Vacuum Electronics.," April 29–30, Garmisch-Partenkirchen, VDE-Verlag, Berlin, 1998.

4.19 Lights

D. W. LEPORE R. A. WILLIAMSON (1985) **B. G. LIPTÁK** (1995, 2005)

Types:	A. Incandescent B. Neon C. Solid-state (LED); Software is also available for virtual instruments, including virtual LEDs. *Note:* In the summary below the letters A to C refer to the above-listed indicator light types.
Operating Power Ranges:	A. 1 to 120 V AC, V DC; 8 mA to 10 amps B. 105 to 250 V AC, 90 to 135 V DC; 0.3 to 12 mA C. 1 to 5 V DC; 10 to 100 mA
Color of Unfiltered Light:	A. White B. Orange C. Red
Relative Brightness:	A. 1.0, 1.0 B. 1.0, 0.5 C. 2.0, 2.0
Average Useful Life:	A. up to 50,000 hours B. up to 25,000 hours C. up to 100,000 hours
Application Limitations:	A. Shock and vibration can cause early failures and generate considerable heat. B. Require high voltages and current-limiting resistors; have relatively low light output. C. Expensive; brightness is high but total light output is low.
Cost:	A heavy-duty, push-to-test, oil-tight pilot light for panel mounting costs about $75 to $100.
Partial List of Suppliers:	Allen-Bradley Co. (www.ab.com) AMP (www.amp.com) Automatic Switch Co. (www.ascovalve.com) Danaher Controls–Eagle Signal (www.danahercontrols.biz/eagsignal.htm) Eaton Corp. (www.eaton.com) GE Fanuc Automation (www.gefanuc.com) Hewlett-Packard Co. (www.hp.com) MicroSwitch, a Honeywell division (www.honeywell.com/sensing) Ronan Engineering Co. (www.ronan.com) Square-D Co. (www.squared.com) R. Stahl Inc. (www.rstahl.com) Westinghouse Process Control (www.westinghousepc.com)

INTRODUCTION

Alarm lights are also discussed in Section 4.1, as status indicators, and in Section 4.18 as components in HMI (human–machine interface) systems.

Lighted indicator lights can convey several types of information to the operator. These include binary information, in which an on/off or open/closed condition can be displayed by a lighted (on) or unlighted (off) indicator, and status information, in which normal or alarm (abnormal) conditions are

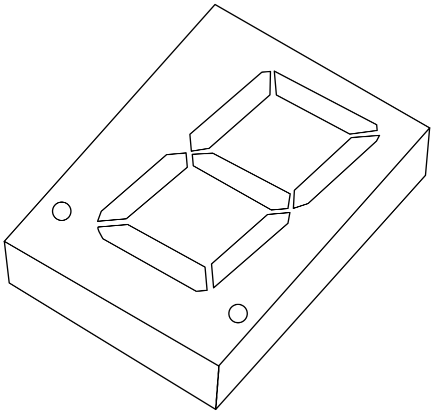


FIG. 4.19a
A seven-segment digital display, designed for good readability in brightly lit control centers. (Courtesy of Hewlett-Packard.)

expressed by legends, colors, and flashing or nonflashing indicators.

Lights can be incandescent, neon, solid state, or software-based virtual displays. The amount of information that can be displayed by a single light is directly proportional to the equipment being monitored and its size or complexity. The use of redundant indicators is not recommended because their use reduces the attention value of all light indicators, and confusion can result if large numbers of them are used.

Light-emitting diodes (LED) can not only be used as solid-state lamps, but also as bar segment (Figure 4.19a) or dot array displays in red, green, or yellow colors, in 0.1- to 1.0-in. (2.50- to 25-mm) sizes. Such indicators are not discussed here because they have already been covered under “Digital Readouts and Graphic Displays” earlier in this chapter (Section 4.15).

LIGHT SOURCE CHARACTERISTICS

The most common types of lighted indicators are incandescent, neon, and solid-state lamps. Figure 4.19b shows their spectral response curves. The relative response ordinate at 100 corresponds to the peak sensitivity of the human eye and to the peak wavelengths emitted by the lamps.

Standard vision, for example, is most sensitive at approximately 560 millimicrons, which is in the yellow and yellow–green band. The gallium arsenide phosphide (GaAsP) solid-state lamp peaks at a wavelength of 650 millimicrons, which is in the red and red–orange band. The curves in Figure 4.19b reflect the relative efficiencies of the light sources. The more efficient lights are the ones that have an output that nearly matches or falls within the standard vision curve. This output characteristic is approximated by both neon and the solid-state lamps. They are efficient because they convert most of their input power into light and emit little heat.

While LED cluster-type lights are more expensive and their total light output is less than that of incandescent lights, their life expectancy is much longer (up to 100,000 hours) and their energy consumption is less because their operating

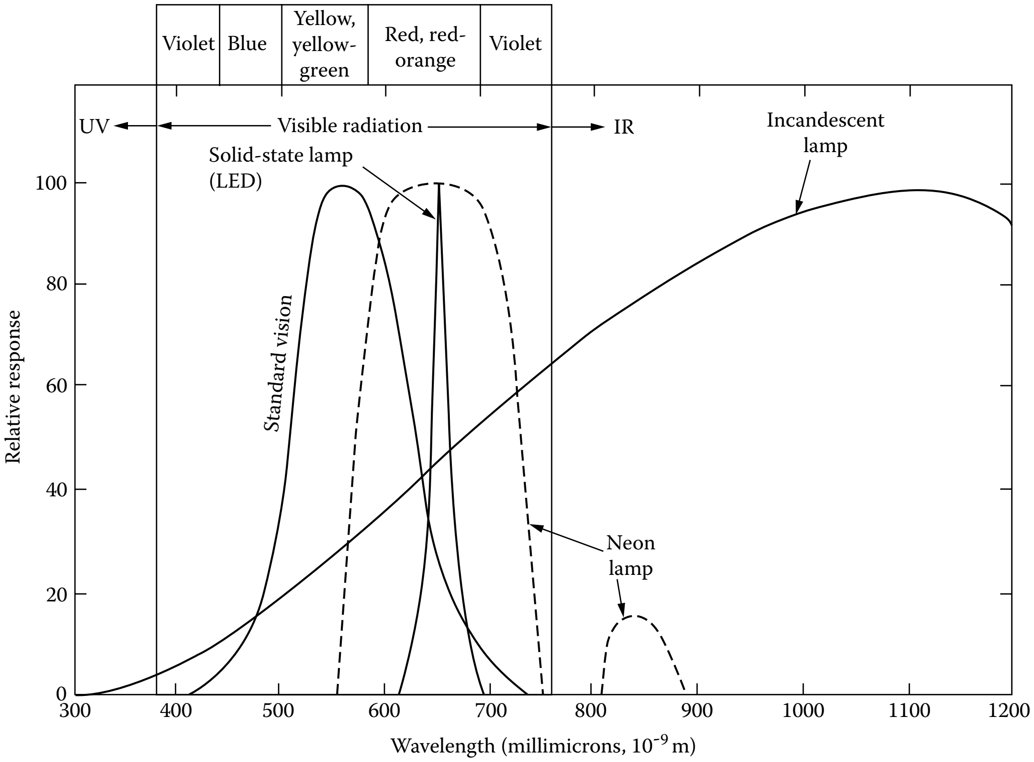


FIG. 4.19b
Spectral response curves of the human eye to various common lamp designs.

temperature is low. LEDs are also insensitive to vibration or shock and can be connected in an intrinsically safe manner. In addition, in the case of LEDs, defective bulbs can be identified (reduced luminescence) while the light is off.

Light Selection

Important human factors in the selection and use of lighted indicators are visibility and arrangement. To transmit information, the indicator must be clearly visible to the operator. Variables that affect visibility are location, brightness, contrast, color, size, and whether the indicator is flashing or nonflashing.

Critical indicators should be located within 30 degrees of the line of sight and should be at least twice as bright as the surface of the mounting panel. The use of dark panels is recommended because they furnish strong contrast to the indicator and reflect little light to the operator. When the ambient light levels are high, alarm legends should have dark characters imprinted on a light background. Inconsequential and routine messages should use the reverse combination.

Colors and Flashing

Colors can be powerful tools when properly used for lighted indicators. To avoid confusion, however, only a few colors should be used to code the different operating conditions.

General information should be lighted in white, while normal conditions should be green. For abnormal conditions or in cases where caution is required, amber (yellow with a reddish tint) is a good choice because it affords maximum visibility. Red color should be used only for critical alarms that require immediate operator response. The use of blue or green lenses should be kept to a minimum.

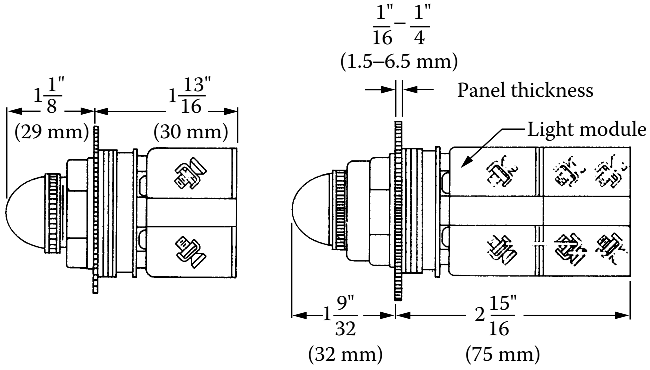
All lamps emit most strongly in the red and red–orange band. Consequently, much light is lost if it is filtered so as to appear blue or green. For important indicators, one should use the largest size that is compatible with the panel space.

Flashing greatly improves visibility, but its use should be limited to critical alarms. The rate of flashing should be 3 to 10 flashes per second with the "on" time approximately equaling the "off" time. Light indicators should be arranged according to a functional format. Indicators associated with a manual control device (pushbutton or switch) should be placed right above the control device.

It is best to locate related indicators on separate subpanels. Displays requiring sequential operator actions should be arranged in the normal reading pattern—from left to right or from top to bottom. Critical indicators should have dual lamp assemblies for additional reliability. A lamp check switch should be supplied to test for and to locate burned-out lamps for replacement.

Lenses and Operating Environments

The selected lenses should be diffusive and should eliminate glare or hot spots. The lens should also provide a wide angle of view (120 degrees minimum), and if side visibility is required, it should protrude over the mounting surface. The lens must be large enough to accommodate the required legends.



Standard pilot light

Push-to-test pilot light

FIG. 4.19c
Pilot lights for panel mounting. (Courtesy of Square-D Co.)

Legends are commonly produced by hot stamping, engraving, or photographic reproduction of transparencies. Ordinarily, hot stamping is the most economical, whereas photo-transparencies furnish the sharpest characters and are the most versatile.

Environmental parameters also affect the operation of indicator lights. Special designs are available for shock, vibration, or high-temperature applications. Rapid dissipation of heat is important if the indicator generates heat. Dripproof or watertight designs should be selected if the indicators are to operate in high humidity or corrosive atmospheres or if the lights will be placed on panels that are periodically washed down.

Figure 4.19c illustrates some heavy-duty oil-tight pilot lights designed for panel mounting applications. They are available in both standard and push-to-test versions.

Light Components

The main components of most indicator light assemblies (Figure 4.19d) include the lamp-holder, the lamp, and the lens. Panel light types are usually secured to a panel by a nut and a lock-washer. Cartridge models can be held in place by a speed-nut type friction clip. Snap-in lights are usually retained by expandable latching fingers. Power can be supplied through wire leads and solder, screw, or quick-connect terminals.

The heart of all lighted indicators is the lamp or light source itself. The three types (Figure 4.19e) of lamps in common use are incandescent, neon, and solid-state designs. The major parts of a lamp are the bulb (containing the light emitter) and the base. Lamps are also classified according to bulb shape and size and by the type of base.

Bulb shape and size are designated by a letter that describes the shape and by a number that gives the nominal diameter in eighths of an inch. For example, a T-1 lamp has a tubular-shaped bulb that is one eighth of an inch in diameter. Common bases are bayonet, screw, flanged, grooved, and bi-pin. Some

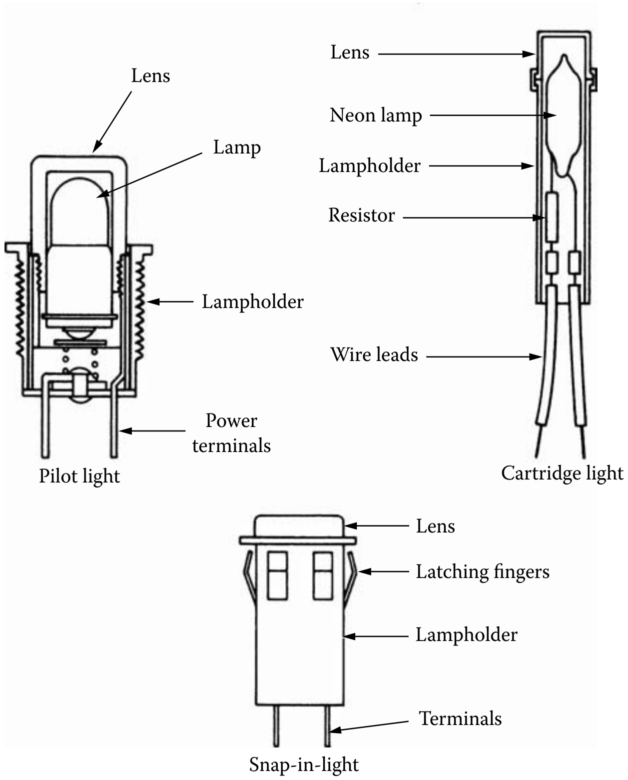


FIG. 4.19d
Indicator light assemblies.

lamps have no base at all and are only supplied with wire terminals.

LAMP TYPES

Incandescent Lamps

The incandescent lamp shown in Figure 4.19e consists of a coiled tungsten filament mounted on two support wires in an evacuated glass envelope. When current is passed through the filament, its resistance causes it to glow and to emit both light and heat. In Figure 4.19b it can be seen that only about one third of the radiation emitted from this lamp is in the visible band (white light), while the rest is in the infrared band (heat). This means that approximately two thirds of the input power is emitted as heat.

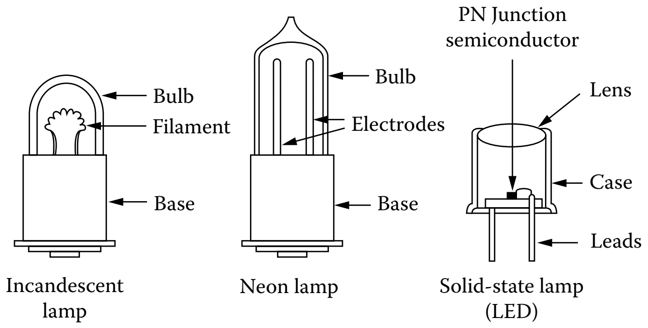


FIG. 4.19e
Common lamps.

If large quantities of incandescent lamps are to be operated continuously and mounted closely together, special allowance for adequate heat dissipation is necessary.

If the lamp is likely to experience shock and vibration, a low-voltage, high-current design should be used because it has stronger filaments. Lamps of 6 volts or less usually have short, thick filaments, whereas lamps of more than 6 volts generally have longer and thinner filaments. In all cases, however, the lamp should be tested under simulated operating conditions.

Incandescent lamps can operate from 1 to 120 volts AC or DC. Current drain will be from 10 milliamperes to 10 amperes. Figure 4.19f shows the relationship of the applied voltage, lamp life, current, and light output. Variations in applied voltage have a drastic effect on lamp life. It is common practice to improve life and sacrifice some light by operating the lamp at slightly below rated voltage. For example, a 6-volt lamp operated at 5 volts will have roughly eight times the normal life and will still provide 60% of normal light.

Over-voltage results in nearly the opposite effect: A 6-volt lamp operated at 7 volts will have only about one sixth of its normal life and will provide one and one half times the normal light. Therefore, controlling applied voltage is very important.

Space permitting, a large lamp is preferred to a small one because its cost will be lower and its life and reliability higher. The larger lamp will also emit less heat than the smaller one of equal light output because the former has more surface area and its filament operates at a lower temperature.

Neon Lamps

The neon lamp shown in Figure 4.19e consists of two closely spaced electrodes mounted in a glass envelope filled with neon gas. When sufficient voltage is applied across the electrodes, the gas ionizes, conducts a current, and emits light and heat. All neon lamps require a current-limiting resistor in series with the lamp to guarantee the designed life and light characteristics, an example of which is the cartridge light in Figure 4.19d.

The orange light emitted by this lamp is easily seen because a large portion of it falls within the standard vision curve (see Figure 4.19b). Since most of its emitted radiation is in the visible band, little heat is emitted. An important consideration, however, is that although the neon lamp is an efficient light source, its total light output is low. A clear or lightly diffusing lens should be used with it so that only a small portion of the light is absorbed.

Neon lamps are very satisfactory for use under conditions of severe shock and vibration. The rugged mechanical construction avoids the use of the fragile filament of the incandescent lamp. Neon lamps will operate only on high voltages

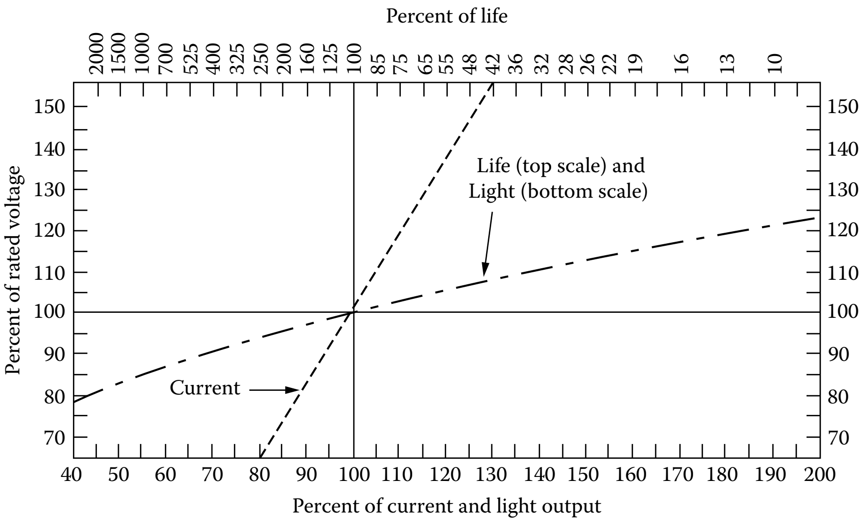


FIG. 4.19f
Incandescent lamp characteristics.

and commonly run directly from standard 120 or 240 volt AC line voltages. Because of the high voltage, they require little current and power.

Solid-State Lamps (LEDs)

The solid-state lamp shown in Figure 4.19e is commonly called a light-emitting diode (LED), and it is a valuable byproduct of semiconductor technology. It is basically a P–N junction diode mounted in a hermetically sealed case with a lens opening at one end.

Light is produced at the junction of P and N materials by two steps. First, a low-voltage DC source increases the energy level of electrons on one side of the junction. In order to maintain equilibrium, the electrons must return to their original state. Therefore, they cross the junction and give off their excess energy as light and heat.

The light output of the popular gallium arsenide phosphide LED (see Figure 4.19b) has a very narrow bandwidth centered in the red band. These lamps are very efficient and have an exceptionally long life but are small and have low light output.

The electrical characteristics of the LED are similar to those of the silicon diode. They are compatible with integrated circuits and operate directly from low-level logic circuits. Solid-state lamps, like neon lamps, perform satisfactorily under shock and vibration.

Virtual Lights

Virtual displays can be programmed to resemble all the familiar analog instruments, including running lights. The Microsoft Visual Basic programming language is popular in depicting virtual instruments. This language is popular because it is easy for third-party vendors to develop products that their customers can use with their Visual Basic programs. The virtual LED displays can mimic the functions of stand-alone LEDs.

CHECKLIST

1. Determine operating voltage.
2. Select lamp type and size.
3. Select lens for type, color, size, and shape. The last two features should be large enough to hold the necessary legends.
4. Select a lamp holder that is compatible with both lamp and lens. Also consider the allowable panel space and the methods of mounting and providing electrical connections.
5. Test the indicator under simulated operating conditions.

CONCLUSIONS

The possible uses of indicator lights to display information are limited only by the designer's imagination. There is usually one particular combination of lamp, lens, and lamp holder that is best suited for an application. Incandescent lamps are preferred for most applications because they are available in the widest range of light output, sizes, and voltages. The low light output of both neon and solid-state lamps limits their use to on/off indicators.

Amber lenses should be widely used because they absorb little lamp light and because amber is the most visible of all colors. Snap-in lamp-holders should be used wherever possible because they require no mounting hardware and take little assembly time.

Bibliography

Bylander, F. G., *Electronic Displays,* New York: McGraw-Hill Book Co., 1979.

Chichibu, S. F. (Ed.), *Nitride Semiconductor Blue Lasers and Light Emitting Diodes,* London: Taylor & Francis, 2000.

Degani, A., et al., "Modes in Human–Machine Systems," Human Factors Research and Technology Web page, NASA, http://human-factors.arc.nasa.gov, 2001.

Gadberry, B. E., "Designing Integrated Control System Displays," Paper #91-0354, 1991 ISA Conference, Anaheim, October 1991.

Jutila, J. M., "Guide to Selecting Alarms and Annunciators," *InTech,* March 1981.

McCready, A. U., "Man–Machine Interfacing for the Process Industries," *InTech,* March 1982.

National Electrical Code, Article 410, National Fire Protection Association, 1981.

NUREG/CR-6633, "Advanced Information Systems Design," U.S. Nuclear Regulatory Commission (USNRC), 2000.

Schubert, E. F., *Light-Emitting Diodes,* Cambridge: Cambridge University Press, 2003.

Shirley, R. S., et al., "What's Needed at the Man/Machine Interface?" *InTech,* March 1981.

Technical Report 1001066, "Human Factors Guidance for Digital I&C Systems and Hybrid Control Rooms," Palo Alto, CA, Electric Power Research Institute (EPRI), 2000.

Zukauskas A., et al., *Introduction to Solid-State Lighting,* New York: Wiley-Interscience, April 2002.

4.20 Recorders, Oscillographs, Loggers, Tape Recorders

F. D. MARTON (1970) **G. F. ERK** (1985) **B. G. LIPTÁK** (1995)
P. M. B. SILVA GIRÃO (2005)

Types of Designs and Movements:	1. Portable or laboratory, bench-top or flatbed 2. Strip-chart 3. Multipoint 4. Circular chart 5. *X–Y* recorder 6. Oscillograph/paperless recorder (light beam recorder) 7. Tape recorder 8. Loggers 9. Charts and accessories
Sensor mechanisms:	A. Potentiometric B. Galvanometric C. Linear array recorder
Costs:	A portable drum chart recorder for temperature or humidity costs $500. A portable 2.375-in. (60-mm) strip-chart recorder for event, mA, mV, V DC, V AC, or thermocouple inputs costs from $400 to $650. A 4-in. (102-mm) portable circular chart recorder is available with 2% full-scale error for temperature ($250), pressure ($350), or temperature and humidity ($500). An 8-in. (203-mm) portable, circular chart recorder with 0.75% full-scale error for temperature or 4 to 20 mA costs $750. A 12-in. (305-mm) digital circular chart recorder for direct temperature, 4 to 20 mA, or voltage costs from $1000 for one channel up to $2000 for four channels. A programmable 4-in. (102-mm) folding strip-chart recorder for millivolts, volts, or direct temperature with 0.5% error outfitted with one pen costs $1250; with two pens $1700; and with three pens $2200. A portable bench-top flatbed recorder with a 10-in. (254-mm) chart and 0.35% full scale error can be obtained with mV and V input for one ($1000), two ($1500), three ($2200), or four ($3100) channels; built-in RS-232C interface adds $500. *X-Y* recorder/plotters with 0.25% error and RS-232C and IEEE-488 interface cost about $4000; 24-point, 10-in. (254-mm) strip-chart potentiometers range from $4000 to $7000, depending on features. A data logger for eight channels costs around $100 and for 96 channels about $12,000. Paperless recorders cost from $650 to $4000 or more.
Partial List of Suppliers:	ABB Group (1, 2, 3, 4, 5, 6, 8, 9, A) (www.abb.com/global/USABB/usabb045.nsf!OpenDatabase&mt=html&l=us) Agilent Technologies (8) (www.agilent.com) Ampex Data Systems (7) (www.ampexdata.com) Anderson Instrument Co. (4, A) (www.andinst.com) Astro-Med Inc. (1, 2, 3, 8, 9, C) (www.astro-med.com) Avalon Electronics Ltd. (7) (www.avalon-electronics.com) Barton Instrument Systems LLC (1, 4, B) (www.barton-instruments.com) Bristol Babcock (1, 2, 3, 4, A, B) (www.bristolbabcock.com) Campbell Scientific Inc. (7, 8) (www.campbellsci.com) Chart Specialties Inc. (9) (www.chartspecialties.com/products.htm) Chino Corp. (1, 2, 3, 6, A) (www.chinoamerica.com) Cole Parmer Instrument Co. (1, 2, 3, 4, 5, 6, 8, 9, A) (www.coleparmer.com) Cooper Instruments & Systems (1, 2, 5, A) (www.cooperinstruments.com) Cybernetics (7) (www.cybernetics.com) Dickson (4, 8, 9) (www.dicksonweb.com/info/home.php)

Dresser Instruments (4,8) (www.dresserinstruments.com)
Dwyer Instruments Co. (1, 2, 4, 8, 9, A) (www.dwyer-inst.com/htdocs/DATA/qs-toc-data.cfm)
Endress+Hauser (2, 3, 6, 9, A) (www.us.endress.com)
Enercorp Instruments Ltd. (2, 3, 6, A, B) (www.enercorp.com)
Graphtec Instruments (1, 2, 3, 5, 6, 8, 9, A, B, C) (www.westerngraphtec.com)
Gulton (9) (www.gulton.com)
Honeywell Industrial Measurement and Control (1, 2, 3, 4, 5, 6, 9, A) (www.content.honeywell.com/imc)
Invensys Eurotherm (1, 2, 3, 4, 5, 6, 7, 9, A, C) (www.eurotherm.com)
Invensys Foxboro (1, 2, 3, 4, 6, A, C) (www.foxboro.com/instrumentation)
Jumo Process Control Inc. (1, 2, 3, 6, 9, A, B) (www.jumoprocesscontrol.com)
Kipp & Zonen (1, 2, 5, 9, A) (www.kippzonen.com)
Linseis International (1, 2, 3, 4, 5, 8, 9, A, C) (www.linseis.com)
Martel Electronics Sales Inc. (7) (www.martelelectronics.com)
Metrum Information Storage (2, 6, 7, 8, A) (www.metrum.co.uk)
Nicolet Gould Instrument Technologies (1, 2, 6, 8, 9, C) (www.niti.com/docs/home.php?id=1)
Ohkura Electric Ltd. (2, A) (www.ohkura.co.jp)
Omega Engineering Inc. (1, 2, 3, 4, 5, 6, 8, 9, A) (www.omega.com)
Palmer Wahl Instruments Inc. (4, 8, A) (www.instrumentationgroup.com)
Pyrometer Instrument Co. (1, 2, 3, 6, 8, A) (www.pyrometer.com)
The Recorder Co. (1, 2, 5, 8, 9, A) (www.recordercompany.com/page1.htm)
Ronan Engineering (8) (www.ronan.com)
RMS Instruments Ltd. (1, 2, 7, 8, 9, A) (www.rmsinst.com)
Siemens Energy & Automation (6) (www.sea.siemens.com/default.asp)
Soltec Corp. (1, 2, 3, 4, 5, 6, 7, 8, A, C) (www.solteccorp.com)
Sony Corp. (7) (www.sony.com)
Teac America Inc. (7, 8) (www.teac.com)
Thermo Electron Corp. (2, 3, 6, 8, A) (www.thermo.com)
Toshiba America Inc. (6, 7, 8) (www.toshiba.com)
Yokogawa Electric (1, 2, 3, 5, 6, 8, A, B, C) (www.yokogawa.com/cms/com/ep/home.do)

INTRODUCTION

The recording of events, trends, or variables is an important part of process control. With the increase of computer-based process control and information digitized and displayed mainly in cathode-ray tubes (CRTs) and liquid crystal-based displays (LCDs), temporary recordings are often made using the computer's memory and permanent recordings are made using a mass-storage unit (e.g., hard disk, CD, tape).

When information on paper is required, the easiest and cheapest way is to use printers connected to the computer. This type of solution is surely the state of the art and probably seen in control rooms of distributed control systems. Nevertheless, both for local and for less complex applications, other means of variable recording for event and trends detection are used.

Information can be recorded either in analog or digital format. In the first case the quantity to be recorded is continuous in time and the recorder (analog type) must have the bandwidth required for reproduction without distortion. Digital recording uses sampled data and the information recorded corresponds to discrete values of time (digital recorders). Some recorders, called hybrid, are of the analog type insofar as printing is concerned but also output data in digital format. Both digital and hybrid recorders operate under microprocessor control, and features presently available include remote operation through instrumentation interfaces (e.g., IEEE488, Fieldbus).

This section describes the various recorder designs; printers are covered in Sections 4.17, "Human–Machine Interface Evolution," and 4.24, "Workstation Designs and Features"; CRTs are discussed in Section 4.5, "CRT Displays."

Recorders can be classified according to their measurement signals (electronic, electrical, pneumatic, thermal), according to their sensor mechanism (potentiometric, open loop, galvanometric, linear array), according to their application (portable, laboratory, flatbed, benchtop, industrial, panel-mounted), according to the type of chart used (strip roll, fanfold, circular, multipoint, *X–Y* plotter), or according to the recording technique applied (pressure, electrostatic, light beam, ink, electric, thermographic). The discussion below begins with descriptions of the most widely used sensing mechanisms.

SENSOR MECHANISMS

Practically all the indicator movements that have been discussed in Section 4.18, "Indicators, Analog Displays" can also be used to operate analog recorders. Here only the most widely used designs — the galvanometric (including light-beam),

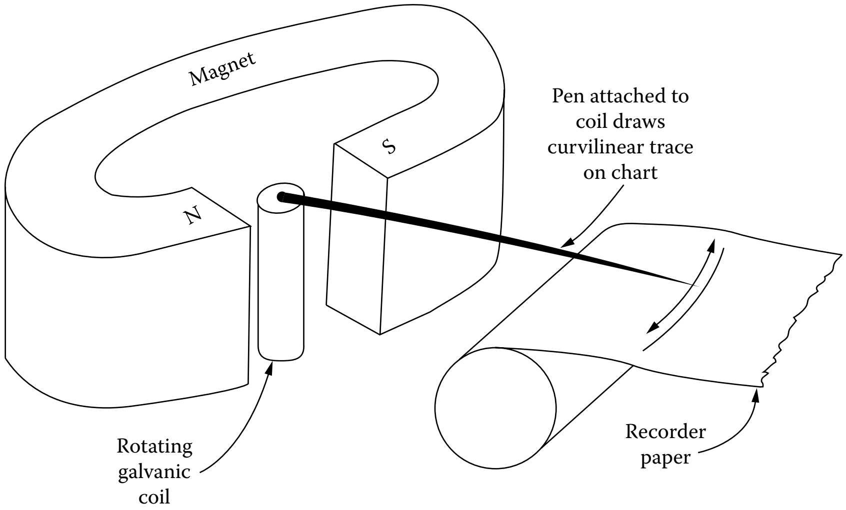


FIG. 4.20a
The recorder pen of a galvanometer can be connected directly to the rotating armature.

potentiometric, open loop, and linear array type recorders—will be described.

Galvanometric Recorders

Figure 4.20a illustrates a direct-writing galvanometer on which the pen is connected to the armature of the galvanometer and therefore rotates with the rotating-coil or moving-iron type armature. The writing system can use carbon transfer, thermal, pressure, electro-discharge, ink, or ink-jet methods. The error sources in galvanometers include linearity, tangential error, drift, and hysteresis. The torque capability of galvanometers limits their frequency response to about 100 Hz in a 10 mm (0.4 in.) wide channel. Multichannel recorders can be obtained by placing several galvanometers side by side. Galvanometers are relatively inexpensive, require little maintenance, and are suited for operation in extreme temperature (high or low) and humidity environments. Nevertheless, they are seldom used today.

Light-Beam Recorders (Oscillographs)

When the armature of the galvanometer is used to rotate a mirror, the unit is called a light-beam-type recorder (Figure 4.20b). As a change occurs in the electric measurement signal, which travels through the coil, this change causes the mirror to rotate. The main components include a light source (usually ultraviolet), the galvanometer with a rotating mirror, an optical system, and a light-sensitive recording chart. A recorder that has several channels, using several galvanometers, can record several measurements simultaneously.

When ultraviolet-sensitive recording film or paper is used, a permanent record is obtained instantaneously, without a need for any further processing. A galvanometer (oscillograph) with several channels can print numbers on the edge of the record to identify the corresponding channels. The light-beam-type galvanometer has a much higher frequency response (up to 25,000 Hz) than does the direct recording galvanometer (under 1000 Hz).

Recorders based on the above-mentioned principles are also rarely seen in use today. In present usage the word “oscillograph” refers mainly to an instrument that incorporates the functions of an oscilloscope and of a printer and, less commonly, to paperless recorders.

Potentiometric Recorders

Today, most analog recorders are of the potentiometric type. Figure 4.20c illustrates a servo-operated null-balance potentiometric recorder. Here the amplifier drives the balancing servomotor (usually a DC motor), which is mechanically connected to the writing pen and the feedback slide-wire. Because there is no current flow when the unit is in balance, load resistance variations or lead-wire resistances have no effect. This allows the sensor to be located a long distance from the recorder.

Potentiometers are sensitive down to the microvolt level and are accurate to within 0.25% of span. Because most process variables can be detected in terms of micro- or millivolt signals, potentiometers can be used to record most variables. When used with a null-balancing bridge input, it can measure resistance, inductance, and capacitance, in addition to electromotive force (emf).

The main limitations of potentiometric recorders are cost and speed of response. The inertia inherent in the recording system limits conventional null-balance recorders to a full-scale pen travel of about 0.5 seconds. Although at low frequencies (few Hz), the potentiometric recorder is the most accurate unit available, its speed of response is a limitation. If the belts, pulleys, gears, and rotary servomotors of a conventional potentiometer are replaced by a high-speed servo (the armature of this servomotor moves back and forth as a shuttle), the speed of response can be increased by a factor of 2 to about 0.2 seconds and hysteresis can be reduced.

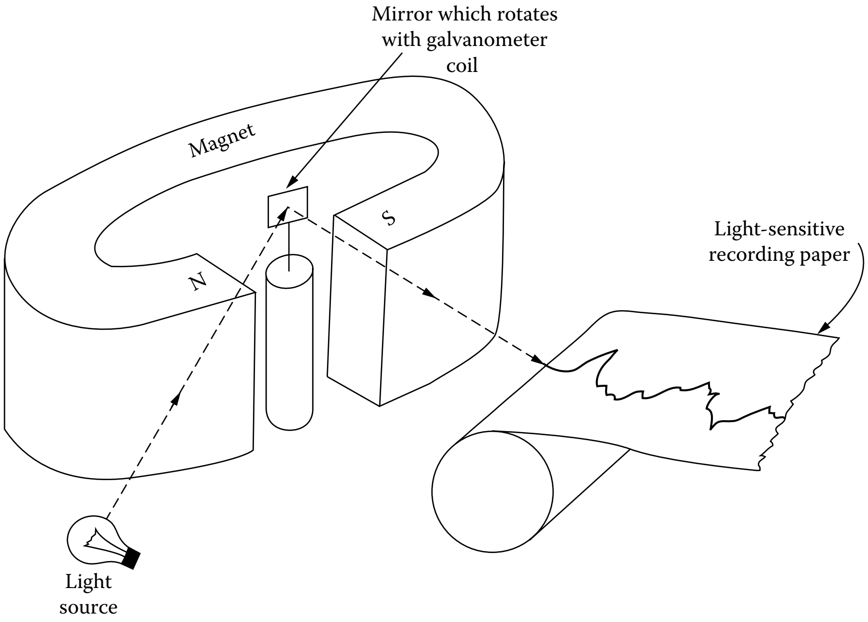


FIG. 4.20b
A recorder design in which a mirror is connected to the rotating armature of the galvanometer and a UV light detects the mirror position is often called a light-beam-type or oscillograph-type recorder.

Open Loop Recorders

In digital recorders, signals to be plotted are digitized. The mechanisms that actuate the writing element(s) are then usually driven not by DC servomotors but by step-motors. The drivers are microprocessor controlled and the system operates in open loop. In a typical arrangement, a step-motor produces paper movement in one direction and one or several step-motors actuate the writing elements in an orthogonal direction. This setup permits the representation not only of time-varying quantities but also an *X–Y* plot if the paper is allowed to move in either sense. In fast-responding strip-chart digital recorders the paper can be alternatively actuated by a DC motor.

Linear Array Recorders

In these recorders the recorder pen has been replaced by a fixed (nonmoving) array of small recording elements. The recorder chart moves under this linear array of “stylii,” each one of which corresponds to a particular measurement value to be recorded. If a 0.25% resolution is desired, a 0 to 100% range will be incremented among 400 fixed stylii, each of

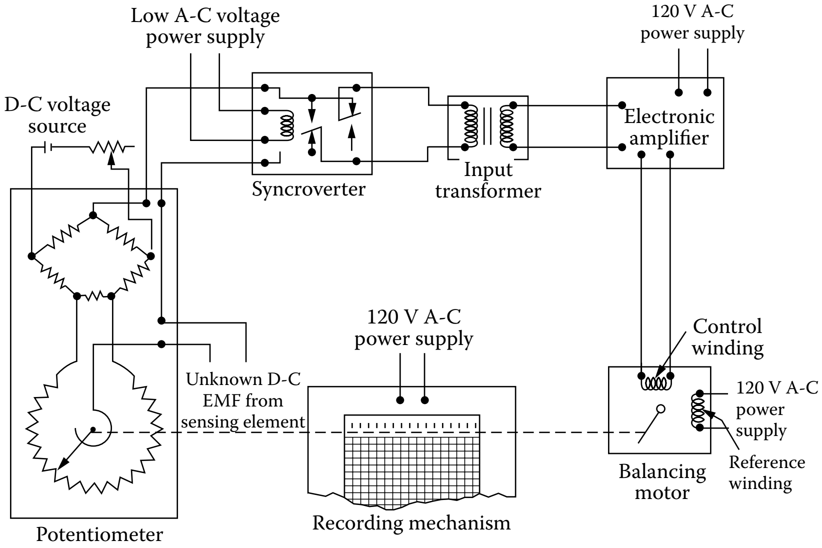


FIG. 4.20c
One of the traditional potentiometric recorder designs.

which represents a particular measurement signal increment. Every print cycle the drivers select the stylii to be activated to print a dot pattern forming the alphanumerics, grids, and signal traces. As this process is repeated every cycle, a continuous recording results.

The main advantage of linear array recorders is the absence of moving parts (other than the paper). The array elements can be thermal or electrostatic. When thermal elements are used, the paper is coated with a thin film of thermoreactive material on a substrate. When electrostatic elements are used, the chart paper is dielectric; in addition to the paper, a liquid toner is required.

The records resulting from the electrostatic technique last longer and require less energy than do thermally obtained recordings. Paper costs are similar for the two methods: about 5 to 10 cents per foot of paper. The maximum width of the print ranges from 10 to 15 in. (254 to 314 mm), chart speeds range from 1 mm/hr to 0.5 m/sec, and the printing speed can approach 2000 lines per second.

RECORDING METHODS

There are many ways of leaving a mark on a chart, including the various ink-writing systems and the inkless methods of thermographic, electric, pressure, light-beam, and electrostatic marking.

Ink-Writing Systems

The early pen-and-ink recorders used bucket, V, fiber-tipped, or ballpoint pens and pressurized or gravimetric capillary-type feeding systems. Refilling was performed by the use of disposable ink cartridges, and maintenance was performed by the use of replacement pen assemblies. In high-speed oscillographs either pressurized inking systems are used (requiring special coated paper) or high-cost ink-jet hardware is installed.

Impact printing initially used a carbon ribbon and later slow-drying inks stored in porous pads; when the pointer mechanism presses down, it leaves a single-color or multicolored mark. More recently, the pointer has been replaced with a traveling print wheel, which is engraved with numbers or symbols that identify the variables being recorded. This type of marking is used on most multipoint strip chart recorders, which are usually provided with an 11-in. (279-mm) chart and can simultaneously record about 20 variables.

Inkless Systems

Thermographic recorders use a heated stylus that melts or chemically changes the heat-sensitive coating on the paper. In the linear array version, some 420 fixed thermal-printing elements (stylii) can be used on a 4.5-in. (114-mm) paper chart. Each element is capable of generating some 100 dots per inch (4/mm). When used with up to six channels as a multipoint recorder, the channels are identified by different letter characters, while the date and time are periodically printed onto the recording.

Pressure-type recorders were one of the earliest designs, in which a sharpened stylus was used to remove smoke or other special coating from the chart, thereby exposing a contrasting color underneath.

In electric recorders an electric discharge generated by the stylus etches the record into an aluminum coating on a black paper, thereby exposing the black substrate.

The light-beam-type recorder is illustrated in Figure 4.20b. In this recorder a photosensitive paper or film is used to record the location where the high-intensity light impinges on the chart. In order to eliminate the effects of ambient light, ultraviolet sources rather than visible light sources have been used. If a programmable light gate array is used to record directly on ultraviolet (UV)-sensitive paper, the error contributions of linearity, beam deflection, tangential error, inertia, and overshoot can be reduced.

In the electrostatic recorder, there is an imaging head with some 1000 wire elements, spaced at 4/mm, giving a total length of 10 in. (254 mm). As the recording chart moves over the image head, a negative voltage is applied to selected wires while positive voltage is applied to “shoes” on the other side, giving a positive point charge to the paper chart at the activated points. Next the chart proceeds to the toner head, where negatively charged ink particles adhere to the charged points on the paper. Following this step a vacuum knife removes the excess toner particles, and when the remaining particles are exposed to air, they permanently bond to the paper.

Paperless Systems

Paperless recorders (Figure 4.20d) are multichannel data acquisition systems that present the information on a display similar to a computer monitor, usually of the LCD type. Some of them have also paper printing and permanent data storage capabilities. They are light-beam-type recorders, much like digital sampling oscilloscopes (DSOs), but as they are acquisition systems they can also be considered data loggers. Thus, paperless recorders can be found in suppliers identified both by 6 and 8 on the partial list of suppliers at the beginning of this section.

CHARTS AND COORDINATES

The process variable actuates a recording mechanism, such as a pen, that moves across a chart. The chart moves constantly with time. These two motions produce an analog record of variable versus time. Any point on the continuous plot obtained in this manner can be identified by two values, called coordinates. Many coordinate systems are in use, with the Cartesian coordinates most widely used in industry. If the reference lines are straight and cross each other at right angles, they are called rectilinear coordinates. If at least one of the reference lines is an arc of a circle, the coordinates are curvilinear.

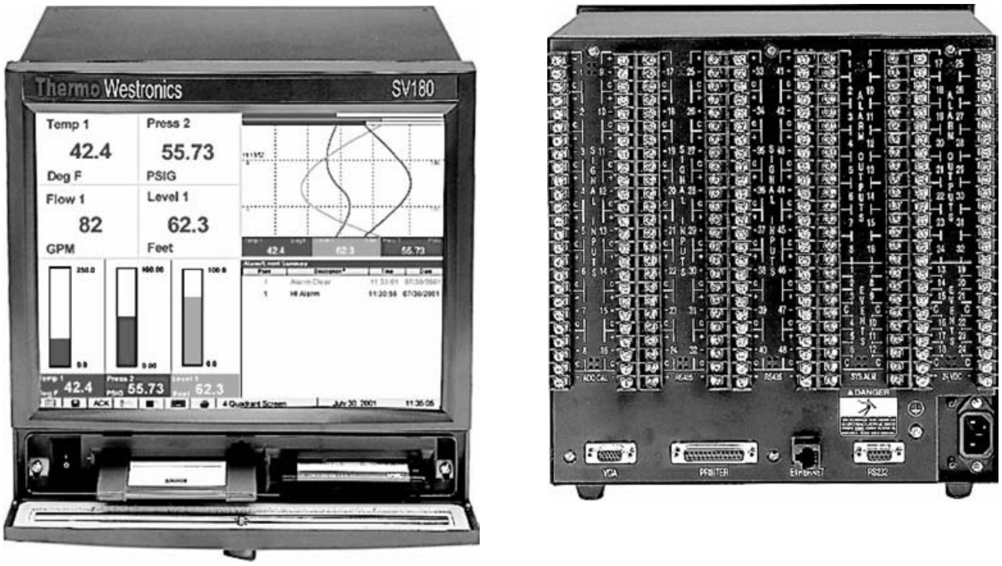


FIG. 4.20d
Front and rear views of a touch-screen paperless recorder with input/output handling and communications capabilities, as well as flexible data storage options. (Courtesy of Thermo Electron Corporation.)

The shape of the chart provides a primary means of classification into (1) circular charts and (2) rectangular charts, in sheet or strip form. Strip charts can be torn off and can be stored in rolls or folded in Z folds. Strip chart lengths vary from 100 to 250 feet (30.5 to 76.2 m).

Circular Chart Recorders

Circular chart recorders are still used because they are simple and, consequently, low priced. Chart speeds are a uniform number of revolutions per unit time. Chart drives making one revolution in 24 hours are standard, but instruments are also equipped with drives of 15 minutes; 1, 4, 8, 12, or 48 hours; and 7, 8, or 28 days per revolution.

The best applications for portable circular chart recorders are temperature or humidity recordings used in heating and ventilating systems.

The circular chart recorders shown in Figure 4.20e produce concentric circles crossed by circular arcs whose center, when the pen crosses the arc, is the same as that of the recording arm. The tangents at the intersection points of the two curves should be as near to a right angle as possible for best readability. The concentric circles form the scale on which the variable is read. The “time arcs” divide the

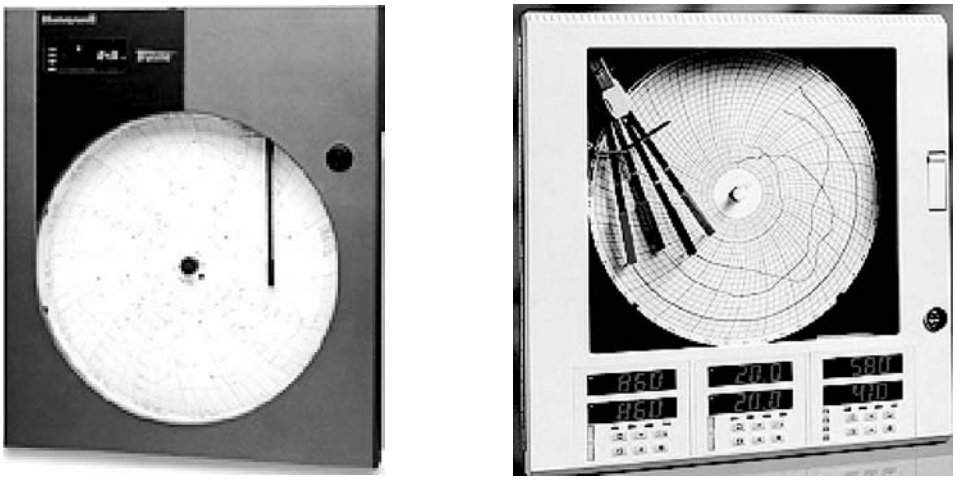


FIG. 4.20e
Circular chart recorder. (Courtesy of Honeywell Industrial Measurement and Control and Omega Engineering, Inc.)

full circles into appropriate uniform intervals of the total period.

Chart diameters vary from 3 in. (75 mm), 4 in. (100 mm), and 8 in. (200 mm) to up to 12 in. (300 mm). Circular charts offer the advantage of a flat surface. Special features, such as automatic chart changers, allow collection of daily records in a continual manner. Contacts can also be mounted in the pen mechanism for alarms on preset high or low signal conditions. One of the limitations of circular chart recorders is that the recording of more than two variables on one chart poses design problems.

Strip-Chart Recorders

Strip-chart recorders are characterized by the uniform linear motion of the paper either horizontally (Figure 4.20f) or vertically (Figure 4.20g). The measurement lines can be straight or curved, furnishing either rectilinear (Figure 4.20g) or curvilinear recordings.

The curvilinear method uses very simple linkage geometry and offers better readability than is obtained with circular charts. Figure 4.20h shows a two-pen recorder (second from left) and a recorder controller. Whereas the two-pen recorder records two independent variables, the recorder controller accepts one sensor signal, compares that signal with a manually operated set point, and provides a signal output for closing a control loop by actuating a final control element.

Strip-chart recorders are available in several standardized sizes, including 4-in. (Figure 4.20i), 6-in., and 11-in. (102-mm, 152-mm, and 279-mm) charts. The standard miniature strip-chart recorder is used either in the wide (6 by 6 in. or 152 by 152 mm) or in the narrow format (3 by 6 in. or 76 by 152 mm). Chart speeds are available from 1 in./hr to 30 in./min in adjustable increments (1.0 in. or 25.4 mm). These strip-chart recorders are available with up to three pens and with up to two alarm contacts per pen. The accuracy of recording is usually within 0.5 to 1% of span.

Multiple Recorders

When several variables are to be recorded on the same chart, such as several temperatures from thermocouples in various locations, multiple recorders are used. Circular chart recordersmay handle only up to four variables, whereas strip-chart recorders handle as many as 24 to 36 measurements. To identify each variable, symbol or numerical coding or color printing is used, as well as full digital alphanumeric printouts on the chart margin.

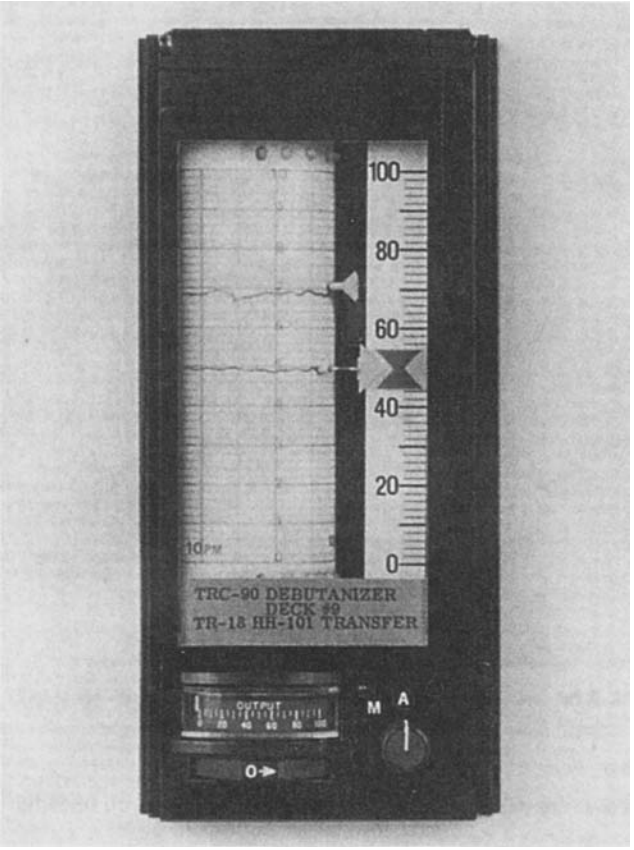


FIG. 4.20f
Horizontal strip-chart recorder.

With the advent of microprocessor technology, multivariable recorders became available. These allow recording of a multitude of variables, such as flow and the associated temperature and pressure. Likewise, for comparison of several variables on the same time scale without line crossing or overlapping, multichannel recorders are available.

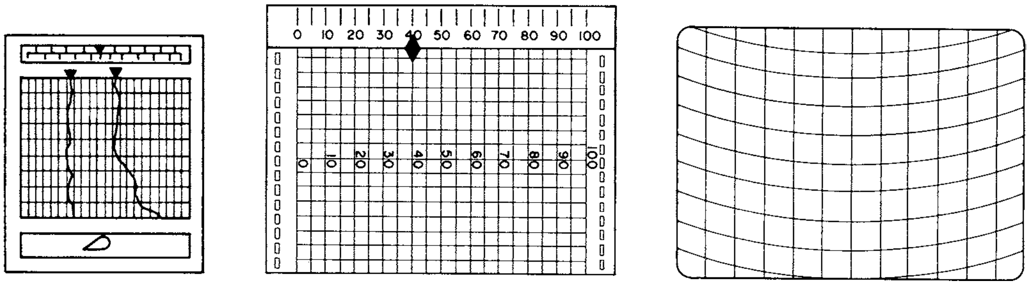


FIG. 4.20g
Strip-chart recorders with vertical movement.

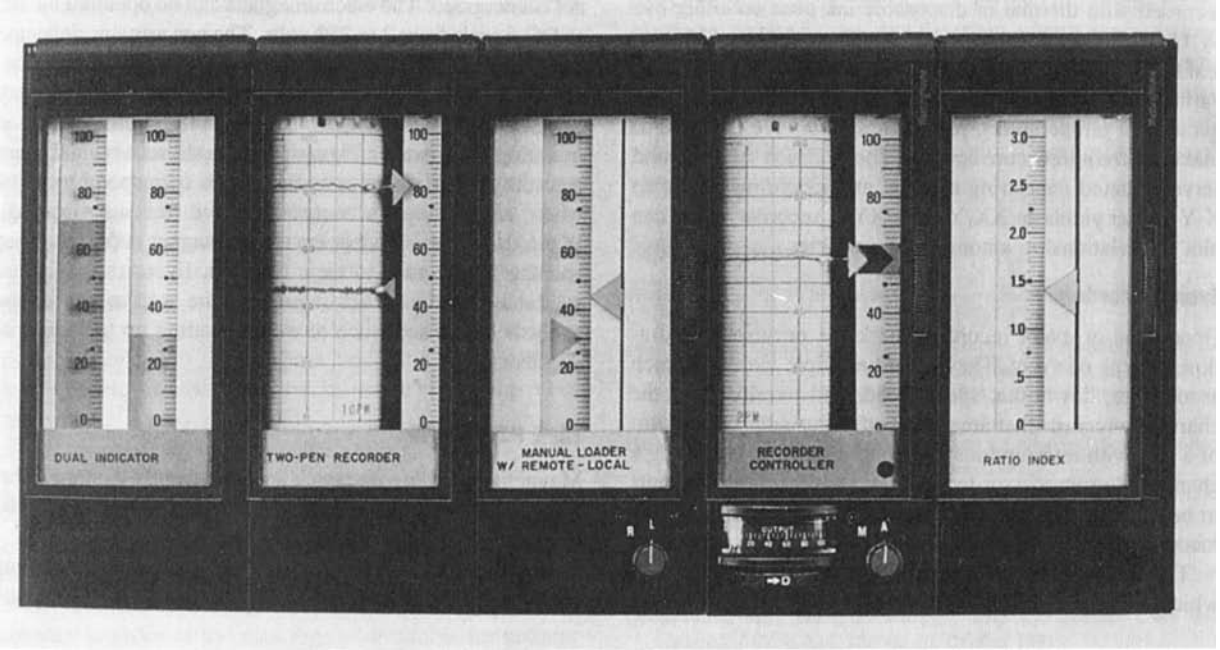


FIG. 4.20h
Two-pen recorder and recorder controller.

Figure 4.20j shows the cut-out requirements and the front dimensions of typical 6-in. and 11-in. (152-mm and 279-mm) multipoint strip-chart recorders. With the addition of microprocessors, high-speed hybrid recorders became available with up to 30 recording channels.

Inputs can be DC volts or millivolts, thermocouples, and resistance temperature detectors. On these units the digital measurement data are printed on the left margin of the chart in up to six colors. The basic accuracy of these units is 0.25% of span. Most hybrid recorders are provided with a keyboard on the front panel, which allows for the programming of chart speed, alarm set points, full scale range, and other features.

FIG. 4.20i
Four-inch (100-mm) strip-chart, multipoint recorder. (Courtesy of Honeywell Industrial Measurement and Control.)

X–Y Recorders

X–Y recorders plot two variables simultaneously, such as stress versus strain or temperature versus pressure. Either the chart is stationary and the scriber is moved along both the abscissa and the ordinate by the two signals, or the chart is moved in one direction while the stylus slides on an arm in the other direction.

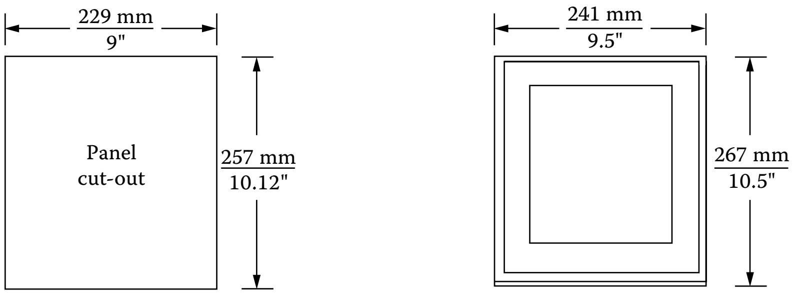
The signals entering the function plotter can be analog or digital. Digital signals require transducers to obtain an analog plot. Likewise, digital records can be provided with analog-to-digital conversion and conventional digital printout.

There are also combination function plotters, called X–Y–Z_1 recorders. Some allow the pen to be driven along either axis at a constant speed, thus making recordings of X versus Y, Y versus T (time), and X versus T possible. Recorders with three independent servo systems allow the recording of two variables against a third.

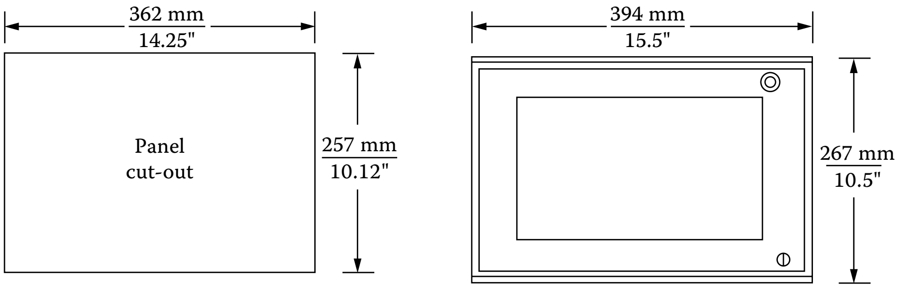
X–Y recorders usually have flat beds with measurements ranging from $8^1/_2$ by 11 in. (216.5 by 279 mm) to 45 by 60 in. (1143 by 1524 mm). Some record on drums. Recorders that print from the back or on a glass screen do not obstruct visual observation.

X–Y plotters can be connected to computers using either RS-232C (serial) or IEEE-488 (parallel) ports and can send or receive data with 0.02 mm resolution. Data are most often recorded with thermal or disposable ink pens on either $8^1/_2$ by 11 in. (216 by 279 mm) or 11 by 17 in. (279 by 432 mm) leaf or roll paper.

When the X–Y recorder is used to measure variables other than voltage (for example, current, frequency, or temperature),



6" (150 mm) strip chart



11" (275 mm) strip chart

FIG. 4.20j
Panel cutout and front dimensions of multipoint strip-chart recorders. (Courtesy of Honeywell Industrial Measurement and Control.)

plug-in amplifiers are installed to make the required conversions. The addition of a second servo-actuated measuring element and recording pen to an X–Y plotter yields an X–X_1–Y or an X–Y–Y_1 recorder that can plot the relationship among three variables.

Event Recorders

Operations or event recorders mark the occurrence, duration, or type of event. They record multiple incidents, such as on-time, downtime, speed, load, and overload, on the chart. The records that are produced are usually in the form of a bar, with interruptions in a continuous line indicating a change. Microprocessor technology allows scores of points to be scanned every millisecond, with high-speed printouts made for the events that occur.

The classic event recorder is a simple instrument (Figure 4.20k) in which the pen is deflected when the associated electromagnet is energized. The electromagnets can be operated by AC or DC signals from 2 to 250 volts. The pen remains deflected until the associated electromagnet is deenergized. The resulting recording shows not only that an event has occurred, but also its time, duration, and possible consequences in initiating other events.

Figure 4.20k also shows a typical event recording chart where open rectangles correspond to times when electromagnets were energized (events occurred). When the event is ended, the electromagnet is deenergized and the pens return to their baseline. Event recorders are available with up to 128 inputs and are well suited for the analysis of the sequence of events leading up to industrial accidents.

TAPE RECORDING

Today, permanent storage of information, either in analog or digital formats, uses different technologies such as battery-operated random access memories, magnetic discs, tapes, and compact optical discs (CDs). In process control, magnetic recording on tape is frequently used either for analog or digital form signals.

Videotape recording is similar to magnetic tape recording with the additional feature of reproducing both picture and sound. Industrial application is found in closed-circuit television monitoring.

Magnetic tape recording is done by magnetizing fine iron oxide particles, which are coated onto a polyester plastic tape. The data can be recorded by fixed heads on multitrack longitudinal recordings in analog or pulse-code-modulated (PCM) form, or the data can be recorded by helical scan techniques, which record analog or digital data in VHS, Beta, or R-DAT formats. For process-control-related recording the multi-track fixed-head recording has been the most popular. On a 0.5 in. videocassette tape, usually no more than 24 recording tracks are available. Helical scanning reduces this track width of 25 mils to about 2 mils and thereby allows a track density of about 250 tracks per 0.5 in. in the VHS format.

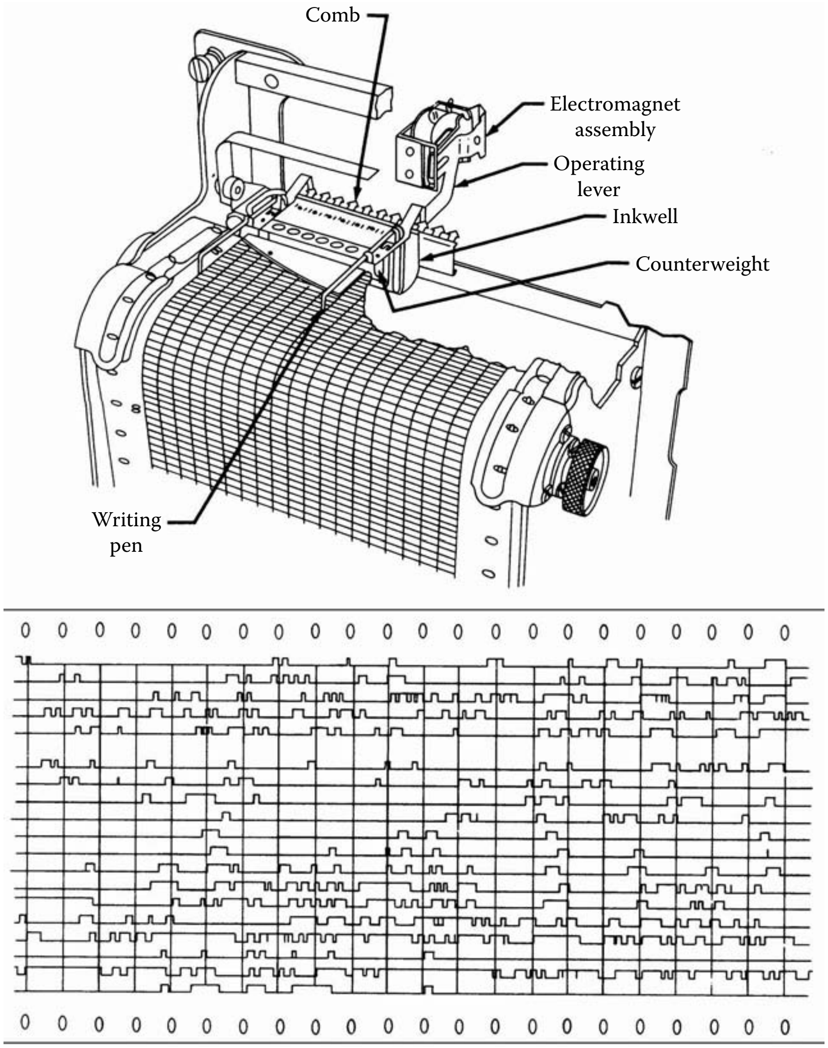


FIG. 4.20k
The design of an event recorder and the appearance of a typical event record chart. (Courtesy of Esterline Angus Instrument Corp.)

DATA LOGGERS

Data loggers are electronic instruments that record measurements of different quantities such as temperature and relative humidity or events over time. They can be software packages loaded into personal computers, DCS systems, or other computing devices that serve to operate printers, but more often they are stand-alone battery-powered units that are equipped with a microprocessor, data storage, a printer and sensors.

Most data loggers utilize turn-key software on a personal computer to initiate the logger and view the collected data. High-speed recording is achieved by combining electronic readouts with electrostatic printing. Digital recorders can print as many as 1 million characters per minute on paper in digital form. For applications such as electric power generation, the high-speed digital recorder is the best choice.

Some stand-alone microprocessor-based data loggers, commonly called paperless recorders, record data in a processor's memory; they do not incorporate printing capabilities but are provided with a screen where charts are presented.

A low-cost solution to data logging is to use a personal computer to control the printer, which shows the trend of the process or the results of a test. Most data-logging systems are provided with capability both for front panel programming and for switch-selectable remote programming, implemented over RS-232C, IEEE-488 fieldbus connections from remote terminals. The logged data can be linearized or scaled; the data can be averaged by time or by groups; and up to four alarm set points can be programmed onto each channel being monitored.

The typical scanning capacity of stand-alone data loggers is up to 128 channels. These units are also provided with an automatic restart feature that functions after a power failure and with RS-232C and passive current loop ports for computer interconnection. Central data loggers are usually integrated into the total control system of the plant and consist not only of a complete data acquisition system but also of some control capability.

Stand-alone battery-operated data loggers are an interesting alternative to analog chart recorders because they are smaller, less expensive (base price and no consumables required), more accurate, and more reliable. On the other hand, because information is digitized, data analysis, presentation, and storage are simplified. The reduced size of some models allows their operation in areas inaccessible to other recording systems.

Bibliography

Anderson, L., "X-Y Recorders," *Measurements and Control,* October 1992.

Bailey, S., "Recorders and Indicators Exert a Cohesive Force in Online Quality," *Control Engineering,* January 1991.

Chiranky, L., "A Case for Linear Array Recording," *Instruments and Control Systems,* July 1982.

Dobrowolski, M., "Guide to Selecting Strip and Circular Chart Recorders," *InTech,* Vol. 28, No. 10, October 1981.

Dyer, S.A. (Ed.), *Survey of Instrumentation and Measurement,* New York: John Wiley & Sons, 2001.

Gadberry, B. E., "Designing Integrated Control System Displays," Paper # 91-0354, 1991 ISA Conference, Anaheim, October 1991.

Gambino, R. J., and Suzuki, T. (Eds.), *Magneto-Optical Recording Materials,* New York: IEEE Press, 2000.

Harold, B., "Development of a Realtime Mass Balance Data Acquisition System," 1992 ISA Conference, Houston, October 1992.

Jorgensen, F., *The Complete Handbook of Magnetic Recording,* New York: McGraw-Hill, 1996.

Krigman, A., "Multipen and Multipoint Recorders: Making the Technology Work for You," *InTech,* August 1984, pp. 39–48.

Mallinson, J. C., *The Foundations of Magnetic Recording,* Boston: Academic Press, 1993.

Mason T., and Kayes, E., "Tape Recorders," *Measurement and Control,* October 1992.

McDowell, W. P., "Microprocessor Controlled Multipoint Recorder with Graphical Video Output for Reactor Core Temperatures," *IEEE Transactions of Nuclear Science,* Vol. NS-27, No. 1, February 1980.

Morris, A. S., *Measurement and Instrumentation Principles,* Oxford: Butterworth-Heinemann, 2001.

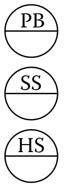
Nylen, P., "Electronic Device for Compensating for Pen Displacement in a Multiple Channel Strip-Chart Recorder," *Review of Scientific Instruments,* Vol. 51, No. 3, March 1980.

Plumer, M. L., van Ek, J., and Weller, D. (Eds.), *The Physics of Ultra-High-Density Magnetic Recording,* New York: Springer, 2001.

Smith, G. M., "Data Management Systems," *Measurements and Control,* September 1991.

4.21 Switches, Pushbuttons, Keyboards

D. W. LEPORE, R. A. WILLIAMSON (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbols

Types:	A. Switching components: A1. Pushbuttons A2. Toggle switches A3. Rotary switches A4. Thumbwheel switches B. Data entry systems: keyboards
Features:	A1. Keyboard, panel, industrial oil-tight A2. Lever rocker, thumbwheel selectors A3. Selector, adjustor A4. Selector
Actuation:	A1. Push in A2. Pivot A3. Twist A4. Push up and down
Sizes:	A1. Standard, miniature A2. Standard, miniature, subminiature A3. Standard, miniature A4. Standard
Costs:	A1. $30 to $150. A single, heavy-duty, general-purpose (NEMA 1) button in flush mounting pull box costs about $125; for NEMA 4 watertight requirements add $25, and for NEMA 7 explosion-proof requirements add $50 A2. $10 to $50 A3. $50 to $300 A4. $15 to $30 B. Keyboards are usually included in the cost of workstation terminals
Partial List of Suppliers:	Ad Products Co. (A) (www.adproductsco.com) Adalet (A) (www.adalet.com) Allen-Bradley Co. (A,B) (www.ab.com) Ann Arbor Technologies (B) (www.a2t.com) ASCO (A) (www.ascovalve.com) Automated Control Concepts Inc. (B) (www.opstation.com) Cutler-Hammer (A,B) (www.cutlerhammer.eaton.com) Fuji Electric Corp. of America (A,B) (www.fujielectric.com) GE Fanuc Automation (A) (www.gefanuc.com) HMW Enerprises (B) (www.hmwent.com) Honeywell Sensing and Control (A,B) (www.honeywell.com/sensing) IEE Industrial Electronic Engineers (B) (www.ieeinc.com) Intellution Inc. (B) (www.intellution.com) Nematron Corp. (B) (www.nematron.com) NKK Switches (A) (www.nkkswitches.com) Pepperl + Fuchs Inc. (A) (www.am.pepperl-fuchs.com) Schneider Electric/Square D (A,B) (www.squared.com) Siemens Energy & Automation Inc. (A,B) (www.sea.siemens.com) StacoSwitch (A) (www.stacoswitch.com) Westinghouse Process Control (A,B) (www.westinghousepc.com)

INTRODUCTION

In this section both the traditional and the more advanced manual switching elements are discussed. The traditional switches and pushbuttons are used to start and stop or open and close final control elements, while the more advanced microprocessor-based touch-panel keyboards are used in all types of information-handling systems and in computer interfaces. In keyboard-type applications, solid-state switching circuits operate either on the Hall effect or by using contact or capacitance-type membrane switches. Light-emitting diodes (LEDs) illuminate switching devices or provide status indication.

In selecting switching devices, one might consider such factors as the requirement for panel and display area. Other factors include the type of switching action required (SPST, SPDT, DPDT, momentary or maintained), contact ratings (from microamperes to 15 A, and from 5 V DC to 600 V AC, with 5 A at 125 V AC being a common industrial selection), service life (ranging from 1 million to 30 million operations), the type of terminations (quick-connect, screw, solder, push on) and the type of housing required (NEMA 1, 4, 7, etc.).

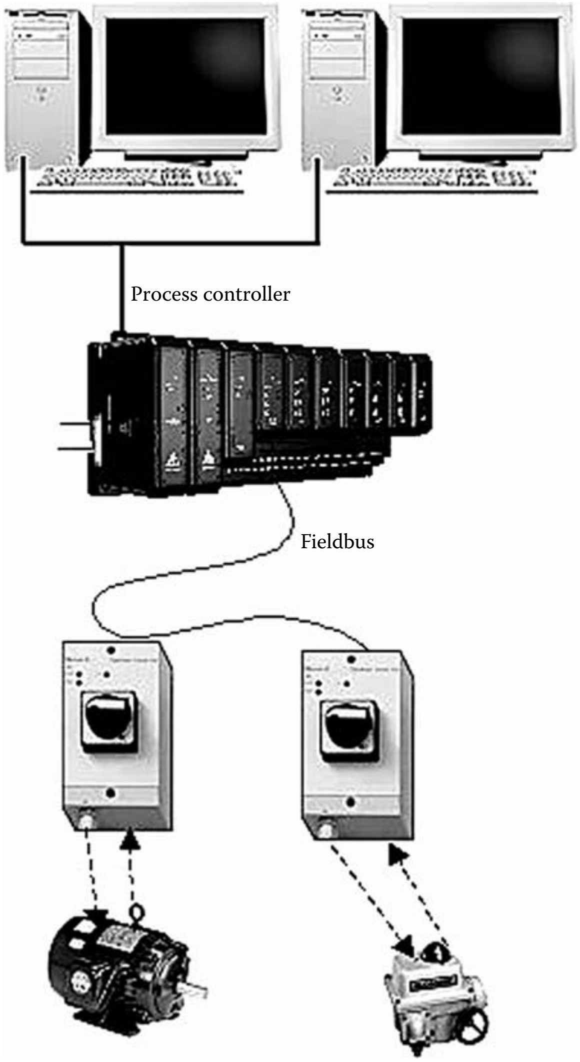
In the area of motor-starter designs, significant advances have been made. These days the packaging is more compact, and the motor starter includes some form of intelligence. The fault diagnostic capability usually includes the detection of power loss, short-circuit sensing, phase imbalance and phase current detection, full-load current, and thermal capacity. These features contribute to the ease of troubleshooting and quick restarting. Most motor-starter manufacturers also provide integrated interfaces for communication on the plant's bus network (Figure 4.12a), while the hand-off-auto switch or pushbuttons remain local, near the motor. This design reduces the amount of wiring and therefore the cost of installation, while the simplified connections and the self-diagnostic features tend to lower the cost of maintenance.

SWITCH DESIGNS AND OPERATION

A manual electric switch (Figure 4.21b) consists of the switching contacts, an actuator to bring them into proximity, terminals to connect the contacts to the conductor of an electric circuit, the insulating mounting provisions, and the enclosure, either separate from or integral to the mounting for contacts and terminal.

The gap between contacts is filled by air or by an inert gas. The control is operated by moving the contacts closer together until the applied voltage causes an electrostatic failure of the gap. The ensuing electron flow heats the cathode and cools the anode contact until molecular welding occurs, creating a continuous metallic path and completing the circuit.

Contact resistance is a measure of the degree to which the insulating gas remains between the contacts, preventing welding over their entire bearing area. It is directly proportional to contact pressure.



Fieldbus motor starters

FIG. 4.21a
Motor starter with local controls and fieldbus interfacing for self-diagnostics and automatic control.

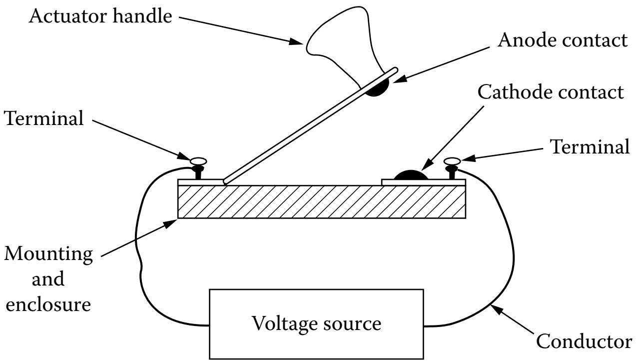


FIG. 4.21b
Operating elements of a manual switch.

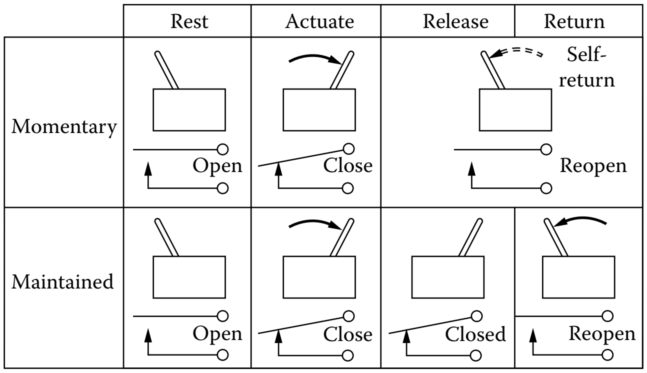


FIG. 4.21c
Basic switching actions.

Switching Action

Switching action refers to the positions assumed by the contact in response to actuator motion. The two basic types (Figure 4.21c) are momentary action, which provides one contact closure and subsequent reopening with one actuation, and maintained action, which requires separate actuation to transfer the contacts from one extreme position to another and back.

Other special types include:

1. Mechanical bail, which allows operation of only one control at a time, physically inhibiting depression of others. Automobile radio selectors are a typical example.
2. Electrical bail, which allows delayed or remote actuation, including sequencing of several controls. Solenoid operators frequently sequence rotary selectors or ganged pushbuttons.
3. Sequential action, which opens one contact set in a specified order in relation to other contacts in the same control. This type may involve simultaneous operation of several contact sets and repeated opening and closing of one contact set during one actuation.

Contact Arrangements

Switching capacity refers to the number, type, and arrangement of contacts and circuits within the control. Figure 4.21d contains the related terminology:

Normally open contacts (NO). The circuit is open (current does not flow) when the contacts are not actuated. Moving them to the other extreme position completes (closes) the circuit.

Normally closed contacts (NC). The circuit is complete when the switch is not operated. Moving the contacts opens the circuits, interrupting current flow.

Number of poles (P). The number of conductors in which current will be simultaneously interrupted by one operation of the switch.

Number of throws (T). The number of positions at which the contacts will provide complete circuits. Since contacts usually have only two extreme positions, switches are either single-throw or double-throw.

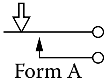
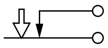
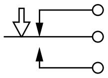
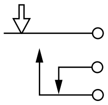
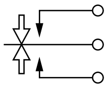
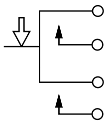
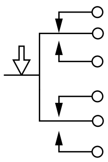
	Diagram	Operation	Features
SPST NO	 Form A	Make (pulse)	Pulses when actuator is depressed, but not when released.
SPST NC	 Form B	Break (pulse)	Same as Form A except opens circuit instead of closing it
SPDT NO/NC	 Form C	Break-make (transfer)	Provides complete circuits at both contact positions
SPDT NO/NC	 Form D	Make-before-break (Continuity transfer)	Momentary overlap of contacts provides circuit continuity
SPDT NO	 Form K	Center off (pulse)	Provides pulses in two conductors from one switch. Useful for "up, down" jogging.
DPST NO	 Form AA	Double make (Double pulse)	Interrupts current in two independent conductors of a circuit. Replaces two Form A switches.
DPDT NO/NC	 Form CC	Double break-double make (Double transfer)	Simultaneously operates two circuits. Several wiring possibilities. Replaces two Form C switches.

FIG. 4.21d
Switching contact configurations.

Switching Elements and Circuits

Switching circuits are classified as low energy or high energy. Low-energy circuits do not develop sufficient volt-amperes to break down insulating contaminant film (that may have built up on the contacts) by contact heating. Consequently, infrequent use degrades the controls in these circuits. Accordingly, reliable low-energy switching elements must maintain low contact resistance. Solutions include:

1. Wiping action under high contact pressures
2. Contact materials with high oxidation resistance
3. Hermetically sealed enclosures, either evacuated or filled with an inert atmosphere

High-energy circuits generate sufficient power to sustain an arc between contacts. Therefore, a reliable high-energy switching element must perform under very high contact temperatures. With these circuits, unlike low-energy circuits, frequent operation degrades reliability. Large contact areas, thick plating, and thermally conductive alloys combat high temperatures. Special features divert, vent, or extinguish the arc.

Switching elements are summarized in Figure 4.21e. Mechanical contactors either wipe the contacts over each other or strike them together. Wiping contacts move in a series of jerks, during which the contacts alternately weld and break loose. Striking contacts are either elastically or plastically deformed on impact.

Magnetic Reed Magnetic reed contacts are overlapping ferromagnetic beams, cantilevered from a glass tube filled with a dry inert gas. A small gap separates the overlapping

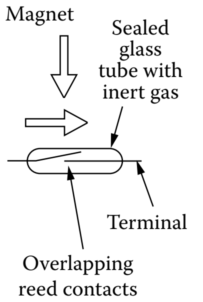
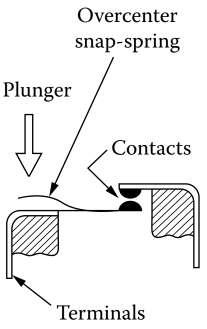
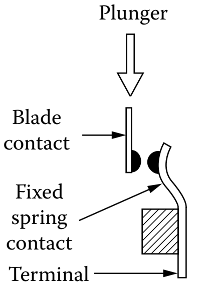
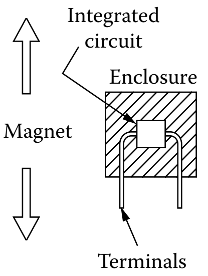
	Name	Diagram	Operation	Operating characteristics	Advantages and disadvantages	Applications
Mechanical contactors	Magnetic reed		Magnetic induction bends reeds. Striking contacts.	Actuation force independent of contact pressure. High contact pressure, fast precise timing. Excellent for low-energy circuits.	Hermetically sealed contacts. Bounce is problem, requires external compensating circuitry. Can be actuated by vibration. Requires shielding from external magnets. Long life, inexpensive.	Keyboard pushbutton, panel pushbutton, (integral).
	Snapaction		Leaf spring snapped. Over-center. Striking contacts.	Irreversible action no "teasing". Fast, precise, short, repeatable timing and travel. Best for simultaneous switching of many circuits with one actuator. High contact pressure, high actuation force.	Enclosed contacts. Suitable for low- or high-energy circuits. Relatively inexpensive. Very small sizes available.	Toggle, panel pushbutton (built-up), industrial pushbutton (medium and heavy duty).
	Wiping blade		Cantilever blade contact wipes past fixed spring contact.	Low contact pressure. Wiping cleans contacts, but wears away plating, timing and travel not precisely repeatable.	Hard silver plating used to resist abrasion, thus not generally suited to low energy use. Rotary is exception since infrequent use allows gold plating. Relatively inexpensive.	Integral panel pushbutton, rotary, industrial pushbutton (light duty).
Contactless	Integrated circuit		(Hall effect) Voltage is developed across edges of current—carrying conductor by magnetic induction.	Actuator force independent of switching action. Very fast switching, precise timing.	No contact bounce. Always-on bias voltage sometimes required. Excellent for repetitive, long-life use in low-energy circuits. Expensive, sealed.	Keyboard pushbutton.

FIG. 4.21e
Basic switching elements.

ends, and magnetic induction in the gap eventually overcomes their stiffness, bending them until they touch. The circuit is reopened by removal of the magnetic influence. Reed contacts provide:

1. High contact pressure with low actuator force, since the actuator does not directly move the contacts
2. Very fast response
3. Long life

Since the reeds are cantilevered springs, a major problem is contact bounce.

Snap-Action Snap-action switches are very rapidly struck together by an overcenter spring mechanism. They provide:

1. Irreversible switching, since the operator can do no more than start the switching action
2. Precise timing, including virtually simultaneous switching of many circuits with one actuator
3. Precise, consistently repeatable travel because there is only one moving part—the spring
4. High load capacity and long life, since arcing is limited by the very fast contact transfer time

Wiping Blade Wiping blade contacts employ a stiff, blade-shaped contactor that passes over spring-mounted stationary contacts under high pressure. Alternately, the moving blade is the spring, and the fixed contacts are stiff. The contacts are sometimes knurled to assist the wiping action. Hard silver plating alloys are required to resist wear, which generally restricts the use of wiping blade contacts to high-energy circuits.

Integrated Circuit Contact-less switches are useful in low-signal-level electronic circuits. The absence of contact bounce eliminates the need for external filtering circuits. Other advantages are virtually unlimited life, very fast switching, and freedom from the effects of contamination. Most contact-less switches employ semiconductors, which require an always-on bias voltage. This can be a disadvantage when compared with a mechanical switch.

Grades of Switching Devices

Table 4.21f summarizes the features of four types of manual control devices—pushbutton, toggle, rotary, and thumbwheel—all of which are available in several grades and types of construction:

Appliance-grade switches are not suitable for process control; they are low-cost devices designed for light, nonabusive environments in which precise timing and current control are not required, performance degradation is not critical during operating life, and exacting operator feedback is not necessary.

Commercial-grade switches are suitable for the control room. They provide consistent actuation and switching performance at specified reliability in average environments. Some degree of contact sealing is present, and corrosion-resistant contact materials are used. Actuation and switching mechanisms afford precise control of light-to-medium electrical loads.

Industrial-grade switches are specialized for local installation in harsh environments. Construction is rugged in order to accommodate abusive actuation and unusually heavy electrical loads. Actuators and contacts are sealed against liquid and solid contaminants, especially oil.

Military-grade switches are suitable for process control use but are usually overspecified in some areas and underspecified in others and carry a price premium.

Built-up construction uses separate housings for the actuator and the switching element. Circuit flexibility is the principal advantage, since switching modules can be ganged at will, although standard capacities are available. Disadvantages are unsealed interface between actuator and switch modules, back-of-panel volume, and price.

Integral construction combines the actuator and the switching element in a common housing. Advantages include satisfactory sealing and smaller overall dimensions. The major disadvantage is fixed switching capacity.

TYPES OF SWITCHING DEVICES

Pushbuttons

Pushbuttons provide the simplest, most naturally comfortable actuation motion. They alone accommodate extensive labeling without the use of separate nameplates. Disadvantages include a switching capacity that is lower than that found in other types because there are only two contact positions and lack of inherent indication of switching status (such as by handle position) other than lighting. Two methods of producing colored light are by transmission and by projection. Color selection has been discussed in Section 4.19.

Transmitted colors are produced by transmission of white light through colored button-lenses; they should be used whenever possible. Colors are intense, saturated, and uniformly distributed over the button; do not degrade from lamp heat; and can be read both from great distances and through wide viewing angles.

Projected colors are produced by projection of colored light on a white button-screen. The colored light is produced by transmission of white light through colored filters that slip over the lamps. Projected colors are weak, dilute, and easily overpowered by normal room ambient light; have restricted viewing distances and angles; and degrade as the colored filter is progressively destroyed by lamp heat.

TABLE 4.21f
Types and Characteristics of Electric Switches

	Pushbutton	*Toggle*	*Rotary*	*Thumbwheel*
Typical applications	Keyboard, motor, machine tool, instrument; best for pulsing; action dynamics match digital display	Computer data register; power on-off, test; "center off" type suitable for jog, adjustment about reference	Computer control panels, ammeter adjustment (potentiometer); action dynamics match fine analog adjustment	Encoded data entry to digital computing circuits; instrument settings; exact value setting
Advantages and disadvantages	Easiest, most comfortable operation; self-contained label flexibility	Least expensive 2,3 position selector; simple nonprecision operation, reflex action; best for actuation of multiple switches simultaneously	Most capacity in one control; selector available for 4+ positions in high-energy circuits; self-contained encoding optional; also mounting for components	Most capacity in least panel area; self-contained encoding, dial readout of value setting; single alphanumeric character-position or complete messages
Handles and feedback features	Integral handle, lighting very useful; no feedback from handle position	Integral, special "decorator" style available; handle position positive feedback (indicator) or status; no light or label except rocker (marginal)	Separate, many shapes and color options; label on skirts; no lighting	Integral; small tabs adjacent to indicator dial; same number as number of positions; lighting optional
Environmental protection	Sealed switch elements standard; sealed actuators optional; standard industrial grade	Enclosed switch elements standard; sealed actuator handle optional	Enclosed switch elements optional; sealed actuator shaft optional	Enclosed switch elements and actuator handle optional
Number of positions	1,2	2,3	2 to 24 (potentiometer)	8 to 16
Comparable cost, position	2.5	1.0	5.0	—
Ratings	0.5 to 60A @ 125 V AC or 30 V DC; 1A @ 600 V AC; 0.5A @ 600 V DC	20A @ 125 V AC 12A @ 30 V DC	0.3A @ 125 V AC 1A @ 30 V DC	0.13A @ 125 V AC or @125 V DC
Life	10 million operations	1 million operations	100,000 360° cycles	100,000 360° cycles
Mounting	Front or back of panel; printed wiring board	Back of panel; printed wiring board	Back of panel; printed wiring board	Front or back of panel; printed wiring board

Panel Pushbuttons A variety of integral and built-up configurations is available (Figure 4.21g), incorporating molded or metal housings, various force and travel options, and a wide range of lighted and unlighted buttons. Square and rectangular buttons are most easily mounted and provide the most labeling area. Labeling is accomplished by hot-stamping, engraving, two-shot molding, silk-screening, or by use of photographic film inserts. Panel pushbuttons are usually mounted in individual openings; built-up configurations mount from the front of the panel, using threaded bushings and lock nuts. Buttons and lamps in each configuration are accessible from the front of panel.

Each mounting method has definite panel thickness limits, which can, however, be circumvented by subassembly to a bracket that in turn mounts on the back of the panel. The buttons are available for both low-energy and high-energy circuits.

Industrial Pushbuttons These buttons are ruggedly made to withstand abusive environments and to resist abusive operation (Figure 4.21h). Compact and standard sizes are available, and both kinds use a built-up construction to accommodate custom switching for specific applications. Actuators and contacts are both liquid tight and dust tight, and contacts are available for three grades of duty:

1. Electronic (wiping action, gold plating)
2. Standard (wiping action, silver plating)
3. Heavy-duty (striking action, silver, silver alloy, or cadmium alloy plating)

Keyboard Pushbuttons These buttons are integral pushbuttons used in low-energy electronic circuits (Figure 4.21i). Force-travel characteristics are carefully designed for precise tactile feedback during repetitive, high-speed touch typing. Simplified, durable, low-cost actuator mechanisms and switching elements furnish a high degree of reliability for many operations.

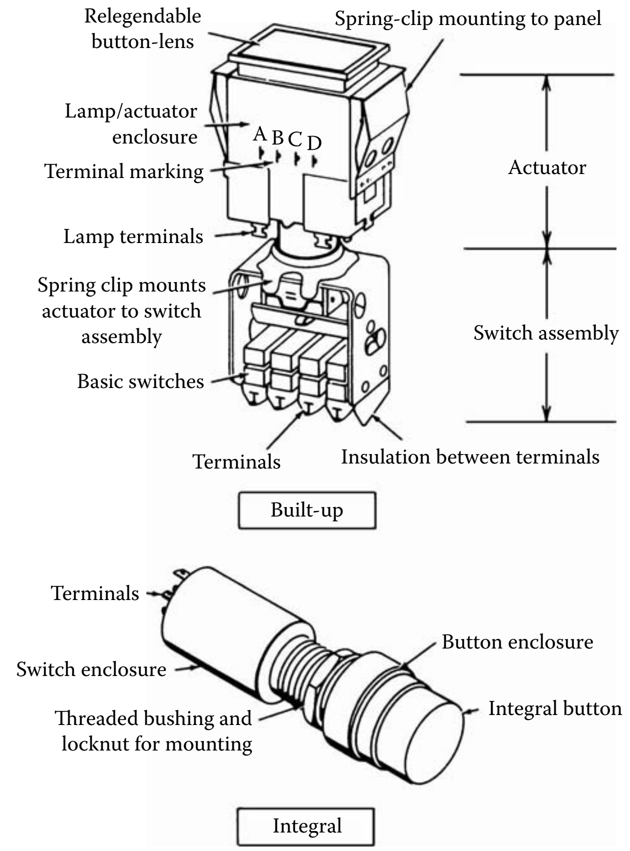


FIG. 4.21g
Panel pushbuttons.

Construction is as functional as possible for mass production, group mounting, and group connecting. Techniques include molded housings, two-shot molded buttons and labels, and printed wiring board mounting and connecting. Both assembled keyboards and a variety of standard and custom arrangements, encoding, operating forces, and buttons are available from most manufacturers of keyboard pushbuttons.

The inscriptions and function buttons of the keyboards are usually highly customized and reflect the purpose that the workstation serves (Figure 4.21j). This customized, process-specific design of keyboards makes the operator's work more efficient.

Hall-Effect Pushbuttons The early keyboard pushbutton designs used reed switches. In modern keyboards either the traditional Hall-effect plunger switches are applied or more modern membrane-type switches using either contact or capacitance-type sensors are employed. Hall-effect devices are snap-acting, solid-state sensors that consist of a generator and a trigger circuit integrated on a single silicon chip.

The Hall-effect chip (Figure 4.21k) detects both the strength and direction of the magnetic field and therefore generates a signal as the magnet attached to the plunger assembly passes the chip. The integrated circuit produces an analog control voltage, usually 0.4 V DC maximum at 4 milliamperes per output. The power supply is 5 V DC, and the rise or fall time of the signal is about 1 microsecond. The purpose of the trigger circuit is to convert this voltage into a digital output. Hall-effect pushbuttons are available with LED-type lighting, and one of their advantages is that they eliminate the problem of contact bounce.

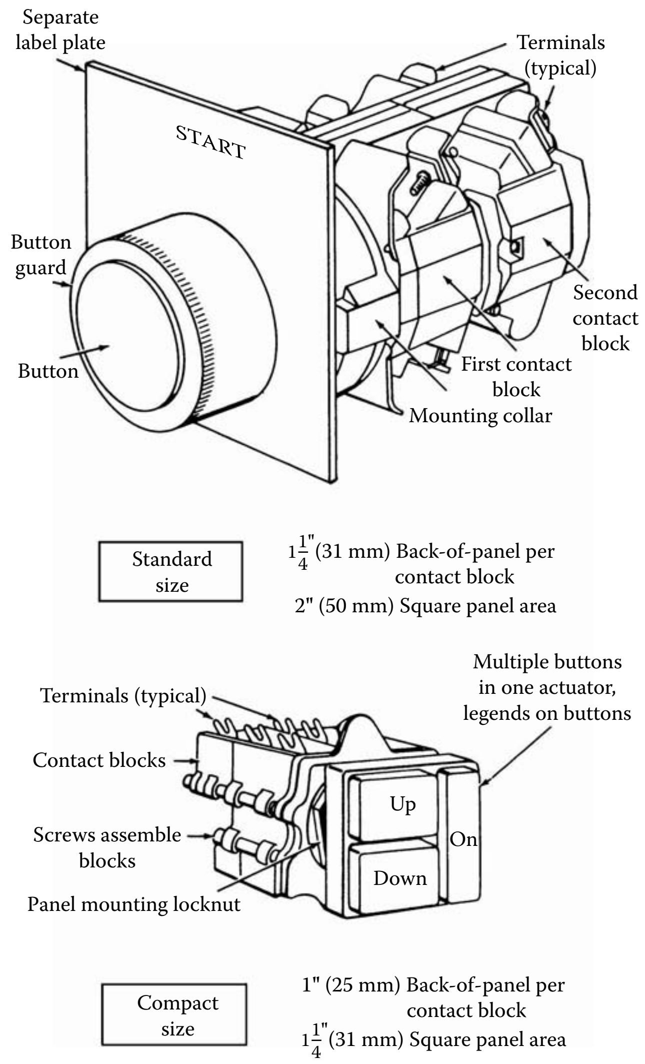
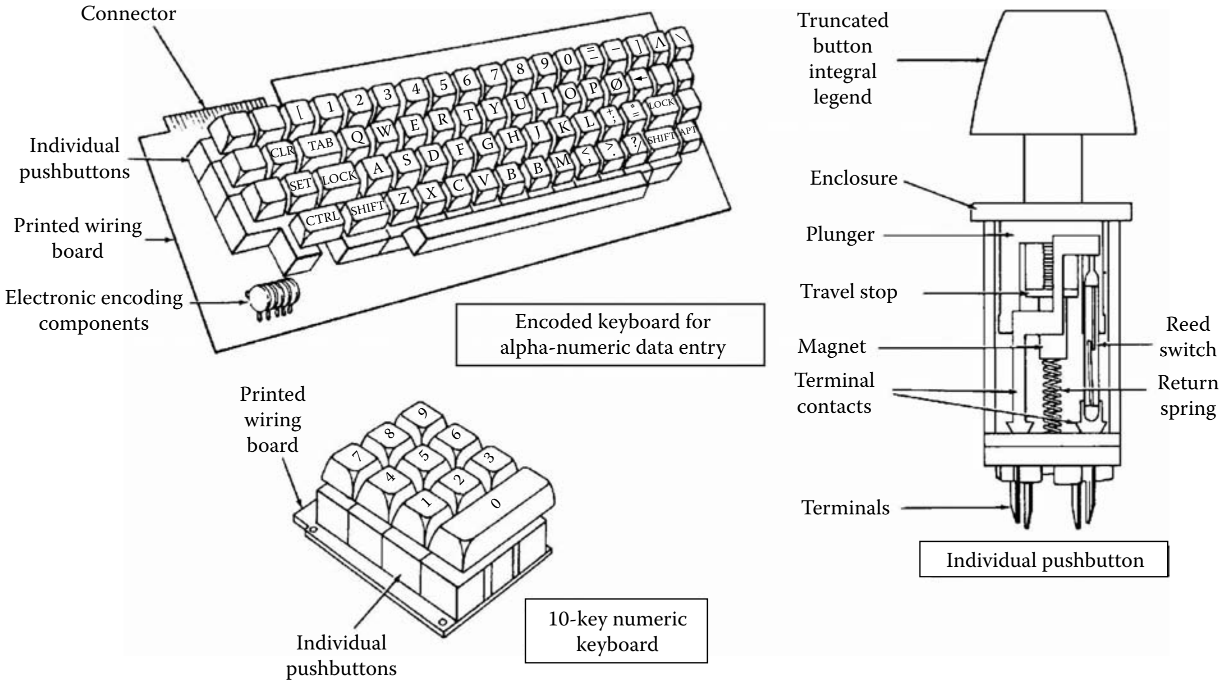


FIG. 4.21h
Industrial pushbuttons.

Membrane Pushbuttons Touch panels or membrane switches are inherently rugged and immune to dirty environments (Figure 4.21l). They are manufactured in a sandwich-like configuration with the circuit layers separated by spacers with holes for the actuation points. When force is applied at the actuation points, the membranes are flexed. In the contact design, a SPST-NO (Single-pole single-throw, normally open) circuit is closed, while in the capacitance design the drive pad approaches the sense pad, resulting in an increase in the variable capacitance, which triggers the detector circuit. The fixed capacitance between the float and sense pads is used as a reference. When the finger pressure is removed, a snap spring reopens the circuit and a compression spring self-returns the key top.



Encoded keyboard for alpha-numeric data entry

10-key numeric keyboard

Individual pushbutton

FIG. 4.21i
Keyboard pushbuttons.

The operating life of such keys is between 2 and 5 million operations, the key size is usually 0.5 by 0.5 in. (13 by 13 mm), and the key travel is 0.01 to 0.02 in. (0.25 to 0.5 mm).

The membrane switches are allowed to “breathe” through venting openings, so that the development of pressure differences across the membranes, due to the pressing of a key or because of altitude or temperature changes, will not affect performance.

The keys of various keyboards are wired in a matrix format. When a key is depressed, one sense circuit and one drive circuit are energized, and the unused circuits are automatically grounded, shielding the keyboard from electrical noise, which could generate erroneous code.

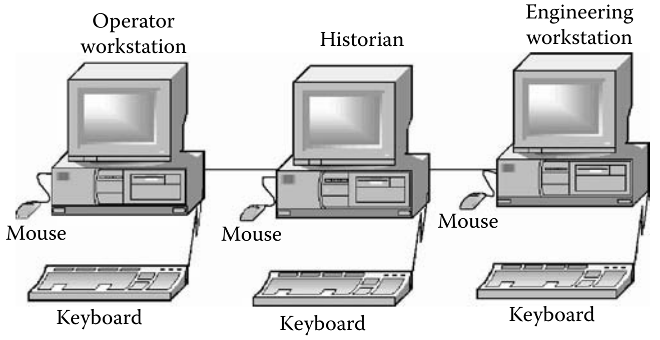


FIG. 4.21j
Custom keyboards are designed to fit the nature and purpose of the workstation.

Toggle Switches

Toggle switches are the cheapest two-position or three-position selectors for a large number of circuits of any level of complexity or energy, and integral and built-up configurations are available. A major advantage is positive indication

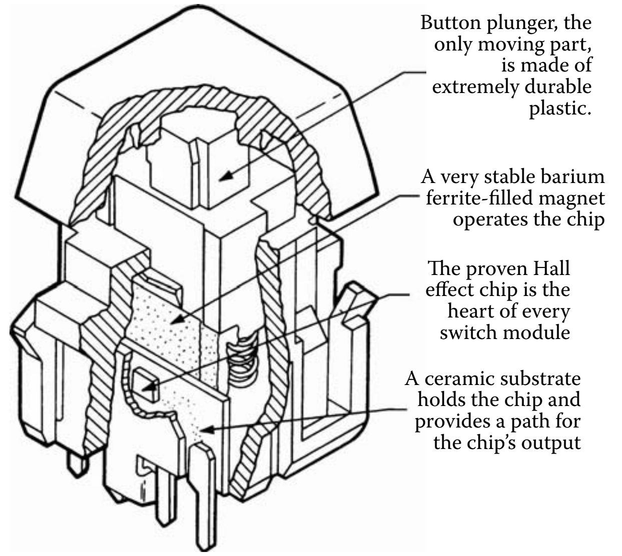


FIG. 4.21k
Solid state Hall-effect pushbutton design. (Courtesy of Pepperl + Fuchs Inc.)

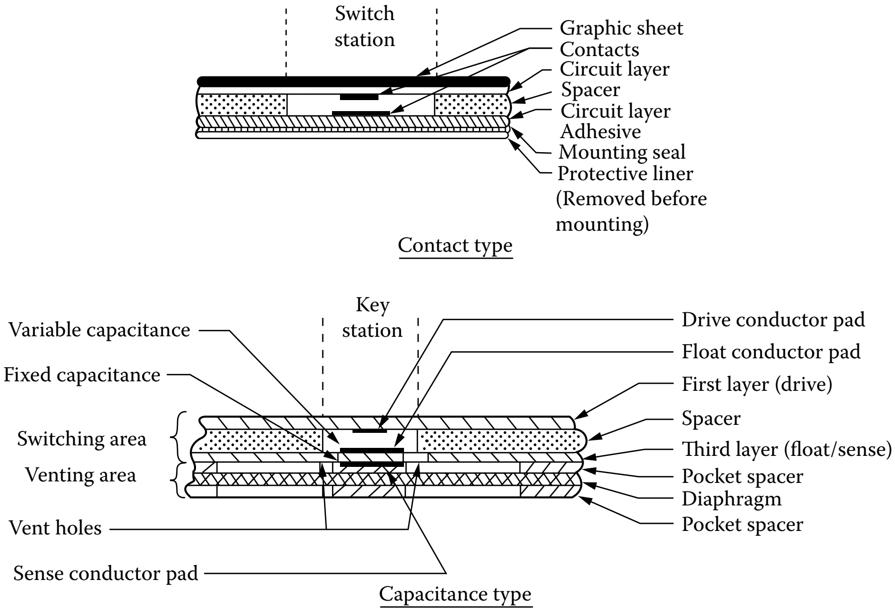


FIG. 4.21l
Membrane-type pushbutton designs. (Courtesy of Pepperl + Fuchs Inc.)

of switching status from handle position. The three basic handle configurations are lever, rocker, and thumbwheel.

The lever handle configuration (Figure 4.21m) is the most common and gives the most positive handle position indication. Careful consideration of length, shape, protrusion above the panel, and panel graphics design is required before the operator can capitalize on this advantage. A wide selection of shapes, sizes, colors, and trim hardware is available to create various front-of-panel appearances, accentuate handle positions, and safeguard against accidental actuation.

The rocker handle (see Figure 4.21m) combines positive handle indication with the operating simplicity of pushbuttons. Some space for labeling and lighting are also available, but both features are marginal owing to space limitations.

The thumbwheel handle (see Figure 4.21m) is occasionally used to reduce front-of-panel protrusion and to accommodate special actuation motion. The disadvantages of the thumbwheel handle are that handle position does not provide a positive indication of switch position and that control of travel and force is difficult since the finger moves in and out at the same time that it moves from side to side. The latter characteristic makes the handle especially awkward for use as an adjustment control about a spring-loaded "center off" position.

Rotary Switches

Rotary controls (Figure 4.21n) are selectors or adjustors for low-energy circuits and afford greater switching capacity in less volume than any other control does. Switching capacity is expanded by addition of switching modules, and a major advantage is that entire operational sequences can be contained on one control. The disadvantages of rotary controls are that operating force is increased with increased switching capacity and that there is no direct access to individual positions without actuation of intermediate positions.

Built-up construction is almost universal, either enclosed or open-frame. Mounting is either to printed wiring boards or back-of-panel, using standoffs or threaded bushings and lock-nuts. Switch modules consist of fixed contacts mounted on round printed wiring board decks. Each deck has a movable contactor attached to a common shaft. The printed wiring affords flexibility in contact arrangement, contact spacing, total rotational travel, switching action, and mounting of electronic components for self-contained encoding.

Users commonly abuse the inherent flexibility of these controls by packing too much switching capacity into one control, based on the rationale of saving the cost of another rotary switch. Angular spacing of contacts becomes so small, and operating force so high, that actuation is awkward and imprecise.

Very high contact pressures result and can mechanically damage contacts and the actuator mechanism. Considerable panel area is required to label a large number of positions. Large-diameter handles, although they decrease force, also consume panel area. The other extreme of excessively wide contact spacing causes awkward wrist and arm movements.

The guidelines generally accepted include:

1. Minimum angular spacing = 15°
2. Maximum angular spacing = 90°
3. Minimum positions = 4 in 360°
4. Better minimum = 4 in 240°

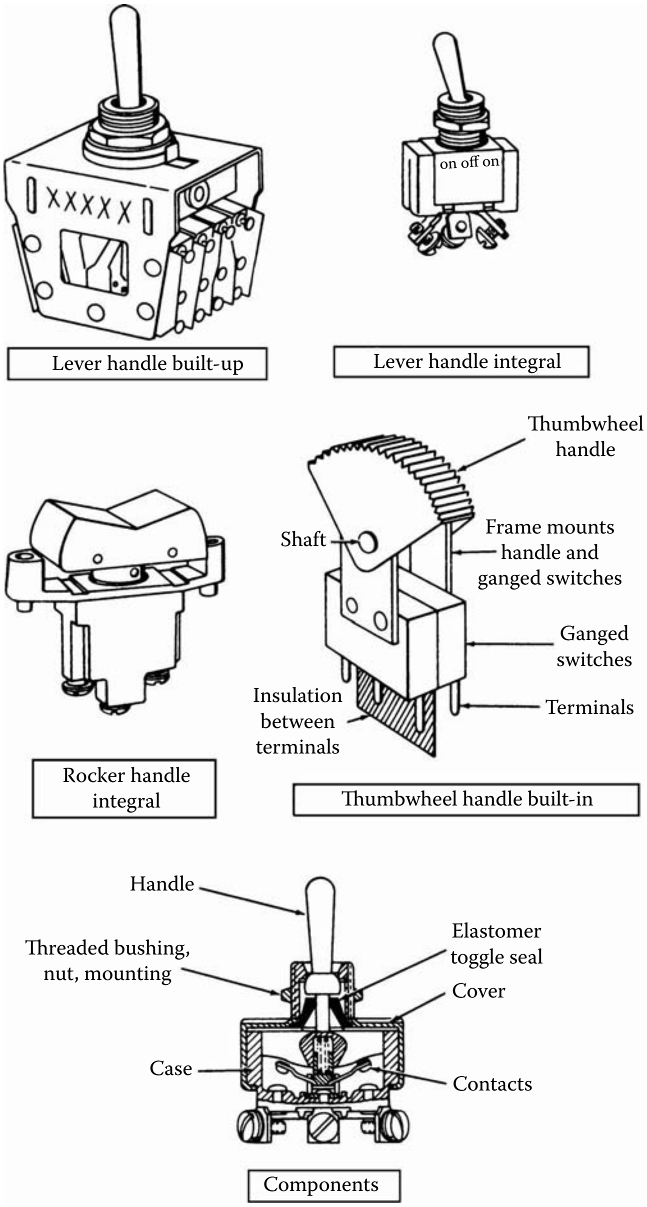


FIG. 4.21m
Toggle switches.

5. Maximum positions = 24 in 360°
6. Better maximum = 20 in 300°

Thumbwheel Switches

Thumbwheel switches (Figure 4.21o) are a form of rotary selector for encoded data entry to low-energy electronic circuits. Encoding is accomplished by internally mounted electronic components. Thumbwheels take up very little front-of-panel area owing to the parallel arrangement of contact disk, moving contactor and handle and self-contained indicator dials. Contact disks can be stacked horizontally to create controls of any size. Thus, thumbwheel switches can do the job of several rotaries, pushbuttons, or toggles and use considerably less panel space.

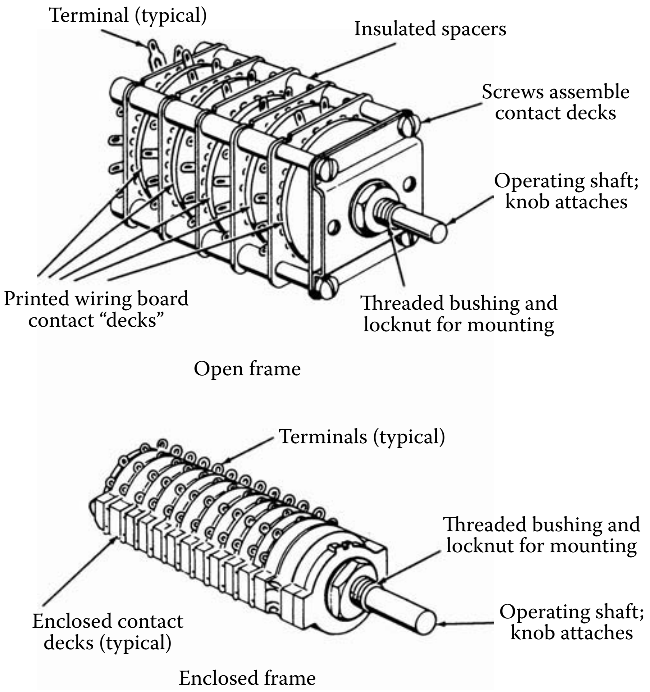


FIG. 4.21n
Rotary selectors.

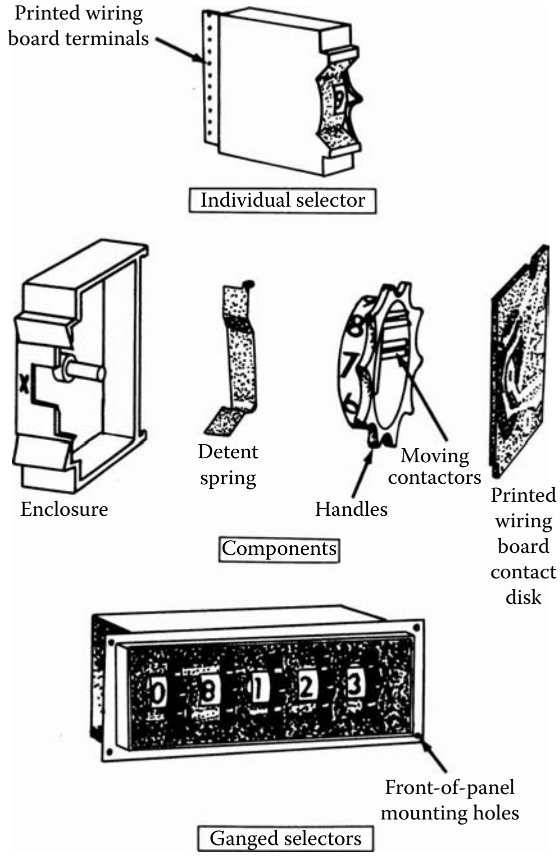


FIG. 4.21o
Thumbwheel selector.

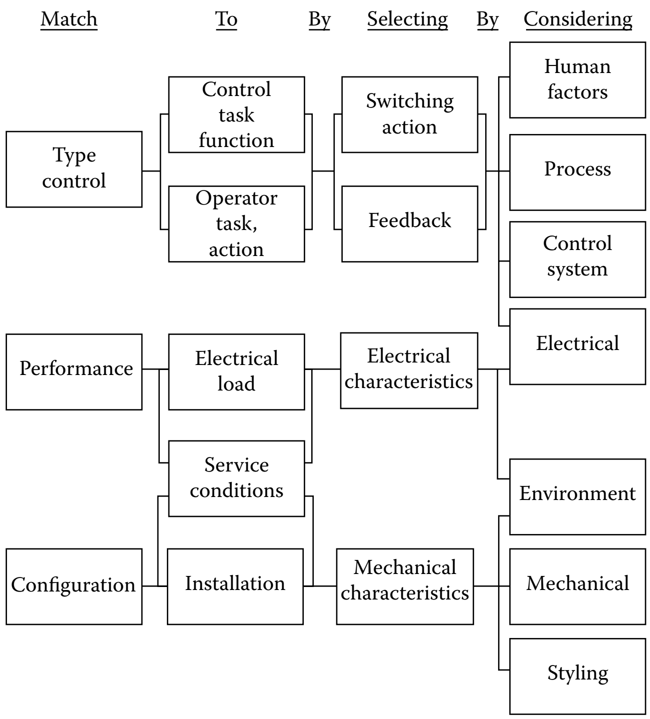


FIG. 4.21p
Selection procedure.

A major disadvantage of their use is that considerable concentration and dexterity are required for efficient operations. Also:

1. Actuator motion is usually opposite dial motion.
2. Handles are very small and must be poked at with fingertips.
3. Incremental spacing on the dial is small.
4. Horizontal spacing between handles is small.

APPLICATION AND SELECTION

The goal of the switch selection process (Figure 4.21p) is to provide the most direct link between operator and process by considering the following factors:

1. What is being controlled—the process
2. How it is being controlled—the type of control system
3. How the control is used—human factors
4. The service conditions—electrical, mechanical, and environmental

Process response to control action determines the operator's subsequent action. Important variables include time response, amplitude response, linearity, and damping.

Control system translation separates the process from the operator's control action and the operator from the process response. Important factors affecting selection of manual control involve:

1. Type of control—analog, digital
2. Resolution—proportional, derivative, integral, and combinations
3. Accuracy—physical losses through transmitters
4. Data processing errors—chop-off and round-off
5. Visual feedback—display resolution, speed, accuracy
6. Safety provisions

Human Factors

Basic to the human factors is mind–eye–hand coordination. Figure 4.21q depicts the role of this neural aptitude in manipulating controls. Its efficiency depends on how well the controls are matched to:

1. The control task and actions required to perform it
2. The workstation design
3. The displays associated with the controls

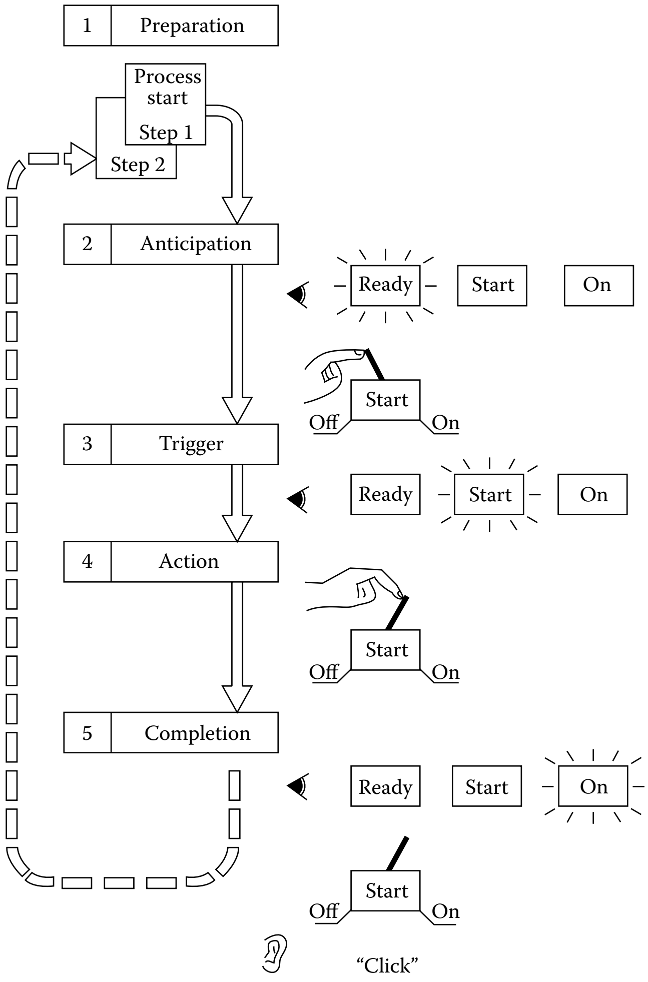


FIG. 4.21q
Mind–eye–hand coordination in control manipulation.

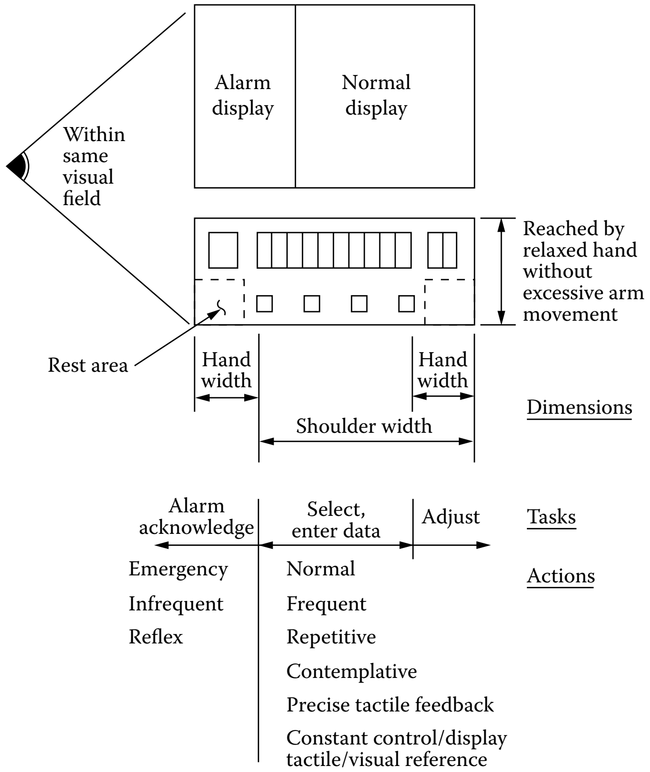


FIG. 4.21r
Workstation layout considerations.

Tasks (Figure 4.21r) are classified according to:

1. Selection of alternates: start, stop, automatic, and manual
2. Adjustment: gross, fine, fast, and slow
3. Tracking: adjustment about a moving reference
4. Data entry: setting values, loop address, and loading computer registers

Actions (see Figure 4.21r) required to perform these tasks may be either:

1. Sequential or simultaneous
2. Continuous or intermittent
3. Exact or approximate
4. Repetitive or different each time
5. Frequent or infrequent
6. Single-pulse or hold-down
7. Normal or emergency

The workstation design objective is to match the physical characteristics of controls and displays to the nature of the control task, the actions required to perform it, and the operator's capabilities.

Anthropometry and general arrangement are different depending on the operator's position (sitting, standing, stationary, or mobile). Figures 4.21s and 4.21t show proper physical relationships. The angles formed by hand, fingers, and controls affect the "feel" of the control.

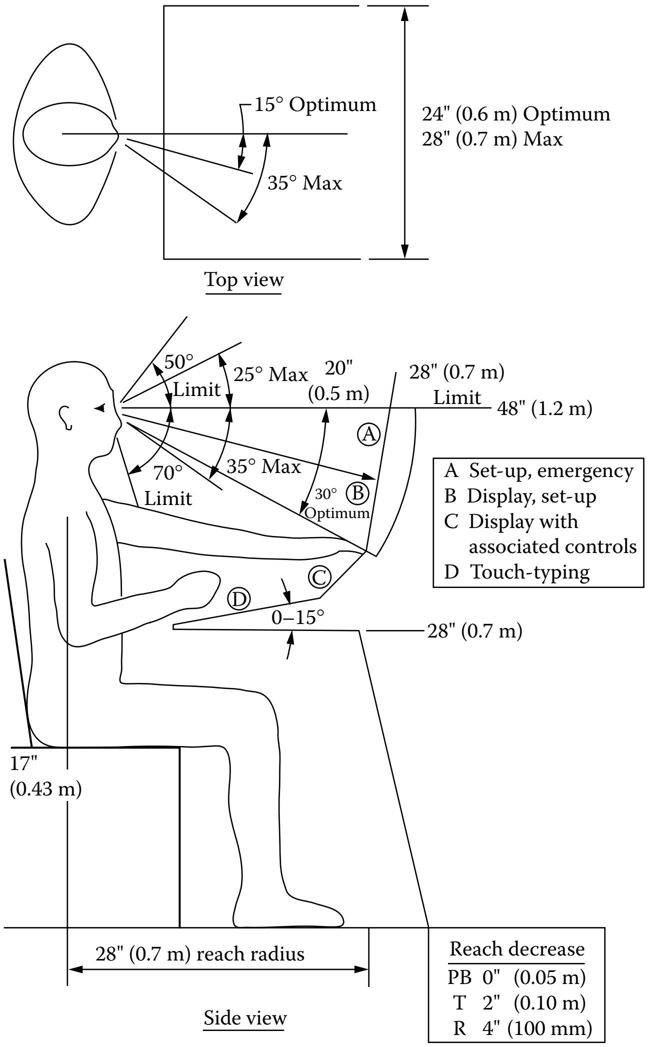


FIG. 4.21s
Hand and eye coordination considerations for an operator who is working in the sitting position.

The relaxed hand assumes natural angles that determine the most comfortable and strongest line of force (Figure 4.21u). Controls should be oriented and arranged so as to achieve these angles over the entire control panel. Compromises are necessary to reach extremities of the panel without inordinate body movement. Detailed arrangement should include the following considerations:

1. The same type of control should be placed differently for sitting and standing positions.
2. A sequence of time or order of use should be clearly depicted. The sequence should be maintained regardless of operator or panel orientations.
3. Workload should be distributed properly between left and right hands with consideration given to the fact that most people are nimbler with their right hand than with their left hand.

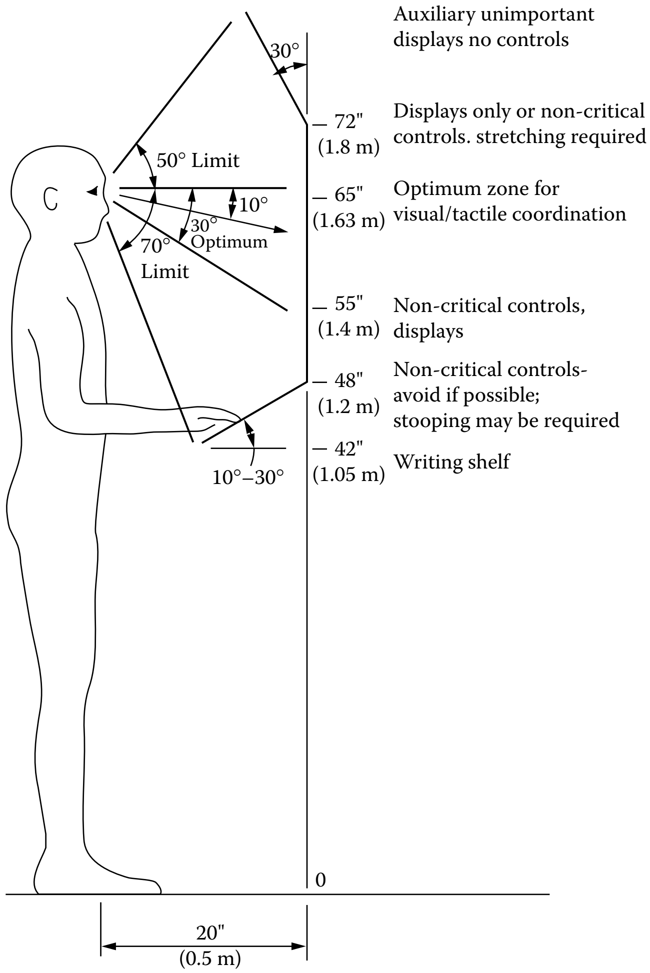


FIG. 4.21t

Hand and eye coordination considerations for an operator working in the standing position.

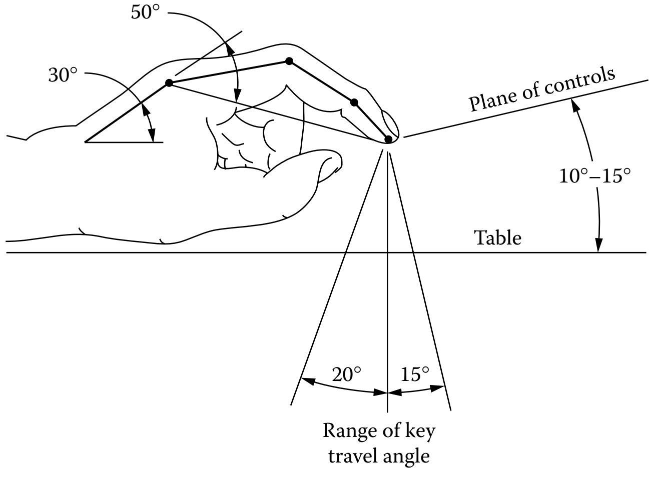


FIG. 4.21u

Natural hand and finger angles.

4. Related controls and displays should be clearly associated and located within the same visual field to minimize head and eye movement. Actuation of controls should not obscure related displays, labeling, or lighting.
5. Controls should be located below or to the left of the display for left-hand operation and below or to the right for right-hand operation.
6. Similar controls performing similar functions should operate in similar directions; similar controls performing dissimilar functions should operate in distinctly different directions.
7. Critical controls may require safeguards against accidental actuation by clothing, general body contact, and falling objects.

Display Movement

Control and display movement should be coordinated for direction, rate, accuracy, resolution, ratio, and linearity. Certain natural relationships are required by human habit patterns and reflexes (Figure 4.21v). Small, precise display adjustments require long-travel, low-force control motion. Short

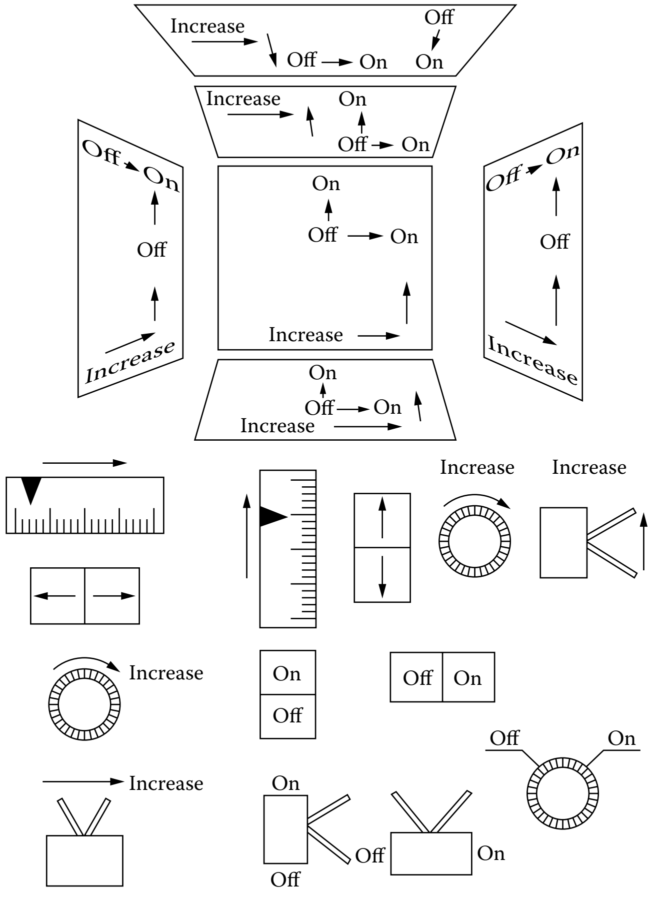
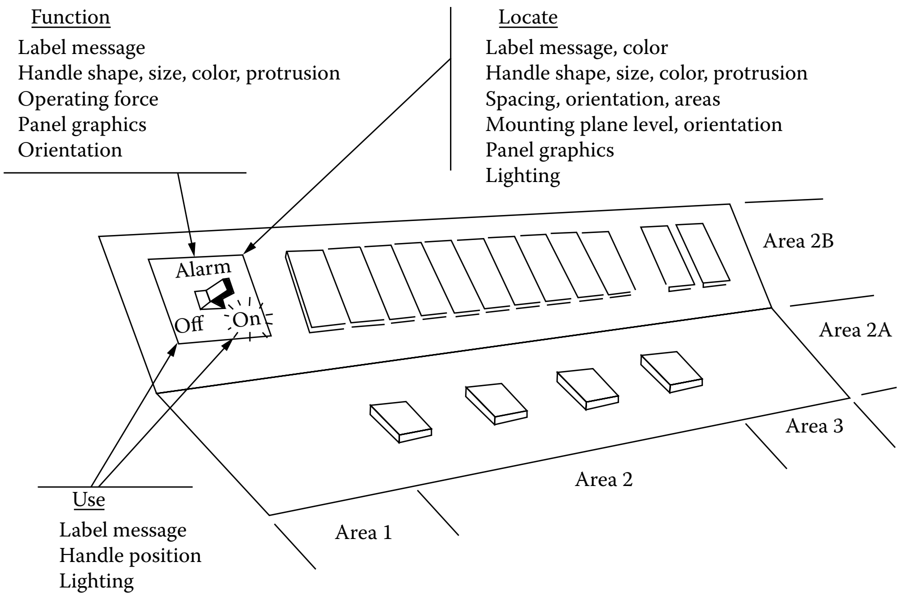


FIG. 4.21v

Natural relationships between the directions of motion.



Labeling guidelines

Too much information is as useless as too little
Be specific, concise, clear, but not elaborate
Avoid "engineering language," identify what is being controlled not the name of the control
Labels should be on controls whenever possible. When on panel, association with identified control should be obvious.
Adjacent labels should not create additional. "false" label
Place labels consistently above or below identified control
Orient labels horizontal to reading line-of-sight, avoid vertical, slanted, or curved orientations
Use capital letters, correlate size with viewing distance
Colors of labels should provide maximum contrast with label background

FIG. 4.21w
Guidelines for labeling and identification.

travel and low force should be used for fast displays. Pulse controls are well suited to incremental displays.

Analog displays are tracked or adjusted most efficiently by a rotary potentiometer or by a control with maintained "on" positions on either side of a spring-return center "off." Efficient identification of the control, its function, and its use is a keystone of efficient operation. Considerable care is required to achieve this seemingly commonplace goal. It is important to recognize that most identification techniques involve double duty and that redundancy of controls can both confuse the operator and lend a "busy" appearance to the panel (Figure 4.21w).

Control design should blend with the overall styling theme of the workstation. Handles and mounting hardware are available in a variety of shapes, sizes, textures, materials, labeling, and lighting (Figure 4.21x).

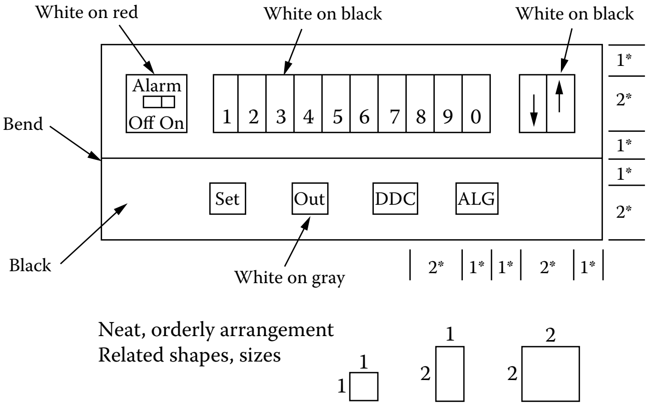
Error Prevention

Error prevention techniques are valuable in order to forestall simultaneous and missed actuations. Both activities are especially prevalent in repetitive, high-speed keyboard operation and can be traced to speed limits imposed by switching mechanisms, actuator mechanisms, and the scanning speed of external monitoring equipment.

Mechanical interlock inhibits two types of simultaneous actuations so that:

1. Two keys cannot be simultaneously depressed to the switching point.
2. A depressed key must be released before another can be depressed.

Electronic interlocks artificially increase the actuation speeds by delaying scanning of consecutive key outputs. "Two-key roll over" generates codes during depression of one key and release of another by blocking the second until



Modular spacing*
Related colors-black, gray
Key grouping by spacing, color, shape
Labels concise, unambiguous, consistent
Label colors uniform, high contrast
Special color coding for habit-pattern
reflex-red = alarm

FIG. 4.21x
Guidelines for label styling.

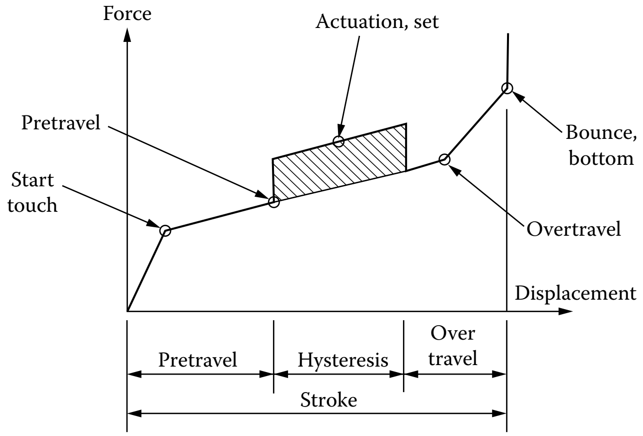


FIG. 4.21y
Best tactile feedback, the best force–travel relationship for the rhythm of a skilled operator.

the first is released. "One-character memory" stores a character code until another is generated, erasing the first.

Tactile feedback in the form of a properly designed force-travel relationship is the best guarantee against missed actuations. Skilled operators develop a rhythm of depressing keys. They keep several steps ahead of their actual manipulation, counting on a change in their rhythm as a signal to stop. This change is usually the absence of a completion cue for an intended actuation. Completion is signified by a sudden increase in operating force near or at the end of travel. Supplementary audible or visual cues also occur at this point. The force-travel relationship shown in Figure 4.21y is generally regarded as desirable, is available in solenoid-assisted electric keyboards, and is closely approximated by some manual keyboard controls.

However, the curve shown in Figure 4.21z is the most economical (a simple linear spring) and therefore the most available. Tests have shown that poor force–travel relationship is a primary source of missed keystrokes in high-speed data entry.

Environmental variables affecting manipulation of controls include temperature, relative humidity, ambient illumination, and noise. Sweating may limit the amount of force the operator can apply, degrade dexterity, and cause fingers to slip off handles. Excessive light can mask lighted buttons, causing errors and loss of time and reducing safety. Noise detracts from concentration and may preclude audible feedback.

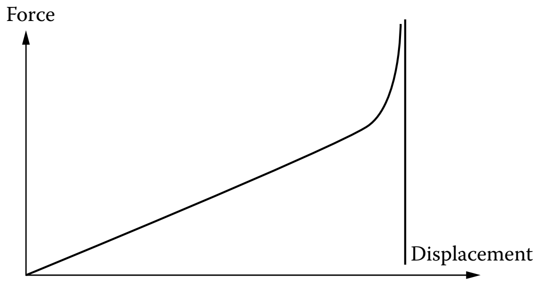


FIG. 4.21z
Force–travel relationships with linear spring feedback.

Electrical Rating and Performance Electrical performance of the control is determined by current and voltage ratings, contact bounce, switching sequence, and electrical interference from other equipment. Current rating measures resistance to welding of contacts. Underrated contacts either fail to open or do not open the required number of times during a specified interval. Current ratings are determined by contact configuration, plating materials, and transfer time.

Selection should account for potential overloads, which can exceed normal currents by as much as 30 times. Common sources of overload are:

1. Inrush starting currents of motors and tungsten-filament lamps
2. Relays and solenoid
3. Equipment malfunctions, such as short circuits and unstable power suppliers
4. Lightning

Voltage rating reflects resistance to arcing between contacts or terminals due to voltage surges. Ratings are determined by air gap spacing between contacts and between terminals and the insulation resistance of their mounting materials. Terminal spacing can cause a problem on subminiature switches, requiring supplementary insulating strips between terminals.

Low insulation resistance can cause voltage leakage to ground or excessive dissipation within the dielectric itself. High-impedance analog computing circuits are especially sensitive to these phenomena. Certain insulating materials support formation of conductive surface films by repeated arcing, leading to eventual flashover between contacts or terminals.

Alternating current resistive loads are easily interrupted because the cyclical current reversals prevent arcing caused by continuous current buildup across the contacts. However, high AC frequencies (400 Hz) can approach DC in this respect. Direct current ratings are usually lower than the ratings for AC service, and DC arcs tend to sustain themselves since the applied voltage does not reverse as it does with AC. Special arc suppression techniques are sometimes required and are incorporated either within the switch using surge cavities, vents, or magnets whose field opposes that generated by the arcing current or within external circuitry using resistors, capacitors, and rectifiers to reduce or surmount the arcing current.

Contact bounce occurs partially in all mechanical contactors. Reed contacts are especially vulnerable owing to their cantilevered spring construction. Snap-action contacts have little bounce since the over-center spring provides damping. Contactless keyboard switches, of course, are free from bounce. Bounce is equivalent to a sequence of contact openings and closings, the consequences of which are:

1. Reduced contact life
2. Generation of extraneous signals, which requires external filter circuitry
3. Radio frequency interference

The switching sequence should be chosen to minimize the circuit complexities that can occur during contact transfer. Make-before-break sequences furnish momentary contact overlaps—useful maneuvers when attempting simultaneous switching of several fast, low-inductance circuits with a multi-pole switch.

Mechanical Features

Mechanical features of electrical switches and of parent equipment that affect workstation design include:

1. Type of mounting (front-of-panel, back-of-panel, printed wiring board, individual, in groups, and vibration isolation)
2. Front-of-panel dimensions (spacing and size of openings)
3. Back-of-panel dimensions (spacing and access)
4. Access requirements (repair, replacement, maintenance, and rearrangement of lamps, handles, and wiring)
5. Wiring (terminal configuration and location)

Special precautions are occasionally required to preclude undesirable interaction between controls and parent equipment. Communications equipment operating at radio frequencies requires shielding from electrical noise generated by contact bounce. Reed switches must be shielded from heavy magnetic ferric materials (steel) and from magnetic fields stronger than their actuating magnet.

Environmental Considerations

Environmental factors pertinent to process control application include:

1. Temperature
2. Relative humidity
3. Contamination from gases, liquids, and solids
4. Barometric pressure
5. Vibration and shock

Most commercial and industrial electrical switches function within specification in environments that the operator can also withstand.

However, infrequently attended controls may be exposed to much more severe operating conditions. Electrical codes sometimes require low-energy circuits in hazardous locations, a circumstance that may necessitate intermediary isolation circuitry between control and controlled equipment, or sufficient air gap spacing to ensure potential differences below spark-generation levels.

The contact temperature is the sum of the room temperature, the rise within the parent equipment, and the rise within the switch. High temperatures accelerate contact corrosion. Unequal expansion of components can cause cracking and binding. Forced cooling may be required for densely spaced lighted pushbuttons that operate continuously.

Relative humidity and temperature cycling produce condensation. Moisture accelerates contact corrosion (decreasing current rating) and decreases insulation resistance (decreasing voltage rating). Contamination from moisture, salts, oils, corrosive gases, and solid matter degrades contacts and actuator mechanisms.

Contacts are protected by glass-to-metal or plastic seals or by inert atmospheres. Sealing can be done at the control panel, between the actuator and the contact mechanism, or within the contact enclosure. Controls should remain in their original packing materials until just before installation. Special materials, such as low-sulfur-content papers, retard contact corrosion during shipping and storage.

Vibration and shock can cause contact chatter, arcing, and outright structural failure of the contact mechanism and can affect the choice of operating force, handle shape, weight, and mounting method. Vibration damping is provided by snap-in spring mounting and the parent equipment structure. Orientation on the control panel should place handle travel at right angles to the vibration force. Barometric pressure affects the ability of air to extinguish arcs and dissipate heat. Current and voltage ratings must be reduced for use at high altitudes; hermetically sealed enclosures are occasionally needed for such atmospheric applications.

Bibliography

Bylander, F. G., *Electronic Displays,* New York: McGraw-Hill Book Co., 1979.

Degani, A., Shafto, M., and Kirlik, A., "Modes in Human–Machine Systems: Review, Classification, and Application," Human Factors Research and Technology Web page, NASA, http://human-factors.arc.nasa.gov, 2001.

Gadberry, B. E., "Designing Integrated Control System Displays," Paper #91-0354, 1991 ISA Conference, Anaheim, California, October 1991.

Keller, P. A., *Electronic Display Measurement Concepts, Techniques, and Instrumentation,* New York: John Wiley & Sons, 1997.

Kwok, T., "Advanced Workstations as DCS Operator Consoles," 1992 ISA Conference, Houston, October 1992.

MacDonald, L. M., and Lowe, A. C., *Display Systems: Design and Applications,* New York: John Wiley & Sons, 1997.

McCready, A. K., "Man–Machine Interfacing for the Process Industries," *InTech,* March 1982.

National Electrical Code, Articles 380 and 430, National Fire Protection Association, 1981.

Shirley, R. S., et al., "What's Needed at the Man/Machine Interface?" *InTech,* March 1981.

Stephanidis, C., and Sfyrakis, M., "Current Trends in Man–Machine Interfaces: Potential Impact on People with Special Needs," In Patric, R., and Roe, W. (Eds.), *Telecommunications for All,* Brussels: ECSC-EC-EAEC, 1995.

4.22 Touch-Screen Displays

BABAK KAMALI (2005)

Sizes:	From 12.1 in. to 19 in.
Costs:	Up to 15-in. monitors are under $1000 (CRT or LCD). Capacitive and resistive overlays are the cheapest ones. CRT displays are more expensive than LCD displays. 15-in. resistive or capacitive LCD is $400 to $500, CRT is $750 to $800. 15-in. surface acoustic-type LCD is $900 to $1000. Serial connectors are cheaper than USB connectors.
Partial List of Suppliers:	Elo TouchSystems Inc. (www.elotouch.com) 3M United States (www.3m.com) Touch International Inc. (www.touch-international.com) GVision Inc. (www.gvision-usa.com) Posiflex Inc. (www.posiflexusa.com)

INTRODUCTION

Touch screens are widely used in industrial applications that require accuracy, touch sensitivity, and durability. Today's computers come in many different shapes and sizes. Displays such as touch screens and single-board computers have made it possible to mount computers in locations as diverse as gas pumps and shopping carts. Touch-input devices fit over any size display and occupy little additional space, making them perfectly suited to the current trends in computers.

It is not only computers that have evolved over time, but the perception of them by the general public as well. These are the reasons that made touch technology so popular. In addition, the use of graphical icons and images makes the touch applications easy to understand, requiring little learning and no complicated instructions. There simply is no easier way to access a computer than by touching its screen. Pointing at an object is natural and intuitive for people all over the world because pointing transcends the barriers of language and culture.

TOUCH TECHNOLOGY

Advantages

Touch technology is especially beneficial if *space is critical.* If the available space is too limited for input devices such as a keyboard or mouse, touch technology is an excellent alternative. The traditional detached input devices are not suitable for use with handheld computers or personal digital assistants (PDAs).

In certain applications, such as during driving, *time is critical* because it is dangerous to divert one's eyes from the road. Touch screens eliminate the time needed for the operator to switch attention back and forth between a display and an input device, and the graphical target displays tend to minimize or eliminate the need for the operator to read any instructions.

Touch technology is also *accurate.* For example, an air traffic controller or a chemical plant operator can touch a radar image on a display rather than typing data to receive additional information on a particular aircraft or unit process.

Another advantage is that touch technology can provide a *simple interface* to an otherwise complex process. One example is a maintenance diagnostics computer. This type of computer, with a traditional interface, may be frustrating to mechanics with no prior computer experience and costly to their employers while they are learning to use it. Graphical touch interfaces can reduce or eliminate both the frustration and the learning curve by guiding the mechanics through the process with a series of touch-active menus. (In addition, many touch screens are impervious to grease and other substances that would damage a keyboard.)

Touch-Screen Designs

In order to select the touch technology that best fits the needs of a particular process control application, it is important to take a brief look at how each technology functions.

There are six basic types of operating principles in touch technology:

1. Capacitive overlay
2. Guided acoustic wave
3. Resistive overlay
4. Scanning infrared
5. Near field imaging (NFI)
6. Surface acoustic wave

Each type of touch technology has attributes that are desirable for specific applications.

All the touch screen designs are attached to a display unit, which can be a terminal, CRT, flat-panel display, static graphic, or combination of flat-panel display and static graphic. The differences between the technologies lie largely in the way the touch is detected and the method used to process the touch input. The next paragraphs describe the touch technologies that are available today.

Capacitive Overlay A capacitive overlay touch screen consists of a glass panel coated with a charge-storing thin coating. To activate the system, the operator must touch the overlay with a conductive stylus, such as a finger. Circuits located at the corners of the screen measure the capacitance and current flows resulting from the operator's touching the overlay and the current flows, which are proportional to the distance of the finger from the corners. The ratios of these current flows are used to locate the touch.

Capacitive touch screens are very durable and have a high clarity (Figure 4.22a). They are widely used, including industrial control applications.

Guided Acoustic Wave The guided acoustic wave design is based on transmitting acoustic waves through a glass overlay placed over the display's surface. Here a transducer is mounted at the edge of the glass and emits an acoustic wave. The wave packet travels along the reflector array and is redirected across the overlay to the reflecting edge, from which it returns to the array, where it is reflected back to the transducer. The first reflector will send a signal back first, then the second, and so on.

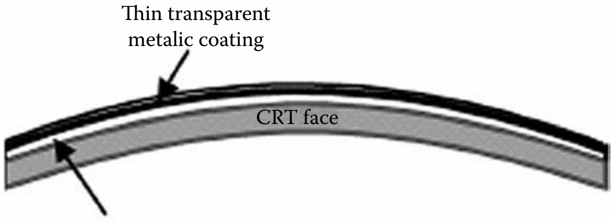


FIG. 4.22a
The design of a capacitive overlay screen.

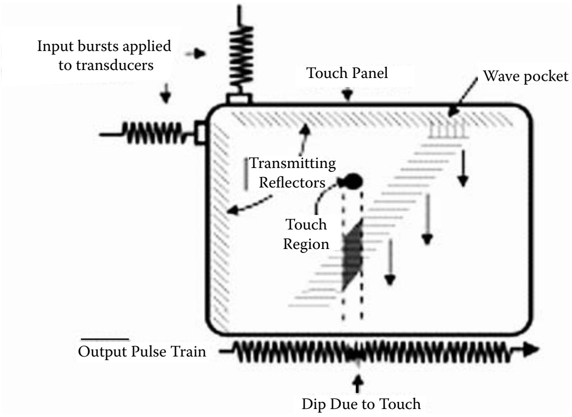


FIG. 4.22b
The guided acoustic wave touch screen design.

When a stylus such as a finger comes into contact with the wave, it attenuates its motion by absorbing part of the wave (Figure 4.22b). Control electronics detect the location of the dip in the wave amplitude, thus determining the location of the touch.

Resistive Overlay A resistive touch sensor consists of a glass or polyester panel that is coated with electrically conductive and resistive layers. The polyester layer has a similar metallic coating on the interior surface. The thin layers are separated by invisible separator dots.

An electrical current travels through the screen. When pressure is applied, the layers are pressed together, causing a change in the current flow, which is detected to locate the touch.

Such resistive touch screens are generally the most affordable. Although their clarity is less than of other touch-screen designs, resistive screens are very durable and are able to withstand a variety of harsh environments (Figure 4.22c). This screen design is recommended for control and automation systems, medical use, and more.

Scanning Infrared The scanning infrared (IR) design relies on the interruption of an IR light grid on the display screen. The touch frame or opto-matrix frame contains a row of IR-light emitting diodes (LEDs) and photo transistors, each mounted on the two opposite sides to create a grid of invisible infrared light. The frame assembly is comprised of printed wiring boards on which the opto-electronics are mounted and

FIG. 4.22c
The design of a resistive overlay touch screen.

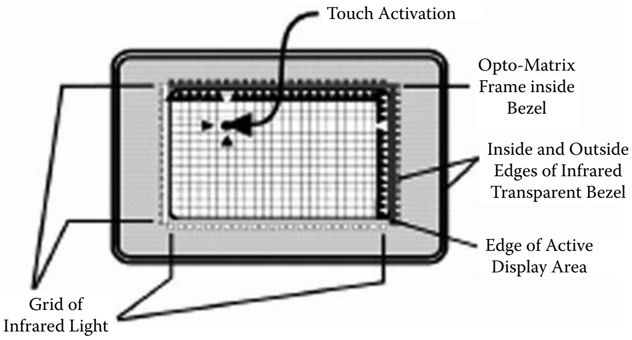


FIG. 4.22d
Scanning infrared touch-screen design.

is concealed behind an IR-transparent bezel. The bezel shields the opto-electronics from the operating environment while allowing the IR beams to pass through.

The IR controller sequentially pulses the LEDs to create a grid of IR light beams. When a stylus, such as a finger, enters the grid, it obstructs the beams (Figure 4.22d). One or more phototransistors detect the absence of light and transmit a signal that identifies the *x* and *y* coordinates of the pouch.

Surface Acoustic Wave The surface acoustic wave design is one of the most advanced touch-screen types. It is based on sending acoustic waves across a clear glass panel, which is provided with a series of transducers and reflectors. Since the speed of the wave is known and the size of the glass overlay is fixed, the first reflector will send the first signal back first, then the second, and so on.

When a stylus such as a finger comes into contact with the wave, it attenuates the wave motion by absorbing part of the wave. This is detected by the control electronics and determines the touch location (Figure 4.22e). This design is the most durable and provides the highest clarity because the panel is all glass and has no layers that could be worn. This technology is recommended for public information kiosks, computer-based training, or other high-traffic indoor environments.

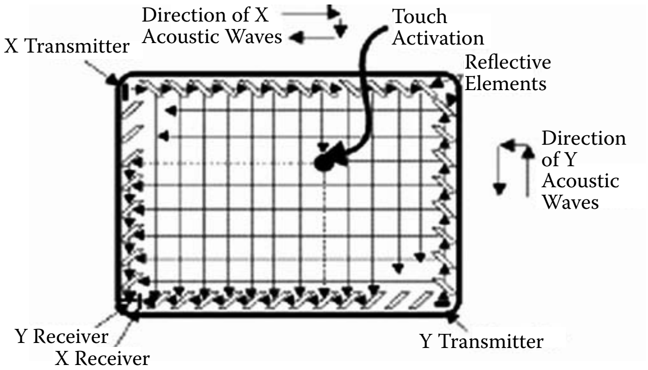


FIG. 4.22e
Surface acoustic wave touch-screen design.

Near Field Imaging (NFI) Near-field imaging (NFI) is based on a proprietary topology/imaging technology. The sensor layout is a piece of glass coated with a pattern of indium tin oxide (ITO) on the front and is completely ITO coated on its back side. The front of the sensor is optically laminated to a layer of passive glass, typically 0.043 in. thick.

An excitation waveform to the conductive sensor is supplied and generates an electrostatic field, which becomes the baseline. When a finger or other conductive stylus comes into contact with the sensor, a change occurs in the electrostatic field. The control electronics then subtract the change from the baseline and determine the peak imaging shape and location to establish the *x* and *y* coordinates of the touch location.

Simply put, in this design the screen itself is the sensor. NFI uses a sophisticated sensing circuit that can detect a conductive object—a finger or conductive stylus—through a layer of glass, as well as through gloves or other potential barriers (moisture, gels, paint, etc.). This is done with a high degree of accuracy using data acquisition and image processing techniques that generate a precise profile of the touch.

The NFI touch-screen sensor uses a transparent conductive film patterned with a proprietary topology applied to the base layer of glass. The front layer of glass is bonded over the base layer with an optical adhesive. An excitation waveform is supplied to the conductive layer by the controller to generate a low-strength electrostatic field in the front layer of glass. The near field is modulated by finger contact with the front layer on the glass, and a differential signal is created, making it possible to accurately resolve the electrostatic loading on the face of the screen.

The system firmware recognizes and decodes the location of the touch. The controller scans continuously until it receives signs of an impending touch. At this point it shifts into a different mode and subtracts the baseline associated with the conditions immediately preceding the touch. This way static and noise do not affect the image of the touch. The profile of the touch is constructed from a dynamic array of data points and is resolved to an actual touch point through continuous reimaging of the electrostatic field. Touch coordinates are fed back to the operating system as fully compliant Microsoft mouse coordinates.

Any long-term changes in the electrostatic image are compensated for, allowing the system to ignore unwanted loading effects from large or distant objects, such as hands or arms, and reject false touches. Sophisticated data acquisition and image processing techniques ensure that NFI can consistently and precisely control the associated equipment, yet it is sensitive enough to detect finger touches through gloves and work in the presence of moisture and other contaminants.

The sensor's glass construction provides superior optical performance and will continue to operate despite scratching, pitting, and other surface damage from abrasives, chemicals, or vandals (Figure 4.22f). NFI touch screens are reliably sealed for applications that require high pressure wash-down or for protection from contaminant-filled environments.

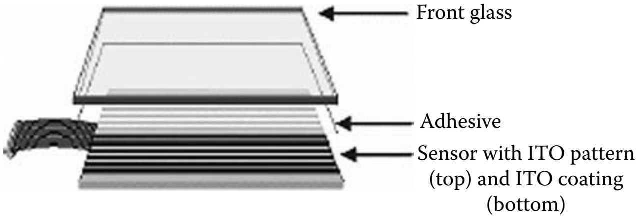


FIG. 4.22f
Near-field image (NFI) touch-screen design.

Evaluating Touch Technologies

Each touch-input technology has both advantages and disadvantages. These result both from the physical factors associated with the technology and from the ability of each design to withstand the environment of the particular application.

Physical Factors Factors to consider include: resolution, transmissivity, response time, stylus type, calibration, integration, and reliability. These factors are specified in Table 4.22h below for each of the touch-screen designs discussed.

The *resolution* of a touch screen refers either to the number of touch-active points or to the physical spacing between the adjacent touch coordinates. When considering the resolution of the touch system, it is important to keep in mind the intended application. For many applications, such as control panels, public access, or computer-based training, fine resolution is not required. In some other applications, such as signature verification, a very high touch-system resolution is desired.

Another variable to consider is image clarity or *transmissivity.* The display image can be affected by the placement of any material between the display image and the viewer's eye. All designs that require an overlay over the display screen (such as resistive, capacitive, surface acoustic wave, near-field imaging, and guided acoustic wave) result in some visual obstruction between the operator and the image on the screen. Transmissivity is defined as the percentage of the light generated by a display that passes through the coating or layered material.

Response time is another important consideration because the faster a touch screen system can respond to an input, the better that design is. Response time is defined as the time required to determine the location of the touch and to transmit that information to the host system. Several factors contribute to the total response time of the application, including the touch-system response time, the host's processing speed, access time to the host's electronics, and the processing time of the application software. Response times specified in Table 4.22g give the response time of the touch system only.

The *stylus type* that can be used is also a selection consideration. A stylus can be an object or an instrument used to activate the touch system and can include a finger, pen, gloved hand, etc.

Calibration is the process of adjusting the active area of a touch system by physically modifying the calibration parameters of the touch (i.e., adjusting potentiometers, setting EEROM parameters, etc.). Although all touch systems require an initial calibration during installation, only systems that are subject to drift (where touch targets gradually move away from the desired locations) require routine or periodic calibration. Capacitive and resistive overlay systems are subject to drift and require calibration, which generally consists of adjusting the offset and scaling parameters to make the touch area equal to, or greater in size, than the display image. Surface acoustic wave, scanning infrared, and guided acoustic wave designs are not subject to drift and do not require calibration after installation.

Integration is the process of attaching the touch system to the display. Invasive integration requires the disassembly of the display to attach the touch system. Typically, this type of integration results in voiding both the manufacturer's warranties and the Federal Communications Commission (FCC) certification. This type of integration requires from 15 to 90 minutes of a skilled technician. Noninvasive integration does not require disassembly of the display. This type of integration can be performed by unskilled people and usually takes less than 10 minutes.

Reliability of a touch system is the time or the number of touches it is expected to last before it fails. Touch systems with polyester or conductive coatings will fail after an anticipated number of touches, which will wear the coatings off. The reliability of other touch systems that do not wear by use is measured in *mean time between failures* (MTBF). This number is a function of the average life expectancy of the electronic components built into the touch system.

Environmental Factors The ability of the various touch-screen designs to withstand a variety of environmental conditions is summarized in Table 4.22h.

Sealability is the ability to seal a touch system (including display electronics) from dirt, liquids, etc. If the system is going to be located in an area where contaminants are present, sealability can be an important consideration. In environments such as surgical operating rooms, sealability is critical. All touch-screen designs can be sealed to meet NEMA 12, which only requires that a system be operational after accidental splashing or cleaning. Capacitive overlay, guided acoustic wave, resistive overlay, near-field imaging, and scanning infrared technologies can be sealed to also meet NEMA 4, which requires that the system continue to operate even after it has been exposed to a hose-down by water.

Durability is the ability to withstand millions of touches over many years and *resistance to vandalism* is the ability to resist defacement—scratching, breaking, theft, etc. Both are evaluated in Table 4.22h.

An excessive buildup of *dust, dirt, or other contaminants* can adversely affect the performance of some touch technologies. Capacitive overlay, scanning, infrared, and surface acoustic wave technologies will operate with low-to-moderate

TABLE 4.22g
*Touch System Comparison Chart 1**

Type of Design	*Resolution and Z-Axis*	*Transmissivity*	*Activation, Parallax and Response Time*	*Stylus Type*	*Sensor Drift and Calibration*	*Integration*	*Reliability*
Capacitive overlay	1024 × 1024 physical, no *z*-axis	85 to 92%	Tactile activation, no parallax, 15 to 25 ms	Requires conductive stylus; unable to simultaneously detect gloved and ungloved finger	Subject to drift; requires repetitive calibration	Invasive and noninvasive; optical bonding required for optimum display clarity	Sensor: 20 million touches per point; controller: > 186,000 hours MTBF**
Near-imaging	1024 × 1024 physical, no *z*-axis	85%	Tactile activation, parallax, 15 to 25 ms	Requires conductive stylus	Not subject to drift	Invasive	Sensor: unlimited. controller: >180,000 hours MTBF
Guided acoustic wave	21,904 points/square inch physical, plus *z*-axis	92%	Tactile activation, no parallax, 18 to 50 ms	Requires soft, energy-absorbing stylus	Not subject to drift	Invasive; optical bonding required for optimum display clarity	Sensor: unlimited; controller: >180,000 hours MTBF
Resistive overlay	256 × 256 to 4096 × 4096 physical, no *z*-axis	55 to 78%	Tactile activation, no parallax, 13 to 18 ms	No stylus limitation	Subject to drift; requires repetitive calibration	Invasive; optical bonding required for optimum display clarity	Sensor: 2 million touches per point; controller: 86,000 to 180,000 hours MTBF
Scanning infrared	0.25 in. physical, 0.125 in. logical, no *z*-axis	100%	Proximity activation, parallax, 18 to 40 ms	No stylus material limitation. Minimum stylus diameter 5/16 in.	Not subject to drift	Invasive and noninvasive	>138,000 hours MTBF
Surface acoustic wave	0.030 in. physical, plus *z*-axis	92%	Tactile activation, no parallax, 53 to 59 ms	Requires soft, energy-absorbing stylus	Not subject to drift	Invasive; optical bonding required for optimum display clarity	Sensor: 50 million touches per point; controller: 86,000 to 118,000 hours MTBF

[*] Manufacturer's published data.
[**] Mean time between failure.

TABLE 4.22h
*Touch System Comparison Chart 2**

Type of Design	*Durability/Resistance to Vandalism*	*NEMA Ratings, Moisture Resistance*	*Dust and Dirt Resistance*	*Chemical Resistance*	*Vibration and Shock Resistance*	*Ambient Light*	*Temperature, Humidity, and Altitude*
Capacitive overlay	Difficult to scratch; conductive layer is subject to wear; glass overlay is breakable	NEMA 12; NEMA 4	Will operate with moderate dust and dirt; excessive accumulation will affect performance	Not affected by general-purpose cleaning solutions	Tolerant of vibration; thick glass overlay moderately susceptible to shock	Unaffected by ambient light	0 to 70°C; 0 to 95% nonconducting humidity; 30,000 ft (9000 m)
Near-field imaging	Difficult to scratch; glass overlay is breakable	NEMA 12; NEMA 4	Not affected by dust and dirt	Not affected by general-purpose cleaning solutions	Tolerant of vibration and shock	Unaffected by ambient light	0 to 50°C; 0 to 95% nonconducting humidity; altitude not specified
Guided acoustic wave	Difficult to scratch; glass overlay is breakable	NEMA 12; NEMA 4	Not affected by dust and dirt	Not affected by general-purpose cleaning solutions	Tolerant of vibration; glass overlay susceptible to shock	Unaffected by ambient light	0 to 50°C; 0 to 95% nonconducting humidity; altitude not specified
Resistive overlay	Sensor is vulnerable to scratches and abrasion; glass overlay can be broken	NEMA 12; NEMA 4	Not affected by dust and dirt	Not affected by general-purpose cleaning solutions; chemicals that affect polyester should not be used	Tolerant of vibration; glass overlay susceptible to shock	Unaffected by ambient light	0 to 50°C; 0 to 95% nonconducting humidity; 15,000 ft (4500 m)
Scanning infrared	Not susceptible to scratching; no overlay to break; completely solid state; no exposed parts	NEMA 12; NEMA 4	Will operate with moderate dust and dirt; excessive accumulation may affect performance	Not affected by general-purpose cleaning solutions; chemicals that affect polycarbonates should not be used	Tolerant of vibration and shock	Varies by manufacturer	0 to 50°C; 0 to 95% nonconducting humidity; altitude not specified
Surface acoustic wave	Difficult to scratch; glass overlay is breakable	NEMA 12	Will operate with moderate dust and dirt; excessive accumulation may affect performance	Not affected by general-purpose cleaning solutions	Tolerant of vibration; glass overlay susceptible to shock	Unaffected by ambient light	0 to 50°C; 0 to 95% nonconducting humidity; altitude not specified

* Manufacturer's published data.

accumulations of dust, dirt, and other contaminants. Excessive levels will affect their performance. Guided acoustic wave, resistive overlay, and near-field imaging designs are not affected by dust, dirt, or other contaminants.

General-purpose cleaning solutions have no harmful effects on any of the touch-screen designs. However, some of the designs can be attacked by certain *chemicals*. A resistive overlay touch system has an exposed polyester overlay. Therefore, such design should not be used if chemicals that attack polyester will be used in the area. A scanning infrared touch system has exposed polycarbonate bezels. Therefore this design should not be used if chemicals that attack polycarbonate, such as petroleum-based chemicals, are going to be present in the area where the screen is to be located. More resistant materials are also available for the construction of special IR touch systems but are not widely used and therefore must be clearly specified.

Vibration and shock is an important consideration if the application is installed on a moving device or on equipment that is subject to significant vibration or shock.

Ambient light is the level of visible and invisible light in the operating area; most designs are not affected by it. On the other hand, some infrared designs are so packaged that they can be adversely affected. The levels of ambient light that are found in well-lit indoor environments do not present any problems.

The *temperature, humidity, and altitude* at which the touch screen is to operate have effects on the operation and durability of any touch system. Plastics and electronics are affected by temperature. Humidity enhances the corrosion of circuitry. Altitude affects the dissipation of the heat generated by functioning electrical circuitry. It is safe to say, however, that the impact of these environmental factors on touch systems will be less (or at least no greater) than their impact on the circuitry of the display or flat panel into which the touch system is integrated.

Specific data regarding the impact of these environmental factors is provided in Table 4.22h.

OVERALL SYSTEM DESIGN

The systems integrator must pay attention to the interrelated considerations of mechanical and physical attributes and also to programming considerations.

Mechanical Considerations

Display Selection Selecting the computer display is one of the major decisions affecting the cost of the complete touch system. Today, the display choices include both flat panels and CRTs. Flat-panel displays tend to present the fewest mechanical design concerns for touch systems, while CRTs typically require some degree of mechanical design compensation to correct for the curvature of the display surface. As a general rule, touch applications that require a large number of targets on a single screen should use a larger size display.

Touch System Integration Touch-system designers can choose between invasive and noninvasive integration options. An invasive integration typically requires that the display be disassembled. This is very time-consuming and may void the factory warranties of the display. Noninvasive integration requires very little time to assemble and will not affect the manufacturers' warranties. Maintenance is simplified as well.

Space Constraints Many touch systems are located in areas where space is limited. Flat panels require very little space, while CRTs tend to be bulky. The amount of physical space required for the touch system varies for the various touch screen designs. The type of integration method selected can also affect the amount of space required for the system. In general, invasive integration tends to require less space than noninvasive integration.

Environmental Factors Touch-systems are designed to operate under a wide range of environmental conditions. When designing a touch-based system, one should consider the sealability, durability, reliability, and vulnerability (to vandalism) of the system, as was discussed in connection with Table 4.22h.

Sealing Depending on the environmental conditions, the touch system may require different degrees of sealing. Applications that may require special consideration include industrial and process control applications and outdoor applications.

Physical Attributes

System Capabilities The system designer must match the application requirements to the capabilities of the touch-screen design. Factors to be considered include glare, transmissivity, resolution, stylus types, and aesthetics. For instance, systems that require high-quality graphics would gravitate toward those touch technologies with the best transmissivity and the least glare. Applications requiring handwriting recognition would need those technologies with the highest resolution.

Another consideration is the type of stylus that will be used in the specific application. Other factors to consider include the availability of features such as multiple operating and reporting modes, improved software resolution, fault tolerance, and diagnostics.

Touch systems can be programmed to detect multiple styli, calculate the size and center of the stylus, reject a stylus that is larger or smaller than the specified limits, or require that the stylus remain in the touch-active area for a specified amount of time before the touch is considered to be a valid hit.

Another operating feature available with some touch technologies is the capability to report a z-axis coordinate, which measures the amount of pressure applied to the sensor. Typically, the harder the user presses on the sensor, the higher the z-axis value. This z-axis coordinate is often used to emulate mouse button events. This is done by comparing the z-axis coordinate to the threshold value.

Programming Considerations

For each application, the programmer must consider the interaction or communication:

1. Between the touching operation and the user
2. Between system hardware units
3. Between the system software packages

Hardware Interface Options Touch systems are designed to interface with the host in a variety of different configurations.

One configuration involves communicating through serial ports, using the RS-232, the RS-422 or other protocols. Another configuration involves parallel communication via a bus standard. Examples include ISA/EISA, micro-channel, and PC/104™.

Another interface method uses the mouse port. In this method, the touch system is connected to the host computer's mouse port and uses either a standard mouse interface protocol or a proprietary touch-system interface protocol.

Another interface configuration involves the use of daisy-chained input devices, such as the Apple Desktop Bus™. Some vendors have chip sets available that can be added to the host's electronics and packaging. The chip sets typically contain all of the functions required to control and communicate with the touch system as well as the host's electronics. Touch vendors have proprietary software interface protocols unique to their touch systems. All of the touch vendors' software protocols report an x and y coordinate; some technologies permit the reporting of a z coordinate as well.

Software Interface Options In the *direct interface method,* the touch application communicates directly with the touch system using the proprietary interface protocol of the touch system.

Software drivers are available from touch vendors that assist the touch application developer in interfacing to the touch system. These drivers usually provide calibration and scaling support, plus they communicate with the touch system using the touch vendor's proprietary interface protocol and with the touch application via a simple application program interface (API).

An *authoring system* is a program that the application developer can use to create a touch application without writing programming code. Authoring systems either include direct support for the touch system or run under a graphical user interface (GUI), which provides support for the touch system. Hypercard™ for the Macintosh and Asymetrix Toolbook™ for the IBM PC™ are examples of such authoring systems.

Mouse emulator drivers that make the touch system appear to be a mouse to the application code are available from touch vendors. The driver emulates the standard Microsoft® Mouse driver protocol. With the touch system connected and the mouse emulator loaded, applications that use a mouse may be used with touch instead. Little or no modification of the application is required. If the application uses targets that are of sufficient size to be used with a mouse but are too small to be used with touch, the application would have to be modified to enlarge the targets for touch use.

Graphical user interfaces (GUIs) such as Apple Macintosh™ or Microsoft Windows™ are operating systems that use icons, pull-down menus, windows, etc., instead of keyboard-entered commands. GUIs typically support installable pointing device drivers. The touch vendor supplies the pointing device driver, which generates the pointing device event messages that are sent to the application. Applications are written to use these standard pointing device event messages, and are therefore independent of any particular pointing device, such as a mouse, touch system, graphics tablet, or other pointing devices.

The *touch user interface* application program is the interface between the user and the computer system. The application program presents displays, accepts user input, and takes action based on that input. The design and organization of the program are critical to the successful use of the touch system, especially when the end users are likely to be novices.

Interface Design Factors To maintain the natural simplicity of the touch interface and to lead the user easily through the program, the interface designer should be aware of the following factors:

Touch *target location* on the screen is determined by the relative importance of the target. Consistency must also be considered. The user will locate targets with greater speed and with less confusion or errors when targets of the same or similar function are consistently located in the same relative location on the screen.

The *number of targets* per screen should be limited to as few as possible, balanced by the difficulty of switching screens. Nesting and prioritizing relieve the need to crowd targets on the screen, hence reducing the potential for human error. In the case of menus, more items can be put on the screen. But if menus are nested too deeply, users will soon tire of searching through the menus. The use of graphic symbols (icons) for touch targets can be effective in helping the user in quickly identifying the targets.

The *size of targets* is limited by the stylus size. The number of errors can be reduced by increasing the size of the targets. When designing targets for finger activation, research has shown that few fingertips are more than 22 millimeters across. Each target should be surrounded with a guard band or dead zone, where touches are not recognized. Guard bands reduce the possibility of user confusion and frustration by eliminating the possibility of activating an adjacent target.

Touch activation mode refers to the behavior of the target when it is touched. Slides, switches, and buttons are typical variations of touch targets. A typical button target has three states: unarmed, armed, and activated. The simplest activation mode is the activation of the target whenever a finger is over the target. The target proceeds directly from the unarmed state to the activated state in much the same manner as a mouse pointer might activate by simply pointing to an area without clicking a button.

This method of activation is the least desirable because it is a one-step process that lacks a means of canceling the activation of a target. This makes unintended activation likely. A variation of this method is to activate the target when the finger is removed from the screen over the target. This activation mode is only slightly better, since it also means that there is no way of canceling the activation; the result can be inadvertent activation.

The most frequently used activation mode is to cause the target to go to the armed state whenever the finger is over the target. If the finger is over the target when it is withdrawn from the screen, the target is activated. If the finger moves away from the target before the finger is withdrawn from the screen, the target is not activated. This allows the user to cancel the selection.

Touch feedback is an integral part of most well-designed activation modes. The user must receive immediate feedback to know for certain when a target has been armed and/or activated. Highlighting, changing color, or depressing a chiseled button are all good visual feedback techniques for indicating that a touch target is armed or activated.

The screen should not be allowed to go blank, even if dummy screens must be designed to fill up the display screen. To the inexperienced user, a dark screen is a sure sign that the system has failed. Audio feedback is an appealing complement to visual feedback, with various tones or sounds being used to indicate the target activation state. This effect can be especially impressive on multimedia computers.

Bibliography

Handbook of Touch Technology, Elo TouchSystems, Inc, www.elotouch.com, 1998.

Mason, J. R., *Switch Engineering Handbook,* New York: McGraw-Hill Professional, 1992.

4.23 Uninterruptible Power and Voltage Supplies (UPS and UVS)

S. P. JACKSON (1970, 1985) **B. G. LIPTÁK** (1995, 2005)

Types:	The energy source of uninterruptible power supplies (UPS) can be batteries or engines. In uninterruptible voltage sources (UVS), capacitors allow the system to ride through an outage of several cycles.
Costs:	A UVS rated for 1500 VA costs around $2500. A UPS rated for 3 KW costs about $7500. For the costs of batteries, alternators, chargers, and inverters, refer to the Figures in the "System Components" paragraph of this section.
Partial List of Suppliers:	Acopian (www.acopian.com) Constant Power Manufacturing (www.constantpowermfg.com) Desert Microsystems Inc. (www.desertmicrosys.com) EnerSys Inc. (www.enersysinc.com) Foxboro-Invensys (www.foxboro.com) Franek Technologies Inc. (www.franek-tech.com) GE Fanuc Automation (www.gefanuc.com) Graybar Electric Co. (www.graybar.com) Ice Qube Inc. (www.icequbе.com) ITS Enclosures (www.itsenclosures.com) Kontron (www.kontron.com) Leviton Manufacturing Co. (www.leviton.com) MGE UPS Systems (www.mgeups.com) Omron Electronics Inc. (www.omron.com/oei) Phoenix Contact Inc. (www.phoenixcon.com) Phonetics Inc. (www.sensaphone.com) Quest Inc. (www.questinc.com) Rice Lake Weighing System (www.rlws.com) SatCon Power Systems (www.satcon.com) Schneider Electric/Square D (www.squared.com) Sola/Hevi-Duty (www.solaheviduty.com) Superior Electric (www.superiorelectric.com) Toshiba International Corp. (www.tic.toshiba.com) Transtector Systems (www.transtector.com) Westinghouse Process Control (www.westinghousepc.com) Woodhead Connectivity (www.connector.com)

INTRODUCTION

After this introduction, the section will first give a general description of the features of uninterruptible voltage and power supplies (UVS and UPS), followed by a brief discussion of the requirements of buses and networks. This will be followed by a discussion of the classes of power failures that are likely to occur, the components of UPS and UVS systems, and the types of standby and redundant power supplies.

Electric power lines are commonly assumed to be perfectly reliable sources of constant voltage. This assumption is valid when complete source reliability, particularly on a short-term basis, is not important. Control and instrument engineers frequently make this assumption for power supply systems that do not satisfy this criterion. Obvious long-time power outages are fairly rare in modern power systems. Notable exceptions, however, have called attention to the complete spectrum of possibilities of power failure.

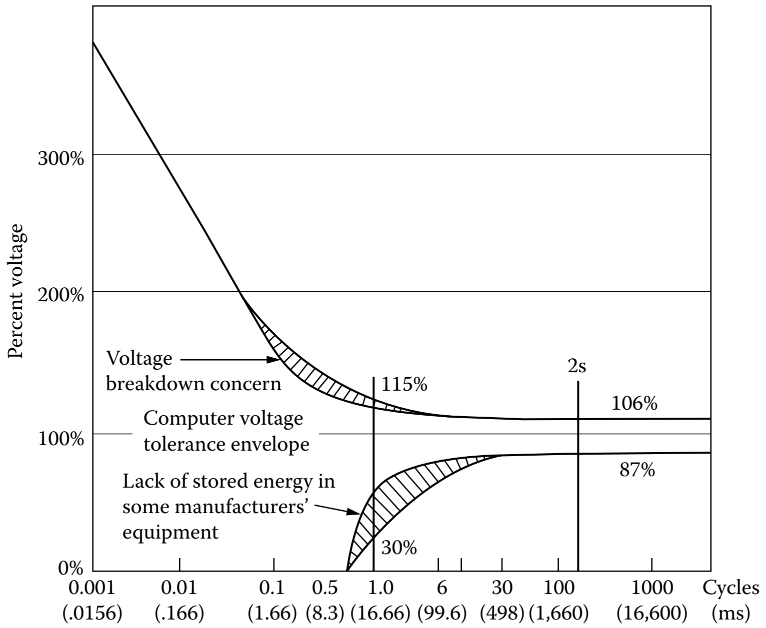


FIG. 4.23a
The safe operation area (SOA) of computers defined in terms of allowable input voltage variations and their duration. (Source: ANSI/IEEE Standard 446–1987.)

Voltage dips too short to be noticed by human senses can occur frequently during the lightning season on exposed suburban and rural power systems. These systems serve many industrial plants. In addition to the obvious power failures caused by weather, there are many transient variances resulting from electrical, mechanical, and human occurrences. A system designer who assumes a perfectly reliable power source is responsible for any loss of production and damage to machines and plant facilities resulting from power irregularities.

The trend in control system design is to use faster components operating on smaller signals. This results in increasing system sensitivity to line voltage variation and in particular to short transients. As control instrumentation technology advances, the need for standby power sources increases. Frequently, the cause of the transients that result in control or instrument circuit complications is not the lack of voltage alone. Phase shift, a change in frequency, inadequate transient response, and noise can be equally damaging. Therefore, an adequate standby power supply system must consider all types of power failure.

UNINTERRUPTIBLE VOLTAGE SOURCES (UVS)

The IEEE-446 standard defines the "Safe Operating Area" for computer-based equipment in terms of the power disturbances that are allowable. As the time duration of the disturbance increases, the allowable magnitude of the disturbances that can be tolerated diminishes (Figure 4.23a).

The purpose of uninterruptible voltage sources (UVS) or line conditioners is to provide protection against short-duration brownouts or outages by guaranteeing that the output voltage is not lost and its waveform remains stable for several cycles after utility loss. This "ride-through" time can extend to several cycles (Figure 4.23b) and can prevent processor crashes. As the ride-through capability is provided by programmable capacitors, it can be extended by the addition of capacitors.

UVS units are a little less expensive than the battery-based UPS units (to be discussed later) and are designed for the hostile environment of the factory floor. They regenerate new AC power in pure sine wave form and keep the output voltage and frequency within 3% of normal, even if the input voltage or frequency experiences wide variations. The ride-through capability enables the user to continue operation when the

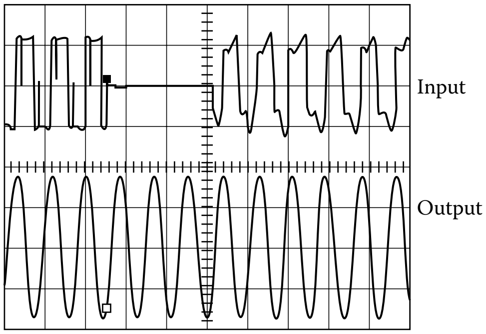


FIG. 4.23b
An uninterruptible voltage supply (UVS) can allow the computer equipment to "ride through" an input outage of several-cycle duration. (Courtesy of Falcon Electric.)

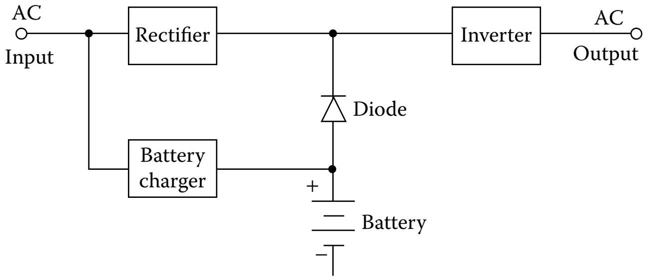


FIG. 4.23c
Standby system with battery charger redundancy.

outage is short or, when the duration of the outage is longer, to anticipate the loss of power in time to signal the processor to stop at a known state.

UNINTERRUPTIBLE POWER SUPPLY (UPS) FEATURES

Standby systems (UPS) can be quite complex. They involve the use of a number of components that may be of the static type, the electromechanical type, or some combination of both. The selection of the standby system and the components for it should be based upon the degree of integrity required of the application.

It is frequently possible to improve standby system reliability and to decrease cost. An example is that of the redundant input circuit shown in Figure 4.23c. By separating system functions, it is possible to purchase the component that does the precise job required.

Much attention has been focused on the problem of failure of the incoming power line. Although this is an important consideration, it is not the only consideration. Attention should also be given to the equipment constituting the standby power supply system. Should critical pieces of the standby equipment be less reliable than the incoming power line, failure will be more frequent, and little improvement will have been accomplished.

Also of importance is the load circuit. If a number of load branches are connected to the output of a standby power supply system, the failure of any one load may result in the failure of the remaining loads. Proper design of the load system will minimize or eliminate this possibility.

The importance of system redundancy cannot be overemphasized. Once designed, the system must be evaluated to determine its weakest link. The cost of redundant equipment may then be assessed in light of the importance of maintaining the load.

Table 4.23d summarizes the standby system classifications and lists the most commonly used components. Obviously, special arrangements may be found that extend some of the components into an additional classification. This table is intended as a means of outlining the discussion that follows.

TABLE 4.23d
Summary of Standby System Classifications and "Most Used" Component Parts

	Standby System Classification		
System Componet	*Multi-Cycle*	*Sub-Cycle*	*"No-Break" (Minimum Transient)*
Secondary power source	Engine-driven alternator or generator *starting* on primary source failure	Engine- or motor-driven alternator or generator *running* with flywheel	Battery
Inverter	Rotating or static	Rotating or static	Static
Bus transfer switches	Electromechanical	Static	Static

Various degrees of *redundancy* may be designed into the standby system. This redundancy may occur on the input side of the inverter, on the output side, or on both sides. A tabulation of components in a single- and a double-source redundant system is provided in Table 4.23e. Both input redundancy and output redundancy are included.

NETWORKS AND BUSES

Fieldbus devices work with supply voltages in the range of 9 to 32 V DC. The 9 V DC is a minimum and a margin of at least 1 V (a minimum of 10 V DC) is recommended to be maintained. The documentation of the bus segment should always show the minimum voltage, and a warning concerning additional loads should be posted on segments operating under 15 V.

TABLE 4.23e
Standby System Component Redundancy

Level of Redundancy	*Input Side*	*Output Side*
One source	Battery and battery charger	Inverter
One source with some equipment redundancy	Battery, battery charger, and rectifier	Inverter with transfer to line
Two sources	Battery, battery charger, and engine-driven generator	Inverter with transfer to alternate inverter
Two sources with some equipment redundancy	Battery, battery charger, rectifier, and engine-driven generator	Inverter with transfer to alternate inverter and then to line

If an ordinary power supply were used to power the fieldbus, it would absorb the communication signals on the cable, due to its attempt to maintain a constant voltage level. Fieldbus power supplies are therefore conditioned by placing an inductor between the power supply and the field cable,

TABLE 4.23f
Examples of Power Component Used in Foundation Fieldbus and PROFIBUS-PA Systems

Type	U_s	I_s	R_Q	*Remarks*
MTL 5053	18.4 V	80 mA	105 Ω	IS power supply with power conditioner and switchable terminator
MTL 5995	19.0 V	≤350 mA	<2 Ω	Non-IS power supply with power conditioner and switchable terminator
Relcom PCS-PC	$V_{input} - 5$ V	≤330 mA	—	Power conditioner
Relcom FCS-PCT	$V_{input} - 5$ V	≤330 mA	—	Power conditioner with terminator
Siemens 6ES7-157-0 AD00 0XA0	12.5 V	100 mA[a]	—	PROFIBUS segment coupler EE*x* [in] IIC
Siemens 6ES7-157-0 AC00 0XA0	19.0 V	400 mA[a]	—	PROFIBUS segment coupler for safe areas
Pepperl + Fuchs KFD2-BR-EX1.2PA.93	13.0 V	110 mA[a]	—	PROFIBUS segment coupler EE*x* [ia] IIC
Pepperl + Fuchs KFD2-BR-1PA.93	25.0 V	380 mA[a]	—	PROFIBUS segment coupler for safe areas

Note: Inclusion in this table does not imply a recommendation; other suppliers exist.

[a] The bus current has already been subtracted from these specifications.

and a resistor is added to the inductor to prevent ringing. The power conditioner might consist of a 50-ohm resistor and a 5-mH inductor in series. The fieldbus conditioning unit is normally installed at the start of the fieldbus segment and often contains a bus terminator. If the segment runs into an explosion-proof area, a barrier is also required. Table 4.23f gives some examples of conditioning unit components.

Several layers of UPS might be required in a plant-wide SCADA process control system. First, the clock and memory of all digital equipment, computers, and remote terminal units (RTUs) must be continuously supplied by power. A small dedicated battery with a life of several years can be used for this. Next, the critical sensors, transmitters, and RTU logic need to be supplied until normal power is restored. In some installations the requirement is for 35 days of UPS power.

POWER FAILURE CLASSIFICATIONS

Standby power systems can be characterized according to the time it takes to achieve full output from the standby power source after failure of the primary source. This transfer time might be as long as many cycles—for large electromechanical switching devices or for engine-driven alternators or generators that start up on failure of the primary source—or as short as fractions of a cycle—for some of the solid-state switching devices or for motor-alternator or generator sets with flywheels. Additionally, no-break systems are available.

Most early standby systems necessitated an interval during which there was no voltage to the load on transfer from the primary source to the standby source or on retransfer from the standby source to the primary source. As control and instrumentation circuits have become more critical, standby systems have been developed that include new techniques for transferring power sources such that essentially no transfer time occurs. For the most part, these no-break systems cost little more than those requiring a significant amount of time to transfer. In general, this characterization of the standby power system by transfer time is disappearing, since a large percentage of present standby systems are of the no-break variety.

Another means of characterizing standby systems is by type of failure. One's first thought is to protect the critical loads from failure of the commercial power line. A careful scrutiny into the system suggests that there also are other points worthy of consideration. Among these is the failure of standby power supply components and the failure of the load.

Source Failure

A very simple standby power system is shown in Figure 4.23g. It consists of an AC power line feeding a battery charger. The battery charger, in turn, floats a battery that provides power to the inverter. The inverter provides an AC output through a distribution panel to a number of loads. Should the AC line fail, the battery charger will cease to provide the current to the inverter. The current will then be provided by the battery that is floating on the system.

In this fashion, the inverter supplies the loads until such time as the AC line is reenergized, at which point the battery charger again provides the power for the inverter and for the loads and at the same time provides recharge current to the battery. Thus, the simple standby system of Figure 4.23g protects against a line failure since there is no cessation of power to the loads when the AC line fails.

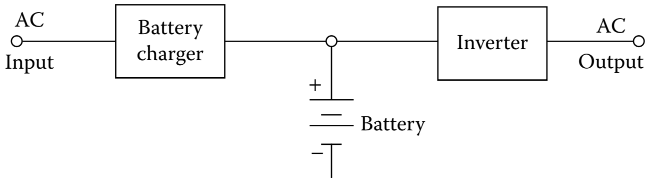


FIG. 4.23g
Basic AC standby system.

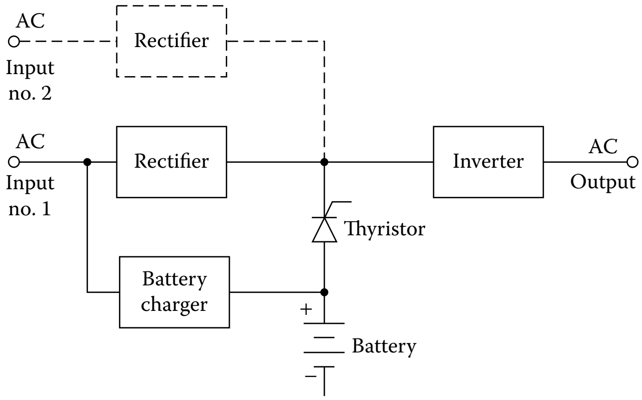


FIG. 4.23h
Standby system with multiple AC inputs.

A suggested improvement in the basic system is shown in Figure 4.23h. Because the functions of the battery charger are separated into (1) supplying steady-state running current to the inverter and (2) supplying recharge current to the battery, two rectifiers can be used. An unregulated rectifier is adequate to supply load current by means of the inverter. The battery charger rectifier must be regulated to ensure long battery life. Since the unregulated supply is less likely to fail, some additional reliability is gained. This can be seen when one considers the results of failure of the battery charger.

In the circuit of Figure 4.23g, should the battery charger fail, the system is no longer operable after the energy stored in the battery is consumed by the load. In the case of Figure 4.23c, however, should the battery charger fail, the system continues to function as long as the AC input is available. Should the AC input source fail, the system will continue to operate until the energy stored in the battery is consumed.

If, prior to this time, the AC input source is restored, the system continues to function properly but the battery charger is not capable of recharging the battery. While the system continues to operate properly, the battery charger may be repaired if it is possible to do so between the time that the failure of the battery charger is noticed and the next failure of the AC input source. Indeed, if nothing else can be done, it is generally possible to bypass the diode with some available resistance (even a light bulb) that will restore some energy to the battery. Even a small amount of battery capacity is ample for a number of short, transient outages.

In addition to the increased reliability of the two sources noted in Figure 4.23c, a lower cost for this system frequently results. This lower cost is attained because the unregulated rectifier in most cases is providing a larger current than the battery charger. Thus the rectifier capacity is greater than that of the battery charger. Since it is less expensive to buy unregulated power than to buy regulated power, it is possible under many circumstances to achieve a lower cost. This combination of lower cost and increased reliability is the optimum objective of the system designer.

The system of Figure 4.23c can be extended to include more than one AC source, as illustrated by Figure 4.23h. Alternative sources may include other AC lines and the output from engine-driven alternators, or, indeed, from any alternator regardless of the number of phases, the voltage, the frequency, or the variation in frequency.

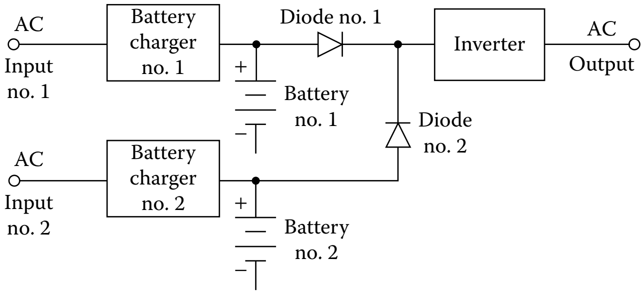


FIG. 4.23i
Standby system with multiple-input redundancy.

It may be desirable to "stagger" the input voltage ranges of the sources to favor one or another source. Since this "staggering" of sources results in an increase in input voltage variation over which the inverter must operate, a thyristor has been included to provide a dynamic switching of sources, minimizing the input voltage variation. The peak value of the alternative sources when rectified must be greater than the battery potential in order to "turn off" the thyristor.

The ultimate in input redundancy occurs with more than one complete power source. Shown in Figure 4.23i is a system with source, battery charger, and battery redundancy. It is also wise to separate the power feeds to the inverter. This system can be extended to any desired degree.

The systems of Figures 4.23c, g, h, and i offer a continuous source of power to a load without regard to the state of the AC input source as long as the standby source has sufficient energy to supply the load. The options noted provide any degree of redundancy desired for the standby power source. These figures show that this redundancy is adequate to ensure the most critical loads. The diagrams also make it clear that, should a failure occur in the inverter, the load source is no longer protected. Thus, our next concern must be the failure of equipment within the inverter block.

Equipment Failure (Inverter)

In many applications of standby power, the integrity of the line must be maintained in spite of equipment failure. Failure to preserve this integrity can result in loss of output, scrap material, plant damage, or loss of life. The degree of the protection necessary depends on the damage that can result. Process control computers are particularly important in plant operations because failure of the computer system results in an uncontrolled process. Loss may be sufficiently high to justify greater system redundancy.

The simplest form of output redundancy is illustrated by Figure 4.23j. A bypass switch is provided from the output of the inverter to the AC input line. In this diagram, and in many to follow, the symbol for an electromechanical switch (relay)

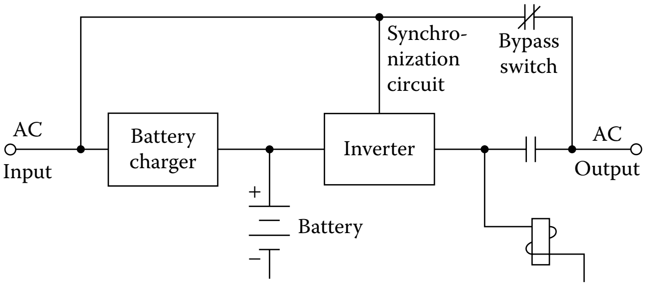


FIG. 4.23j
Output redundancy to AC input.

will be used. This symbol should be construed to include both static and electromechanical devices.

Two items are essential in the operation of the circuit of Figure 4.23j: a synchronization circuit and a means of sensing source failure. A synchronization circuit has been added to ensure that both the AC input and the inverter are in phase in order to minimize the switching transient. The switching device or devices have also been added, together with appropriate sensing circuitry.

At this point, it is easy to gloss over an essential discussion masked by the obviousness of the preceding remark. Consider for a moment the fact that both voltage waveforms must be of the same frequency and in phase. On failure of the input AC line, no transfer occurs since the output of the inverter is not impaired. In the event of a loss of output from the inverter, transfer to the AC source takes place. Retransfer to the inverter can occur when the inverter output is established and when synchronization of the outputs has been restored. The retransfer may occasion an output voltage transient, as well as the transfer, since stored energy in the filter or the inertia of the rotating alternator used in the inverter requires some time to bring up to full output current.

Note that the addition of the line synchronization capacity has increased the number of components in the inverter, thereby decreasing its reliability. The synchronization circuit must be carefully designed to eliminate any AC transients that may cause a failure in inverter output.

The point of detection of inverter output failure is instant. Sensing as early in the circuit as possible provides a better transfer since energy stored in the filter may be used to reduce the transient. Early detection of thyristor failure or of abnormal vibration is preferable to the simple detection of reduced output voltage.

Difficulties can occur in providing a sensing circuit that operates on an adjustable reduced output voltage level. If no delay is built into the sensing circuit, transfer can occur on line transients, causing frequent operation. If the normal delay is included in the retransfer circuit, no system redundancy occurs in the interval defined by the time of transfer and the delay before the retransfer.

To minimize transfers on simple line transients, which may be caused by sudden load demands, an integrating circuit may be inserted in the voltage level sensor. Although this eliminates the transfer resulting from short transients, it causes a greater output transient when a transfer is made because of equipment failure.

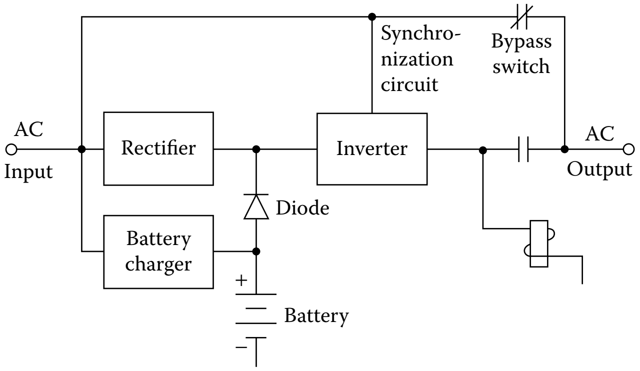


FIG. 4.23k
Input and output redundancy.

The difficulties noted strongly suggest that consideration be given to early failure detection. The importance of this portion of the standby power supply system cannot be overemphasized. It is wise to provide a double sensor system, which on either the failure of a component such as thyristor or on the failure of a bearing (resulting in abnormal vibration) will provide a delay, as well as an integrating sensor circuit and detecting the reduction of the output voltage.

Input and output redundancy is shown in Figure 4.23k. Any of the previously discussed redundant input schemes can be used.

With AC input redundancy, the integrity of the AC line must be considered. Naturally, if frequent line disturbances occur, little is achieved by such an arrangement. The only gain is the possibility of not having a line failure until something can be done to reestablish source redundancy.

The degree of redundancy can be improved if a backup inverter standby system is provided. Figure 4.23l illustrates this emergency power supply backing up an emergency power supply but with dual loads. Here, each inverter is capable of

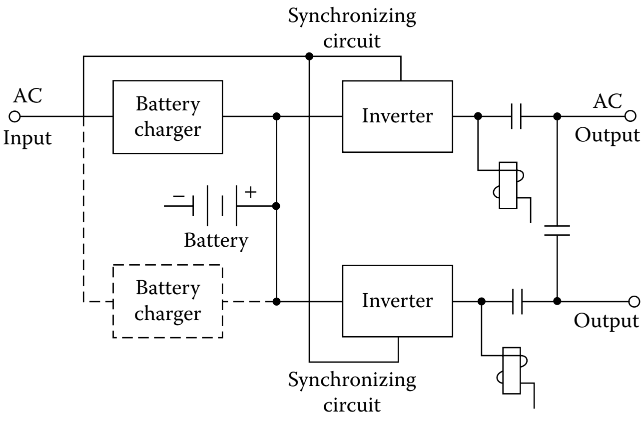


FIG. 4.23l
Redundancy with dual loads.

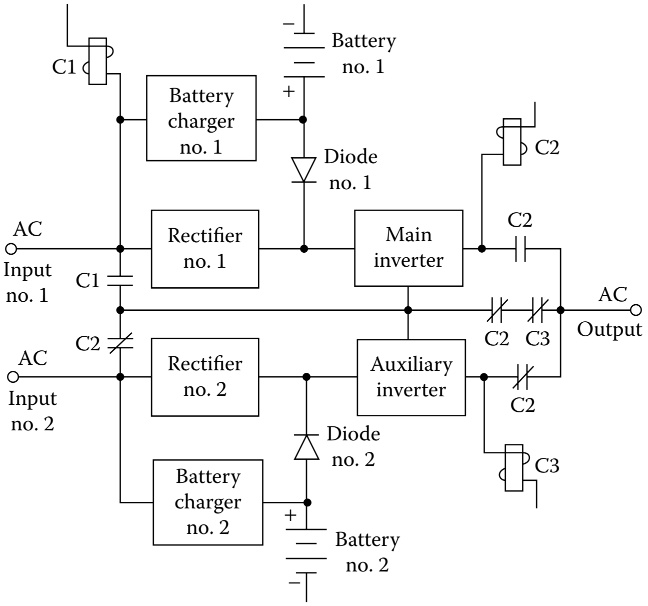


FIG. 4.23m
Threefold redundancy.

handling the full output capacity (both loads). On failure of either, the remaining unit assumes the full load. Momentary paralleling is possible to use the stored energy in the filter of the unit going off. As noted previously, this energy can help to minimize the transfer transient.

Complete input and output redundancy are diagrammed in Figure 4.23m. Battery, battery charger, rectifier, and inverters are repeated. Each is sized to handle the full load. Further redundancy is provided by the AC line. If the main set fails, the auxiliary set supplies the load. If it, in turn, fails before the main is repaired, the load is supplied by the line.

Common Bus Branch (Load) Failure

In addition to providing a means of transfer from the output of an inverter that has failed to an alternate power source, the transfer switch provides a means of clearing branch circuit fuses sufficiently fast to protect other loads from a faulted load. The inverter is frequently current-limited to provide a finite overload capability. Thus, on failure of one load, the current limit provides a known amount of current for opening the fuse in the faulted branch.

By referring to available fuse characteristics, one derives the information that a number of branches are required with even the fastest of available fuses in order to provide load clearing within one half cycle. By using a transfer switch, which allows the transfer of the load from the inverter output to the AC line, one can, in effect, increase the short-circuit current capability, thus providing a means of rapidly clearing the fuse in the branch circuit. After clearing the faulted branch, one can retransfer back to the inverter standby systems.

Ideally, a short circuit in one fused branch of an n-branch load, such as in Figure 4.23n, would have no effect upon the power supplied to the remaining branches. In reality, however, a fuse requires a finite amount of time to clear. If the power source is current-limited, the supply voltage will drop to a value near zero until the fault is cleared. Unless rapid opening of the fuse occurs, other loads will become inoperable.

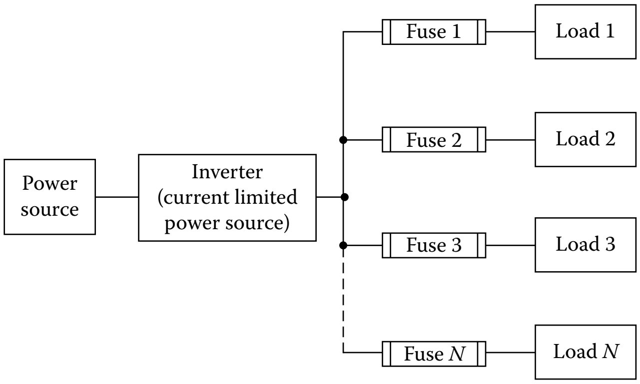


FIG. 4.23n
Inverter system, which is supplying N fused load branches.

Assuming, for any given supply output capacity, that all branches of an n-branch load consume nearly equal parts of the supply power, it follows that the larger the number of load branches, the smaller the average branch fuse rating. A 10 KVA power source, for example, may have five equally loaded (2 KVA) branches. Another 10 KVA supply may have ten equally loaded (1 KVA) branches. In the latter case, the average branch fuse has approximately one half the current rating of those in the former case.

In general, the larger the current capacity of a power source, the shorter the time required to clear a fuse in a short-circuited branch. Looking at the same relationship in a slightly different way, we can see that the smaller the fuse rating of a short-circuited branch, the less time required to clear it—assuming the short-circuited supply capacity remains constant.

The preceding two paragraphs lead to the generalization that the smaller the fraction of total supply capacity carried by a fused load branch, the smaller the time required to clear a branch fuse and the less serious the load power disturbance. Conversely, the larger the fraction of total capacity carried by each branch, the larger the fuses in each and the longer the time required to clear the fault.

Figure 4.23o demonstrates this generalization. Assuming that (1) all branches consume equal fractions of load power, (2) the source always current-limits at 15 amperes, and (3) the short-circuit load impedance is zero, fuse clearing time is shown as a function of the number of load branches. The graph illustrates these data for three different fuse "speeds."

A well-designed standby power supply system requires that consideration be given to all types of failure. The most often neglected area of consideration is the load bus system. Selection of the type of branch circuit protector must be coordinated with the short-circuit characteristic of the inverter as well as the requirements of the loads.

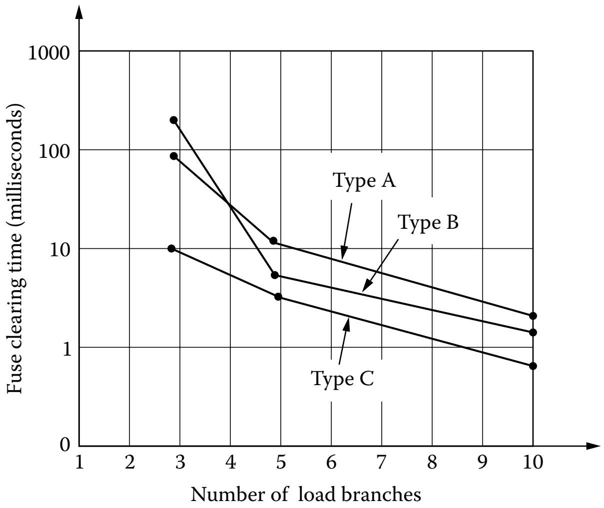


FIG. 4.23o
Fuse clearing time expressed as a function of the number of load branches.

SYSTEM COMPONENTS

It is frequently true that the characterization of the system itself is dependent on the components available for use in the system. For each system function, there are a number of components from which to choose. Arbitrary selection of any single component without regard to the others can result in an unworkable or, at best, an inefficient system.

Thus, the designer's problem is to choose a compatible set of components that will satisfy the requirements. Specifically, the designer seeks the optimum compatible set of components. In order to select the appropriate component for a given function, it is necessary to understand the characteristics of the components from which one must select.

Rotating Equipment

Rotating equipment may be subdivided into two general classes. The first class includes all devices operating from a source of electric power. This set includes motors and, because of their intimate relationship, alternators and generators. The other general category of rotating equipment includes those devices driven by engines that have the ability to operate from liquid or gaseous fuels.

The preceding discussion has not covered the nature of the equipment in the blocks. As an example, the battery charger could be a motor generator set with appropriate controls. The inverter could, of course, be a DC motor driving an alternator. The selection of these components is determined by economic considerations. The economics involve not only the initial cost, operating cost, and maintenance cost of the equipment itself but also an evaluation of the need for reliability based on the importance of the load.

If load failure results in a vacant lot characterized by a hole or by the need to repipe a plant because of the solidification of material that under normal circumstances would have been a usable product, the greatest possible reliability should be built into the standby power source. On the other hand, if the loss of power results in some annoyance but not in the loss of plant capacity or deterioration of product quality, then the ultimate in redundancy and reliability is not warranted.

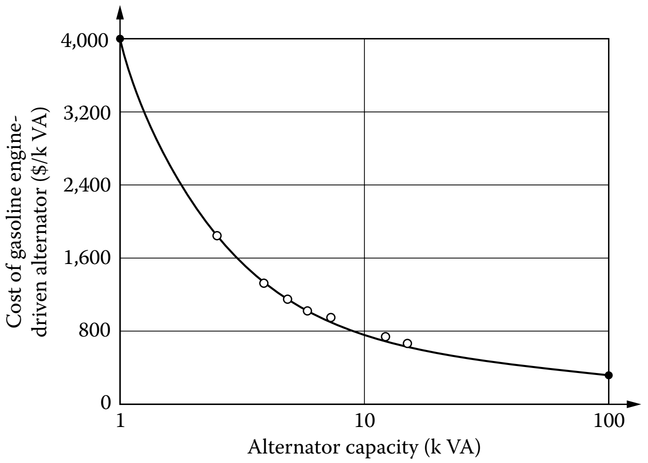


FIG. 4.23p
Cost of gasoline engine-driven alternators as a function of their capacity.

The costs of some standby power supply systems using engines are shown in Figure 4.23p. The engine has been proved to be a reliable device in many situations and in very adverse environments. Unfortunately, the engine itself is not the most frequent cause of failure. Most complaints of poor reliability of engine-driven sets can be traced to unreliable but necessary peripheral equipment, such as fuel pumps, cooling systems, and so forth. Care should therefore be exercised in the specification of all the engine system components.

Two typical standby power supply systems involving motors, generators, and alternators are shown in Figures 4.23q and 4.23r. In Figure 4.23q, the AC line provides power until it fails. On failure, the battery supplies power to the alternator, which, in turn, supplies the output voltage.

In Figure 4.23r, the AC motor has been replaced by a rectifier of the static variety. Now referring to Figure 4.23r and to Figure 4.23g, it is easy to see that the two are identical, with the motor–alternator set replacing the block marked "inverter." A system involving an engine is shown in Figure 4.23s.

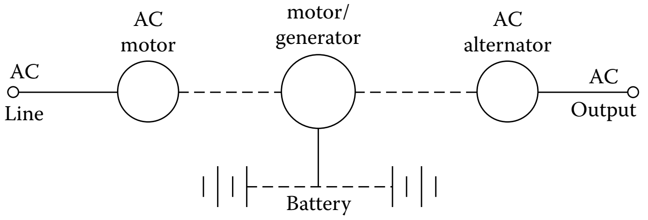


FIG. 4.23q
Standby power supply system consisting of AC motor, generator, and alternator.

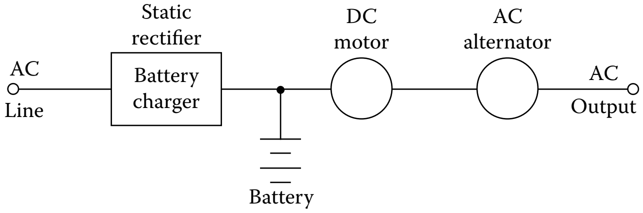


FIG. 4.23r
Standby power supply system utilizing a static rectifier.

Batteries

Three types of batteries are in general use for standby systems: lead antimony, lead calcium, and nickel cadmium. For their approximate costs refer to Figure 4.23t.

Lead Batteries The lead antimony and lead calcium are lead acid batteries deriving their name from the hardening material in the lead alloy. Both have approximately the same ampere-hour characteristics on discharge. The lead antimony construction costs less, requires more maintenance because of its higher internal losses, and evolves more hydrogen than the lead calcium. Life expectancies of 14 to 30 years are frequently quoted. The life depends on the construction of the plate and the plate thickness. It also depends in large measure on the care given the batteries in service.

Both types can operate over a temperature range of –10 to 110°F (–23 to 43°C). The lead calcium cell may be floated at 2.25 volts per cell without the necessity for equalizing charge or, at worst, with long periods between equalizations.

Nickel Cadmium Batteries Nickel cadmium batteries are the alkaline type. They differ from lead acid batteries in that they have a larger short-time current capability, higher cost, and lower volts per cell. Little hydrogen is generated by this cell, and frequent overcharge is recommended. Life expectancy and operating temperature range are similar to those of the lead acid types.

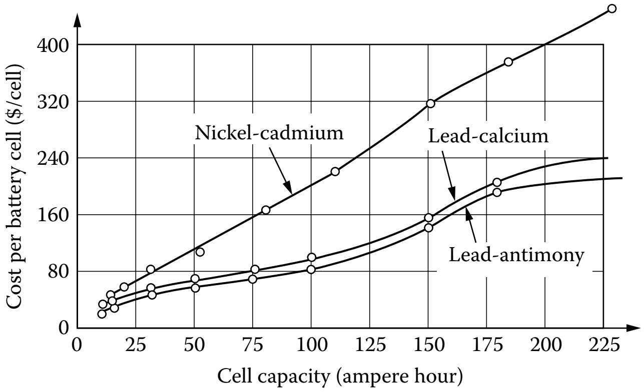


FIG. 4.23t
Battery cell unit costs as a function of battery type and capacity.

Battery Chargers The battery chargers generally used for standby systems are "float" chargers. Their cost is a function of the rated charging current and the float voltage as shown in Figure 4.23u. They are characterized by relatively constant output voltage to recharge the battery because their output current varies from almost zero to rated value. Beyond their rated current, output voltage drops rapidly with increased load current. This current limit protects the charger when applied to the battery in the discharge state.

Satisfactory battery life is dependent on the design and operation of the battery charger, and so are maintenance costs. The feedback techniques used to maintain the constant output voltage and current limit of battery chargers are well known and will not be discussed in detail. The rectifiers themselves may take the form of either a polycrystalline cell, such as selenium or copper oxide, or a monocrystalline cell, such as germanium or silicon. The trend is toward the silicon rectifier. Control devices include the magnetic amplifiers and thyristors.

Motor-generator battery chargers are also available for standby system recharging service. The use of motor generators in this application predates that of the drive-type rectifiers.

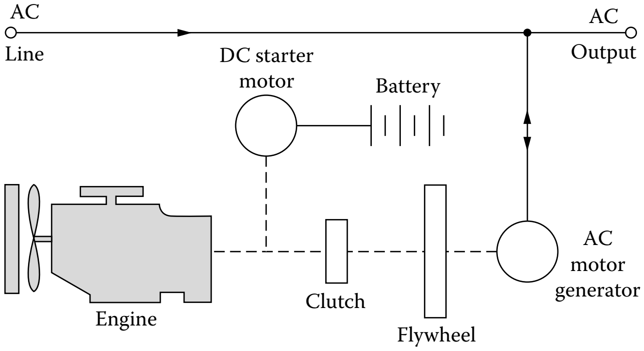


FIG. 4.23s
Standby power supply system using an internal combustion engine as the source of the power.

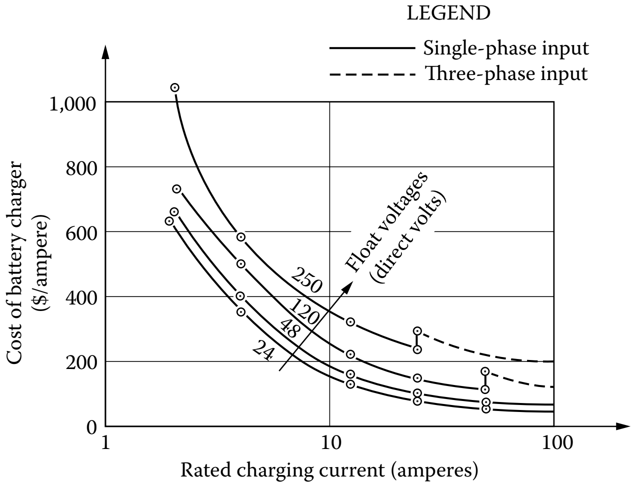


FIG. 4.23u
Battery charger costs as a function of their current and voltage ratings.

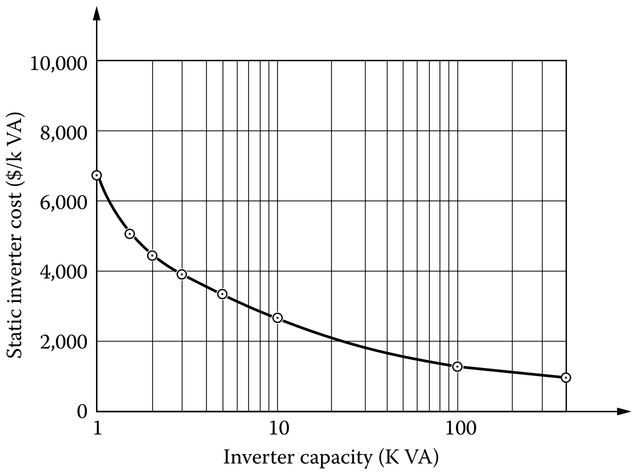


FIG. 4.23v
Static inverter costs as a function of its capacity.

Static Inverters

The static inverter used in a standby system tends to be the most complex piece of equipment in that system. The approximate costs of static inverters are given in Figure 4.23v, and a brief discussion of their operation is provided here.

Transistor inverters are the least expensive of the static inverters. Their principal area of usefulness is the low input voltage (24 direct volts and less) and low output capacity (500 volt-amperes and less) range. This type of inverter can operate at high frequency and can cease operation under dangerously high output current overloads. Figure 4.23w shows a typical circuit for the center-tap transistor inverter; N_1 and N_2 are feedback windings, R_1 is the feedback resistor, and R_2 is the starting resistor.

The operating cycle can be traced by assuming Q_1 closed and Q_2 open. Substantially all the supply voltage E appears across N_3 causing a change in flux level by Faraday's law:

$$\Delta\phi = \frac{10^8 Et}{N_3} \qquad \textbf{4.23(1)}$$

where

$\Delta\phi$ = change in flux level
N = turns in N_3 and N_4 coil (identical)
E = supply voltage (appearing across N_3)
t = time

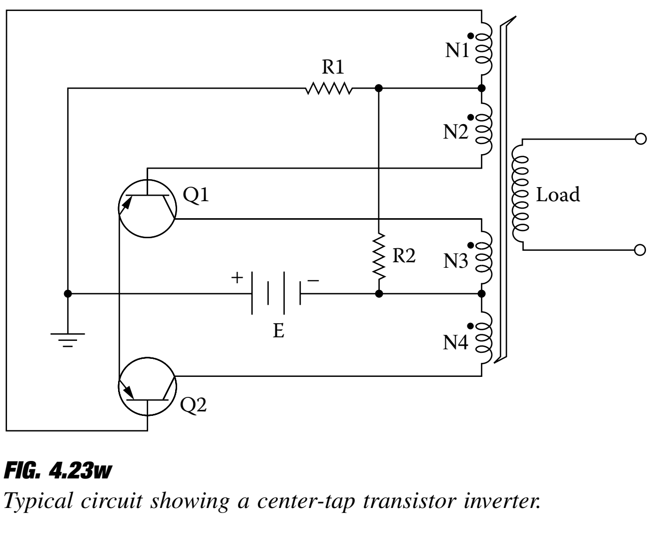


FIG. 4.23w
Typical circuit showing a center-tap transistor inverter.

Eventually the core saturates, requiring an increase in exciting current. To supply this increased collector current, the transistor must have an increased base current. This cannot be supplied because of the decreased coupling between N_2 and N_3 resulting from core saturation; thus, Q_1 begins to open, reducing exciting current. At this point, the change of flux reverses, reversing the coil voltage polarity. Q_2 is then turned on by N_1 and Q_1 is turned off by N_2. The half cycle begun in this fashion is similar to that described until a reversal occurs again, completing the cycle.

Starting resistor R_2 provides enough base bias current to allow exciting current to flow. Natural circuit imbalance ensures that only one transistor closes, thus starting the oscillator.

Should the load be short-circuited, no feedback is provided by N_1 and N_2, and oscillations cease. Other modes of operation are also possible. Care must be taken to provide the correct base current (which is dependent on transistor current gain) so that loading does not excessively shift the frequency. The amount of this shift is also dependent on the "rounding" of the B-H loop.

In the center-tap inverter circuit, each transistor must withstand a voltage equal to or greater than twice the supply potential. Transistors having a sufficiently high rating to withstand a 48-volt source are more expensive. Usually, above 24 volts it is less expensive to use the bridge circuit shown in Figure 4.23x.

The bridge circuit operates in a manner similar to that described for the center-tap circuit. Two starting resistors, R_{s1} and R_{s2}, are necessary. The diodes CR 1, 2, 3, 4 provide transient voltage suppression for unsymmetrically wound transformers.

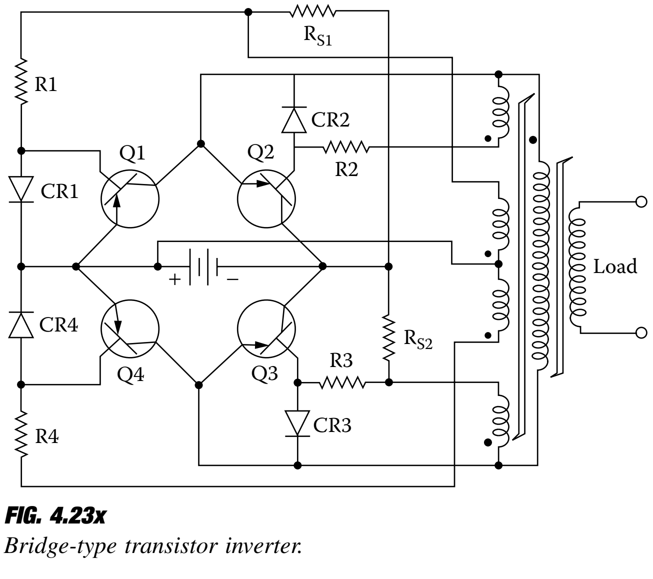


FIG. 4.23x
Bridge-type transistor inverter.

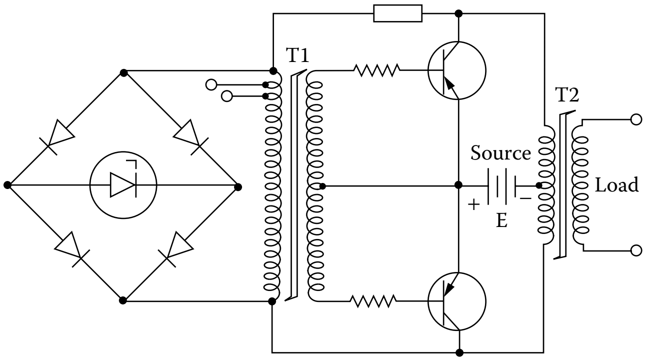


FIG. 4.23y
Typical means of stabilizing inverter frequency.

A typical means of stabilizing frequency is shown in Figure 4.23y. Since the saturation flux density and turns are relatively constant, frequency is controlled by the supply voltage *E*. Use of a "zener" diode stabilizes the voltage and therefore, frequency.

Silicon-controlled rectifier-inverters are the "work horse" of static inverters. They operate efficiently and reliably at high input voltage (130 to 600 direct volts) and high output capacities (500 volt-amperes and larger). Proper specification of the equipment is essential to obtaining reliable operation.

The operation of a static inverter may be stimulated by switches as shown, in Figure 4.23z. Switches 1 and 1′ are operated in unison, as are switches 2 and 2′. When 1 and 1′ are closed and 2 and 2′ are open, load current flows in a direction shown by the arrow in Figure 4.23z. With 2 and 2′ closed and 1 and 1′ open, load current flows in the reverse direction. Thus, whereas the source current i_s flows in the same direction when either set of switches is closed, the load current i_l reverses polarity as each set is alternately closed and opened. Consequently, the current has been inverted by the circuit.

The switches used in Fig. 4.23z are not static since a switch contains moving parts. In Figure 4.23aa, the mechanical switches have been replaced with electrical switches (silicon-controlled rectifiers). Also shown are the commutating indicators L and communicating capacitor C. These components are necessary to turn the controlled rectifiers off. Turn-on is accomplished by the application of the voltage to the gate leads of the controlled rectifiers by the oscillator.

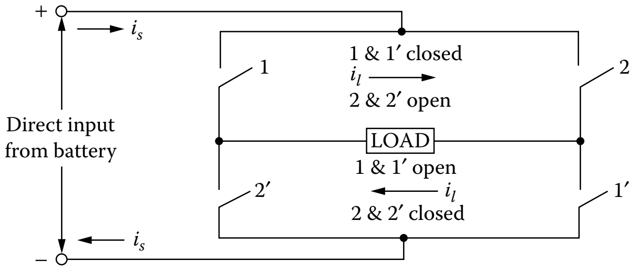


FIG. 4.23z
Simulated static inverter using mechanical switches.

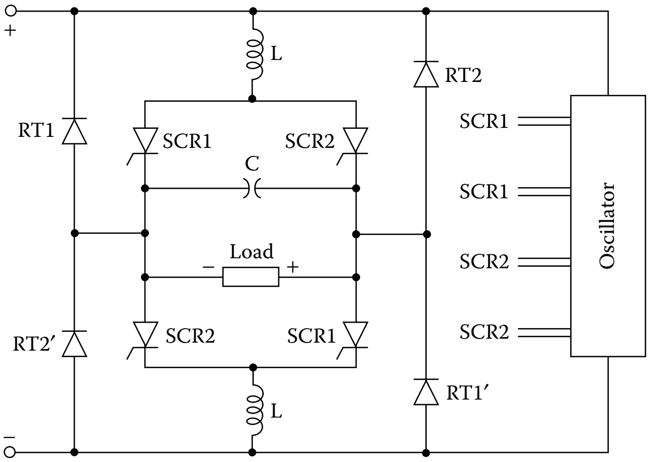


FIG. 4.23aa
Bridge-type SCR static inverter.

Rectifiers diodes RT_1, RT_1, RT_2, and RT_2 are *not* a part of the basic inverter switching circuit. They serve to clamp the amplitude of the load voltage to a value approximately equal to the magnitude of source voltage.

Figure 4.23aa shows the diagram of a bridge-connected static inverter. This arrangement is frequently used for source voltages of 130, 260, and 600 volts. For source potentials of 12, 24, and 48 volts, the circuit of Figure 4.23bb is frequently used. Its operation is seen to be similar to that of the bridge circuit. It differs in that half the number of controlled rectifiers is used, and each must hold off a voltage approximately equal to twice the supply voltage.

Various types of output waveforms may be obtained from the square wave, which is the basic output waveform of the static inverter. Sinusoidal waveforms are most common, but triangular, sawtooth, and many rectangular combinations are also possible. Voltage stabilization may be a welcome bonus provided by the output wave-shaping circuitry. Current limiting is also possible.

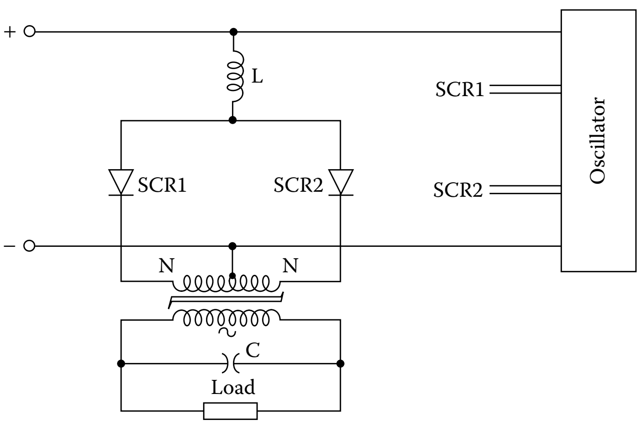


FIG. 4.23bb
Center-tap-type SCR static inverter.

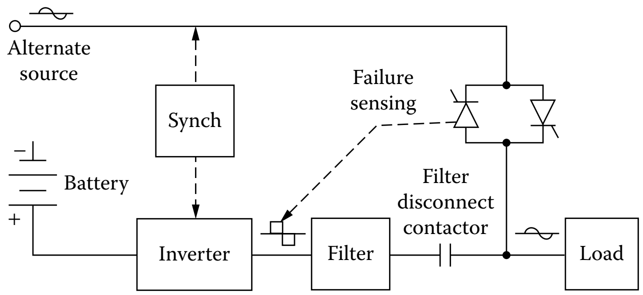


FIG. 4.23cc
No-break power system with hybrid transfer switch.

Bus Transfer Switches

Bus transfer switches have historically been of the electromechanical type. Various techniques have been used to speed the transfer from one source to the other. These techniques have included a pulsing arrangement on the coil of the electromechanical switch in order to overcome the inherent inertia. "Make-before-break" sequences have also been used.

A newer development has been the use of the static switch for more rapid transfers. Generally, thyristors have been used as the power-handling switching component. Figure 4.23cc represents a hybrid transfer switch on the output side of the inverter. Momentary paralleling of sources is achieved by the "drop out" time of the electromechanical contactor. Fast turn-on is achieved by the thyristors. Sensing is performed at the output of the inverter switch (prior to the filter) in order to anticipate output failure.

A static set that serves the same function is shown in Figure 4.23dd. In this case, momentary paralleling is achieved by the logic circuit that supplies the gating pluses.

Protective Components

It is essential to the proper operation of the emergency power supply system that adequate thought be given to the protective devices. Available as protective devices are fuses of various speeds and circuit breakers. The application of fuses to the load circuit was discussed earlier in this section.

Although it is not possible to provide a clear definition of the components to be used for the specific applications within a system, it is necessary to urge that one pay careful attention to the selection. In-rush current on start-up, transient variation of the input voltage, and transient load variation each present particular problems. A frequent experience is that the protective system may be the least reliable of the system components. This is not intended to mean that the component itself fails, but rather that the system fails through action of the protective device under normal operating conditions. Whenever the standby power supply system fails to provide output, whether the components are damaged or not, the resulting damage is the same.

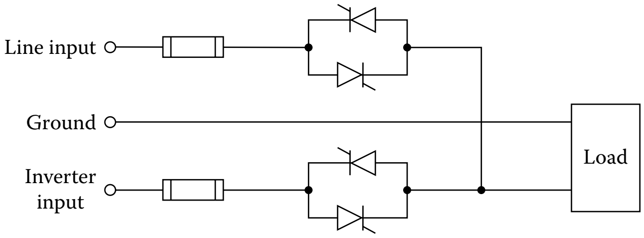


FIG. 4.23dd
Static no-break power system.

STANDBY POWER SUPPLY SYSTEMS

A number of standby power supply systems have been presented in the previous discussion. These have been categorized in terms of the type of power failure but not in terms of the class of system they represent. It was noted earlier that the systems themselves could be classified by virtue of the time it takes to transfer from the primary to the secondary power source. This grouping is more nearly akin to the thinking of the purchaser than is the classification by power failure noted previously.

Multicycle Transfer System

In general, multicycle transfer systems involve either electromechanical transfer switches or engine- or motor-driven equipment that must start up. Figures 4.23ee and 4.23ff show two composite systems using both rotating and static equipment. In Figure 4.23ee both a battery input and an engine-driven generating input are provided to the inverter.

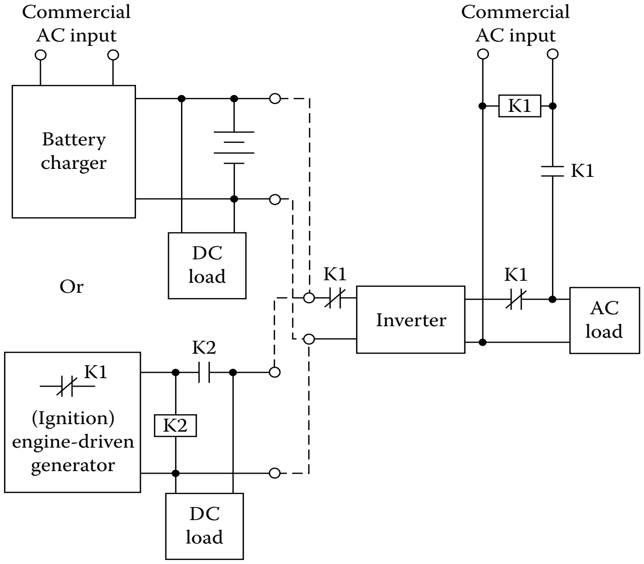


FIG. 4.23ee
Composite multicycle transfer system using both batteries and an engine-driven generator.

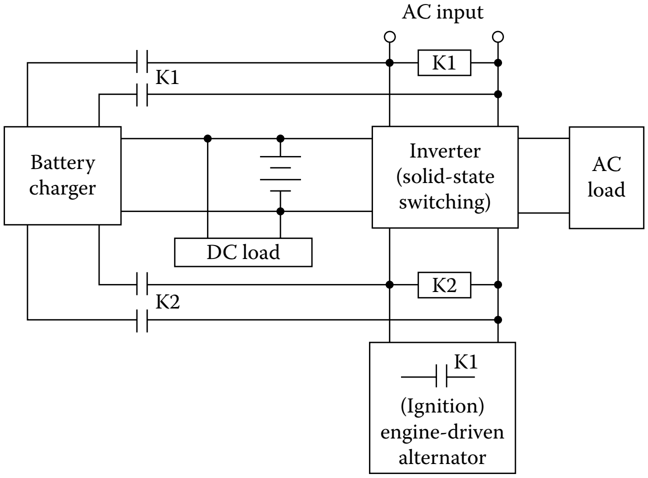


FIG. 4.23ff
Alternative arrangement of a multi-cycle transfer system.

If the engine-driven generator is started prior to the time the energy contained in the battery is completely used by the inverter and the load, *no* interrupt time occurs. Note that the commercial AC power input provides the load normally. On failure of the power lines, the contactor K_1 operates to insert the inverter. Since the inverter is started on the fallout of the contactor K_1, some start-up transient must occur in the inverter, requiring some time between the failure of the commercial AC input source and the inverter source. Note that the use of a static switch in position K_1 would increase the speed of operation of the electromechanical contactor but would *not* materially affect the start-up of the inverter.

Note that a rearrangement of the components of Figure 4.23ee results in the system shown in Figure 4.23ff. This system may not have a transfer time since the battery would normally be chosen to have a capacity sufficient to cover the energy required by the inverter and load during the time it takes to start the engine-driven alternator. Should that interval be long, multicycle startup will result. Note, however, that in this system there is no protection for inverter failure. The system of Figure 4.23ff does have a bypass to the AC commercial line, providing some protection in the event of component failure in the system.

Sub-Cycle Transfer System

Sub-cycle transfer systems are generated by the use of static switches on the output side of the inverter. Thus, the circuits of Figures 4.23j, k, l, m, dd and ee may be sub-cycle transfer systems depending entirely on the arrangement of the switch itself and, in the case of Figure 4.23ee, the start-up time of the inverter.

No-Break Transfer System

The so-called no-break transfer system may be no-break in the sense that if the AC commercial source fails, no cessation in output power results. No-break may also be applied to those systems having a redundant source on the output side of the inverter. Figures 4.23c, g, h, i, and 4.23ff are examples of no-break systems with redundant input sources so that no break occurs in the output power should the AC commercial source fail.

The switching systems shown in Figures 4.23cc and dd can be used in many of the previously defined circuits to provide no-break switching under the right sequence of operations. Figures 4.23j, k, l, and m are of this type.

SYSTEM REDUNDANCY

The subject of system redundancy has been frequently mentioned in the previous discussions. Two basic classifications of system redundancy are made on the basis of input and output. Redundancy in the input circuit to the inverter results from the use of multiple battery sources, multiple input lines coming from separate power feeders, or various types of rotating alternators and generators with either motor or engine drives. Figures 4.23c, h and i illustrate the various types of input redundancy.

Redundancy in the output circuit of the inverter is obtained by use of a switch that will provide a path to an alternate source. Output redundancy may involve switching from the inverter output to the power line. More complex schemes involve switching from one inverter standby system to another inverter standby system. Figures 4.23j, k, l, and m provide examples of output redundancy.

The previous examples have shown the wide number of choices available for component redundancy in standby power supply systems. In order to select the best system for a given application, it is necessary first to evaluate the degree of integrity required for the application. It is then essential that the standby system be evaluated to determine which component is most likely to fail. A decision can then be made, balancing the cost of redundancy versus the needs of the application.

SPECIFICATIONS

Considerable activity is underway to attempt to provide specifications for standby power supply systems and system components. It is essential that the user provide in the specifications certain types of information that are important in ensuring system reliability.

It is suggested that consideration be given not only to the immediate load requirements but also to future load requirements. It is generally less expensive to purchase additional capacity when the system is first acquired than to add capacity to that system at a later date.

In many cases, the characteristics of the loads are most important. Power factor as well as transient characteristics should be clearly defined. These definitions are particularly important if static equipment is involved.

Transient data on the input sources are important in proper design of the system. If a battery source is used, it is desirable to know the transients that can exist on the battery bus. These transients should be specified in terms of their maximum voltage as well as their energy content. If an existing battery installation is used, any loads that are switched will generally institute a transient voltage because of the inductance of the lines themselves. A knowledge of this transient voltage is particularly important in the proper design of static equipment.

Although it is difficult to obtain any meaningful data regarding the number and duration of outages of the utility power lines, it is necessary to provide information concerning the length of time the power supply must produce power for the loads without having the primary input source available. This evaluation can best be performed on the time required to "shut down" the load system rather than with reference to the input failure.

Particular note should be taken of the characteristics of static inverters. Three overload ratings are important. In order that sufficient commutating capacity can be designed into the inverter, it is necessary to know the maximum instantaneous current. It is also necessary to know the overload current for a one- to two- minute interval in order that ample cooling be provided to the semiconductor devices. The final overload rating of importance is the one- or -two hour overload necessary in order to provide ample thermal capacity in the magnetic components. This third overload rating is also of importance in defining rotating equipment.

In the event that the standby system includes a means of transferring the load of the inverter output to an alternate source, the characteristics during transfer and retransfer should be amply defined. Among those characteristics is the length of time, during which switching between the two sources, that zero voltage can be tolerated. Phase shift in voltage from one source to the other should also be stated, together with the transient voltage characteristics on transfer or retransfer. These characteristics can be defined by an evaluation of the sensitivity of the loads to varying phase angle, frequency, and transient voltage. If dynamic loads are included, the transient response and time constant of these loads should be stated.

Bibliography

Conger, N. L., "How to Install and Maintain UPS Systems," *Instrumentation Technology,* April 1970.

Dyer, S. A., *Survey of Instrumentation and Measurement,* New York: John Wiley & Sons, 2001.

Graf, R. F, *Measuring Circuits,* Oxford: Butterworth-Heinemann, 1997.

"Guide to Selecting Uninterruptible Power Systems," *Instruments and Control Systems,* March 1973.

Horowitz, P., and Hill, W., "The Art of Electronics," New York: Cambridge University Press, 1980.

Jazinski, G., "Uninterruptible Power Systems for Industrial Computers," 1991 ISA Conference, Anaheim, October 1991.

Johnson, W. J., "Use of Uninterruptible Power Supplies in Chemical Plants," ISA Conference Paper #76-835, Instrument Society of America, Houston, Texas, October 1976.

"Key Items for UPS Selection," *Instruments and Control Systems,* October 1977.

Lapuh, R., "Simple Power Measurement System," Proc. of the 18th IEEE 1035, Instrumentation and Measurement Technology Conference, IMTC 2001, Vol. 2, 2001.

McPartland, J. F., and McPartland, B. J., *National Electric Code Handbook,* New York: McGraw-Hill, 1999.

"National Electrical Code 1999" (NFPA 70) Quincy, MA: NFPA, 1998.

Northrop, R. B., *Introduction to Instrumentation and Measurements,* Boca Raton, FL: CRC Press, 1997.

Teets, R., "Don't Let Power Failures Plague Your System," *Instruments and Control Systems,* January 1977.

Warren, E. G., "Uninterruptible Power Systems in Refineries," *Instrumentation Technology,* March 1977.

Waterbury, R. C., "UPS Protects Electric Utility," *InTech,* September 1992.

Witte, R. A., *Electronic Test Instruments: Analog and Digital Measurements*, 2nd ed., Englewood Cliffs, NJ: Prentice Hall, 2002.

Yuen, M. H., "How to Specify and Select UPS Systems," *Instrumentation Technology,* April 1970.

4.24 Workstation Designs

G. B. SINGH (2002) **B. G. LIPTÁK** (2005)

Cost:	$7,500 to $10,000 for a complete workstation server without any software loaded.
Partial List of Suppliers:	ABB (www.abb.com) Advanced Systems & Designs Inc. (www.spcanywhere.com) Allen-Bradley Co. (www.ab.com) Aydin Displays Inc. (www.aydindispalys.com) Cuttler-Hammer (www.cuttlerhammer.eaton.com) Emerson Process Management (www.emersonprocess.com) Fisher Controls International Inc. (www.fisher.com) Foxboro-Invensys (www.foxboro.com) GE Fanuc Automation (www. gefanuc.com) Honeywell Automation and Control (www.honeywell.com/acs) ImageVision Inc. (www.imagevisioninc.com) Kessler Ellis Products (www.kep.com) Nematron (www.nematron.com) Rockwell Automation (www.ab.com/sensors) Ronan Engineering (www.ronan.com) Rosemount Inc. (www.rosemount.com) Siemens Energy & Automation (www.sea.siemens.com) Systel (www.systelusa.com) Toshiba International Corp. (www.tic.toshiba.com) Westinghouse Process Control (www. westinghouse.com) Xycom Automation Inc. (www.xycom.com) Yaskawa Electric America (www.yaskawa.com) Yokogawa (www.yokogawa.com/us)

CLASSIFICATION OF WORKSTATIONS

Operator stations can be divided into various categories, based on the hardware architecture and on the functions they perform.

Hardware Architecture

Those workstations classified based on the hardware architecture can be further divided into two categories: (1) diskless workstations or dumb terminals and (2) intelligent or stand-alone workstations.

Diskless Workstations This type of workstation does not have a hard disk. Its major task is to act as a dumb interface between the application program and the end user. This type of workstation was popular in the days of mainframe computers, where the main processor, which had enormous processing power, was centrally located, and the terminals were used as an interface for the end users.

Diskless workstations are defined by *webopedia* as: A workstation or PC on a local-area network (LAN) that does not have its own disk. Instead, it stores files on a network file server. Diskless workstations can reduce the overall cost of a LAN because one large-capacity disk drive is usually less expensive than several low-capacity drives. In addition, diskless workstations can simplify backups and security because all files are in one place—on the file server. Also, accessing data from a large, remote file server is often faster than accessing data from a small, local storage device. One disadvantage of diskless workstations, however, is that they are useless if the network fails (Figure 4.24a).

Intelligent Workstations Intelligent or stand-alone workstations are fully equipped with hard disks and application software (see Figure 4.6m). These operator stations may also have databases and may perform various critical operation and control functions. The advantage is that intelligence is distributed across the network, and each workstation can do its piece of a task, which helps reduce processing overhead

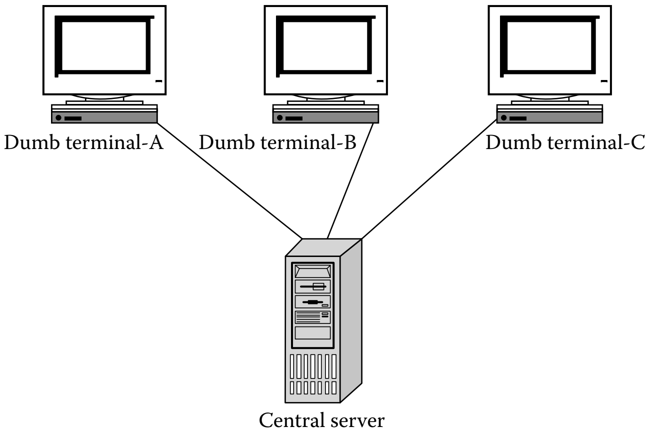


Fig. 4.24a
Diskless workstations.

on a single computer. Also, if one of the operator stations is down, only the functionality related to that workstation is affected and all other tasks remain unaffected. The disadvantage of this type of workstation is that it requires additional hardware and software. The licenses of each software (application or database) have to be purchased for each operator station wherever the particular software is loaded. Certain processes become redundant and are duplicated over various workstations. Client/server architecture is based on this philosophy (Figure 4.24b).

Function

The second category of classification is based on the functions workstations perform; in this grouping, workstations are known by the task to which they are primarily assigned:

- *Operator Workstations*: These workstations are used solely to operate, monitor, and control plant parameters.
- *Engineering Workstations*: The main task of engineering workstations is to perform engineering functions: creating new loops; adding various input and output points, modifying sequential and continuous control logic, simulating the logic offline, database engineering, and preparing such engineering documentation as input/output (I/O) lists and device summaries, etc.
- *Historian Workstations*: These workstations are specifically reserved for performing plantwide acquisition, storage, and retrieval of historical process and system information such as alarms, events, operator actions, and system diagnostics.
- *Application Workstations*: A hybrid of all the above-mentioned functions is performed at a single station, known as an application workstation. The workstation may be capable of database management, historical data management, plant and process control, third-party interface functions, etc. The application workstation performs operations such as online configuration and provides a broad range of configuration capabilities, including database, display, system definition, control strategy configuration, and system status monitoring.
- *X-Terminals/Application Processors/Gateways*: These workstations are dedicated servers. They are preassigned to perform a predefined task and normally they act as a black box for the users. Their wide functionality may include conversion of protocols to enable connection of various types of networks (token ring, token bus, Ethernet, etc.), DDE (dynamic data exchange), and OLE (object linking and embedding) servers, and X-terminals.
- *Gateways*: A gateway is a special host that interfaces with two or more distinct networks, to provide for extension of the address cover of each. Some gateway hosts provide protocol translation between the nets they connect. There are two fundamental types of gateways: a gateway that has the capability of performing routing functions when interconnected networks have a consistent message format, and a gateway that must translate services between interconnected networks when hosts use dissimilar message formats. In this case, the internetwork can only support services that can be mapped between the two dissimilar networks. End-to-end flow control is lost with this type of gateway.
- *Portable Workstations*: These are mobile workstations with a special wireless network interface card, which allows operation of the plant from anywhere in the plant (based on the distribution of the antennas). In contrast to the fixed type of workstation, these stations provide the convenience of remote access to ease operation and maintenance. This type of workstation is very useful in plant start-ups and troubleshooting. However, portable workstations are still not very popular in industrial mission-critical applications.

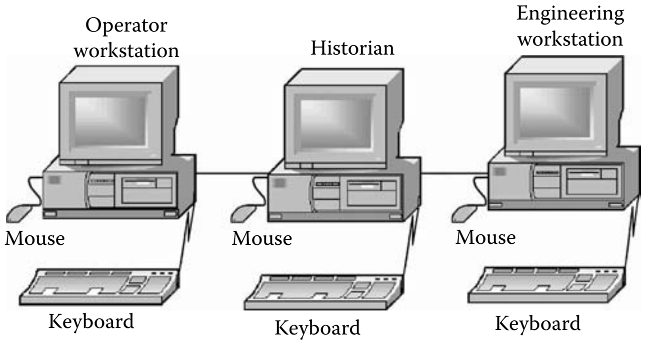


FIG. 4.24b
Intelligent workstations.

Photographs of diskless workstation-based DCS and intelligent workstation-based DCS are shown in Figures 4.24c and 4.24d, respectively.

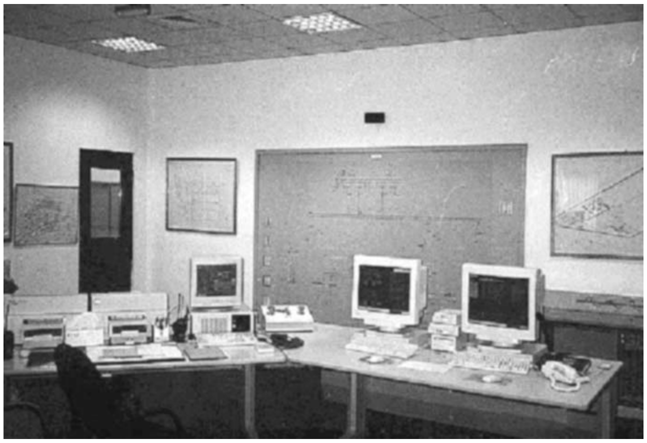


FIG. 4.24c
A diskless workstation-based DCS.

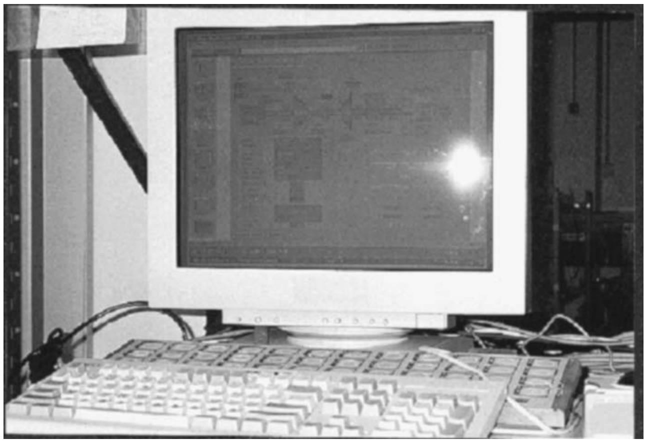


FIG. 4.24e
A workstation with an alarm annuciator panel and keyboard.

HARDWARE COMPONENTS

Any workstation may have the following components:

- Central processing unit (CPU)
- Workstation display (with/without touchscreen)
- Alphanumeric keyboard
- Annunciator and annunciator/numeric keyboards (Figure 4.24e)
- Mouse
- Trackball
- Permanent storage device (e.g., a hard disk)
- Temporary storage device (e.g., a floppy drive, CD-ROM drive, or tape drive)
- RAID (redundant array of inexpensive disks) for backups
- RAM

Several of these hardware components affect the performance of any workstation:

FIG. 4.24d
An intelligent workstation-based DCS.

CPU: The central processing unit is the main workhorse of the workstation. The higher the number of bits, the higher the processing power. RISC processors and 64-bit processors are popular in workstation applications.

RAM: This is a very critical component of any operator station. Because all operator stations will have a lot of graphic capabilities and software, it is critical that the size of the RAM be adequate to achieve a faster refresh rate for the graphics. Generally, the accepted refresh rate for switching between one graphic to another is 1 s or less, and apart from various software methods that can be used to reduce this time, the size of the RAM plays a very important role.

Monitor: The display resolution of the monitor is very important for the aesthetics of a workstation. CRTs (cathode ray tubes) have been the mainstay of computer technology. CRT monitors are measured diagonally from the outer corners of the glass tube as in 14, 17, 19, 20, or 21 in., but the viewable image size on a CRT is smaller because the corners of the glass have a radius, so they are not available as viewable area. Active matrix TFT (thin-film transistor) LCDs (liquid crystal displays) or flat-panel LCD monitors are popular in applications where CRTs have traditionally been used, thanks to their improved color and brightness, better appearance, higher resolution, lesser power consumption, and longer life. Stray magnetic fields generated by motors, high-power machinery, welders, buses, and transformers are common on the plant floor and can result in screen noise and degraded color purity, or can cause severe geometric distortions in CRT images. Flat-panel monitors are not affected by magnetic fields, and hence TFT screens are becoming more and more popular. Figure 4.24f shows a dual-screen workstation.

The RAM size of the video card also plays a crucial role. The optimum size must be selected based on the graphics and their orientation.

Workstations may also consist of special utilities, such as touchscreen utilities, speech synthesizer, etc. They must be correctly selected for the application, and additional software utilities must be matched to these applications.

FIG. 4.24f
Typical dual-screen workstation. (Courtesy of Foxboro Systems.)

SOFTWARE FEATURES

Most of the time the operator station acts as a gateway between the user and the plant. Thus, it must be robust enough to sustain crude handling by operators and be failproof to avoid catastrophe due to errors committed unknowingly.

There are a number of standard applications on any operator station:

- Provision of equipment information for the workstation and its associated I/O devices, buses, and printers
- Capability for change actions directed to the associated equipment
- Processing of station alarm conditions and messages and maintenance of the system date and time
- Database management, storage, retrieval, and manipulation of system data files
- Maintenance of a history of values for process-related measurements, which have been configured for retention by the historian
- Subsetting of the overall operating environment according to the type of user process engineers, process operators, and software engineers who have access to specialized functions and databases suited to their specific requirements and authorizations
- Dynamic and interactive process graphics

Following are the key features available in a Windows NT-based/Windows 2000/X-Windows/UNIX-based workstation:

- Operating system: The operating system (NT/UNIX/Linux/OS2, etc.) is the basic platform of operation of the workstation over which all other software is ported.
- System software: System management function software (data acquisition, control, and management) performs the real-time data collection, management, and storage of the process control messages and data, database management software, and its interface with the historian software.
- Graphical user interface: The graphical window interface to software applications, which is also known as display manager software, determines the user-friendliness and the aesthetics of the workstation.
- Application programming interface (API): API, used for interface with third-party devices, defines the capability of performing control and data acquisition functions for direct connection to a variety of I/O devices.
- Database management system: Database management systems such as DB/2, Informix, Access, Oracle, etc. are the workhorses where the raw data are stored. These database management systems provide front-end software such as SQL, Access, Developer 2000, and Visual Basic for manipulating the data.

Optional packages may include:

Historian package for viewing the history of various parameters
Sequence of event recording function
Trend display and troubleshooting functions
Special engineering software such as performance monitoring systems, efficiency calculators, maintenance management software, etc.

SELECTION OF CORRECT PLATFORM

With the advent of various operating systems such as Windows 95, Windows 98, Windows NT, Windows 2000, Windows XP, Linux, UNIX, OS/2, etc., the question arises: Which is the best?

NT can be considered a *de facto* open operating system because it is a widely used operating system and supports a wide variety of products; however, it is still a proprietary operating system, and most of the internal source code related to the system registry is not released to the public. This makes it cumbersome for administrators to secure NT positively because of the obscurity built into the operating system.

UNIX is an old but highly tested operating system. UNIX is an open operating system in which the source code is available and reviewable. A skilled team of administrators can secure UNIX systems more positively than NT.

Each of the operating systems has advantages and disadvantages. For example, a UNIX-based system comes with a very high level of security provisions, but it requires a great deal of effort and a special skill set to develop and maintain such a system. At the same time a system on Windows NT may not have the robustness of UNIX but may be easier to handle. UNIX has an advantage at the high end (enterprise-level automation) whereas for a small to medium level of enterprise automation this difference is negligible.

However, all these issues are very much debatable. Each operating system has certain advantages and disadvantages over others; the right question a system designer must ask is what is right for a particular application.

The following are questions that may help any control system engineer decide on the ideal choice:

Is a highly open system or a closed system required?
Need the system have a very high level of security provision?
Is the special type of skilled labor available to operate the system that is sought?
Is user-friendliness built into the system?
What support is available for a particular system in the marketplace?
What are the third-party interfaces supported by the system?
What is the proven track record of the system?
Is the system upgradable to the next version?
What is the history of the product under consideration?
Does the vendor support upgrading or will it be necessary to change to a new system altogether?

Technical superiority is not the only consideration in the choice of an operating system. Personal, economic, and environmental factors also play important roles in the decision-making process of platform selection for a particular system. A particular operating system may be ideal technically, but too expensive.

Often a particular feature may be absolutely essential to one company, important to another, trivial to a third, and an unacceptable impediment to a fourth.

Hardware and the interactions between hardware and software can be significant factors in both quantifiable and anecdotal results. Variations in software, even software that shares the same name and version number, can play a significant role.

Variations in the personnel creating, administering, and maintaining the computer system (the human factor) can be a significant factor in results. Many times considerable emphasis is placed on the selection of the platform, but how well the applications are designed and interwoven with the platform is equally important. Many times the platform may be very stable but the application program is unstable.

COMPARING VARIOUS OPERATING PLATFORMS

A brief comparison of the various attributes of each type of operating system follows (Table 4.24g).

TABLE 4.24g
Comparison of Architectural Features of Windows NT Server with SCO UNIX

Architecture	*Windows NT*	*SCO UNIX*
Multiuser operating system	Yes	Yes
Preemptive multitasking	Yes	Yes
Support to intel processor	Yes	Yes
Symmetric multiprocessing	Yes	Yes
Asymmetric multiprocessing	No	Yes
Paged virtual memory	Yes	Yes
Maximum number of user connections	Unlimited	Unlimited
Single login to network	Yes	No
Dynamic loading of services	Yes	Yes
Memory protection	Yes	Yes
Audit alerts	Yes	No
Structured exception handling	Yes	No
Installable file system	Yes	Yes
Hardware abstraction layer	Yes	No
Microkernel-based architecture	Yes	No

Cost

For any buying decision, the bottom line always is determining the cheapest solution available. However, choosing an operating system and evaluating its cost is a different ballgame, because one should consider not only the initial investment but also the maintenance cost over the long run. Normally, the major cost categories are hardware and software costs, but often there are also hidden costs, such as costs of after-sales support agreements, upgrades/service packs, hardware upgrades, profits lost due to downtime, and last, but not least, costs for systems engineers and supporting staff.

It is very important that when the costs of systems are compared the features that are standard in one system must be compared with those in the other systems. It should be determined if these features are also available as a standard package or if they must be purchased at additional cost. NT is often chosen for budget reasons because many customers are not willing to pay for the more expensive hardware required by most commercial versions of UNIX.

Reliability

Reliability is generally considered important by end users. Even if one operating system offers more functionality, is more scalable, and offers greater ease of system management, it would be of no use when a server processing real-time mission-critical process application is plagued by frequent crashes resulting in costly downtimes.

In addition, an operating system may be extremely reliable at one kind of task and extremely unreliable at another. UNIX has been operating in the server environment for more than 30 years and now has proved its mettle as far as reliability is concerned. However, Windows NT and Windows 2000 are still trying to prove their credentials. Many system integrators and manufacturers have shifted from OS/2 or UNIX to Windows NT over the past few years but still UNIX is considered a more reliable system than NT. Only the coming years will establish the reliability of Windows NT as compared with UNIX.

Manageability and Administration

This is a very important issue for system engineers who influence the buying decision of a system. For a distributed control system, manageability and administration are very important, and it is generally recognized that the Windows-based system provides easier manageability and administration because of its graphical user interface (GUI). NT has long enjoyed an intuitive user interface for managing single systems, largely benefiting from the exceptional familiarity of the Windows look-and-feel adopted by the NT GUI. For the administration of a control system, the administration functions are frequently much simplified by providing special administration programs with a user-friendly GUI. Such functions as backup, creating users, sharing resources, defining priorities in the execution of multiple tasks, defining access, and security are a few of the core management and administrative functions.

Scalability

Scalability is a measure of the ability of an operating system to *scale* from use on a very small personal computer to a very large network of computers. The existence of scalability allows for the smooth growth of a computer system: using the same operating system on the entire installed system base.

If the operating system is truly scalable, it is possible (but not always achievable) to reduce administrative, maintenance, and training costs by standardizing on a single operating system. For control system users, scalability is of prime importance because the ever-expanding nature of any business always demands the system to grow larger and larger. If this topic is discussed in the context of the commercial systems, then the issue of multiprocessing capability would arise. However, for a control system manufacturer, scalability is not only the power of the operating system to support multiple processors but also relates to the capability of the hardware that is used in the system, the network topology adopted, the media used for communication, etc. Both Windows NT– and UNIX-based systems are highly scalable. Another viewpoint of scalability is the capability of a system to support physical memory and processors. UNIX systems can, in general, support more physical memory and processors than can NT. This, in turn, leads to higher performance even though both systems can be supported on the same microprocessors.

Security

The *security* of a system is a very debatable subject. No system can claim perfect security. Normally, for a control system security is designed with the following aspects in mind:

1. Security features to avoid corruption or loss of data
2. Security features to avoid misuse (intentional or by operator error)

Regarding the first aspect, the major culprits are computer viruses, worms, or Trojan horses that may enter the system and corrupt the data of the system. To prevent this, *never* allow floppy disks or tape drives from unknown sources to be used in the system. Also, the system must be guarded with strong antivirus software, which must be upgraded regularly. Apart from this, the user rights for editing, copying, etc. must only be accessible to the system engineer. It is generally believed that UNIX-based systems are less prone to virus attacks as compared with Windows NT–based systems.

Regarding the second aspect, system programmers must strive to make the system failproof by anticipating the various combinations that a naive operator may perform on the system. Ideally, all the critical functions must be password-protected or at least they should not be activated by a single key press. There should always be an acknowledgment to each task. These issues may not be directly related to the operating system; however, they are definitely linked indirectly via the built-in security features of each operating system. For example, the Windows NT system has a built-in security feature such that, unless files are allowed to be shared, by default the files remain not shared.

Error Handling

Systems are prone to failures; therefore, some disaster management plan must always be in place. Many of the operating systems provide for annunciation of the failure of their hardware; however, if there are problems in the software/services they also need to be annunciated to the administrator. What is important for a control system engineer is that all the hardware- and software-related alarms and messages appear in the system alarms. A separate group of alarms called system alarms must always be created, and this group must be different from the process alarms. Once errors are reported, it is important to know how to recover from these errors, and here lies the real test of a system engineer. Error recoveries are predominantly governed by the hardware architecture that has been selected for the system and how the system administration tasks are done. If regular backups of the latest changes are kept, then it may be easier to restore the system to good health. Such hardware architecture as disk mirroring and RAID, normal standby configuration, etc. may be very helpful in handling the errors. Both Windows NT- and UNIX-based systems support various error recovery tools very well.

Integration of Software and Hardware

This is the crux of a computer system. A computer system consists of both hardware and software. A key recurring question is how well the software and hardware are integrated. The choice here is between close integration of hardware and software versus the availability of a wide number of choices of vendors, which forces purchase managers to collect the different parts from different parties and then ask system engineers to synchronize the system.

It takes more than a GUI to make a computer easy to use. It takes tight integration between software and hardware.

> The Mac is still more elegant and stylish, still more tightly integrated, with better links between software and hardware, because a single company makes both the computer and operating system.—*The Wall Street Journal*

Both Windows NT and UNIX support a large variety of hardware; however, performance may vary with each type of hardware used. Also, the same application can run quite differently on different platforms, even though the platforms have the same speeds and feeds on performance. This is because optimization levels differ, and sometimes hardware vendors assist software vendors in tuning applications for their particular environments.

The ideal solution is to have a total solution from a single system integrator; for example, buying branded hardware and software from SUN, IBM, HP, or the equivalent would be a better choice than collecting an operating system from one source, a workstation from another source, and other applications from a third source.

Openness

Distributed control systems are the central supervisory and monitoring system for the plant, and there may be many third-party interfaces required to connect with the system. Thus, openness is crucial to the system. Both UNIX and NT are highly open systems. The NT system is considered more open because of the possibility of connecting it to various subsystems available commercially. Normally, these interfaces are a major cause of downtime of the system and the compatability of these interfaces must always be checked before considering them. It must be kept in mind that many times these interfaces require special hardware (gateways, bridges, etc.) and software. Integration and management of these additional gateways is often very cumbersome. If these interfaces are redundant, then the overall redundancy of the system must be reviewed after considering the redundancy of these interfaces in tandem with system redundancy.

Another measure of connectivity is the ability to run programs from other operating systems in emulation. Emulated software runs more slowly than native software, but allows for easy trading of data and use of obscure programs available on a limited number of platforms. Hardware emulation can attain the same speed as the actual system being emulated. It also has the possible advantage of sharing some computer resources (sharing hard drives, sharing monitors, etc.) and perhaps even the ability to copy and paste between systems or other levels of direct sharing, as well as saving desk space by requiring only one computer setup to run multiple systems.

CONCLUSIONS

The bottom line is that there is little to distinguish between UNIX and Windows NT. This choice is much more dependent on the environment in which the operating system has to run, the industry and the application to be served, the technical support and programming staff, the client environment, and the existing infrastructure. Eventually many of the glaring technical differences between these operating systems will become nonissues because of the industry trend toward better integration of these two operating systems, with products such as OpenNT.

As process control tends to move away from single-loop control toward unit-operation control, the workstation designs will also reflect that trend (see Figure 4.6m).

GLOSSARY

CLIENT/SERVER ARCHITECTURE An architecture where processing is shared between client and server applications. The server accepts requests from the client, processes the requests, and sends information back to the client.

DATA A representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by people or automatic means; any representation, such as characters, to which meaning might be assigned.

DATA ACQUISITION The process of identifying, isolating, and gathering source data to be processed by some central agency into a usable form.

DRIVER A series of instructions the computer follows to reformat data for transfer to and from a particular network device. Software drivers standardize the data format between different kinds of devices. Multiplexing of messages through a driver, for example, is made possible by supervisor synchronizing commands, or primitives, that permit several sender processes to send messages to a single-receiver process port associated with the driver.

EMULATION The use of programming techniques and special machine features to permit a computing system to execute programs written for another system.

END USER A person, process, program, device, or system that functions as an information source or destination and/or employs a user–application network for the purpose of data processing and information exchange.

ETHERNET A baseband local-area network specification developed jointly by Xerox Corporation, Intel, and Digital Equipment Corporation to interconnect computer equipment using coaxial cable and transceivers.

HOST (1) In a network, a computer that primarily provides services such as computation, database access, special programs, or programming languages; the primary or controlling computer in a multiple-computer installation. (2) An abstraction of an operating environment wherein a set of processes interacts with a supervisor. The supervisor contains system processes and manages the operating environment, which includes input/output devices, directories, and file systems. A host is a convenient boundary for containing specific resources needed by other hosts. A host is virtual, and several hosts may reside on the same computer.

INTERFACE (1) A shared boundary. For example, the physical connection between two systems or two devices. (2) Generally, the point of interconnection of two components, and the means by which they must exchange signals according to some hardware or software protocol.

SERVER A processor that provides a specific service to the network, for example, a routing server, which connects nodes and networks of like architectures; a gateway server, which connects nodes and networks of different architectures by performing protocol conversions; and a terminal server, printer server, and file server, which provide an interface between compatible peripheral devices on a local-area network.

TERMINAL (1) A point at which information can enter or leave a communication network. (2) An input/output device to receive or send source data in an environment associated with the job to be performed, capable of transmitting entries to and obtaining output from the system of which it is a part.

TOPOLOGY (1) The physical arrangement of nodes and links on a network; description of the geometric arrangement of the nodes and links that make up a network, as determined by their physical connections. (2) The possible logical connections between nodes on a network, indicating which pairs of nodes are able to communicate, whether or not they have a direct physical connection. Examples of network topologies are bus, ring, star, and tree.

Bibliography

Arsonson, R. L., "CRT Terminals Make Versatile Computer Interface," *Control Engineering,* April 1970.

Barnard, H., "Engineering and Ergonomics in Control Room Design," *Measurement & Control News,* Vol. 33(4), page 121, 1999.

Compton, K., *Image Performance in CRT Displays,* SPIE Press, Bellingham, 2003.

Ellis, R. K., "Color Graphics CRT's Provide Window into Factory Operations," *Instruments and Control Systems,* February 1983.

Gadberry, B. E., "Designing Integrated Control System Displays," Paper # 91–0354, 1991 ISA Conference, Anaheim, October 1991.

Guerlain, S., and Bullemer, P., "User-Initiated Notification: A Concept for Aiding the Monitoring Activities of Process Control Operators," Honeywell Technology Center 1996.

Hwang, F. H., and Hwang, S. L., "Design and Evaluation of Computerized Operating Procedures in Nuclear Power Plants," *Ergonomics,* Vol. 46(1-3), pages 271–284, 2003.

IEEE 1289, "IEEE Guide for the Application of Human Factors Engineering in the Design of Computer-based Monitoring and Control Displays," Piscataway, NJ: IEEE Press, 1998.

ISA-TR77.60.04-1996, "Fossil Fuel Power Plant Human–Machine Interface CRT Displays," 1996.

Keller, P. A., *Electronic Display Measurement: Concepts, Techniques, and Instrumentation,* John Wiley & Sons, New York, 1997.

Lanier, G. W., "Developing CRT Graphics Displays," 1992 ISA Conference, Houston, October 1992.

Lueder, E., *Liquid Crystal Displays: Addressing Schemes and Electro-optical Effects,* John Wiley & Sons, New York, 2001.

Lundstrom, J. E., "CRT Terminal Uses Mosaics to Build Process Graphics," *Control Engineering,* February 1981.

MacDonald, L. W., and Lowe, A. C., Display Systems: *Design and Applications,* John Wiley & Sons, New York, 1997.

MaCready, A. K., "Man–Machine Interfacing for the Process Industries," *InTech,* March 1982.

NUREG-0700, "Human–System Interface Design Review Guideline," Vol. 1, Washington, DC: The NRC Public Document Room, 1996.

NUREG/CR-6637, "Human Systems Interface and Plant Modernization Process," USNRC, March 2000.

Reinhart, S., "Are Your Control Displays User Friendly?" ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Rumph, P., "Improved Information Interfacing Turns Operators into Process Managers," *Pulp & Paper,* August 1998.

Shirley, R. S., et al., "What's Needed at the Man/Machine Interface?" *InTech,* March 1981.

Stubler, W. F., et al., "Human System Interface Modernization Process," NUREG/CR-6637, Washington, DC, Office of Nuclear Regulatory Research, 2000.

Technical Report 1001066, "Human Factors Guidance for Digital I&C Systems in Hybrid Control Rooms," Palo Alto, CA, EPRI, November, 2000.

Tollmar, K., and Sundblad, Y., "The Design and Building of the Graphic User Interface for the Collaborative Desktop," *Computers & Graphics,* Vol. 19(2), 1995.

Vicente, K. J., "Operator Adaptation in Process Control: A Three-Year Research Program," *Control Engineering Practice,* Vol. 5(3), 1997.

PLCs and Other Logic Devices

5

5.1
BINARY LOGIC DIAGRAMS FOR PROCESS OPERATIONS 880

5.2
LADDER DIAGRAMS 893

5.3
OPTIMIZATION OF LOGIC CIRCUITS 898

5.4
PLCs: PROGRAMMABLE LOGIC CONTROLLERS 906

5.5
PLC PROGRAMMING 944

5.6
PLC SOFTWARE ADVANCES 966

5.7
PRACTICAL LOGIC DESIGN 976

5.8
PROGRAMMABLE SAFETY SYSTEMS 993

5.1 Binary Logic Diagrams for Process Operations

Standard formatted for publication in this handbook by:

J. E. JAMISON (2005)

Reprinted by permission. Copyright © 1976, Instrument Society of America. From *ANSI/ISA-S5.2-1976 (R 1992)*, "*Binary Logic Diagrams for Process Operations*," reaffirmed July 13, 1992. Also, permission was granted by ISA for inclusion of the latest thinking on binary logic, memory, and time functions from *Draft 4 of ISA Draft 5.01.01 Instrumentation Symbols and Identification, Copyright 2000 ISA,* which now includes the previous S5.2.[1]

1. PURPOSE

1.1 The purpose of this Standard is to provide a method of logic diagramming of binary interlock and sequencing systems for the start-up, operation, alarm, and shutdown of equipment and processes in the chemical, petroleum, power generation, air conditioning, metal refining, and numerous other industries.

1.2 The Standard is intended to facilitate the understanding of the operation of binary systems and to improve communications among technical, management, design, operating, and maintenance personnel concerned with the systems.

2. SCOPE

2.1 The Standard provides symbols, both basic and nonbasic, for binary operating functions. The use of symbols in typical systems is illustrated in the appendices.

2.2 The Standard is intended to symbolize the binary operating functions of a system in a manner that can be applied to any class of hardware, whether it be electronic, electrical, fluidic, pneumatic, hydraulic, mechanical, manual, optical, or other.

3. USE OF SYMBOLS

3.1 By using the symbols designated as "basic," logic systems may be described with the use of only the most fundamental logic building blocks. The remaining symbols, not basic, are more comprehensive and enable logic systems to be diagrammed more concisely. Use of the nonbasic symbols is optional.

3.2 A logic diagram may be more or less detailed depending on its intended use. The amount of detail in a logic diagram depends on the degree of refinement of the logic and on whether auxiliary, essentially nonlogic, information is included.

As an example of refinement of detail: A logic system may have two opposing inputs, e.g., a command to open and a command to close, which do not normally exist simultaneously; the logic diagram may or may not go so far as to specify the outcome if both the commands were to exist at the same time. In addition, explanatory notes may be added to the diagram to record the logic rationale.

Nonlogic information may also be added, if desired, e.g., reference document identification, tag numbers, terminal markings, and so on.

In these ways, the diagram may provide the level of detail appropriate, for example, for communication between a designer of pneumatic circuits and a designer of electric circuits, or may provide a broad-view system description for a plant manager.

3.3 The existence of a logic signal may correspond physically to either the existence or the nonexistence of an instrument signal, depending on the particular type of hardware system and the circuit design philosophy that are selected.[2] For example, a high-flow alarm may be chosen to be actuated by an electric switch whose contacts open on high flow; on the other hand, the high-flow alarm may be designed to be actuated by an electric switch whose contacts close on high flow. Thus, the high-flow condition may be represented physically by the absence of an electric signal or by the presence of the electric signal. The Standard does not attempt to relate the logic signal to an instrument signal of any specific kind.

3.4 A logic symbol that is shown in Section 4 with three inputs—A, B, and C—is typical for the logic function having any number of two or more inputs.

3.5 The flow of intelligence is represented by lines that interconnect logic statements. The normal direction of flow is from left to right, or top to bottom. Arrowheads may be added to the flow lines wherever needed for clarity and shall be added to lines whose flow is not in a normal direction.

---

[1] A disclaimer to any future ISA Standards documents is hereby stated. The reader is cautioned that the Draft ISA Document that provided Table 5.1a has not been approved as of the time of this writing. Therefore, it cannot be presumed to reflect the position of ISA or any other committee, society, or group.

[2] In process operations, binary instrument signals are commonly either *ON* or *OFF*. However, as a more general case, logic systems exist that make use of binary hardware having signals with two alternate real values, e.g., +5 V and –3 V. In *positive logic,* the more positive signal, +5 V, represents the existence of a logic condition, e.g., *pump stopped*. In *negative logic*, the less positive signal, –3 V, represents the existence of a logic condition of *pump stopped.*

3.6 A summary of the status of an operating system may be put in the diagram wherever it is deemed useful as a reference point or landmark in the sequence.

3.7 There may be misunderstanding of binary logic statements involving devices that are not recognizable as inherently having only two specific alternative states. For example, if it is stated that a valve is not closed, this could mean either (a) that the valve is open fully, or (b) that the valve is simply not closed, namely, that it may be in any position from almost closed to wide open. To aid accurate communication between writer and reader of the logic diagram, the diagram should be interpreted literally. Therefore, possibility (b) is the correct one.

If a valve is an open-close valve, then, to avoid misunderstanding, it is necessary to do one of the following:

1. Develop the logic diagram in such a way that it says exactly what is intended. If the valve is intended to be open, then it should be so stated and not be stated as being not closed.
2. Have a separate note specifying that the valve always assumes either the closed or the open position.

By contrast, a device such as a motor-driven pump is either operating or stopped, barring some special situations. To say that the pump is not operating usually clearly denotes that it has stopped.

The following definitions apply to devices that have open, closed, or intermediate positions. The positions stated are nominal to the extent that there are differential-gap and dead band in the instrument that senses the position of the device.

Open position: a position that is 100% open.
Not-open position: a position that is less than 100% open. A device that is not open may or may not be closed.
Closed position: a position that is 0% open.
Not-closed position: a position that is more than 0% open. A device that is not closed may or may not be open.
Intermediate position: a SPECIFIED position that is greater than 0% and less than 100% open.
Not-at-intermediate position: a position that is either above or below the SPECIFIED intermediate position.

For a logic system having an input statement that is derived inferentially or indirectly, a condition may arise that will lead to an erroneous conclusion. For example, an assumption that flow exists because a pump motor is energized may be false because of a closed valve, a broken shaft, or other mishap. Factual statements, that is, statements based on positive measurements that a certain condition specifically exists or does not exist, are generally more reliable.

3.8 A process operation may be affected by loss of the power supply[3] to memories and to other logic elements. In order to take such operating eventualities into account, it may therefore be necessary to consider the effect of loss of power to any logic component or to the entire logic system. In such cases, it may be necessary to enter power supply or loss of power supply as logic inputs to a system or to individual logic elements. For memories, the consideration of power supply may be handled in this manner or as shown in sections 10–12 in Table 5.1a.

By the same token, it may be necessary to consider the effect of restoration of power supply.

Logic diagrams do not necessarily have to cover the effect of logic power supplies on process systems but may do so for thoroughness.

3.9 It is recommended, for clarity, that a single time-function symbol, as appropriate, be used to represent each time function in its entirety. Though not incorrect, the representation of a complex or uncommon time function by using a time-function symbol in immediate sequence with a second time-function symbol or with a *NOT* symbol should be avoided (see Table 5.1a).

3.10 Process instrument symbols and designations follow ANSI/ISA Standard S5.1-1984 (formerly American National Standards Institute Standard Y32.20-1975), "Instrumentation Symbols and Designations." However, these symbols are included for illustrative purposes only, and are not part of Standard S5.2.

3.11 If a drawing, or set of drawings, uses graphic symbols that are similar or identical to one another in shape or configuration and that have different meanings because they are taken from different standards, then adequate steps shall be taken to avoid misinterpretation of the symbols used. These steps may be to use caution notes or reference notes, comparison charts that illustrate and define the conflicting symbols, or other suitable means. This requirement is especially critical if the graphic symbols used, being from different disciplines, represent devices, conductors, flow lines, or signals whose symbols, if misinterpreted, may result in danger to personnel or damage to equipment.

4. SYMBOLS

The symbols for diagramming binary logic are defined in Table 5.1a and are the latest thinking of the ISA SP5.1 subcommittee:

The symbols in Table 5.1a are never used in piping and instrument diagrams (P&IDs) and are used to help document and diagram logic control designs and narratives. The present Standard ISA S5.2 (ANSI/ISA-S5.2-1976(R1992) is now being revised and rolled into the new ANSI/ISA-5.01.01 standard as proposed in the current (as of this writing) Draft 4. Symbols, Truth Tables, Definitions, and Graphs used here in Section 5.1 are in accordance with Draft 4 and are very different from S5.2. These are given here to illustrate to the reader the latest thinking in this area, including expanded timing functions. Application information and examples on

---

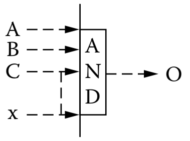
[3] The term "power supply" covers the energizing medium, whether it be electric, pneumatic, or other.

TABLE 5.1a

Instrument and Control System Functional Diagramming Symbols-Binary Logic, Memory, and Time Functions (Proposed for the next revision of ISA S5.1 (now ANSI/ISA-5.01.01) at the time of this writing)

No	*Symbol* *Truth Table*	*Definition* *Graph*

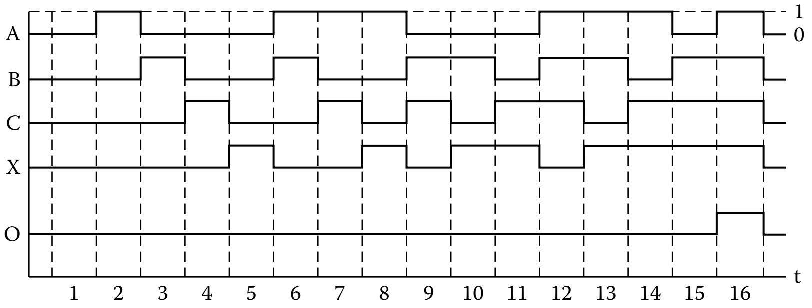
01



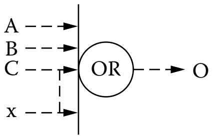
	A	B	C	x	O
1	0	0	0	0	0
2	1	0	0	0	0
3	0	1	0	0	0
4	0	0	1	0	0
5	0	0	0	1	0
6	1	1	0	0	0
7	1	0	1	0	0
8	1	0	0	1	0
9	0	1	1	0	0
10	0	1	0	1	0
11	0	0	1	1	0
12	1	1	1	0	0
13	1	1	0	1	0
14	1	0	1	1	0
15	0	1	1	1	0
16	1	1	1	1	1

1. AND gate.

2. Output true only if all inputs are true.



02



	A	B	C	x	O
1	0	0	0	0	0
2	1	0	0	0	1
3	0	1	0	0	1
4	0	0	1	0	1
5	0	0	0	1	1
6	1	1	0	0	1
7	1	0	1	0	1
8	1	0	0	1	1
9	0	1	1	0	1
10	0	1	0	1	1
11	0	0	1	1	1
12	1	1	1	0	1
13	1	1	0	1	1
14	1	0	1	1	1
15	0	1	1	1	1
16	1	1	1	1	1

1. OR gate.

2. Output true if any input is true.

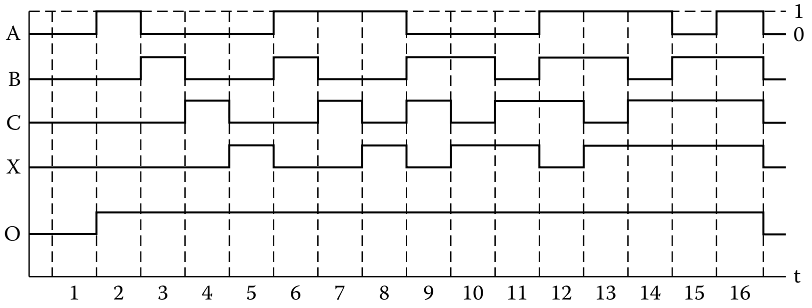
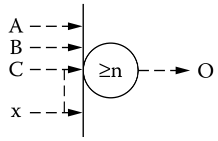
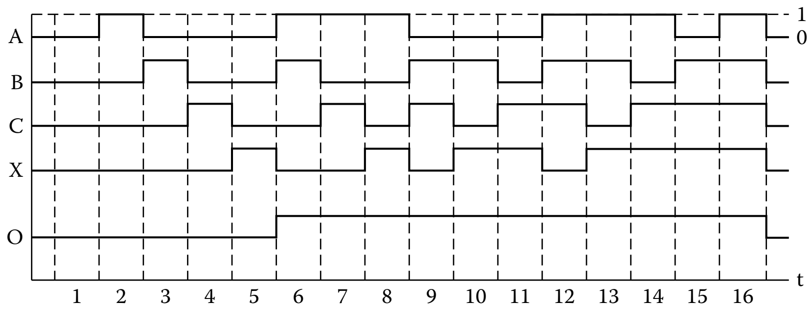
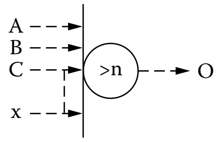


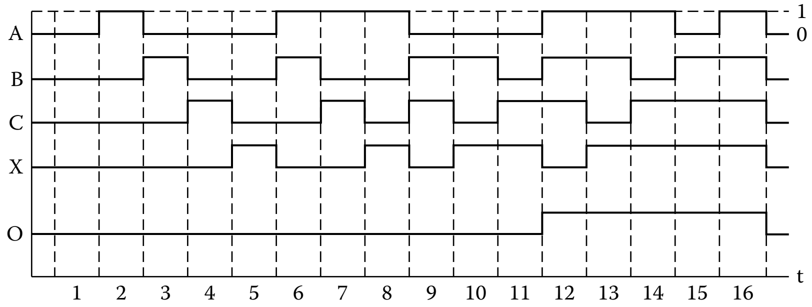
TABLE 5.1a (continued)

Instrument and Control System Functional Diagramming Symbols-Binary Logic, Memory, and Time Functions (Proposed for the next revision of ISA S5.1 (now ANSI/ISA-5.01.01) at the time of this writing)

<table>
<tr><th rowspan="2">No</th><th>Symbol</th><th>Definition</th></tr>
<tr><th>Truth Table</th><th>Graph</th></tr>
<tr><td>03</td><td>


<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>x</th><th>O</th></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>2</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>3</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>6</td><td>1</td><td>1</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>7</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>8</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>9</td><td>0</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>10</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>11</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>12</td><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>13</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>14</td><td>1</td><td>0</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>15</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>16</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
</table>
</td><td>
1. Qualified OR gate with greater than or equal to qualifications.
2. Output equals “1” if number of inputs equal to “1” are greater than or equal to “n” inputs.
3. Truth table and graph are for “n” equals 2.


</td></tr>
<tr><td>04</td><td>


<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>x</th><th>O</th></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>2</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>3</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>6</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>7</td><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>8</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>9</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>10</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>11</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>12</td><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>13</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>14</td><td>1</td><td>0</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>15</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>16</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
</table>
</td><td>
1. Qualified OR gate with greater than qualifications.
2. Output equals “1” if number of inputs equal to “1” are greater than but not equal to “n” inputs.
3. Truth table and graph are for “n” equals 2.
.


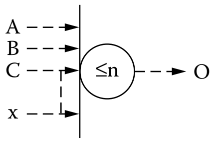
</td></tr>
</table>

TABLE 5.1a (continued)

Instrument and Control System Functional Diagramming Symbols-Binary Logic, Memory, and Time Functions (Proposed for the next revision of ISA S5.1 (now ANSI/ISA-5.01.01) at the time of this writing)

No	Symbol Truth Table	Definition Graph

05

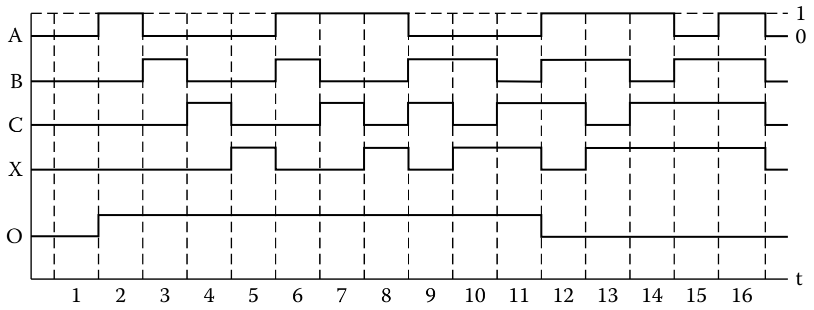


	A	B	C	x	O
1	0	0	0	0	0
2	1	0	0	0	1
3	0	1	0	0	1
4	0	0	1	0	1
5	0	0	0	1	1
6	1	1	0	0	1
7	1	0	1	0	1
8	1	0	0	1	1
9	0	1	1	0	1
10	0	1	0	1	1
11	0	0	1	1	1
12	1	1	1	0	0
13	1	1	0	1	0
14	1	0	1	1	0
15	0	1	1	1	0
16	1	1	1	1	0

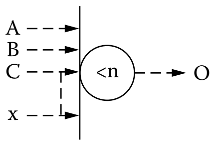
1. Qualified OR gate with less than or equal to qualifications.

2. Output equals "1" if number of inputs equal to "1" are less than or equal to "n" inputs.

3. Truth table and graph are for "n" equals 2.



06



	A	B	C	x	O
1	0	0	0	0	0
2	1	0	0	0	1
3	0	1	0	0	1
4	0	0	1	0	1
5	0	0	0	1	1
6	1	1	0	0	0
7	1	0	1	0	0
8	1	0	0	1	0
9	0	1	1	0	0
10	0	1	0	1	0
11	0	0	1	1	0
12	1	1	1	0	0
13	1	1	0	1	0
14	1	0	1	1	0
15	0	1	1	1	0
16	1	1	1	1	0

1. Qualified OR gate with less than qualifications.

2. Output equals "1" if number of inputs equal to "1" are less than but not equal to "n" inputs.

3. Truth table and graph are for "n" equals 2.

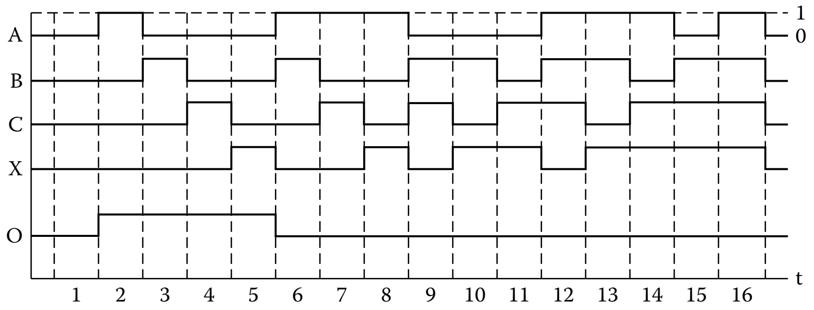


TABLE 5.1a (continued)

Instrument and Control System Functional Diagramming Symbols-Binary Logic, Memory, and Time Functions (Proposed for the next revision of ISA S5.1 (now ANSI/ISA-5.01.01) at the time of this writing)

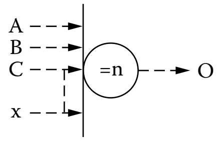
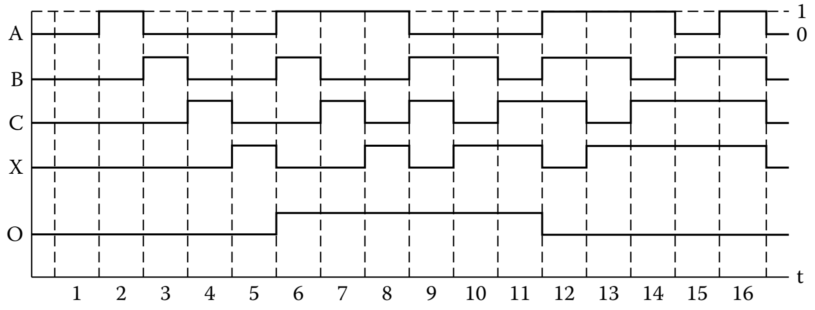
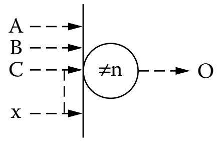
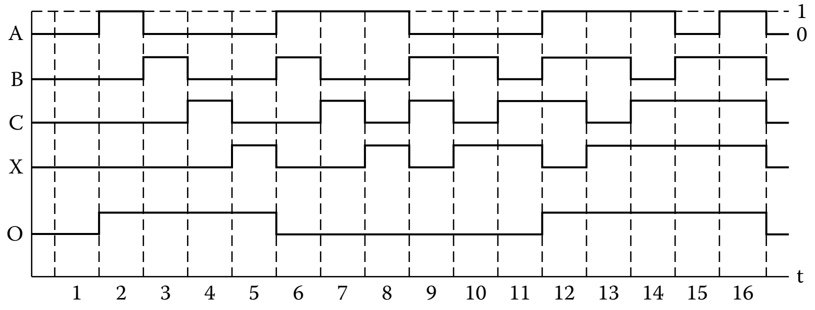
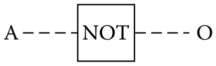
<table>
<tr><th rowspan="2">No</th><th>Symbol</th><th>Definition</th></tr>
<tr><th>Truth Table</th><th>Graph</th></tr>
<tr><td>07</td><td>
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>x</th><th>O</th></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>2</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>3</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>6</td><td>1</td><td>1</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>7</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>8</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>9</td><td>0</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>10</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>11</td><td>0</td><td>0</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>12</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>13</td><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>14</td><td>1</td><td>0</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>15</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>16</td><td>1</td><td>1</td><td>1</td><td>1</td><td>0</td></tr>
</table></td><td>1. Qualified OR gate with equal to qualifications.
2. Output equals “1” if inputs equal to “1” are equal to “n” inputs.
3. Truth table and graph are for “n” equals 2.
 </td></tr>
<tr><td>08</td><td>
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>x</th><th>O</th></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>2</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>3</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>4</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>6</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>7</td><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>8</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>9</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>10</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>11</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>12</td><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>13</td><td>1</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>14</td><td>1</td><td>0</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>15</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>16</td><td>1</td><td>1</td><td>1</td><td>1</td><td>1</td></tr>
</table></td><td>1. Qualified OR gate with not equal to qualifications.
2. Output equals “1” if inputs equal to “1” are not equal to “n” inputs.
3. Truth table and graph are for “n” equals 2.
 </td></tr>
</table>

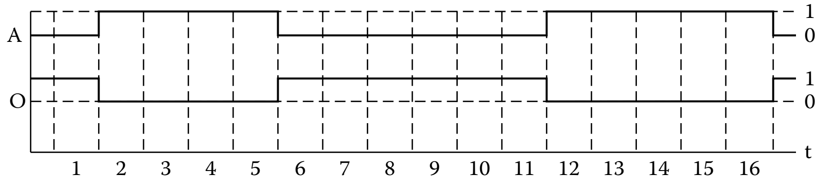
TABLE 5.1a (continued)

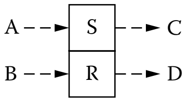
Instrument and Control System Functional Diagramming Symbols-Binary Logic, Memory, and Time Functions (Proposed for the next revision of ISA S5.1 (now ANSI/ISA-5.01.01) at the time of this writing)

<table>
<tr><th>No</th><th><i>Symbol</i>
<i>Truth Table</i></th><th><i>Definition</i>
<i>Graph</i></th></tr>
<tr><td>09</td><td>


<table>
<tr><th>A</th><th>O</th></tr>
<tr><td>1</td><td>0</td></tr>
<tr><td>0</td><td>1</td></tr>
</table>
</td><td>

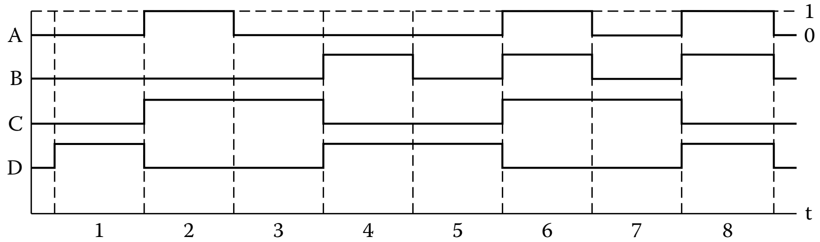
1. NOT gate.
2. Output false if input true.
3. Output true if input false.

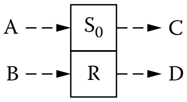


</td></tr>
<tr><td>10</td><td>


<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>2</td><td>1</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>4</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>6</td><td>1</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>7</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>8</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
</table>
</td><td>

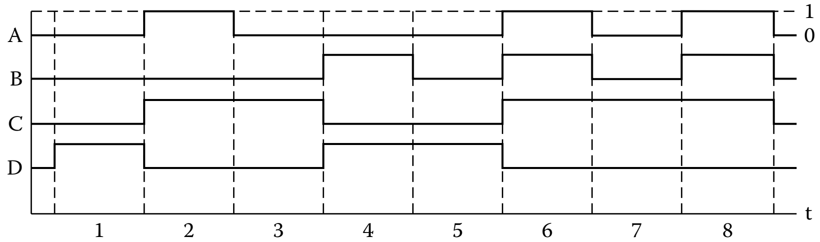
1. Basic memory.
2. Outputs C and D are always opposite.
3. If input A equals “1” then output C equals “1” and D equals “0.”
4. If input A changes to “0” output C remains “1” until input B equals “1,” then C equals “1” and D equals “0.”
5. If input B equals “1” then output D equals “1” and C equals “0.”
6. If input B changes to “0” output D remains “1” until input A equals “1,” then D equals “1” and C equals “0.”
7. If inputs A and B are simultaneously equal to “1” then outputs C and D change state.



</td></tr>
<tr><td>11</td><td>


<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>2</td><td>1</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>4</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>6</td><td>1</td><td>1</td><td>1</td><td>0</td></tr>
<tr><td>7</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>8</td><td>1</td><td>1</td><td>1</td><td>0</td></tr>
</table>
</td><td>

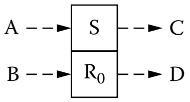
1. Set dominant memory (“S_o Dominant”).
2. Outputs C and D are always opposite.
3. If input A equals “1” then output C equals “1” and D equals “0.”
4. If input A changes to “0” output C remains “1” until input B equals “1,” then output C equals “1” and D equals “0.”
5. If input B equals “1” then output D equals “1” and C equals “0.”
6. If input B changes to “0” output D remains “1” until input A equals “1,” then output D equals “1” and C equals “0.”
7. If inputs A and B are simultaneously equal to “1” then output C equals “1” and D equals “0.”

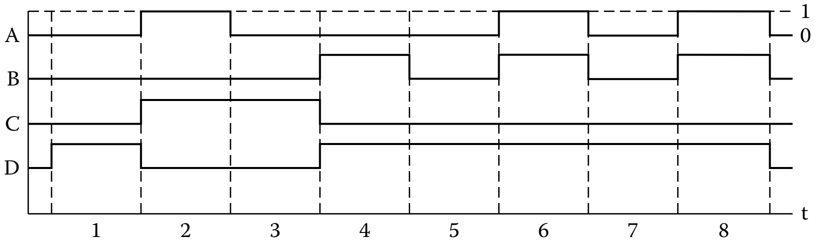


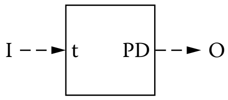
</td></tr>
</table>

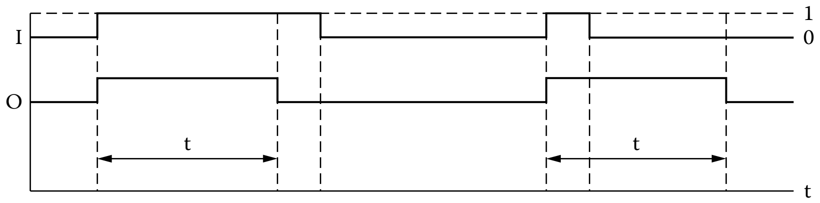
TABLE 5.1a (continued)

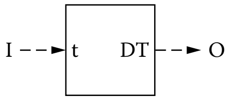
Instrument and Control System Functional Diagramming Symbols-Binary Logic, Memory, and Time Functions (Proposed for the next revision of ISA S5.1 (now ANSI/ISA-5.01.01) at the time of this writing)

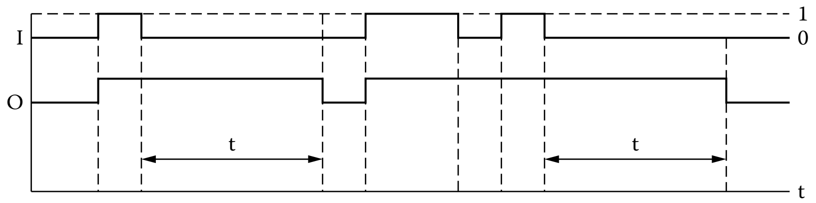
<table>
<tr><th rowspan="2">No</th><th>Symbol</th><th>Definition</th></tr>
<tr><th>Truth Table</th><th>Graph</th></tr>
<tr><td>12</td><td>


<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th></tr>
<tr><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>2</td><td>1</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>1</td><td>0</td></tr>
<tr><td>4</td><td>0</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>5</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>6</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
<tr><td>7</td><td>0</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td>8</td><td>1</td><td>1</td><td>0</td><td>1</td></tr>
</table>
</td><td>
1. **Reset dominant memory (“R_o Dominant”).**
2. Output C and D are always opposite.
3. If input A equals “1” then output C equals “1” and D equals “0.”
4. If input A changes to “0” output C remains “1” until input B equals “1,” then output C equals “1” and D equals “0.”
5. If input B equals “1” then output D equals “1” and C equals “0.”
6. If input B changes to “0” output D remains “1” until input A equals “1,” then output D equals “1” and C equals “0.”
7. If inputs A and B are simultaneously equal to “1” then C equals “0” and D equals “1.”


</td></tr>
<tr><td>13</td><td>


NONE
</td><td>
1. **Pulse duration, fixed.**
2. Output O changes from “0” to “1” and remains “1” for prescribed time duration “t” when input “I” changes from “0” to “1.”


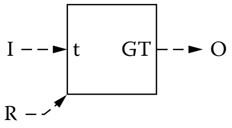
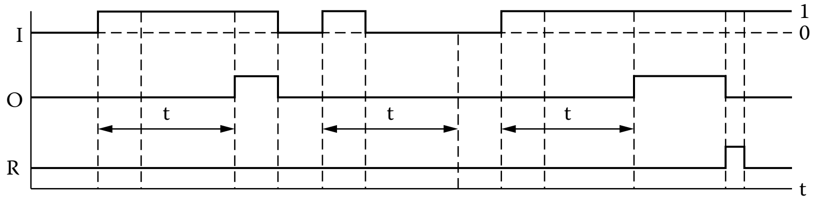
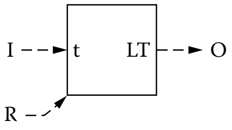
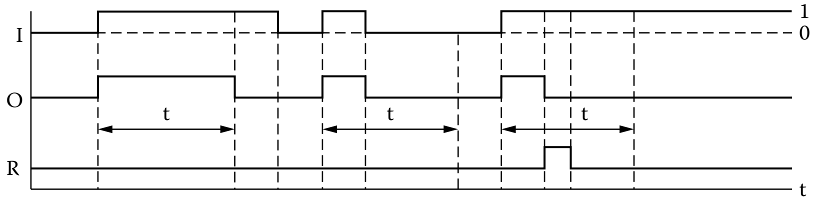
</td></tr>
<tr><td>14</td><td>


NONE
</td><td>
1. **Off time delay.**
2. Output O changes from “0” to “1” when input “I” changes from “0” to “1.”
3. Output O changes from “1” to “0” after input I changes from “1” to “0” and has been equal to “0” for time duration “t.”


</td></tr>
</table>

TABLE 5.1a (continued)

Instrument and Control System Functional Diagramming Symbols-Binary Logic, Memory, and Time Functions (Proposed for the next revision of ISA S5.1 (now ANSI/ISA-5.01.01) at the time of this writing)

No	*Symbol* *Truth Table*	*Definition* *Graph*
15	 NONE	**1. On time delay.** 2. Output O changes from "0" to "1" after input "I" changes from "0" to "1" and "I" remains "1" for prescribed time duration "t." 3. Output O remains "1" until: a. Input "I" changes to "0." b. Reset R changes to "1." 
16	 NONE	**1. Pulse duration, variable.** 2. Output O changes from "0" to "1" when input "I" changes from "0" to "1." 3. Output O changes from "1" to "0" when: a. Input "I" has equaled "1" for time duration "t." b. Input "I" changes from "1" to "0." c. Reset R changes to "1." 

the use of the binary symbols given in Appendices *A*, *B*, and *C* are direct extracts from and utilize the current Standard ANSI/ISA-S5.2-1976(R1992).

Binary logic switching and memory functions are used in analog or sequential control schemes. In truth tables and graphs, Logic One ("1") is "true" and Logic Zero ("0") is "false."

APPENDIX A: GENERAL APPLICATION EXAMPLE

A1. Introduction

This example uses a representative process whose instruments are denoted by the symbols of ANSI/ISA-S5.1-1984 (R 1992) (ANSI Y32.20-1975). The process equipment symbols are included only to illustrate applications of instrumentation symbol. The example is not a part of Standard S5.2.

A2. Simplified Flow Diagram

Figure 5.1b provides the flow sheet representation of the logic involved in a tank filling operation. A written explanation of the various symbols follows:

A3. Word Description

A3.1 Pump Start Feed is pumped into either tank A or tank B. The pump may be operated manually or automatically, selected manually on a local maintained output select switch, *HS-7*, which has three positions: *ON*, *OFF*, and *AUTO*. When the pump is operating, red pilot light *L-8A* is on; when not operating, green pilot light *L-8B* is on. Once started, the pump continues to operate until a stopping command exists or until the control power supply is lost.

The pump may be operated manually at any time provided that no trouble condition exists: The suction pressure must not be low; the seal water pressure must not be low; and the pump motor must not be overloaded and its starter must be reset.

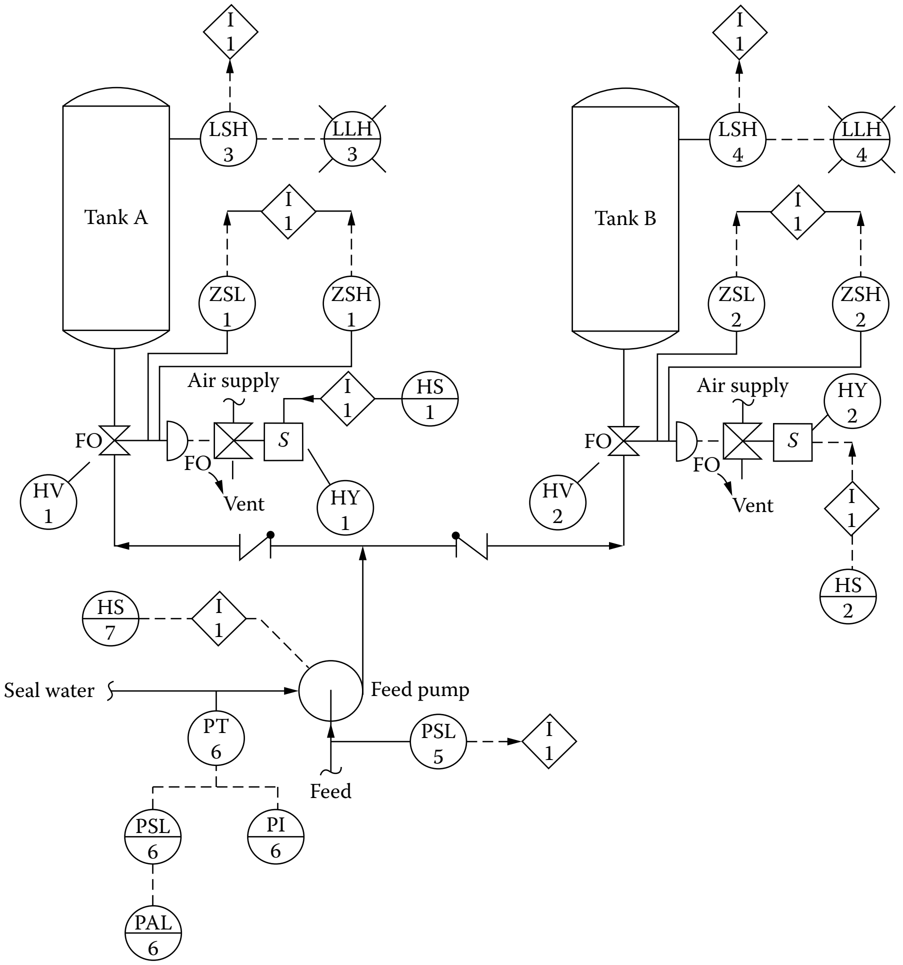


FIG. 5.1b
Tank filling operation-simplified flow diagram.

In order to operate the pump automatically, all the following conditions must be met:

A3.1.1 Board-mounted electric momentary-contact hand switches, *HS-1* and *HS-2*, start the filling operation for tanks *A* and *B*, respectively. Each switch has two positions, *START* and *STOP. START* de-energizes the associated solenoid valves, *HY-1* and *HY-2*. De-energizing a solenoid valve causes it to go to the fail-safe position, i.e., to vent. This depressurizes the pneumatic actuator of the associated control valves, *HV-1* and *HV-2*. Depressurizing a control valve causes it to go to the fail-safe position, i.e., to open. The control valves have associated open-position switches, *ZSH-1* and *ZSH–2*, and closed-position switches*, ZSL-1* and *ZSL-2*.

The *STOP* position of switches *HS-1* and *HS-2* causes the opposite actions to occur so that the solenoid valves are energized, the control valve actuators are pressurized, and the control valves close.

If starting circuit power is lost, the starting memory is lost and the filling operation stops. The command to stop filling can override the command to start filling.

To start the pump automatically, either control valve *HV-1* or *HV-2* must be open and the other control valve must be closed, depending on whether tank *A* or tank *B* is to be filled.

A.3.1.2 The pump suction pressure must be above a given value, as signaled by pressure switch *PSL-5*.

A.3.1.3 If valve *HV-1* is open to permit pumping into tank *A,* the tank level must be below a given value, as signaled by level switch *LSH-3*, which also actuates a board-mounted high-level pilot light, *LLH-3*. Similarly, high-level switch, *LSH-4*, permits pumping into tank *B,* if not actuated, and actuates pilot light *LLH-4*, if actuated.

A.3.1.4 Pump seal water pressure must be adequate, as indicated on board-mounted receiver gage, *PI-6*. This is a non-interlocked requirement that depends on the operator's attention before the operation starts. Pressure switch, *PSL-6*, behind the board, actuates the board-mounted low-pressure alarm, *PAL-6.*

A.3.1.5 The pump drive motor must not be overloaded and its starter must be reset.

A.3.2 Pump Stop The pump stops if any of the following conditions exists:

A.3.2.1 While pumping into a tank, its control valve leaves the fully open position, or the valve of the other tank leaves its fully closed position, provided that the pump is on automatic control.

A.3.2.2 The tank selected for filling becomes full, provided that the pump is on automatic control.

A.3.2.3 The pump suction pressure is continuously low for 5 seconds.

A.3.2.4 The pump drive motor is overloaded. It is immaterial to the process logic whether or not the memory of the pump motor overload is retained on loss of power in this system because the maintained memory that operates the pump is defined as losing memory on loss of power, and this by itself will cause the pump to stop. However, an existing motor-overload condition prevents the motor starter from being reset.

A.3.2.5 The sequence is stopped manually through *HS-1* or *HS-2*. If stop and start commands for pump operation exist simultaneously, then the stop command overrides the operate command.

A.3.2.6 The pump is stopped manually by *HS-7*.

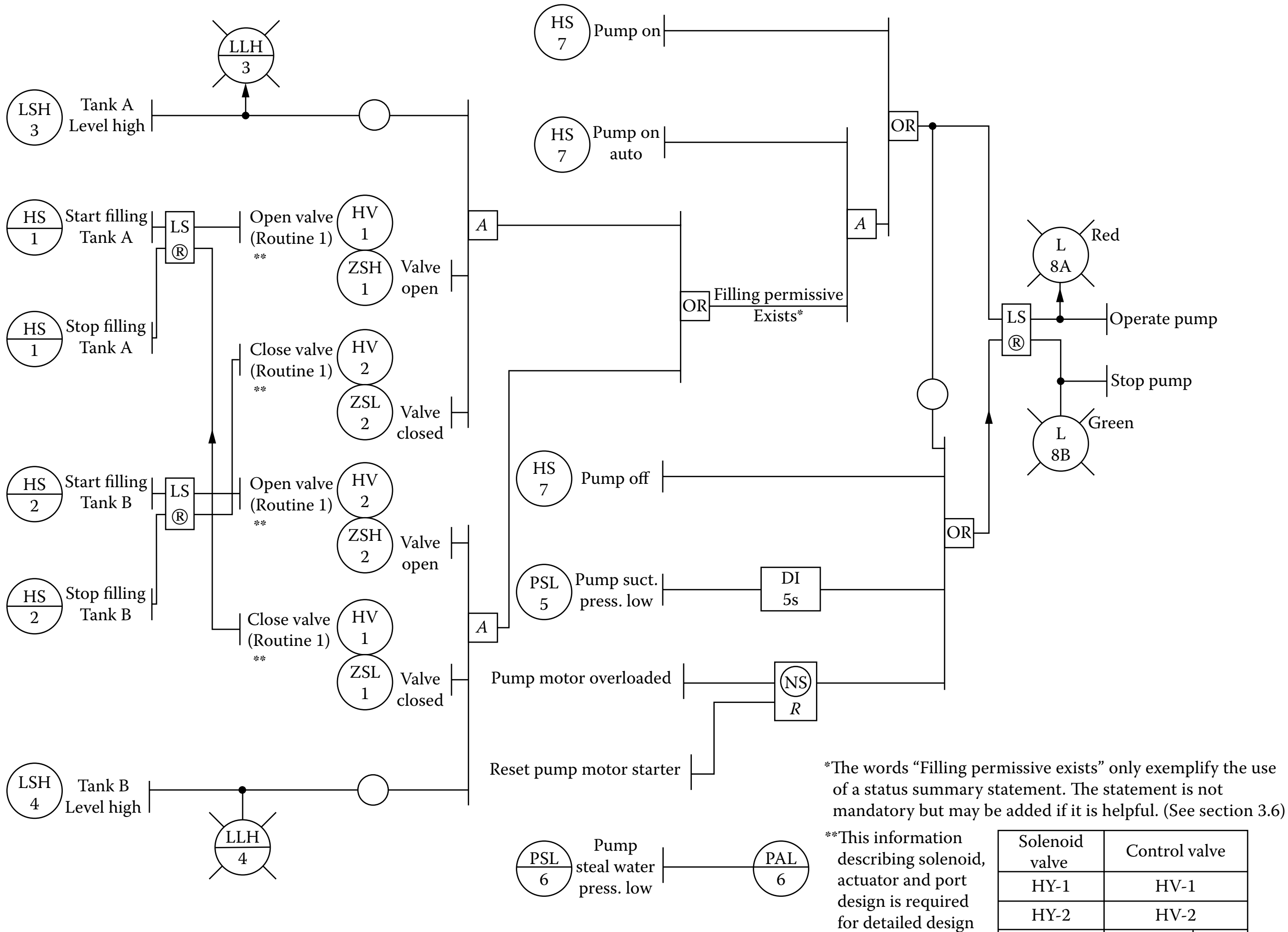
A.3.2.7 The pump seal water pressure is low. This condition is not interlocked and requires manual intervention to stop the pump.

A4. Logic Diagram

The equivalent of the flow sheet representation shown in Figure 5.1b is the logic diagram provided in Figure 5.1c.

Comments on the logic diagram for Interlock 1 described in Figure 5.1c:

1. The diagram may be simplified by using general notes (GN) for a project, especially for repetitive items. For example, the operating light for the pump may be omitted from the diagram by using a general note that states: “All pumps have red and green pilot lights to denote that the pump motors are operating or not operating, respectively.”

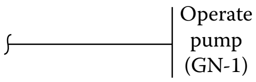


*The words “Filling permissive exists” only exemplify the use of a status summary statement. The statement is not mandatory but may be added if it is helpful. (See section 3.6)

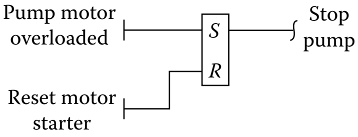
**This information describing solenoid, actuator and port design is required for detailed design work

		Solenoid valve	Control valve	
		HY-1	HV-1	
		HY-2	HV-2	
			Actuator	Port
Operation	Open valve	De-Energized	Vented	Open
	Close valve	Energized	Pressurized	Closed

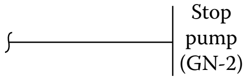
FIG. 5.1c
Tank filling operation-interlock 1 logic diagram.



As another example, the motor lockout detail



will commonly be simplified by referring to a general note that states: "The motor starter locks out when tripped,"
Thus:



2. The memory function that keeps the pumps in operation may be, but is not necessarily, provided by a circuit breaker for the pump motor. The other maintained-memory function in the diagram may be provided by pneumatic or electric latching relays or other types of hardware. This illustrates the essentially hardware-free nature of the operational logic portion of the diagram and the emphasis on logic function.

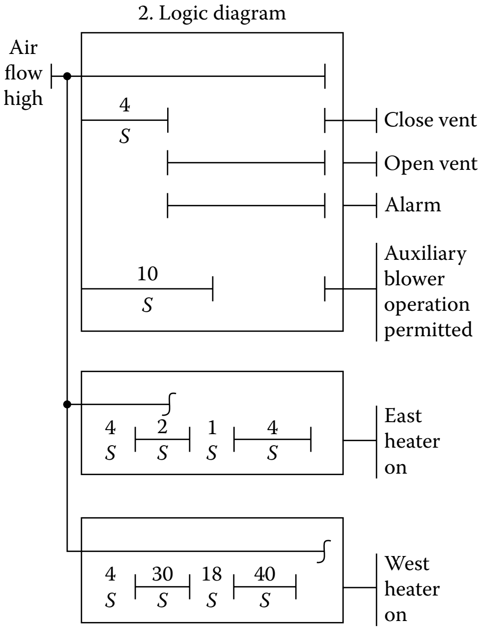


FIG. 5.1d
Logic diagram equivalent of the "word description."

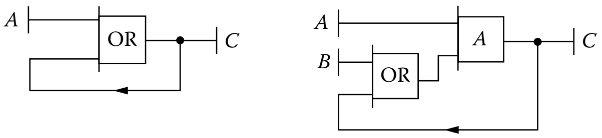
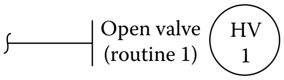
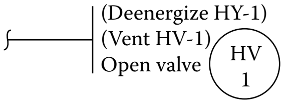


FIG. 5.1e
This symbology is not recommended to be used for depicting memory loss due to power supply failure.

3. The logic diagram emphasizes the operating logic of the process by not detailing the system mechanism for opening and closing the control valves. Thus, this information is provided by means of Routine 1, which may apply to similar hardware of an entire project as well as to Interlock 1. However, if it is desired to make the diagram more self-contained by including hardware functions, this can be done as follows, using an excerpt from the diagram as an example:



Alternative:



APPENDIX B: COMPLEX TIME-ELEMENT EXAMPLE

B.1 Word Description

For an illustration of the logic equivalent of the word description (Figure 5.1d), assume a process operation, as follows:

If air flow becomes high and is so sustained for 4 seconds, then open vent, actuate alarm, and initiate heating by east and west heaters. If heating by east heater is initiated, the heater goes on for 2 seconds, off for 1 second, and on again for 4 seconds, regardless of whether the air flow remains high while this is occurring. If heating by west heater is initiated, then heater goes on for 20 seconds, off for 18 seconds, and on for 40 seconds, but only if the air flow remains high while this is occurring.

If high flow of air is sustained for 10 seconds, stop the auxiliary blower if it is running.

When air flow is no longer high, close the vent and permit the auxiliary blower to be restarted and the alarm to be reset.

APPENDIX C: LOSS OF POWER SUPPLY FOR MEMORY

There are no symbols in the new Table 5.1a from Draft 4 of the new proposed Standard to indicate how to symbolize memories that are lost in the event of loss of power supply. The use of a logic feedback to symbolize a memory is deprecated. Thus, the symbolisms shown in Figure 5.1e shall not be used:

Bibliography

American National Standards Institute Standard Y32.14-1973, "Graphic Symbols for Logic Diagrams (Two-State Devices)," 1973.

American National Standards Institute Standard X3.5-1970, "Flowchart Symbols and Their Usage in Information Processing," 1970.

ANSI/ISA-S5.2-1976 (R 1992), "Binary Logic Diagrams for Process Operations," reaffirmed July 13, 1992.

Draft 4 of ISA Draft 5.01.01, "Instrumentation Symbols and Identification," Research Triangle Park, NC: ISA, 2000.

International Electrotechnical Commission Recommendation, Publication 117-15, "Binary Logic Elements," 1972.

National Electric Manufacturers Association Standard ICS 1-103, "Static Switching Control Devices."

National Fluid Power Association Standard T.3.7.68.2, "Graphic Symbols for Fluidic Devices and Circuits."

5.2 Ladder Diagrams

R. A. GILBERT (1985) **W. P. DURDEN** (1995, 2005)

INTRODUCTION

Ladder diagrams are one of the traditional methods for describing logic control circuits, whether they are electrical, pneumatic, or hydraulic. Ladder diagrams sometimes are referred to as elementary diagrams or schematics. They are called so because of the vertical power rails and the horizontal rungs that are individual parallel circuits between the common power rails.

For the purpose of this discussion electrical symbols will be used. Boolean logic is used in these diagrams to represent switch and relay contacts. The basic ladder diagram symbols that are used in electrical ladder diagrams are described in this section. Also presented are examples that illustrate the connection between the process and the ladder diagram and that indicate the procedure for analyzing a ladder diagram to determine its control functions.

For the standard that describes the binary logic symbols and diagrams recommended by the Instrument Society of America, refer to Section 5.1. The ISA standard compiled for ladder diagrams and other process logic depictions is ISA-5.2-1976 (R1992), "Binary Logic Diagrams for Process Operations."

LADDER DIAGRAM SYMBOLS

Figure 5.2a describes a few of the common symbols that are used in ladder diagrams and will be discussed in this section. The top row of the symbols in this figure include two manually operated pushbuttons (P.B.s) and a selector switch.

In the symbol for the three-position, two-pole selector switch, "X" represents the closed and "O" represents the open state. The "XOO" notation adjacent to the first level (top) of the selector switch represents the states of the switch contact in each of the three selector switch positions. The multiple lines from the top of the switch show the number of positions the switch can take, and the line that is drawn heavily represents the position that corresponds to the state drawn in the figure.

The second row in Figure 5.2a shows a coil, which can be that of a starter, a relay, or a solenoid coil (relays are discussed in Section 5.9). This symbol often includes a letter to describe the type of coil used. Also shown in the second row of the figure are the symbols for a Normally Open (NO) and a Normally Closed (NC) relay contact.

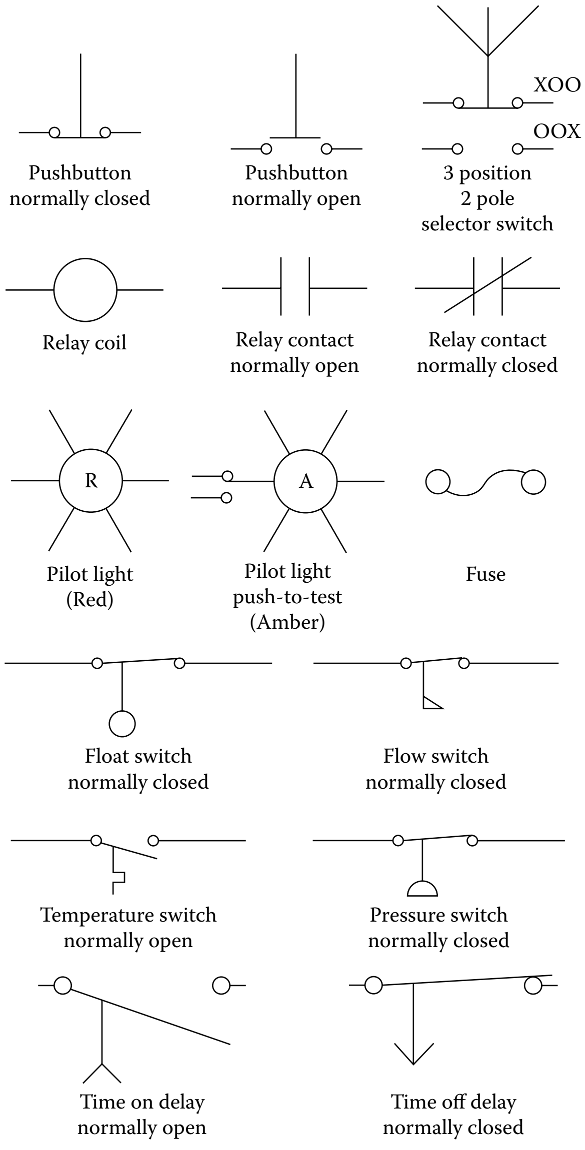


FIG. 5.2a
Sample of symbols used in ladder diagrams.

The third row shows two forms of a pilot or indicator light and a symbol for a fuse. The fourth and fifth rows describe switch contacts of level, flow, temperature, and pressure switches that are activated by the process. The last row depicts two time delay relay contacts, one normally open, the other normally closed. If the time delay is used to delay the turning on of a device, it has an arrow pointing into the contact (left), and if it serves to delay the turning off of a device, it has an arrow pointing away from the contact (right).

Although there are a great number of other ladder diagram symbols, the set presented here is sufficient to enable the reader to understand the drawings presented in this section.

The wires in the ladder diagram should be labeled with unique alphanumeric identifiers. The rule of labeling is that an electrically common point of a circuit that will never be switched or disconnected will carry the same label. A variation of the above rule is to have a single wire number for each continuous circuit, but added to this number is a "subnumber" (-#) that changes at each termination point. For example, in the wire labels *D*1-47-1 and *D*1-47-2, *D*1 is the loop or device number, 47 is the wire number, and 1 and 2 are electrically the same wire but are connected at a terminal block.

DEVELOPING A LADDER DIAGRAM

Figure 5.2b illustrates the automatic control of a pressurized water tank and the ladder diagram of its automatic control circuit. The figure shows a water storage tank, a pressurized water tank, a pump, and some level switches and other devices required to control the system.

The ladder diagram of the system illustrates a control circuit, which has both an automatic and a manual mode of operation. The stop P.B. is usually located near L_1 before any

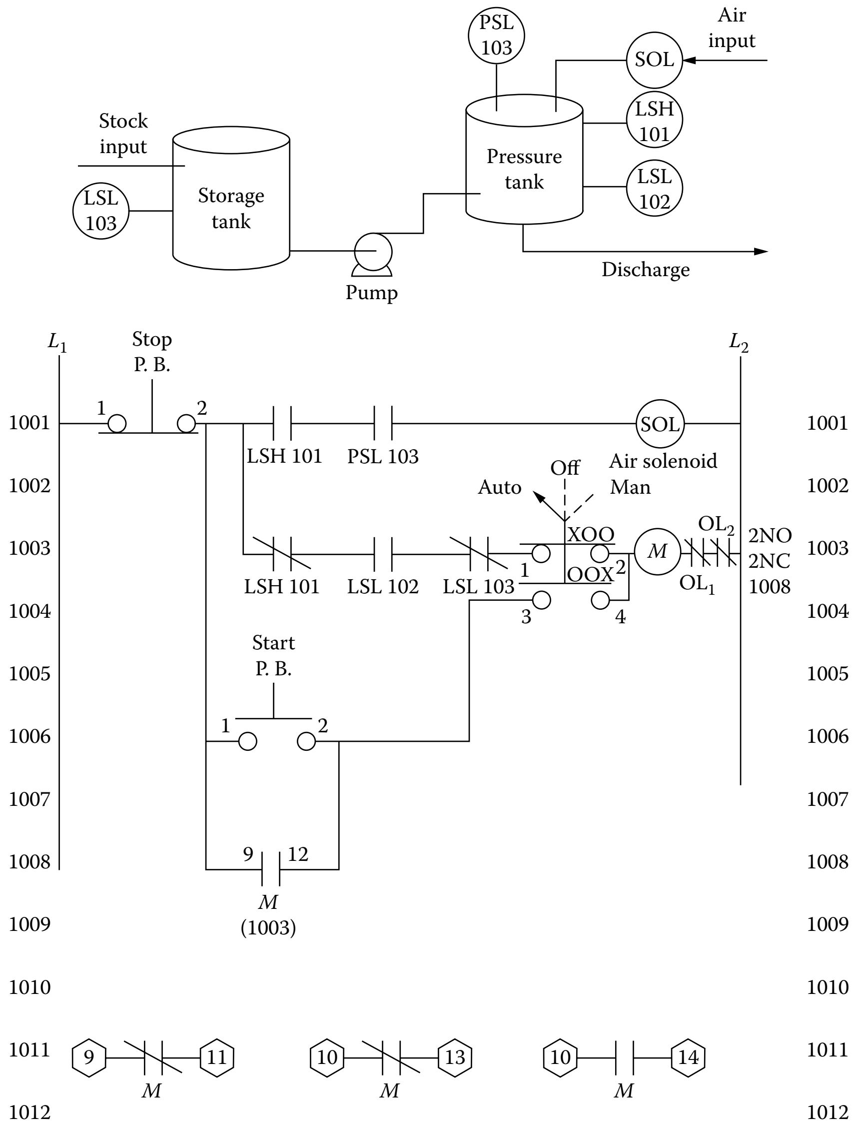


FIG. 5.2b
Automatic control of a pressurized water tank.

branches are added to the drawing. The manual control mode in this instance consists of only a start-and-stop pushbutton and the safety overload protection (OL_1 and OL_2) for the pump. These elements are shown in the ladder diagram and have been arranged so that the electrical connection from L_1 to L_2 will be made when the stop P.B. contact is closed, the start P.B. is pressed (contact closed), and the operating mode selector switch is in the "Auto" position.

L_1, on the left or power side of the ladder diagram, and L_2, on the right or neutral side, represent the power feed to the circuit. The normally open contacts (9 and 12), labeled "*M*," are activated (closed) when the pump relay is energized and keep the pump relay energized (engaged) even after the start P.B. has been released and therefore contacts 1 and 2 are open. This circuit is called a seal-in or latching circuit.

The common terminology for relay contacts is NO (Normally Open or Form *A*) and NC (Normally Closed or Form *B*). *A* Form *C* relay is one where the common power supply is connected to both an NO and an NC set of contacts. Refer to Section 5.9 for a more thorough description.

Automatic Mode of Operation

In the control system there is a high- and a low-level switch (LSH-101 and LSL-102) on the pressurized tank and a low-level switch (LSL-103) on the water storage tank. In addition there is a low-pressure switch (PSL-103) on the pressurized water tank. In preparing the ladder diagram for the corresponding control system, one should make sure that the water pump will be energized only when each of the control constraints have been satisfied and the electrical path from L_1 through the pump starter relay to L_2 has been completed.

The automatic portion of this water control circuit is presented in the top half of the ladder diagram in Figure 5.2b. In this control circuit, the electrical connection between L_1 and L_2 will only be made if the switch contacts shown in the figure are closed.

The high-level switch on the pressure tank (LSH-101) has two sets of contacts: a normally open set to control the operation of the air solenoid and a normally closed set to provide a limit condition for pump operation. The pump can only operate if, in addition to the level not being high (the NC LSH-101 contacts closed), the level is not low in the storage tank (the NC LSL-103 contact is closed) and the level is low in the pressurized tank (the NO LSL-102 contacts are closed). The pump starter coil will, therefore, be energized when the water level is low in the pressurized tank but will be deactivated (the pump will be stopped) when the water level is low in the storage tank.

This configuration will both protect the pump and maintain the pressure in the pressurized tank. To keep the water supply pump running, once the level in the pressurized tank drops to the low level switch (the NO contacts of LSL-102 are momentarily closed), an interlock with a set of pump relay contacts must be provided. The normally open contacts (9 and 12), labeled "M," are activated (closed) when the pump relay is energized, and they keep the pump relay energized (engaged) even after the level in the pressurized tank is no longer low (the NO contact of LSL-102 is open).

In Figure 5.2b the mode selection switch is placed near the starter relay and the normally closed contacts of LSL-103 and LSH-101 are placed in series with the normally open LSL-102 contacts. As it was the case in the manual mode of operation, a set of normally open starter relay contacts (9 and 12) ensure that the starter relay remains energized when the liquid level rises above the minimum set by LSL-102, but remains below the maximum set by LSH-101. Therefore, LSH 101 disengages (de-energizes) the relay when the level in the pressurized tank reaches its high set point.

When the NO contacts of this high water level switch (LSH-101) close, that will energize (open) the air solenoid, if the tank pressure is also below the setting of PSL-103 and therefore the NO contacts of that switch are closed.

SUMMARY

In summary, the steps required in developing a ladder diagram for a process control system include (1) to review the possible process conditions and the responses required, (2) to select the various switches and other control elements that can accomplish these required tasks, and (3) to arrange the system components in a sequential fashion between the voltage lines of the ladder diagram.

When the required response to a momentary condition is to latch in the state of operation, interlock contacts should be provided around start pushbuttons and other momentary contacts. Stop pushbuttons and other safety interlock contacts should be arranged so that they are electrically close to the source voltage line.

Each component of the drawing should be labeled by the same tag number that is used for the corresponding actual device in the process. Particular attention should be paid to identifying all of the contact sets of a specific relay coil. Finally, all automatic contacts of all switches and other actuating devices should be shown in a state (normally open or closed) that will exist when the device is in a de-energized (idle or inactive) state.

LADDER DIAGRAM ANALYSIS

The basic procedure for the analysis of a control circuit ladder diagram is to consider the diagram one component at a time and to decide what will happen if a pushbutton contact is closed or a switch contact is opened. If the ladder diagram is reviewed from this point of view, it will be noted that such contact activity usually closes or opens complete circuits and then a particular coil will be energized or de-energized, depending on the continuity of the circuit.

When the circuit is complete to any particular coil, that contactor, relay, or starter will be energized and the state of its contacts will be the opposite to its normal, de-energized position. If they are normally closed contacts, they will now be open, and if they are normally open, they will be closed.

When a time delay relay is used in the circuit, its contacts are opened or closed after some delay. If relays are used in the circuit, it is important to consider every contact that is operated by that relay whenever the coil is energized. Failure to consider all contact sets of a relay will result in a misconception of the function of the circuit. Finally, when analyzing a circuit, one should be certain that every component is considered in both its de-energized (normal) and energized position so that the whole operation of the complete circuit can be comprehended.

START-UP AND SHUTDOWN

In the design of a complex ladder logic it is very important to consider power-up and power-down sequences. In an electrical circuit the start-up and the shutdown are the times when strange and undesirable conditions can evolve. Typically, on power-up, if a circuit has a normally closed contact, this contact could cause a motor to start before it should, due to the dynamics of the circuit.

On the other hand, if there is a normally open circuit that prevents a device from running, this could prevent the device from ever starting unless manually started the first time. Also, when time delay relays are used, the circuit could be "locked out" until the time has elapsed and the relay has either energized or de-energized. This delay can be either intentional or unintentional.

Dynamic Breaking of a Motor

The ladder diagram of a motor's dynamic breaking circuit is shown in Figure 5.2c. This diagram can be reviewed by using the ladder diagram analysis procedures presented in this section. The example also provides insight into the concept of dynamic braking.

Dynamic braking can be applied to any equipment when a smooth, fast stop is required or when it is desired to prevent the manual rotation of the motor shaft when the power is disconnected. A dynamic braking system provides a stop without any tendency to reverse and it also produces less shock to motor drive components than does plug stop braking.

In a dynamic braking system, a DC voltage is applied to the rotating motor to provide a smooth but positive braking action and to bring the motor to a rapid stop. The DC signal is removed when the motor is almost stopped to prevent any motor winding damage caused by overheating as a result of excessive current flow in the low-resistance windings.

The diode rectifier element that produces the DC voltage is shown in the lower portion of the ladder diagram in Figure 5.2c. This part of the diagram also shows that the rectifier is isolated from motor terminals T_1 and T_3 by two sets of normally open brake coil contacts, *B*. There are also two sets of normally open brake contacts that isolate the primary side of the transformer from power lines L_1 and L_2. The fifth contact controlled by the brake coil is a normally closed set that is in series with the motor relay coil and the start pushbutton.

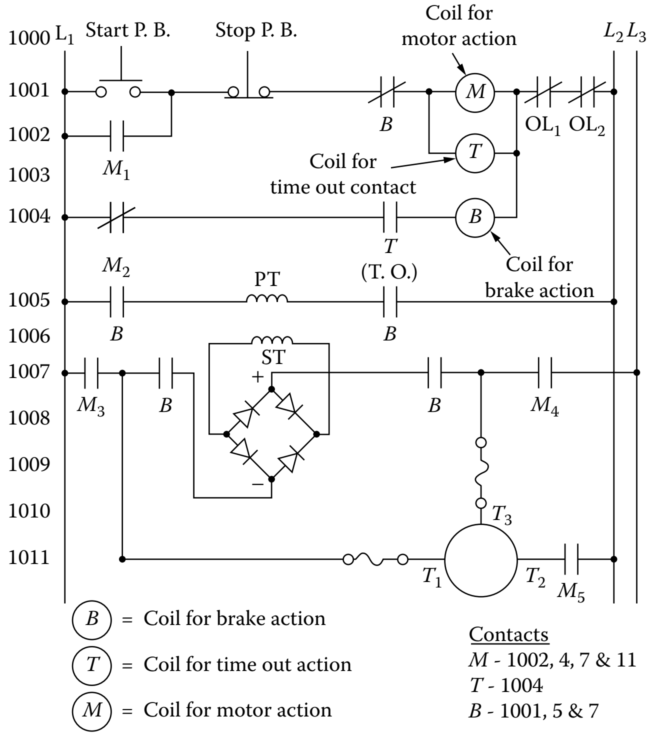


FIG. 5.2c
Ladder of a motor diagram for a dynamic braking circuit.

If the motor start P.B. is depressed (provided that the safety overload contacts are closed) and the stop P.B. is not also depressed, then the motor relay coil is engaged and the five sets of motor relay contacts are activated. The first of these motor relay contacts, M_1, is illustrated in the top horizontal line of the ladder diagram. This normally open contact closes when the coil is energized to guarantee that the coil remains hot even after the start P.B. is released.

The second motor contact set, M_2, which is located on the drawing just below the first, is normally closed but is opened on motor coil activation to prevent the brake coil from receiving the L_1 signal. It should be noted that the normally open time-out contact, *T*, is closed because the time-out relay is also energized when the motor coil is active. The last three normally open motor contact sets are connected to the motor terminals and, when closed, provide a connection from the power lines to the three motor terminals.

The braking portion of the ladder diagram includes symbols for the brake relay, the time-out relay, the rectifier, the primary and secondary transformer elements, and the five brake contact sets. Figure 5.2c indicates that motor terminals T_1 and T_3 are connected to the DC terminals of the rectifier when the stop P.B. is pressed. When the stop P.B. is engaged, the motor relay is deactivated, the motor is isolated from the

AC lines, the time-out relay is dropped from the circuit, and the brake contacts are all activated.

The brake relay is energized because M_2 has returned to its idle state and the time-out contact, *T*, remains closed for the time-out period. Thus, the activated brake coil keeps the four normally open *B* contacts associated with the transformer and the rectifier closed until the time delay relay has timed out to force *T* to its normally open position. The actual time-out period is determined as a function of the time needed to bring the motor to a near stop.

FAIL-SAFE DESIGN

Another consideration in ladder design is the "what-if" aspect. What if a wire breaks or vibrates loose? Will an alarm condition occur that will go unnoticed because the alarm indicator never receives power, or will the alarm indicator become energized as a "default"? This part of the ladder diagram design is one of the most difficult to grasp. It can become a tail chasing the dog problem.

The general guiding rule is this question, "Which situation is worse, more catastrophic, or potentially life treating?" If an alarm is initiated when either the process is in alarm or when the alarm circuit has failed, that would be better than having a pumps seal water flow switch go into alarm but not realizing it until the motor bearings overheat and seize. This may require an additional relay, or a time delay relay, but the protection of safety and capital equipment will normally far outweigh any cost increase.

DOCUMENTATION

As in any design, good documentation is important. Several things can be done to allow for easier troubleshooting and for thorough documenting of design. Simple ladder diagrams as shown in the examples do not require any additional steps, but complicated diagrams do. The first thing that can be done is to run a number string down the side of the neutral and power side of the diagram with a number corresponding to each line space, whether a rung exists there or not. In other words, if rung spacing is every 0.75 in., the spacing (placement) of the numbers would also be 0.75 in.

To use these numbers, make note of the location of the use of each contact of a relay beside the coil location. This allows easier analysis of the logical use of the coil's activation. Second, it is good practice to note how many normally open and normally closed contacts are available on all coils. This enables the designer to know if there are any spare contacts available for modifications to the design.

Third, mark the terminal numbers for each contact or coil as well as any unused contacts. The unused contacts should be shown at the bottom of the rung column where the coil is used. This completely documents the circuit and the use of the components in it. Uniquely label each electrically common point in the diagram. These labels will be used on the wires in the physical circuit.

CONCLUSIONS

In summary, the proper procedure for ladder diagram analysis is to consider the circuit one component at a time and to decide what happens to the component when a contact is closed. In this process, one should determine the function of each component in its de-energized (normal) and energized positions and, finally, one should review the role of each component in relation to the other elements in the system.

It is important to perform a complete analysis of the diagram without jumping to conclusions halfway through the analysis. A hurried analysis can be disastrous, because the action of just one additional contact set can easily change the response or the basic nature of the circuit.

Bibliography

American National Standards Institute Standard X3.5-1970, "Flowchart Symbols and Their Usage in Information Processing," 1970.

American National Standards Institute Standard Y32.14-1973, "Graphic Symbols for Logic Diagrams (Two-State Devices)," 1973.

Basics of Ladder Logic, http://techonline.com/community/home/32150, December 2003.

"Binary Logic Elements," International Electromechanical Commission Recommendation, Publication 117-15, 1972.

"Electrical Diagrams and Schematics," *Department of Energy Fundamentals Handbook, Volume 1, Engineering Symbology, Prints, Drawings,* DOE-HDBK-11016/1-93, January 1993.

Ellermeyer, W., "Manipulate Logic without Boolean Algebra," *Control Engineering,* December 1969.

Kuphaldt, T. R., *Lessons in Electric Circuits, Volume IV (Digital),* Chapter 6, http://www.faqs.org/dos/electric/digital/DIGI_6.html, 2000–2003.

McIntyre, R. L., *Electric Motor Control Fundamentals,* New York: McGraw-Hill, 1960.

National Electric Manufacturers Association Standard ICS 1-102, "Graphic Symbols for Logic Diagrams."

National Electric Manufacturers Association Standard ICS 1-103, "Static Switching Control Devices."

National Fluid Power Association Standard T.3.7.68.2, "Graphic Symbols for Fluidic Devices and Circuits."

Pelloso, P., *Practical Digital Electronics,* New York: John Wiley, 1986.

Platt, G., "Describing Interlock Sequences," *ISA Journal,* January 1966.

Porat, D. L. and Arpad, B., *Introduction to Digital Techniques,* 2nd edition, New York: John Wiley, 1987.

Strong, J. A., *Basic Digital Electronics,* London: Chapman and Hall, 1991.

5.3 Optimization of Logic Circuits

P. M. KINTNER (1970) **R. A. GILBERT** (1985, 1995, 2005)

Instrumentation and controls is an engineering discipline, expertise, and practice that relies heavily on information presented in P&ID function, and circuit diagrams. An interesting but challenging aspect of function and circuit diagrams is the fact that the logic circuits depicted have often been optimized. Process control loop element cost and assembly drive this optimization. Since logic simplification always means less hardware, circuits that are less expensive and more reliable, and simpler circuits for troubleshooting, new control scheme circuit designers are encouraged to expand their knowledge in this area beyond the introductory material presented here.

Extensive optimization accents the need for two skills. Today, most personnel involved in instrumentation and controls comprehend optimized function diagrams and, to some extent, create optimized logic circuits. This section presents the building blocks for the optimization of basic function diagrams. The section begins with a brief review of logic elements and the impact of the inversion operator on logic functions.

Classically, the design and optimization of logic circuits have centered on the synthesis of switching element (static or relay-type) interconnections that will satisfy the switching functions required by the process. With the development of computer control there is an increased need for instrument and control engineers who understand and can apply logic techniques for process control. In addition, new computer-aided design software produces function diagrams equivalent but not visually similar to the initial design idea. The ability to recognize this equivalency is essential to avoid lost time, energy, and money altering something that does not need to be altered. This awareness begins with a sound appreciation of the fundamental building blocks.

OPTIMIZATION BUILDING BLOCKS

The basic logic components and operations are straightforward and have their origins in ancient Greek philosophy. Modern engineering uses of logic expand the application of the AND, NAND, OR, and NOR operations to decision making—e.g., whether equipment is to be turned on or off, a voltage level is to be high or low, or a valve is to be opened or closed. Figure 5.3a illustrates these four cornerstone operations with respective truth tables.

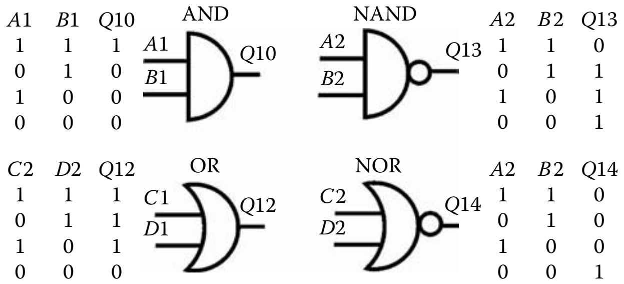


FIG. 5.3a
Summary of fundamental logic operations.

Equation 5.3(1) is a succinct statement of a control scheme that only permits a backup pump to engage if both the liquid level and tank pressure are below sensor alarm values.

$$Z0 = 1 \text{ only if LSL} = 1 \text{ and PSL} = 1 \qquad \textbf{5.3(1)}$$

This equation presents logic signals or logic variables, such as the ones into or out of logic switching elements, as alphanumeric characters with possible values of "1" or "0." In this example, *Z*0 represents the pump, and LSL and PSL stand for level sense low and pressure sense low, respectively, and symbolize the low-level and low-pressure alarm status within the tank. Thus, LSL = 1 indicates the liquid has dropped below the low-level limit while PSL = 0 indicates that the pressure within the tank is above an established minimal value.

The AND symbol with its truth table in the upper left side of the Figure 5.3a also summarizes the idea presented in Equation 5.3(1). The reader must be comfortable with that fact and also recognize the unique properties of the other symbols in the figure. For example, the NOR is the only logic operation that provides a logic "0" when any of its inputs are at logic "1." The unique output situation for the NAND, also logic "0," occurs when all of its inputs are high, while the OR only goes to logic "0" when all its inputs are at logic "0."

Figure 5.3b provides a nonoptimized function diagram for an application that involves a timing requirement. The diagram presents two AND devices, one NAND device, and a 555 pulse-generating device. The circuit also has two inverter devices, shown as triangles with circles attached. Although the discussion of this diagram is delayed, it does provide a segue to review the properties of the inverter.

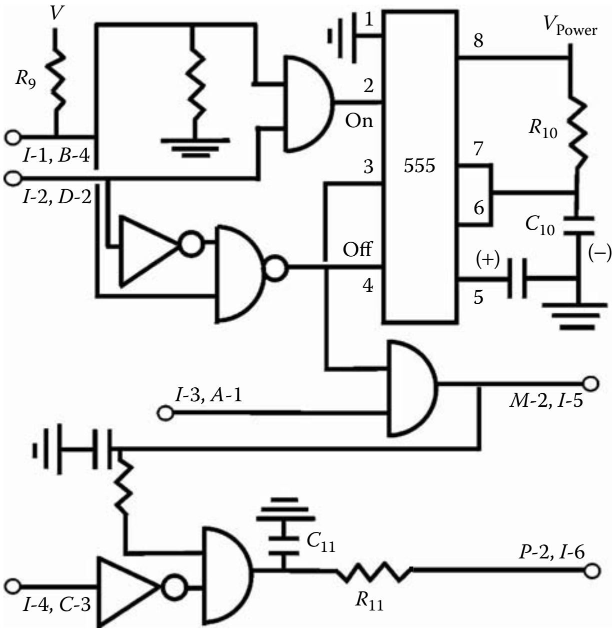
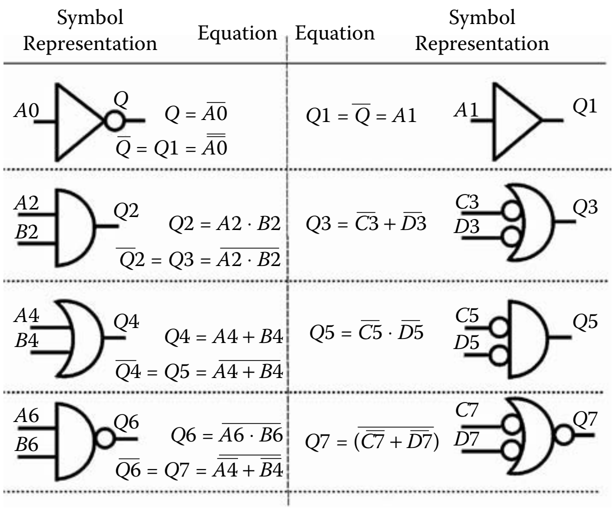


FIG. 5.3b
Nonoptimized logic circuit application.

The inverter is a circuit element that performs the inversion operation. This mathematical operator is often known as the NOT, and the top left section of Figure 5.3c provides its symbolic as well as mathematical representation. The value of the input variable, $A0$ in this case, is inverted and expressed as the variable Q. The line, or bar, over the $A0$ is a visual reminder that if $A0$ is at logic "1," then output Q will be at logic "0." Thus, the NOT operation is merely the inverse of the current value of the input variable; i.e., if $A0 = 0$, then Q will equal 1 after the NOT operation.



Q1 is the inverse of Q
Q5 is the inverse of Q4

Q3 is the inverse of Q2
Q7 is the inverse of Q6

FIG. 5.3c
Logic function negation examples.

The inversion operation can also be performed on sets or subsets of logic variables. For example,

$$Z1 = \overline{(A \cdot B) + (C \cdot D)} \qquad \textbf{5.3(2)}$$

is an extremely succinct method of stating that two outputs from two AND operations based on variables (A, B and C, D, respectively) are the inputs to an OR operation; the results of this are inverted to provide a value for the variable $Z1$. The reader should confirm that Equation 5.3(2) is fundamentally a NOR idea because the only way $Z1$ can be at logic "1" is when the results of both AND operations are "0."

The mathematics of logic circuit optimization involves the use of the inversion (negation) operation. Figure 5.3c summarizes the results when the negation (inversion) operation is performed on a set of basic logic operations. The figure is vertically divided, with the results of the negation shown to the right of the dashed line. For example, the symbol and equation for an inverter are presented as the top left entry in the diagram. If a negation is performed on output variable Q, the resulting symbol, a triangle without an attached circle, and the equality, $Q1 = A1$, are shown on the right. Now $Q1$ also equals $A0$ and $Q1$ is the inverse of Q. Although this is an obvious result, the other examples suggest why inversion can appear to be confusing.

The symbol and equation on the bottom left side of Figure 5.3c represent the NAND operator. A review of the AND and NAND truth tables in the top row of Figure 5.3a indicates that the inverse of the NAND is an AND. However, the symbol in the bottom right section of Figure 5.3c does not match the AND symbol in Figure 5.3a. This is not a mistake, discrepancy, or deception but recognition that even though there is more than one way to symbolize the results of an inversion, the truth table results are identical. The value of $Q7$ is the inverse of $Q6$. Likewise, the symbol to the right of the AND symbol in the second row of Figure 5.3c is equivalent to a NAND symbol, and the value of $Q3$ is always the inverse of $Q2$. The reader should develop the truth table for the equation for $Q5$ on the right side of the third row of the figure to confirm that the complex logic symbol associated with $Q5$ also represents a NOR, with its equivalent truth table presented in Figure 5.3a.

Figure 5.3d summarizes the results if an additional negation is performed.

The operational principle is the fact that two cascaded inversions cancel, because this amounts to a double negation. Stated in symbolic form

$$\overline{\overline{A}} = A \qquad \textbf{5.3(3)}$$

where it is indicated that the logic signal A has been inverted twice, returning it to its original value. The progression of

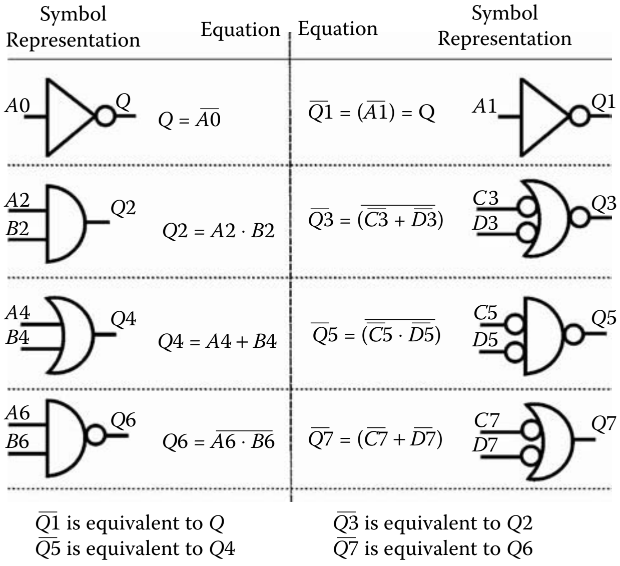


FIG. 5.3d

Examples of cascaded negation logic operations (DeMorgan equivalent circuits).

negations from left to right in the first row of Figure 5.3c and then in the right side of the first row of Figure 5.3d also illustrates this double negation effect. The inverter output variable, $Q1$, shown in Figure 5.3c is the inverse of Q, while the subsequent inversion of $Q1$ in Figure 5.3d returns the value of $Q1$ to the original Q value state. Following the same presentation pattern, the reader should confirm that the remaining three logic functions illustrated on the right side of Figure 5.3d are identical to the respective AND, OR, and NAND in the left portion of the corresponding row of Figures 5.3c and 5.3d.

As an instructional example of how a double inversion cascade is performed, consider Equation 5.3(4).

$$Z2 = (A \cdot B) + (C \cdot D) \qquad \textbf{5.3(4)}$$

A single inversion of Equation 5.3(4) can certainly be represented as Equation 5.3(2), where $Z1$ is the inverse of $Z2$. However, the expression on the right side of Equation 5.3(2) is not the only possible representation. To develop an addition equivalent logic statement, the inversion of every input signal and simultaneous interchange of the AND and OR operations (DeMorgan's theorem) is performed on Equation 5.3(2). Equation 5.3(5) summarizes these operations as

$$Z1 = \overline{Z2} = (\bar{A} + \bar{B}) \cdot (\bar{C} + \bar{D}) \qquad \textbf{5.3(5)}$$

Following this same set of rules, the second inversion is performed on $Z1$ to return to Equation 5.3(4). In this case, the individual inversion symbols above variables A, B, C, and D are removed, and the OR operation within each parenthesis becomes an AND operation, while the AND operation that separates the parentheses now becomes an OR operation.

This cascade inversion process can be used to obtain equivalent forms of both the NOR operation

$$\overline{A + B} = \bar{A} \cdot \bar{B} \qquad \textbf{5.3(6)}$$

and the NAND operation

$$\overline{A \cdot B} = \overline{\overline{\bar{A} + \bar{B}}} = \bar{A} + \bar{B} \qquad \textbf{5.3(7)}$$

The respective left and right sides of Equation 5.3(6) and Equation 5.3(7) are called DeMorgan equivalents. The right side of Figure 5.3d shows the DeMorgan equivalent symbols for the NAND as well as the OR, AND, and the inverter. Because the inverter only has a single input, the DeMorgan equivalent for the inverter is sometimes depicted as a triangle with the circle moved to its input side. Although the DeMorgan equivalent for the NOR is not included in Figure 5.3d, the reader, after re-examination of Figure 5.3c, should realize that the NOR equivalent symbol is associated with variable $Q5$.

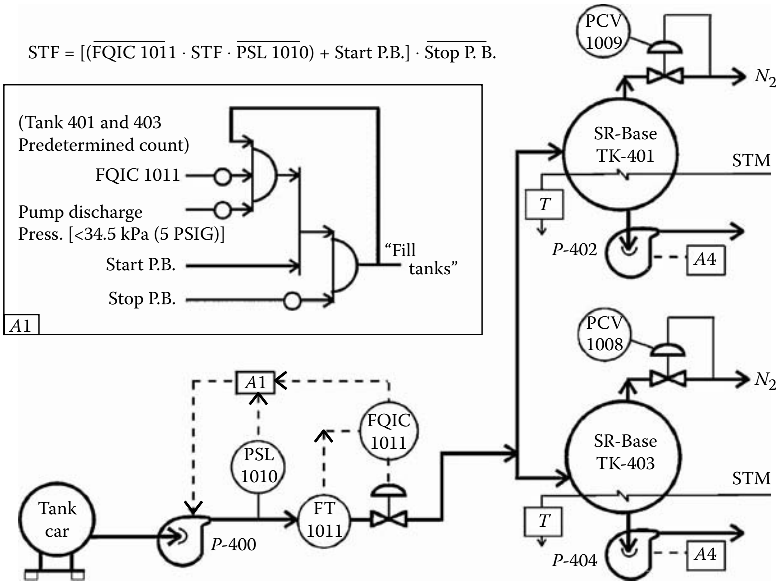
Graphic Logic Functions

Control logic schemes are mathematical relationships between input and output variables that can be represented graphically through logic diagrams. Figure 5.3e illustrates a chemical storage and filling station. The figure provides a description of the tank filling process from three distinct perspectives. The equation at the top of the diagram connects the storage tank fill (STF) operation to all of the variables affecting that filling step. The P&ID diagram provides the visual details of the interconnections among the equipment, sensors, and actuators. The panel in the center left portion of the figure presents the function diagram that emphasizes the seal-in action included in the control scheme.

A good way to appreciate the utility of logic diagrams and the complication optimization can create is to review Figure 5.3e. The example illustrates the tank section for a chemical process. The drawing shows three pumps (equipment 400, 402, and 404), a low-pressure sensor (PSL 1010), two pressure control valves (PCV 1008 and 1009), and a flow quantity indicator and controller (FQIC 1011). For simplicity, consider the filling operation for this system because it involves only one pump and two control elements, PSL 1010 and FQIC 1011. In this example it will be assumed that the logic function for this part of the operation is

$$\begin{aligned} STF = {} & ((\overline{FQIC\,1011} \cdot STF \cdot \overline{PSL\,1010}) \\ & + START\ P.B.) \cdot \overline{STOP\ P.B.} \end{aligned} \qquad \textbf{5.3(8)}$$

where START P.B. and STOP P.B. are the variables assigned to the start and stop pushbuttons in the system and STF is the variable associated with the operation of the pump. STF



$$STF = [(\overline{FQIC\ 1011} \cdot STF \cdot \overline{PSL\ 1010}) + Start\ P.B.] \cdot \overline{Stop\ P.\ B.}$$

FIG. 5.3e
Chemical storage and tank filling station.

equals "1" when the pump is on and has a value of "0" when the pump is off. Although Equation 5.3(8) is accurate, it is difficult to comprehend, particularly because the output variable, STF, appears on both sides of the equation. Such implicit logic functions are common and are best interpreted by use of a logic diagram.

The panel in Figure 5.3e is the logic diagram for the storage tank fill system. The drawing uses a second graphic symbol convention common to industrial applications. A three-input AND, a two-input AND, a two-input OR, and three inversion operations are illustrated in the figure. The inversions are represented by the three circles; the OR symbol is the short vertical line in the center of the drawing.

As suggested above, this logic diagram is not as complex looking as Equation 5.3(8). Variable STF is present twice in the equation because of the safety interlock for the stop pushbutton. In addition, the diagram clearly shows that for this specific example, once the pump is running, the pump will not continue to function if the pressure sensor is in alarm OR the predetermined count has been obtained OR the stop action has been initiated.

Ladder Diagrams from Logic Diagrams

Although some form of ladder logic language is often used to program a programmable logic controller (see Section 5.2), the logic diagram is still a very popular way to illustrate the target control scheme. Figure 5.3f presents a logic diagram and ladder program for the pump station illustrated in Figure 5.3e. A review of these two drawings illuminates the procedure for interchanging ladder diagram elements with logic diagram elements.

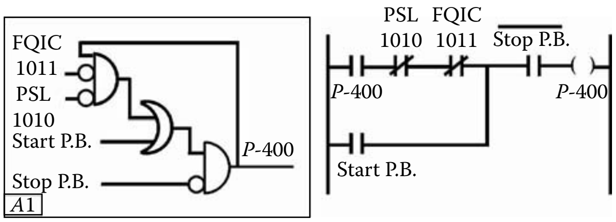


FIG. 5.3f
Function diagram and ladder logic program for tank filling station.

The three features of a ladder diagram to be considered are the serial, parallel, and interlock portions. The serially arranged elements in the ladder diagram are used as inputs to an AND operation in the logic diagram. The parallel components of ladder diagrams become inputs to OR operations in logic diagrams. Thus, for the present example, the serially arranged elements (FQIC 1011, PSL 1010, and P-400) in the ladder program reflect the three inputs to the AND operation shown in the top left section of plate *A*1, while the START P.B. in the parallel rung reflects the OR operation. Finally, the interlock or seal-in is accomplished by associating the ladder program output variable, P-400, with the AND operation.

OPTIMIZED LOGIC CIRCUIT CONSTRUCTION

In general, logic diagrams are useful tools for the instrument engineer to use in debugging logic functions and circuits. These function diagrams indicate the role of the digital variables and the priority each variable may have. In Figure 5.3g,

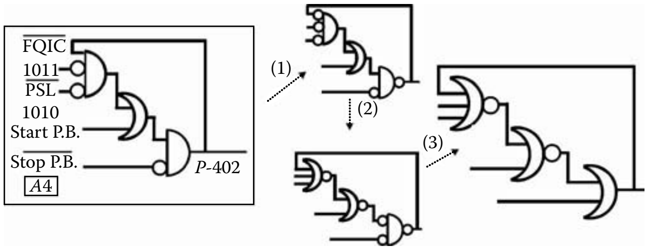


FIG. 5.3g
Visualization of circuit optimization process.

for example, it is clear that the stop action shown in plate A4 has the highest priority over the other three input variables depicted. This same message is shown in the ladder diagram as well. The stop action will occur no matter what the logic states of the other variables are. The confusion might develop when the function diagram reflects the optimized logic circuit instead of an original control scheme motivated diagram. The former diagram is accurate and useful from an electrical engineering standpoint but might not easily reflect the intended control function of the circuit.

This situation is demonstrated with two examples. In the first example, the function diagram on the side of Figure 5.3g reflects a control scheme for pump 402 that is graphically transformed into an optimized form, shown on the right side of the figure. The initial diagram includes two AND devices, an OR device, and three inverters, while the optimized version of the same control circuit is constructed from three OR devices and two inverters. The actual circuit would be easily constructed from the optimized function diagram but the recognition that the stop action overrides all the other variables is not immediately obvious. The second example returns to Figure 5.3b. The details for this circuit are provided in Section 5.10, and the optimized version of this circuit is presented in that section as Figure 5.10g. Again, the optimization focus was to reduce the number of devices needed to accomplish the task. In this case, the control circuit is TTL based (see Section 5.10), and it can be constructed with two TTL packages, 7400 and 7404, in its optimized form instead of three packages, 7408, 7400, and 7404, as dictated by the function diagram presented in Figure 5.3b.

Logic Circuit Synthesis

The optimization of a complex logic function demands mastery of logic circuit synthesis methods. Comfort level with and interpretation of these complex logic functions require knowledge and use of these same tools. Maxterm, minterm, Boolean algebra, and logic maps are popular examples of logic circuit synthesis tools.

The minterm and maxterm are two special functions used in logic circuit development and analysis. As a facilitating example, consider a complex logic function that has the truth table summarized as Equation 5.3(9). As this truth table suggests, logic "1" is the

LSH	*PSH*	*Z*3
1	1	0
0	1	0
0	0	0
1	0	1

5.3(9)

unique output for the logic function but this output is not the result of an AND or an OR logic function. It will be some combination of the AND, OR, and inverter logic operations such that when LSH and PSH are both active the output of the logic function is at logic "1."

The minterm for this system is developed from the row in the table in which there is a "1" in the output column. In general, a minterm is generated in a two-step fashion. First, an AND operation links the elements on the row where the output "1" appeared. Second, all of the inputs in this row that have "0" indicate the variables that have a NOT operation applied to them. For this example, the last row of the truth table is the only minterm; it is defined by Equation 5.3(10):

$$Z3 = \overline{PSH} \cdot LSH \qquad \textbf{5.3(10)}$$

Note that the NOT operation is applied to variable PSH and that this variable has a value of "0" in the last row of the table.

The synthesis of complex functions using the maxterm is a process that is the inverse of that for the minterm. An OR operation is used, and those inputs with a "1" in the row where the "0" output appears have the NOT operation applied to them. When more than one maxterm exists, the logic function is synthesized through AND connections among these maxterms. (If multiple minterms were present, the complex logic function would be created by OR connections among the minterm contributions.) Thus, Equation 5.3(9) can also be expressed by the following logic function:

$$Z3 = (\overline{LSH} + \overline{PSH}) \cdot (LSH + \overline{PSH}) \cdot (LSH + PSH) \qquad \textbf{5.3(11)}$$

There are three separate maxterms in this equation, and they correspond to the first three rows of the truth table shown as Equation (tabulation) 5.3(9).

The synthesis of a complex logic function can also be accomplished using NAND operations that are based on the minterm procedure explained earlier. The procedure is to construct the minterm function as before and then to replace both the OR and the AND operations by NAND operations. For logic function synthesis with NOR operations, the maxterms are used, then the NOR operation replaces both the OR and the AND operations.

As an example of function synthesis with NAND operations, consider an alternate control scheme that still uses the

LSH and PSH limit sensors. This modification is summarized as follows:

LSH	*PSH*	*Z5*
1	1	0
0	1	0
0	0	1
1	0	1

5.3(12)

where the third row of the table now reflects the modification in the desired operation of the final control element associated with Z5; i.e., this element is now also active when LSH = 0 and PSH = 0. The logic function for this new system is

$$Z5 = (\overline{LSH} \cdot \overline{PSH}) + (LSH \cdot \overline{PSH}) \qquad \textbf{5.3(13)}$$

where the first AND term represents the new minterm in the system. Now, following the procedures stated previously, the AND operations $(LSH \cdot PSH)$ and $(LSH \cdot PSH)$ are replaced by NAND operations. The OR operation is then also replaced by a NAND operation, to yield the final NAND synthesized function

$$Z5 = \overline{\overline{(\overline{LSH} \cdot \overline{PSH})} \cdot \overline{(LSH \cdot \overline{PSH})}} \qquad \textbf{5.3(14)}$$

Equation 5.3(14) can be manipulated further, with one option being an overall NAND of two OR functions, one of which is just $(LSH + PSH)$.

Logic Simplification with Boolean Algebra

Boolean algebra is the class of mathematics used to manipulate logic expressions. The maxterm and minterm circuit synthesis methods outlined above are based on Boolean algebra theorems. A few of these theorems are given below as examples. These particular theorems are easy to remember, because they parallel familiar arithmetic and conventional algebraic operations. They are also important to remember because they represent common algebra manipulations used to construct or comprehend optimized logic circuits.

The following are some examples of Boolean algebra equivalents:

$$1 \cdot A = A \qquad \textbf{5.3(15)}$$

$$0 \cdot A = 0 \qquad \textbf{5.3(16)}$$

$$1 + A = 1 \qquad \textbf{5.3(17)}$$

$$0 + A = A \qquad \textbf{5.3(18)}$$

$$\bar{A} + A = 1 \qquad \textbf{5.3(19)}$$

$$A + A = A \qquad \textbf{5.3(20)}$$

$$A \cdot B + A \cdot C = A \cdot (B + C) \qquad \textbf{5.3(21)}$$

To show the use of these relationships for logic simplification, consider the three input variables, $A4$, $B4$, C_4, and one output variable, $Z4$, expression defined by Equation (tabulation) 5.3(22):

$A4$	0101	0101
$B4$	0011	0011
C_4	0000	1111
$Z4$	0001	0001

5.3(22)

The numerical pattern arrangement of the possible input variable values in the table is by design. The numerical value for the column sum of the binary entries in each column of the table correspond to the numbers 0, 1, 2, 3, 4, 5, 6, and 7 when the top row, the $A4$ row, in the table is assigned the least significant bit place. Thus, the eight columns in the table clearly include all the possible three variable value patterns. Using the $Z4$ row of the table, the minterm expression for Equation 5.3(22) is

$$Z4 = A4 \cdot B4 \cdot \bar{C}_4 + A4 \cdot B4 \cdot C_4 \qquad \textbf{5.3(23)}$$

where the first and second logic functions are derived from column 4 and column 7, respectively. This term is examined to determine if it has an algebraic equivalent with fewer variables. For this illustrative example the sequence of steps is brief and to the point. If Equation 5.3(23) is "factored," Equation 5.3(24) results

$$Z4 = A4 \cdot B4(C + \bar{C}_4) \qquad \textbf{5.3(24)}$$

and Equation 5.3(24) then becomes

$$Z4 = A4 \cdot B4 \cdot 1 \qquad \textbf{5.3(25)}$$

It is clear that the expression for $Z4$ in Equation 5.3(25) is mathematically equivalent to the expression for $Z4$ in Equation 5.3(23) and that variable C is not needed.

Logic Simplification through Logic Maps

The synthesis of logic functions based on minterms or maxterms often produces functions that are unnecessarily complex. The comparison of the two equivalent expressions, Equation 5.3(10) and Equation 5.3(11), dramatically demonstrates this point. In this example, the minterm synthesis was more efficient, but in many cases the designer does not intuitively select the better synthesis route. Another problem common to Equation 5.3(11) and other complex functions is the possible presence of extra logic elements.

The existence of unnecessary inputs in a logic expression is called redundancy. Some cases of logic redundancy are easily seen by inspection of the logic function or logic table, as in the example just given. However, many cases of redundancy are difficult to recognize. A common tool used to find redundant terms is a method known as the logic map, or the Karnough map. The logic map consists of an array with entries corresponding to elements from a logic table.

As an example, consider the logic circuit described by Equation (tabulation) 5.3(22). For a three-input system, the table is first rearranged as presented in Equation (tabulation) 5.3(26):

$A4$	0011	0011
$B4$	0110	0110
C_4	0000	1111
$Z4$	0001	0001

5.3(26)

The numerical value for the column binary sums in Equation 5.3(26) are, from left to right, 0, 2, 3, 1, 4, 6, 7, and 5, respectively. The rules for the use of this pattern, or map, can be interpreted such that any detected square or rectangular pattern of logic "1" values indicates a possible logic variable redundancy. In this case, the redundancy can be removed if the variable C is removed from the set. Naturally, Karnough mapping has more table arrangement possibilities than shown above; however, computer programs are now available that will map all the possible column arrangements for any logic table size and identify the variables that are redundant candidates.

NEGATIVE VS. POSITIVE LOGIC USAGE

The graphic presentation of complex logic functions that describe the control of an industrial process is an essential aspect of a designer's responsibilities. Large process logic control schemes cannot be understood or maintained by the engineering support group unless a graphic version of the logic circuit is supplied. Once the designer has developed the logic circuit, removed the redundancies, and constructed the optimized circuit, a decision must be made to present the circuit graphically with positive or negative logic. The choice of logic is based on the purpose of the circuit and the general desire to explain the circuit graphically in a clear, concise fashion.

A positive logic circuit is one that illustrates logic elements that output "1"s as a result of the input of "1"s into the element. An AND is an example of a positive logic element. The only way an AND operation produces a "1" is when all of its input variables are "1." Logic functions described with AND as well as OR operations are considered positive logic functions. Conversely, negative logic circuits illustrate a logic function by using logic elements that require logic "0" values at inputs to produce logic "0" values at their outputs.

The distinction between and significance of positive and negative logic may be initially difficult to understand. The second row of Figure 5.3d presents one graphic example of positive and negative logic elements. As stated earlier, these two logic functions have the same logic table and illustrate the AND operation. It should also be clear from previous discussions that $Q2$ has the unique value of "1" only if $A2$ and $B2$ also have a value of "1." (All other combinations of values for $A2$ and $B2$ result in a value of "0.")

The values of $Q3$ do not seem as simple to understand when viewed from a positive logic perspective. The addition of the three inversion circles complicates our thought process when we attempt to determine the output logic of the drawing associated with $Q3$. However, this logic function is easily interpreted from a negative logic perspective; i.e., output variable $Q3$ has a value of "0" when either input C3 or input D3 has a value of "0." (The remaining combination of values for C3 and D3 results in a value of "1.")

The merit to negative logic is that it facilitates the creation of control circuits for many real systems. For example, many alarm sensors are classified as active low sensors because they provide a low electrical signal when they indicate an alarm situation. Under these circumstances, it is convenient to assign the logic value of "0" to the sensor output when it provides a low electrical signal. Stop pushbuttons are another example of an active low input. If a stop button is pushed it provides the low signal while the button is engaged. Plate $A4$ in Figure 5.3g reflects this negative logic point of view and uses a bar over the appropriate variables to indicate an active low characteristic of the sensor or contact set in question.

SUMMARY

In any process environment, a major fraction of instrumentation and control personnel time and energy is spent keeping the process running. The quality of the logic circuit documentation and the ease in which it is interpreted contribute significantly to the reduction of the time spent on this task. Most process engineers develop control schemes but do not necessarily design control circuits.

However, they must be able to comprehend the control logic circuits that are provided, especially if the circuit has been optimized. To this extent, knowledge of circuit optimization and the various logic functions available to do the same logic task must be appreciated. Although this section has presented the fundamental tools used to create or comprehend optimized logic circuits, it is anticipated that most readers will most often use the skills for logic circuit comprehension.

Virtual optimization is one example where optimization skill is required even though no circuit is physically constructed. In these situations large logic control schemes are to be installed in a programmable controller or process computer. The optimization is performed to reduce the number of instructions, improve scan time, and diminish "debugging" efforts. In these situations documentation is extremely important because the final virtual circuit seldom has the appearance of the original logic scheme. Thus, it is a very good practice to provide function diagrams of both the virtual and original versions of the logic scheme.

Bibliography

Ellermeyer, W., “Manipulate Logic without Boolean Algebra,” *Control Engineering,* December 1969.

Grace, J. D., “Graphic Batch Processing Language Simplified,” *Instruments and Control Systems,* December 1975.

Harrison, H. L., “Optimize Switching Circuits Using Karnough Maps,” *Control Engineering,* August 1969.

Jain, A. K., Bolton, R. J., and Abd-El-Barr, H., “CMOS Multiple-Valued Logic Design-Part I: Circuit Implementation,” *IEEE Transactions on Circuits and Systems,* Vol. 40, 1993.

Jain, A. K., Bolton, R. J., and Abd-El-Barr, H., “CMOS Multiple-Valued Logic Design-Part II: Function Realization,” *IEEE Transactions on Circuits and Systems,* Vol. 40, 1993.

McAloon, K. and Tretkoff, C., *Optimization and Computational Logic,* New York: John Wiley & Sons, 1996.

Mergler, H., “The Key to Optimum Logic Design,” *Control Engineering,* January 1967.

Pelloso, P., *Practical Digital Electronics,* New York: John Wiley & Sons, 1986.

Porat, D. L. and Arpad, B., *Introduction to Digital Techniques,* 2nd edition, New York: John Wiley & Sons, 1987.

Sandige, R. S., *Digital Concepts Using Standard Integrated Circuits,* New York: McGraw-Hill, 1978.

Sasao, T. (Ed.), *Logic Synthesis and Optimization, Kluwer International Series in Science*, Vol. 212, 2002.

Stoffels, R. E. “Shift Registers and Stepping Switches for Sequence Control,” *Control Engineering,* January 1967.

Strong, J. A., *Basic Digital Electronics,* London: Chapman and Hall, 1991.

Williams, G. E., *Digital Technology,* Chicago: SRA, 1982.

5.4 PLCs: Programmable Logic Controllers

W. N. CLARE, G. T. KAPLAN, D. R. SADLON, A. C. WIKTOROWICZ (1985)

R. A. GILBERT (1995) **C. W. WENDT** (2005)

Reviewed by B. G. Lipták (1995)

Types of Input/Output (I/O):	*Discrete I/O: 120 VAC, 220 VAC, 0 to 5 V DC, 0 to 24 V DC, transistor-transistor logic (TTL)* *Analog I/O:* 4 to 20 mA, 1 to 5 V DC, 0 to 10 V DC, –5 to +5 V DC, –10 to +10 V DC *Special I/O:* Thermocouple, RTD, stepper motor (pulses), strain gauge, high-speed counters, PID
Typical Specifications:	*Scan Time per 1000 Words (1 K) of Logic: 0.1 to 50 ms, depending on manufacturer and enhanced software features* *Word Size*: 4, 8, 16, or 32 bit (typical) *Amount of I/O:* Nano PLC — under 15 I/O, Small or Micro PLC—15 to 128 I/O; Medium PLC—128 to 512 I/O; Large PLC—512 and greater I/O *Size of Memory:* Small PLC—256 to 2 K words (K = 1024 bits, bytes, or words of digital data); medium PLC—2 K to 12 K words; large PLC—12 K words and larger *Type of Memory:* CMOS (complementary metal oxide semiconductor) RAM (random access memory); BBRAM (battery-backed RAM); EEPROM (electrically erasable PROM); and Flash memory *Environmental Conditions:* 0 to 60°C (32 to 140°F), relative humidity, to 95% noncondensing, 115 VAC ± 15% and 230 VAC ± 15%
Costs:	*Nano PLC Hardware: Basic PLC with CPU, 12 I/O points costs $99 to $200* *Small PLC Hardware:* Basic PLC with CPU, 2 K to 8 K RAM memory, 20 to 40 I/O points costs $200 to $1000 (extra discrete I/O when available costs $10/point) *Medium PLC Hardware:* Basic system with 256 I/O points costs $5,000 to $8,000; with 512 I/O points costs $10,000 to $20,000; and with 1024 points costs $14,000 to $28,000 (RAM quantity adjusted to handle designated I/O count) *Large PLC Hardware:* $15,000 to $70,000 (cost driven by network and information display requirements) *Software and Engineering:* Software costs range from 50 to 100% of the hardware cost. System engineering costs and documentation costs each range from 25 to 50% of hardware cost. Therefore, the total cost (without installation labor) is about twice (if not more) the hardware cost.
Support Equipment Cost:	*Handheld programmers cost $100 to $500; PC-supported programming software costs $2500 to $5000*
Partial List of PLC Suppliers:	ABB (Elsag-Bailey Controls) (www.ABB.com) Allen-Bradley/Rockwell Automation (www.AB.com) Automation Direct (www.Automationdirect.com) Danaher (Eagle Signal Controls) (www.Dancon.com) Eaton (Cutler-Hammer) (www.EatonElectrical.com) Emerson (Westinghouse) (www.EmersonProcess.com) Fuji Electric Corp. (www.FujiElectric.com) G.E. Fanuc Automation (www.GEIndustrial.com) Giddings & Lewis (www.GLControls.com) Idec Corp. (www.Idec.com)

International Parallel Machines Inc. (www.ipmiplc.com)
Mitsubishi Electric (www.meau.com)
Modicon/Schneider Electric (www.Modicon.com)
Moeller Corp. (www.Moeller.net)
Omega Engineering (www.Omega.com)
Omron Electronics Inc. (www.Omron.com)
Reliance Electric Co./Rockwell Automation (www.Reliance.com)
Siemens (www.sea.siemens.com)
Toshiba Inc. (www.Toshiba.com)
Triconex/Invensys (www.Triconex.com)
Uticor Technology Inc. (www.Uticor.com)

(*Note*: The most popular PLCs are Allen-Bradley, Modicon, Siemens, and GE Fanuc.)

Partial List of HMI Software Suppliers:	GE Fanuc (iFix & Cimplicity) (www.gefanucautomation.com) Rockwell Software (RSView) (www.software.rockwell.com) Invensys (Wonderware InTouch) (www.wonderware.com)

INTRODUCTION

A programmable logic controller (PLC) is an industrially hardened computer-based unit that performs discrete or continuous control functions in a variety of processing plant and factory environments. Originally intended as relay replacement equipment for the automotive industry, the PLC is now used in virtually every industry imaginable. Though they were commonly referred to as PCs before 1980, PLC became the accepted abbreviation for programmable logic controllers, as the term "PC" became synonymous with personal computers in recent decades.

PLCs are produced and sold worldwide as stand-alone equipment by several major control equipment manufacturers. In addition, a variety of more specialized companies produce PLCs for original equipment manufacturer (OEM) applications.

Typically, PLC vendors can supply large volumes of application notes for their products. Most major PLC vendors also publish detailed articles about applications in technical journals and prepare papers for engineering societies and industrial symposia on control, automation, and so forth. Each manufacturer's software package usually has its own application programming techniques. Vendors also are a valuable source of "how-to" information, providing training courses in their local office or at the factory as well as actual hands-on experience to help users gain familiarity with the PLC. Most vendors offer an applications or programming manual that provides insight on how to use available programming features. Of course, familiarity with one brand of PLC generally helps the engineer learn to use other brands quickly.

HISTORY

The automotive industry fostered the development of the PLC primarily because of the massive rewiring required each time a model change occurred. Solid-state logic is much easier to change than relay panels, and this advantage was reflected in the cost of installing and operating PLCs instead of traditional relay systems. Table 5.4a lists some of the milestones in the development of PLCs. Table 5.4b provides a summary of the differences among various technologies that perform logic control.

Bedford Associates, the forerunner of Modicon, designed the first PLC in 1968 for the Hydramatic Division of General Motors Corporation to eliminate costly scrapping of assembly-line relays during model changeovers and to replace unreliable electromechanical relays. The objectives of the program were to:

- Extend the advantages of static circuits to 90% of the machines in the plant
- Reduce machine downtime related to controls problems
- Provide for future expansion
- Include full logic capabilities, except for data reduction functions[1]

Richard Morley, a Bedford engineer, is credited with the original PLC design and the creation of ladder logic programming.[2] Morley says the diagrams on which ladder logic is based, however, probably originated in Germany years before to describe relay circuitry.[3] Describing the technology of the day, Morley explains, "Automatic control of industrial metalworking and assembly operations in 1969 were mostly by relays and clock-driven electromechanical devices. Relays ran hot and tended to beat themselves to death with their persistent 60-cycle hum and contact arcings. Electromechanical timers/sequencers were much like enlarged music box mechanisms and were maintenance headaches."[4] Describing the early years of the PLC, Morley recounts, "We had some real problems in the early days of convincing people that a box of software, albeit cased in cast iron, could do the same thing as 50 feet of cabinets, associated relays, and wiring."

Morley recounted that in 1969, "all computers required a clean, air-conditioned environment, yet were still prone to frequent malfunctions. … Thus, even though PLCs were and are special, dedicated computers, considerable effort was made to not identify PLCs as computers due to the poor reliability of computers and the fact that they were not things procured

TABLE 5.4a

History of Programmable Logic Controller (PLCs)

1968	PLC designed for General Motors Corporation to eliminate costly scrapping of assembly-line relays during model changeovers.
1969	First commercial PLCs manufactured for automotive industry as electronic equivalent of relays.
1971	First application of PLCs outside the automotive industry.
1972	Allen-Bradley (AB) adds the first computer interface to its PLC, read-write controller, and off-line software documentation package.
1973	Introduction of "smart" PLCs for arithmetic operations, printer control, data move, matrix operations, CRT interface, etc.
1974	First CRT-based program panel for PLC.
1975	Introduction of analog proportional, integral, derivative (PID) control, which made possible the accessing of thermocouples, pressure sensors, etc. Allen-Bradley coins the PLC acronym.
1976	First use of PLCs in hierarchical configurations as part of an integrated manufacturing system. AB introduces Remote I/O.
1977	Introduction of PLCs based on microprocessor technology.
1978	PLCs gain wide acceptance; sales approach $80 million. USDATA introduced REACT, the industry's first microprocessor-based, user-configurable, interactive color graphics workstation for use with PLCs.[2]
1979	Integration of plant operation through a PLC communications system with AB's DataHighway.
1980	Introduction of intelligent input and output modules to provide high-speed, accurate control in positioning applications.
1981	IBM introduces PC/XT using DOS. Data highways enable users to interconnect many PLCs up to 15,000 feet from each other. More 16-bit PLCs become available. Color graphics CRTs are available from several suppliers.
1982	Larger PLCs with up to 8192 I/O become available.
1983	"Third-party" peripherals, including graphics CRTs, operators' interfaces, "smart" I/O networks, panel displays, and documentation packages, become available from many sources.
1985	IBM-PC programming terminal introduced. AB PLC-5/15 introduced with built-in RIO and peer-to-peer DH+ directly on processor card.[6] Intellution introduces the FIX software, the first DOS-based process control system for IBM PCs.[2,5]
1986	Various other PC HMI software introduced.
1990	Wonderware introduces InTouch, the first MS Windows-based HMI application, and starts a major transition to Windows-based "open" technology.[2,5]
1991	Small (Brick) PLCs enter the market place.
1992	Networked block I/O developed. Moore Products manufactures APACS, the first control system using IEC 1131-3 standard allowing DCS function blocks, PLC ladder, sequential function chart, and structured text to be used and combined within a single controller.[2]
1993	Ethernet and TCP/IP connectivity appear on PLCs.
1994	AB launches DeviceNet, an open device-level network.

by manufacturing operations."[4] Unlike computers of that era, the programmable controller was designed to be reliable. Morley explains, "No fans were used, and outside air was not allowed to enter the system for fear of contamination and corrosion. Mentally, we had imagined the programmable controller being underneath a truck, in the open, and being driven around—driven around in Texas, driven around in Alaska. Under those circumstances, we wanted it to survive. The other requirement was that it stood on a pole, helping run a utility or a microwave station which was not climate controlled, and not serviced at all. Under those circumstances, would it work for the years that it was intended to be? Could it be walled in? Could it be bolted in a system that was expected to last 20 years?" Upon arriving for a client demonstration, the PLC was accidentally dropped on the floor. The client was surprised that not only did the PLC work perfectly, but that Modicon staff expected it to work after being dropped.[3]

Bedford Associates demonstrated the Modicon 084 solid-state sequential logic solver at GM in 1969.[5] The fundamental advantage that PLCs introduced was that they used programming rather than rewiring to configure for a new application. PLC programming could be done much quicker than the traditional rewiring methods. In addition, solid-state devices offered greater reliability, required less maintenance, and had a longer life than mechanical relays, all in a smaller footprint.[2,6]

The 1970s and 1980s were dominated by proprietary systems and software. Odo Struger, often called the father of Allen-Bradley's PLC, is credited with the PLC acronym.[5] The early

TABLE 5.4b
Cost Advantages over Relay

	Relays	*Solid-State Controls*	*Microprocessor*	*Minicomputer*	*PLCs*
Hardware cost	Low	Equal	Low	High	High to low, depending on number of controls
Versatility	Low	Low	Yes	Yes	Yes
Usability	Yes	Yes	No	No	Yes
Troubleshooting maintainability	Yes	No	No	No	Yes
Computer-compatible	No	No	Yes	Yes	Yes
Arithmetic capability	No	No	Yes	Yes	Yes
Information gathering	No	No	Yes	Yes	Yes
Industrial environment	Yes	No	No	No	Yes
Programming cost	(Wiring) High	(Wiring) High	Very high	High	Low
Reusable	No	No	Yes	Yes	Yes
Space required	Largest	Large	Small	OK	Small

1980s saw a cross pollination between PLCs and distributed control systems (DCSs). PLCs, for example, had already begun incorporating distributed control functions so they could be linked much in the way that DCSs were linked. Building on the trend, software companies sprang up in great numbers during this time.[2]

During the 90s, standardization and open systems were the main themes. Ethernet peer-to-peer networking became available from virtually all PLC manufacturers. EEPROM and Flash memories replaced the EPROMs of the 1980s. PCs and CRTs in general became accepted and started to replace switches and lights on control system panels. Small PLCs called “Bricks” were introduced to the marketplace. Redundancy for PLCs became a standard product. Many PLC systems were upgraded to address Y2K concerns. The first few years of the 21st century have seen a consolidation of PLC manufacturers. Very small nano or pico PLCs, some as small as industrial relays, have appeared. Safety PLCs featuring triple redundancy were introduced. LCD base operator interface panels have largely displaced CRTs, especially on the plant floor.

PLC SIZES

Today’s PLC is at best a distant relative of the first- and second-generation PLCs built during the 1970s and 1980s. There are now many PLC sizes to select from, ranging from nano- and micro-size devices with 12–30 I/O, to large supervisory control units with built-in PC and networking capabilities.[7]

The modern medium-sized PLC performs all the relay replacement functions expected of it but also adds many other functions—including counting, timing, and complex mathematical applications—to its repertoire. Most medium-sized PLCs can perform proportional, integral, derivative (PID), feedforward, and other control functions as well (see Chapter 2). In addition, medium-sized and large-scale PLCs have data highway capabilities and can function well in DCS environments.

Nano PLCs

In the 1990s, small PLCs were introduced with limited functionality. Today’s small PLCs generally contain all of the software and networking power that used to be confined to the larger units.[8] A small PLC can be amazingly inexpensive. This small, dedicated controller is enclosed in a single-mounted hardened case. It is intended to be a relay replacement unit and provide reliable control to a stand-alone section of a process. Small PLCs offer many of the same functions as larger models but with limited flexibility of adding I/O points. They are especially suited to limited applications with only basic communications requirements. “For instance, a $99 PLC with four PID loops and two serial ports, matched with a $79 four channel analog plug-in option card, is a cost-effective choice over traditional relays and a single-loop controllers.”[9]

BASIC PLC COMPONENTS

A PLC manufactured by virtually any company has several common functional parts. Figure 5.4c illustrates a generic PLC architecture. The diagram shows a central processor, memory, I/O, power supply, and programming and peripheral device subsections. Each is discussed below.

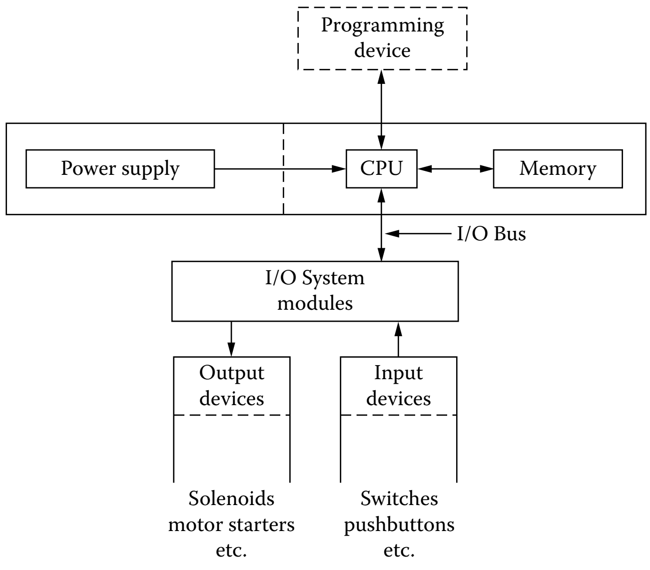


FIG. 5.4c
PLC architecture.

Central Processor Unit (Real Time)

The central processor unit (CPU) performs the tasks necessary to fulfill the PLC function. Among these tasks are scanning, I/O bus traffic control, program execution, peripheral and external device communications, special function or data handling execution (enhancements), and self-diagnostics. Central processing units that use microprocessor-based (VLSI) systems are both more powerful and more flexible than earlier technologies such as TTL or CMOS. It should be noted that the CPU and memory units are considered separate functions.

One common method for rating how a PLC performs these tasks is its scan time. Scan time is roughly defined as the time it takes for the PLC to interrogate the input devices, execute the application program, and provide updated signals to the output devices. Scan times can vary from 0.1 ms per 1 K (1024) words of logic to more than 50 ms per 1 K of logic. Although scan times are often given as performance measures, there are factors that make them misleading. Word size varies from 4 bits to 32 bits depending on the PLC model and manufacturer. Special features, which vary from preprogrammed drum times to full floating-point mathematics, have different processing times and may generate longer scan times. Because of these variables, other performance factors must be considered when selecting a PLC. The user should take into account the application as well as the speed of the controller. Generally speaking, process applications need to take advantage of microprocessor power, whereas machine control applications are usually more concerned with program execution speed.

Newer CPUs are equipped with communications ports such as ASCII (RS-232), remote I/O (RIO), and Ethernet built into the CPU module. Some CPUs now even support timed interrupts.

Redundancy The new breed of data highways can provide a process with a "hot" backup PLC to take over in the event of a PLC processor failure. Figure 5.4d depicts a PLC system with redundant CPUs and I/O.

Until recently, it was difficult or impossible to configure an ordinary PLC into a redundant hot backup system. Siemens was one of the first PLC makers to introduce redundant configurations for PLCs. For some people, redundancy is very important. Why is redundancy important? Safety in areas containing explosives or flammables, unattended operation, and environmental liabilities lead the list of concerns addressed in part by redundancy.[10]

Safety PLCs For several years, Triconex was one of a few manufacturers that offered a triplicate PLCs for use in safety systems. Fault tolerant systems called programmable safety systems (PSSs) are now offered by a number of PLC manufacturers. "The big difference between a regular PLC and a PSS is that all the triple redundant features needed in both hardware and software are built into the latter. That can be done with regular PLCs but it's a significant engineering challenge."[11] See Section 5.8 for further information on safety PLC systems.

It's important to acknowledge that having redundant CPUs doesn't necessarily provide system-wide redundancy. Nor does it protect against all types of errors. A programming error, such as divide by zero, or an infinite loop will stop a redundant CPU just as fast as a nonredundant CPU.

Memory Unit

The memory unit[12] of the PLC serves several functions. It is the library where the application program is stored. It is also where the PLC's executive program is stored. An executive program functions as the operating system of the PLC, which serves as the program that interprets, manages, and executes the user's application program. Finally, the memory unit is the part of the PLC where process data from the input modules and control data for the output modules are temporarily stored as data tables. This includes I/O status bits, counter values, timer preset and accumulated values, and other stored constants or variables. Typically, an image of these data tables is used by the CPU and, when appropriate, sent to the output modules.

The basic PLC memory element is the word. A word is a collection of 4, 8, 16, or 32 bits that is used to transfer data about the PLC. As word length increases, more information can be stored in a memory location. Even with the ambiguity associated with word length, PLCs that provide the equivalent of 32 K of 8-bit memory locations can execute application programs that are moderately complicated and interact with 50–100 discrete I/O points.

Memory can be volatile or nonvolatile. Volatile memory is erased if power is removed. Obviously, this is undesirable, and most units with volatile memory provide battery backup to ensure there will be no loss of program in the event of a power outage. This is often referred to as battery-backed

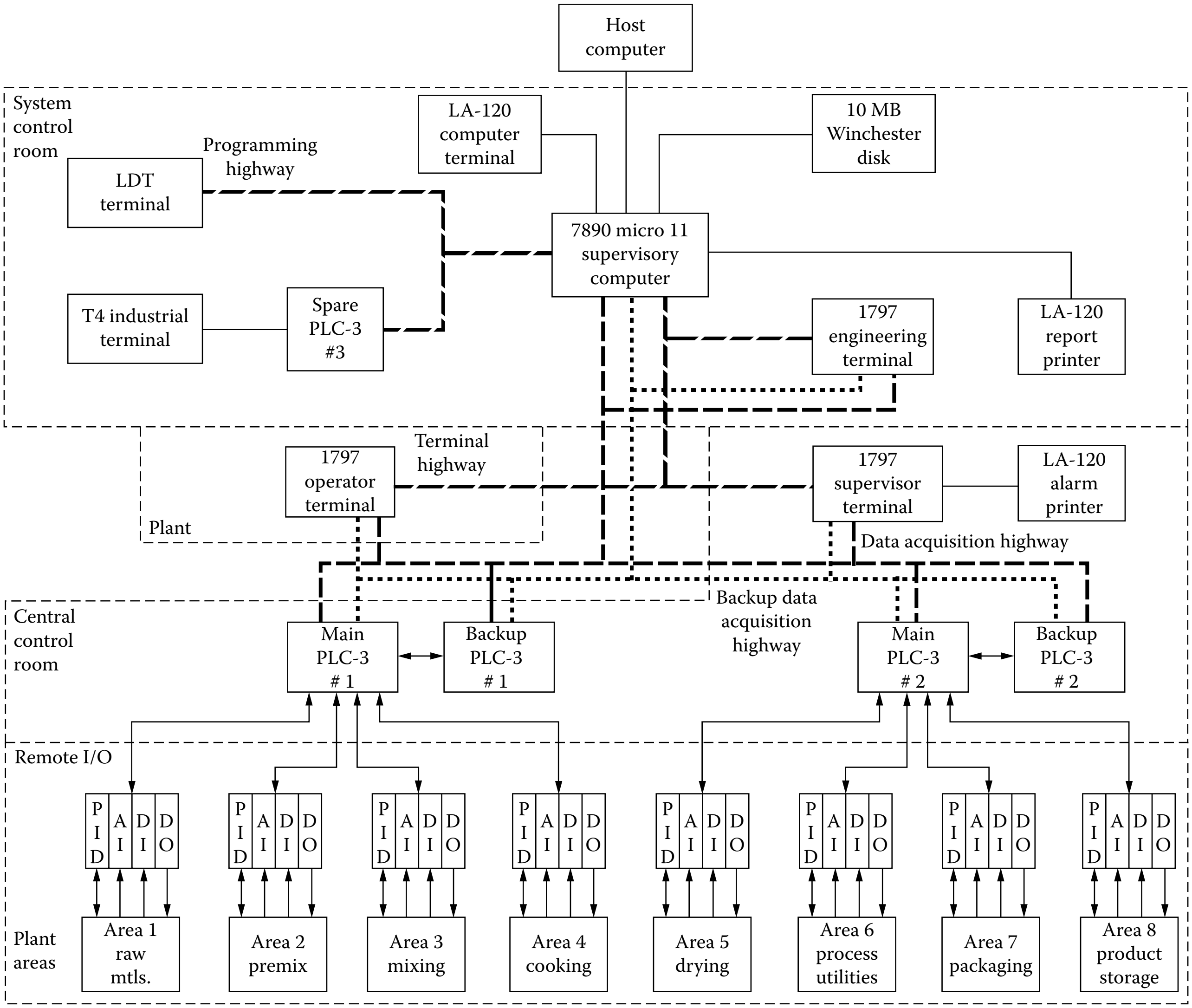


FIG. 5.4d
Hot backup PLC networking. (Courtesy of Allen-Bradley.)

RAM. When batteries are used, they must be replaced on a regular basis, typically 1–2 years. Some small PLCs use a large capacitor instead of a battery to avoid this maintenance issue.

Nonvolatile memory does not change state on loss of power and is used in cases in which extended power outages or long transportation times to job sites (after program entry) are anticipated. Core memory was widely used in the 60s and early 70s but is no longer used. In the late 70s a solid-state nonvolatile memory called programmable read-only memory (PROM) became popular. The user programmed the PROM by applying a high current pulse in order to fuse internal links to set individual bits to 1 or 0. This procedure was irreversible.[13] This required a separate piece of equipment called a PROM burner. Then erasable PROMs became available. Like the PROMs, EPROMs could be programmed by the user. The advantage was that EPROMs could be erased optically, through exposure to ultraviolet light through a quartz window on the chip, and then rewritten electrically. A drawback to EPROM devices was that they had to be removed from the equipment to be reprogrammed.

EEPROM and Flash Memory New memory technologies, electrically erasable (or Alterable) PROMs (EEPROMs, E^2PROMs, or EAPROMs) and Flash memories, became available in the 90s and have essentially replaced the earlier types. EEPROM and Flash memories can be electrically reprogrammed by the user, but without removal from the system, and without the use of exterior programming devices such as PROM programmers.[14] Flash is also used for the firmware used by PLCs and intelligent modules. Revision changes to the firmware can be “Flash upgraded” in the field, without equipment disassembly. EEPROM and Flash memory are the current choices for nonvolatile memory.

I/O Systems

The I/O group includes all types of modules, from dry contact modules to intelligent I/O to remote I/O capabilities. Inputs are defined as real-world electrical signals that give the controller real-time status of process variables such as level, pressure, flow, temperature, weight, analysis results, position, or status (hand switches, pushbuttons, alarm contacts, and so on). These signals can be analog or digital, low or high frequency, maintained or momentary. Typically, they are presented to the PLC as a varying voltage, current, or resistance value.[15,16] Based on the status sensed or values measured, the PLC controls various output modules to operate devices such as valves, motors, pumps, and alarms. PLC manufacturers are providing more and more types of input and output capabilities in their products. There are, however, many third-party peripherals that aid the PLC in interfacing to the field devices.

One of the most important functions of the I/O is its ability to filter, condition, and isolate real-world signals (0–120 VAC, 0–24 V DC, 4–20 mA, 0–10 V, and thermocouples) and convert these to the low signal levels (typically 0–5 V DC max) in the PLC I/O bus. This is accomplished by use of optical isolators. A light-emitting diode (LED) generates a light that optically turns on a phototransistor. Because the only thing normally connecting the input to the output is light, there is an excellent isolation between input and output voltages. Only the electrical breakdown of the isolator (typically over 1000 V) limits the voltage isolation achieved with these devices. Typical discrete I/O schematics are shown in Figures 5.4e, 5.4f, and 5.4g.

Most I/O systems are modular in nature; that is, a system can be arranged by use of modules that contain multiples of I/O points. These modules can be composed of 1, 4, 8, 16, or 32 points and plug into the existing bus structure. The bus structure is really a high-speed multiplexer that carries information back and forth between the I/O modules and the central processor unit. Higher point densities are possible, but their selection may involve a trade-off in wire size used, as well as the ease of wire harness installation to the module.

In small PLCs, the CPU and I/O are generally contained in a single enclosure and may or may not be expandable. These have been nicknamed “Bricks” because the size of the enclosure is similar to the size of a brick. In most medium and large PLCs, plug-in I/O modules are used to convert the I/O signal level to one that is compatible with the bus architecture. These modules can be composed of 1, 4, 8, 16, or 32 points, depending on the manufacturer’s standard design. For small projects (20–256 I/O), I/O requirements are usually easy to define and group. A systematic approach is required for medium-sized projects (256–1024 I/O), however, in order to avoid confusion of I/O allocation. Obviously, the organization of I/O for large systems (1024 I/O and above) requires careful planning.

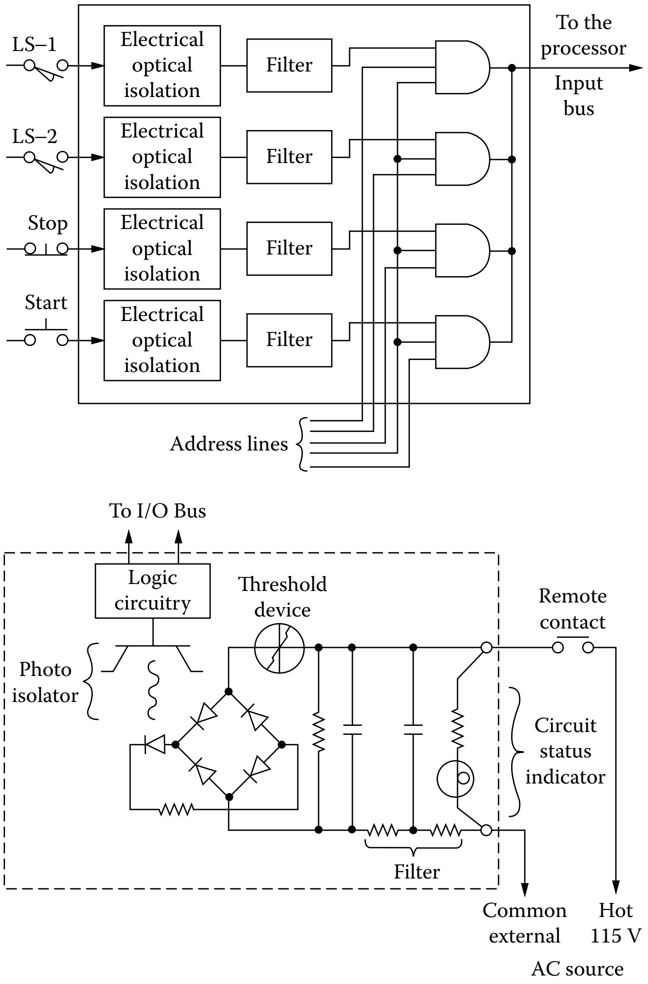


FIG. 5.4e
Typical input area; typical AC input unit.[4]

PLC manufacturers are typically very good at backward compatibility. When new models of CPUs are introduced,

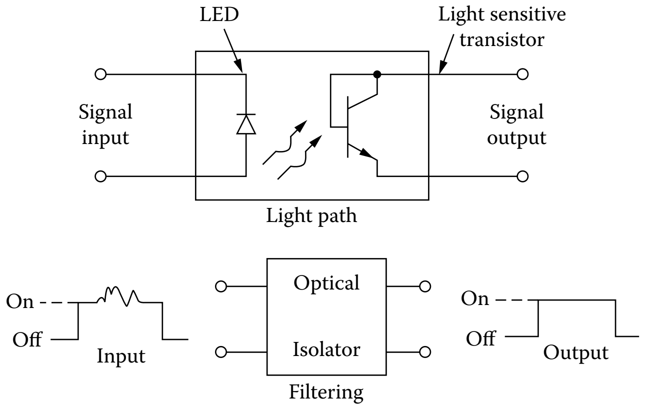


FIG. 5.4f
Electrical optical isolator.[4]

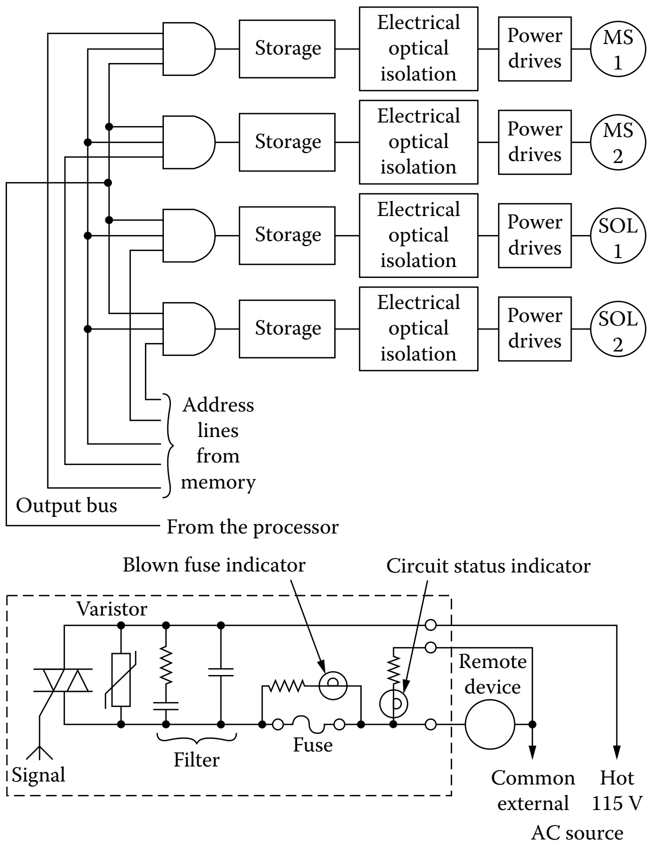


FIG. 5.4g
Typical output area; typical AC output unit.[4]

they are generally able to use the I/O modules made years before. For example, at least one PLC manufacturer's current PLC still works with I/O modules introduced nearly 30 years ago — an impressive feat considering the advances in technology during that time span. In other cases, when manufacturers have changed lines, other companies have stepped in to support and expand the older PLC equipment.

Some manufacturers assign I/O memory locations based on I/O type and position in the rack. Care must be taken in these systems when adding I/O to these systems, as it might cause an automatic readdressing of existing I/O modules. Because the program doesn't alter itself, it would need to be updated for it to function like it did before the I/O module addition.

Local I/O There are at least three different types of racks: local, expansion, and remote. A local I/O base can hold the CPU module, power supply, and optionally I/O modules. Expansion racks that are close by (usually within the same panel) are either plugged into the side of the local I/O base or connected by a short parallel cable. Standard cabling is provided for connection from the remote I/O base to the PLC processor. These cables are generally high-speed serial cables to simplify installation. The PLC processor generally plugs into a dedicated slot in the rack. Some manufacturers allow multiple processors in a single rack.

The I/O base (rack or housing) is used to hold the I/O module in place and to provide a termination point for the wiring. The bases may be mounted anywhere in the control enclosure; however, there are cable length and free space requirements that must be met. Some racks are sized to mount directly into a standard 19-in. rack enclosure. The majority of bases mount horizontally to allow proper module cooling. Some manufacturers define how much space must be reserved around the rack to permit proper circulation. A terminal strip is built into the mounting base for field connections so that no wiring need be disturbed in order to remove or replace a module. These bases typically hold various quantities of I/O — anywhere from 1 to 128 I/O points.

Some PLC vendors provide wire-way as part of the rack structure to "clean up" the panel design. As each new generation of I/O becomes increasingly dense, the systems become progressively harder to wire. Interposing terminal blocks between the field connections and the rack may become standard in the future. Some manufacturers already offer prewired field terminals that mount in a panel and are equipped with a cable that plugs directly onto the I/O module.

Digital Inputs Pushbuttons, limit switches, or even electromechanical relay contacts are familiar examples of digital, contact closure type signals. Input modules are typically transistor-triggered and have built-in time delays to protect against contact bounce. The input signal from a field device (limit switch) has to be energized for some amount of time in order for the module to notify the processor of a true "on" condition. The modules are available in a number of voltages including DC voltages (14, 15–30, 24, 120, 240), TTL (5 V DC), and AC voltages (24, 48, 120, 240). They come in 8, 16, or 32 points per module.

Discrete I/O modules come equipped with an LED indicator light to indicate the status of each I/O point in the module (on or off) for troubleshooting. The input module LED indicates field side status of the pushbutton.

Modules, especially high-density modules, often share a common neutral between multiple inputs. This means that the input voltages must be referenced to the same neutral. AC signals powered from different power sources cannot be mixed within the shared group. Using a different module for the different sources or using isolating relays solves the problem.

Digital Outputs Digital or discrete outputs (DOs) can be used to power pilot lights, solenoid valves, or annunciator windows (lamp box). The discrete output module uses a solid-state switch (triac) to power a field device, such as a motor starter, a valve, or lights. Outputs are available for voltage ranges of 5–240 V at currents up to 5 A, with typical 120 V outputs operating at 1–2 A maximum. Solid-state drivers of this type are not intended to drive large loads directly (e.g., a large motor starter). Highly inductive loads or those with a high surge current may also require an interposing dry contact relay in order to power the field device or a resistor-capacitor (RC) network to control transients.

Both manufacturers and third-party vendors offer dry contact modules. These modules solve the problems normally associated with triacs: low power and uncertainty of failure state. Triac outputs also "leak" some current when they are off. When attached to high impedance devices such as Sonalerts or LED pilot lights, this leakage current can cause continuous or intermittent operation. A voltage-dropping resistor can be added across the load to solve this, but the resistors are often large and run hot. Some designers simply add inexpensive isolation relays as needed to address this problem.

Manufacturers rate their equipment output current at different temperatures. All current ratings for solid-state outputs will vary with ambient temperatures. Therefore, one should be sure to check the PLC output rating for each application and manufacturer. Built-in fusing on output modules is becoming standard in the industry and provides good protection for overload conditions. The type of fuse depends on the module, and the way in which fuses are accessed varies from one PLC manufacturer to another. Some wiring arms have individual fuses for each output.

Discrete I/O modules come equipped with an LED indicator light to indicate the status of each output on the module (on or off) for troubleshooting. The output module LED generally indicates logic side status.

Analog Inputs Analog modules are sometimes referred to as A/D (analog-to-digital) modules. They provide optical isolation for electrical noise protection and are typically arranged in a quad module or an eight-point module. The analog I/O system is designed to interface with analog field devices such as flowmeters, pressure transmitters, and valve positioners. This system accepts inputs of 0–5 V DC, 0–10 V DC, –5 to +5 V DC, –10 to +10 V DC, 0–20 mA, or 4–20 mA. Precisions of 1 in 4,096 (12-bit), 1 in 32,768 (15-bit), or 1 in 65,536 (16-bit) are commonly available.

Direct thermocouple and RTD input modules typically accept eight to ten points each. On-board cold junction compensation is often included with thermocouple input modules, and many different types of thermocouples (B, C, E, J, K, N, R, S, T), RTD (100 Ω platinum, 120 Ω nickel, 10 Ω copper), and mV signals (0–50 mV, 0–100 mV) can be accommodated.

Analog Outputs Analog outputs can drive signals to variable-speed drives, control valves, and other analog devices. The output analog module can output a signal of 4–20 mA current loop as well as 0–10 V DC, 0–5 V DC, –5 to +5 V DC, and –10 to +10 V DC. Most source their power from the back plane. Some loops require external loop power supplies.

Register I/O Systems One additional type of input signal, the register input, reflects the computerized nature of the PLC. The register input is particularly useful when the process condition is represented by a collection of digital signals delivered to the PLC at the same time. A binary-coded decimal (BCD) thumbwheel is a good example of an input device that is compatible with a register input port. If the thumbwheel represents three and one-half digits of process data, then all 14 data output wires from the thumbwheel would provide their digital signal directly to the PLC register input unit, which would, in turn, signal the condition and transfer the data to the central processor unit.

The register I/O system provides direct interface to multibit data field devices, such as thumbwheel switches, position encoders, and digital readouts to the PLC. These devices are typically TTL level, which allows interface to other types of electronic hardware as well. Intelligent I/O and other special-purpose I/O requirements are becoming increasingly common.

Specialty Modules Specialty modules are designed to solve a single interface problem. X/Y positioner modules can be included in this category, as well as servo axis controllers, stepper motor outputs, and even maintenance access modules. These modules are a further extension of the distributed technology. Combination I/O modules combine either digital or analog inputs and outputs on a single module. They are particularly useful in small systems.

High-Speed Modules Fast-response I/O is currently offered in both discrete and analog versions. Discrete rapid response modules are facilitated by the PLC logic, but the output does not rely on ladder logic scan times to get updated. High-speed analog modules provide quicker analog-to-digital and digital-to-analog (D/A) conversions. This gives PLCs the ability to control faster PID loops and to make analog measurements of assembly-line parts (weight, for example). High-speed pulse counter modules provide the ability to interface with turbine meters, stepper motors, and optical encoders. High-speed pulses cannot normally be interfaced to PLC inputs because of the scan time of the ladder logic. These modules provide an interface that does not rely on the scan time, so the PLC is able to monitor fast pulses that indicate position or flow. They can often be configured as high-speed up counters, up/down counters for quadrature use, high-speed interrupts, pulse catchers for very fast pulses, and adjustable filter inputs. PID loop coprocessor and temperature controller modules are also available.

Intelligent I/O Modules Intelligent I/O modules include all modules that are able to perform processing functions. Because the tasks performed by the PLC are further distributed, greater speed and reliability for the overall system can be realized. Intelligent I/O modules give the PLC multiple additional capabilities, which may include memory storage and retrieval, computing tasks, and communications. Memory modules provide additional room to store data points, alarm messages, lookup tables, and the like. This approach leaves the main operating memory free for the control tasks. Computing modules give PLCs the ability to perform true computer functions using a language like Basic. Basic modules can be used for a variety of things, such as complex math abilities.

Because they usually come with two or more serial ports, they can be used for handling specialized ASCII protocols and they can demultiplex a multiplexer data stream. Of course they can be programmed as Modbus masters or slaves. Again, the real-time tasks are left in the main memory, but tasks such as setpoint calculation, formation of data, and some operator interface tasks may be placed in the computer module. Other in-rack modules include complete computers, state language coprocessors, and real-time interrupt modules.

Clock modules that fit into the I/O bus may be considered part of this group. These modules provide real-time and day/date functions upon interrogation from the PLC. Most are backed up by a battery to ensure time keeping during power outages.

I/O Simulators Some I/O simulators used to develop and debug programs can be categorized in the I/O enhancement group. These specific devices are typically hardware modules that can be plugged into the PLC. I/O simulators generally look like input cards except switches are built into the module for testing.

PLC Power Supply

The power supply may be integral or separately mounted. It always provides the isolation necessary to protect solid-state components from most high-voltage line spikes. The power supply converts power line voltages to those required by the solid-state components. All PLC manufacturers provide the option to specify line voltage conditions (typically 120 VAC, 240 VAC, or 12/24 V DC). In addition, the power supply is rated for heat dissipation requirements for plant floor operation. This dissipation capability allows PLCs to have high-ambient-temperature specifications and represents an important difference between PLCs and PCs for industrial applications.

The power supply drives the I/O logic signals, the central processor unit, the memory unit, and possibly some peripheral devices. As I/O is expanded, some PLCs may require additional power supplies in order to maintain proper power levels. The power supply generally does not provide power to field devices, though some manufacturers provide external terminals to permit this use. The additional power supplies may also be separate or part of the I/O structure. The power requirements of all the I/O modules and CPU or remote I/O adapters are added together for each voltage used to make sure that the power supply is large enough.

ADDITIONAL PLC COMPONENTS

Communications Modules

Communications modules can provide the PLC with a range of capabilities, from simple ASCII output strings to communications networking. The storage of ASCII messages for a printer or display can be contained outside the main memory of the PLC, and the data can be output when required. Full-system communications networking capabilities are provided with network modules, giving the designer the ability to multidrop PLCs off a single operator interface device or a supervisory computer. A wide variety of devices communicate using ASCII, including weight scales, bar code scanners, some operator interfaces, and modems. Modules included in this category also include telephone and radio modem modules. Remote I/O and peer-to-peer are specialized communications modules that are discussed below.

Remote I/O

Robust industrial networks make distributed I/O feasible. Networks enable I/O modules to be located close to field devices, resulting in labor and material savings for wiring.[8,11] Often manufacturers provide a remote I/O distribution panel or module to serve the efficient multiplexing of the modules on the remote I/O rack back to the CPU. A remote I/O base is similar to the local I/O base except it holds a remote I/O adapter (or has it built-in) instead of the CPU. Most medium-sized PLCs can support several remote racks, which in turn contain 4, 8, or 16 I/O modules. It may or may not be possible to mix various modules in a remote rack. Specific information about module compatibility and remote I/O multiplexing is available from the manufacturer. This information is required to facilitate PLC selection and sizing for specific applications.

Whereas in most systems the module has the intelligence to communicate with the CPU, some systems require the use of serial interface modules. In any case, some provision is made to accept register input data from the input modules and to send this data (on or off status of field device) in serial format to the PLC processor. Serial data are also converted into register data to be sent to the output module. This is normally a proprietary protocol.

Remote I/O adapters need to be configured for a particular drop, either by programming or a switch setting on the module itself. Remote I/O is very often an RS-485-based system. The end modules in the RIO network are usually terminated with a resistor that matches the impedance of the cable used. RIO communications speeds range from 56 K to 230 Kbps. Remote I/O is broken down into two distinct types: the integral type, which allows a limited transmission distance (up to 15,000 ft, or 4,500 m); and the transmitter/receiver type, which allows virtually unlimited transmission capability. Most PLC manufacturers and third-party peripherals manufacturers can provide some form of either type. Technology is advancing rapidly in this area, as systems change from fiber optics to microwave and radio transmission. Many manufacturers are now offering flexible remote I/O solutions. Often these are modular, plug-together systems instead of being rack based. Designed to be located next to the wireway, they incorporate built-in terminal blocks for direct wiring.

Open vs. Proprietary I/O Networks Proprietary networks have a compelling advantage: They are easy to put together. If you buy all your network components from one supplier, even if they're for an open network like DeviceNet, you avoid potential configuration problems. Even so, offering both open and proprietary systems is important to most manufacturers.[11]

Besides proprietary remote I/O networks, a variety of open protocols exist to enable I/O data interchange. These include Modbus, DeviceNet, Profibus, HART, and Fieldbus, among others. Because these are open, a number of different manufacturers offer products based on them. Ethernet is becoming a major factor by providing a backbone that can carry the diverse protocols of nonproprietary I/O communications.[17]

Peer-to-Peer Communications

Many PLC manufacturers have responded to the increasing industry demands for equipment that allows communications among multiple process areas.[18,19,20,21] The original master control layout shown in Figure 5.4h is adequate for PLC control of moderate-size applications of perhaps 100–500 I/O points, but as the application grows the PLC becomes overburdened under this arrangement. The hierarchical plan illustrated in Figure 5.4i provides a supervisory PLC controlling network that reduces this burden. The plan allows for a "master" PLC that oversees the process and controls a set of "slave" PLCs that control the actual process activities in the plant.

An alternate approach is illustrated in Figure 5.4j. This distributed control system allows dedicated PLCs to control sections of the process with an interface to management. As indicated in the figure, this interface could be to a human operator. This person would monitor selected data from the collection of PLCs. Any management decisions generated from this data are passed back to the PLCs by an input terminal. Under normal conditions it is expected that each dedicated PLC can monitor and control its section of the process.

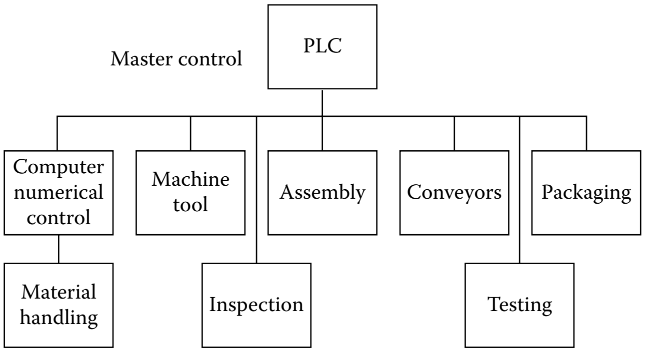


FIG. 5.4h
With master control, one PLC controls a number of related machines and processes. This system is simple but may require long runs of multiple-wire cables, and the entire system is vulnerable to the failure of the one PLC. (Courtesy of Allen-Bradley.)

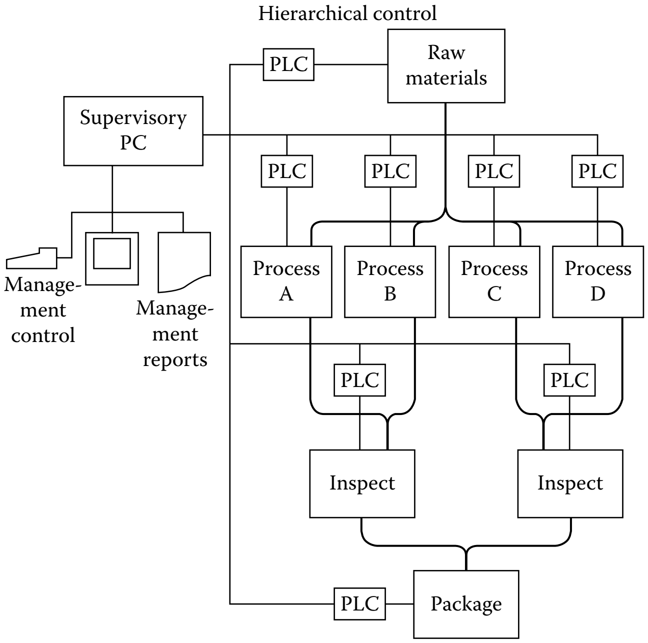


FIG. 5.4i
Hierarchical control. (Courtesy of Allen-Bradley.)

Most major PLC manufacturers currently offer their own network communications for their products (e.g., DataHighway+, ControlNet, Modbus+, and TI-WAY). Various network protocols exist and efforts to generate standards are gaining momentum. Instrument and control organizations such as the Instrumentation, Systems, and Automation (ISA) Society constantly exert pressure on the PLC industry to agree on one or at least a few communications standards. In any event, there are several communications philosophies of interest. Some of these are outlined below.

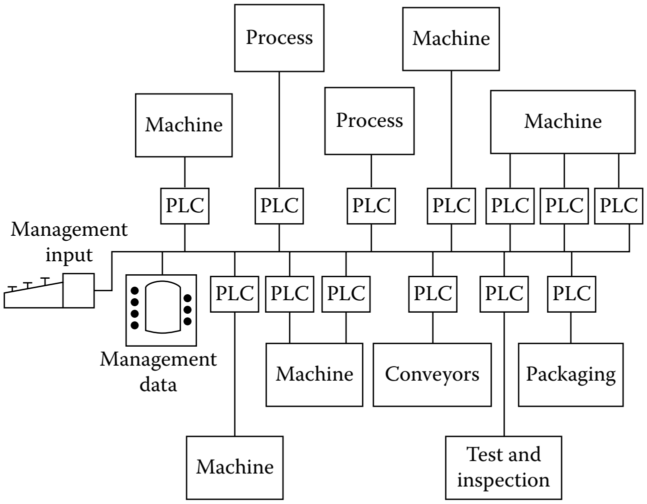


FIG. 5.4j
Distributed control. (Courtesy of Allen-Bradley.)

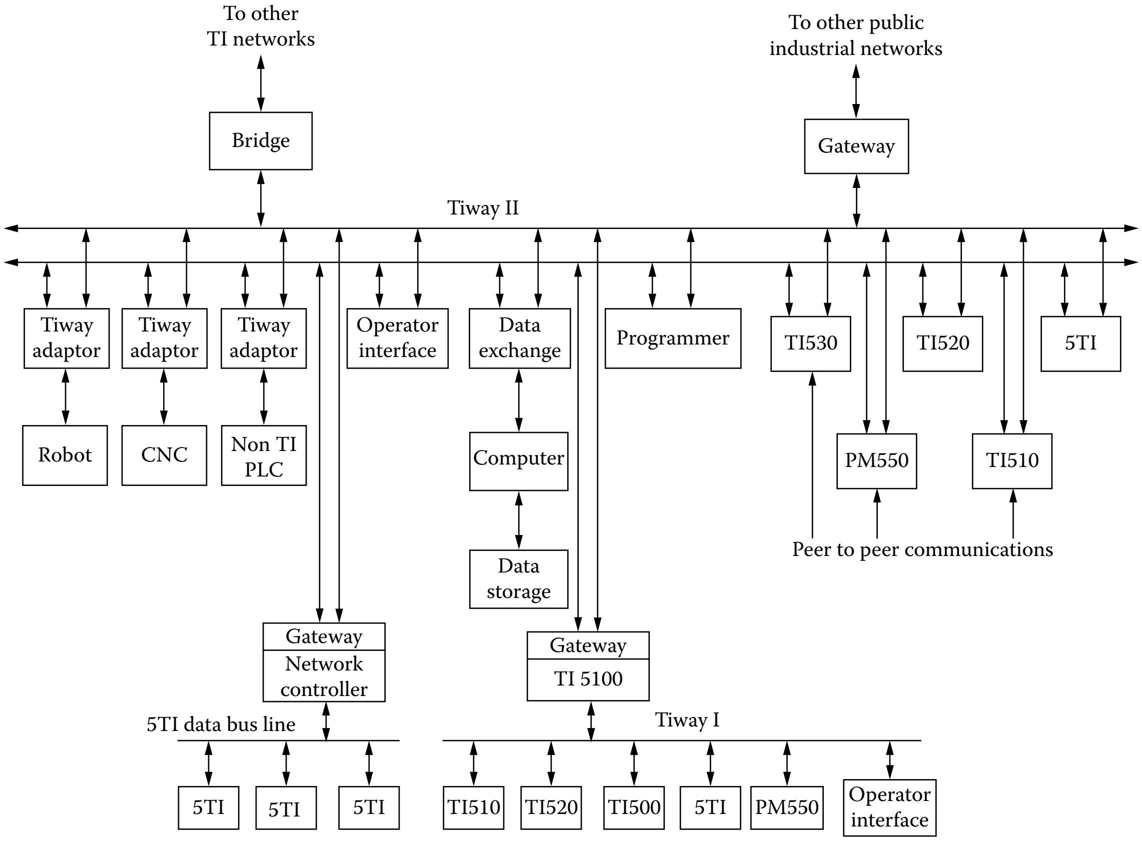


FIG. 5.4k
Peer-to-peer communications. (Courtesy of Texas Instruments.)

The computer interface device group is a rapidly expanding section of PLC peripheral devices. These devices allow peer-to-peer communications (i.e., one PLC connected directly to another), as well as network interaction with various computer systems. In fact, this group of devices will certainly expand in number as communications standards become commonly accepted and more and more products are provided to facilitate such network interactions.[22]

Most PLC manufacturers have already addressed the need for peer-to-peer communications among PLCs on a distributed network. Without this feature, every time one PLC needs to know the status of another part of the machine or plant, it must interrupt the activities of the supervisory computer in order to get the information (Figure 5.4k) or interface through status I/O points. The ability of one PLC to “talk” to another along the data highway greatly speeds the control activities of each machine and allows the supervisory computer to “concentrate” on its tasks.

Some PLC manufacturers and third parties are offering universal communications networking based on open protocols such as Modbus. This is likely to be the favored method in the future, because the need for networking different brands of PLCs together will increase.

More PLC vendors are adding Ethernet compatibility to their systems. Ethernet has become the open architecture of choice for a variety of industries. However, care must be taken to ensure that the PLC is never connected to the Internet. Doing so opens the door to all sorts of security issues, especially hackers.[11] Using a PC in combination with firewalls, encryption, and passwords helps address some of these issues.[23]

Peripheral Devices

The popularity of PLCs has led to the creation of a strong third-party peripheral manufacturing industry. These companies are always developing new products that assist the PLC user with interfacing a specific application to a PLC. Two categories of these products— operator stations and programming/documentation tools—are presented below. The operator stations facilitate operator interface with the PLC-controlled process to monitor process variables, to alter program parameters, to conduct on-line program alterations, and to conduct troubleshooting procedures. Programming and documentation tools include products supplied by the manufacturers or made available by third-party vendors.[24,25]

Local Operator Interface

Operational aids include a variety of resources that range from switches and lights, to color graphics CRTs, to equipment or support programs that can give the operator specific access to processor parameters. In this situation the operator is usually allowed to read and modify timer, counter, and loop parameters but not access the program itself. Some aids facilitate the interaction between the PLC and printers to deliver process information in a desired format. Some devices have the ability to set up an entire panel and plug into the PLC through external RS-232-C ports, saving enormous panel and wiring costs.

The most basic of operator interfaces is switches and lights. These are directly wired to the input and output cards and relay information back and forth to the PLC. Analog signals can also be used with potentiometers, panel meters, and chart recorders. Individual loop controllers can be wired to accept a remote set point from the PLC. Register I/O is used with BCD thumbwheels and BCD displays. Annunciator panels can also be used. Although just about all control panels have some switches and lights, the cost of using a display has fallen to such an extent that it is often not cost-effective to use switches and lights as the primary interface. PLC backup systems that are used if the PLC is being serviced still use these simple devices.

Annuniciator panels are still used in some industries. They are light boxes that have alarm descriptions illuminated when an output is activated. The PLC generally supplies one output point per alarm. The lamp test, silence, and acknowledge functions are built into the annunciator assembly. Text display units are also available in a number of sizes, which display a preprogrammed message when signaled by the PLC. Other message systems are available that can dial a preprogrammed phone number and deliver a page, a text message, or a verbal description of the triggering event. Some switches and light assemblies became available in standard arrangements that connected directly using remote I/O technology. Some are available that communicate using the Modbus protocol.

Operator stations include those provided by manufacturers intended to be used with their particular PLC and those offered by third parties for use with either a particular brand or anyone's PLC. These stations may include devices such as timer/counter access modules (TCAMs), loop access modules (LAMs), data terminals, color graphics consoles, computers, printers, and manual backup stations. Most PLC manufacturers provide an operator interface unit (OIU) designed specifically for their PLC. These are either part of the standard system or offered as an option. They are sometimes mounted directly on the PLC or can be panel-mounted and cabled back to the controller. Functions include access to read/write register data, simple programming, and diagnostics.

Some specialized devices, such as TCAMs, LAMs, and OIUs, provide operator interaction with PLC internal registers and loop tables. This gives the systems designer the ability to provide real-time changing of variables, loop tuning and inspection, manual control of analog outputs, and the ability to provide batch- or menu-type information at low cost. Communications with the PLC are multidropped over an RS-422 or a similar differential line. Unauthorized data entry is prevented with software locks, keylock protection, or both. TCAMs and LAMs have become less popular as the cost of graphic interface panels has dropped, because their functions are easily added to these newer interface panels. Some PLCs can support communications directly with dumb data terminals. However, dumb terminals have been replaced almost completely by PCs.

Many PLCs are able to provide communications directly to printers. A stand-alone PLC system can often provide performance reports, alarm logging, and the like without ever involving a computer. This feature is usually somewhat limited, because PLCs were designed primarily to control the process machine. Large amounts of data, sophisticated print logs, and multiple alarms are not really within the realm of a stand-alone PLC system. This type of data manipulation is too cumbersome and requires too much memory for most PLCs. In addition, many PLCs are located in the process floor, which is generally not a good environment for printers.

Small touch screen panels have become very popular in recent years. They provide basic reporting of status information and commands in a small format. Some also provide built-in alarm management functions. They are meant to replace the traditional switches, lights, and meters. The main disadvantage these small screens have is that they are fundamentally different from the computers in the control room. The screen graphics for the local panels must be developed separately. They generally are limited to the individual control panel they are associated with.

With the advent of plant floor Ethernet and relatively inexpensive computers, it has become practical to provide the same screens at the local control panel that are displayed in the control room. An industrial computer or a regular desktop computer in a sealed computer enclosure is used. The same software that is used in the control room is loaded onto the local computer, which uses the local area network (LAN) to communicate with the supervisory control and data acquivision (SCADA) server. This has many advantages. Because the screens are shown on the same type of hardware, they can be simply copied from the control room computers. Obviously, this helps reduce development and maintenance costs. Because these communicate with the entire plant, an alarm can be investigated at any control panel throughout the plant.

Human-Machine Interface

Intelligent human-machine interfaces (HMIs) can be networked into a total distributed system to give a redundant or local interface to the system (Figure 5.4l). This technique can be used to provide the control room interface, while the supervisory computer supports its own interfaces. Another powerful technique is to allow the networking of a few PLCs together solely for the purpose of providing a single human-machine interface to all parts of the system (Figure 5.4m). This solution is ideal for small PLC users because it is very economical yet allows expansion as the plant grows.

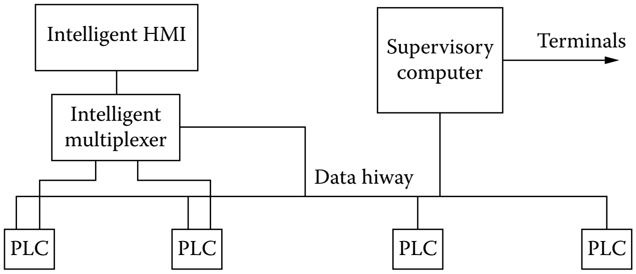


FIG. 5.4l
Local or redundant HMI in network.

Color graphics consoles offer process graphics and communications facilities simultaneously to many brands of PLCs. Because of this ability, HMIs are sometimes used to tie multiple PLC systems together. These systems range from those that can simply be purchased and put on-line with a minimum of engineering effort to those that require some programming. The basic differences are in flexibility. Those that do not require programming may not be able to provide the custom menus and graphics that are required. The ease of communications with different types of PLCs also varies according to manufacturer. Finally, the method of generating the graphics pages differs greatly. Most color graphics consoles offer multiple graphics pages that are animated by reading data tables in the PLCs. Operators enter the data by means of standard keyboards, user-configurable industrial keyboards, light pens, touch screens, and the like. Different graphics pages may be selected with preformatted menus or custom menus programmed by the user or the systems house. Development stations are often required to give the final user the ability to change graphics menus or key commands after the initial project is completed.

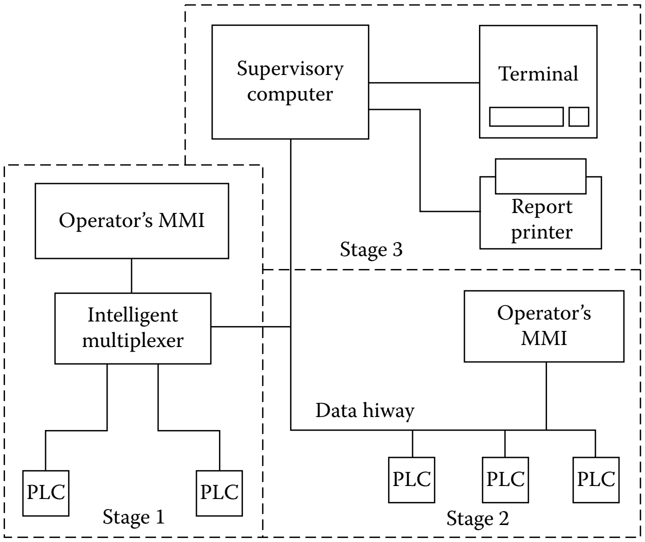


FIG. 5.4m
Entry-level networking grows with user's needs.

Computer systems can be made to perform human-machine interface functions. Indeed, the color graphics consoles described in the previous paragraph are simply computers with standard graphics and communications software packages. Most PLC manufacturers provide board-level additions or modules that give the PLC the ability to converse via the RS-232 or Ethernet with nearly any computer. Of course, both the communications software and the particular applications software must be generated to provide an interface. Many vendors and systems houses are providing communications packages for various PLCs to run on microcomputers and PCs. These small systems offer low-cost operator interfaces to PLCs, providing data handling capabilities and the ability to be networked into a true distributed architecture. Microcomputers that have the ability to multitask, access large amounts of both RAM and nonvolatile memory, have proper software support, and are able to be networked will provide a good investment in terms of operator interface functions as well as total system capability.

Printers

Printers have always been an important part of the PLC system both as a development tool and for handling some of the operator interface functions. Two types of printers are commonly found in control rooms: dot matrix printers and laser printers. The dot matrix printers are used to print alarms and events as they occur. They are unique because as soon as a line of text is sent to the printer, the operator can read it. This is not possible with most types of printers currently popular. The laser printers are used for reports or screen printout and are either color or black and white.

Programmers and Workstations

The programmer unit provides an interface between the PLC and the user during program development, start-up, and troubleshooting. The instructions to be performed during each scan are coded and inserted into memory with the programmer.

Programmers vary from small handheld units the size of a large calculator to desktop stand-alone intelligent CRT-based units. These units come complete with documentation, reproduction, I/O status, and on-line and off-line programming ability.[26,27] Most PLCs can use a PC as the programming tool. A program for the PC allows it to interface with a serial or Ethernet port in the PLC.

Programming units are the liaison between what the PLC understands (words) and what the engineer desires to occur during the control sequence. Some programmers have the ability to store programs on floppy disks or on a hard drive. Another desirable feature is automatic documentation of the existing program. This is accomplished by a printer attached to the programmer. With off-line programming, the user can write a control program on the programming unit, then take the unit to the PLC in the field and load the memory with the new program, all without removing the PLC. Selection of these features depends on user requirements and budget. On-line programming allows cautious modification of the program while the PLC is controlling the process or the machine.

Manufacturers of PLCs provide two basic programming tools. These are handheld programmers and programming software that runs on PCs. The programming software may not be an option if a small PLC is purchased. Dedicated CRT programmers, once manufactured for the sole purpose of programming PLCs, have been replaced with PCs running PLC programming software.

The handheld programmer enables the operator to enter a program one contact at a time. These units are widely used because they are rugged, portable, and easy to operate. They are very cost-effective and give an engineer the capability to enter a program and to diagnose trouble in logic and field devices.

The programming software on a PC provides the engineer with a visual picture of the program in the PLC. Ladder diagrams are drawn on the screen, just as they would be drawn on paper. Design and troubleshooting time is reduced. With menu-driven software, programming training time is decreased. Information is stored on either floppy disks or the hard drive. Programs can be copied from one disk to another and then verified without need for loading the PLC's memory. With stand-alone programming an engineer can develop a program on the PC and then load it into the PLC.

PLC programming software generally provides complete documentation capability, including ladder diagrams, cross-reference listing, I/O listing, user-defined contacts, and coil names along with commentary above each network or rung. All of this is can be sent to a printer for hard copy documentation. The screen typically shows eight rungs of ladder logic by 11 contacts across. The ladder diagrams can be placed into the real-time mode, which allows visual contact status. A whole screen of contacts and coils can be updated in as fast as a few seconds. Laptop PCs are often used for portability. With a modem connection, these programs can be used at remote locations for programming and troubleshooting.

PLC programming software may be restricted for use on only one PC. Security keys are one way to obtain this isolation. Security keys are devices that plug into the back of the computer or a special key disk. Without the key the software will not run. Some manufacturers use alternate security schemes such as providing extra keys at group prices or issuing site license agreements. In any case, it is unlikely that a PLC manufacturer will ever allow the use of programming software on an unlimited number of PCs.

Some PC-compatible software allows the PLC to be emulated by the PC. This software is sold by the PLC manufacturer or a licensee and is often model-specific. If the software also offers on-line programming and troubleshooting characteristics, it might be usable on only a single specific PLC. This restriction is achieved by means of software or hardware keys that come with each copy of the software purchased.

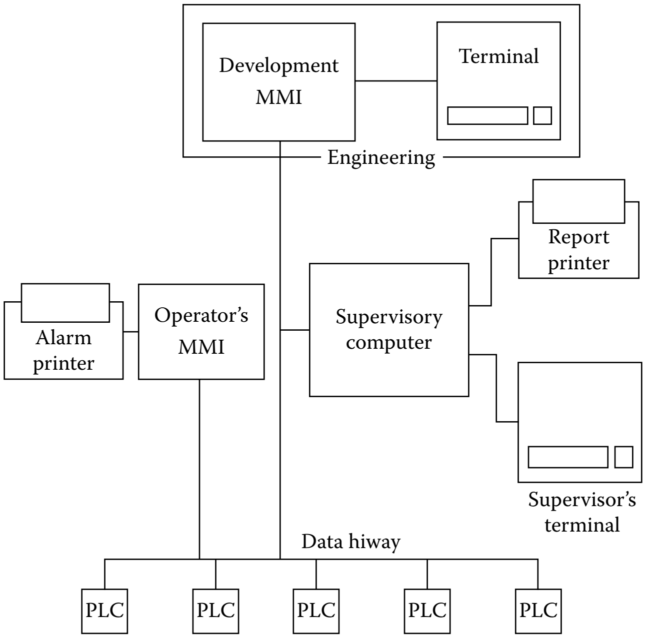


FIG. 5.4n
Networking with on-line engineering.

On-line engineering of PLC systems can be configured from a remote location like a control room with the combination of the programming/documentation tools and the distributed network. Systematic start-up and debugging of processes are available with this technique. Figure 5.4n depicts on-line engineering as part of a distributed system.

JUSTIFICATION FOR THE USE OF PLCs

For a given control problem, several technologies besides PLCs can be applied to achieve a solution: relays and stand-alone controllers, DCSs, and software running on PCs. PLCs differ from these technologies as described below.

PLCs vs. Relays and Stand-Alone Controllers

The main purpose of any control system is to get the process variable under control effectively and reliably. When control options are available, several factors can be taken into consideration in making an implementation decision. Some of these factors include the controller's cost, versatility, flexibility, and expandability.

Cost The ideas associated with cost are discussed first because this is usually the first issue users consider when selecting a control technology. In most cases, the two things initially known about a problem are the results desired and the budget available for fixing the problem. Unfortunately, at the beginning of a project the understanding of the problem is often limited. If the user's focus on the problem is too narrow, the solution might solve only the perceived symptoms of the problem while the problem itself may reappear in an alternate form at some other point in the process.

The budget assigned to solve a problem should be based not only on the initial investment required for the PLC hardware, but also on the costs associated with labor, maintenance, and downtime. Purchasing a larger and therefore more expensive PLC might in some situations be cost-effective when labor, maintenance, and downtime costs are also considered. Similarly, while PLCs may be more expensive than individual solid-state control units, if the indirect costs are also considered then the PLC becomes a cost-competitive alternative. As the cost of PLCs continues to drop, the cost of small PLCs may be equal to or even be less expensive than the cost of industrial relays.

Although significant, cost reductions alone are not the only reason that PLCs are the major replacement candidate for traditional relay logic. Compared with electromechanical relay systems, PLCs offer the following additional advantages:

- Ease of programming and reprogramming in the plant
- A programming language that is based on relay wiring symbols familiar to most plant electrical personnel
- High reliability and minimal maintenance
- Small physical size
- Ability to communicate with computer systems in the plant
- Moderate to low initial investment cost
- Rugged construction
- Modular design

Versatility The multifunction capability of a PLC allows control logic decision-making, enabling versatility rarely possible with other systems. The ability to combine discrete and analog logic is a powerful tool for the controls engineer. This is particularly evident in the control of batch processes. Entire start-up and shutdown sequences can be performed by the sequencer logic, and analog logic can be brought in during the batch run. Control of critical start-up parameters, such as temperature and pressure, can be precisely preprogrammed for each start-up step. Temperature stepping is easily programmed, as are the feedforward calculations that are used in some polymer reactors. All of these types of PLC applications are currently in use today and are well documented.[28,29,30,31]

Flexibility As a process goes on-line and is refined, the control equipment should be easily reconfigured to accommodate such modifications. The multifunction use of the PLC has already been discussed. In addition, digital blending applications, boiler control of either carbon monoxide or excess oxygen, and some other forms of optimizing control are also within the capabilities of PLCs. Because one common device performs multiple functions in a plant, fewer spare parts are needed, and the programming language is technician-friendly. In addition, the digital nature and self-diagnostic capabilities are strong additional justification for the PLC.

Expandability As a process matures, it is inevitable that enhancements will be added. These usually require more inputs and outputs. For hand-wired relay systems this usually necessitates extensive panel changes, which generally are problematic. A PLC easily accommodates the additional I/O without requiring changes in the existing wiring: The new points are merely placed in the system. If a PID loop or two is being added, no panel rework is necessary; only the wiring of the new points and some reprogramming to incorporate them are required. Conversely, if the initially selected PLC is "tight," additional I/O bases might be necessary. For this reason, most manufacturers recommend sizing the system to allow for 10 to 20% expansion.

Another advantage of the PLC is that it allows piecemeal implementation of projects. Systems can be brought on-line quickly and can be gradually converted to the PLC while on-line. The ability of the PLC to be reprogrammed while operating permits automation of processes that are too expensive to shut down. This technique is valuable to new as well as to retrofit projects (revamps).[32,33]

PLC vs. DCS

The capabilities of PLCs and DCSs have changed to the extent that today many applications that used to be the exclusive province of one or the other can be handled by both. The technology differences between PLCs and DCSs have evolved to be subtle or nonexistent. Manufacturers that had been making relays for logic and interlock applications developed PLCs, while DCS systems were developed by process control manufacturers with substantial experience in PID-type analog control. In the past it made good sense to use each type of controller in its area of superior experience. If the bulk of the I/O was digital (discrete), the logical choice was to use a PLC. If the I/O was mostly analog, a DCS system was selected. This logic is no longer universally true, and personal preference and end-user familiarity have become decisive factors in system selection. PLC and DCS comparisons have been taking place for as long as the two systems have existed. What's odd about this debate is that PLCs define only the controllers, whereas DCSs also include the software, operator interface, and network.[34]

In terms of pros and cons between the two designs, PLC I/O is likely to be more rugged and PLCs are likely to handle discrete logic faster than DCS systems. PLCs are also likely to be more desirable because their languages, such as ladder logic, are usually more familiar to plant personnel, and therefore there is less resistance to using them. On the other hand, ladder logic-type languages can be undesirable in some situations because they are not well suited to analog process control.

PLCs tend to be sold in a more piecemeal fashion as compared to the integrated systems approach common to DCSs. While PLCs are less expensive than DCSs, the end user often becomes more deeply involved with programming and systems integration issues. Nonetheless, many users find

PLCs to be a cost-effective solution for small to medium-sized process plants.[35]

Some users have overcome the limitations of PLCs by coupling them to PCs using custom-coded programming. The disadvantage of this approach is that such a nonstandardized system is usually understood fully only by its designer, and when that individual leaves the organization, the system can be ruined. When it comes to communications redundancy and data security, the DCS systems are superior. The DCS systems are also superior in their programming library, in advanced or optimizing control, in self-tuning algorithms, and, particularly, in their total plant architecture and information management capabilities.

PLC vs. Personal Computers

The PC is designed to be flexible and handle thousands of different types of applications, but PLCs are especially designed for control.[35] PCs started showing up on the factory floor in the mid-80s in programming and HMI applications. However, now the PC is migrating toward actual control. In response, PLC manufacturers have adapted PC technology, such as Ethernet. Some systems actually provide a “PC on a card” to plug into the PLC back plane.[11] PLCs are often thought of as computers. To a certain extent this is true; however, there are important differences between PLCs and computers.

Real-Time Operation/Orientation The PLC is designed to operate in a real-time control environment. The first priority of the CPU is to scan the I/O for status, make sequential control decisions (as defined by the program), implement those decisions, and repeat this procedure all within the allotted scan time. Most PLCs have internal clocks and “watchdog timers” built into their operations to ensure that a software error like “divide by zero” or an endless loop does not send the central processor into an undefined state. When the watchdog time is exceeded, the processor shuts down in a predetermined manner and usually turns off all outputs. In real-time systems, reliability is a big concern. PLC manufacturers' experience shows mean time between failures (MTBF) ranging from 20,000 to 400,000 hours. “This is far in excess of almost any other type of electronic or control equipment.”[3,6]

Environmental Considerations PLCs are designed to operate near the equipment they control. This means they function in hot, humid, dirty, noisy, and dusty industrial environments. Typical PLCs can operate in temperatures as high as 140°F (60°C) and as low as 32° F (0°C), with tolerable relative humidity ranging from 0 to 95% noncondensing. In addition, they have electrical noise immunities comparable with those required in military specifications.

Programming Languages and Techniques PLC languages are designed to emulate the popular relay ladder diagram format. This format is read and understood worldwide by maintenance technicians as well as engineers. Unlike computer programming, PLC programming does not require extensive special training. Applications know-how is much more important. Although certain special techniques are important for programming efficiency, they are easily learned. The major goal is the control program performance. Another difference between computers and PLCs is the sequential operation of the PLC. Program operations are performed by the PLC in the order they were programmed (Figure 5.4o). This is an extremely useful feature that allows easy programming of shift registers, ring counters, drum timers, and other useful indexing techniques for real-time control applications.

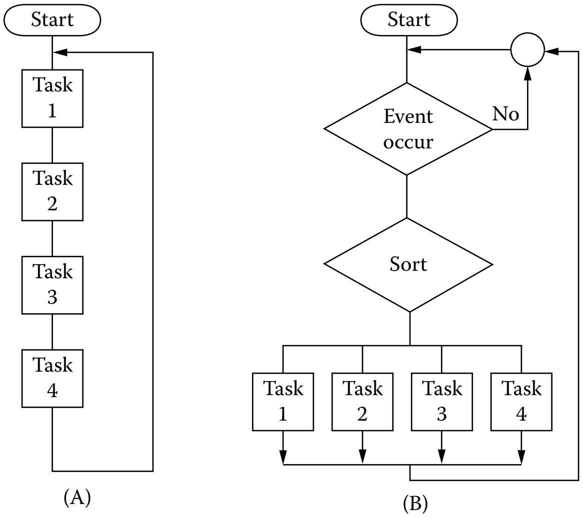


FIG. 5.4o

PLC versus computer. Program structure for a PLC (A) requires sequential execution with a scan, starting with task 1 and proceeding through task 4. The program structure for a general-purpose computer (B) permits task execution in any order.

Maintenance and Troubleshooting As a plant floor controller, the plant electrician or the instrument technician must maintain the PLC. It would be highly impractical to require computer-type maintenance service. To this end, PLC manufacturers build in self-diagnostics to allow for easy troubleshooting and repair of problems. Most PLC components are modular and simple to isolate; remove-and-replace (system modules) diagnostic techniques are usually implemented.

SUMMARY

Table 5.4p together with Figures 5.4q and 5.4r provide illustrative summaries of the PLC characteristics discussed thus far. Table 5.4p illustrates an I/O list for a milling application, while Figure 5.4q shows in some detail the PLC and all its key parts. The CPU is shown in the center of the figure as the PLC processor. Power for this unit is delivered by the

TABLE 5.4p
Example of a Detailed I/O List for a Milling Machine

Input	*Definition*			*Used in Rung(s)**					
XO		*MOA*	*Automatic*	*1–46*	*40–47*	*40–48*	*43*	*44*	*45*
X1	Right Shot	Pin In	L.S. 1	3	–3	9			
X2	Left Shot	Pin In	L.S. 2	3	–3	9			
X3	Machine	Slide Adv	L.S. 3	7	–22				
X4	Machine	Slide Ret	L.S. 4	3	8	–39			
X5	Shot Pins	Out	L.S. 5 & 6	6	–52				
X6	Pump	#1	3M	1					
X7	Right Bore	Slide Ret	L.S. 7	11	46				
X8	Left Bore	Slide Ret	L.S. 8	12	47				
X9	Right Bore	Slide Adv	L.S. 9	3	15	–19			
X10	Left Bore	Slide Adv	L.S. 10	3	15	–21			
X11	Swing	Clamp On	P.S. 1	6					
X12	Pull	Clamp On	P.S. 2	5					
X13	Swing	Clamp Off	P.S. 3	16					
X14		Cycle	Start	2	3	–17	–36	–36	–37
				–45					
X15		Cycle	Stop	1	2				
X16		Auto Lube	On/Off	40	48				
X17	Lube	Complete	Pin	–40	–48				
X18		Hi Lube	Level	37	–40				
X19		Lo Lube	Level	–40	–48				
X20		Manual	Light	–1	49	50	51	52	
X21	Pull	Clamp	On/Off	49	–52				
X22		Shot Pins	In/Out	50	–52				
X23	Swing	Clamp	On/Off	51	–52				

* Rung is a grouping of PLC instructions that controls one output or storage bit. This is represented as one section of a logic ladder diagram.

power supply shown to its right. The programming unit connected directly to it on the left is the way a ladder logic program line, like the one shown just above the CPU, is entered into the controller.

Representations of two I/O modules are shown in Figure 5.4r. The input module is on the left of the drawing and indicates a variety of contacts directly attached to seven of the eight points on the module. The point at address zero is shown as a spare for use as a replacement or future enhancement site. The output module shown has all eight of its output addresses (from address 140 through 147) in use.

Figure 5.4q shows an example of a remote I/O rack. Although the rack shown (rack No. 2) is powered by the main power supply, that is not a requirement. A remote I/O rack functions the same way as its direct I/O counterpart does. Various modules can be inserted in the rack to match the application's control needs.

PROJECT EXECUTION

Like any major enterprise, the PLC project must take into account the important considerations of schedule and budget. The PLC can facilitate the transition, however, by simultaneously pursuing several activities, thereby condensing the overall project schedule. A review of each major activity is presented in the following paragraphs.

Systems Analysis

The control system should be analyzed as a whole to determine plant control requirements. The PLC plays an integral part in these analyses, and its capabilities should be thoroughly understood by the controls engineer. Vital to systems analysis are the process and instrument diagram (P&ID), the descriptive operational sequence, and a logic diagram

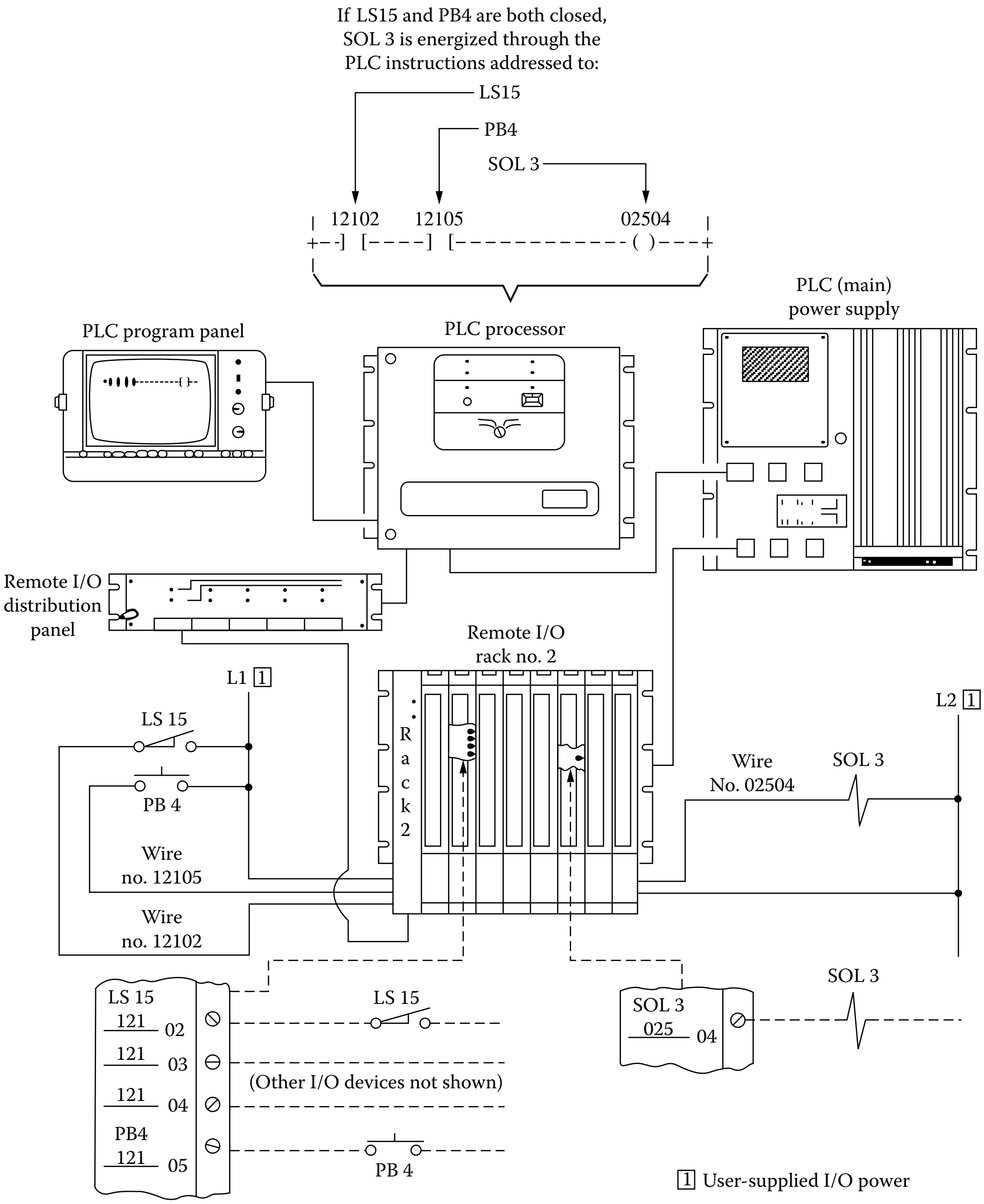


FIG. 5.4q
Configuration drawing. (Courtesy of Allen-Bradley.)

or electrical schematic. Part of this evaluation will be system sizing and selection. Once the appropriate PLC is selected and purchase orders are placed, two activities should begin immediately: engineering design and software development.

Systems Integrators One of the first decisions is whether to perform the work in-house or have a systems integrator do it. The general trend in the industry is that systems are more likely to be sold to OEMs and system integrators than to end users.[10] If an integrator will be used, it should be selected carefully. Some things to consider include:

- How well does the integrator know the PLC/HMI/reporting software?
- How well does it know your industry?
- How familiar is the integrator with your operations?
- How well can you define the operation of your project?
- How good is its training?
- Will it have time and budget to fine-tune the system?

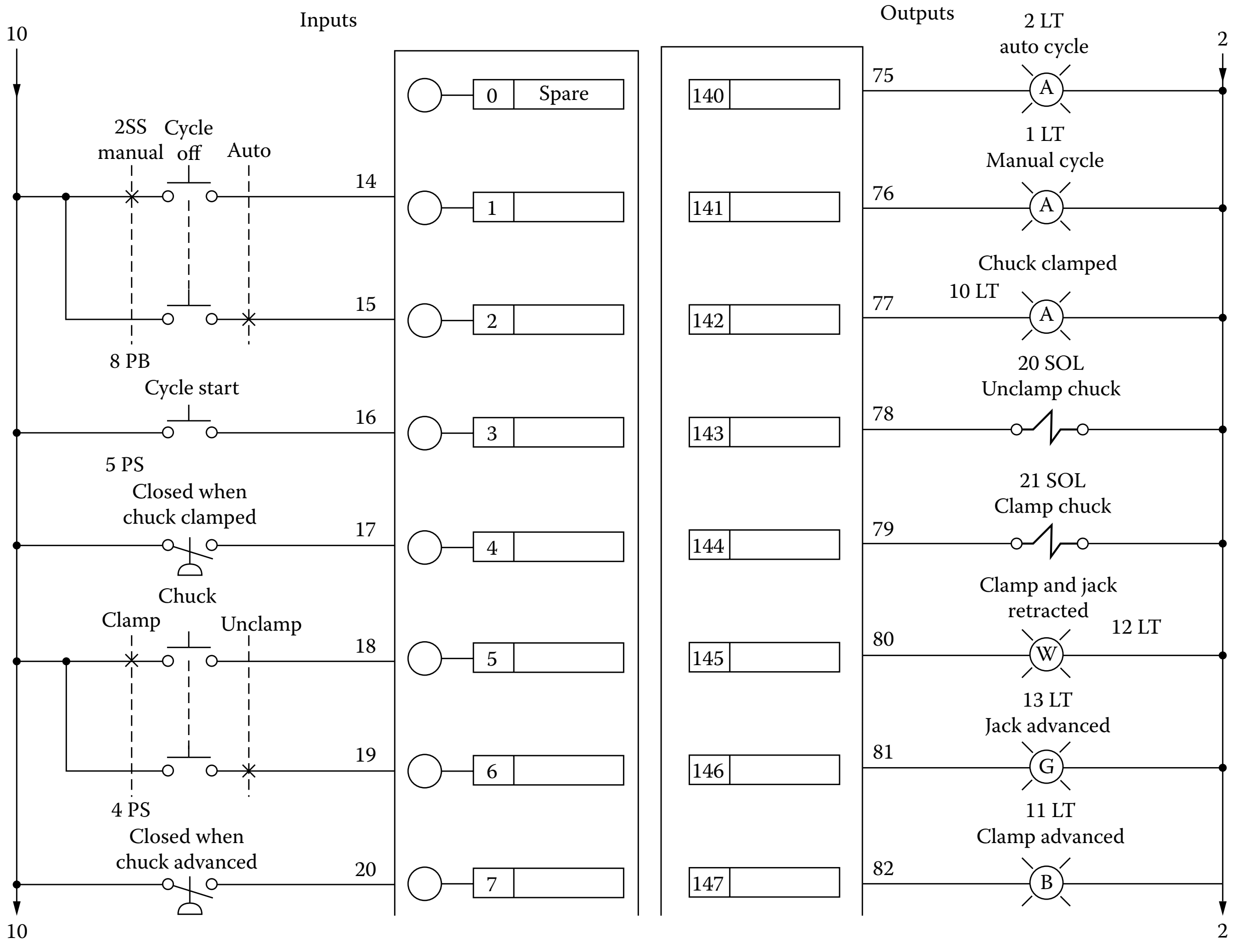


FIG. 5.4r
Point-to-point wiring diagram.[14]

- After warranty, will you be able to make changes without the integrator or will you have to pay for it to return?[36]

For budgeting, whether the work is done in-house or by an integrator, software costs range from 50 to 100% of the hardware cost. System engineering costs and documentation costs each range from 25 to 50% of hardware cost. Therefore, the total cost (without installation labor) is about twice (if not more) the hardware cost.

Open Systems

Another consideration is whether to use open systems or proprietary. A major benefit of using open systems is that a number of different manufacturers can supply equipment for the system. “The trend today is definitely towards open systems and open languages, but the bulk of the PLC market is still proprietary—about 75% proprietary and 25% open or standards-based. Even so, a large number of proprietary PLCs are shipped with Profibus, DeviceNet, or other open networking capabilities.”[7]

Distributed Control “One of the effects of better networks and intelligent devices is that control function no longer has to reside in a central controller. It can be scattered about, either in a network of small PLCs or in intelligent I/O bases.… Some distributed processing is good to have because you can divide an operation into logical pieces, but it becomes difficult to partition programs and logic and so forth into the machines.”[11] A potential benefit dividing a single large system into separate areas of control is that starting and troubleshooting a smaller piece of the system is more manageable.[37]

Redundancy Another issue is the amount of redundancy that will be built into the control system. Some areas to consider when evaluating the need for redundancy include preventing production impacts, equipment damage, business interruption and quality impacts, liability costs, public health and safety, site safety, and environmental impacts.[38] Redundant PLCs are available and I/O can be made redundant but it is often better to provide diversity in redundancy planning. Instead of using PLCs, manual controllers can be used. Manual control stations are important as backups in case of failure of the PLC controlling PID loops. Operator stations and HMI often provide manual control capabilities but still rely on the integrity of the PLC, so they are not truly manual in the hardwired sense. A manual control station is an important part of the distributed control system because it gives true manual control of the loops locally or in the control room, even when the local control systems are down. It is important to look at the overall system when evaluating redundancy. Unfortunately, the term "control-system redundancy" has come to refer to only the CPU, possibly the I/O but not the field devices. Even redundant PLCs can "fail" if they are powered from a single power source.

PLC Hardware, System Sizing, and Selection

Despite the variety of available PLC models, system sizing is relatively simple. Hardware and system size can be determined by an analysis of the following system characteristics:

1. I/O quantity and type
2. Remote I/O requirements
3. Memory quantity and type
4. Programming requirements
5. Programmers
6. Peripheral requirements

Although sizing is generally straightforward, selection of the right PLC requires considerable judgment regarding trade-offs between future requirements and present cost. The first three are examined below. The others are addressed elsewhere in this chapter.

I/O Quantity and Type (Example 1) For this simple system, the first step is development of the I/O list (Table 5.4p). This detailed document will be used extensively and should be developed with great care. (Once I/O numbers are assigned, it becomes very difficult to change all references to these numbers.) If switches and pilot lights are used to operate the controls, include I/O for them. Include any alarm outputs used for alarm annunciators. An intrusion switch on the panel door is useful for larger control panels. Also, consider allowing 20% spare I/O of each type used. The I/O list is followed by the configuration drawing.

The configuration drawing (Figure 5.4q) shows the arrangement of the I/O and support hardware. The point-to-point wiring diagram (Figure 5.4r) is used by the panel shop and the installation contractor to make the I/O device interconnect. Panel, or enclosure, design should now be coordinated with the addition of panel instrumentation, such as light switches, meters, and recorders. Once these steps are completed, panel fabrication and assembly can begin.

I/O Quantity and Type (Example 2) In this somewhat more complicated example, the user should arrange any special I/O types as well as the commonly available modules according to an I/O matrix by logical area, as shown in Table 5.4s. In this table, I/O types are listed across the first row and plant areas are listed down the first column. In this way, one can accurately reconstruct the decision-making process concerning I/O quantity and type. It is important to include at least 10 to 20% spare rack space in all I/O considerations.

For example, consider a typical process application. Assume a total I/O count of 764, broken down into 436 inputs and 328 outputs. This application falls into the *medium* PLC category. Because the majority of the field devices are located a good distance from the CPU, a PLC with remote I/O is desirable. The I/O requirements by locations are as follows:

Process Area: Total of 390 inputs and outputs, of which 230 are 24 V (discrete DC inputs), 24 are 4–20 mA analog inputs, and 136 are 24 V DC discrete outputs.

Tank Farm #1: Total of 98 I/O, of which 62 are 120 VAC discrete inputs (DIs) and 36 are 120 VAC discrete outputs.

Tank Farm #2: Total of 86 I/O, of which 52 are 120 VAC discrete inputs and 34 are 120 VAC discrete outputs.

Loading Station: Total of 190 I/O, of which 68 are 120 VAC discrete inputs and 122 are discrete outputs (of which 24 are 120 VAC and 98 are 240 VAC).

Let us assume that the PLC system being considered has the following features: (1) No constraints on input and output mixture; (2) the I/O modules are available in two formats, 16 points per module and 8 points per module; and (3) 10% spare I/O is required. For the sake of illustration, we will use the 16 points per module I/O structure in the process area and the 8 points per module structure in the tank farms and loading station. The I/O distribution (including spares) per location would now be as follows:

Process Area: Total of 390 I/O points, which require the following modules:
- 230 + 10% = 253 24 V DC inputs with 16 points/module = 15.8 modules; use 16 modules
- 24 + 10% = 26.4 4–20 mA analog inputs at 16 points per module = 1.6; use 2 modules
- 136 + 10% = 149.6 24 V DC discrete outputs at 16 points per module = 9.35; use 10 modules

Tank Farm #1: Total of 98 I/O points, which require the following modules:
- 62 + 10% = 68.2 are 120 VAC discrete inputs at 8 points per module = 8.5; use 9 modules
- 36 + 10% = 39.6 are 120 VAC discrete outputs at 8 points per module = 4.9; use 5 modules

Tank Farm #2: Total of 86 I/O points, which require the following modules:
- 52 + 10% = 57.2 are 120 VAC discrete inputs at 8 points per module = 7.2; use 8 modules
- 34 + 10% = 37.4 are 120 VAC discrete outputs at 8 points per module = 4.7; use 5 modules

TABLE 5.4s
I/O Matrix

Plant Area	*Analog*	*In*	*Model No.*[2]/*Qty*	*Analog Out*	*Model No.*[2]/*Qty*	*Discrete In*/*Voltage*	*Model No.*[2]/*Qty*	*Discrete Out*/*Voltage*	*Model No.*/*Qty*
Process area[1]	24	4–20 mA	—/2	0		230/24 VDC	—/16	136/24 VDC	—/10
Tank farm #1[3]	0			0		62/120 VAC	—/9	36/120 VAC	—/5
Tank farm #2[3]	0			0		52/120 VAC	—/8	34/120 VAC	—/5
Loading station[3]	0			0		68/120 VAC	—/10	24/120 VAC	—/4
								98/240 VAC	—/14

Notes: (1) Discrete in 16 pts/card
Discrete out 16 pts/card
Analog in 16 pts/card
Analog out
(2) Add model numbers after award of contract.
(3) DI and DO are 8 points per card

Loading Station: Total of 190 I/O points, which require the following modules:

- 68 + 10% = 74.8 are 120 VAC discrete inputs at 8 points per module = 9.4; use 10 modules
- 24 + 10% = 26.4 are 120 VAC discrete outputs at 8 points per module = 3.3; use 4 modules
- 98 + 10% = 107.8 are 240 VAC discrete outputs at 8 points per module = 13.5; use 14 modules

If we assume that one remote communications channel can service up to 128 I/O in groups of sixteen 8-point modules or eight 16-point modules, the system becomes:

- Process Area—390 I/O; 28 16-point modules; four remote channels
- Tank Farm #1—98 I/O; 14 8-point modules; one remote channel
- Tank Farm #2—86 I/O; 13 8-point modules; one remote channel
- Loading Station—190 I/O; 28 8-point modules; two remote channels

Remote I/O vs. Distributed Control A unique feature of the PLC is the multiplexed nature of the I/O bus. This can be used to great advantage to reduce overall wiring cost. If I/O racks are centralized in logical clusters, plant wiring requirements can be greatly reduced. Wiring between racks and the CPU can be reduced to a few twisted pairs of wires or a single cable. The tremendous cost savings that result can be realized without a compromise of control accuracy or capability.

A system configuration diagram (such as that shown in Figure 5.4q), when used in conjunction with the I/O matrix in Table 5.4s, aids in keeping track of the overall system configuration.

It is important to remember the major weakness of remote I/O systems. If the bus is cut or interrupted, the effects of I/O failure will be relatively unpredictable. One must consider the effect of a possible system failure on each step in the sequence. Some users install a redundant version of remote I/O communications to guard against the loss of remote I/O communications. For this reason, duplication of smaller CPUs at each remote location is often considered preferable to a large central CPU. This is actually an extension of distributed control within the network of the PLC itself. This approach can be very cost-effective, because requirements for the central unit size can be reduced. Serious consideration should be given to distributed versus centralized architecture in remote I/O systems in which control system integrity is important.

Memory Quantity and Type The type and quantity of PLC memory used depends on the controller's size and the company that manufactured it. Most small PLCs come with a fixed quantity of RAM. Although this is usually 2 K to 4 K of memory, the actual number of memory locations is not as important as the average size application program the PLC can be expected to handle. (In this case, size refers to the number of I/O points that are to be controlled and the average number of logic, timer, counter, and mathematics operations that are to be performed.) Some manufacturers may provide an extra-expense option of EEPROM or Flash memory with their small PLCs.

Midsize and large PLCs provide users an option for almost any type of memory desired. This includes various types of nonvolatile memory. Quantity limits imposed by the PLC will exceed most application demands. When this is not the case, it usually suggests that a more efficient control scheme is in order or that the application really does not belong on a PLC in the first place.

Total memory, as stated in the manufacturer's literature, does not necessarily mean the entire content is available to the user. Some manufacturers reserve large blocks for the PLC executive. A system with 4 K of 16-bit words of user memory may comfortably accommodate a program, whereas another system with 8 K of 8-bit words may have too little memory for the same program.

Special programming language features are an important aspect of memory sizing, especially in process control. The PID algorithm is a perfect example: One manufacturer requires 33 words of user-available memory, whereas another may need in excess of 1000 words. Obviously, the memory sizing for a loop control program would vary in these two systems. Another example is the use of special functions, such as shift registers. An alternative way of developing a shift register in ladder logic is to use a special function shift register or handling data to require less user memory. Word (or register) moves are also powerful in terms of memory efficiency. Programming languages, which can be binary- or octal-based or alphanumeric Boolean, affect memory use. The closer the language is to a machine code (binary-based), the more user memory is required to perform the more complex functions. The closer the language is to alphanumeric Boolean, the less memory will be required for complex functions.

The best way to determine program memory prerequisites is to write a representative sample program reflecting some actual project requirements and to request information about user memory size from the various manufacturers. If the manufacturer's suggestions are followed, the user can be reasonably assured that the memory will not be undersized.

The final area of caution about memory size concerns the consideration of data storage. Data tables, scratch pads, and historical data retrieval requirements can inflate the size of the PLC memory. It should be remembered that the primary task of a PLC is control of the process. If data requirements are large, connection to auxiliary devices, such as mini- and microcomputers, should be given serious consideration. Many of these devices are currently available and are of an industrial grade; furthermore, the price of these systems is coming down rapidly. It is not good engineering practice to degrade control capabilities by burdening the PLC with excessive data acquisition functions. As a plant goes online, operational requirements for data generally increase astronomically. These will be easily accommodated by a mini- or microcomputer but not by the PLC memory.

PLC Installation and Panel Design

Installation of PLC systems is not a difficult or mysterious procedure, but the following general rules will save time and trouble for the systems designer or installer. The basic principles of PLC installation are the same as those for the installation of relay or other control systems. Safety rules and practices governing proper use of electrical control equipment in general should be observed. These include correct grounding techniques, placement of disconnect devices, proper selection of wire gauge, fusing, and logical layout of the device. PLCs can often be retrofitted into existing hardwired relay enclosures because they are designed to withstand the typical plant environment. PLC vendors provide installation manuals upon request.

After the PLC equipment is selected, the support equipment will need to be identified. Some panel shops that build PLC panels can assist or perform this additional design. At the panel level this includes terminal blocks, wireway, wire type and size both for discrete and analog, loop power supply (usually 24 V DC), circuit breakers and fuses, isolation relays, intrinsic safety barriers (if I/O wiring enters flammable or explosive environments as defined by national electric code), main isolation transformer, pilot lights, pushbuttons and switches, internal panel work lights, and receptacle. Also included are any door-mounted equipment like chart recorders, digital panel meters, and operator interface panels. A UPS may be required. Separate isolated grounding is sometimes used for loop shields. Motor starters and small variable frequency drives may also be in the same enclosure.

After all the equipment in the panel is determined, a heat calculation should be done for the ambient extremes expected, particularly if the cabinet is outdoors, to verify the internal temperature is well below the maximum temperature of all components. Generally the PLC is not the limiting factor here, but other electronics and power supplies may limit the maximum allowable temperature. If the interior temperature is too high, a heat exchanger or air conditioner may be required. If the panel is installed in a cold area, a heater may be needed to prevent condensation, which can ruin electronic equipment.

Safety Considerations Perhaps the most important safety feature, which is often neglected in PLC system design, is emergency stop and master control relays. This feature must be included whenever a hardwired device is used in order to ensure operator protection against the unwanted application of power. Emergency stop functions should be completely hardwired (Figure 5.4t). In no way should any software functions

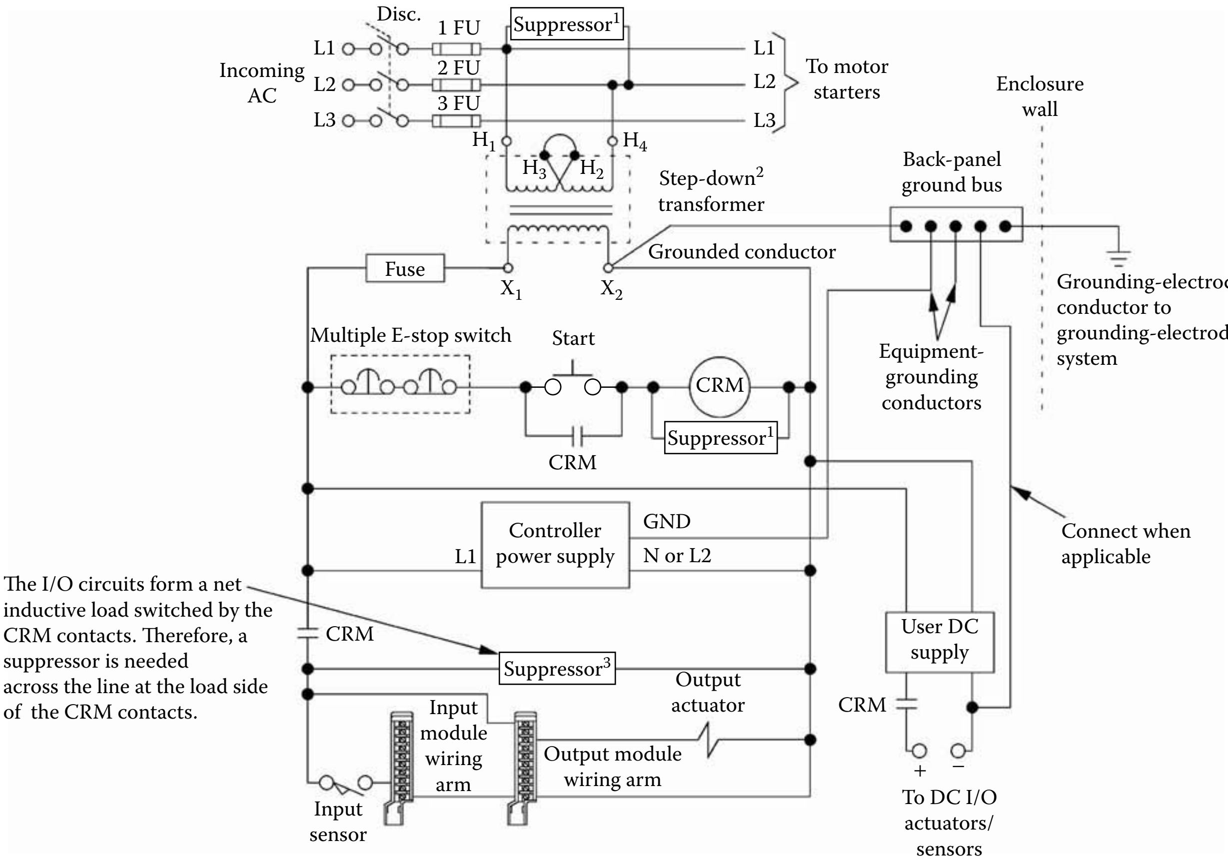


FIG. 5.4t

Grounded AC power-distribution system with master-control relay.

Notes:

1. To minimize EMI generation, connect a suppressor across an inductive load.
2. In many applications, a second transformer provides power to the input circuits and power supplies for isolation from the output circuits.
3. Connect a suppressor here to minimize EMI generation from the net inductive load switched by the CRM contacts. In some installations, a 1 µf 220 ohm suppressor or 2 µf 100 ohm suppressor has been effective. (Courtesy of Allen-Bradley)

be relied upon to shut off the process or the machine. Disconnect switches and master control relays should be hardwired to cut off power to the output supply of the PLC. This is necessary because most PLC manufacturers use triacs for their output switching devices, and triacs are just as likely to fail on as off. This feature is often required by local or national codes. In the United States, NFPA 79, "Electrical Standard for Industrial Machinery," covers this subject.[39,40,41,42]

Implementation Planning ahead is every bit as important in designing a complete PLC system as in laying out a relay logic panel. Care in counting I/O points in the beginning—and leaving a safety factor—will save headaches in the panel fabrication stage. Panels should always have plenty of expansion room left over, because I/O is invariably added as the job progresses and the operators see the advantages of PLCs. The designer should refer to the layout considerations provided by the manufacturer. Extra space should be left to provide access to the boards and connectors of the PLC. The diagnostic and status indicators should all be visible. The designer should leave room between I/O racks for wireways and large hands.

One good technique for ensuring efficient panel layouts is to involve maintenance personnel in the design procedure. This not only optimizes the layout but also introduces the staff to the hardware (Figure 5.4u).

In general, the best defense against creating a tangled mess when designing a PLC system is to follow proper documentation techniques. A little more time spent documenting panel layout, I/O counts, and wiring diagrams results in a lot less time spent starting up the system. PLCs can handle large amounts of I/O points with varying electrical characteristics, so things can get pretty confusing in a hurry. Cable requirements between hardware boxes vary from one type of PLC to another, so this is an important consideration in panel layout.

Enclosure Enclosures should nearly always be provided for the PLCs themselves. This protects the electronics from moisture, oil, dust particles, and unwanted tampering. Most

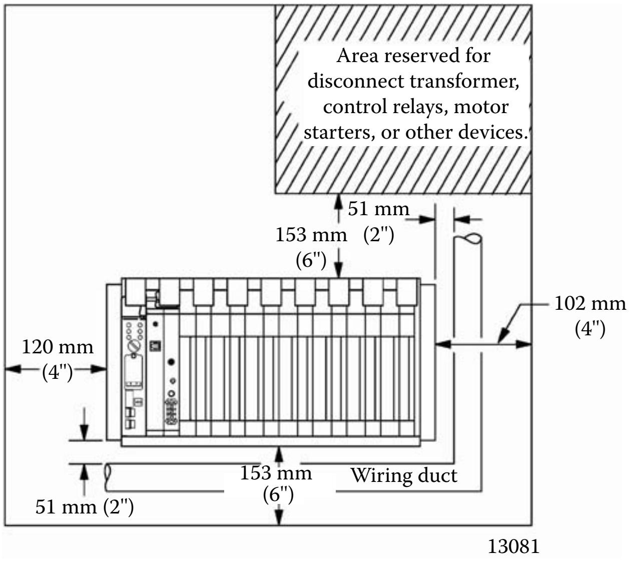
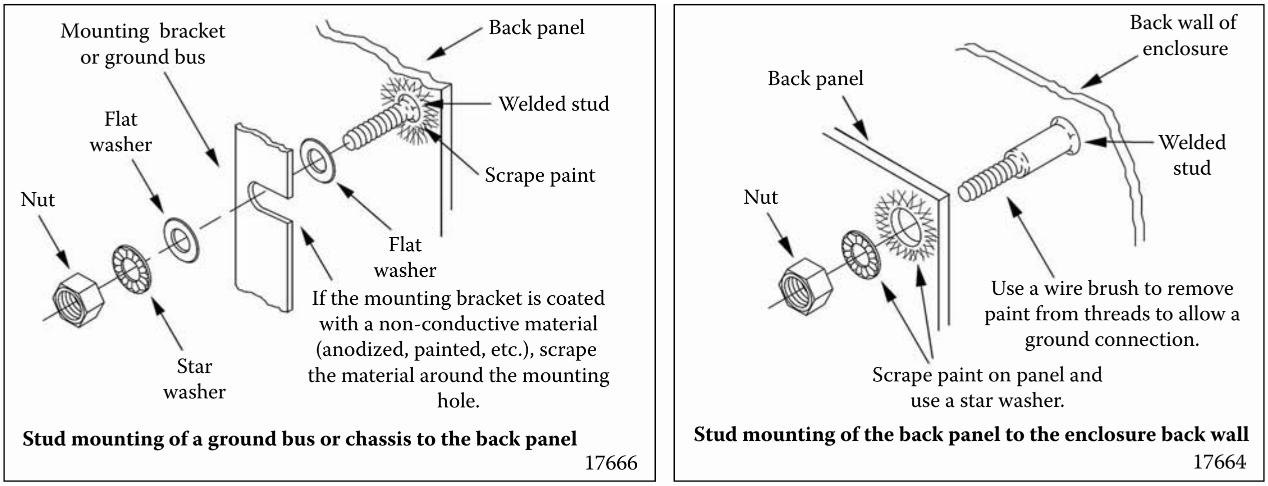


FIG. 5.4u
Typical enclosure layout. (Courtesy of Allen-Bradley.)

manufacturers recommend a NEMA 12 enclosure for the standard industrial environment or a NEMA 4X for outdoor or corrosive environments. A NEMA 3R is not recommended because it is not sealed against dust and moisture. This type of enclosure is readily available in a variety of sizes and, in fact, may be already included with a new system.

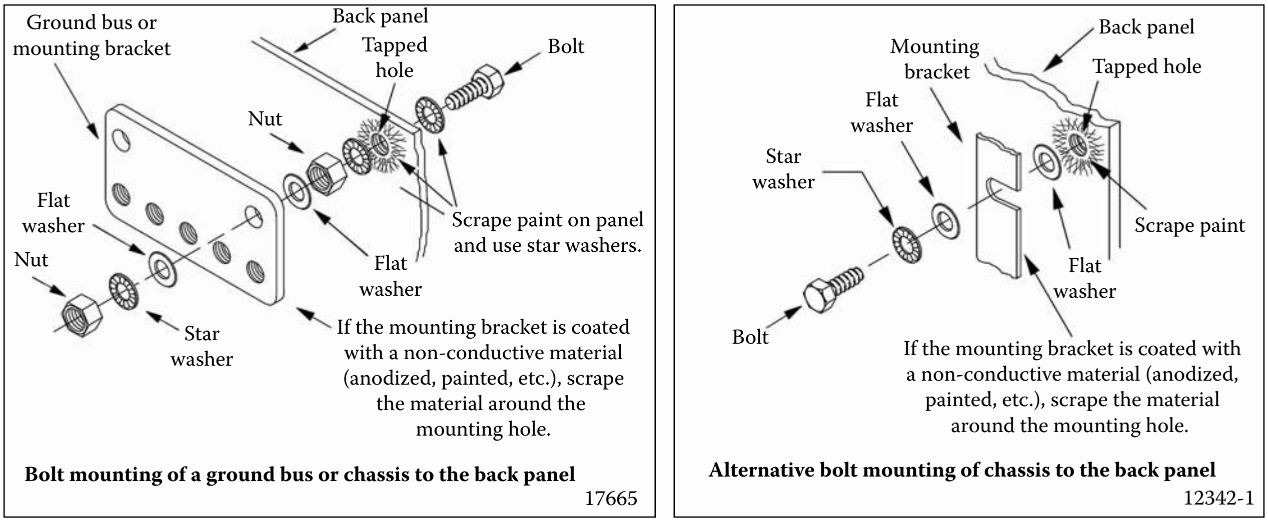
PLCs are designed to be located close to the machine or the process under control. This keeps the wiring runs short and aids in the troubleshooting procedure. At times, however, mounting the PLC directly on the machine or too close to the process is not advisable, such as in cases of vibration inherent in the machine, electrical noise interference, or excessive heat problems. In these situations, the PLC must be either moved away or successfully protected against these environmental conditions.

Temperature Considerations Installing any solid-state device requires paying attention to ambient temperatures, radiant heat bombardment, and the heat generated by the device itself. PLCs are typically designed for operation over a broad range



Stud mounting of a ground bus or chassis to the back panel

Stud mounting of the back panel to the enclosure back wall



Bolt mounting of a ground bus or chassis to the back panel

Alternative bolt mounting of chassis to the back panel

FIG. 5.4v
Ground connection details. (Courtesy of Allen-Bradley.)

of temperatures, usually from 0 to 60°C. When analyzing the proposed PLC environment, however, one should remember that enclosure temperatures usually run a few degrees higher than ambient temperatures. Radiant heat on an enclosure from surrounding tanks can raise the internal temperature beyond that specified by the manufacturer.

Heat generated by the PLC is a key issue when the device is placed in ambient temperatures close to the extreme mentioned in the specifications. The temperature rise caused by the power consumption of the PLC itself is not hard to estimate. In addition, most manufacturers will provide a notation of the power consumption of the triacs driving field loads. When designing the hardware layout within the panel, one should adhere to the manufacturer's suggestions regarding ways to minimize heating problems. Most PLCs use convection over fins to take heat away from particular areas within the hardware. Care must be taken to ensure that no obstruction to air flow over these fins is introduced by placement of the PLC in the enclosure. Wireways are typically provided with holes to allow air to pass through. Generally, one can avoid problems with PLC enclosures by simply leaving plenty of air space around the heat producers.

Should all of these factors combine to cause a temperature problem, the panel can be vented, air conditioned, or moved to another location. Usually, simply blowing filtered air through the enclosure will resolve minor difficulties. If air conditioning is required, small units that are designed for cooling electronic enclosures are readily available.

Noise Noise or unwanted electrical signals can generate problems for all solid-state circuits, particularly microprocessors. Each PLC manufacturer suggests methods for designing a noise-immune system. These guidelines should be strictly followed in the design and installation phases, because noise problems can be very difficult to isolate after the system is up and running. I/O systems are isolated from the field, but voltage spikes can still appear within the low-voltage environment of the PLC if proper grounding practices are not followed.

A well-grounded enclosure can provide a barrier to noise bombardment from outside. Metal-to-metal contact between the PLC and the panel is a must, as is a good connection from the panel to the ground (see Figure 5.4v). Noise producers within the panel should be noted during the panel design phase, and the PLC must not be located too close to these devices. Wiring within the panel should also be diverted around noise producers to avoid picking up any stray signals. Often, it is necessary to keep AC and DC wiring bundles apart, particularly when high-voltage AC is used at the same time that low-level analog signals are present. Refer to Tables 5.4w and 5.4x and Figure 5.4y for recommendations.

TABLE 5.4w
Follow These Guidelines for Grouping Conductors with Respect to Noise (Courtesy of Allen-Bradley)

Group Conductor Cables Fitting This Description	*Into This Category:*	*Examples:*
Control and AC power-High-power conductors that are more tolerant of electrical noise than category 2 conductors and may also cause more noise to be picked up by adjacent conductors Corresponds to IEEE levels 3 (low susceptibility) and 4 (power)	Category 1	AC power lines for power supplies and I/O circuits. High-power digital AC I/O lines—to connect AC I/O modules rated for high power and high noise immunity. High-power digital DC I/O lines — to connect DC I/O modules rated for high power or with input circuits with long time-constant filters for high noise rejection. They typically connect devices such as hard-contact switches, relays, and solenoids.
Signal and communications—Low-power conductors that are less tolerant of electrical noise than category 1 conductors and should also cause less noise to be picked up by adjacent conductors (they connect to sensors and actuators relatively close to the I/O modules) Corresponds to IEEE levels 1 (high susceptibility) and 2 (medium susceptibility)	Category 2	Analog I/O lines and DC power lines for analog circuits. Low-power digital AC/DC I/O lines — to connect to I/O modules that are rated for low power such as low-power contact-output modules. Low-power digital DC I/O lines — to connect to DC I/O modules that are rated for low power and have input circuits with short time-constant filters to detect short pulses. They typically connect to devices such as proximity switches, photo-electric sensors, TTL devices, and encoders. Communications cables (ControlNet, DeviceNet, Universal remote I/O, extended-local I/O, DH+, DH-485, RS-232-C, RS-422, RS-423 cables) — to connect between processors or to I/O adapter modules, programming terminals, computers, or data terminals.
Intra-enclosure — Interconnect the system components within an enclosure Corresponds to IEEE levels 1 (high susceptibility) and 2 (medium susceptibility)	Category 3	Low-voltage DC power cables — to provide backplane power to the system components. Communications cables — to connect between system components within the same enclosure.

TABLE 5.4x
Follow These Guidelines for Routing Cables to Guard against Noise (Courtesy of Allen-Bradley)

Route This Category of Conductor Cables:	*According to These Guidelines*
Category 1	These conductors can be routed in the same cable tray or raceway with machine power conductors of up to 600 VAC (feeding up to 100 hp devices).
Category 2	If it must cross power feed lines, it should do so at right angles. Route at least 5 ft from high-voltage enclosures, or sources of RF/microwave radiation. If the conductor is in a metal wireway or conduit, each segment of that wireway or conduit must be bonded to each adjacent segment so that it has electrical continuity along its entire length, and must be bonded to the enclosure at the entry point. Properly shield (where applicable) and route in a raceway separate from category 1 conductors. If in a contiguous metallic wireway or conduit, route at least 0.08 m (3 in) from category 1 conductors of less than 20 A; 0.15 m (6 in) from AC power lines of 20 A or more; but only up to 100 kVA; 0.3 m (1 ft) from AC power lines of greater than 100 kVA. If not in a contiguous metallic wireway or conduit, route at least 0.15 m (6 in) from category 1 conductors of less than 20 A; 0.3 m (1 ft) from AC power lines of 20 A or more; but only up to 100 kVA; 0.6 m (2 ft) from AC power lines of greater than 100 kVA.
Category 3	Route conductors external to all raceways in the enclosure or in a raceway separate from any category 1 conductors with the same spacing listed for category 2 conductors, where possible.

Line voltage variations can cause hard-to-trace problems in the operation of any computer-based system. PLCs are no exception, even though they are designed to operate over a much larger variation in supply voltage. Large spikes or brownout conditions can cause errors in program execution. Most manufacturers protect against this, enabling the controller to come up running after a brownout, but these measures may not be acceptable in all applications. The designer may wish to add an isolation transformer to a proposed PLC system, sized for twice the anticipated load. This is cheap insurance, and PLC manufacturers will help determine the required load.

Triac outputs require some special attention that will be new to relay users. Triacs used for AC loads typically leak a small amount of current. In the case of triac outputs from a PLC, this leakage may be enough to keep panel lamps glowing or small relays energized. When a triac is used to switch the input on a PLC, the leakage may be enough to make the PLCs "think" the input is on. A dummy load (shown in Figure 5.4z) can be used to drain this leakage when the input should be off. Whenever a mechanical contact is used in series with a load energized by a triac (as shown in Figure 5.4aa), a resistance-capacitance (RC) network should be used as shown to protect the triac from inductive kickback. A varistor should be provided in parallel with a load whenever the load can be "hot-wired" around the triac (Figure 5.4bb). The user should check with the PLC manufacturer for the suggested RC and metal oxide varistor (MOV) types for the particular application. Triacs cannot directly drive large motor starters and similar devices. PLC manufacturers provide surge specifications for the various I/O cards. Sometimes an interposing relay or dry contacts will be required for large loads.

PLCs are similar to most electrical control systems. To be sure, solid-state devices, microprocessors, and triacs require some special considerations during the design, installation, and start-up phases of a project, but these concepts are not too complex or difficult to assimilate. As always, good design habits in the beginning will ensure a safe and reliable control system.

Hookup PLC panels can be very neat and orderly if all the terminals are arranged in a logical fashion. The actual result is a direct function of the time spent during the design process. Interposing terminal blocks between the PLC I/O structure and the field is suggested, because the terminations provided by PLC manufacturers are shrinking in the race to provide higher-density I/O. This also gives the panel designer the ability to place the field termination points where they are easily accessed. Wiring ducts keep the panel neat and protect the wire from mishap.

Following good wiring practices can avert many noise problems. Low-voltage signal wiring should be kept away from noise sources. Analog signals should be shielded, with the shield terminated at an isolated ground in the panel only (to prevent shield ground loops). Again, these analog signals should be separated from power wiring.

Software (Program) Development

The I/O list mentioned previously will be used to begin the program development. Basic control philosophy decisions need to be made at this point. Should valves fail open or closed? What fail-safe provisions are necessary for analog control? These philosophical decisions should be documented and included with the process operational descriptions.[15] Often this document will be referred to as the software functional specification. Its purpose is to define, as precisely as possible, the operation of the controls.[43] It also has several other functions:

1. It communicates the functional requirements of the control system to those writing the PLC code.

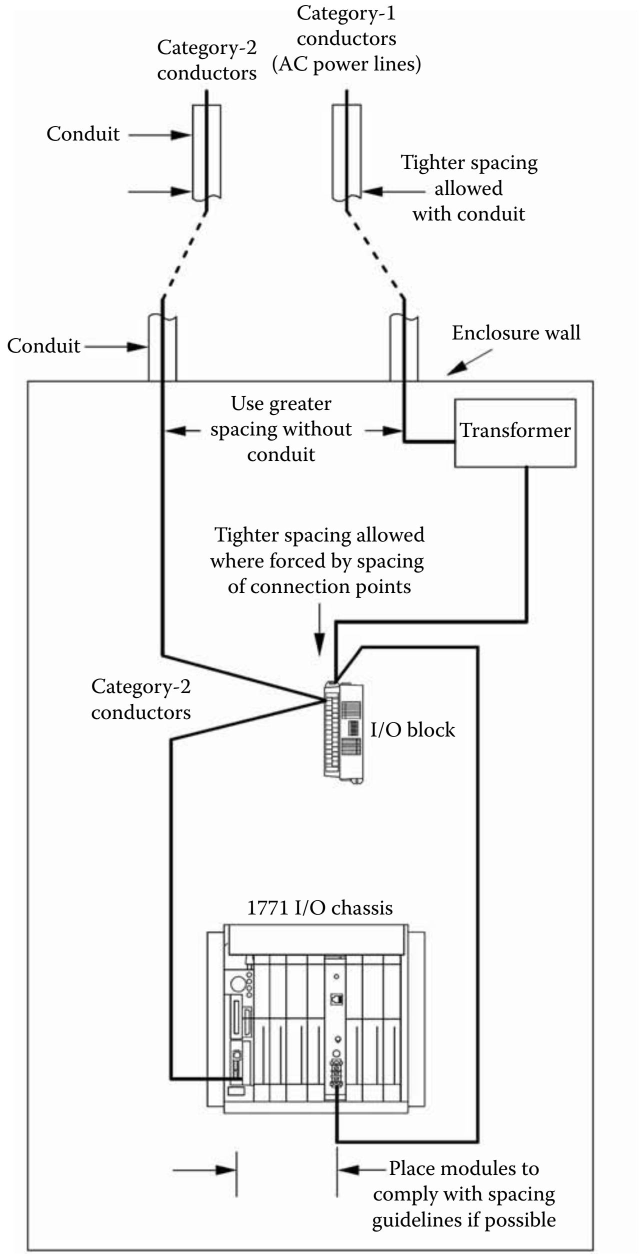


FIG. 5.4y
Mounting and wiring details. (Courtesy of Allen-Bradley.)

2. It records the thought process (regarding control) of the system designer to be used in the event of a personnel change. Such information can be invaluable.[44]
3. It provides a review document for personnel working in other capacities (mechanical, process, and project management) to ensure that they understand the operation of the controls.
4. It provides a guide for developing the operational description for the operator's manual.

After the functional specification has been reviewed and approved, a detailed operational sequence chart, timing

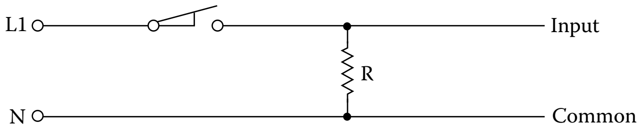


FIG. 5.4z
Dummy load for leakage. If a "leaky" triac, such as a proximity switch, is used to switch the hot side of the line, then a pull-down resistor (R) is required to counteract the leakage.

diagram, logic diagram, flowchart, or electrical schematic is developed from it. This schematic is translated or coded into the appropriate PLC language, cross-referencing I/O with PLC designator tags. The piping and instrument diagram is also cross-referenced with PLC designators. In this way, future cross-referencing of system drawings and PLC codes is facilitated.

As the code is entered, a memory map or register index is kept by the programmer (Table 5.4cc). This map is useful for organizing program data in logical arrangements and will prove invaluable during start-up, when the programmer may need to locate available blocks of memory quickly for program revisions. Most PLC programming software can generate a cross-reference; however, certain types of instructions are difficult for the program to cross-reference, so a list is still desirable.

The use and understanding of PLC programming depends on knowledge of the process to be controlled, an understanding of electrical schematics, and an appreciation for logic operations and for various types of logic and relay devices. Other sections of this chapter provide information on each of these topics. A review of those topics might be useful at this point.

Although the programming style and language used is, to some extent, dictated by the size of the PLC used, there are fundamental programming elements, including logic operations, timers, counters, and arithmetic capabilities, that are provided in all models. Some of the characteristics of these important elements are briefly discussed below (see Sections 5.5 and 5.6 for more information). Ladder logic continues to be popular because it was purposely created to look like hardwired relay logic drawings already in use. It is this similarity between ladder diagrams used for relays and the programming version that eased the change from using relays to using PLCs. In addition, ladder logic is relatively easy to learn.[9,45]

Boolean Logic Relay type instructions are the most basic of all PLC instructions. Boolean logic functions (AND, OR, and NOT) are diagrammed as combinations of normally open and normally closed contacts. A coil symbol represents the result of the logic function. In ladder logic, contacts can be shown in series or parallel to achieve the "AND" or "OR" functions, respectively. A normally closed contact represents a "NOT" function.[45]

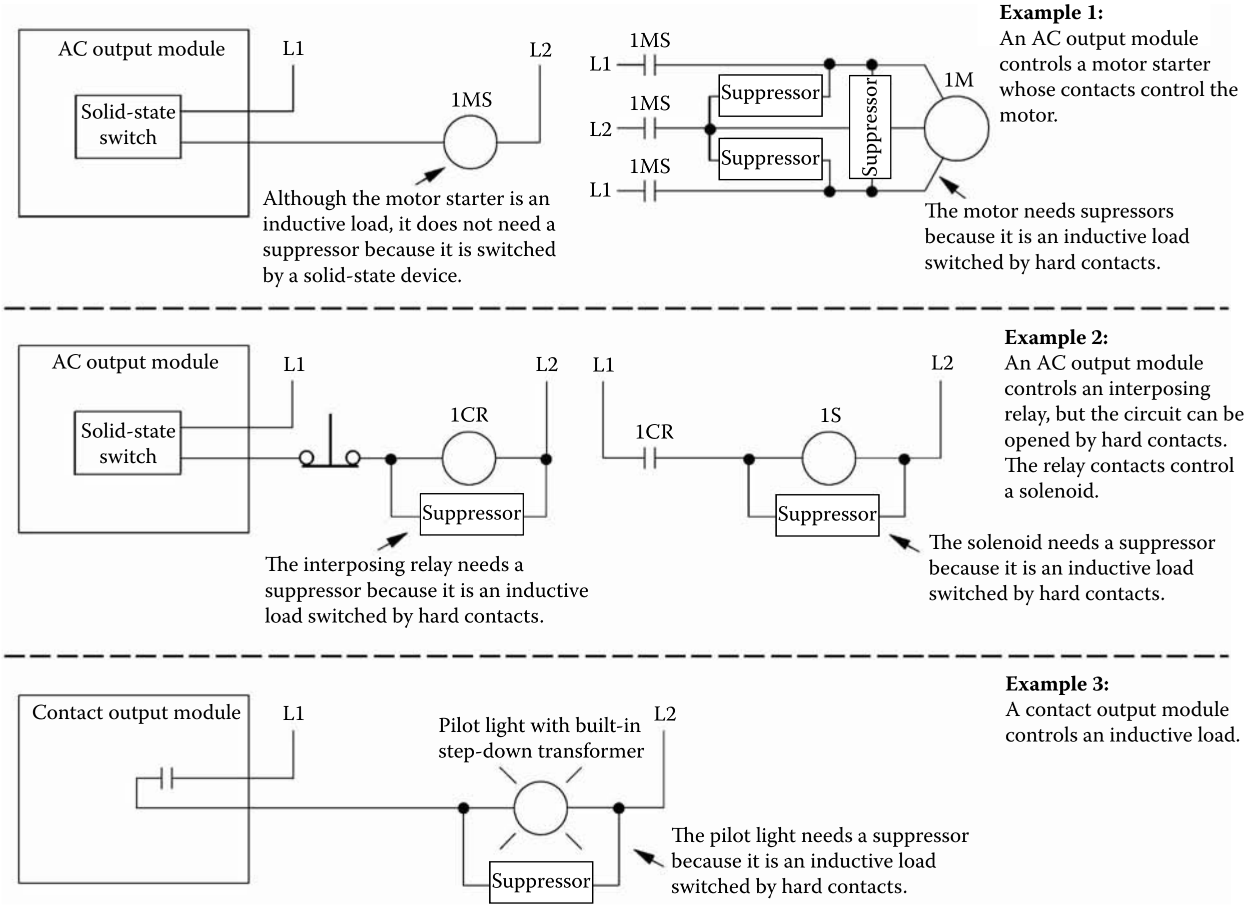


FIG. 5.4aa

Examples of where to use suppression. (Courtesy of Allen-Bradley.)

One popular programming technique involves defining the sequential logic in electrical schematic format, using actual tag numbers, and then translating this diagram into the appropriate programming language. Figure 5.4dd shows the translation of some examples of typical circuits to ladder diagrams, Boolean algebra, and mnemonics. Because this translation is relatively simple, maintenance and engineering personnel have accepted PLCs, although they have not accepted computers as readily. The unknown has been replaced with the familiar.

Timing and Counting Figure 5.4ee is a schematic representation of a timer and a counter. Although their formats differ, the principles are the same. The legs of the timer represent start/stop and reset. Timers can be on-delay or off-delay and can be cascaded (that is, linked together in series). Counters can be up or down and have a count leg (in which the number of switch closures is the count) and an up/down leg (in which the position of the switch determines up count or down count). A PLC with arithmetic capability can use a combination timer and counter as an integrator. Figure 5.4ff shows a turbine meter pulse counter turned into a low-cost integrator within the PLC. Obviously, the scan rate of the PLC (scans per second) must be twice the pulse rate of the turbine (pulses per second).

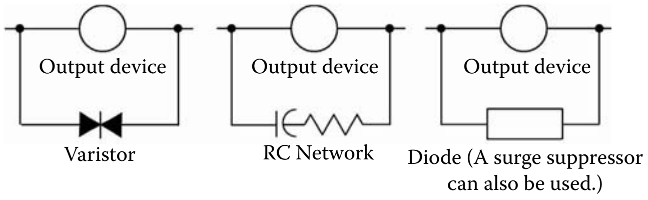


FIG. 5.4bb

Surge suppression for inductive AC load devices. (Courtesy of Allen-Bradley.)

Arithmetic Capabilities Figure 5.4ff shows an arithmetic program that permits the rapid addition of pulse counts from two counters hooked to two electric meters. The resulting sum is displayed through a panel meter. This logical addition is performed using integer mathematics (that is, no decimal calculations can be performed). Most PLCs use an approximation technique called “double-precision integer mathematics” to do calculations of greater complexity (such as PID). Some PLCs have true floating-point mathematics capability.

Floating-point mathematics is a powerful tool for process applications. For example, in the integrator example in

TABLE 5.4cc
Example of Memory Map for Milling Machine

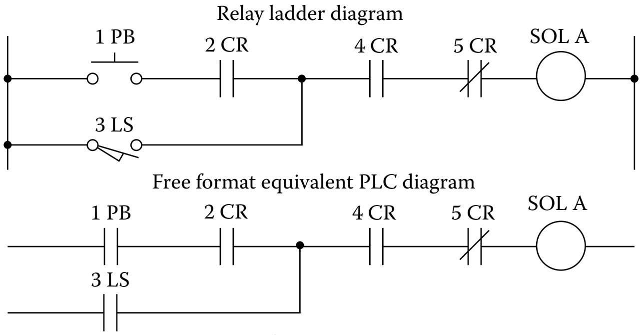
Coil	*Definition*			*Used in Rung(s)*					
CR0		R.H. Head	Retract	13	15	19	−20	−33	−46
CR1		L.H. Head	Retract	14	15	21	−33	−38	−47
CR2		Clamp	Part	4	5	−34			
CR3	Cycle	Part	Unclamp	2	2	25	25		
CR4	Shot	Pins	Out	5	5	6	27	−28	
CR5	Machine	Slide Adv	Milling	6	6	7	22	24	29
				30	32	35			
CR6	Machine	Slide	Retract	7	7	8	−22	39	
CR7	Milling	Spindles	Off	9	9	10	−24	−24	−32
CR8	Start	Boring	Spindles	10	10	11	12	20	23
CR9		R.H. Bore	Complete	11	11	13	13	46	
				31	33	38			
CR11	Unclamp	Swing	Clamps	15	15	16	−23	28	–29
				−30	−31	−31			
CR12	Unclamp	Pull	Clamps	−3	16	16	17	26	
CR13		Hold	Light On	3	3	4	4	−26	34
CR15	Milling	Complete	Shot Pins In	8	8	9	−27	−39	
CR17		Pump #1	Pressure	25	26	28	52		
CR18	Unclamp	Pull	Clamp	26	49				
CR19		Retract	Shot Pins	27	50				
CR20	Unclamp	Swing	Clamps	28	51				
CR29		L.H. Bore	Complete	12	12	14	14	47	
CR31	Pressure Off	Pump Time	Start	17	17	18	18		
CR32	Pressure	Off	Pump #1	18	−52				
CR101		Auto Lube	Cycle	36	36	40	40	41	41
				−43	48				
CR102			Lube Off	41	42	42	−43	−44	45
				−48					
CR103			Lube On	−41	−41	42			
CR104	60 Min.	Lube	Restart	43	44				
CR105		Start/	Disable	−40	45	45			
CR106	Contin.	Lube	Restart	40	44				

Source: Xcel

Figure 5.4gg, the division of pulses by elapsed time can be expressed as a decimal number rather than as a truncated integer. Feedforward calculations and PID can be performed in double-precision integer mathematics but are more memory-intense than in floating-point mathematics. Floating-point mathematics may require the use of a separate microprocessor within the CPU and usually involves two adjacent memory locations to store the mantissa and the abscissa in a form of scientific notation. The programmer automatically translates when the memory location is followed by a period (.), indicating a floating point number.

Programming Documentation The following technique can be used for PLC programming applications. There are many advantages to this approach.

1. Develop detailed I/O lists. Table 5.4p shows an I/O cross-reference relating tag numbers to I/O points. This list should be used extensively; starting without it will cause confusion and errors resulting from inevitable changes.
2. Develop a detailed descriptive operational sequence of events. Figure 5.4ii shows a sample sequence using a process batch application.



Boolean Statement
$((1PB \bullet 2CR) + 3LS) \bullet 4CR \bullet 5\overline{CR} = SOL\ A$

Code or mnemonic language

```
Load    1PB
And     2CR
Or      3LS
And     4CR
Nand    5CR
Store   SOL A
```

FIG. 5.4dd
Ladder translation. Here is a comparison of programming languages that are used with various PLCs. The most popular is still the relay ladder diagram because plant personnel are more familiar with it.[54]

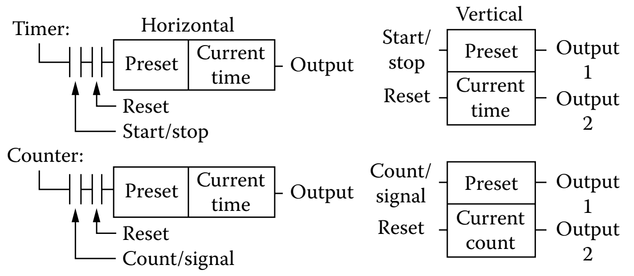


FIG. 5.4ee
Timer/counter schematic. A key part of any PLC programming is its capability to do timer/counter functions. The instructions are entered either horizontally or vertically, depending on the make. Horizontal programming, however, is more commonly used.[54]

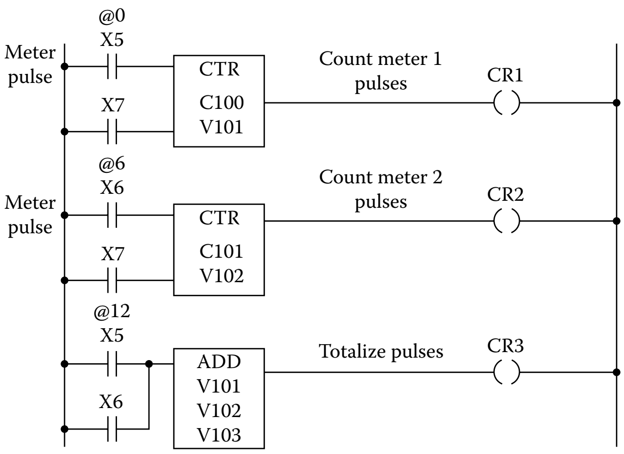


FIG. 5.4ff
Pulse counter totalizer.

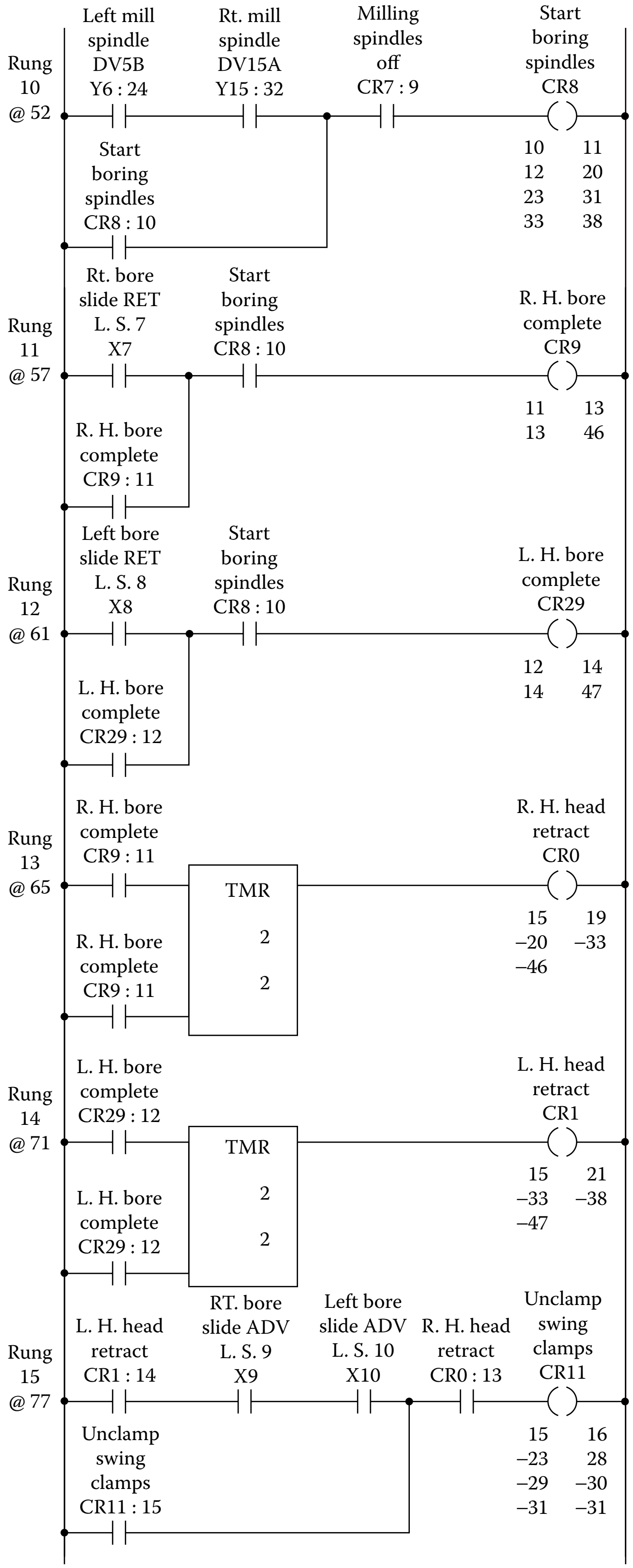
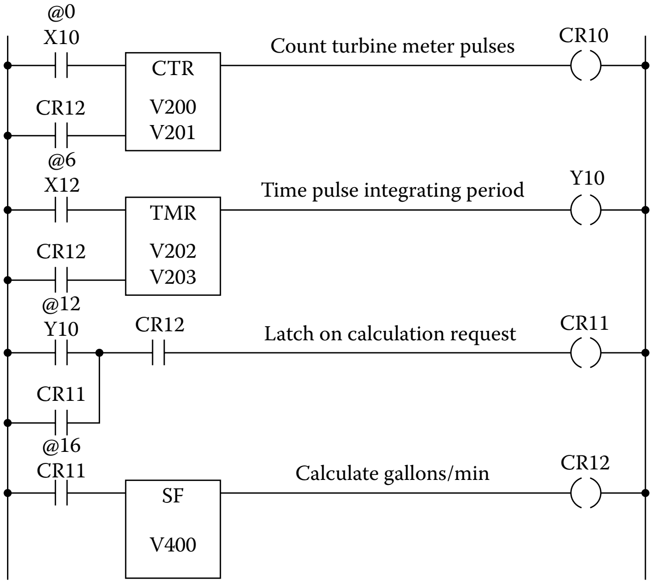


FIG. 5.4gg
Sample ladder diagram—annotated. (Courtesy of Xcel.)



```
                    Special function
                       user math
    Start address: V400                    Next address : V406
Error output ? . . . . . . . . . . . . . . . . . . . . . . . N
Error output designator . . . . . . . . . . . . . . . . .
Entry format:
(Math term: operator, math term: operator, . . . = result)
V201      V202      C200      =V205
No.       time      const.    gal/min
Pulses    period
```

FIG. 5.4hh
Floating-point integrator.

3. Develop electrical schematics or ladder diagrams for sequential control.
4. Develop piping and instrumentation drawings (or a logic diagram) for process control. Figure 5.4jj shows a diagram that will allow maintenance personnel to find important activities quickly. Note the I/O matrix index, showing what happens inside the PLC software, and the cross-referenced I/O memory locations and tag numbers.
5. Translate the drawings in steps 3 and 4 to the PLC language.
6. Enter the program code using a memory map (see Table 5.4cc).
7. Debug the program at the programmer's facility. Use a simulator to debug the program. Run through the operational sequence defined in step 2. If it has changed, be sure to ascertain how that change has affected other parts of the program. Rewrite the sequence description to reflect current operations, if necessary.
8. Save and document the program. Reproduce the program on transportable media, such as floppy disks. (Do this daily.) Document programming changes using printouts.
9. Enter and debug the program in the field. It is essential to note all changes made on the documentation. Some of the biggest problems with relay systems are undocumented field modifications.
10. Redocument and reproduce the final program.[14,21,46,47]

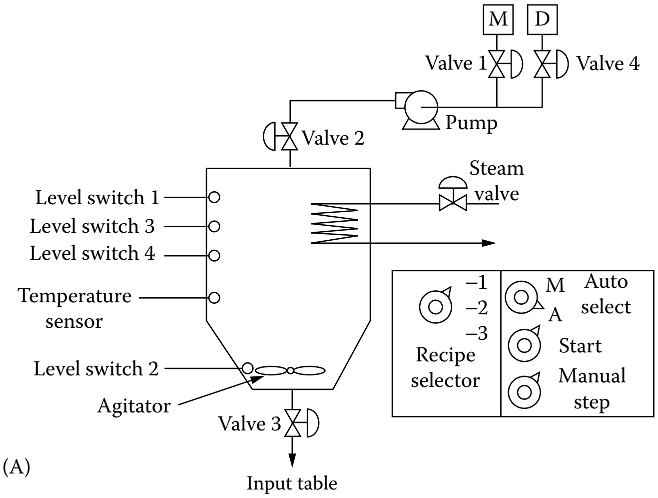
The final documentation package should include the following: I/O list and cross-reference, descriptive operational sequence, electrical schematics, process schematics, program listing (see Figure 5.4gg for annotated final document), memory maps (showing the memory areas that have been used and those that are available), and notes for future program changes or additions.

One important word about the documentation package: A major advantage of the PLC is its ability to be reprogrammed as plant requirements change. Without proper documentation, previous programming efforts will have to be reproduced in order to make the changes that are required. Poor documentation results in wasteful efforts at reconstruction. The PLC, like all engineering tools, requires good control systems engineering practices in order for its full potential to be realized. Good documentation is an essential element of any PLC project.[22,48]

Programming and Documentation Tools Both PLC and aftermarket parties offer programming and documentation tools for the system designer or user. Programming software is typically provided by the PLC manufacturer and is designed to program a specific machine or family of machines. Some third parties are offering universal programming software. These software packages vary greatly in price and capabilities but offer on- and off-line programming to many different types of PLCs, real-time status, and some very sophisticated annotation. Communications to different PLCs are usually supported by different software packages. Each vendor's product offers different types and amounts of ladder and contact comments. Again, many types of cross-references are available to be printed out. Often, other PLC design documentation problems may be solved, such as the generation of panel configuration drawings, point-to-point wiring diagrams, and I/O layout. One system even prints out the wire labels.

HMI Software (Program) Configuration The operator interface needs to be designed and implemented. Generally, the same process is followed whether the interface is a local panel or the SCADA server in the control room. Sketches of the displays should be made and reviewed with operations staff. These should emphasize operator-friendly graphics for the plant operators, who will spend the most time with them.[49] Most of the time, operators need to view an operations-oriented view of the process. Only when an abnormal event occurs will they go to the detail screens to find the cause.[50]

Once a consensus is reached on how the system should be operated, the screens can be configured on the interface equipment. "Linking" is the term used for defining how the screen will change based on the contents of PLC memory locations.



Output table

Step	Description	Input conditions to advance	Outputs Valve 1 300	Valve 2 301	Valve 3 302	Pump 303	Steam 304	Agitate 305	Valve 4 306
Home	Start	Auto select and start							
1	Fill	Tank full (Level switch 1)	X	X		X			X
2	Heat	Temperature setpoint or timer		X			X	X	
3	Cool	Timer							
4	Agitate	Timer						X	
5	Empty	Tank empty (Level switch 2)			X				
6	End	Automatic reset							
		Conditional interlocks:				Valve 2 open and Valve 1 or 4 open	Temperature not at setpoint or tank not empty	Tank not empty (Level switch 2 not set)	

X = Output energized

(B)

Input table

Input no.	*Description*	*Input no.*	*Description*
001	Auto select switch	010	Recipe 1 select
002	Start pushbutton	011	Recipe 2 select
003	Level switch 1	012	Recipe 3 select
004	Temperature sensor	013	Level switch 3
005	Level switch 2	014	Level switch 4
006	Valve 1 open limit switch	015	Manual step pushbutton
007	Valve 2 open limit switch		
008	Valve 3 open limit switch		
009	Valve 4 open limit switch		

(C)

FIG. 5.4ii

Batch sequence. (A) Graphic representation of a simple batch process reactor vessel, with inputs and outputs. (B) Outputs for each step in the process are described in the "Output Table." The advance conditions for each step are shown along with input interlocks (bottom), which affect certain outputs. (C) The "Input Table" correlates the connection terminal number with the word description of each input.[29]

Reports Configuration Some simple reports can be directly set up in the operator interface software. More extensive reports usually require separate reporting or database software. These reports should be set up early so that they can be checked before installation of the equipment.

Software/Hardware Integration

Once the program is entered, a simulation is recommended, and the program checkout process begins "on the bench." This process uses the functional specification to prove the software is acceptable. A large percentage of the program can be proved in this manner. Program debugging can be completed before field installation. Field corrections will then be minimized, and high-salaried electrical and installation personnel will not be standing around waiting. The savings that can be realized are quite significant.[51] In addition, it's important to keep in mind that there's no substitute for a bench-simulated program check. The software simulation proves the program and allows acceptance by the customer. The program should be reproduced and documented after it has been checked.

Once the panels are built, and all configuration and programming are complete, as much of the system as possible should be set up in a staging area for a complete checkout. Final control elements (valves, motor starters) function can be checked for proper outputs. Analog sensors (pressure, level, flow, and so on) are usually simulated with a 4–20 mA calibrator. Using the PLC and programming aids, the panel wiring can be "rung out" (that is, checked point by point) through the PLC. Each I/O point should be activated separately to the terminal block or from the panel controls (buttons, lights, switches). In this manner, the electrical integrity of the panel from the terminal blocks inward is ensured. If any continuity problems exist thereafter, they will be located in the field wiring. All remote I/O should be connected. The data highway or Ethernet networks should be temporarily connected. All displays should also be checked to make sure that the I/O points trigger the correct part of the screen or activate the correct alarm.

Some organizations prefer to perform a simulated operation checkout at this time. This is a highly useful approach and can be implemented if the simulators and the I/O point arrangement are organized to simulate the process outside the panel. Some I/O points may need to be jumpered for simulation. For instance, if run contacts are required to close a few seconds after the motor is called to run, they can often be temporarily connected to the output for the test. It is generally felt that each hour spent troubleshooting at this point will save two or three hours during system start-up. Occasionally, operations personnel are brought in

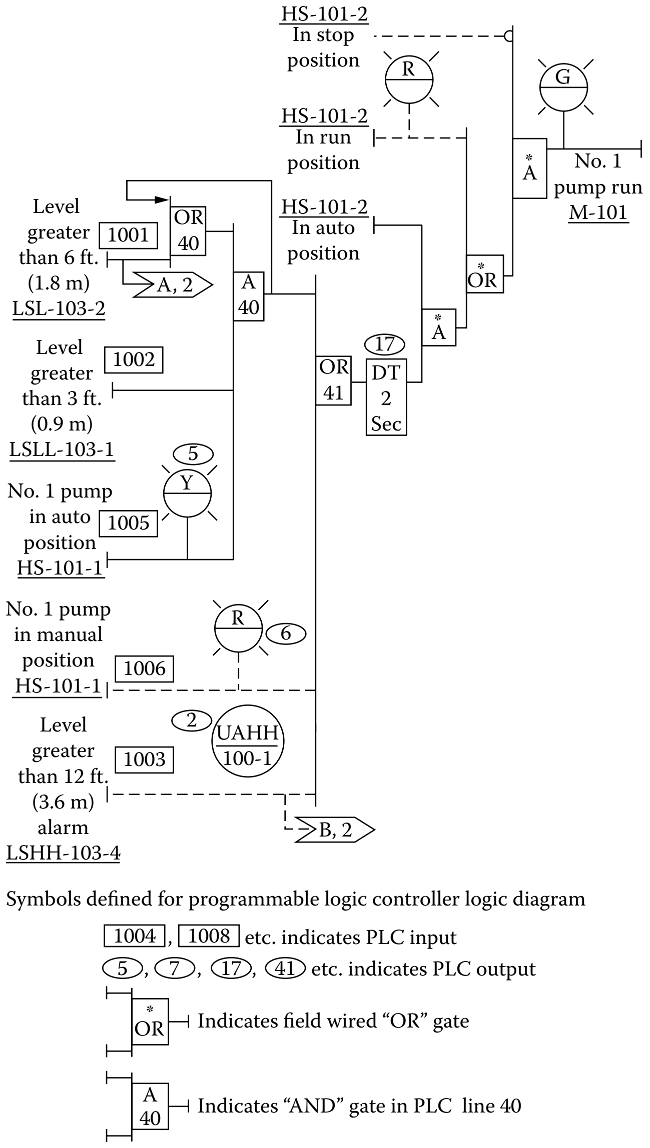


FIG. 5.4jj
Process and logic diagrams. This logic diagram is for a sump pump.[40]

to demonstrate this process. This is an effective way for the operations personnel to become familiar with the new system, especially in new automation projects in which there may be some "fear of the unknown" to overcome. Occasionally, operations personnel find operational flaws or make suggestions for improvements that can be easily implemented in the panel shop but would be difficult or impossible in the field.

Test software intelligently by trying to break it. Testing all aspects of the project during start-up will save time and money in the long run. Also it is recommended to wait until after major milestones in the project to perform software and hardware upgrades. This makes upgrade troubleshooting easier.[49] At the conclusion of this exercise the panel is accepted by the customer and is shipped to the job site.

System Checkout and Start-Up

After electrical interconnections are made and point-to-point wiring is completed (mechanical completion), the system is ready for start-up. The ability of the PLC to operate step by step through the start-up becomes very useful at this stage.

Experienced PLC personnel may provide temporary switches in the back of the panel in order to facilitate the start-up procedures. These switches can be key-locked, software-locked, or disconnected for normal operation. They are also very useful as future maintenance and troubleshooting tools to diagnose future problems.

Unanticipated circumstances are always a factor during start-up. Wiring errors, program errors, and mismatches between PLC and HMI databases are common problems at this stage. For this reason it is not uncommon to have electrical and programming personnel available at this time for implementation of any changes that might be necessary. Quite often, program changes can be accomplished over long-distance telephone lines using modems. These changes are not easily implemented without adequate documentation. After successful start-up, the plant is signed off.

After Start-Up

Once again, it is imperative for future successful plant operation that complete, current documentation be available. This documentation should include the items discussed previously. Especially useful for future changes or additions are the start-up notes and notes pertaining to future modifications.

Training of operations, maintenance, and engineering personnel should be timely and "hands-on." It is useful to videotape these training sessions for future reference (e.g., for training of new personnel). Suggested programs for PLC training can be obtained from the PLC vendor. This is an important function provided locally at the job site, at a nearby metropolitan area, or at the PLC vendor factory.

Troubleshooting When troubleshooting a PLC system it is good to remember that almost 80% of the time the problem is either outside the PLC entirely or in a broken I/O module. Program errors and wiring errors and mismatch of PLC/HMI databases can also cause these problems, but these are usually resolved during start-up. Some simple troubleshooting tips can be followed to decide if a problem is in PLC or elsewhere.

Discrete Inputs: Because DIs are usually high impedance, check the input voltages with a high impedance multimeter. Using a low impedance solenoid voltage checker on digital inputs can sometimes cause confusing results. For a discrete input, an indicator on the module shows the state of the input. This should match what is observed on the HMI screen or PLC programmer monitor. The commons for the DI card may be grouped or individual. Check the voltage between common for that point on the I/O card terminal block (or swing arm) and the point in question. If possible actuate the field device and verify that the PLC logic status and voltage change together. If the module shows the correct I/O status, but the PLC monitor or HMI does not, the I/O module or its configuration is suspect and may need replacing.

Discrete Outputs: Because triacs-type outputs "leak" a small amount of current, they may show a voltage on a multimeter if they are not connected to a load. For discrete outputs, an indicator on the module shows the state of the output. This should match what is observed on the HMI screen or PLC programmer monitor. The power to the DO card may be grouped or individual. Check the voltage for that point on the I/O card terminal block (or swing arm). Then check the voltage for the point in question. If possible cause the output to actuate and verify that the PLC logic status and output voltage change together. If the module shows the correct I/O status, but the measured output voltage does not, the I/O module or its configuration is suspect and may need replacing.

Analog 4–20 mA Inputs: For analog I/O, the input values are not displayed on the module. Temporarily remove field wiring and place a 4–20 mA calibrator directly on the point and select either passive or sourced as appropriate for the input configuration. Verify that 4 mA gives the minimum value, that 20 mA gives the maximum value, and that 12 mA produces the value exactly between the min and the max values at the HMI or the PLC. If this works, then the problem is outside the PLC hardware.

Analog 4–20 mA Outputs: For analog I/O, the output values are not displayed on the module. Some modules require an external loop power supply. Verify its operation. Put a multimeter in current mode in series with the point in question. Have an operator set the output to 0%, 100%, and 50%. This should produce readings of 4 mA, 20 mA, 12 mA, respectively. If this works, then the problem is outside of the PLC hardware.

Erratic Problems: To troubleshoot erratic PLC operations that affect more than a single I/O module, check PLC grounding, power to PLC power supply, DC power output of the power supply, and batteries. Electromagnetic interference (EMI) or radio frequency interference (RFI) generated by large motors starting, arc welding, lightning strikes, handheld radios, or transmitters (a common issue) can produce erratic operation. Improvements in power conditioning, grounding, and shielding can usually resolve these issues. If the program has been affected, it may need to be reloaded.[52]

CONCLUSION

PLCs are durable, delivering real-time control in a rugged and dependable package with no moving parts.[7] They can be installed on either the factory floor or outdoors, withstanding temperature swings up to 60°C PLCs have the critical ability to process sequential logic without experiencing faults in the operating system. Moreover, users continue to use PLCs because they know how to support them and understand the simple ladder logic language they use.[10]

The sheer number of PLC applications is enormous. According to a recent *Control Engineering* magazine poll, "The major applications for PLCs include machine control (87%), process control (58%), motion control (40%), batch control (26%), diagnostic (18%), and other (3%)."[53] The results don't add up to 100% because a single control system generally has multiple applications. Many sources documenting various PLC applications are listed in the Bibliography of this section.

References

1. Dummermuth, E., "A Bit Twiddler's History of the PLC," *A-B Journal,* November 2002.
2. Strothman, J., "More than a Century of Measuring and Controlling Industrial Processes," *InTech*, June 1995.
3. "The History of the PLC as Told to Howard Hendricks by Dick Morley," R. Morley Inc., 2002, http://www.barn.org/FILES/historyofplc.html.
4. Morley, R., "PLCs According to Morley," *Industry.Net Report,* September 27, 1995.
5. Strothman, J., "Leaders of the Pack," *InTech,* August 2003.
6. Hart, D., "The PLC-5, The Evolution of an Industry-Leading Solution," *A-B Journal,* November 2001.
7. "PLCs Provide Plenty of Possibilities," *Control Magazine,* November 2000.
8. Mintchell, G. A., "Distributed Control Goes Discrete," *Control Engineering,* October 2000.
9. Johnson, D., "Nano Devices Lead Assault on Traditional PLC Applications," *Control Engineering,* August 2002.
10. Merritt, R., "PLCs Persist and Prosper," *Control Magazine,* April 2001.
11. Cleaveland, P., "PLC Manufacturers Make Their Solutions a Lot More Open," *Instruments and Control Systems,* April 1999.
12. Korarek, L. A., "Sizing Criteria for the Evaluation of Programmable Controllers," Instruments and Control Systems Seminar Workshop (available from Allen-Bradley).
13. Hordeski, M. F., *Microprocessor Cookbook,* Blue Ridge Summit: Tab Books Inc., 1979.
14. Intel Corporation, "2816 E2PROM Data Sheet," Santa Clara, CA, 1981.
15. "The PC User Buyer's Guide," *PC User.*
16. "Fast Growing PC Market Encourages Wide Range of Product Offerings," *Control Engineering,* January 1983.
17. Waterbury, B., "Distributed Intelligence Goes to Work," *Control Magazine,* August 2000.
18. Eckard, M., "Tackling the Interconnect Dilemma," *Instruments and Control Systems,* May 1982.
19. ESD Vendor Workshop, "Control Networks vs. Data Acquisition (A Case for Dividing the Tasks)," 12th Annual Conference and Equipment Display, ESD, March 1983 (available from ISSC).
20. Hoag, D. S., "Programmable Controllers in Distributed Process Control Systems," International Conference on Power Transmissions, Houston, TX, June 1982 (sponsored by *Power Magazine*).
21. Jannotta, K. L., "Factory Communications: How to Talk to the Machines," International Conference on Power Transmissions, Houston, TX, June 1982 (sponsored by *Power Magazine*).
22. Whitehouse, R. A., "The Future of Programmable Controllers," ISA 1976 Annual Conference, Advances in Instrumentation, Vol. 33, Part I, Instrument Society of America, October 15–19, 1976.
23. Wallace, D., "How to Put SCADA on the Internet," *Control Engineering,* September 2003.

24. Laduminsky, A. J., "Peripherals Enhance PC Performance," *Control Engineering,* January 1983.
25. Lutz-Nagy, R., "The Dawn of Intelligent Motion," *Power Transmission Design,* February 1980.
26. "PC Users Guide," *Instruments and Control Systems.*
27. Londerville, S., "Programmable Controllers in Boiler Control Applications," Instrument Society of America, Northern California Section, *ISA Days,* June 2, 1982.
28. Bennett, L. E. and Lockert, C. F., "Evaluating Controls for Batch Processing," *InTech,* July 1976.
29. Dietz, R. O., "Programmable Controllers in Batch Processing," *InTech,* May 1975.
30. Larsen, G. R., "A Distributed Programmable Controller System for Batch Control," *InTech,* March 1983.
31. Melville, H. J., "PLC Applications in Batch Digester Control," ISA 1980 Annual Conference, Advances in Instrumentation, Vol. 35, Part II, Instrument Society of America, October 20–23, 1980.
32. Hawkinson, G. M., "Selecting a Programmable Controller for a Retrofit," 9th Annual Programmable Controllers Conference Proceedings, ESD, March 1980.
33. King, J. and Ptak, J., "Toothpaste Processing Using a PC Mini Microcomputer Network," 12th Annual Conference and Equipment Display Proceedings, ESD, March 1983.
34. Harrold, D., "Controllers — Heartbeat of Production," *Control Engineering,* January 2002.
35. Larson, K., "The Shifting Sands of DCS, PLC, & PC Technology," *Control,* May 1995.
36. Stafford, T., "Taking Back Control," *InTech,* July 2002.
37. Keskar, P., "Sailing to the Islands — Islands of Automation among Goals in Budget-Strapped Upgrades," *InTech,* February 2003.
38. Harrold, D., "Figure Out When Enough Is Enough," *Control Engineering,* June 2002.
39. Bryant, J. A., "Is Your Plant's Control System Safe?" *Power,* August 1979.
40. Klostermeyer, W. H. and Thurston, C. W., "PCs Prove Reliable in Chemicals and Plastics," *Control Engineering,* April 1976.
41. Hoffman, E. L., "Redundant Control Systems," 10th Annual Conference and Equipment Display Proceedings, ESD, March 1981.
42. Sykora, M. R., "The Design and Application of Redundant Programmable Controllers," *Control Engineering,* July 1982.
43. Penz, D. A., "Organizing PC Software Development," *Instruments and Control Systems,* February 1982 and March 1982.
44. Heider, R. L., "The Programmable Controller as a Process Link to a Distributed Computer Network," ISA 1980 Annual Conference, Advances in Instrumentation, Vol. 35, Part II, Instrument Society of America, October 20–23, 1980.
45. Rosandich, R. G., "What to Know About PLC Ladder Diagram Programming," *EC&M,* June 1996.
46. Conrad, W., "Document Your PC Design," 9th Annual Programmable Controllers Conference Proceedings, ESD, March 1980.
47. Stockweather, J. W., "Proper PC Documentation for the User," 11th Annual Conference and Equipment Display Proceedings, ESD, March 1982.
48. Langhans, J. D., "The Programmable Controller in a Continuous Process," *Control Engineering,* July 1972.
49. Dunlap, R. A., "8 Ways to Improve Control System Projects," *Control Engineering,* April 2003.
50. Harrold, D., "Find the Right Engineering Tools," *Control Engineering,* September 2001.
51. Kraemer, W. P., "Testing and Start-up of Programmable Controller Systems," Paper PC 179–14, *IEEE Transactions on Industry Applications,* Vol. 1A-16, No. 5, September/October 1980.
52. Rosandich, R. G., "Troubleshooting PLCs," *EC&M,* May 1996.
53. McBride, A. E., "Programmable Logic Controllers—Get More than What You See," *Control Engineering,* February 2001.
54. Programmable Controller Course, *Instruments and Control Systems,* Radnor, PA: Chilton Co., 1981.

Bibliography

Albert, C. L. and Coggan, D. A., *Fundamentals of Industrial Control,* Research Triangle Park, NC: ISA, 1992.

Amy, L. T., *Automation Systems for Control and Data Acquisition,* Research Triangle Park, NC: ISA, 1992.

Asfahl, C. R., *Robots and Manufacturing Automation,* 2nd edition, New York: John Wiley, 1992.

Baille, N. B., "The Application of Programmable Controllers to Compressor Systems for Offshore Applications," *Measurement and Control: Journal of the Institute of Measurement and Control,* England, Vol. 14, No. 9, September 1981.

Baker, A. T. and Rahoi, R. J., "Programmable Controller Applications for Power Plant Waste Water Treatment Facility," ISA Power Instrumentation Symposium, Chicago, May 14–16, 1980.

Barrett, L. E. and Lemmon, R. H., "Use of Programmable Controllers to Control Series Slurry Pumping Systems," 10th Annual Programmable Control Conference and Equipment Display Proceedings, ESD, March 10–12, 1981.

Bartlett, P. G., Jr., "Installing a PC," *Instruments and Control Systems,* August 1980.

Batten, G. L., *Programmable Controllers: Hardware, Software and Applications,* 2nd edition, New York: McGraw-Hill Companies, 1994.

Bauman, D. E. and Mellott, M. T., "Automatic Small Refinery Blending Operations," *Hydrocarbon Processing,* November 1981.

Bedworth, D. D., Henderson, M. R., and Wolfp, P. M., *Computer Integrated Design and Manufacturing,* New York: McGraw-Hill, 1991.

Bergwall Productions Inc., *Introduction to Programmable Logic Controller: E19*, Albany, NY: Delmar Publishers, 1992.

Bertrand, R. M., *Programmable Controller Circuits,* 1st edition, Clifton Park, NY: Delmar Learning, 1995.

Boehm, B. W., *Software Risk Management,* Piscataway, NJ: IEEE (Computer Science), 1990.

Bolton, W., *Programmable Logic Controllers,* 3rd edition, Burlington, MA: Newnes, 2003.

Booty, D. J., "So You're Buying a Programmable Controller," *InTech,* September 1980.

Boyer, S. A., *SCADA: Supervisory Control and Data Acquisition,* 2nd edition, Research Triangle Park, NC: ISA, 1993.

Bryan, E. A. et al., *Programmable Controllers: Workbook and Study Guide,* 2nd edition, Atlanta: Industrial Text & Video, 1997.

Bryan, L. A. and Bryan, E. A., *Programmable Controllers—Theory and Implementation,* Atlanta: Industrial Text Company, 1997.

Burr, J., "Smart Programmable Controller: Key to Pipeline Pump Station Control" (available from Gould Modicon).

Campbell, A., *The Allen-Bradley SLC 5/01-2 Programmable Controller,* New York: McGraw-Hill/Glencoe, 1998.

Carlisle, B. H., "More Power for Programmable Controllers," *Machine Design,* March 6, 1980.

Cherba, D. M., "Redundancy in Programmable Control Systems," *Instruments and Control Systems,* April 1983.

Clare, W. and Wilson, G., "Programmable Controllers Application to Water Treating Plants: What, Why, When, Where, How," 7th Annual Joint Instrumentation Conference and Workshop ISA and AWWA, Santa Ana College, CA, August 20, 1980.

Cleaveland, P., "Programmable Controllers: A Technology Update," *Instruments and Control Systems,* May 1983.

Coelho, W. A., "Putting Programmable Controllers to Work on Material Handling," *Instrumentation Technology,* March 1974.

Collins, D. and Lane, E., *Programmable Controllers,* New York: McGraw-Hill Book Co. Ltd., 1995.

Cooke, E. F., Halajko, J. S. et al., "Supervisory Programmable and Setpoint Control for Automating a Batch Plant," *InTech,* July 1980.

Costanzo, M., *Programmable Logic Controllers,* London: Hodder & Stoughton, 1997.

Crispin, A. J., *Programmable Logic Controllers and Their Engineering Applications,* 2nd Edition, New York: McGraw-Hill, 1996.

Daiker, J., "Operator Interface: Link-up of a Color-Graphic CRT and PLC," 9th Annual Conference and Equipment Display Proceedings, ESD, March 1980.

Dechont, G. R. and Ware, W. E., "Performing PID with a Programmable Controller on an Extrusion Line," 11th Annual Conference and Equipment Display Proceedings, ESD, March 1982.

Deltano, D., "Programming Your PC," *Instruments and Control Systems,* July 1980.

Dropka, E., *Toshiba Medium PLC Primer,* book and disk edition, Oxford: Newnes, 1995.

Dunning, G., *Introduction to Programmable Logic Controllers,* 2nd edition, Clifton Park, NY: Delmar Learning, 2001.

Ellis, R., "Color Graphics CRTs Provide 'Window' into Factory Operations," *Instruments and Control Systems,* February 1983.

Farrar, L., "Discrete Control Products for Reliable Automation," *Instruments and Control Systems,* April 1982.

Farrar, L., "Programmable Controllers in the Factory," *Instruments and Control Systems,* July 1983.

Filer, R. F. and Leinonen, G., *Programmable Controllers and Designing Sequential Logic (Saunders College Publishing Series in Electronics Technology)*, Boston: International Thomson Publishing, 1997.

Filer, R. F. and Leinonen, G., *Programmable Controllers Using Allen-Bradley SLC500 and Control-Logix,* 1st edition, New York: Prentice Hall, 2001.

Fischer, G. I., "Use of PCs for Energy Management," 9th Annual Programmable Controllers Conference Proceedings, ESD, March 1980.

Franck, D. T., "Programmable Controllers Today and Tomorrow," *Electrical Consultant,* 1974.

Frost, H. C., "New Processes Improve Food Quality, Yields Help Energy Efficiency, Labor Utilization," *Food Product Development,* April 1980.

Geller, D. A., *Programmable Controllers Using the Allen Bradley SLC-500 Family,* New York: Prentice Hall, 1999.

Gilbert, R. A. and Llewellyn, J. A., *Programmable Controllers: Practices and Concepts,* Industrial Training Co., 1985.

Girard, B., *Programmable Logic Controllers: Architecture and Application,* New York: Wiley & Sons, 1990.

Hackworth, F.D., *Programmable Logic Controllers: Programming Methods and Applications,* book and CD edition, New York: Prentice Hall, 2003.

Harold, R. A., "Designing Automated Systems for High Reliability and Maintainability," 12th Annual Conference and Equipment Display Proceedings, ESD, March 1983.

Hasenfuss, F. C., "Programmable Controllers Control Natural Gas Underground Storage," 10th Annual Programmable Controller Conference and Equipment Display Proceedings, ESD, March 10–12, 1981.

Henry, D., "PC Input/Output Consideration," *Instruments and Control Systems,* June 1980.

Herb, S.M., *Understanding Distributed Processor Systems for Control,* Research Triangle Park, NC: ISA, 1999.

Herndor, W., "Advantages of Digital Control in Small Process Loop Applications," ISA 1980 Annual Conference, Advances in Instrumentation, Vol. 35, Part I, October 20–23, 1980.

Hickey, J., "Programmable Controller Update," *Instruments and Control Systems,* July 1979.

Hill, C., "Microprocessors Simplify an Adaptive Gain Problem," ISA 1980 Annual Conference, Advances in Instrumentation, Vol. 35, Part I, October 20–23, 1980.

Hollopeter, W. C., "Man/Machine Interfaces: Yesterday, Today, and Tomorrow," ISA Regional Show 1979, Los Angeles, CA, Andover, MA: Gould Modicon.

Huber, R. F., "Programmable Controls: Where the Action Is," *Production,* September 1971.

Hughes, T. A., *Programmable Controllers,* 3rd edition, Research Triangle Park, NC: ISA, 2000.

Hyde, J. F., "Sophisticated PCs: The Right Direction but Don't Stop Now," 11th Annual Conference and Equipment Display Proceedings, ESD, March 1982.

Institution of Electrical Engineers, *Semsplc Guidelines: Safety-Related Application Software for Programmable Logic Controllers,* Stevenage, UK: IEE 1999.

Jeffcoat, W. and Clements-Jewery, K., *The PLC Workbook: Programmable Logic Controllers Made Easy,* New York: Prentice Hall, 1996.

Kalpakjian, S., *Manufacturing Engineering and Technology,* 2d edition, Reading, MA: Addison-Wesley, 1992.

Kirchof, G., *FX Cascading Logic: A Machine Control Methodology for Programmable Logic Controllers,* Research Triangle Park, NC: ISA, 2002.

Kissell, T.E., *Industrial Electronics: Applications for Programmable Controllers, Instrumentation and Process Control, and Electrical Machines and Motor Controls,* 3rd edition, New York: Prentice Hall, 2002.

Larson, K., "The Great DCS vs. PLC Debate," *Control,* January 1993.

Mandado, E. et al., *Programmable Logic Devices and Logic Controllers,* 1st edition, New York: Prentice Hall, 1996.

Martin, P.G., *Bottom-Line Automation,* Research Triangle Park, NC: ISA, 2001.

Michel, G. et al., *Programmable Logic Controllers: Architecture and Application,* New York: John Wiley & Sons, 1990.

Morriss, S.B., *Programmable Logic Controllers,* 1st edition, New York: Prentice Hall, 1999.

Nelson, J. E. and Menefee, W. F., "Application of Programmable Controllers in the Loop Supervisory Control and Monitoring System" (detailed paper available from Gould Modicon).

NFPA 79, Electrical Standard for Industrial Machinery, Quincy, MA: National Fire Protection Association, Inc., 1997.

Obert, P., "The 12 Most Often Asked Questions About PLCs," *Instruments and Control Systems,* August 1976.

Otter, J. D., *Programmable Logic Controllers: Operation, Interfacing and Programming,* New York: Prentice Hall, 1988.

Owens, D. R., "A Case for Custom Software," *Instruments and Control Systems,* November 1981.

Palko, E. (Ed.), "The Solid State Industrial Programmable Controller," *Plant Engineering,* April 1973.

Parkin, T. O., "How Cost-Effective Are PCs?" *Instruments and Control Systems,* April 1980.

Parr, E.A., *Programmable Controllers: An Engineer's Guide,* 3rd edition, Boston: Newnes, 2003.

Persson, N. C., "Programmable Controllers," *Design News,* April 1978.

Petruzella, F. D., *Programmable Logic Controllers*, 2nd edition, New York: McGraw-Hill Science/Engineering/Math, 1997.

Petruzella, F. D., *Programmable Logic Controllers, Activities Manual,* 2nd edition, New York: McGraw-Hill Science/Engineering/Math, 1996.

PLC-5 Programming Software, Allen-Bradley, 1991.

Poisson, N. A. and Anderson, R., "Batch Process Control with Programmable Controllers," (available from Gould Modicon).

Programmable Controller Handbook, Lebanon, OH: Cincinnati Milacron Inc., Electronics Systems Division, 1972.

"Programmable Controllers," *Measurements and Control,* September 1993.

Quatse, J. T., "Programmable Controllers of the Future," *Control Engineering,* Vol. 33, No. 1, January 1986.

Rabiee, M., *Programmable Logic Controllers: Hardware and Programming,* Teacher edition, South Holland, IL: Goodheart-Willcox Co., 2002.

Ridley, J., *Introduction to Programmable Logic Controllers: The Mitsubishi FX,* Oxford: Butterworth-Heinemann, 2001.

Rohner, P., *PLC: Automation With Programmable Logic Controllers,* Seattle, WA: University of Washington Press, 1996.

Rosandich, R. G., *Fundamentals of Programmable Logic Controllers,* Overland Park, KS: EC&M Books, 1997.

Saitta, G. N., "PC Manages Vapor Recovery System at Alaska Pipeline Terminal," *InTech,* May 1978.

Sampson, W. H., "Controlling Electric Demand with a Programmable Controller," 11th Annual Conference and Equipment Display Proceedings, ESD, March 1982.

Savelyev, M. K., "How to Use PCs for Energy Management Systems," *Control Engineering,* February 1979.

Simpson, C.D., *Programmable Logic Controllers,* 1st edition, New York: Prentice Hall, 1998.

Standard ICS3-1978, "Industrial Systems," National Electrical Manufacturers Association, 1978.

Stenerson, J., *Fundamentals of Programmable Logic Controllers, Sensors, and Communications,* 2nd edition, New York: Prentice Hall, 1998.

Stenerson, J., *Programming PLCs Using Rockwell Automation Controllers,* book and CD-ROM edition, New York: Prentice Hall, 2003.

Stockweather, J. L., "Suppliers to Users: What Is Their Role?" *PC User,* March 1982.

SYMAX 650 Programming Manual, Square D, 1991.

Taebel, T. A., "An Introduction to Programmable Controllers," presented at the ESD PC Conference, May 1977.

Thumann, A., "For Greater Control Sequence Flexibility, Use the Programmable Controller," *Electrical Consultant,* 1974.

Treseler, M., *Designing State Machine Controllers Using Programmable Logic (Prentice Hall Series in Innovative Technology),* New York: Prentice Hall, 1991.

Urmie, M., "Programmable Controllers for Batch Processing," *Measurements and Control,* December 1982.

Waite, R., "PC Study Reveals User Wants and Needs," *Instruments and Control Systems,* February 1983.

Walker, F. H., "Use of Programmable Controllers for Industrial Power Distribution System Monitoring, Control and Protection," IEEE Proceedings Conference Record 79CH1423-3 1A, Paper Number PCI-79-25, 1979.

Ward, P. T. and Mellor, S. J., *Structured Development for Real-Time Systems,* Yourdon Press, 1991.

Webb, J. W. and Reis, R. A., *Programmable Logic Controllers: Principles and Applications,* 5th edition, New York: Prentice Hall, 2002.

Whitehouse, R. A., "Programmable Controllers: Cost and Maintenance," 1979 Plant Engineering and Maintenance Conference, Chicago, May 8, 1979.

Wiktorowicz, A. C. and Jeffreys, J., "Boiler Combustion Control with Programmable Controllers," International Conference on Power Transmission, Houston, TX, June 1982 (sponsored by *Power Magazine*).

Williams, R. I., "Fundamentals of Pipeline Instrumentation, Automation and Supervisory Control," ISA 1981 Annual Conference, Advances in Instrumentation, Vol. 36, Part I, March 23–26, 1981.

5.5 PLC Programming

V. A. BHAVSAR (2005)

Costs of Programming:	PLC programming cost depends on the complexity and the length of the project. The hourly rate of programming is between \$40 and \$80. Typically, the programming cost is 50 to 100% or more of the hardware cost depending on the size and complexity of the PLC system.
Partial List of Suppliers:	Most PLC manufacturers also do programming for their clients or they have systems integrators who can do the programming for them. Listed below are some of the major PLC manufacturers:

ABB (www.abb.com)
Allen-Bradley (www.ab.com)
GE Fanuc (www.gefanuc.com)
Klockner-Moeller (www.km-medialab.com)
Mitsubishi (www.mitsubishielectric.com)
Modicon (www.modicon.com)
Omron (www.omron.com)
Siemens (www.siemens.com)
Triconex (www.triconex.com)

INTRODUCTION

This section covers the basics of PLC programming. It provides the readers with enough information about PLC system hardware and programming techniques to enable them to write simple programs and understand complex programs. It starts with a basic understanding of PLC, followed by an introduction to major hardware components and operation of PLC system, and various programming languages available.

The main focus of this section is on ladder logic, which is the most widely used PLC language. The rest of the section gives details about basic ladder logic instructions, programming devices used to write ladder logic, ladder logic structure, how to develop ladder logic and program a PLC, programming considerations, and documentation. The section concludes with a typical ladder logic program example and recent developments in PLC programming.

What Is a PLC?

PLC is an acronym for programmable logic controller. It is also known by such other names as programmable controller or logic controller. Basically, a PLC is an electronic device that was invented to replace the electro-mechanical relay logic circuits used for machine automation in the automobile industry. It was first introduced in late 1960s by Bedford Associates as Modular Digital Controller (MODICON) to a major U.S. car manufacturer. Modern definition of PLCs: They are small computers, dedicated to automation tasks in an industrial environment.

There are three major kinds of PLCs:

Compact PLC: Monolithic construction, monoprocessor, fixed number of I/Os
Modular PLC: Modular construction, one or multiprocessor, expandable I/Os
Soft-PLC: Windows NT- or CE-based, direct use of CPU or coprocessors

Today, PLCs are used in many industries including machining, packaging, material handling, automated assembly, and countless others. Usually PLC manufacturers provide their customers with various applications of their products. Other useful sources to learn more about existing and new PLC applications are technical journals, magazines, and papers published on control and automation.

SYSTEM HARDWARE AND OPERATION[1]

A modular PLC system consists of the following major components:

- Rack
- Power supply module

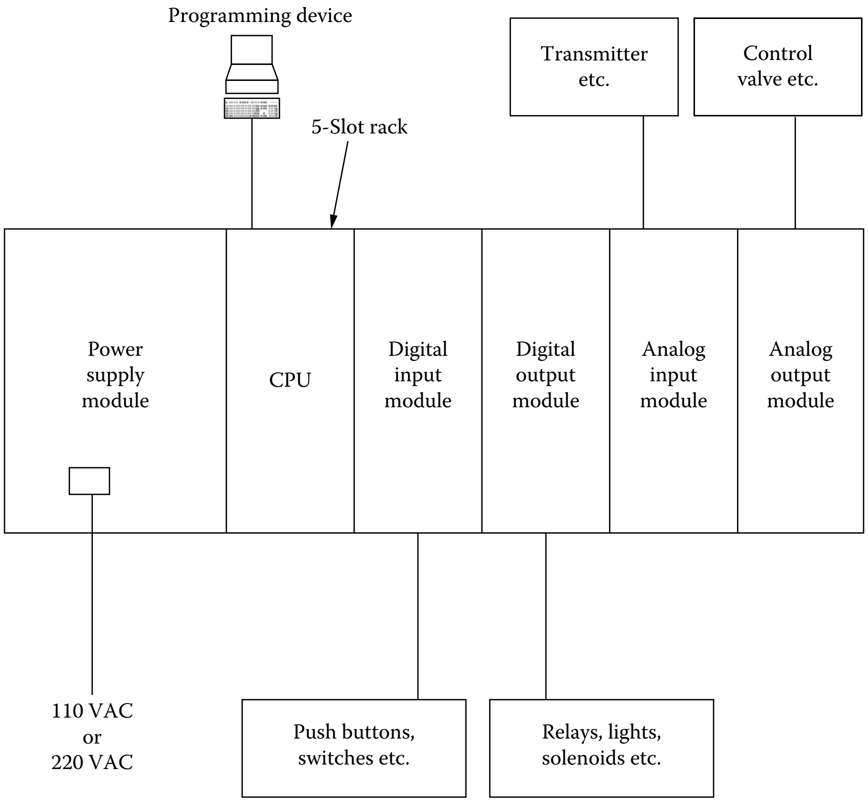


FIG. 5.5a
Modular PLC system.

- CPU (processor/memory)
- I/O modules
- Programming device

Figure 5.5a shows a typical modular PLC system.

A PLC is a sequential device; i.e., it performs one task after another. Figure 5.5b shows how a conventional PLC system operates. There are three major tasks that a PLC performs, in the following order:

- Task 1: Read inputs. PLC checks status of its inputs to see if they are on or off and updates its memory with their current value.
- Task 2: Program execution. PLC executes program instructions one by one sequentially and stores the results of program execution in the memory for use later in task 3.
- Task 3: Write outputs. PLC updates status of its outputs based on the results of program execution stored in task 2.

After PLC executes task 3 it goes back to execute task 1 again. The total time taken by a PLC to perform these three tasks is called the PLC scan time. Scan time depends on CPU clock speed, user program length, and number of I/Os. Typically scan time is in milliseconds. The smaller the scan time, the faster the updates of the I/O and the program execution.

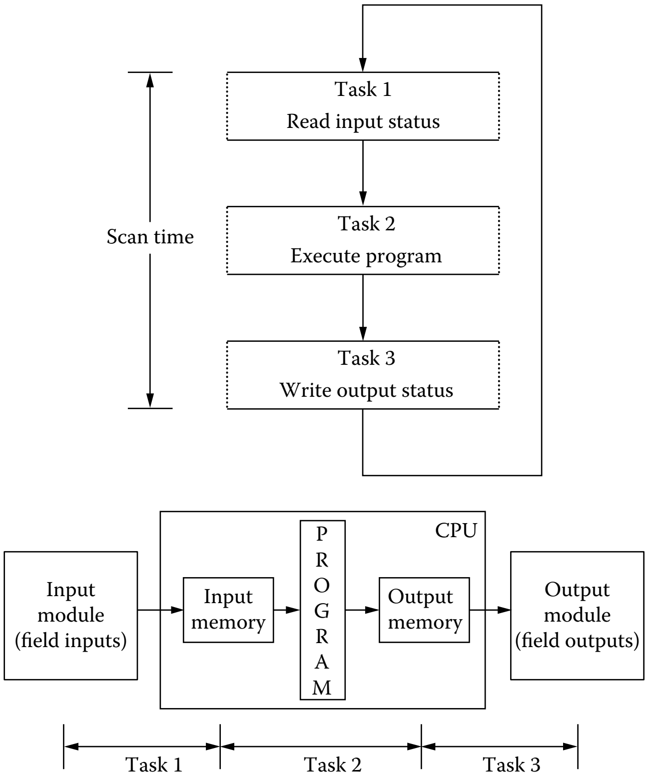


FIG. 5.5b
PLC operation.

As a general guide, PLC scan time should be less than half the time it takes the fastest changing input signal to change in the system.

The following are the functions of a PLC:

Measure: Measure process values, both discrete and continuous
Command: Execute logic and control programs
Regulation: Control field devices
Communication: Communicate with other PLCs, remote I/Os, MMIs, and other network peripherals

PROGRAMMING LANGUAGES

The two forms of PLC programming languages are text and graphic languages:

Text languages are the instruction list (IL) and the structured text (ST) types. The graphic languages are the sequential function charts (SFCs), function block diagrams (FBDs), and the ladder logic types.

Different PLCs support one or more of the above languages for programming. These languages have their advantages and limitations, and they complement one another to provide programmers with more programming power. A brief understanding of these languages is given below, followed by a detailed understanding of ladder logic.

Instruction List

Instruction list is a low-level language. It is mainly used for smaller applications or for optimizing parts of an application. IL is one of the languages for the description of the actions within the steps and conditions attached to the transitions of the SFC language (this is discussed in more detail later). Instructions always relate to the current result (or IL register value). The operator indicates the operation that must be performed on the current value of the IL register and the operand. The result of the operation is stored again in the IL register.

An IL program is a list of instructions. Each instruction must begin on a new line and must contain an operator and, if necessary, for the specific operation, one or more operands, separated with commas (,). A label followed by a colon (:) may precede the instruction. If a comment is attached to the instruction, it must be the last component of the line. Comments always begin with (* and end with *). Empty lines may be entered between instructions. Comments may be put on empty lines. Listed below are examples of instruction lines:

Label	*Operator*	*Operand*	*Comments*
Start:	LD	IN1	(* start pushbutton *)
	AND	MD1	(* mode is manual *)
	ST	Q2	(* start motor *)

LD-Load, ST-Stove

Structured Text

Structured text is a high-level structured language designed for automation processes. This language is mainly used to implement complex procedures that cannot be easily expressed with graphic languages. ST is one of the languages for the description of the actions within the steps and conditions attached to the transitions of the SFC language (this is discussed in more detail later). An ST program is a list of ST statements. Each statement ends with a semicolon (;) separator. Comments may be freely inserted into the text. A comment must begin with (* and end with *). The basic types of ST statements are

- Assignment statement (variable := expression;)
- Subprogram or function call
- Function block call
- Selection statements (IF, THEN, ELSE, CASE ...)
- Iteration statements (FOR, WHILE, REPEAT ...)
- Control statements (RETURN, EXIT ...)
- Special statements for links with other languages such as SFC

Sequential Function Charts

Sequential function chart is a language used to graphically describe sequential operations. The process is represented as a set of well-defined steps, linked by transitions. A sequence of steps is called a task, and many such tasks make up the whole process. Figure 5.5c shows basic components (graphic

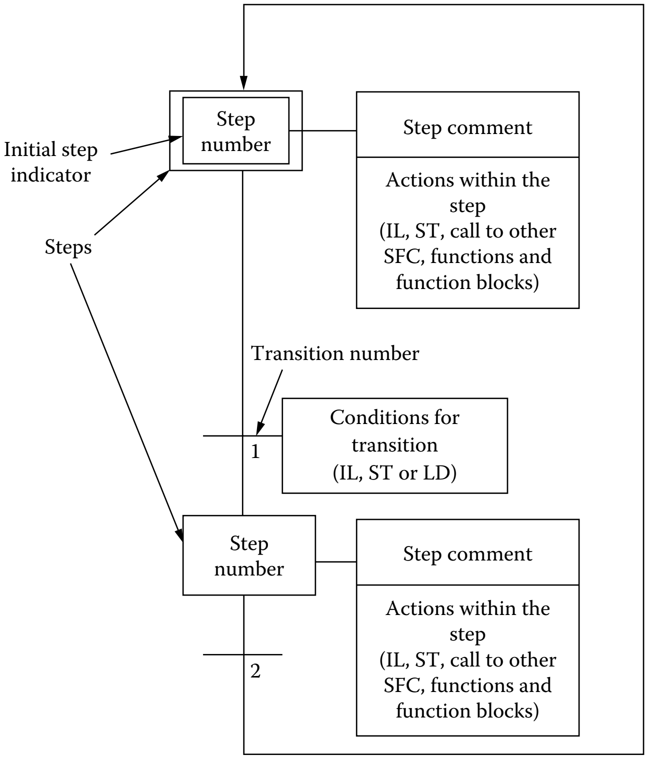


FIG. 5.5c
Basic components of SFC language.

symbols) of the SFC language: steps and initial steps, transitions, and links between steps and transitions.

An SFC program is activated by the system when the application starts. At run time, the step that is active is highlighted. When program execution starts, the initial step becomes the current step by default. When the conditions for transition become true, the sequence advances to the next step in the sequence. Actions within the step can be Boolean, ST or IL statements, calls to other SFCs, or function blocks or functions. A Boolean condition is attached to each transition using other languages, such as ST, IL, or ladder logic (LD).

Function Block Diagrams

The functional block diagram (FBD) is a language used to graphically build complex procedures by taking existing functions and wiring them together. Figure 5.5d shows basic components (graphic symbols) of the FBD language. An FBD diagram describes a function between input variables and output variables. A function is described as a set of elementary function blocks. Input and output variables are connected to blocks by connection lines. An output of a function block may also be connected to an input of another block.

Each function block has a fixed number of input connection points and a fixed number of output connection points. A function block is represented by a single rectangle. The inputs are connected on its left border. The outputs are connected on its right border. An elementary function block performs a single function between its inputs and its outputs. The name of the function to be performed by the block is written in its rectangle symbol. Each input or output of a block has a well-defined type. Input variables of an FBD program must be connected to input connection points of function blocks.

The type of each variable must be the same as the type expected for the associated input. An input for an FBD diagram can be a constant expression, any internal or input variable, or an output variable. Output variables of an FBD program must be connected to output connection points of function blocks. The type of each variable must be the same as the type expected for the associated block output. An output for an FBD diagram can be any internal or output variable, or the name of the program (for subprograms only). When an output is the name of the currently edited subprogram, it represents the assignment of the return value for the subprogram (returned to the calling program).

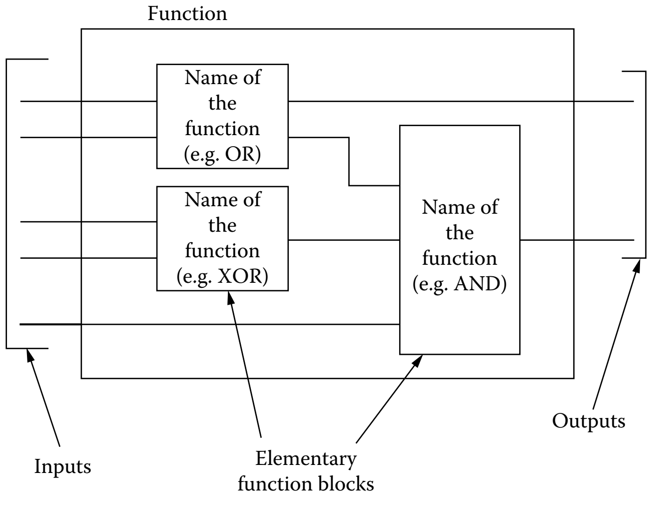


FIG. 5.5d
Basic components of FBD language.

LADDER LOGIC PROGRAMMING

Ladder logic is one of the most popular and widely used programming languages by electricians and programmers, since it emulates the old relay-based ladder logic structure. For this reason, the rest of this section discusses ladder logic programming in greater detail.

Ladder Logic Structure

As ladder logic is an extension of the relay logic, it makes sense to see how the ladder logic and relay logic structures are analogous to each other. Also, a basic understanding of Boolean logic will help in writing ladder logic, since many times a control scheme is first developed using Boolean logic and then converted into a ladder logic in PLC. Figure 5.5e shows three different representations of a simple logic for starting/stopping a motor.

In relay logic, the lines L1 and L2 represent the power applied to the relay circuit. Start pushbutton is normally open type and stop pushbutton is normally closed type. When the start pushbutton is pressed, the start pushbutton's contact closes, the current flows through start and stop pushbuttons' contacts to the relay coil, and it energizes the relay coil. The contact of the relay coil gets closed and provides an alternate current path to the relay coil, thus keeping it energized. When the stop pushbutton is pressed, the stop pushbutton's contact opens, the current path to the relay is opened, the relay gets de-energized, and the relay contact opens. The motor runs as long as the motor relay stays energized.

In Boolean logic, the start pushbutton and motor relay contact are represented as inputs to the OR logic block, output of the OR logic block and stop pushbutton are represented as inputs to the AND logic block, and motor relay coil is represented as output of the AND logic block. Normally closed stop pushbutton is shown by inverting the stop input (logical 1). When the start pushbutton is pressed (logical 1), output of the OR block becomes logical 1. With both inputs of the AND block being logical 1, output of the AND block becomes logical 1 and turns the motor relay on. Status of motor relay is fed back to the OR block to keep its output on. Motor relay turns off as soon as the stop button is pressed.

In PLC ladder logic, two vertical lines represent virtual power lines, and the actual electrical current is replaced by

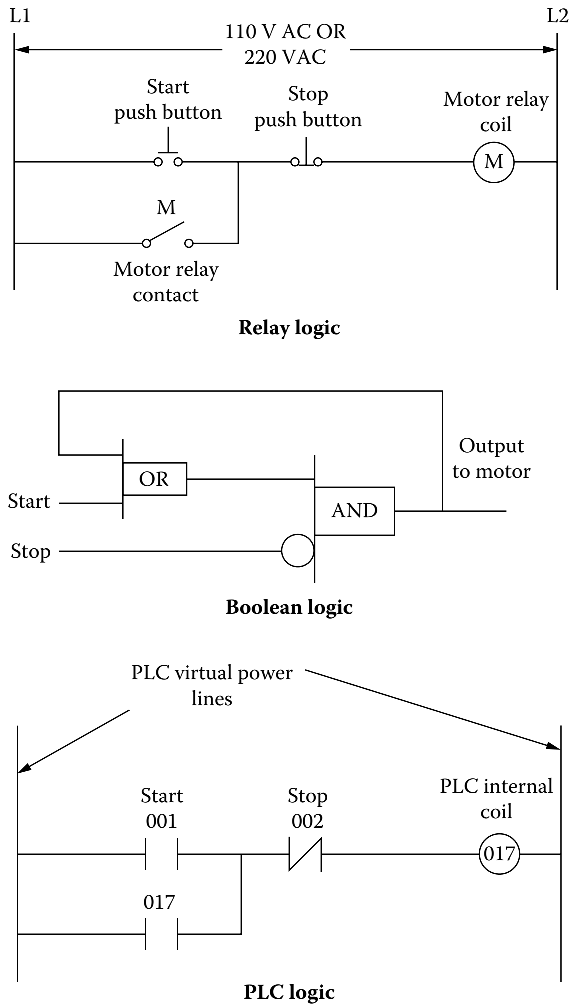


FIG. 5.5e
Three types of logic representation.

logical path. A horizontal line represents a logical rung. Inputs are shown near the left vertical line and outputs are shown near the right vertical line on a rung. When the inputs have logical 1 or TRUE state, logical path to the outputs completes, and outputs get logical 1 or TRUE state.

In this example, start and stop pushbutton contacts are shown as inputs that represent the physical pushbuttons connected to the PLC input module. Inputs have their PLC memory addresses where their current values are stored (here they are 001 and 002). As the pushbuttons are operated, the PLC memory is updated with their current status. An open contact represents logical 0 or FALSE, and a closed contact represents logical 1 or TRUE in PLC memory. In this example the relay coil is shown as an internal coil (PLC memory address 017) that represents a physical relay connected to the PLC output module. As the internal coil's status changes based on the pushbuttons status, PLC memory is updated with the internal coil's current status.

An energized internal coil represents logical 1 and a de-energized internal coil represents logical 0 in PLC memory. The physical relay coil gets energized when the PLC internal coil is energized (logical 1) and de-energized when the PLC internal coil is de-energized (logical 0). The internal coil has a large number of virtual contacts limited only by the PLC memory capacity.

Ladder Logic Programming Basic Instructions

Figure 5.5f and Figure 5.5g show graphical symbols of ladder logic basic instructions discussed below.

Contact This is an input instruction. It can be used to represent an external digital input or a contact of internal (soft) relay. There are two basic types of contacts:

1. Normally open contact (NO) is used to represent an input signal that is normally off and will become on when operated, e.g., a physically connected pushbutton with normally open contact. For an NO contact, logical 0 represents OFF or FALSE condition and logical 1 represents ON or TRUE condition.
2. Normally closed contact (NC) is used to represent an input signal that is normally on and will become off when operated, e.g. a normally closed contact of an internal (soft) relay coil. For an NC contact, logical 0 represents ON or TRUE condition and logical 1 represents OFF or FALSE condition.

Coil This is an output instruction. It can be used to represent an external digital output or an internal (soft) relay. There are two basic types of coils:

1. Normally de-energized coil is used to represent an output signal that is normally off or de-energized and will become on or energized when all inputs on a rung preceding this coil are TRUE.
2. Normally energized coil is used to represent an output signal that is normally on or energized and will become off or de-energized when all inputs on a rung preceding this coil are TRUE.

Timer This is a timing instruction. Timer is used to delay an event or to time an event. Generally, a timer instruction has enable input, preset time value, elapsed time value, and an output signal. Preset and elapsed times are stored in PLC memory registers. There are two basic types of timers. Figure 5.5h shows the timing diagram for these timers.

1. On-delay timer is used to delay turning on an output. When the enable condition becomes true, timer starts timing, its current time value starts to go up, and after its current time reaches the preset time value, output of the timer is turned on. Output stays on as long as the enable input signal is on. As soon as the enable signal becomes false or logical 0, the output goes off

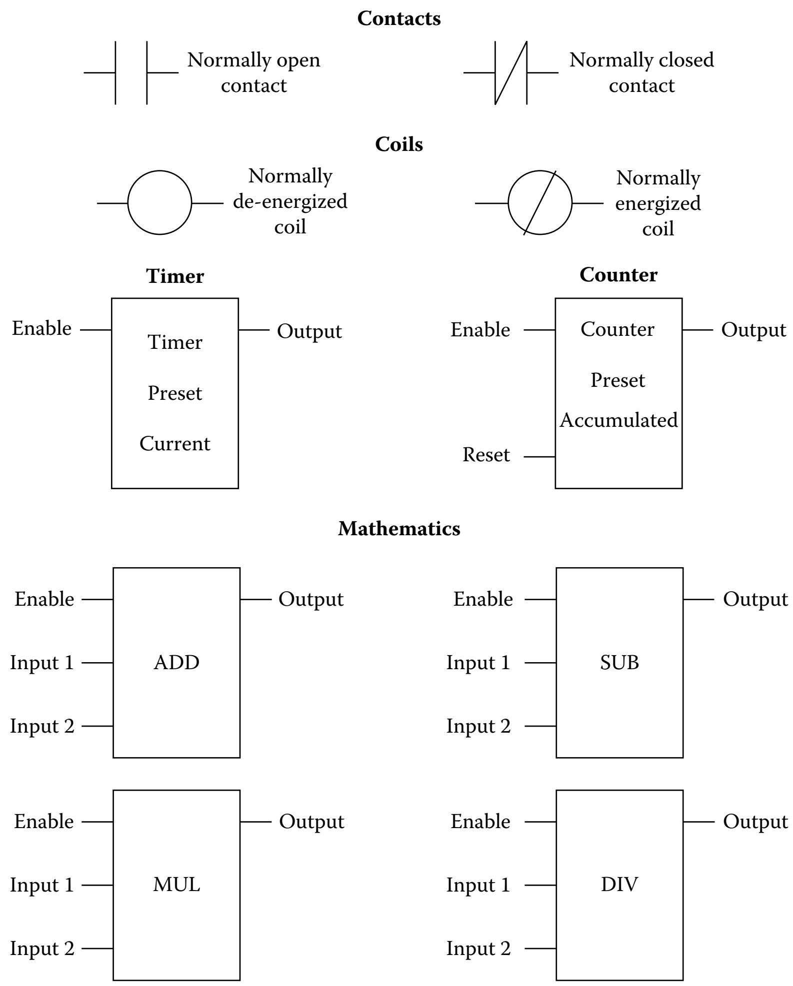


FIG. 5.5f
PLC ladder logic instruction symbols.

and the timer current value becomes 0. There are other variations of timers that have a reset input connection to reset the timer current value.

2. Off-delay timer is used to delay turning off an output. As soon as the enable condition becomes true, output of the timer turns on. When the enable condition becomes false or logical 0, timer starts timing and its current time value starts to go up. Timer output stays on until its current time reaches the preset time value. As soon as the timer current time reaches the preset time, its output goes off and the timer stops timing and its current value stops increasing. When enable input becomes true the next time, timer current value is reset to 0.

Counter This is a counting instruction. Counter is used to count the number of occurrences of an event. Generally, a counter instruction has enable input, preset count value, accumulated count value, reset input, and an output signal. Preset and accumulated counts are stored in PLC memory registers. There are two basic types of counters. Figure 5.5i shows a timing diagram for these counters.

1. Up-counter is used to count up. When the enable condition becomes true, i.e., transitions from logical 0 to logical 1, the counter accumulated value is incremented by 1. Every time this transition takes place, the counter accumulated value increases by 1. When the accumulated value reaches the preset value, the counter output turns on. Output stays on until the counter is reset.
2. Down-counter is used to count down. Counter accumulated is set at the preset value. When the enable condition becomes true, i.e., transitions from logical 0 to logical 1, the counter accumulated value is decremented by 1. Every time this transition takes place, the counter accumulated value decreases by 1. When the accumulated value reaches 0, counter output turns on. Output stays on until the counter is reset.

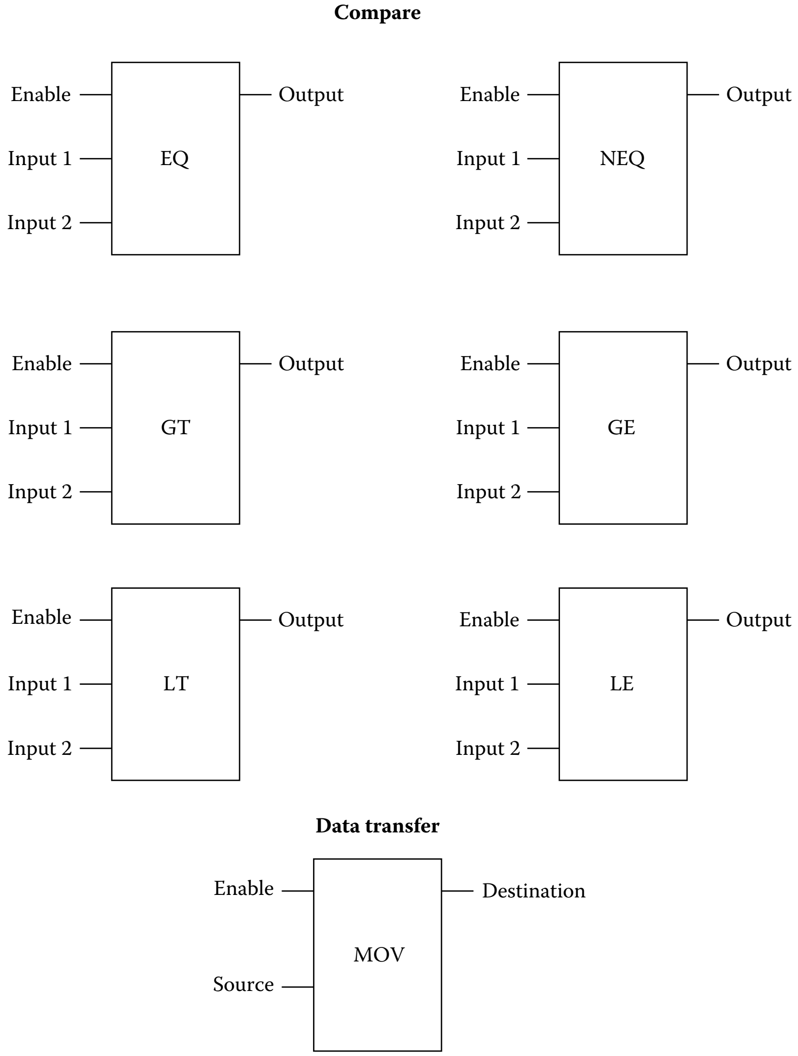


FIG. 5.5g
PLC ladder logic instruction symbols.

Mathematics Math instructions are used for data manipulations. Generally, a math instruction has enable input, two data inputs, and one output. Data inputs are register locations or constants and output is a register location where the result of the math operation is stored. There are four basic types of math instructions:

1. ADD is used to perform addition. When the enable condition becomes true, input 2 data are added to input 1 data and the result is stored in output.
2. SUB is used to perform subtraction. When the enable condition becomes true, input 2 data are subtracted from input 1 data and the result is stored in output.
3. MUL is used to perform multiplication. When the enable condition becomes true, input 1 data are multiplied with input 2 data and the result is stored in output.
4. DIV is used to perform division. When the enable condition becomes true, input 1 data are divided by input 2 data and the result is stored in output.

Although most advanced PLCs support real math, some PLCs only support integer math, i.e., math performed on integer numbers. Hence, when you perform division on integers, the result is an integer and loose remainder (fraction part of the result). To deal with this, you must first multiply the number being divided with 10 for one decimal digit accuracy and then perform the division in PLC. Thus, the result is an integer number with one implied decimal digit. To display this number on a man-machine interface (MMI), which has only one decimal digit, you have to read the integer value from PLC and display it after dividing it by 10 in the MMI.

Compare Compare instructions are used for comparing data. Generally, a compare instruction has enable input, two data inputs, and one output. Data inputs are register locations or constants. There are six basic types of compare instructions:

1. EQ is used to check if input 1 and input 2 are equal. When the enable condition becomes true, input 1 data

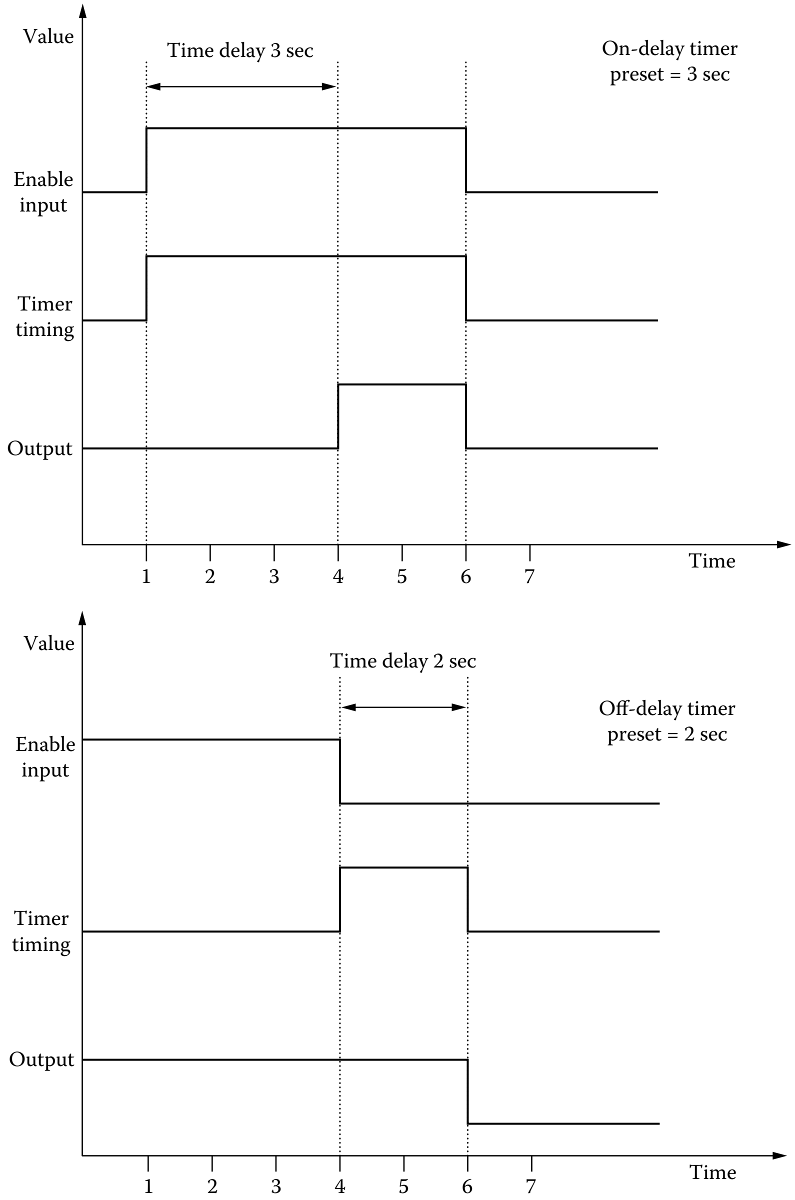


FIG. 5.5h
Timing diagram for PLC timers instruction.

are compared with input 2 data to see if they are equal and if so, output of EQ instruction becomes on.

2. NEQ is used to check if input 1 and input 2 are not equal. When the enable condition becomes true, input 1 data are compared with input 2 data to see if they are not equal and if so, output of NEQ instruction becomes on.
3. GT is used to check if input 1 is greater than input 2. When the enable condition becomes true, input 1 data are compared with input 2 data to see if input 1 is greater than input 2 and if so, output of GT instruction becomes on.
4. GE is used to check if input 1 is greater than or equal to input 2. When the enable condition becomes true, input 1 data are compared with input 2 data to see if input 1 is greater than or equal to input 2 and if so, output of GE instruction becomes on.
5. LT is used to check if input 1 is less than input 2. When the enable condition becomes true, input 1 data are compared with input 2 data to see if input 1 is less than input 2 and if so, output of LT instruction becomes on.
6. LE is used to check if input 1 is less than or equal to input 2. When the enable condition becomes true, input 1 data are compared with input 2 data to see if input 1 is less than or equal to input 2 and if so, output of LE instruction becomes on.

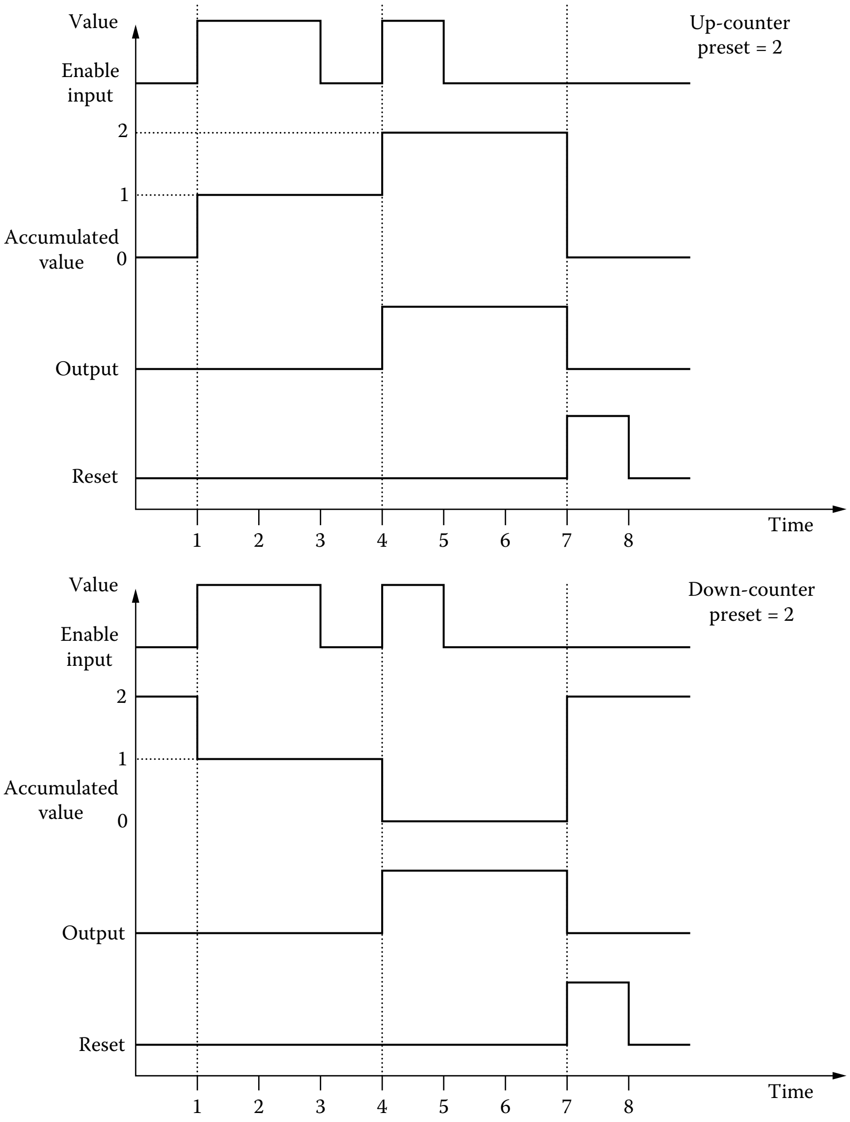


FIG. 5.5i
Timing diagram for PLC counters instruction.

Data Transfer This instruction is used for moving data. It has enable input, a source input, and a destination output. MOVE is a basic data transfer instruction. MOVE is used to transfer data from a source memory location to a destination memory location. When the enable condition becomes true, source data are transferred to destination.

Memory Structure

PLC memory is embedded in CPU. The following types of memory are used in CPU:

1. Read-only memory (ROM) is used to store the firmware or operating system of the PLC. It is not available to the user for data or user program storage.
2. Random access memory (RAM) is used to store user program, data, and results of various arithmetic and logical operations performed by the CPU. RAM is battery-backed to prevent data loss in the event of a power failure.
3. Erasable programmable read-only memory (EPROM) is similar to ROM, but can be erased with ultraviolet light and reprogrammed.
4. Electrically erasable programmable read-only memory (EEPROM) is similar to ROM, but can be erased with voltage and reprogrammed. Newer types of EEPROM, such as Flash memory, are becoming available.

When we refer to the PLC memory structure, we are talking about the user memory, i.e., RAM.

PLC user memory is specified in bytes. One byte represents 8 bits of data. One register represents two bytes or 16 bits of data. If a PLC has 1K memory, it means it has 1056 bytes or 528 registers of memory available for the user. PLC memory is divided in the following major areas and is referenced by register number/bit number in a register by user program:

1. User program memory: This memory stores user ladder logic or any other form of user program.
2. Input status memory: This memory stores the status of inputs physically connected to the PLC. Each digital field input takes 1 bit of input memory. Each analog field input takes 8, 12, or 16 bits of input memory depending on the resolution of analog input module.
3. Output status memory: This memory stores status of outputs physically connected to the PLC. Each digital field output takes 1 bit of output memory. Each analog field output takes 8, 12, or 16 bits of output memory depending on the resolution of analog output module.
4. Timer status memory: This memory stores the status, preset, and elapsed time values of timers in the PLC.
5. Counter status memory: This memory stores status, preset, and accumulated values of counters in the PLC.
6. Numeric data memory: This memory stores numeric data in the PLC. It is used to store data for manipulation by user program and the results of data manipulation.

Ladder Logic Programming Devices

Programming devices are required to write, edit, and monitor user programs, to download them into the PLC through the programming port on the CPU, and to monitor and force I/O status. It is also used for debugging PLC programs and to view PLC diagnostics. There are hand-held programmers and personal computer-based programmers available for PLC programming.

A hand-held programmer is a dedicated programming device manufactured and supplied by a PLC manufacturer for its PLCs. It has a keypad and an LCD display. You can connect the programmer to the PLC using the programming cable that comes with the programmer and do online programming or make program changes. You cannot store programs in this type of programmer.

A personal computer-based programmer needs the following to write PLC ladder logic:

1. PC with DOS or Windows operating system
2. Serial port or interface card
3. PLC programming software supplied by the PLC manufacturer
4. Programming cable supplied by the PLC manufacturer

Personal computer-based programmers are very popular for the following reasons.

1. Off-line programming can be done and programs can be stored on hard disk.
2. Program documentation becomes easier.
3. Big screen makes it easier to see large ladder rungs.
4. Program printouts can be easily made for permanent records.
5. Some programmers even allow logic simulation by executing the program in a PC as if it were being executed in the PLC.

In the industrial environment today, laptops are used for on-line programming and program changes.

Programming Considerations

The following are some of the major considerations while writing any PLC program:

1. Program scan time: This is a very important factor in any PLC program. As explained earlier, scan time is time taken by PLC to complete its three main tasks: read input status, execute PLC program, and write output status. PLC repeats these three tasks continuously. If the total scan time of a program is not less than half the time at which the fastest input changes in the system, there is a possibility that PLC will not be able to detect a change in that input and will fail to respond to the change. The following are ways to reduce the scan time in a PLC system:
   - Use faster CPU
   - Use interrupt routine
   - Make program compact by logic optimization (Reference 1) and simplification
   - Organize program in such a way so that for elements in parallel, the one most likely to be true is highest in the ladder and so on, and for elements in series, the one least likely to be true is leftmost in the ladder and so on. This way, when one of the parallel elements is found true, the rest of the parallel elements are not required to be solved and program execution progresses to the next rung. The same is true for elements in series.
2. Program scan direction: It is very important in PLC programming to know how the PLC program scan is done. Some PLCs, e.g., Allen-Bradley PLCs, scan programs left to right (called rung scanning), whereas others, e.g., Modicon PLCs, scan programs top to bottom (called column scanning). Both methods are appropriate; however, a programmer should be aware of the scan method used by the PLC because it has an impact on whether a coil gets energized or de-energized in the same scan or the next scan after input conditions become true.
3. Fail-safe programming: A program should be written in such a way that it always puts the process it is controlling in safe mode when any of the critical system components fails. This is called fail-safe programming. In the system, first define fail-safe conditions and then determine what outputs are required to be turned off or turned on to achieve those conditions. There are many components in a PLC system, like I/O modules, power supply module, and cables. The program

should be written in such a way that if there is a loss of input signal due to power loss or cabling problems or failure of the input module itself, then the outputs will go to a state that is safe. Also, programs should be designed so that they check for problems, and shut down in safe ways. Use modular well-designed predictable programs and PLC's built-in functions for error and failure detection.
4. Emergency shutdown: There should be a hardwired relay, called master control relay (MCR), and a normally closed (NC) emergency stop button wired in series in the system. MCR is normally kept energized through emergency stop button by the PLC system power. Power to the physical outputs is supplied through the contacts of this MCR. In case of emergency, the emergency stop button is activated and the power to all the outputs is removed.
5. PLC fault handling and diagnostics: A program should also have the ability to use various diagnostics available in the CPU and generate alarms for any fault condition in the system or failure of any hardware that might create unsafe operating conditions. There are status registers in the PLC that can be used to generate alarms or messages to help diagnose the problems and indicate prompt corrective actions.
6. Program structure and memory allocation: Program should be divided into smaller separate blocks by function. This will help in troubleshooting the code. Each functional block should be allocated its own block of memory so that there are no conflicts with shared memory.

Program Documentation

Program documentation is a very important aspect of PLC programming, though many times it is overlooked or given less importance. These days, programming software is available with good program documentation capabilities. Good program documentation has the following advantages:

- Helps in program debugging
- Helps program users understand and maintain/modify the logic
- Provides on-line information about field connections, thus in so many cases it eliminates the need to have the wiring diagrams handy in troubleshooting

Program documentation involves the following main areas:

- I/O names and descriptors: These are the short user-defined meaningful names given to field inputs and outputs with a more detailed description of the I/O signals. The name is usually a shortened field tag name, and the descriptor is a short explanation about the signal.
- Internal memory names and descriptors: These are the short user-defined meaningful names given to internal coils and registers with a more detailed description of their function. The name is usually a shortened identifier and the descriptor is a short explanation about the function.
- Program comments: Rung titles and comments are inserted between the logic rungs in the program to explain the purpose of a block of logic and how that piece of logic works.

Program documentation is a continuous process during PLC programming. The following are the commonly used steps for program documentation:

- Define all the I/O names and descriptors
- Define all the internal memory names and descriptors
- Write comments before writing a piece of logic
- Keep adding/editing comments and other descriptors/names as you write the program

There are utilities available in programming software to import/export names, descriptors, and comments from/to word processors and spreadsheets. Use of Excel spreadsheets in creating PLC documentation saves a lot of time.

A PLC program documentation printout consists of the following:

- List of I/O names and descriptors
- List of internal memory names and descriptors
- Cross-reference list for I/Os and internal memory
- Usage list of I/Os and internal memory
- Total PLC memory used and available
- PLC ladder logic listing
- PLC system configuration

PLC Hardware Configuration

PLC programming also involves hardware configuration. Hardware configuration lets the CPU know what components make up the whole PLC system and their addresses. PLC has to be configured first before it can be programmed with logic. PLC hardware configuration involves the following:

1. Select type of rack (chassis) used (5- or 10-slot)
2. Select type of card that is installed in each slot (digital input/output or analog input/output)
3. CPU configuration for memory allocation of various types of user memory
4. CPU configuration for communication ports (programming port and so on)
5. Address allocation for physical I/Os connected to I/O modules

PLC programming software is used to develop and download PLC configuration into CPU. A typical PLC system hardware and CPU configuration for GE Fanuc PLC is discussed next.

LADDER PROGRAM STRUCTURE

Typical GE Fanuc PLC

Figure 5.5j shows a typical GE Fanuc program structure. A program is contained in a folder and can be divided into various logic blocks. When you start to write a program, first you create a folder with an appropriate name, e.g., TEST. There is a main program block where you write ladder logic and also make calls to other logic blocks. These other logic blocks are declared in the main logic block and are called subroutines. In this example, the main program block is divided in three sections or subroutines (ANALOG, ALARMS, and OUTPUT) and makes calls to these

```
                       I D E N T I F I E R   T A B L E

            IDENTIFIER     IDENTIFIER TYPE         IDENTIFIER DESCRIPTION
            ----------     ---------------     --------------------------------
            ANALOG         SUBROUTINE              ANALOG INPUT SCALING LOGIC
            ALARMS         SUBROUTINE              ALARMS WITH DEADBAND LOGIC
            OUTPUT         SUBROUTINE              ANNUNCIATOR OUTPUT LOGIC
            TEST           PROGRAM NAME            TEST PROGRAM

  |[        BLOCK DECLARATIONS       ]

                   +-------+
          SUBR  1 |ANALOG |        LANG: LD  (* ANALOG INPUT SCALING LOGIC *)
                   +-------+

                   +-------+
          SUBR  2 |ALARMS |        LANG: LD  (* ALARMS WITH DEADBAND LOGIC *)
                   +-------+

                   +-------+
          SUBR  3 |OUTPUT |        LANG: LD  (* ANNUNCIATOR OUTPUT LOGIC  *)
                   +-------+


  |[      START OF PROGRAM LOGIC      ]
  |
  | << RUNG 4  STEP #0001 >>
  |
  |ALW_ON
  |%S0007  +-------------+
  +--] [---+ CALL ANALOG +
  |        | (SUBROUTINE)|
  |        +-------------+

  | << RUNG 5  STEP #0003 >>
  |
  |ALW_ON
  |%S0007  +-------------+
  +--] [---+ CALL ALARMS +
  |        | (SUBROUTINE)|
  |        +-------------+
  |
  | << RUNG 6  STEP #0005 >>
  |
  |ALW_ON
  |%S0007  +-------------+
  +--] [---+ CALL OUTPUT +
  |        | (SUBROUTINE)|
  |        +-------------+
  |
  |[       END OF PROGRAM LOGIC      ]
  |
```

FIG. 5.5j
GE Fanuc PLC program structure.

TABLE 5.5k
GE Fanuc PLC Memory Identifiers and Address Format

Type of Memory	*Memory Identifier*	*Address Format*
DIGITAL INPUT	%I	%I0001
DIGITAL OUTPUT	%Q	%Q0001
ANALOG INPUT	%AI	%AI001
ANALOG OUTPUT	%AQ	%AQ001
INTERNAL COIL	%M	%M0001
REGISTER	%R	%R0001
SYSTEM	%S	%S0001
TEMPORARY	%T	%T0001
GLOBAL	%G	%G0001

subroutines. These subroutines contain specific logic as explained below.

- ANALOG subroutine handles analog input signals and converts the raw analog value into engineering units for alarming purpose. This value in engineering units can also be displayed on any operator interface (MMI).
- ALARMS subroutine generates alarms with deadband.
- OUTPUT subroutine generates physical outputs for annunciator alarms.

This block structure makes it easier to understand and troubleshoot the program.

Table 5.5k shows various types of memory identifiers for PLC addressing.

Table 5.5l shows a typical variable declaration table.

Figure 5.5m shows a typical PLC system hardware and CPU configuration.

Figure 5.5n shows a typical PLC logic for analog input scaling.

Figure 5.5o shows a typical PLC logic for alarming with deadband.

Figure 5.5p shows a typical PLC logic for annunciator outputs.

TABLE 5.5l
GE Fanuc PLC Variable Declaration Table

Reference	*Nickname*	*Description*
%I0001	PSL001	TANK TK-1 LOW PRESSURE
%Q0001	SV001	TANK TK-1 AIR SOLENOID
%AI001	PT001	TANK TK-1 PRESSURE TX
%AQ001	PCV001	TANK TK-1 PRESSURE CONTROL VALVE
%M0001	TK1PLTD	TANK TK-1 LOW PRESSURE FOR10 SEC TIMER DONE
%R0001	TK1PLTMR	TANK TK-1 LOW PRESSURE 10 SEC TIMER

Typical Allen-Bradley PLCs

A program in Allen-Bradley PLC-5 is organized in a file structure. First, you create a project with an appropriate name. In the project, there are program files as explained below:

File #	*Name*	*Description*
0	System	PLC operating system, unavailable to programmers
1	Undefined	PLC operating system, unavailable to programmers
2	Main program	Main program code and subroutine calls
3–999	Subroutines	Subroutine code, called by main program

In a typical project, there are different types of data files, such as:

Data files O0 (output) and I1 (input) are physical I/O status files.
Data file S2 (status) is a general status file.
Data files B3 (binary), T4 (timers), C5 (counters), R6 (control word), N7 (integer), F8 (floating) are reserved for the main program file. Other data files can be designated for subroutines as needed.

Addressing is done in the following format:

XY:WW:ZZ

where
X is the file identifier such as I for input, O for output, S for status, B for binary, T for timers, C for counters, R for real, N for integer, and F for floating point
Y is the file number (in decimal) such as 0 for output, 1 for input
WW is the word number (in decimal) in file such as 00 for word #1
ZZ is the bit number (in octal) in a word such as 00 for bit #1

A program in Allen-Bradley's ControlLogix PLC is organized by tasks. First, you should identify a project with an appropriate name and controller type. In the project, there are a variety of tasks, programs, and routines. For example, the controller contains controller tags, controller fault handler, and power-up handler. The controller tags are global tags that can be used by all programs.

Controller supports 31 separate tasks. Each task can contain multiple programs, status information, and configuration information.

There is always a main task, which is created when you create the project and is by default a continuous task. Other (up to 30) tasks are user-defined and are periodic tasks. There

```
|------+------+--- RACK 0 --+------+------|
|  PS  |  1   |  2   |  3   |  4   |  5   |
|======= PROGRAMMED  CONFIGURATION =======|
|      |      |      |      |      |      |
|PWR321|CPU351|MDL240|MDL340|ALG221|ALG391|
|      |      |      |      |      |      |
|      |      |I AC16|Q AC16|IALGI4|QALGI2|
|      |      |      |      |      |      |
|      |      |RefAdr|RefAdr|RefAdr|RefAdr|
|      |      |%I0001|%Q0001|AI0001|%AQ001|
|      |      |      |      |      |      |
|      |      |      |      |      |      |
|      |      |      |      |      |      |
|      |      |      |      |      |      |
+------+------+------+------+------+------+


SERIES 90-30 MODULE IN RACK 0 SLOT  1
+------+------------------ SOFTWARE  CONFIGURATION ----------------
| SLOT |  Catalog #: IC693CPU351       SERIES 90-30 CPU, MODEL 351
|    1 |
|      |
|CPU351|-----------------------------------------------------------
|      |  IOScan-Stop: NO        Baud Rate  : 19200
|      |  Pwr Up Mode: LAST      Parity     : ODD
|      |  Logic/Cfg  : RAM       Stop Bits  : 1
|      |  Registers  : RAM       Modem TT   : 0  1/100 Second /Count
|      |  Passwords  : ENABLED   Idle Time  : 10 Seconds
|      |  R/S Switch : DISABLED
|      |
|      |  Chksum Wrds:   8       Sweep Mode : NORMAL
|      |  Tmr Faults : DISABLED  Sweep Tmr  : N/A     msec
|      |
+------+
CPU MEMORY CONFIGURATION FOR Model 351 CPU:

      Discrete Input    (%I)  2048  Points
      Discrete Output   (%Q)  2048  Points
      Internal Discrete (%M)  4096  Points
      System Use        (%S)   128  Points
      Temporary Status  (%T)   256  Points
      GENIUS Global     (%G)  1280  Points
                              ----  -----
      TOTAL DISCRETE MEMORY:  9856  Points

      Analog Input     (%AI)  2048  Words
      Analog Output    (%AQ)   512  Words
      Register Memory  (%R)   9999  Words

      TOTAL LOGIC MEMORY     56802  Bytes

      CPU MEMORY TOTAL       81920  Bytes
```

FIG. 5.5m

GE Fanuc PLC rack and CPU configuration.

can be up to 32 programs in a task. The program contains program tags and routines. Program tags are local tags that can be used by routines within an individual program. There is a main routine, which gets executed first in each program. The main routine is used to make calls to other routines.

A program can be scheduled or unscheduled for execution. Scheduled programs are executed and unscheduled programs are not executed. Data types for tags can be user-defined or predefined. There are module-defined data types, as well. There are no physical or internal addresses assigned to tags. Tags are referenced in the programs by their names. Tags have various parameters depending on their data type. I/O configuration is used to configure the physical I/O modules.

```
|
(************************************************************************)
| (* LOGIC TO CONVERT RAW ANALOG INPUT INTO ENGINEERING UNITS FOR REACTOR  *)
| (* TEMPERATURES USING FOLLOWING FORMULA                                  *)
| (* TEMP (IN ENGG. UNIT) = TEMP (RAW ANALOG INPUT)*FULL SCALE RANGE/32000 *)
|
(************************************************************************)
|
| << RUNG 4  STEP #0002 >>
|
|ALW_ON
|%S0007  +-----+                +-----+
+--] [---+ DIV_+----------------+ MUL_+-
|        | INT |                | INT |
|        |     |                |     |
|REACTOR |     | REACTOR REACTOR |     | REACTOR
|TEMP. 1 |     | TEMP. 1 TEMP. 1 |     | TEMP. 1
|ANALOG  |     | IN DEG  IN DEG  |     | IN DEG
|INPUT   |     | C       C       |     | C
|TI_1051 |     | TI1051  TI1051  |     | TI1051
|%AI0001-+I1  Q+-%R0001  %R0001 -+I1  Q+-%R0001
|        |     |                |     |
| CONST -+I2   |          CONST -+I2   |
| +32000 +-----+          +00100 +-----+
|
| << RUNG 5  STEP #0005 >>
|
|ALW_ON
|%S0007  +-----+                +-----+
+--] [---+ DIV_+----------------+ MUL_+-
|        | INT |                | INT |
|        |     |                |     |
|REACTOR |     | REACTOR REACTOR |     | REACTOR
|TEMP. 2 |     | TEMP. 2 TEMP. 2 |     | TEMP. 2
|ANALOG  |     | IN DEG  IN DEG  |     | IN DEG
|INPUT   |     | C       C       |     | C
|TI_1052 |     | TI1052  TI1052  |     | TI1052
|%AI0002-+I1  Q+-%R0002  %R0002 -+I1  Q+-%R0002
|        |     |                |     |
| CONST -+I2   |          CONST -+I2   |
| +32000 +-----+          +00200 +-----+
|
|
|[    END OF SUBROUTINE LOGIC     ]
|
```

FIG. 5.5n
GE Fanuc PLC logic for analog inputs.

Typical Modicon PLC 984

A program in Modicon PLC is organized in networks. First, you create a program file. In the program file, there are segments that are made up of logic networks. In PLC 984, there are 1–32 segments, and you can write logic networks in each of these segments. Network numbers start from 1 and are continuous for the program. Addressing is done in the following format for data tables:

XYYYY

where X is data type identifier such as 0 for discrete o/p or coil (used to drive real outputs through output module or to set internal coils), 1 for discrete input (used to drive contacts in the logic program, controlled by input module), 3 for input register (holds numeric input from external source like analog or high-speed counter input module), and 4 for output holding registers (used to store numeric data for internal use or analog output module).

YYYY is reference number in decimal such as 1, 2, etc., with upper limit as configured.

Total memory for all data types should not exceed processor user memory.

Access and Programming Modes

The following are a few ways of connecting a programming device with a PLC:

1. Serial connection
2. Ethernet
3. Data Highway Plus (DH+)

```
|[   START OF SUBROUTINE LOGIC    ]
| (****************************************************************************)
| (* HIGH ALARM WITH DEADBAND OF 2 DEG C                                      *)
| (****************************************************************************)
| << RUNG 4  STEP #0002 >>
|                                                                        REACTOR
|                                                                        TEMP. 1
|                                                                        > 70
|                                                                        DEG C
|ALW_ON                                                                  TI1051U
|%S0007  +-----+                                                         %M0001
+--] [---+ GT_ |+----------------------------------------------------------( )--
|        | INT ||
|        |     ||
|REACTOR |     ||
|TEMP. 1 |     ||
|IN DEG  |     ||
|C       |     ||
|TI1051  |     ||
|%R0001 -+I1  Q++
|        |     |
| CONST -+I2    |
| +00070 +-----+
| << RUNG 5  STEP #0005 >>
|                                                                        REACTOR
|                                                                        TEMP. 1
|                                                                        > 68
|                                                                        DEG C
|ALW_ON                                                                  TI1051L
|%S0007  +-----+                                                         %M0002
+--] [---+ GT_ |+----------------------------------------------------------( )--
|        | INT ||
|        |     ||
|REACTOR |     ||
|TEMP. 1 |     ||
|IN DEG  |     ||
|C       |     ||
|TI1051  |     ||
|%R0001 -+I1  Q++
|        |     |
| CONST -+I2    |
| +00068 +-----+
| << RUNG 6  STEP #0008 >>
|
|REACTOR                                                                 REACTOR
|TEMP. 1                                                                 TEMP. 1
|> 70                                                                    HIGH
|DEG C                                                                   ALARM
|TI1051U                                                                 TI1051A
|%M0001                                                                  %M0003
+--] [----------+----------------------------------------------------------( )--
|               |
|REACTOR REACTOR|
|TEMP. 1 TEMP. 1|
|HIGH    > 68   |
|ALARM   DEG C  |
|TI1051A TI1051L|
|%M0003  %M0002 |
+--] [-----] [--+
|
|[    END OF SUBROUTINE LOGIC     ]
```

FIG. 5.5o
GE Fanuc PLC logic for alarm with deadband.

```
|[   START OF SUBROUTINE LOGIC    ]
|
| << RUNG 3  STEP #0001 >>
|
|REACTOR                                                                 REACTOR
|TEMP. 1                                                                 TEMP. 1
|HIGH                                                                    HIGH
|ALARM                                                                   ANN. OP
|TI1051A                                                                 TAH1051
|%M0003                                                                  %Q0001
+--] [-------------------------------------------------------------------( )--
|
| << RUNG 4  STEP #0003 >>
|
|REACTOR                                                                 REACTOR
|TEMP. 2                                                                 TEMP. 2
|LOW                                                                     LOW
|ALARM                                                                   ANN. OP
|TI1052A                                                                 TAL1052
|%M0006                                                                  %Q0002
+--] [-------------------------------------------------------------------( )--
|
|[    END OF SUBROUTINE LOGIC     ]
|
```

FIG. 5.5p
GE Fanuc PLC logic for annunciator outputs.

Once a programming device has been connected to the PLC, you can place the CPU in one of the following modes:

1. Run: In this mode CPU does all three tasks: read inputs, execute program logic, write to outputs.
2. I/O Scan Only: In this mode CPU does only one task: read inputs.
3. Stop/Program: In this mode CPU stops all three tasks: read inputs, execute program logic, write to outputs, and the PLC can be programmed in this mode.

The PLC programming and configuration can be done in the following two ways:

1. Off-Line Programming: In this mode, a PLC can be programmed on a PC-based programmer without connecting the programmer to the PLC. The program can be saved on computer hard disk or on a floppy disk for download to the PLC later.
2. On-Line Programming: In this mode, a PLC can be programmed on a PC-based programmer after connecting the programmer to the PLC on its programming port. The program can be saved on the computer hard disk or on a floppy disk and at the same time downloaded to the PLC. Some PLCs allow a program to be downloaded into the PLC only in stop/program mode. If PLC was in run mode prior to program download, a message is given to the programmer that PLC is in run mode and its mode needs to be changed to stop/program mode before the download can be done.

Developing the PLC Program Logic

There are no set rules or guidelines on how to develop PLC program logic. However, the following steps will help in developing PLC program logic:

1. Write a process description or obtain it from the process or operations engineer
2. Make a sketch of the process or obtain it from the process or operations engineer
3. Prepare control strategy narratives
4. Make a list of field devices and their inputs and outputs that will be connected to PLC
5. Assign addresses to inputs/outputs
6. Develop flowcharts (if required)
7. Develop Boolean logic diagrams
8. Develop PLC hardware configuration based on number and type of I/Os
9. Write PLC ladder logic, i.e., convert Boolean logic into actual ladder logic
10. Test PLC ladder logic

Level Control Example Figure 5.5q illustrates a tank level control application where a remotely operated pump is used as the final control element in the system. In this example, a ladder logic program will be developed, using a GE Fanuc PLC, using the steps that are listed above.

1. Process description: Water tank TK-001 is being filled continuously and its level is to be controlled by an on/off drain pump.

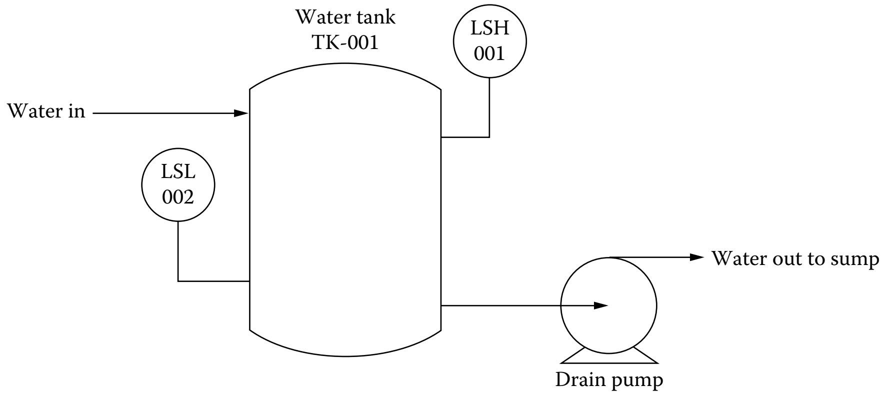


FIG. 5.5q
Tank level control.

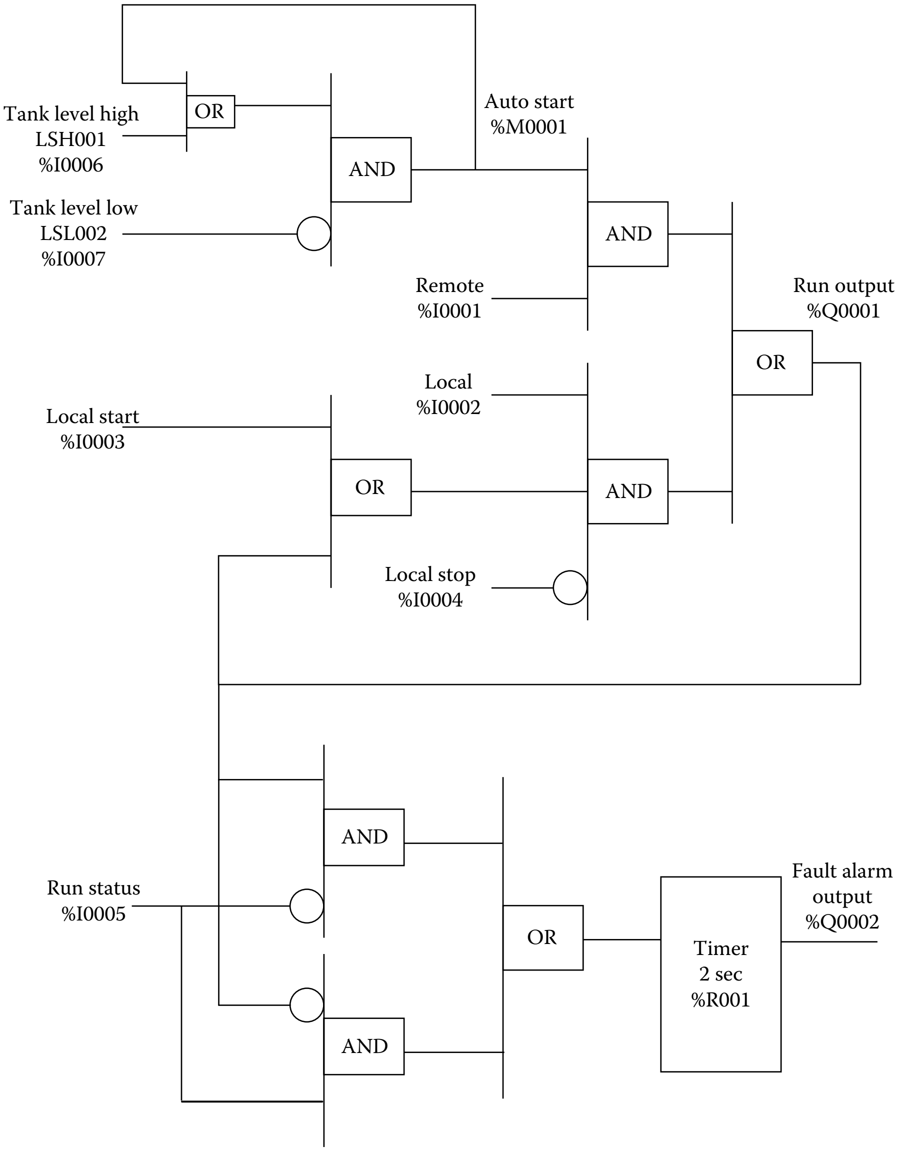


FIG. 5.5r
Boolean logic for tank level control example.

2. Make a sketch of the process: Make a sketch as shown in Figure 5.5q.
3. Prepare control strategy narrative for this pump, which has two modes of operation:
    - Remote mode: In this mode, its operation is controlled automatically by the level in the tank. When the level rises to its high limit (say 90%), the water pump should start to drain the water to the sump. Pump will continue to drain water until the level drops to its low limit (say 10%) in TK-001.
    - Local mode: In this mode, the pump operation is controlled by the operator out in the field, and the automatic level control logic is bypassed.
4. Make a list of field devices (sensors and so on) and their inputs and outputs that will be connected to the PLC.

Field Device	*Input*	*Output*	*Remarks*
Local/remote switch	02	—	For pump operation control
Local start pushbutton	01	—	To start pump in local mode
Local stop pushbutton	01	—	To stop pump in local mode
Water pump	01	01	Input = Run status, Output = Run command
Motor fault alarm	—	01	For annunciation
LSH001	01	—	High-level switch
LSL002	01	—	Low-level switch

5. Assign addresses to inputs/outputs.

Field Device	*Input*	*Output*	*PLC Address*
Local/remote switch	02	—	%I0001 (Remote), %10002 (Local)
Local start pushbutton	01	—	%I0003
Local stop pushbutton	01	—	%I0004
Water pump	01	01	%I0005 (Run status), %Q0001 (Run command)
Motor fault alarm	—	01	%Q0002
LSH001	01	—	%I0006
LSL002	01	—	%I0007

6. Develop Boolean logic diagram as shown in Figure 5.5r.
7. Develop PLC hardware configuration as shown in Figure 5.5m without the analog modules in slots 4 and 5.
8. Write PLC ladder logic, i.e., convert Boolean logic developed in previous step into ladder logic as shown in Figure 5.5s.
9. Download program in PLC and test the program as explained below.

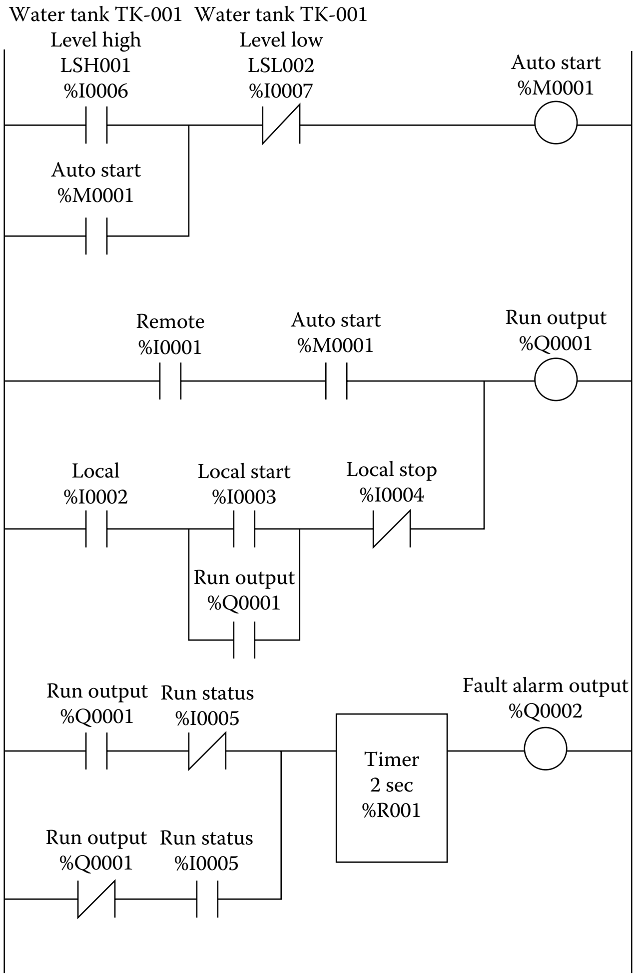


FIG. 5.5s
PLC logic for tank level control example.

TESTING AND SIMULATION

Once the logic is entered in the PLC, the next step is to debug the program to make sure that the logic works the way it is supposed to and to correct any errors made while developing and writing the PLC program.

Input modules with switches to simulate the input signals are used. Such a simulation module is installed to simulate the digital inputs to the PLC. The PLC is placed in the run mode, and its inputs are simulated by operating the corresponding switches. The logic outputs are checked to see if the program properly carries out the required logic. The status of the internal coils is determined by using the PLC data table.

Instead of using simulation modules, it is also possible to use the FORCE I/O function in the PLC to enable or disable inputs in order to simulate field inputs. FORCE ON can be used to turn input on and FORCE OFF can be used to turn input off. Once the inputs are forced on or off you would proceed to check if the program has turned the various

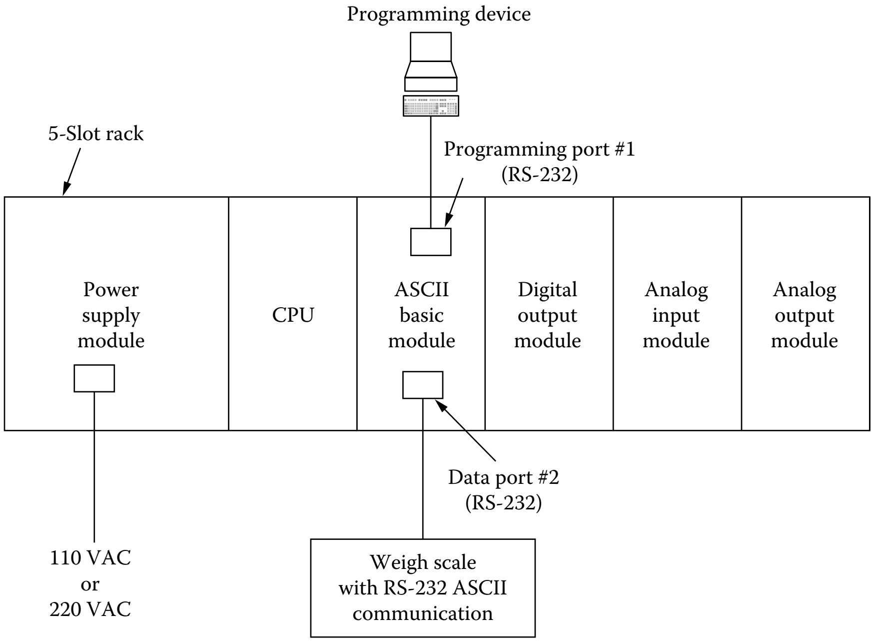


FIG. 5.5t
PLC system with ASCII basic module for weigh scale interface.

```
+-----------------------------------------+
|------+------+--- RACK 0 --+------+------|
|  PS  |  1   |  2   |  3   |  4   |  5   |
|======= PROGRAMMED  CONFIGURATION =======|
|      |      |      |      |      |      |
|PWR321|CPU351|FRGN  |MDL340|ALG221|ALG391|
|      |      |      |      |      |      |
|      |      |      |Q AC16|IALGI4|QALGI2|
|      |      |      |      |      |      |
|      |      |RefAdr|RefAdr|RefAdr|RefAdr|
|      |      |      |%Q0001|AI0001|%AQ001|
|      |      |      |      |      |      |
|      |      |AI0009|      |      |      |
|      |      |%AQ009|      |      |      |
|      |      |      |      |      |      |
+------+------+------+------+------+------+
```

```
SERIES 90-30 MODULE IN RACK 0 SLOT  2
+------+------------------ SOFTWARE  CONFIGURATION ---------------------------
| SLOT |  Catalog #: FOREIGN                FOREIGN MODULE
|    2 |
|      |
|FRGN  |----------------------------------------------------------------------
|      |  Module ID  :    3
|      |  %I Ref Adr :   %I0001  Byte 1      : 00000001  Byte 9      : 00
|      |  %I Size    :      0    Byte 2      : 00000010  Byte 10     : 00
|      |  %Q Ref Adr :   %Q0001  Byte 3      : 00        Byte 11     : 00
|      |  %Q Size    :      0    Byte 4      : 00        Byte 12     : 00
|      |  %AI Ref Adr:  %AI0009  Byte 5      : 00        Byte 13     : 00
|      |  %AI Size   :      8    Byte 6      : 00        Byte 14     : 00
|      |  %AQ Ref Adr:   %AQ009  Byte 7      : 00        Byte 15     : 00
|      |  %AQ Size   :      8    Byte 8      : 00        Byte 16     : 00
|      |  %R Ref Adr :   %R0001  %R Ref Adr :   %R0001
+------+  %R(in) Size:      0    %R(out)Size:      0
```

FIG. 5.5u
GE Fanuc PLC rack and ASCII basic module configuration.

This basic module code reads the weigh scale data string from port #2 and stores the crate weight and tare weight values in PLC memory. Weigh scale data string format is <STX>xxxx yyyy<CR> where xxxx is crate weight and yyyy is tare weight. Programming PC is to be connected at port #1 for programming/debugging.

```
 10   REM DATALINK TO WEIGH SCALE                      -- remark
 20   STRING 1026,40                                   -- allocates string memory
 30   SETCOM 9600                                      -- sets communication parameters for programming (port #1)
 40   SETCOM #9600,E,7,1,N,0                           -- sets communication parameters for weigh scale (port #2)
 50   SETINPUT 1,1,13,17,10,10                         -- input statement configuration
 60   INPUT # $(1)                                     -- inputs data from weigh scale (port #2)
 70   PRINT “WEIGH SCALE RESPONSE IS : “, $(1)         -- prints weigh scale data string on port #1
 80   $(2) = LEFT$ ($(1),1)                            -- assigns 1st character of data string to $(2)
 90   $(3) = MID$ ($(1),2,4)                           -- assigns crate weight data string to $(3)
100   $(4) = MID$ ($(1)7,4)                            -- assigns tare weight data string to $(4)
110   IF $(2) = CHR$ (02) THEN GOTO 160 ELSE GOTO 120  -- if 1st character is <STX> then valid data
120   OUT(2,0) = 1                                     -- sets bad value flag in PLC memory (bit 1 of AI011)
130   WEIGHT = 0                                       -- sets crate weight to 0
140   TWEIGHT = 0                                      -- sets tare weight to 0
150   GOTO 190
160   WEIGHT = VAL ($(3))                              -- sets crate weight to value read from weigh scale
170   TWEIGHT = VAL ($(4))                             -- sets tare weight to value read from weigh scale
180   OUT(2,0) = 0                                     -- resets bad value flag in PLC memory (bit 1 of AI011)
190   PRINT “WEIGHT IS : “, WEIGHT                     -- prints crate weight on port #1
200   PRINT “TARE WEIGHT IS : “, TWEIGHT               -- prints tare weight on port #1
210   OUT (0) = WEIGHT                                 -- stores crate weight in PLC memory (AI009)
220   OUT (1) = TWEIGHT                                -- stores tare weight in PLC memory (AI010)
230   GOTO 50
240   RETURN
```

FIG. 5.6v
Basic program for ASCII basic module.

outputs on or off per the logic. Again, the status of the internal coils is determined by using the PLC data table.

In case of large PLC systems with a lot of I/Os, you should use a graphical process simulator package, which can be developed by using MMI packages. As was discussed in Section 4.13 in Chapter 4, processes can be displayed graphically on the screen and respond to the PLC signals in the same manner as the actual process would. Equipment can be operated from the screen, and their status can be seen on the screen. Naturally, interface cards are needed to connect PLC to computer running simulator programs.

ADVANCES IN PROGRAMMING

Until recently, most of the PLC programming softwares available in the market were DOS-based for ladder logic programming. Recently, as Windows-based laptop computers are becoming inexpensive and popular, PLC manufacturers are coming out with Windows-based programming software.

The advantages of Windows-based programming software include ease of programming, the ability to use multiple windows to do both the programming and the troubleshooting, and an improvement in the on-line help available for programming.

Another significant advancement is the PLC's ability to communicate with intelligent devices using interface cards that are programmable in the Basic language. This interface module is called ASCII Basic module and is physically installed in the PLC rack like any other module; it is programmed using a serial port on the module.

One such system used for weigh scale interfacing is shown in Figure 5.5t. In this system, ASCII Basic module is programmed in Basic language to read the measured weight from the weigh scale using RS-232 serial communication.

Figure 5.5u shows the PLC configuration, and Figure 5.5v shows the Basic code for an ASCII module. This type of application can be used to connect the PLC to any intelligent device that supports ASCII serial communication. Some typical applications include RF tag readers and labeling stations.

As PC costs keep dropping, personal computers are used more and more in place of conventional PLCs. PCs can solve ladder logic, scan I/Os, and communicate with other network devices, if the appropriate software is loaded into them. However, there are safety considerations to be resolved if the PC is shared among several applications. Yet there already are a number of small-scale PC-based control systems in use today, in areas such as data acquisition, process monitoring, and batch control.

An alternative to relay ladder logic (RLL) is state logic. It is a language that forces the user to build the program with finite states. It uses a natural language, known as English Control Language Programming Software (ECLiPS), which allows the programmers to write their control commands in their own words. State logic was created to reduce the program development and modification time.

Reference

1. Lipták, B. G., *Instrument Engineers' Handbook,* 4th edition, Process Control, Section 5.4 and other sections in Chapter 5, Boca Raton, FL: CRC Press, 2005.

Bibliography

Allen-Bradley, *Logix5000 Controllers Programming Documentation Reference Guide.*

Allen-Bradley, *PLC-5 Programming Software, Instruction Set Reference,* Release 4.4.

Allen-Bradley, *PLC-5 Programming Software, Software Configuration and Maintenance.*

Carrow, R. A., *Soft Logic: A Guide to Using a PC as a Programmable Logic Controller,* New York: McGraw-Hill, 1997.

Cox, R. D., *Technician's Guide to Programmable Controllers,* New York: Delmar, 1995.

GE Fanuc, *Logic Master 90-70 Programming Software User's Manual,* August 1992.

GE Fanuc, *Series 90-70 Programmable Controller Reference Manual,* July 1992.

GE Fanuc, *MegaBasic Language Reference and Programmer's Guide,* September 1994.

Horner Electric, *ASCII Basic Module User's Guide,* May 1995.

IEC 61131-3, "Programming Languages," IEC, 1993.

Lewis, R. W., *Programming Industrial Control Systems using IEC 61131-3,* U.K.: British Institute of Electrical Engineers (IEE), 1998.

Modicon, *Ladder Logic Block Library User Guide,* 1994.

Modicon, *984 Controller Loadable Function Block Programming Guide,* February 1992.

Parr, E. A., *Programmable Controllers,* Butterworth-Heinemann, 1999.

Petruzelka, F. D., *Programmable Logic Controllers,* New York: Glencoe, 1998.

Simpson, C., *Programmable Logic Controllers,* Englewood Cliffs, NJ: Prentice Hall, 1994.

Webb, J. W. and Reis, R. A., *Programmable Logic Controllers Principles and Applications,* Englewood Cliffs, NJ: Prentice Hall, 1999.

5.6 PLC Software Advances

B. D. CAMPBELL, P. G. FRIEDMANN (1995) **C. W. WENDT** (2005)

Program Development:	Sequence definition thru flowcharts, state diagrams, or sequential function charts
PLC Programming Languages:	Graphical—Ladder logic and function block Text based—Instruction list and structured text Flowchart like—Sequential function charts
Language Standards:	IEC 61131-3 and IEC 848
Suppliers:	PLC manufacturers are listed in Section 5.4

INTRODUCTION

Programmable controllers were originally introduced to replace large panels of relays used for running automobile assembly lines. Controller flexibility allowed reprogramming for year-to-year model changes. Early controllers were restricted to discrete (ON/OFF) control. Programs were represented by the use of ladder diagrams (Section 5.2), the same way that relay logic is represented.

Programmable controller capability has been extended to include mathematical functions, file manipulations, analog signal manipulations, and techniques for dealing with high-speed signals. Programmable controllers are now routinely linked together in networks, connected to general-purpose computer systems, and connected to specialized intelligent devices, as was shown in Section 5.4. These devices include displays, high-speed control cards, and interfaces to specialized sensors such as bar code readers and RF identifiers. Many programmable controller suppliers support computer software for program development, program annotation, and program storage/retrieval.

These added capabilities have allowed the application of programmable controllers in many areas outside of the traditional automobile assembly-type service. They are used for continuous process control high-speed packaging, energy distribution, and automated warehousing. Many of these applications strain the capability of ladder logic for program design and implementation.

The software available in many programmable controllers has extended ladder logic to serve these diverse applications. To engineer a software system correctly, the application's requirements must be well understood. A graphic representation of these requirements is described below. Among the ladder logic advances are indirect addressing, program flow modification, incorporation of mathematical or scientific instructions, communication with intelligent devices, and methods for working with high-speed signals. Ladder logic segments can be organized using sequential flow chart languages to produce a sequentially oriented program.

Productivity enhancements developed for computer programmers are appearing in PLC programming software. Many program editors emulate MS Visual Studio formats with multiple windows and tool bars, pull-down menus, and visible "tree" structure for the project. Wizards have appeared in programming editors to automate tasks. If fact, Siemens provides one wizard in its S7 (200) programming software that automatically programs a Master-Slave polling task.

PLC logic simulation, a powerful tool for programming and debugging even before the hardware is on hand, is becoming more widely available.[1] Most major PLC manufacturers now offer PC-based PLC simulation software. Even the hardware configuration is easier. PLCs used to be configured with switches and jumpers. Now PLCs can be programmed and configured remotely over the network.

GRAPHIC DESCRIPTION OF CONTROL REQUIREMENTS

There have been many studies of time and money investments in building computer systems.[2,3] One key conclusion of these studies is that spending more time on the planning, design, and study of systems can have several benefits. One benefit is that the risk of project failure is reduced. Another benefit is that the resulting system is more likely to meet real needs and to provide continuing returns.

One of the effective ways of representing program requirements is through the combined use of state diagrams

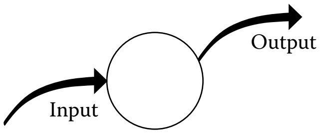
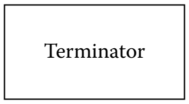
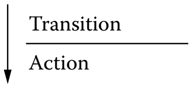
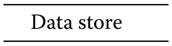
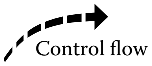
Symbol		*Explanation*
	Data flow	Represents data movement.
	Data transformation	Computes outputs from inputs.
	Terminator	Defines items outside of the context of the system.
	State	A user-identifiable system mode.
	State transition	A transition causes a state change while actions occur during a transition.
	Data store	Shows how data are archived.
	Control flow	Data triggering WHEN things happen.

FIG. 5.6a
Data flow diagram primitives (symbols).

and data flow diagrams.[4,5] These structured documents contain symbols for terminators, data transformations, data storage, triggers for control flows, and state diagrams. Figure 5.6a includes a list of these symbols (primitives). Control flows and transformations describe when actions occur. Data flows and data transformations describe what the system does. Either data transformations are further decomposed into other data transformations, or the input/output character of the transformation is specified.

The left side of Figure 5.6b shows a tank provided with a level transmitter. The sequence of tank filling and emptying are controlled by starting and stopping two pumps. If the level transmitter, pump motors, and display devices are connected to a programmable controller, the software shown on the right side of Figure 5.6b is needed to control the tank. This software must provide such actions as the automatic shutting off of the fill motor when the tank is full.

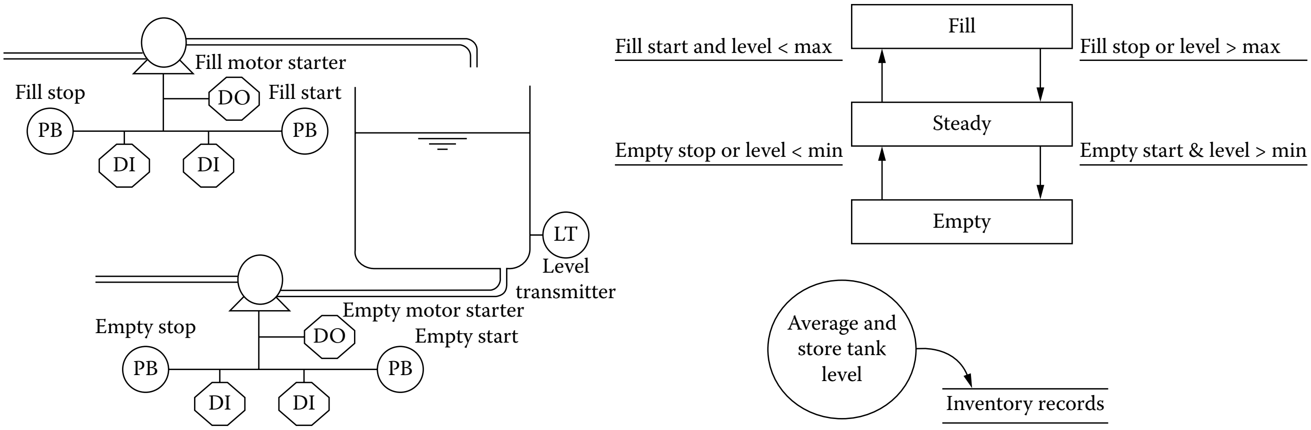


FIG. 5.6b
Tank level control example.

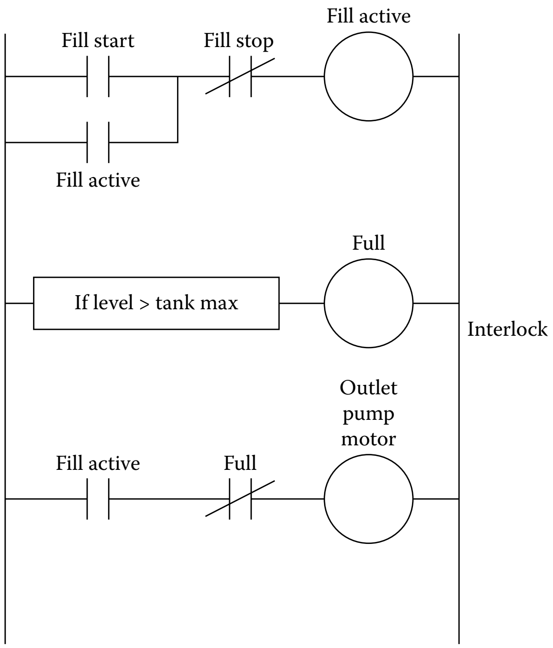


FIG. 5.6c
Portion of ladder logic code.

The programmable controller's ladder logic code for the fill motor starter controls is shown in Figure 5.6c. It shows how the start/stop station and the level interlock functions in order to guarantee that the tank does not overflow.

Some programmers advocate the use of state-machine representation as a planning tool and then implementing the program directly in ladder logic. Sequential function charts (SFCs) can be used the same way. The benefits are:

- Design decision are made before programming begins.
- Project engineers can review and make changes before programming.
- Stakeholders can more easily understand and review the state-machine representation.
- Once the diagram is finished, it is easy to program.
- The diagrams help document the ladder logic.[6]

Another programming technique is based on the cycles of a machine. The cycles for the machine are numbered in sequence and are called states. The set conditions of coils or a counter is used to indicate which step the machine is in. All the actions that take place in that particular step are executed and when completed, the condition is defined to move to the next state. In addition to the cycle states, an initial state and a default safe state should also be defined. If the machine gets stuck in a state because of a malfunction, this programming technique is easy to troubleshoot, because only the logic in that step needs to be examined.[7]

LADDER LOGIC ADVANCES

Section 5.2 discussed ladder diagrams and their generation of commands that energize, latch, and unlatch coils. Other ladder logic commands read live inputs, outputs, or internal bits and perform instructions based on their state. Timers and counters are other sources of ladder logic commands.

In most medium- and large-size controllers, these ladder logic-style commands are augmented by assembler-style instructions. Rather than working on bits alone, they operate on entire words. Some ladder logic extensions take on the character of high-level languages such as C or Basic.

The assembly language is a primitive computer language, which Bryan defines as "a symbolic programming language that can be directly translated into machine language instructions."[8] While a particular assembly language is dependent on the type of machine used, each of the ladder logic extensions has its basis in some type of assembler language. For a survey of the many ways of using assembler (and advanced ladder logic), see the books by Knuth.[9] Table 5.6d lists some typical assembler-style instructions along with their comparable ladder logic instructions.

Program Flow Modification

Some of the ladder logic extensions shown in Table 5.6d modify the program flow. Conventional relay logic commands execute sequentially. In this case there are no changes in program flow, because each instruction is executed in every program scan.

A GO TO command, when invoked, dictates that certain portions of the program will not be executed during the particular (current) scan. It allows the program flow to jump over the commands between the GO TO and the destination that is indicated in the GO TO.

While the GO TO instruction is frequently used in general programming practice, it can make programs difficult to understand and maintain. A better device that serves the same purpose is the subroutine. The subroutine has a beginning, usually marked by a subroutine number, and an end, noted by a RETURN. The subroutine is executed by a JSR (jump to subroutine) or similar command (GOSUB, for example). When this occurs, the program flow goes to the label indicated in the command and executes the instructions that follow.

Programs that consist of many subroutines are referred to as modular programs. After the RETURN is encountered at the end of the subroutine, program flow can return to the portion of the logic following the JSR.

The benefits of the subroutine are as follows:

1. The same code can be run for different purposes in different places of the program without recopying the code. Fewer lines of code mean less time to make modification and maintain the code.
2. The code can be run conditionally. This means that the scan time, or time it takes to run once through the system, can be lowered if portions of the code run conditionally.

TABLE 5.6d
Assembly Language

	Comparable Assembler-Style Instructions		
Instruction Category	*Instruction Name*	*Mnemonics in Allen-Bradley's PLC5 and Square D's 650*	*Definition*
Arithmetic Operators	Add	ADD	Add two registers
	Subtract	SUB	Subtract two registers
	Multiply	MUL	Multiply two registers
	Divide	DIV	Divide one register into another
	Average	AVE	Average contiguous registers
	Negate	NEG	Take opposite sign of source and store in destination
Compare	Equal	EQU	Set bit if source A equals source B
	Greater than or equal	GEQ	Set bit if source A is greater than or equal to source B
	Greater than	GRT	Set bit if source A is greater than source B
	Less than or equal	LEQ	Set bit if source A is less than or equal to source B
	Less than	LES	Set bit if source A is less than or equal to source B
	Not equal	NEQ	Set bit if source A is not equal to source B
	Clear	CLR	Set destination to zero
Logical Operators	And	AND	Set a bit only if two bits are TRUE
	Or	OR	Set a bit if either of two bits is TRUE
	Excusive Or	XOR	Clear a bit if two bits are the same, otherwise set the bit
	Not	NOT	Reverse state of bit
Register Assignment	Move	MOV	Place a bit from one register into another
	Masked move	MVM	Place a bit from one register into another if set in mask
	BCD convert	BCD* convert	Convert a register from BCD into integer
		Indirect read	Read a word, given its address
Shift Register	Bit shift left	BSL	Shift bits in work left
	Bit shift right	BSR	Shift bits in work right
	First in, first out load	FFL	Add an element to a waiting line
	First in, first out unload	FFU	Remove an element from a waiting line
Program Control	Jump	JMP	Skip rungs until the label is found
	Jump to subroutine	JSR	Execute subroutine
	Return from subroutine	RET	Exit subroutine and return to normal ladder execution

*BCD = binary-coded decimal.

3. Details can be hidden. Breaking up the overall program into modules affords the programmer an opportunity to hide details inside of subroutines. Thus, someone other than the original author can follow the overall program better by understanding what each module of code does, without examining module internals.[10]

As an example, it would be possible to code the previously described tank farm example in four separate subroutines. All of the tank logic for a single tank can be written in a single subroutine. The logic that sets up the tank logic by reading the appropriate signals can be written in another subroutine. A third subroutine might communicate with other equipment after the tank logic has run. An executive routine can be used to call on each individual subroutine when they need to be called. This configuration is shown schematically in Figure 5.6e. The executive, in this case, contains only calls to the individual subroutines.

Indirect Addressing

One way to extend the tank-filling example across a tank farm would be to physically add sections of code for each tank. With a large number of tanks in a tank farm, a large number of rungs (lines in a ladder diagram) would be needed.

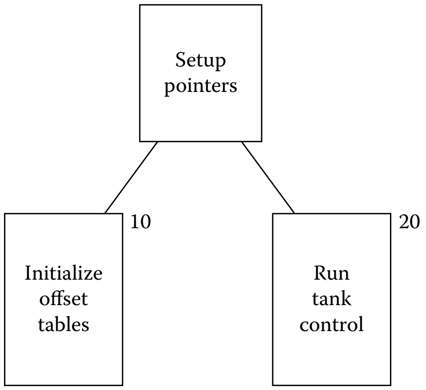


FIG. 5.6e
Tank farm structure.

The larger the number of rungs, the larger the chances are that one of them could contain an error.

Indirect addressing, which is new to most users of programmable controllers, is an effective way of simplifying and reducing the size of large programs. If, for instance, there are 200 tanks in a tank farm and to code one tank requires about 200 rungs, the resulting 40,000-rung program is difficult to maintain. With indirect addressing, if only one rung needs to be changed for each tank, there is only one rung to change in the total system. Without it, modifying the control would require modifications in 200 rungs.

While there are variations in the capabilities of programmable controllers in regard to indirect addressing, by using assembler commands, indirect addressing is usually possible. If one desires to extend the code developed for one tank to work for an entire tank farm, the resulting requirement might look like the diagram in Figure 5.6f.

The code reads the tank in progress, reads the offsets needed to run the code, reads the required information, triggers the code to operate on this information, and then writes out the result and increments to the next tank on the available tank list. Figure 5.6g shows a listing of the necessary codes. Note that this code is “data driven.” By making files larger (B9 in Figure 5.6g is a file) and modifying their contents, indirect addressing provides a means for using fewer instructions than physical coding would use.

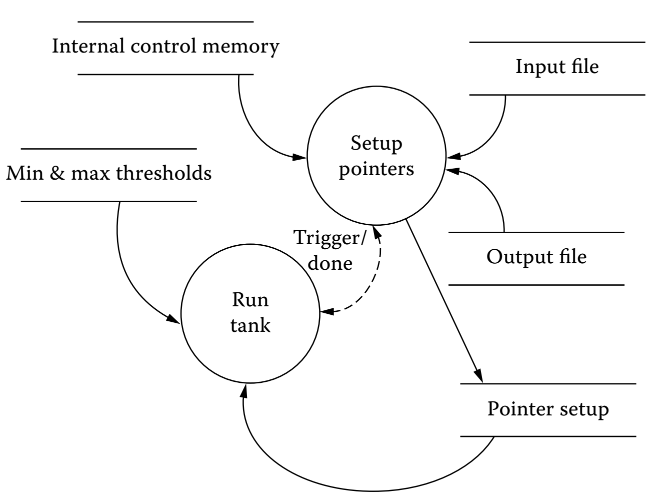


FIG. 5.6f
Partial tank farm requirement.

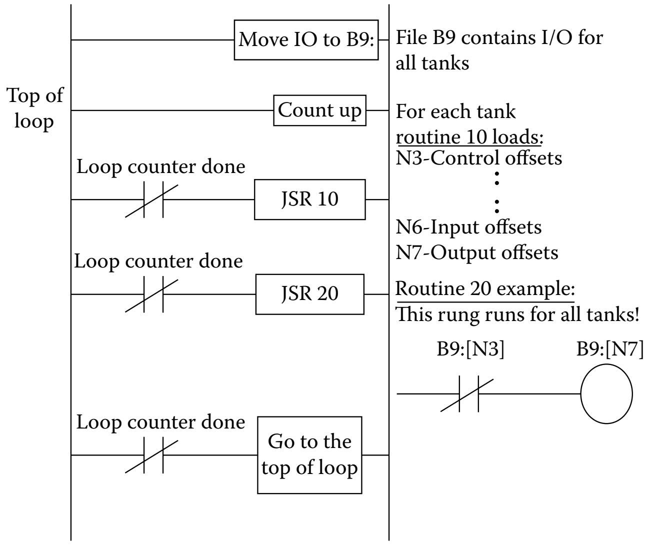


FIG. 5.6g
Pointer setup and indirect addressing.

Figure 5.6g shows the operation of indirect addressing. On the first pass, the value of the counter is 1. Then the code (JSR 10) runs, jumping to the subroutine 10. In this subroutine, integers are placed in registers N3, N4, N5, N6, and N7. Although the details of this subroutine are not shown, there could be a table for each of the registers, and the value of the counter would tell the program which element should be moved from the table to the register.

Since the counter is at 1, the values associated with tank 1 are loaded into the registers. When subroutine 20 is run, it operates the controls for tank 1. For example, if the register that contains tank 1’s inputs is B9:45, then B9:[N3] will resolve to B9:45. Therefore, the same rung can be used for different tanks. When the counter is 2, subroutine 20 will load a different value in register *N*3, so that the rung will inspect a different input.

Trade-Offs of Indirect Addressing Indirect addressing can help reduce the time involved in writing the ladder logic. However, it is more difficult to understand and maintain it than to maintain regular ladder logic. The benefits of indirect addressing must be weighed against some major trade-offs, including:

- Execution times are slower.
- Troubleshooting becomes more difficult.
- Accommodating differences in the group can be difficult.

In time-critical applications, indirect addressing will require more processor time than would physical addressing. Thus, if, for instance, one unit's code takes 20 ms to run, the indirect code for 100 units will take much more than the 2 seconds that would be required to run the code, if done physically. Instead of executing all the codes in a single scan, a counter can be set up to run one or more tanks per scan. In this case, it will still take more than 2 seconds to update all the tanks, but other parts of the system can run normally, and watchdog timeouts can be avoided.

The second trade-off is in using a programming terminal to view the operation of the logic. When using direct I/O, either the ladder logic displays or the display of the data table can be used. With indirect addressing, the state of the logic can only be determined by studying the data table display. This makes troubleshooting difficult.

A third trade-off is really an assumption that all the tanks are the same and will stay the same. It can become burdensome to write and troubleshoot the special logic to accommodate several small differences in the 100 tanks. In addition, other programming and documentation tools (like cross-referencing) may not work as expected, either.

Assembly Language-Like Extensions

Commands shown in Table 5.6d can perform data manipulation, including arithmetic operators, Boolean operations, and register assignment. One way that these are used is in a counting application. If, for instance, one needs to deliver a number of packages to an operator, a register can be assigned to contain the count.

When operators want the programmable controller to read a thumbwheel, they can depress a button. A high state on the button causes the programmable controller to read the thumbwheel setting in binary-coded decimal and place that setting in its count register. Sensors are used to detect when individual packages are delivered to the operator. A change in state in these sensors causes the count to decrement.

A ladder logic program performing these functions is shown in Figure 5.6h. The first rung moves information from the thumbwheel 10:1 to an internal word, $B1:1$. This word represents the required number of packages. The next decrements this number once for every FALSE to TRUE transition of the photoeye.

Commands is another group shown in Table 5.6d to perform queue management. Queue management commands include LIFO (last in, first out) and FIFO (first in, first out) commands, which load and unload the queue. A queue in a programmable controller is usually a series of words that are loaded into contiguous registers to form an array. A queue management command will insert or extract a word from the array.

One use for the FIFO queue is in tracking packages on a conveyor belt. Every package that enters the belt loads the FIFO and every one that exits unloads the FIFO. The word contains, for example, a package identification number. Package sensing can be accomplished with photoelectric devices.

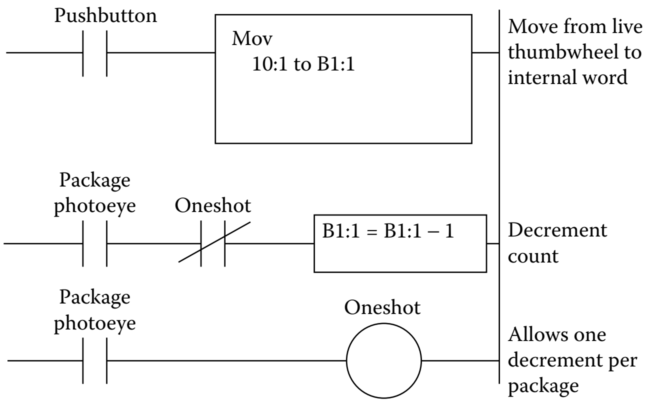


FIG. 5.6h
Ladder logic for package counting.

Communication with Intelligent Devices

Often, there are microprocessor devices that communicate with the programmable controller. Some newer devices do this by mimicking racks and by being scanned as a rack would be scanned. While this scheme affords a simple interface to these devices, it decreases the number of racks with which the programmable controller may connect.

Often these devices physically plug in like an I/O card and communicate using some type of serial communication. In fact, one common type of card is the ASCII card, a card that is able to communicate using RS-422 or RS-232. As long as the supplier is specific about the format of the serial communication, the user can program these cards with ladder logic and jumpers on the card to communicate with any device using this form of serial communication.

One common feature of these applications is that they require a ladder logic program that reads and writes a block of words, often performed continuously. One device that uses this type of programming and is common to many programmable controller vendors is the counter card. Counter cards count pulses that might occur too quickly for a programmable controller to handle. The words communicated to and from the counter cards might include a setpoint, an accumulated count, and a word containing information about the state of the communication itself. The way that this block of words is read from the external card and written to the external card is illustrated in Figure 5.6i.

In Figure 5.6i, "BTR" refers to the statement in Allen-Bradley's PLC-5 called a block transfer read. The rest of the items on this instruction indicate how the read will be done, including the PLC memory to be used, the location of the physical hardware, and the amount of data in the transmission. A similar statement is available in other programmable controllers that provide serial communication with intelligent devices.

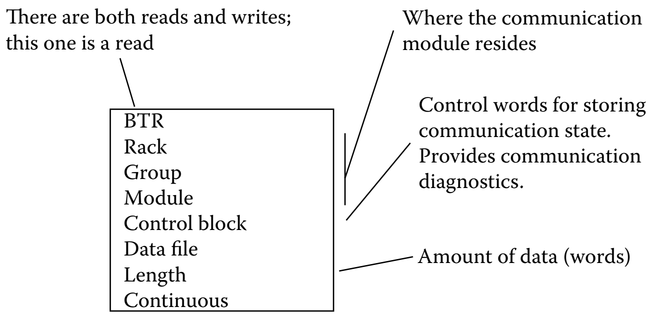


FIG. 5.6i
The block transfer statement.

Fast I/O Updating Methods

In many medium- to large-size programmable controllers there are numerous ways to read signals that occur faster than the scan time of the programmable controller. Conversely, outputs must sometimes be updated quickly, without the error associated with a variable scan time. The four most common methods for fast I/O updating are time-scheduled interrupts, programmable interrupts, latching hardware, and independent (dedicated) programmable controllers.

If, for instance, one desires to capture a pulse of 20 ms duration in a controller with a program scan time of 100 ms, there is a good chance that the program that serves to count the pulses is going to miss some.

One solution is the use of a concurrent time-scheduled routine. The routine contains a counter that reads the pulse input in question. By scheduling this counter to run once every 15 ms, the program would be guaranteed to read the 20 ms pulse, yet it would count each pulse only once, on the leading edge of the pulse. Several suppliers provide this type of routine. Code for such a routine is loaded in some special area of the memory, and the user must provide a time interval for running the concurrent routine.

In many programmable controllers, there is an immediate update input instruction, or a "programmable interrupt." When the instruction is encountered, the controller stops scanning the program, determines the current state of the particular (interrupting) input, and then continues on with the program. Several suppliers provide this type of interrupt on their inputs or outputs.

As another option for reading quick, discrete signals, suppliers provide a special-purpose card for reading pulses, usually called a latching card. It is a useful approach for counting pulses only if the time between pulses is longer than the maximum program scan time. After using the "latched" input, the program needs to rest the card, making it ready for the next pulse.

A final updating method is the use of another programmable controller. A small controller that is fast enough to detect quick pulses can be programmed to act like a counter or a latching card. Some manufacturers supply small controllers to act as cards for the programmable controller rack or with convenient serial communication for communication to other controllers.

GRAPHIC, FLOWCHART-LIKE LANGUAGES

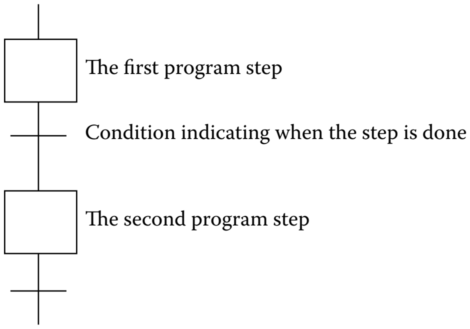
Graphic languages are natural in today's computer environments (e.g., Windows). Since the early 1980s, there have been several versions of graphic, flowchart-like languages, including sequential function charts.

Their goal is to provide a maintenance-related programming language for sequences of operations. Because of this, they are often not very adaptable to continuous process control. As the origin of ladder logic is in relays, it is a language well adapted to describing a large number of relays. By their nature, these devices operate in a nonsequential, highly concurrent manner. Only with great difficulty is a sequence of control possible. These sequences are usually tracked by a counter or by latching relays. The order of the sequence is usually difficult, if not impossible, to quickly determine from the program.

Several attempts have been made to build a language in which sequential control is easily coded. There is an international standard for a graphic programming language.[11] In this language, there are several basic symbol blocks (primitives). They are:

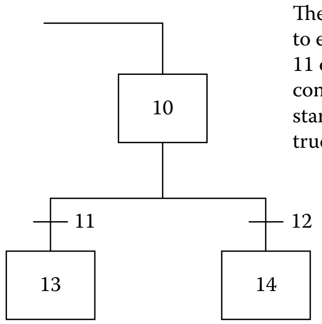
- Steps
- Transitions
- Directed links
- Branching sequences
- Simultaneous sequences

Steps are small programs. The programs might be written in any language, but many are ladder logic–driven. The transitions initiate a move from one step to another. When a step is connected to a transition, steps can be made to occur sequentially, as shown in Figure 5.6j. The direct sequential connection of steps with a transition is known as a directed link.



Note: The detailed description of each step and transfer condition is a separate program listing.

FIG. 5.6j
The use of directed links.

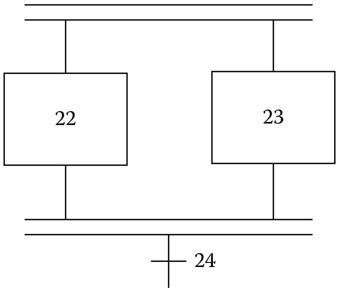


The process will branch from step 10 to either step 13 or step 14. If condition 11 comes true, step 13 will start. If condition 12 comes true, step 14 will start. Condition 11 and 12 cannot be true simultaneously.

FIG. 5.6k
Branching steps.

Steps can also branch, depending on some other condition, allowing alternate sequences to occur. In the example shown in Figure 5.6k, conditions 11 and 12 control the program flow, allowing execution of step 13 or 14. It is possible to program concurrent or simultaneous steps. This means that during a program scan, the code associated with both of the concurrent steps is executed.

In the example shown in Figure 5.6l, so long as exit condition 24 is not true, steps 22 and 23 will be executed during each program scan. If an exit condition becomes true, its associated step will no longer be executed.



In this instance, both steps 22 and 23 run until condition 24 indicates completion.

FIG. 5.6l
Concurrent steps.

An example of using concurrent steps is shown in Figure 5.6m. In this example, each assembled product needs to have a cover placed on it and four screws fastened on the cover. One solution is to design a program for each screw, to be run concurrently. Each program turns a screwdriver driving a screw. When a screw has been inserted completely, as indicated by screwdriver position, that screwdriver is retracted and its program is no longer executed. The flowchart for this example, Figure 5.6m, illustrates the application power of this type of controller programming.

IEC 61131-3 PLC LANGUAGE STANDARD

Programming a PLC is one of the most labor-intensive tasks in assembling a control system. It therefore makes good sense to try to reuse as much of the program as possible, once it is debugged.[12] A uniform programming environment for all manufacturers aids reusability, limits errors, and increases programmer efficiency. Developing the code only once and reusing it in many recurring control process scenarios is highly desirable, and attempts to standardize PLC language are in progress (see Table 5.6n).

The International Electrotechnical Commission (IEC) introduced IEC 1131-3 in 1992[13] and shortly thereafter renamed it IEC 61131-3. Since then, it has become widely accepted by the international user and vendor community. There are eight parts to the IEC 61131-3 standard for programmable controllers:

1. General information
2. Equipment requirements and tests
3. Programming languages

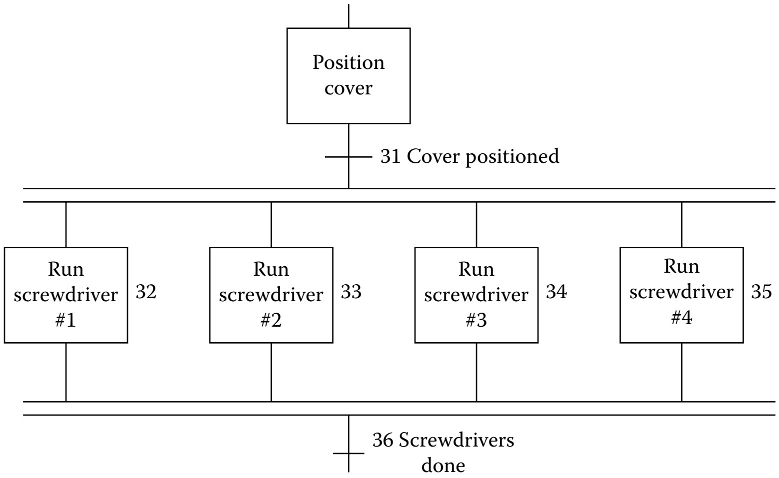


FIG. 5.6m
An example of using concurrent steps in a PLC program.

TABLE 5.6n
Work on a PLC International Language Standard Has Been Ongoing Since 1979[13]

1970	NEMA Programmable Controllers Committee formed
1977	Grafcet (France) DIN 40 719, function charts (Germany)
1978	NEMA ICS-3-304, programmable controllers (U.S.)
1979	IEC SC65A/WG6 formed
1980	DIN 19 239, programmable controller (Germany)
1983	IEC 65A (Sec) 38, programmable controllers MIL-STD-1815 Ada (U.S.)
1985	IEC SC65A (Sec) 49, PC languages
1987	IEC SC65A (Sec) 67 IEC 848, function charts
1988	IEC 65A (Sec) 90
1992	IEC 1131-3

4. User guidelines
5. Messaging service specification
6. Reserved for future use
7. Fuzzy control programming
8. Guidelines for the application and implementation of programming languages

Programming languages (part 3) define, as a minimum set, the basic programming elements, syntactic and semantic rules for the most commonly used programming languages. It also defines major fields of application, applicable tests, and means by which manufacturers may expand or adapt those basic sets to their own programmable controller implementations.[14] The IEC 61131-3 standard defines the following languages:

- Sequential function chart
- Ladder logic diagram
- Function block diagram
- Instruction list
- Structured text

Sequential function chart is a flowchart-like organizational construct based on the French Graphcet (IEC 848). Actual programming for steps and transitions is programmed in the other four languages.

The two graphical languages are ladder logic and function block. Ladder logic was invented in the United States and is still the most widely used there. Function block is popular in Europe and is widely used in DCS systems. It is also well suited to controlling batch operations.

The two text languages are the instruction list and structured text languages. Instruction list is an assembler-type language that was widely used in Germany. Structured text is the only language that was invented for the standard. It has a high-level syntax that resembles Basic. It is unique in that it can combine the SFC in text form and incorporate the control in the same language.

In actual application, rifts exist between standardized versus proprietary versions. Each vendor's implementation of the standard is a little different. This inevitably causes some rewriting of code when moving programs between manufacturers. The degree that reuse is possible depends on how well each vendor complies with the standard.

Also, high-level functions like PIDs are not addressed within the standard, so each manufacturer puts together its own version.[15] To address this, PLC Open was formed in 1992 to promote the adaptation of IEC 61131-3 and to minimize the number of variant adaptations of the standard. It also provides independent certification of conformance.[16]

End users and system integrators want to be able to move easily between different brands and types of PLCs. Software that can configure all of the PLCs of at least one manufacturer and be similar to other manufacturers' programming software would be a step in the right direction.[17]

In the past, sequential and continuous processes were often addressed separately, with PLC used for the sequential and DCS for the continuous process and using their respective languages of ladder diagrams for PLCs and function blocks for DCS. One of the benefits of IEC 61131-3 is that PLC users can use ladders and DCS veterans can use function blocks in the same PLC.

CONCLUSION

Modern PLC software advances are increasing the productivity of PLC programmers. PLCs have matured from simple relay replacement to sophisticated controls involving interfacing to PCs and intelligent controllers. International standards like IEC 61131-3 have produced a unified, manufacturer-independent programming environment.

Graphical languages (ladder logic and function block), text languages (instruction list and structured text), and organizational constructs (sequential function charts) support a wide variety of applications. The ability to mix these languages within a single PLC has also fueled software development. Modern software productivity tools (including wizards) and simulation software have made PLC programming and testing considerably easier than it was just ten years ago.

References

1. Mintchell, G. A., "Toolbox Overflows for Control Programmers," *Control Engineering,* September 2001.
2. Boehm, B., *Software Engineering Economics*, New York: McGraw-Hill, 1981.
3. DeMarco, T., *Controlling Software Projects: Management, Measurement and Estimates,* Englewood Cliffs, NJ: Yourdon Press, 1986.
4. Ward, P. T. and Mellor, S. J., *Structured Development for Real-Time Systems,* Englewood Cliffs, NJ: Yourdon Press, 1991.
5. DeMarco, T., *Structured Analysis and System Specification,* Englewood Cliffs, NJ: Yourdon Press, 1979.

6. Ponizil, D., "Back to Basics: The Essentials of Structured PLC," *Control Engineering,* September 2001.
7. Fairey, W., "A Standard PLC Programming Method for Small Manufacturers," *Programmable Controls,* March/April 1991.
8. Bryan, L. A. and Bryan, E. A., *Programmable Controllers: Selected Applications,* Chicago: Industrial Text Co., 1987.
9. Knuth, D., *The Art of Computer Programming:* Vol. 1, *Fundamental Algorithms,* 1981; Vol. 2, *Seminumerical Algorithms,* 1982; Vol. 3, *Sorting and Searching*, 1983; Reading, MA: Addison-Wesley.
10. Kernigham, B. E. and Plaugher, P. J., *Elements of Programming Style,* New York: McGraw-Hill, 1978.
11. International Electrotechnical Commission, Standard #848.
12. Laduzinsky, A. J., "An Open Look for PLC Programs," *Control Engineering,* May 1993.
13. Christensen, J. H., "Programmable Controller Users and Makers to Go Global with IEC 1131-3," *I&CS,* 1993.
14. "General Introduction into IEC 61131: All Parts," PLC Open Web site, June 2003, www.PLCopen.org.
15. Greenfield, D., "Once Written, Often Used," *Control Engineering,* August 2003.
16. Babb, M., "IEC 1131-3: A Standard Programming Resource for PLCs," *Control Engineering,* February 1996.
17. Hebert, D., "PLCs in Harmony," *Control Magazine,* January 2001.

Bibliography

Berger, H., *Automating with STEP 7 in STL and SCL: SIMATIC S7-300/400 Programmable Controllers,* 2nd Book and CD-ROM edition, Hoboken, NJ: John Wiley & Sons, 2005.

Bergwall Productions Inc., *Statement List Programming PLCs,* Albany, NY: Delmar Publishers, 1992.

Boehm, B. W., *Software Risk Management,* Los Alamitos, CA: IEEE (Computer Science), 1990.

Bryan, L. A. and Bryan, E. A., *Programmable Controllers: Selected Applications,* Chicago: Industrial Text Co., 1987.

Bryan, L. A. and Bryan, E. A., *Programmable Controllers: Theory and Implementation,* Chicago: Industrial Text Co., 1988.

DeMarco, T., *Controlling Software Projects: Management, Measurement and Estimates,* Englewoods Cliffs, NJ: Yourdon Press, 1986.

DeMarco, T., *Structured Analysis and System Specification,* Yourdon Press, 1979.

Fleming, D. W. and Pillai, V. A., *S88 Implementation Guide,* New York: McGraw-Hill Professional, 1998.

Girard, B., *Programmable Logic Controllers: Architecture and Application,* New York: John Wiley & Sons, 1990.

Kernigham, B. E. and Plaugher, P. J., *Elements of Programming Style,* New York: McGraw-Hill, 1978.

Knuth, D., *The Art of Computer Programming, Vol. 1: Fundamental Algorithms,* Reading, MA: Addison-Wesley, 1981.

Knuth, D., *The Art of Computer Programming, Vol. 2: Seminumerical Algorithms,* Reading, MA: Addison-Wesley, 1982.

Knuth, D., *The Art of Computer Programming, Vol. 3: Sorting and Searching Algorithms,* Reading, MA: Addison-Wesley, 1973.

Larson, K., "The Great DCS versus PLC Debate," *Control,* January 1993.

Lewis, R., *Programming Industrial Control Systems Using IEC 1131-3* (IEE Control Engineering Series , Vol. 50), Revised edition, Stevenage, UK: IEE Publishing, 1998.

Parr, E. A., *Programmable Controllers,* 2nd edition, Boston: Newnes, 1998.

Pereira, F., *Application Note: Sequential Function Chart Programming Concepts,* Allen-Bradley, 1985.

Perron, S., "New Generation Language for Control Design," *Control Engineering*, February 1992.

Petruzella, F. D., *Programmable Logic Controllers*, New York: McGraw-Hill, 1989.

PLC-5 Programming Software, Allen-Bradley, 1991.

Ready, L. A., "Programming PLCs with Sequential Logic," *Control Engineering,* November, 1991.

SYMAX 650 Programming Manual, Square D, 1991.

Ward, P. T. and Mellor, S. J., *Structured Development for Real-Time Systems,* Englewood Cliffs, NJ: Yourdon Press, 1991.

Warnock, I. G., *Programmable Controllers: Operation & Application,* Englewood Cliffs, NJ: Prentice Hall, 1989.

5.7 Practical Logic Design

A. TUCK (2005)

EXAMPLE

> An operator notices a demineralized water-forwarding pump rapidly cycling on and off during the regeneration cycle of a demineralized water system. The plant has been in service for years. "Why," the operator wonders, "after all this time, has the pump logic stopped working properly? Who changed the logic?"

The logic has not changed at all but contains an error that appears only under a certain combination of conditions—in this case, when the supply tank level is hovering around the low-low set point and the demineralized water system requires the pump to run.

Many simple logic schemes that were tested and thought to work properly have subtle errors that are discovered only after they are placed in service. It is easy to test for the things the logic *should* do, but it is much harder to test for the things it *should not* do.

To help avoid such problems when designing logic, we have provided standard logic diagrams and solutions for control in this section. We have also provided information, compiled from years of correcting logic in various plants, that will help avoid the most common problems found in logic. This information will be especially useful while redesigning logic during commissioning.

For an explanation of the symbols used for the logic elements in this section, refer to Figure 5.7a

DESIGN PHILOSOPHY

Make It Simple. Somehow, the simplest and most direct way to solve a problem usually works best. When correcting logic that does not work properly, the best way to solve the problem often involves removing extraneous logic rather than adding more. Logic can be made to do almost everything anyone can dream up, provided there is enough input/output (I/O), but such lengths usually are not necessary. Find a simple way of solving the problem and do that instead.

The purpose of the logic is to keep systems controlled at steady state and to protect the equipment—it works best when it does not attempt to make all of the decisions for the operator. It is better for the operator to understand what the logic is doing and to stay involved in the operation of the plant. If too much is controlled *for* the operators, they will not be clear about where the logic stops and their jobs begin, and they are more likely to miss something that needs their attention. The only safe alternative would be to have the logic control everything and replace the operator entirely, and that usually is not practical.

Be Consistent. Writing logic that is consistent throughout the plant is fundamental to good design, and is important for safety as well. All valves and pumps of the same type should work the same way so that the operator knows what to expect when a button is pushed. Keep in mind all of the operators who will be hired in the years to come—they will not have the advantage of learning all the quirks of the plant during its start-up.

OPEN/CLOSE VALVES

> A valve opens automatically upon a steam turbine trip. While the turbine is down, the operator tries to close the valve; however, the steam turbine trip indication is still in place and locks the valve open. The operator then puts the valve in manual mode and closes it. The steam turbine is started again a few days later, and by this time the operator has forgotten about the valve and does not return it to auto mode. When the steam turbine trips again, the valve does not open.

Many would call this operator error, but it also is a design flaw that makes the plant hard to control. There are many different ways to design open/close valve logic, and it is not always clear how difficult it is to control until the plant is up and running.

Following are some tips for creating safe and easy-to-control logic, and some diagrams of standard logic for motor-operated valves (MOV) and solenoid valves.

Definitions

Input to the logic:

Open and close command: The commands from the graphic display, initiated by operator, to open and close the valve. Note that these are always momentary.

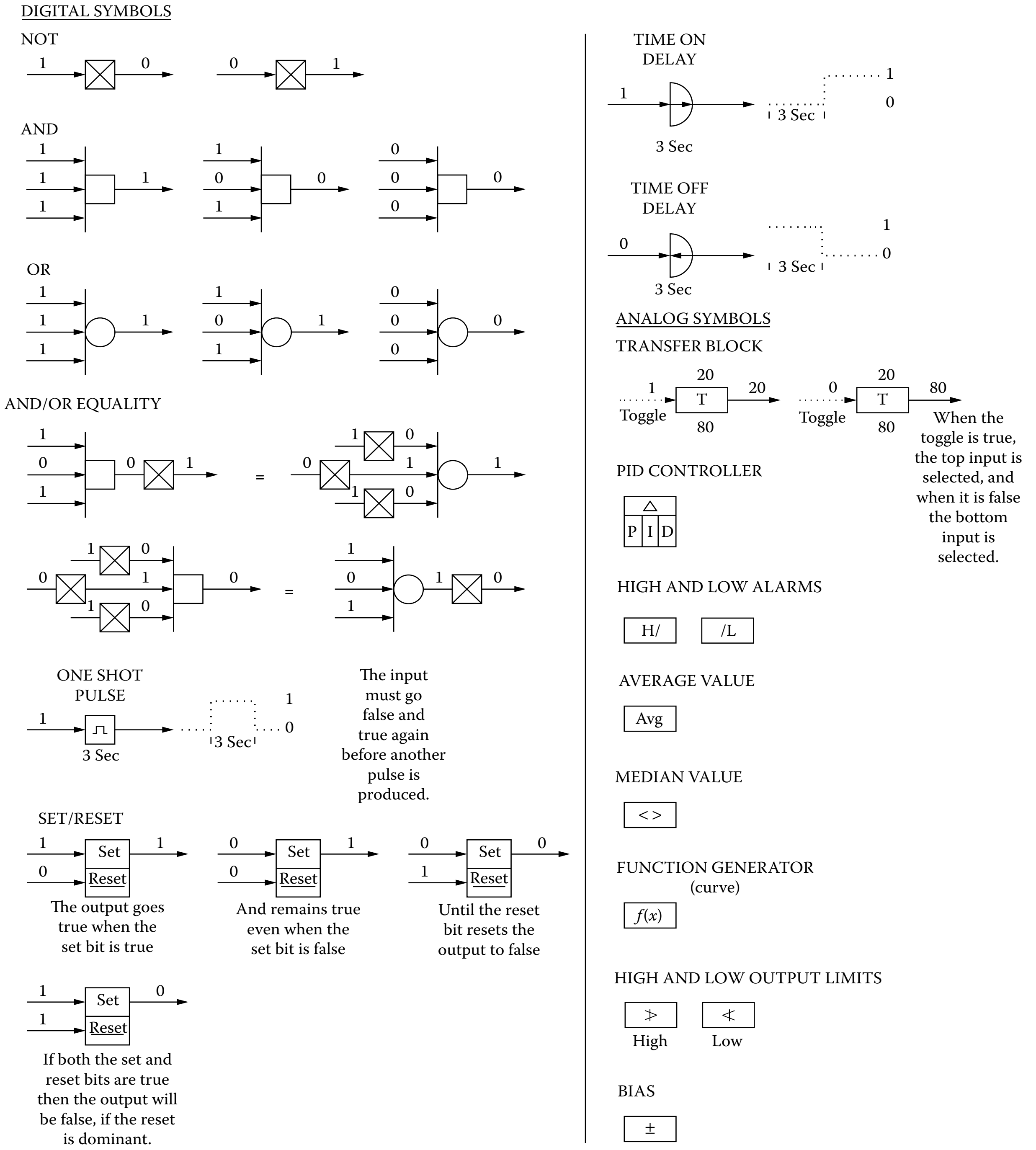


FIG. 5.7a
Logic design symbols.

Open and close limits: The indication from the valve limit switches that tells the operator and the logic when the valve is fully open or fully closed.

Open and close permissive: Process conditions that are required before the valve can be opened or closed.

Auto open or close: Conditions that open and close the valve automatically while the valve is in auto mode.

Output from the logic:

Open and close outputs: The output from the logic to the valve that strokes it open and closed.

Indications and alarms:

Auto mode: Shows that the valve is in auto mode and will be operated automatically by the **auto open** and **auto close** commands.

Failed to open/failed to close: An alarm that shows that the valve was given an **open** or **close** output, but the fully open or fully closed limit was not reached in a specified amount of time.

Mismatch: An alarm that shows that both the open and close limit are true at the same time.

Auto Mode

One common approach to designing auto/manual logic is to design the logic to automatically switch the valve to manual mode upon an open or close command from the operator. Unfortunately, that makes it even easier for the operator to forget to place the valve in auto mode, because it can go unnoticed that the valve switched to manual mode. Logic designers also often elect not to use an auto mode at all to avoid the risk of the operator's forgetting to put the valve into auto mode; however, this makes the operator's job quite difficult since he or she will need to stroke the valve for maintenance when the system is not in service and will have to force the auto commands to do that.

A better way to handle this is not to require that the valve be in manual for the operator to give a command and to reserve manual mode for locking out the auto commands when the system is not in service—this can be done by simply making all auto open and auto close commands momentary as shown in the standard logic for valves in Figures 5.7e, 5.7f, and 5.7g, later in this section.

The operator in the example above would need only to give the valve a close command; then, when the steam turbine trips again, the valve will open as it should.

By making the auto command momentary, one need not worry about conflicting auto open and auto close commands that are True at the same time or about having a valve that strokes in the direction of the auto command as soon as it is placed into auto mode (the operator should always know what to expect when placing a valve into auto mode). Also, if the auto commands are latched, then they will lock the reset of the fail alarm.

If an auto command should hold the valve in place while it is True, then the command should also be a permissive. For example, an auto open command that should hold a valve open should also be inverted and fed into the close permissive. This way, when the auto open condition is True, the valve will open and cannot be closed even when it is put into manual mode (Figure 5.7b).

If there is more than one auto open or auto close command, use a pulse for each individual auto command as shown in Figure 5.7c; otherwise, one of the commands may

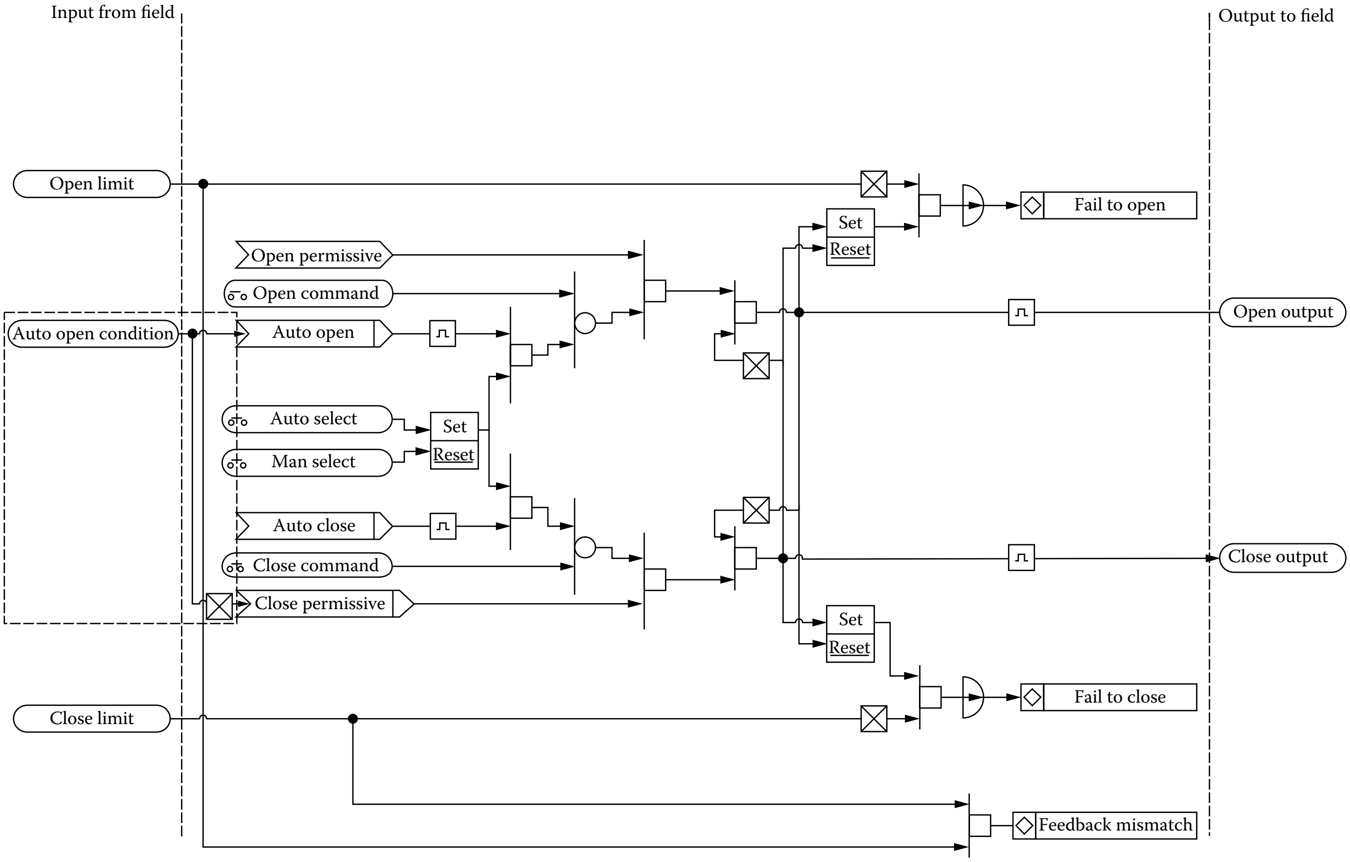


FIG. 5.7b
Open/close valve—use permissive to hold auto command.

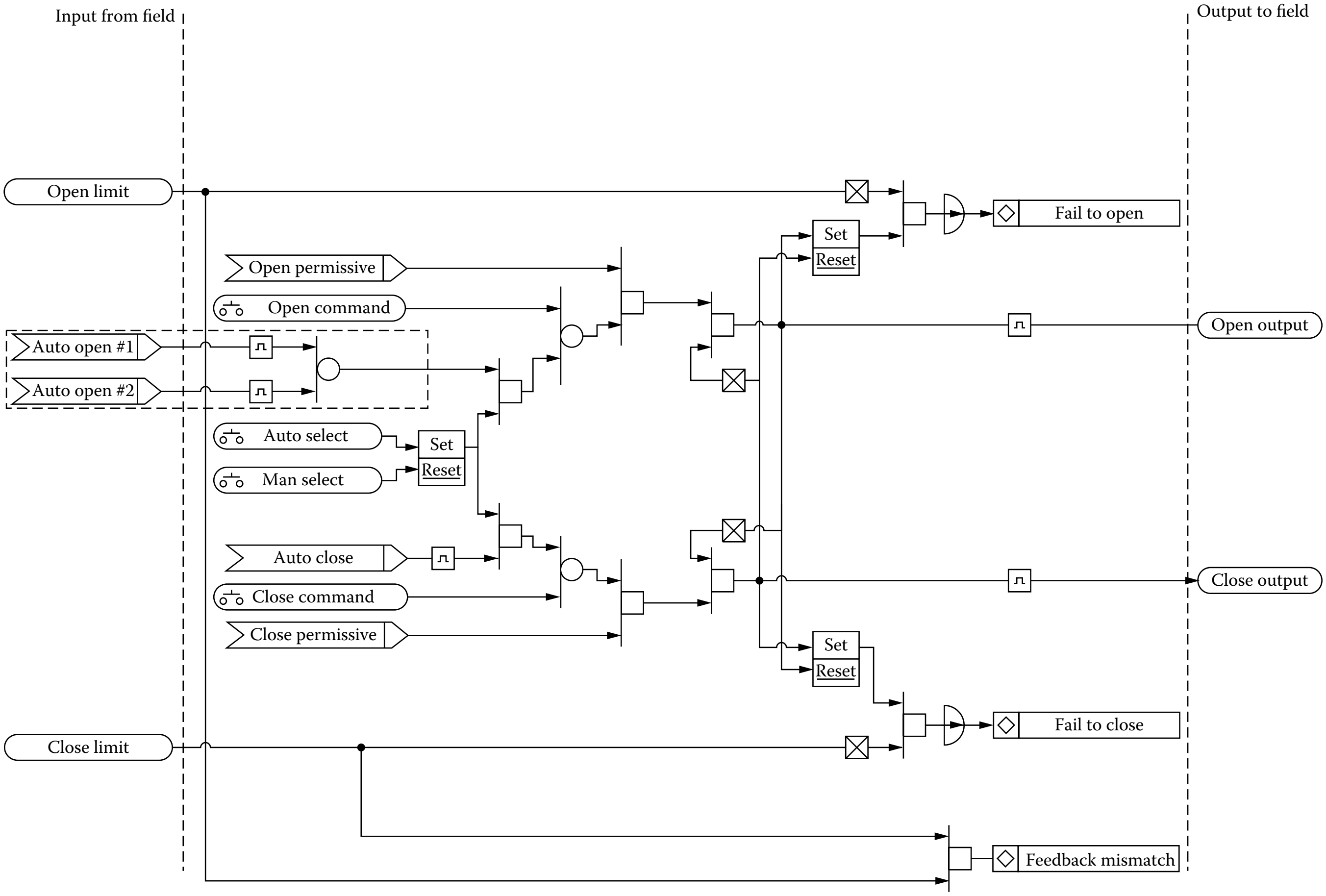


FIG. 5.7c
Open/close valve—more than one auto command.

be ignored. For example, if there are two auto open commands, and one of them opens the valve and remains True, and then an auto close command shuts the valve, when the other auto open command becomes True, the valve will not reopen because the pulse will not have occurred since the first auto open command is still True.

With motor-operated valves in particular, it is important not to write logic that will continually cycle the valve open and closed as this will eventually destroy the motor. Take the time to analyze the logic to make sure the auto open and auto close commands will not cause the valve to cycle.

For example, it is best not to use the same input for both the auto open and (inverted) auto close commands, which can cause the valve to cycle open and closed. If it is necessary, because of hardware restrictions, be sure to add a hefty time delay on either auto mode input, as shown in Figure 5.7d.

Motor-Operated Valves

The standard logic for two of the most common types of motor operated valve (MOV) logic is shown in Figures 5.7e and f. The logic in Figure 5.7e has momentary open and close outputs that will command the MOV to move. The circuit for the MOV will then take over and continue to stroke the valve and will stop stroking when the limit is reached. The logic in Figure 5.7f sends latched outputs to the field that stay True until the limits are reached.

Failure Logic

An operator gives an open command to an MOV on a main steam line. The valve opens, releasing steam to the system, but does not quite reach the open limit. The maintenance engineer who has been troubleshooting the valve thinks he has found the problem and asks the operator to give another open command to the valve to open it fully. Unfortunately, the operator can only give an open command after first closing the valve, which would bottle up the steam and cause the relief valves to open.

The operator should be able to issue a second command to attempt to move a valve that has failed. It is a common problem to leave this option out of the logic, which makes it very difficult to remedy the problem of a failed valve while a system is in service.

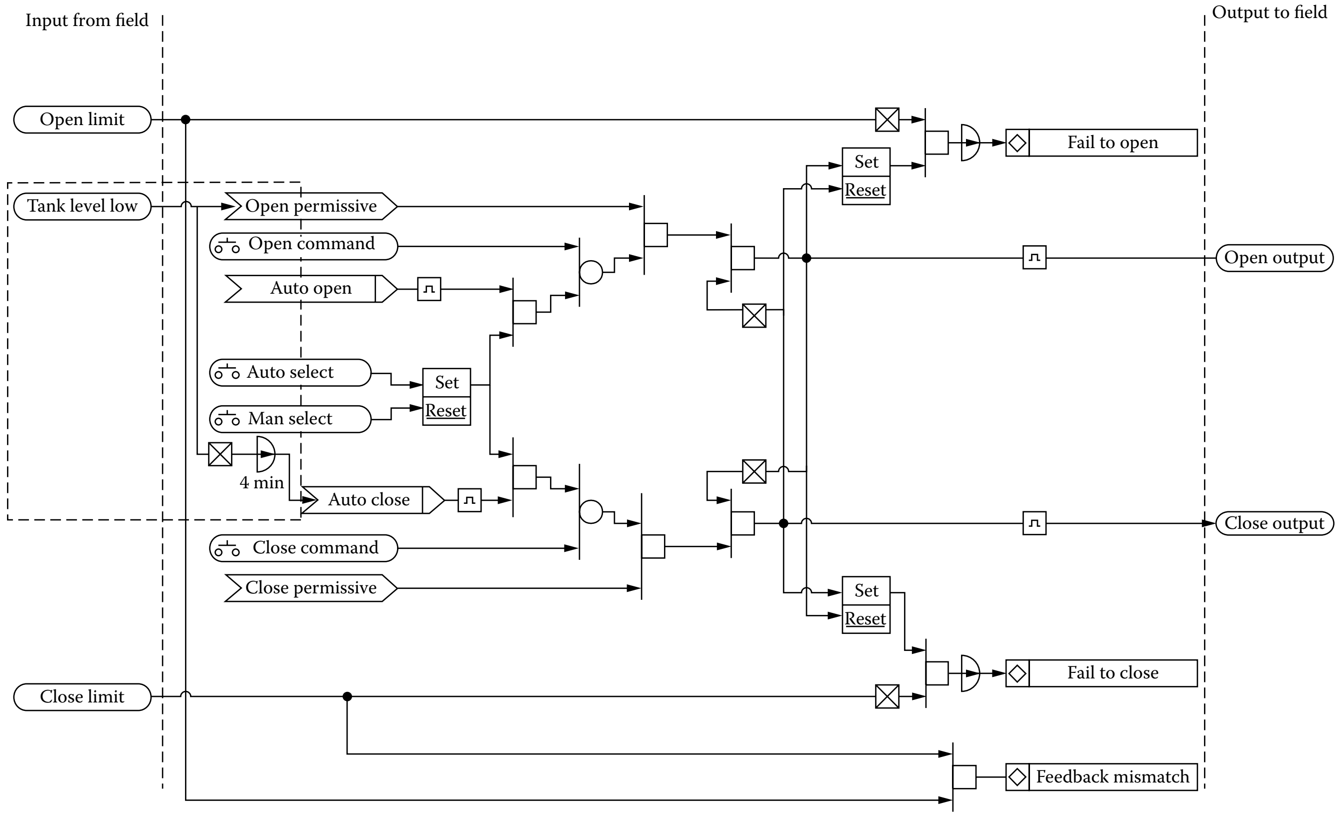


FIG. 5.7d
Open/close valves—auto open and close from the same input.

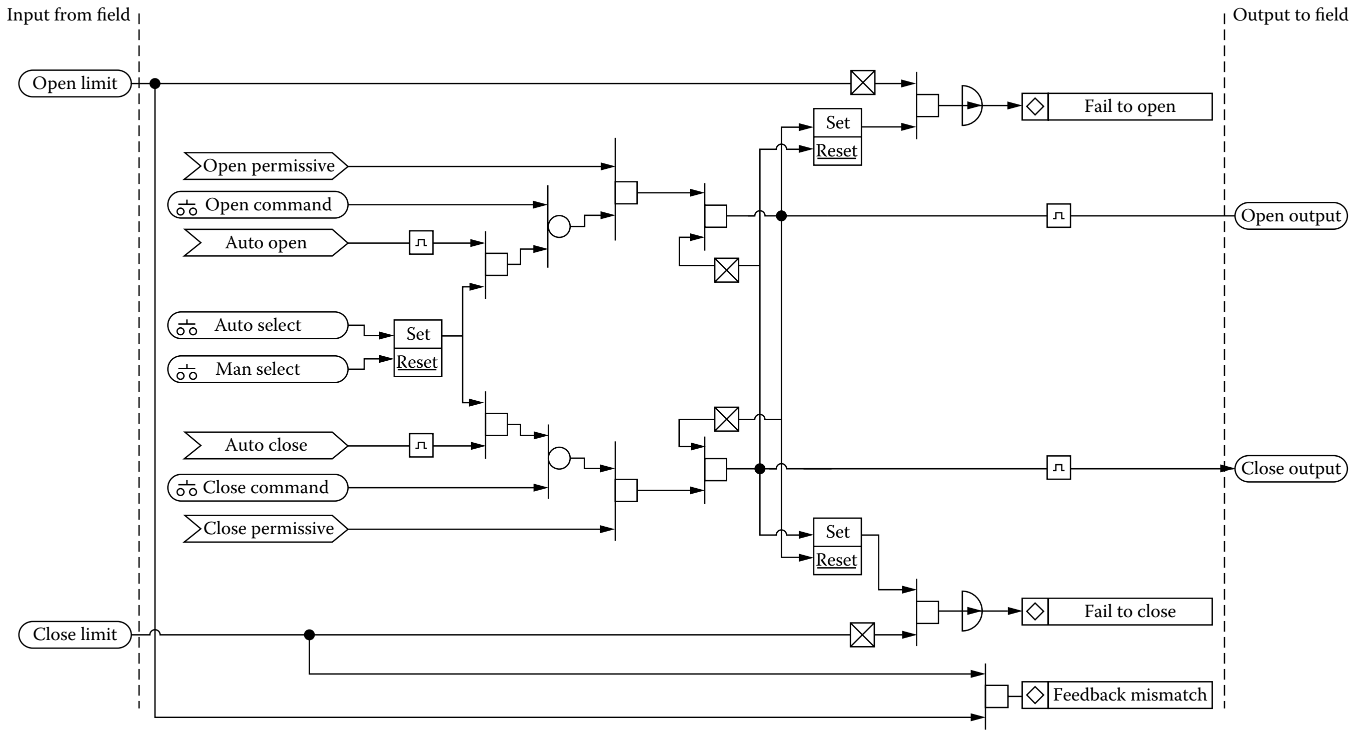


FIG. 5.7e
Standard MOV logic—momentary outputs.

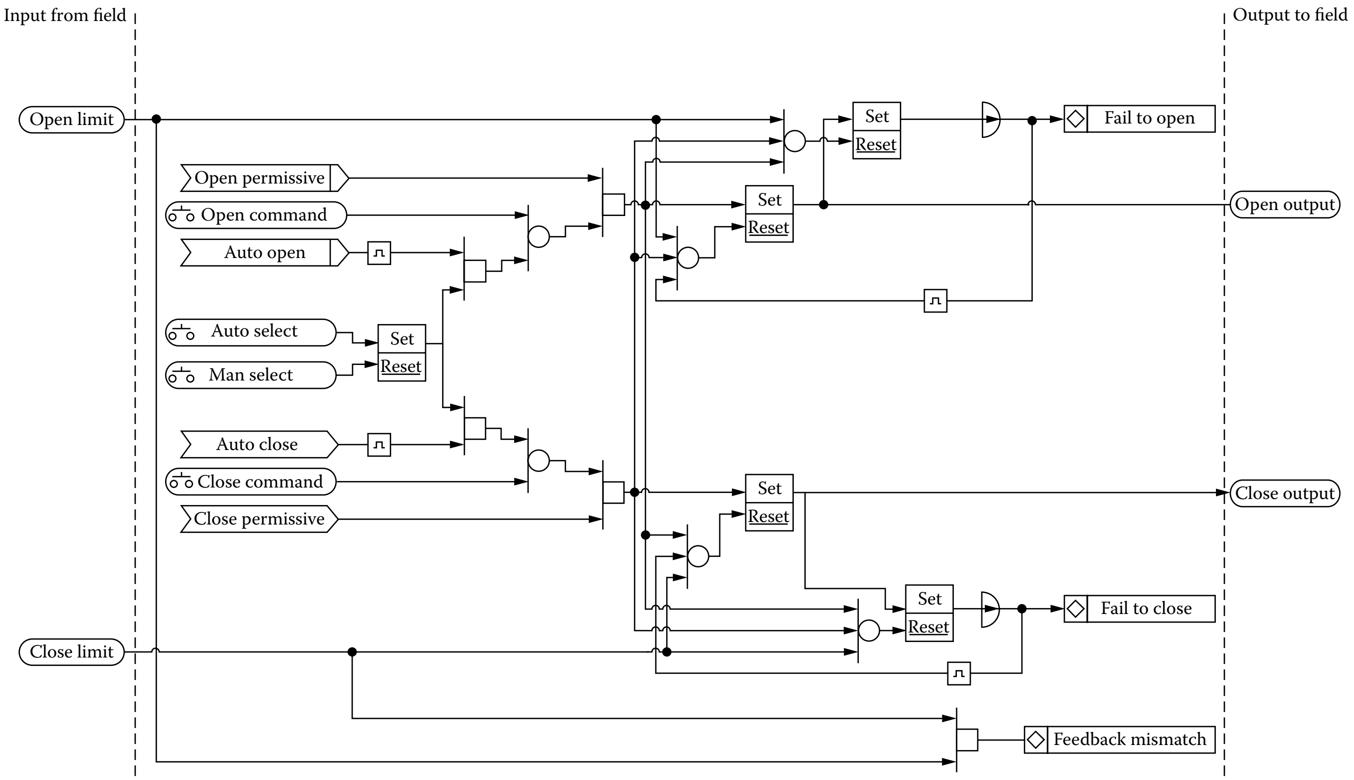


FIG. 5.7f
Standard MOV logic—maintained outputs.

Also, if the valve fails because process flow slows the stroke of the MOV past the stroke time limit of the fail timer, the fail alarm should reset automatically when the limit is eventually reached without requiring the operator to stroke the valve first.

If the MOV is opened (or closed) and later loses its open (or close) limit, then a fail alarm should appear. Often, logic is not written to do this, and many days can pass without the operator knowing that an MOV has failed.

For most MOVs the fail timer can be set to be the stroke time plus about 20 s. Make sure to time the stroke time for both open and close as they are often different. Also, it is wise to power the limit switches from the I/O cabinet instead of the breaker. This way, if the MOV breaker is open, the logic will still see the correct state of the valve and will not react to incorrect information.

For MOVs with latched outputs, as shown later in Figure 5.7l, the outputs must be cleared upon a fail to open or close; otherwise, there will be a chance of personal injury, since the output will remain True while someone is troubleshooting the valve.

Solenoid Valves

Figure 5.7g shows the logic for solenoid valves. Often, only one limit switch or none is used. In these cases, simply omit the failure logic for those limit switches that are missing.

PUMPS

Figures 5.7h and 5.7i show standard logic for the two most common types of pump motor. Figure 5.7h, the most common type of pump motor, has one maintained output for starting the pump (True = start and False = stop). The pump motor with two momentary outputs shown in Figure 5.7i is often used for very large pump motors.

Definitions

Input to the logic:

Start and stop command: The commands from the graphic display initiated by the operator to start and stop the motor. Note that these are always momentary.

Run contact: The indication from the pump motor that tells the logic that the pump is running.

Start permissive: A condition that is required before the pump can be started.

Process trip: Process conditions that shut down the pump motor (i.e., trip the pump when the tank level is too low).

Auto start: A condition that starts a pump automatically while the pump is in auto mode (example: Auto start the pump when another pump fails).

Auto stop: A condition that stops the pump automatically when the pump is in auto mode.

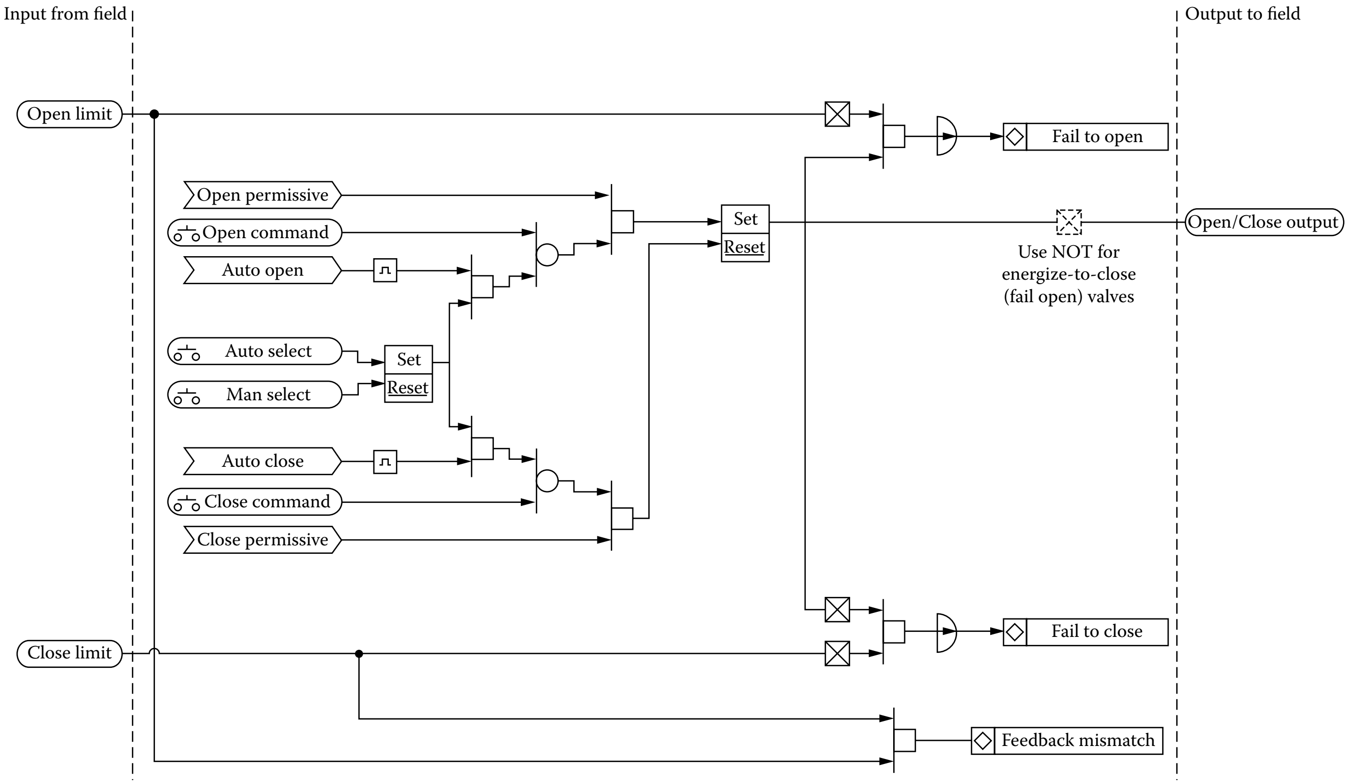


FIG. 5.7g
Standard solenoid valve logic.

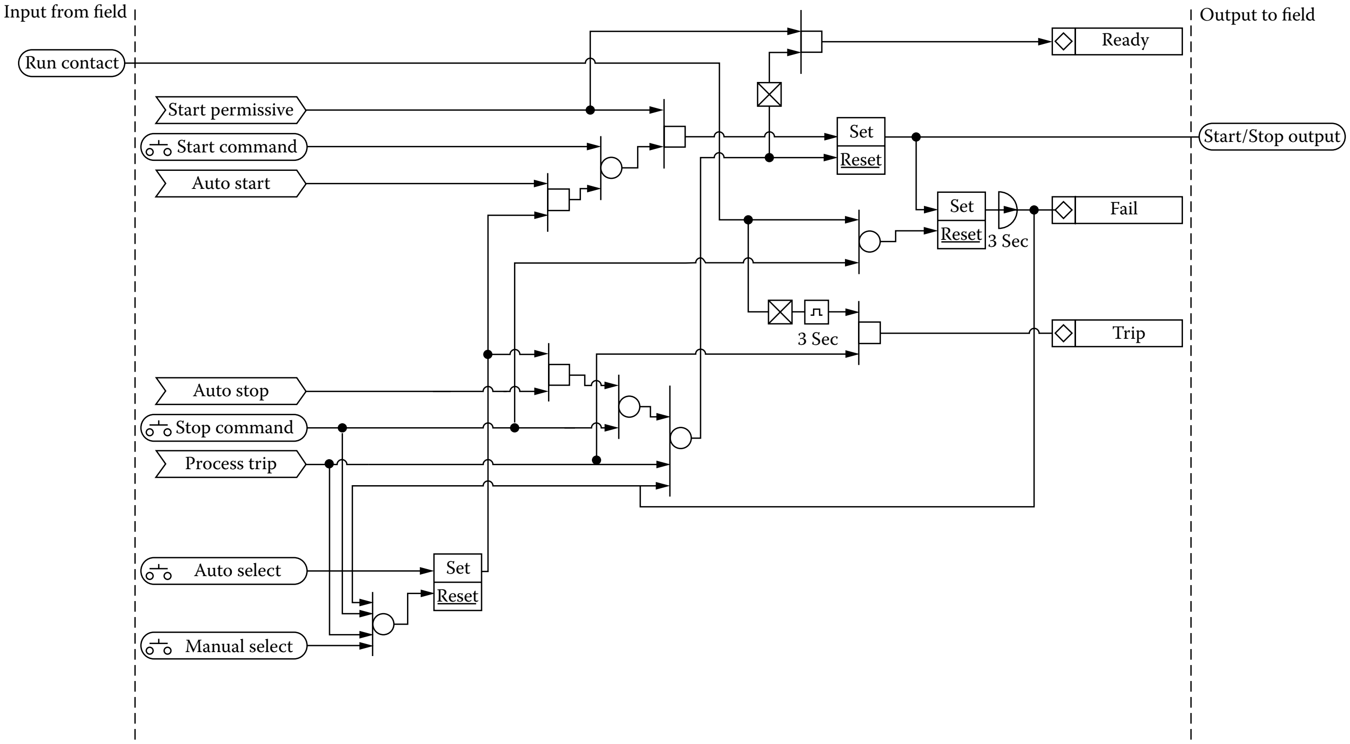


FIG. 5.7h
Standard pump logic—maintained output.

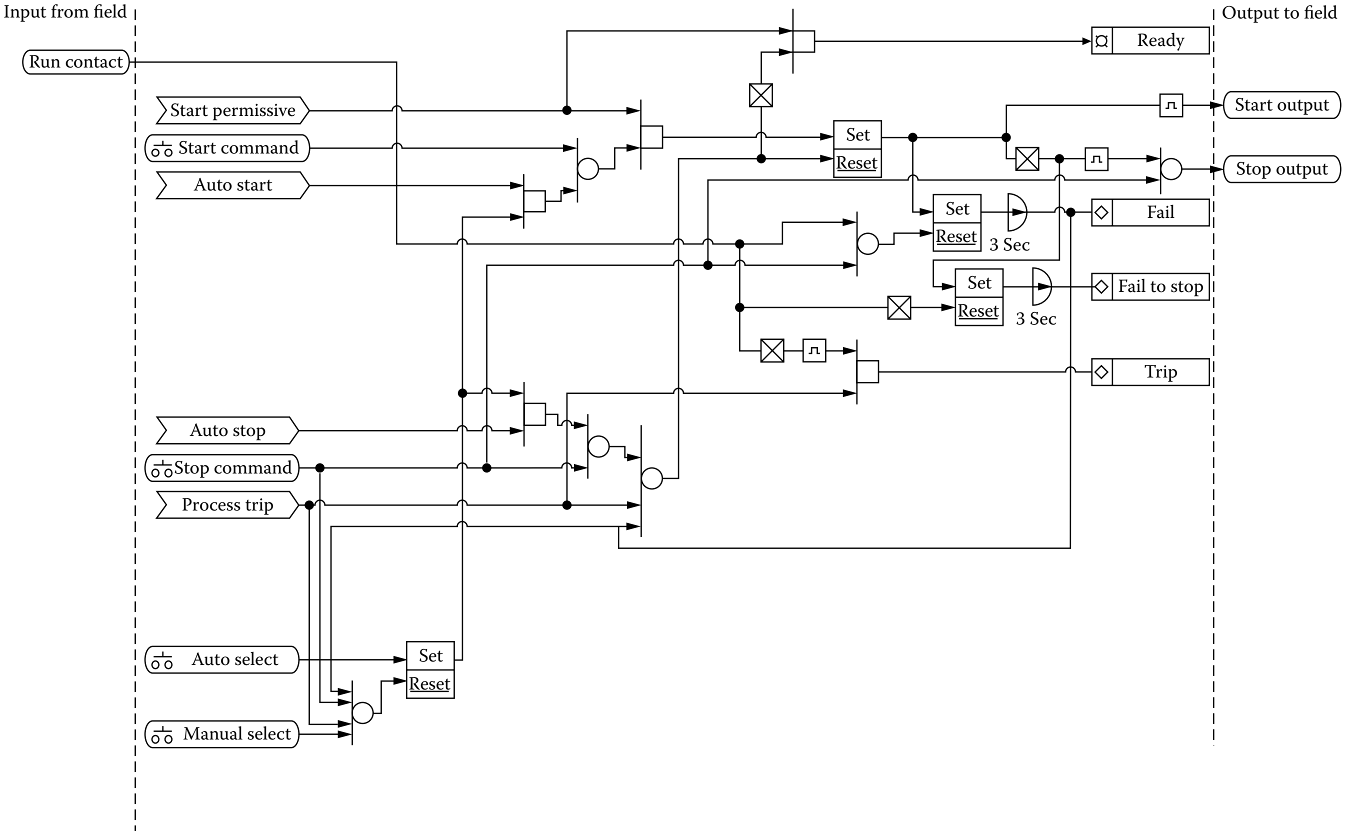


FIG. 5.7i
Standard pump logic—momentary output.

Output from the logic:

Start and stop output: The output from the logic to the pump motor that starts and stops the motor.

Indications and alarms:

Ready: Shows that all permissives have been met and the pump is ready to be started by the operator.

Auto mode: Shows that the pump is in auto mode and will be operated automatically by the auto start and auto stop commands.

Failed: An alarm that tells the operator that the pump was given a start output, but it is not running. This tells the operator that there is a problem with the pump itself and indicates that either the pump did not start when it was told to, or that it stopped after it had been running, *but it was not stopped by the logic*.

Tripped: An alarm that tells the operator that the pump was stopped by the logic via the process trip (not by the operator or an auto stop command). This alarm is not always included in pump logic.

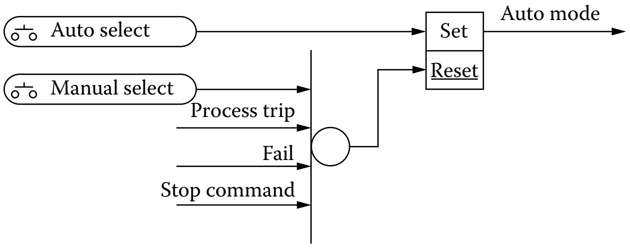


FIG. 5.7j
Typical pump auto mode.

Auto Mode

The “auto” and “manual” push buttons shown in Figure 5.7j set and reset the auto mode that allows the pump to start and stop automatically when the auto start and auto stop commands are True. The auto mode reset should also include pump fail, process trip, and the operator stop command. If the auto mode does not reset upon pump failure, the pump will restart as soon as the failure is reset while the auto start is True. Similarly if the operator stop does not reset the auto mode the pump will restart after the operator stops it if the auto condition is True.

If the process trip does not reset the auto mode, a problem like the demineralized water pump behavior described in the introduction will occur. If the trip condition cycles on and off (as a level trip may do while the level hovers around the set point) while the auto start condition is True, the pump will cycle on and off.

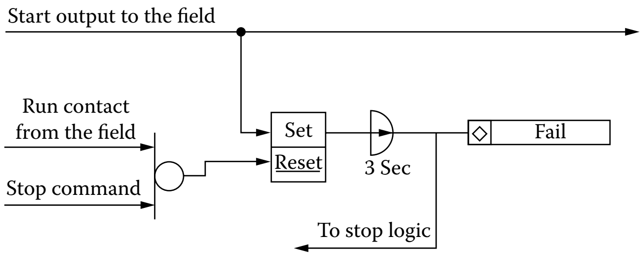


FIG. 5.7k
Typical pump failure logic.

Pump Fail

A pump failure occurs when a start output is sent to the motor and no run contact is detected (Figure 5.7k). This is different from a pump trip, which indicates that the pump was tripped by the logic upon a process condition. Making the distinction between fail and trip eliminates much confusion about what has shut down a motor—something that happened in the field or something caused by the logic. The "fail" alarm tells the operator that the problem is in the field and the "trip" alarm indicates that the motor was shut down by the logic.

For safety purposes, when a motor fails, the start output must be reset so that a person inspecting the failed pump will not be surprised by a sudden start. Although this is accepted as standard design practice for safety purposes, it is often erroneously omitted.

The failure indication will remain True and will inhibit the start of the motor until the operator manually resets it by giving a stop command. The start inhibit is performed by feeding the fail bit into the stop logic (as shown in Figures 5.7h and 5.7i) and should not be pulsed before it is fed to the stop circuit or the pump will be allowed to start again without first clearing the fail indication, and will not clear the start command after it has failed a second time.

Other Common Problems with Pump Logic Design

Quite often, motor logic is written so that when the run contact is seen by the logic, a start output is given to the motor. If, for example, the run contact is forced True during simulation, or if the run contact is accidentally made True when wiring is altered in the I/O cabinet, the pump will start unexpectedly. This is a safety hazard that must be removed from the logic.

The start and stop commands from the operator should always be momentary. Otherwise, the pump will start unexpectedly when the permissives are later satisfied if an earlier attempt was made to start the pump when the permissives were not satisfied. A maintained command will also cause the pump to start unexpectedly when the trip condition that caused the trip is cleared. The momentary pulse is usually executed by the graphic and not the programmed logic. For all of the diagrams given in this section, it is assumed that the command push buttons are momentary.

Problems often appear when permissives are overridden by another input from the logic such as an auto start. Make sure that the permissives are always used as they are intended. If in rare cases it is necessary to override a permissive, then design the logic so that the operator will have to make a conscious decision to perform the override.

Often a "red tag" or logic lockout system is used to show that a pump is undergoing maintenance and to lock out start commands from the operator. Surprisingly often, it is mistaken for a substitute for proper locking and tagging out of the motor breaker, and when used alone it does not provide adequate prevention against starting the motor. It is acceptable to put the motor in manual mode while using proper maintenance procedures for locking out the breaker instead.

After a pump trips upon a process trip, the operator should not be required to initiate a stop command before restarting the pump. This is another common design flaw that is not a safety hazard, but it is a nuisance to the operator and should be fixed.

Pumps with Two Outputs

Figure 5.7i shows standard logic that can be used for pumps with two outputs. When motors of this type fail, it is important to send a stop output to the motor even though it may already have stopped. If the motor fails after losing the run contact but is still running, then the motor will not trip when a process trip occurs because the logic "thinks" that the motor has already stopped.

The stop output for this type of motor can be made fail-safe by inverting it so that it is always True unless a stop command is given. If the stop output is not fail-safe, then it is necessary to include a "fail to stop" alarm in the motor logic. The logic in Figure 5.7b has been designed so that the operator can make a second attempt to stop the motor upon a fail to stop without having to give a start command first.

Controlling Two Pumps Together

Standby Pump Logic Figure 5.7l shows how to start a standby pump upon the failure of the running pump. The fail indication of one pump is simply fed to the auto start input of the standby pump so that when the standby pump is in auto mode, it will start when the running pump fails. When two pumps are required to run at the same time, for example, when the discharge pressure of the running pump is insufficient, a low pressure contact is fed into the auto start. In this case, a time delay is required after the first pump is started to allow time for pressure to build up; otherwise, the standby pump will start as soon as the main pump has started.

Some pump systems should start the standby pump on trip conditions that are not common to both pumps as well as on pump failure. For example, a standby boiler feed pump can start when the running pump trips on high vibration or bearing temperature, low lube oil pressure, and any other noncommon trips, as well as when it fails. Figure 5.7m shows how this is done.

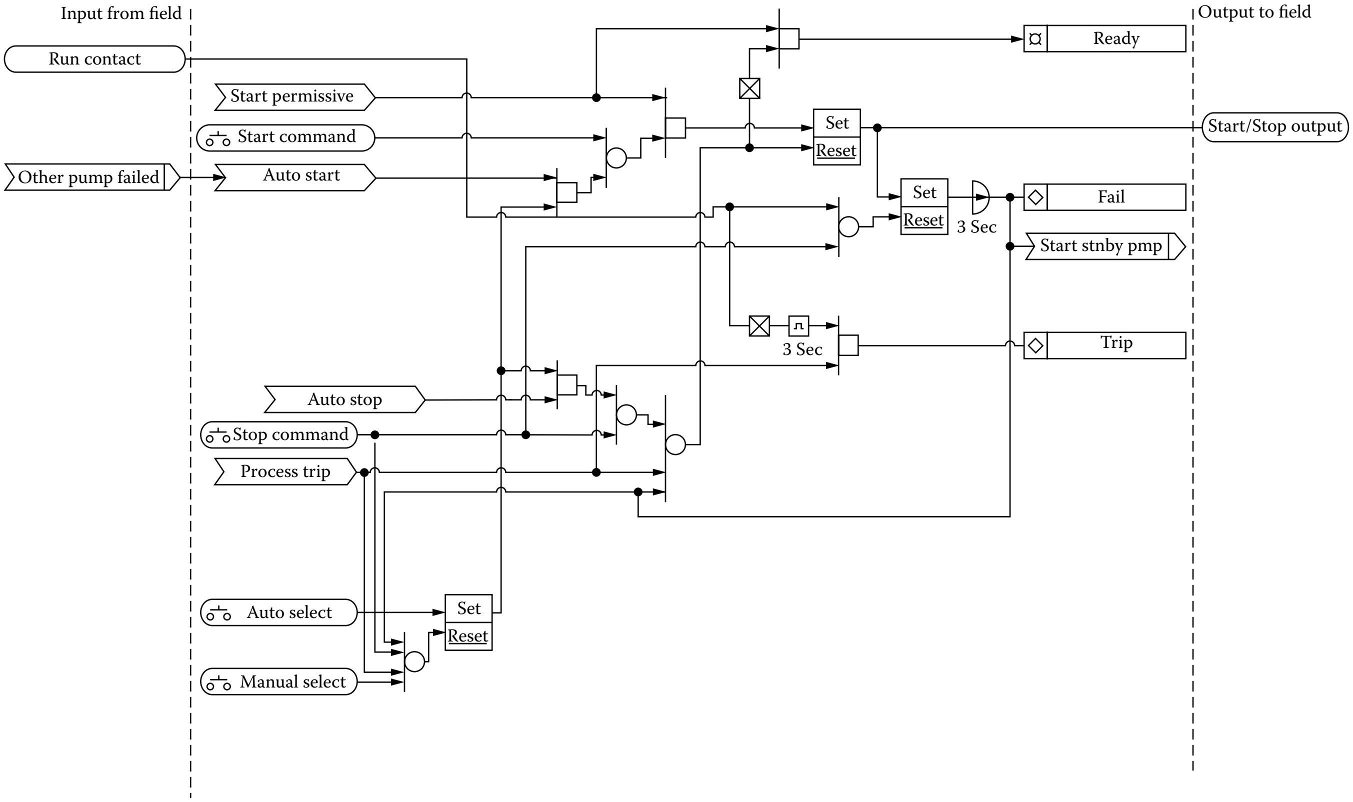


FIG. 5.7l
Pump—standby start.

Main and Auxiliary Pump Logic Figure 5.7n shows logic for starting an auxiliary pump upon the start command of the main pump (for example, a boiler feed pump with an auxiliary lube oil pump, or a vacuum pump with an auxiliary seal water pump). Upon the start command of the main pump, the auxiliary pump is started first, and after a specified condition (such as lube oil pressure sufficient, seal water flow sufficient, or a simple time delay) the main pump is started.

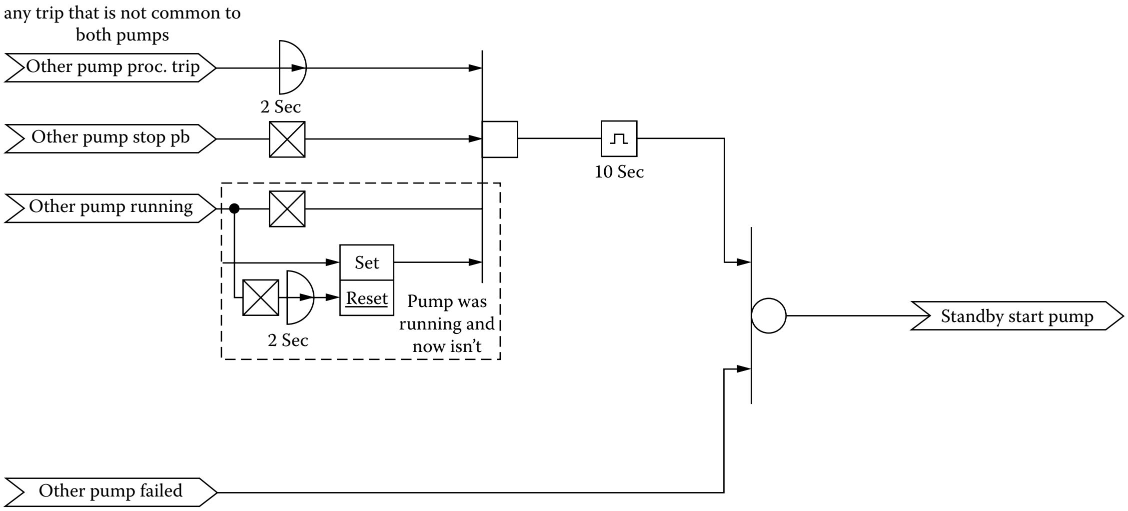


FIG. 5.7m
Standby start upon fail and noncommon trips.

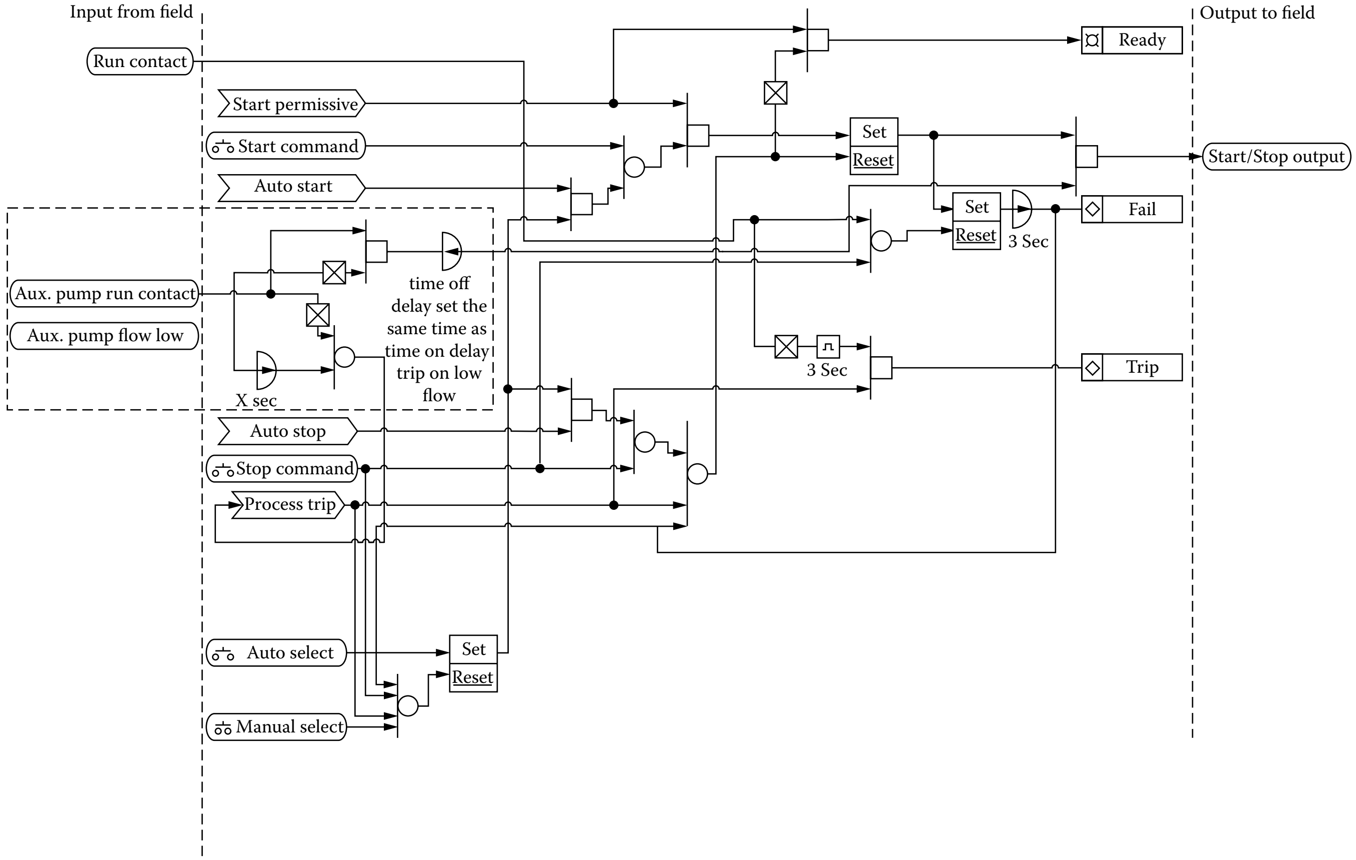


FIG. 5.7n
Main and auxiliary pump start.

BREAKERS

Figure 5.7o shows simple breaker logic. If both the field circuit and the control system logic are controlling the breaker, make sure that the two parts of logic do not interfere with each other. A remote/local switch is often used to remedy this and is fed into the permissive.

ANALOG CONTROLS

A feedwater controller is designed to switch from the start-up feedwater valve to the main valve when the start-up valve reaches 100%—immediately closing the start-up valve and popping the main valve open to 20% during the start-up of the power plant. The operator dreads the swap over to the main valve, because it upsets the drum level control, and it must be caught in manual mode to keep the plant from tripping.

Switchover

Switching over from one valve to another in a process is often a tricky thing to control. Often the logic is written with transfer blocks that immediately close one valve while opening the other. For a smoother transition, curves can be used instead and can be adjusted as needed during the tuning process to better control the way the valves open and close during switchover. For split ranged valves, if the first valve can be kept open after the second one starts to open, then logic as shown in Figure 5.7p can be used. If the first valve must close after the second opens, then the first valve must be ramped closed (not using the curve), while releasing the second valve to open on its own as it controls the process.

For switching between two valves that are controlled with separate controllers, the hold option in the controller can be used to freeze the controller that is not in use.

Pop Open/Clamp Closed

To pop a valve open or clamp it closed upon a specified condition, use the low and high output limit parameters of the controller that specify the range that the output of the controller is allowed—normally 0 to 100%. For example, all that is needed to clamp a valve closed (say, when a pump is not running) is simply to change the high output limit from 100 to 0% so that the range that the output of the controller is allowed will be 0 to 0%. This example, shown in Figure 5.7q, shows how the change from 100 to 0% is handled with a transfer block. The valve is clamped closed the

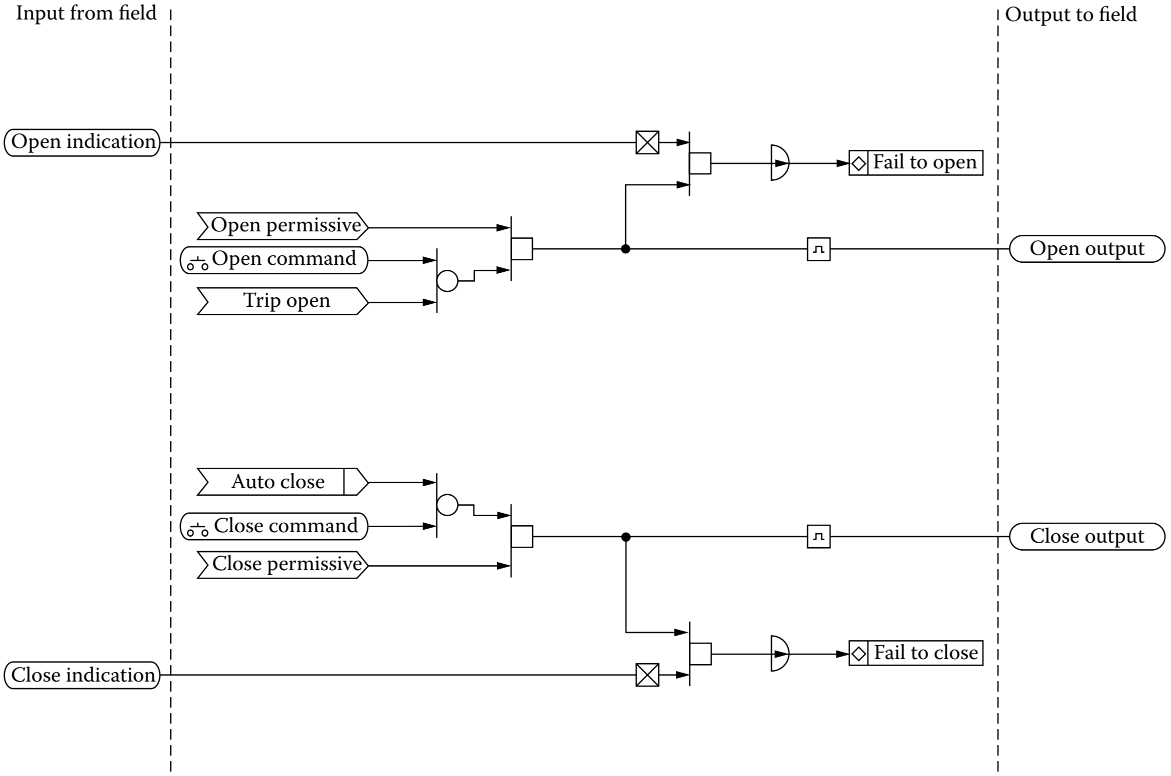


FIG. 5.7o
Standard breaker logic.

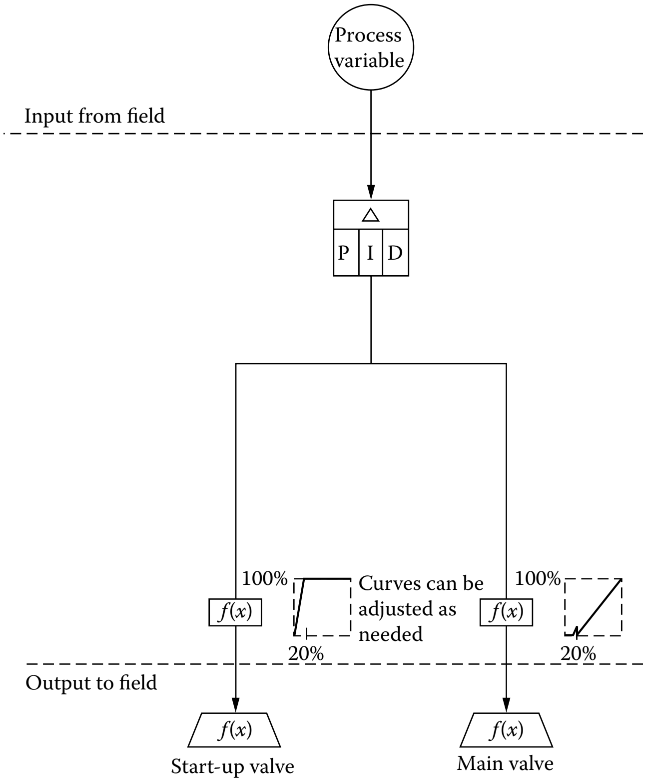


FIG. 5.7p
Analog—split range valves.

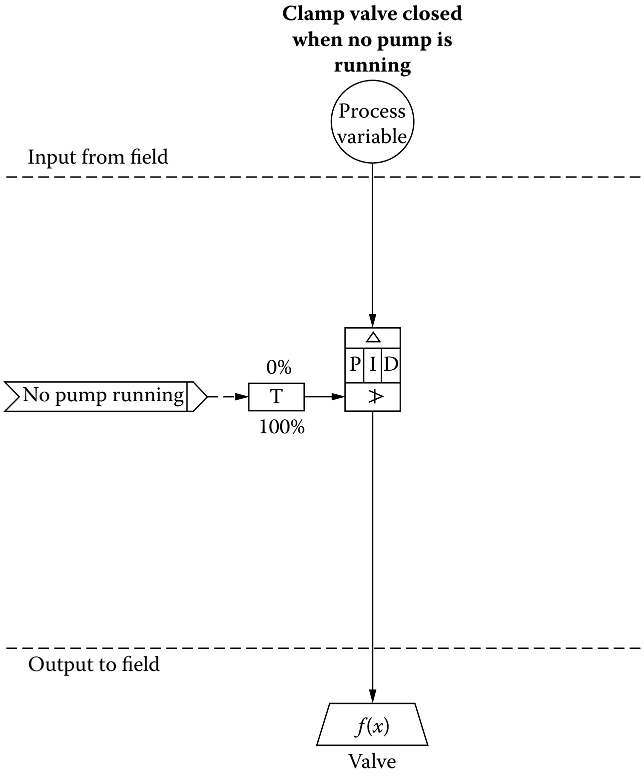


FIG. 5.7q
Analog—clamp valve closed when no pump is running.

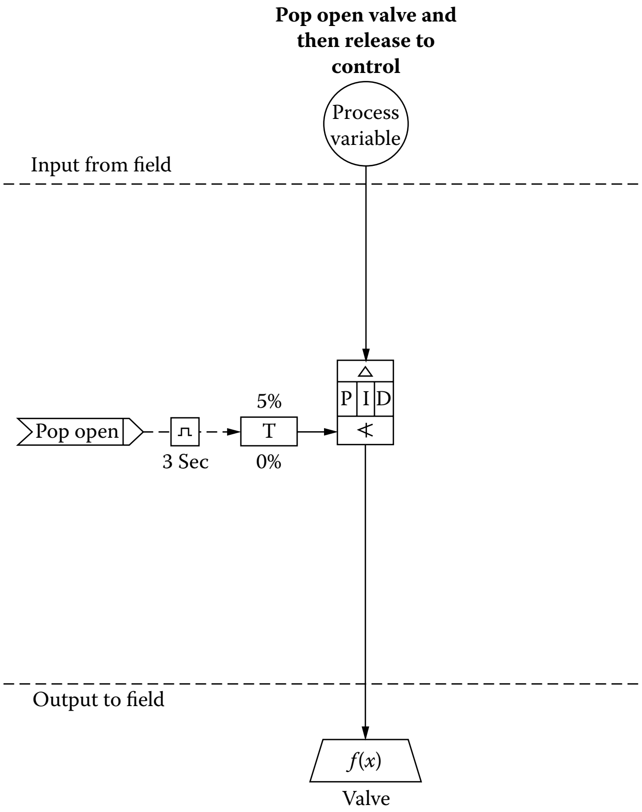


FIG. 5.7r
Analog—pop open valve and then release to control.

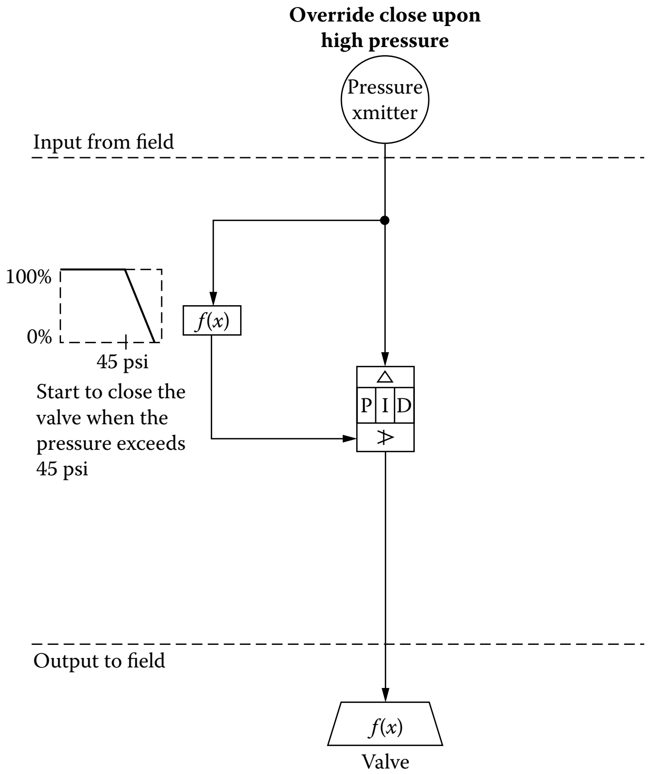


FIG. 5.7s
Analog—override close.

whole time that the pump is not running, so that there is no need also to switch the controller to manual mode, as is often done. When the pump runs again, the controller is automatically released and there is no need to have to remember to reset the controller to auto mode before starting the pump.

To pop open a controller, simply increase the low output limit. The controller will open to the specified position and start controlling from that point. In these cases, it is best to make the pop-open condition momentary so that the operator can manually close the valve if needed (Figure 5.7r). If a smoother transition is desired, a lagged transfer block can be used.

Override Open and Close

To override the controller to ensure that a limit on the process variable is maintained (for example, to push a level controller closed upon high water level in a drum), use a function block to generate a curve to feed the high output limit. As shown in Figure 5.7s, the input to the curve is the process variable to be controlled. When the limit is nearly reached, the output of the curve will start to decrease the high output limit from 100 to 0%, to close the valve until the process is normal. These curves can be adjusted during the tuning stage to best control the limit.

Similarly, one can use a controller to make sure that a limit is not reached on a process variable that one is not directly controlling and feed the controller output into the high or low limit of another related controller. This is different from cascade control (explained below), which controls two process variables at the same time, or feedforward control (explained below), which modifies the main controller as another process variable changes. In this case, one normally controls one process variable and uses the override to control only during the "out of spec" condition. For example, to control recirc flow in a system while making sure not to overpressurize the main discharge pipe, simply feed the overpressure controller output into the low output limit of the flow controller. This will push open the flow controller only when needed, as shown in Figure 5.7t.

Feedforward

If the change of one process affects another process under control, one can use the bias input of the controller to modify the controller output as the other process variable changes. This is especially handy for anticipating the need for more or less output before the controller sees the result of the change. The bias simply adds or subtracts a value from the output of a controller and is often added during the tuning process as needed (Figure 5.7u). The process variable should be passed through a multiplier so that the effect of the feedforward can be adjusted as needed. The feedforward adjustment can be made even more accurate by using a curve instead of a multiplier if

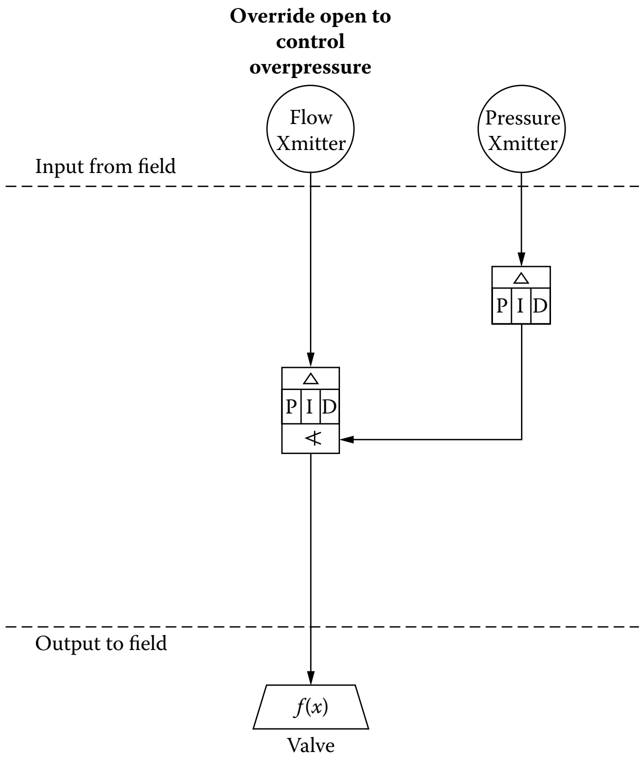


FIG. 5.7t
Analog—override open.

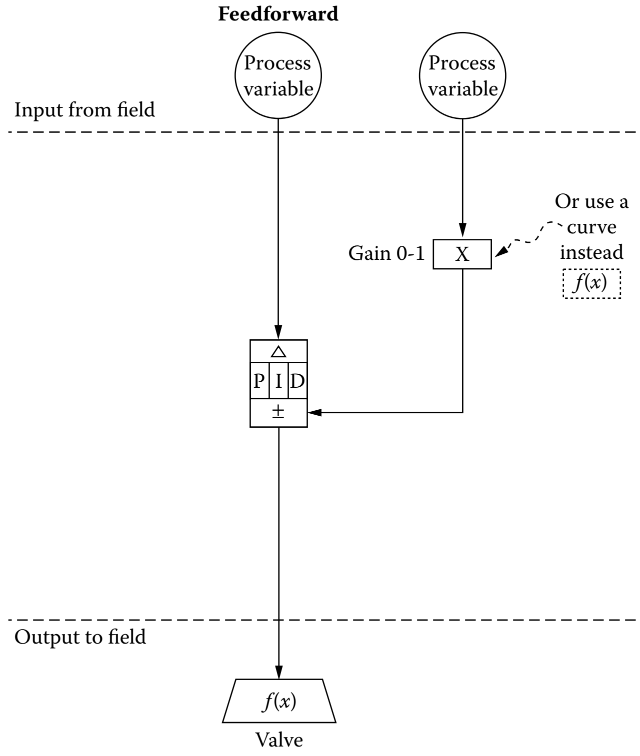


FIG. 5.7u
Analog—feedforward.

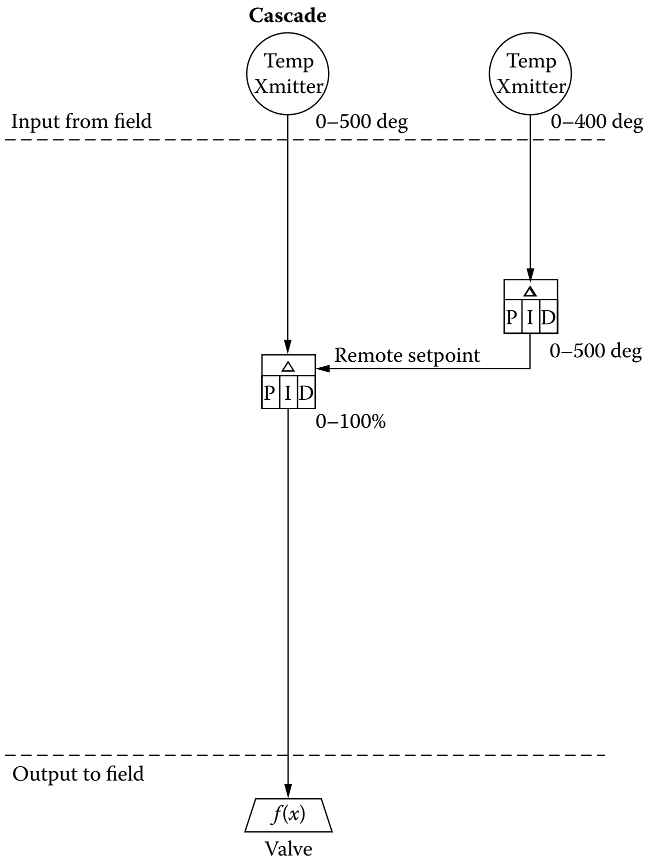


FIG. 5.7v
Analog—cascade.

it is known exactly how the controller should change with the change of the process variable.

Cascade

To control two process variables at once, use two controllers and feed the output of the first controller into the set point of the second controller. Then feed the output of the second controller to the valve (Figure 5.7v). Think of the controller that feeds the valve as the worker, because it is usually located at the point in the process where it can have the most effect or can see process changes sooner. The top controller adjusts that controller as necessary. Remember to adjust the high and low output limits of the controller so that the output is within range of the set point of the worker.

Three-Transmitter Select

When three redundant transmitters are used, a selection method as shown in Figure 5.7w is used to choose the median value, if the data from all three of the transmitters are of good quality, or to use the average of two if one transmitter has a faulty reading, or to select one transmitter. If the output of a transmitter has bad quality (usually specified as below 4 mA or above 20 mA), the logic automatically throws out its value.

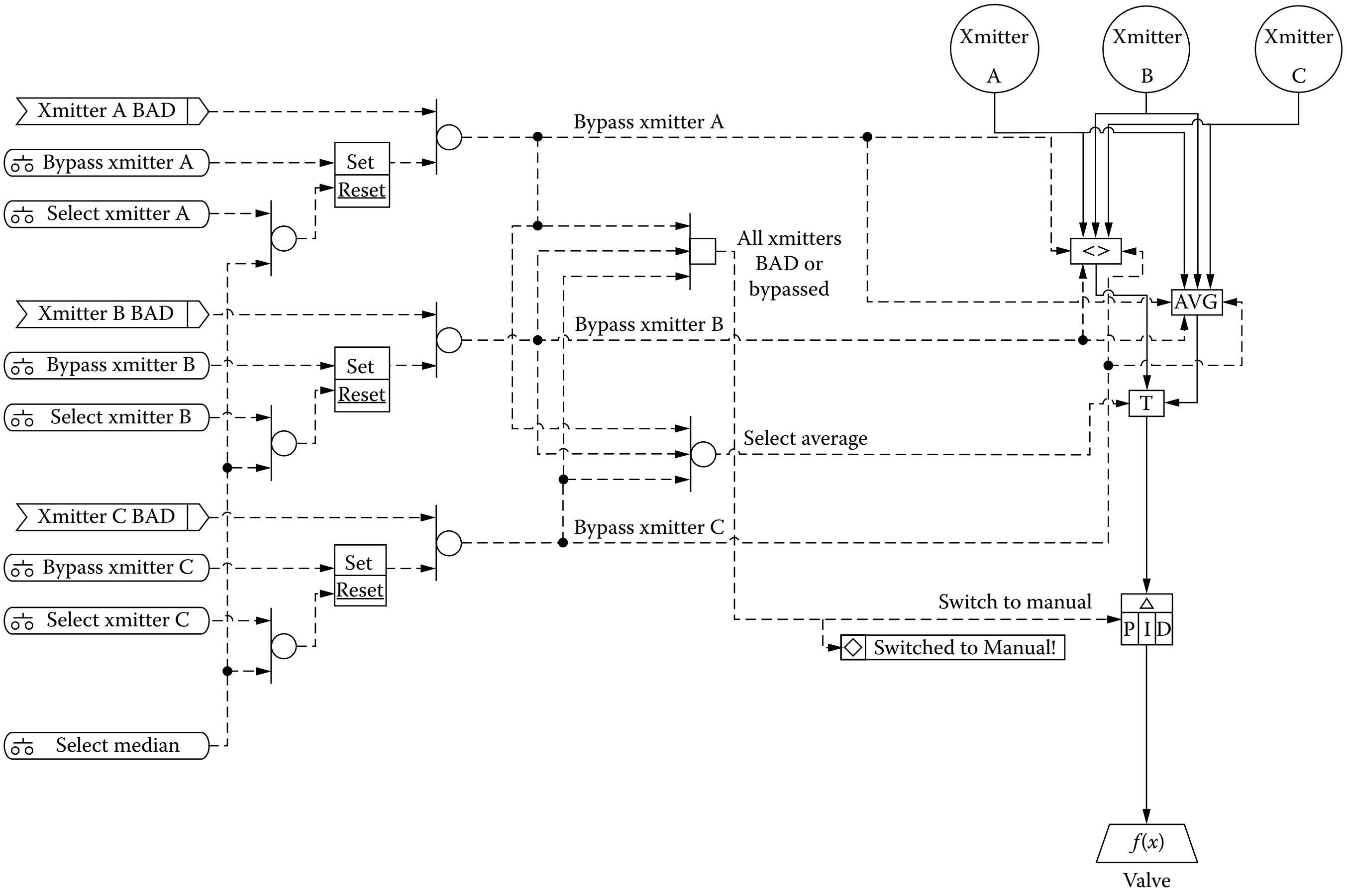


FIG. 5.7w
Analog—three transmitter selector.

The operator can also choose to bypass a transmitter that has a reading that does not "look right." If all of the transmitters are bad or bypassed, any controller that is using the selected value as a process variable may be set to manual mode.

Switch to Manual Mode

If a controller is automatically switched to manual mode by the logic, there should always be an alarm to the operator to show what has happened. Switch to manual mode only if absolutely necessary (for example, where there is no available process variable). Otherwise, it is best to leave the decision to switch to manual mode to the operator's discretion.

Changing the Set Point with Changes in the Number of Pumps Running

The set point of a controller such as a flow controller can be changed as the number of pumps running changes by using the logic shown in Figure 5.7x. If the pump is tripped when the flow is too high or low, one must make sure to add a hefty time delay in the trip logic to allow for the process to meet set point after the set point has been changed.

START-UP AND SHUTDOWN SEQUENCES

Automatic start-up sequences will start an entire plant with the touch of a button. Because they take a lot of control out of the operator's hands, they should be reserved only for those plants that are particularly complicated and would take more time to start manually, or those plants that are started and shut down frequently. Operators often have a difficult time following these fully automatic sequences, especially if the sequence graphics are not sufficiently detailed.

Think of the sequence logic as acting in place of the operator during the start-up sequence. It should not include any safety logic that is normally included in the logic for the pumps and valves themselves so that the safety logic will function even when the auto sequence is not used or is on hold. For example, logic that automatically closes boiler vent valves when the steam pressure reaches 25 psig should reside in the valve logic, not in the sequence, because the automatic start-up sequence could be on hold while the pressure passes the 25 psig set point.

Figure 5.7y shows the basic structure of a start-up sequence. The logic should include a hold button so that, if there is an equipment malfunction during the sequence, it

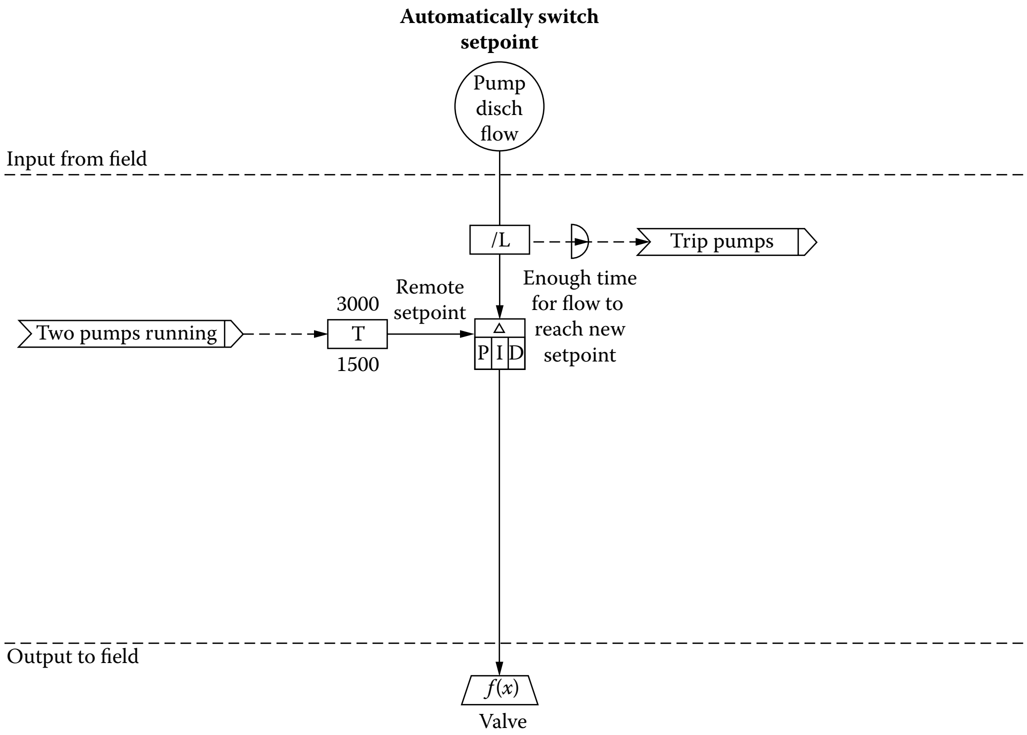


FIG. 5.7x
Analog—automatically switch set point.

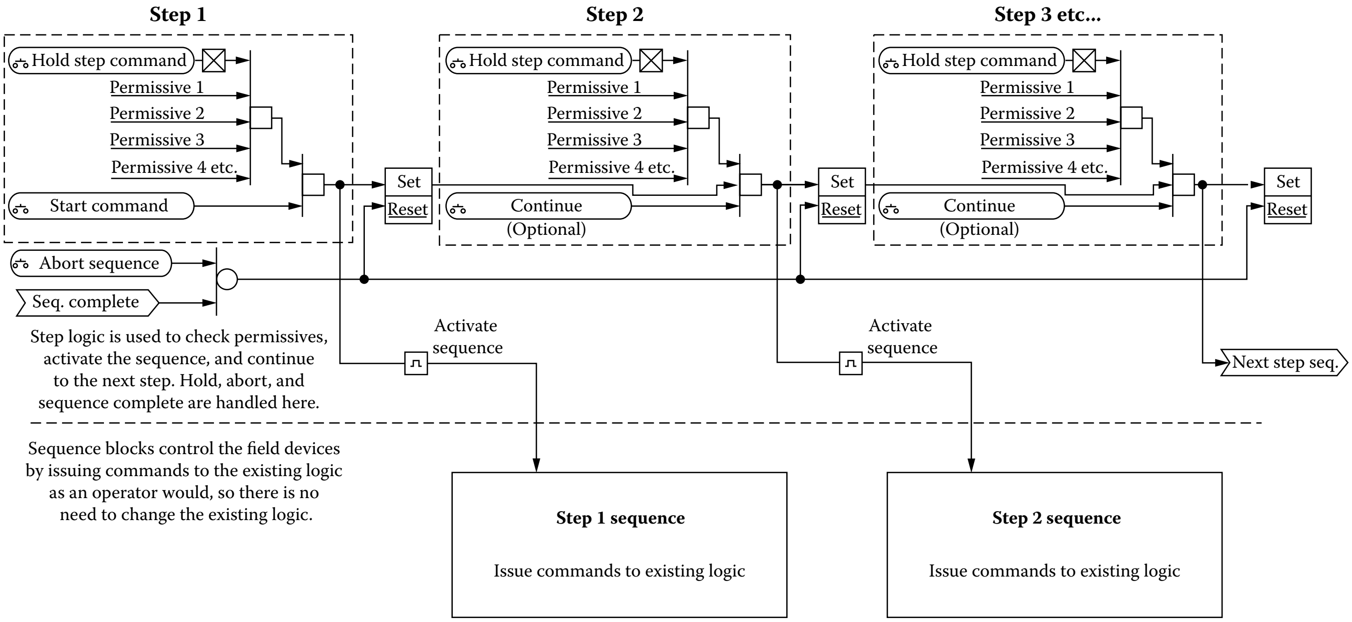


FIG. 5.7y
Auto start sequence logic structure.

will wait while the problem is remedied. In this way, the operator will not have to shut down the sequence and start over. A sequence abort should be included in case the problem cannot be fixed quickly.

OPERATION AND CUSTOMIZATION

Now is the time to operate the equipment with the logic the designer has written. By now the designer should have simulated the logic, independently of the system that it controls, to work out any bugs. This is very important, as it will take years to find all of the bugs as they occur during operation, and *it will actually save time to do all the testing beforehand.* Even logic that the designer has personally programmed, and most certainly logic that has already been tested in a factory acceptance test, should be retested.

Now that the equipment is running, the logic designer will soon find that the systems do not always behave exactly as expected. This stage of the job is a humbling experience because the designer inevitably finds problems with logic that was thought to be well designed for the system. Designers will produce a much better product if they are willing to make changes to the design, customizing it to fit the equipment in the field.

Listen to the operators. The logic designer must rely on operators for information on errors or upsets that have occurred while the designer was away. Operators can often offer important insights on the behavior of the equipment, as well.

A good historian system is a useful tool. If possible, make sure to collect data on all of the process variables and the outputs and set points for all of the controllers—it will make troubleshooting and tuning much easier. Keep all the changes simple. Before adding new logic or feedforward to controllers, make sure the system is well tuned first—additions may not be necessary. Go easy on the fancy stuff—tuning parameters that change automatically and controllers that switch make the whole process less stable.

A NOTE ON SAFETY

Not enough has been said in this section about safety. Fortunately, safety goes hand in hand with good logic design. Safety should be the first priority. When safe systems are designed, well-designed logic will follow.

A few simple points to remember: Ensure that the equipment does not start, open, close, etc. unexpectedly. Remove any outputs to the field when in the failed state. Make sure the logic is consistent throughout the plant so that the operator will know what to expect when a button is pushed.

Document your system well so that others can understand how it works. Design user-friendly graphics that show permissives and trips so the operators can easily determine what is going on. Fewer accidents happen when there is less confusion in the control room.

Bibliography

Babcock and Wilcox Co., *Steam/Its Generation and Use*, New York: Babcock and Wilcox, 1972.

Chesmond, C. J., *Basic Control System Technology,* New York: Van Nostrand Reinhold, 1990.

Dorf, R. C., *Modern Control Systems,* 4th ed., Reading, MA: Addison-Wesley, 1986.

Lipták, B. G., *Optimization of Industrial Unit Processes,* 2nd ed., Boca Raton, FL: CRC Press, 1999.

Polonyi, M. J. G., *Power & Process Control System,* New York: McGraw-Hill, 1991.

Stallings, W., *Data and Computer Communications,* 4th ed., Englewood Cliffs, NJ: Prentice-Hall, 1997.

Ulanski, W., *Valve and Actuator Technology,* New York: McGraw-Hill, 1991.

Woodruff, E. B., Lammers, H. B., and Lammer, T. F., *Steam Plant Operation,* 5th ed., New York: McGraw-Hill, 1984.

5.8 Programmable Safety Systems

ASISH GHOSH (2005)

Partial List of Programmable Safety System Suppliers for Process Industries:

ABB (Elsag-Bailey Controls) (www.ABB.com)
G.E. Fanuc Automation (www.GEIndustrial.com)
HIMA-Americas Inc. (www.hima-americas.com)
Honeywell ACS Service (honey well.com/acs)
ICS Triplex (www.icstriplex.com)
Rockwell Automation (www.rockwellautomation.com)
Siemens (www.sea.siemens.com)
Yokogawa Corp. of America (www.yca.com)
Triconex/Invensys (www.Triconex.com)

Partial List of PLC Suppliers:

ABB (Elsag-Bailey Controls) (www.ABB.com)
Allen-Bradley/Rockwell Automation (www.AB.com)
Automation Direct (www.Automationdirect.com)
Danaher (Eagle Signal Controls) (www.Dancon.com)
Eaton (Cutler-Hammer) (www.EatonElectrical.com)
Emerson (Westinghouse) (www.EmersonProcess.com)
Fuji Electric Corp. (www.FujiElectric.com)
G.E. Fanuc Automation (www.GEIndustrial.com)
Giddings & Lewis (www.GLControls.com)
Idec Corp. (www.Idec.com)
International Parallel Machines Inc. (www.ipmiplc.com)
Mitsubishi Electric (www.meau.com)
Modicon/Schneider Electric (www.Modicon.com)
Moeller Corp. (www.Moeller.net)
Omega Engineering (www.Omega.com)
Omron Electronics Inc. (www.Omron.com)
Reliance Electric Co./Rockwell Automation (www.Reliance.com)
Siemens (www.sea.siemens.com)
Toshiba Inc. (www.Toshiba.com)
Triconex/Invensys (www.Triconex.com)
Uticor Technology Inc. (www.Uticor.com)

INTRODUCTION

Since the publication of the IEC 61508 safety standard[1] and more recently the IEC 61511 standard for process safety,[2] the interest in rigorous safety analysis and in certified safety instrumented systems (SISs) has increased considerably among the users. As users are becoming more knowledgeable about safety issues, they are increasingly focusing on the goal of overall safety.

Users want their safety systems to be cost-effective and to provide closer integration of the safety and control systems. They are looking for flexible architecture with more scalability. They are also looking for increased functionality for modifying alarm limits based on process conditions and on orderly shutdown procedures in case of an emergency. The major trends in safety systems are

- Increased focus on overall safety
- Closer integration with control systems
- Increased flexibility and scalability
- Increased function block capabilities

Both IEC 61508 and 61511 standards are performance-based; as such, they do not mandate any specific safety system architecture or risk assessment procedures. However, they do provide guidance on the analysis of safety life cycle, hazards, and risks, and on methods for determining safety requirements.

TABLE 5.8a
Factors that Increase Risk

- Operating plant and machinery closer to their limits
- Transient operation states
- Use of hazardous raw materials
- Manufacture of hazardous intermediates
- Presence of untrained personnel
- Absence of safety culture

Transient operations include startup, shutdown, shift change, and workforce transitions

Safety system certifications should objectively assess the reliability and availability of critical control and safety shutdown systems and related equipment. Technical Inspection Associations (in German, Technischer Uberwachungs Verin, or TUVs) in Germany have been in the forefront of inspection and certification of safety-related systems worldwide.

In choosing a safety system, users should take into account not only all the features of that system but also the specified restrictions, which are spelled out by the certification authority. This information is often found in the product safety manual. In choosing a system supplier, users should take into account the supplier's knowledge and experience in safety analysis, their application knowledge, and local support.

Risk Reduction

Risk is usually defined as a combination of the severity and probability of an unplanned event. Risk depends on how often that event can happen and how bad it will be when it does. In manufacturing operations, the type of events and their associated risks include loss of life or limb, environmental impact, loss of capital equipment, and loss of production. For many manufacturers, loss of company image can also be a significant risk factor. With increased environmental awareness, regulatory concerns, and threat of litigation, risk reduction is becoming more and more important to most manufacturers (Table 5.8a).

The best way to reduce risk in a manufacturing plant is to design inherently safe processes. However, inherent safety is rarely achievable in today's manufacturing environments. Risks prevail wherever there are hazardous or toxic materials stored, processed, or handled (Figure 5.8b).

Because it is impossible to eliminate all risks, a manufacturer must agree on a level of risk that is considered to be acceptable. After identifying the hazards, a study should therefore be performed to evaluate each risk situation by considering likelihood and severity. Site-specific conditions, such as population density, in-plant traffic patterns, and meteorological conditions, should also be taken into consideration during risk evaluation.

The risk levels that are determined by the safety studies can be used to decide if the risks are within acceptable levels. Basic process control systems, including process alarms and the means of manual intervention, provide the first level of risk reduction in a manufacturing facility. Additional protection measures are needed where a basic control system does not reduce the risk to an acceptable level. They include safety-instrumented systems along with hardware interlocks, relief valves, and containment dikes. To be effective, each protection subsystem should act independently of the others (Table 5.8c).

History

In the early days of process control, commonly used alarming and safety interlocking devices included pressure, flow, level,

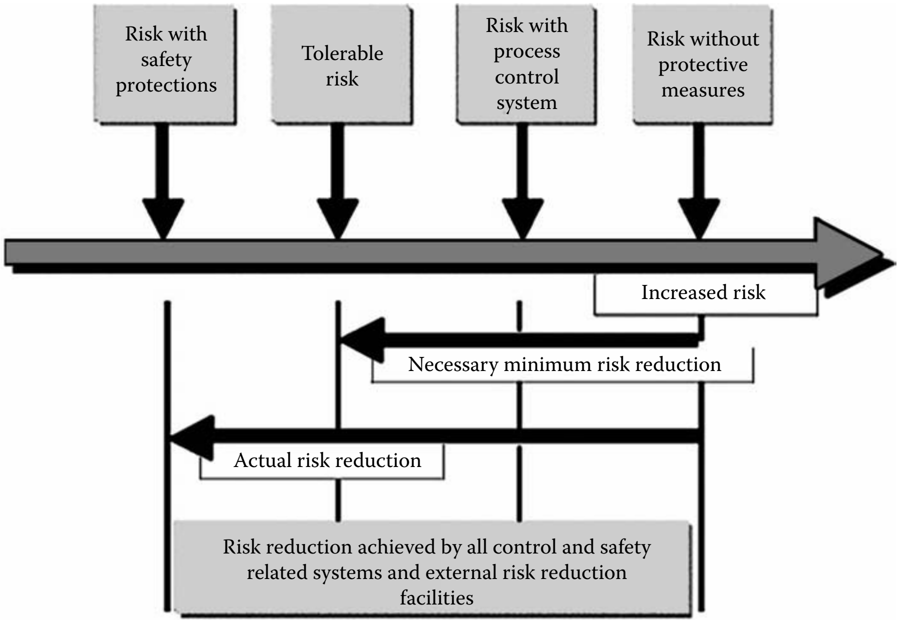


FIG. 5.8b
Reducing risk.

TABLE 5.8c
Driving Forces for Lowering Risks

- Higher environmental awareness
- Increased regulatory considerations
- Emergence of safety standards
- Maintaining company image

and temperature switches. These switches were simple mechanical or electromechanical devices that, upon detection of hazardous conditions, activated valves, motors, and other plant equipment to bring a process to a safe state. Other mechanical devices, which are also still used today, include such physical devices as electrical fuses, safety valves, and rupture disks.

While the electromechanical and solid-state relays could be used to design more sophisticated safety systems, they were difficult to program or to interface with digital computers. Hence, programmable safety systems were developed in the early 1970s. Programmable safety systems provide scalability, flexibility, and ease of configuration (Table 5.8d).

Duplex and Triplex Designs In the late 1970s, August Systems pioneered the development of the programmable safety system, which was followed by systems from Triconex and Triplex. These three suppliers developed the triple modular redundant (TMR) systems, in which three independent, parallel processors with extensive diagnostics are integrated into a single system (2oo3). At each decision point within the system, a two-out-of-three vote is taken to determine failures and guarantee correct operations. Other suppliers of TMR systems for process industries include GE Fanuc and Yokogawa (Figure 5.8e).

TABLE 5.8d
Typical Applications of Safety Systems

- Emergency shutdown (ESD)
- Fire and gas monitoring and protection
- Critical process control
- Burner management and control
- Turbine and compressor control
- Unmanned installations

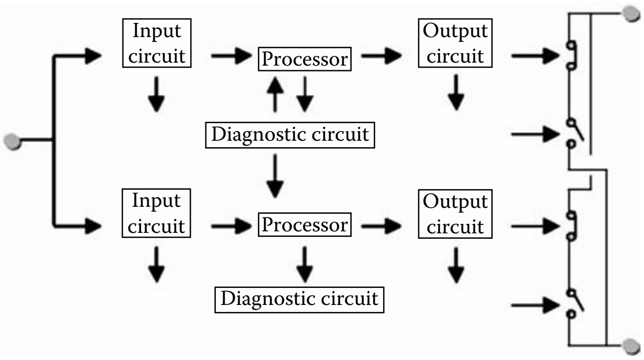


FIG. 5.8f
Typical duplex system.

A dual redundant system with extensive diagnostics (duplex) is another common safety system design. Here, two identical processors are configured as a married pair to check the health of the system (1oo2D). In this arrangement, two identical processors operate in parallel. They use the same inputs, while only one processor controls the output modules at any given time. The outputs of both processors are always compared to ensure that they are synchronized and identical. If they disagree, a diagnostic evaluation is initiated to determine which of the two is still reliable, and that the one used will continue the process in a safe state or shut it down. At the same time, messages are to fix the failed processor (Figure 5.8f). Major suppliers of duplex systems include ABB, Honeywell, Siemens, and Yokogawa.

Quadruple Redundant Systems Another safety system design is the quadruple modular redundant (quad) system. The quad architecture provides four processors — two per channel (2oo4) — which may be viewed as a pair of duplex

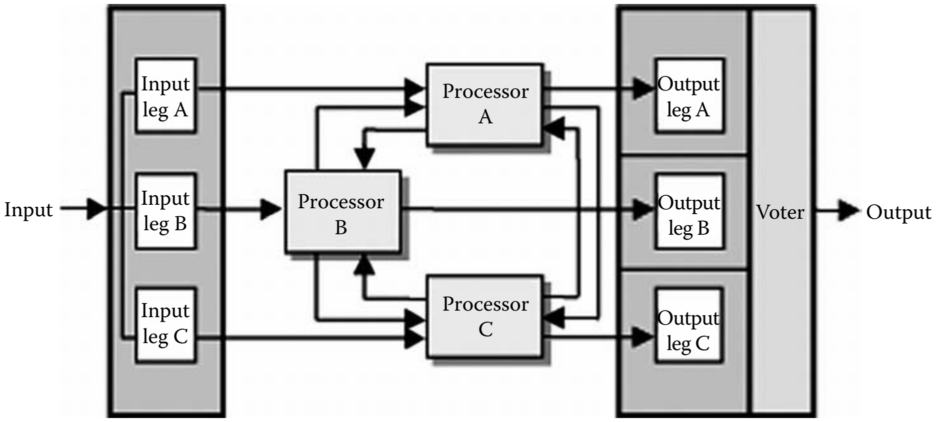


FIG. 5.8e
Typical TMR system.

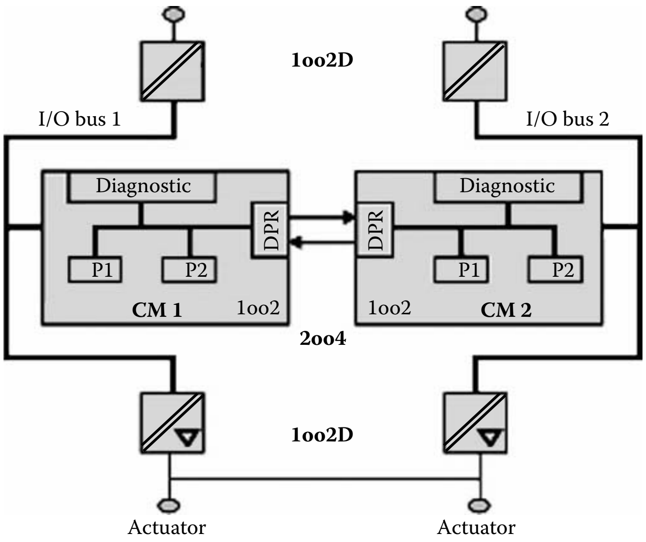


FIG. 5.8g
Typical quad system.

systems with diagnostics. Both pairs of active processors operate synchronously with the same user program. A hardware comparator and a separate fail-safe watchdog monitor the operation of each pair of processors to diagnose and resolve anomalies (Figure 5.8g).

At present HIMA and Honeywell are the two major suppliers of quad systems. The safety and availability of quad, TMR, and duplex systems are comparable. It is the quality of diagnostics and the system implementation that determines their relative performance.

In recent years, the increased awareness of safety, the impact of various regulatory agencies, and the publication of safety standards have led to the rapid growth in demand for safety systems. Many DCS- and PLC-based control system suppliers are competing for a share of this market.

SAFETY STANDARDS

The IEC 61508 safety standard published by the International Electrotechnical Commission (IEC) is applicable to a wide range of industries and applications. The standard is intended both as the basis for the preparation of more specific standards and for standalone use. A more specific international safety standard for process industries (IEC 61511) has also been published.

Since the publication of IEC 61508 and IEC 61511 standards, interest in rigorous safety analysis and in applying certified safety instrumented systems has increased. These standards give guidance on good practice and offer recommendations, but do not absolve its users of responsibility for safety. The standards not only deal with technical issues but also include planning, documentation, and assessment of all activities. Thus, the standards deal with the management of safety throughout the entire life of a system.

IEC 61508: General Safety Standard

The IEC 61508 standard is in seven parts:

- Part 1: General requirements
- Part 2: Requirements for safety-related systems
- Part 3: Software requirements
- Part 4: Definitions and abbreviations
- Part 5: Examples of methods for the determination of safety integrity levels (SILs)
- Part 6: Guidelines on the applications
- Part 7: Overview of techniques and measures

The standard is generic and can be used directly by industry, as a standalone standard, and by international standards organizations as a basis for the development of industry-specific standards, such as for the machinery sector, the process sector, or the nuclear sector. The IEC 61511 standard is more specific to the process industries.

IEC 61511: Safety Standard for Process Industries

While IEC 61508 has seven parts, the IEC 61511 standard has only three parts:

- Part 1: Framework, definitions, system, hardware, and software requirements
- Part 2: Guidelines on the application
- Part 3: Guidance for the determination of the required safety integrity levels

IEC 61511 Part 1 is primarily normative, while Parts 2 and 3 are informative. Part 1 is structured to adhere to a safety life cycle model similar to that in the IEC 61508 standard. The hazard and risk analysis utilizes the notion of protection layers and specifies the safety integrity level concept developed by the IEC 61508 standard. It also lists key issues that need to be addressed when developing a safety requirement specification. Issues like separation, common cause, response to fault detection, hardware reliability, and proven-in-use are also addressed in this part (Table 5.8h).

In this part of the standard, software safety requirement specifications are included, addressing such items as

TABLE 5.8h
Main Differences Between IEC 61508 and IEC 61511 Standards

IEC 61508	*IEC 61511*
Generic safety standard for broad range of applications	Sector-specific safety standard for the process industries
Applies to all safety-related systems and external risk reduction facilities	Applies only to safety-instrumented systems
Primarily for manufacturers and suppliers of safety systems and devices	Primarily for system designers, integrators, and users of safety systems

architecture, relationship to hardware, safety instrument functions, safety integrity level, software validation planning, support tools, testing, integration, and modification. In addition, a section is dedicated to factory acceptance testing requirements, and another section lists the installation and commissioning requirements.

Part 2 of the standard provides "how to" guidance on the specification, design, installation, operation, and maintenance of safety instrumented functions and related safety instrumented system as defined in Part 1 of the standard.

Part 3 of the standard provides guidance for development of process hazard and risk analysis. It provides information on:

- The underlying concepts of risk and the relationship of risk to safety integrity
- The determination of tolerable risk
- A number of different methods that enable the safety integrity levels for the safety instrumented functions to be determined

It also illustrates methods from different countries that have been proven-in-use. It further illustrates good engineering practices across cultural and technological differences, providing the end user with effective methods from which to select.

ANSI/ISA-84.01 Standard

The original ANSI/ISA-84.01 standard was published in 1966; as such, it predates the IEC 61508 safety standard. However, it is being abandoned in favor of the IEC 61511 international standard. A new ISA standard was released in 2004, which was nearly identical to the IEC 61511 safety standard. There is, however, a grandfather clause in the new version that allows the continued use of safety systems following the original version of the standard.

The safety standards give guidance on good practice and offer recommendations, but do not absolve its users of responsibility for safety. The standard recognizes that safety cannot be based on retrospective proof, but must be demonstrated in advance, and there cannot be a perfectly safe system. Therefore, the standards not only deal with technical issues, but also include planning, documentation, and assessment of all activities. Thus, the standard deals with the management of safety throughout the entire life of a system. The standards bring safety management to system management and safety engineering to software engineering.

Safety Integrity Levels Safety integrity is defined as the likelihood of a safety instrumented system satisfactorily performing the required safety functions under all stated conditions, within a stated period. A safety integrity level (SIL) is defined as a discrete level for specifying the safety integrity requirements of safety functions. Whereas a safety integrity level is derived from an assessment of risk, it is not a measure of risk. It is a measure of the intended reliability of a system or function (Table 5.8i).

TABLE 5.8i

Safety Integrity Levels (SIL)

Safety Integrity Level (SIL)	*Probability of Failure on Demand Mode of Operation*	*Probability of Failure on Continuous Mode of Operation*
1	$\geq 10^{-2}$ to $<10^{-1}$	$\geq 10^{-6}$ to $<10^{-5}$
2	$>10^{-3}$ to $<10^{-2}$	$>10^{-7}$ to $<10^{-6}$
3	$>10^{-4}$ to $<10^{-3}$	$>10^{-8}$ to $<10^{-7}$
4	$\geq 10^{-5}$ to $<10^{-4}$	$\geq 10^{-9}$ to $<10^{-8}$

Notes:

1. Demand Mode: Where actions are taken in response to process or other conditions (no more than once per year)
2. Continuous Mode: Functions, which implement continuous control to maintain functional safety

Safety Life Cycle Safety life cycle is a method or procedure that provides the way to specify, design, implement, and maintain safety systems in order to achieve overall safety in a documented and verified way. All major safety standards, such as ANSI/ISA-84-01-1996, IEC 61508, and IEC 61511, have specified safety life cycles, which show considerable similarities, differing only in the details. The safety life cycle specified by the IEC 61511 standard shows a systematic approach to safety starting from hazard and risk analysis to implementation of safety system and finally to its decommissioning (Figure 5.8j).

Perform Hazard and Risk Analysis: Determine hazards and hazardous events, the sequence of events leading to hazardous condition, the associated process risks, the requirements of risk reduction, and the safety functions required.

Allocate Safety Functions to Protection Layers: Allocate safety functions to protection layers and safety systems.

Specify Requirements for Safety System: Specify the requirements for each safety system and their safety integrity levels.

Design and Engineer Safety System: Design system to meet the safety requirements.

Design and Develop Other Means of Risk Reduction: Means of protection other than programmable safety systems include mechanical systems, process control systems, and manual systems. They are not specified in any detail in the standard (Figure 5.8k).

Install, Commission, and Validate the Safety Protections: Install and validate that the safety system meets all the safety requirements of the safety integrity levels.

Operate and Maintain: Ensure that the safety system functions are maintained during operation and maintenance.

Modify and Update: Make corrections, enhancements, and adaptations to the safety system to ensure that the safety requirements are maintained.

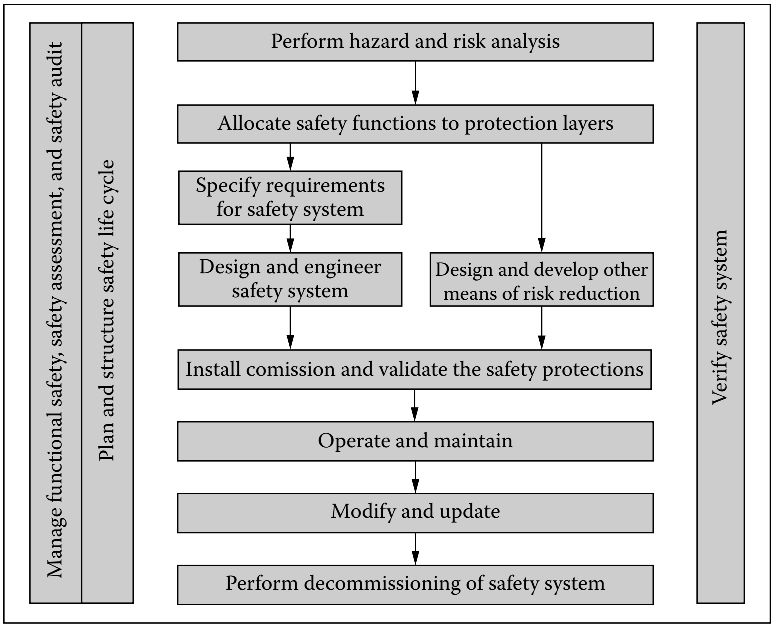


FIG. 5.8j
The safety life cycle.

Perform Decommissioning of Safety System: Conduct review and obtain required authorization before decommissioning of a safety system. Ensure that the required safety functions remain operational during decommissioning.

Manage Functional Safety, Safety Assessment, and Safety Audit: Identify the management activities that are required to ensure the functional safety objectives are met.

Plan and Structure Safety Life Cycle: Define safety life cycle in terms of inputs, outputs, and verification activities.

Verify Safety System: Demonstrate by review, analysis, or testing that the required outputs satisfy the defined requirements for each phase of the safety life cycle.

Like all models, the safety life cycle is also an approximation. While the life cycle phases are listed sequentially, in reality, there are significant iterations between these phases.

Requirements of some of the functions, such as hazard and risk analysis, allocation of safety functions to protection layers, and designing and developing other means of risk

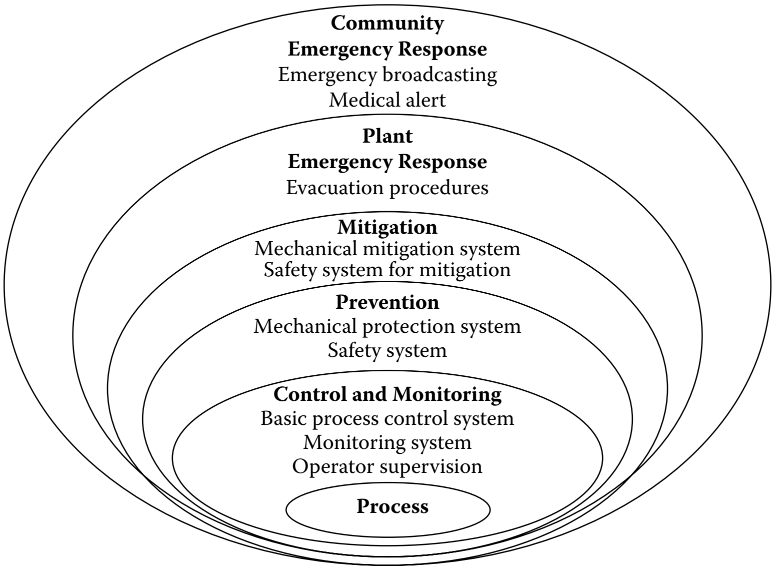


FIG. 5.8k
The protection layers.

reduction, are not specified in any detail in the standard. Certain functions, such as managing functional safety, planning and structuring safety life cycle, and verifying safety requirements at each phase, are carried out continuously during the whole life cycle and are shown as vertical boxes in the figure.

MANAGEMENT CONSIDERATIONS

The standards should be recognized as defining requirements for safety management rather than merely for system development. Not all safety life cycle phases will be relevant to every application, and it is the responsibility of management to define what requirements are applicable in each case. The standards do not prescribe exactly what should be done in any particular case, but only offer advice and guidance to the management. Therefore, management is still responsible for taking appropriate actions and for justifying them.

Management responsibilities include rigorous safety planning, which includes the choice of safety life cycle phases to be used and the activities to be carried out within those phases. However, the users should realize that safety systems by themselves do not achieve safety. People working within a strong safety culture achieve greater safety, and it is management's responsibility to foster and maintain such a culture (Table 5.8l).

TABLE 5.8l
Management Responsibilities

- Put organization structure with necessary authority
- Define applicable requirements
- Provide documentation infrastructure
- Foster safety culture

HAZARD AND RISK ANALYSIS

The standard requires that safety requirements should be determined by analyzing the risks posed by the totality of a manufacturing system and its control system. The analysis consists of three stages: hazard identification, hazard analysis, and risk assessment.

A hazard is defined as a potential source of harm. A manufacturing system and its control system may pose many hazards, each carrying its own risk. In determining the necessary overall risk reduction, the risk posed by each hazard must be considered. The importance of hazard identification cannot be overemphasized, because the risks associated with unidentified hazards cannot be reduced.

Hazard identification is unlikely to be effective if carried out by an individual; therefore, it is preferable to have a team whose members are chosen to bring complementary viewpoints to the process. A well-managed team with defined objectives is more effective in performing hazard analysis than a single individual.

IEC 61508 and 61511 are performance-based standards; as such, they do not mandate any specific safety system architecture or risk assessment procedures. However, they provide guidance in the areas of risk assessment and risk reduction. Following are some of the risk assessment and SIL determination concepts as outlined in the IEC 61511 standard Part 3. Detailed descriptions of these techniques are beyond the scope of this section. Readers are advised to refer to the IEC 61511 standard or the textbooks listed in the References section.

As Low as Reasonably Practicable (ALARP)

The ALARP principle may be applied during the determination of tolerable risk and safety integrity levels. However, it is not in itself a method for determining safety integrity levels. Tolerable risk implies that it is not possible to achieve absolute safety. A level of risk may be considered tolerable, in the light of the benefit gained in taking the risk, provided it is as low as reasonably practicable.

The ALARP triangle is divided into three regions with the width at any point indicating the magnitude of the risk (Figure 5.8m). Risk class I represent risks that cannot be justified except in extraordinary circumstances. Risk class III represents risk that is so low as to be negligible and is thus acceptable without any further risk avoidance measures. Risk class II in the middle represents risk that can only be tolerated if measures have been taken to reduce it to as low as reasonably practicable. This means that the cost of mitigating the risk is disproportionate to the benefits gained.

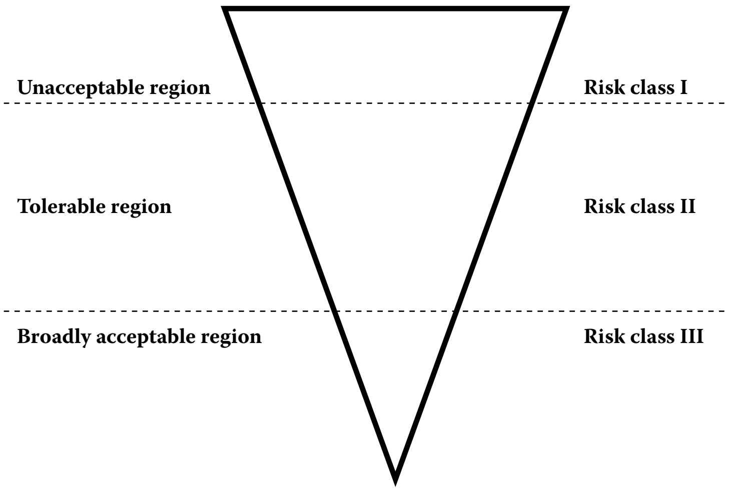
The concept of ALARP can be used when qualitative or quantitative risk targets are adopted. In order to apply the ALARP principle, it is necessary to define the three regions in terms of the probability and consequence of an incident. This definition would take place by discussion and agreement between the interested parties, such as those producing the risks, those exposed to the risks, and safety regulatory authorities.

Table 5.8n is an example of the three risk classes for a number of consequences and frequencies. After determining the tolerable risk target, it is then possible to determine the safety integrity levels of safety-instrumented functions.

Required Safety Integrity Level

The IEC 61511 standard Part 3 specifies a number of ways of establishing the required safety integrity levels for a specific application. The methods selected for a specific application depends on many factors, such as:

- Application complexity
- Guidelines from regulatory authorities
- Nature of the risk and the risk reduction requirements
- Experience and skills of the persons available to undertake the work
- Information available on the parameters relevant to the risk



Risk class I: Risk cannot be justified except in extraordinary circumstances

Risk class II: Risk is tolerable only if:
—Further risk reduction is impracticable or its cost is grossly disproportionate to the improvements gained
—Society desires the benefit of the activity given the associated risk

Risk class III: Level of residual risk regarded as negligible and further measures to reduce not usually required

Note: There is no relationship between risk class and SIL

FIG. 5.8m
Tolerable risk and ALARP.

More than one method may be used in an application. A qualitative method may be used first, followed by a more rigorous quantitative method, if needed. Qualitative methods outlined in the standard include:

- Safety reviews
- Checklists
- What if analysis
- Hazard and operability (HAZOP)
- Failure mode and effects analysis
- Cause-consequence analysis

Hazard and operability analysis is one of the more widely used techniques. It identifies and evaluates hazards in process plants and nonhazardous operability problems that compromise its ability to achieve design productivity. Table 5.8o is an example of the results of a HAZOP analysis.

TABLE 5.8n
Example of Risk Classification of Incidents

Probability	*Catastrophic*	*Critical*	*Marginal*	*Negligible*
Frequent	I	I	I	II
Probable	I	I	II	II
Occasional	I	II	II	II
Remote	II	II	II	III
Improbable	II	III	III	III
Incredible	II	III	III	III

The probability of occurrence are defined as:
Frequent: Many times in the system's lifetime
Probable: Several times in the system's lifetime
Occasional: Once in the system's lifetime
Remote: Unlikely during the system's lifetime
Improbable: Very unlikely
Incredible: Absolutely improbable

The consequences are defined as:
Catastrophic: Multiple loss of life
Critical: Loss of a single life
Marginal: Major injuries to one or more person
Negligible: Minor injuries

Note: The risk classes are application dependent

TABLE 5.8o
Example of a HAZOP Report

Item	*Deviation*	*Cause*	*Consequence*	*Safeguards*	*Recommendations*
Reactor	High level	Failure of control system	High pressure	Operator	
	High pressure	• High level • External fire	Release to environment	– Alarm, protection layer – Fire deluge system	Evaluate conditions for release to environment
	Low flow	Failure of control system	Excess pressure	Operator	Open pressure release valve manually

Semi-Quantitative Risk Analysis Techniques

An estimate of the process risk can be made by a semi-quantitative risk analysis procedure that identifies and quantifies the risks associated with potential process accidents or hazardous events. The results can be used to identify necessary safety functions and their associated SIL in order to reduce the process risk to an accepted level. Following are the main steps of this technique, where the first four steps can be performed during the HAZOP study:

- Identify process hazards
- Identify safety layer composition
- Identify initiating events
- Develop hazardous event scenarios for every initiating event
- Ascertain the frequency of occurrence of the initiating events and the reliability of existing safety systems
- Quantify the frequency of occurrence of significant hazardous events
- Integrate the results for risks associated with each hazardous event

The above exercise leads to a better understanding of hazards and risks associated with a process and leads to the identification of safety functions needed to reduce risks to acceptable levels.

Risk Graphs

The use of a risk graph is a method for evaluating safety integrity levels, which is illustrated in Figure 5.8p.

This method focuses on the evaluation of risk from the point of view of a person being exposed to the incident impact zone. In a risk graph there are four parameters to characterize a potential hazardous event: consequence, frequency of exposure, possibility of escape, and likelihood of events. In assessing the consequence severity, the following are considered:

- Potential for injury or fatality
- Possibility of the exposed person recovering and returning to normal activities
- The effects of injury: acute or chronic

The resulting safety integrity levels are shown in vertical columns.

SAFETY SYSTEM CERTIFICATION

Safety system certifications objectively assess the reliability and availability of critical control and safety shutdown systems and related products. Following are some of its advantages:

- Allows making informed decisions when choosing a product for a specific application
- Allows products and systems certified against standards
- Allows installing certified products and achieves recognized levels of process safety
- Gives the manufacturers of safety systems the opportunity to improve their products
- Gives suppliers of safety systems competitive advantage through documented product quality and reliability

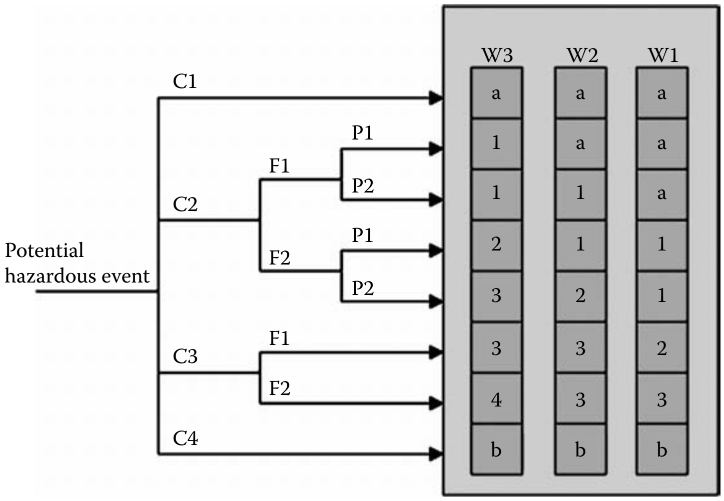
Safety regulations came after the start of industrialization when steam engines and boiler explosions caused many deaths. For over 127 years, TUVs have been in the forefront of inspection and certification of safety related systems in Germany.

The TUV certification process is very exhaustive and covers everything from the formulation and documentation of the original design concepts to the manufactured product and its suitability for a defined application. TUV is not a single entity, but consists of a number of independent regional organizations. Among them, TUV Rheinland, TUV Product Services (a part of TUV Suddeutschland), and TUVIT (part of RWTUV) are most active in certifying safety-related systems.

A research project started in the early 1980s by the TUVs on computer-based safety systems resulted in a document that led to the German safety standard DIN/VDE 0801. Until recently, the TUVs certified systems were based on the DIN standard (AK Class 1 to 7). They are now certifying systems based on IEC 61508 standards.

Every supplier that pursues this market has systems certified by one of the TUVs. This condition has been viewed favorably by both users and suppliers of safety systems, where users wanted reassurance of having a system certified by a qualified agency, and suppliers were willing to pay the necessary fees in order to differentiate themselves from competitors.

It should be emphasized, however, that some of the certified equipment has extensive restrictions. These restrictions effectively represent requirements not met by the manufacturer. When this happens, the user must compensate by adding externally wired equipment or user-programmed logic.



a = No special safety requirement
b = A single safety system is not sufficient
1, 2, 3, 4 = Safety integrity level

Consequence of event:
C1 = Minor injury
C2 = Serious permanent injury to one or more persons, death to one person
C3 = Death of several people
C4 = Very many people killed

Frequency of exposure:
F1 = Rare to more often exposure in the hazardous zone
F2 = Frequent to permanent exposure

Possibility of escape:
P1 = Possible under certain conditions
P2 = Almost impossible

Probability of occurrence:
W1 = Slight probability (less than 0.3 pa)
W2 = Medium probability (greater than 0.3 pa but below 3 pa)
W3 = High probability (3 pa or above)

FIG. 5.8p
Example of a risk graph.

This extra work and extra expense should be estimated before selecting the equipment.

Other organizations involved in safety system certification include INERIS, Factory Mutual, UL, BASEEFA, and many other national organizations. INERIS is a French research organization with close links with French Ministries of the Environment and Industry. It has been involved in safety system certification since its inception in 1992. INERIS has little presence outside France, and its clients are predominantly French manufacturers.

Factory Mutual, whose headquarters are in Norwood, MA, has a long history in product testing and approval. The company services its clients through offices around the globe and interlaboratory agreements with testing facilities in North America, Europe, Asia, Australia, and Latin America.

In late 1997, Factory Mutual and TUV Product Services signed a collaborative agreement enabling safety system manufacturers to obtain global certification of their products at either laboratory. The two labs are also working together to develop common interpretation and certification procedures for the IEC 61508 standard.

As the two organizations have name recognition in the different regions of the world, safety certification from either organization will be more widely accepted than before. The two organizations will be able to work in parallel, when required, to reduce the time required for the certification process and time to market. Users with manufacturing facilities around the world will be able to install safety systems in more countries without recertification.

It is essential that users consider only those systems that have been certified by a TUV or a similar organization for safety applications. To get full details of specific restrictions a user should demand a copy of the safety certification report and safety manual from its prospective supplier.

MAJOR TRENDS

As users are getting more knowledgeable about safety issues, they are performing more thorough safety analysis to determine their needs for each application more accurately. They are looking for reduction of risks by increasing their focus on overall safety. They would like their safety systems to satisfy their needs in a more cost-effective way by closer integration of safety with control systems. They are looking for a flexible architecture and with more scalability; increased functionality for modifying alarm limits based on process conditions; and orderly shutdown procedures in case of emergency.

Overall Safety

The main cause of an SIS failure is not the failure of logic, but the failure of field devices. There has been a significant advance in the development of the architecture of logic solvers with voting circuits and advanced diagnostics. However, they do not address over 90% of the causes for failure, which are due to the failure of field devices. What is needed is an integrated safety approach, where field devices are well integrated with the programmable safety systems.

Today, a protective system should be able to check the health of the process inputs and outputs (I/O). In fact, it needs to incorporate the I/O components in its overall design. They include:

- Sensor validation
- Environment condition monitoring, such as temperature and humidity that can cause sensor degradation
- Impulse line blockage monitoring

Failures of electronic components are frequently due to environmental conditions. Many electronic device failures are due to elevated humidity and temperature, which need to be monitored closely. Sensor calibration is also becoming an integral part of a safety system. Use of protocols, such as HART, Profibus, and Foundation Fieldbus, allows for remote monitoring, diagnostics, and validation.

Intelligent valves and digital valve positioners can also contribute to improved safety. With valves, the most frequent safety problems is when the valve does not respond to the trip demand, which is often caused by stem seizure or packing failure. The newer control valve designs, such as TUV certified valves, have reduced the probability of these and are available on the marketplace. The SIS-based valve design includes the feature of testing for limited valve movement during normal operation.

Separation from the Control System

It has long been the practice of many users to keep controllers used for safety completely independent of those used for control and optimization. In the past, controllers used for safety came from specialized manufacturers that added extensive diagnostics and received TUV safety certification. There was no choice but to use completely different systems for control and safety. Some users even mandated the purchase of control and safety systems from different manufacturers.

There are many other reasons to place the safety and control functions in different control systems. They include

- Independent failures: Minimizing the risk of simultaneous failure of a control system along with the SIS
- Security: Changes in a control system not causing any change or corruption in the associated SIS
- Special requirements for safety controllers: Special features, like diagnostics, certified fail-safe response, special software error checking, protected data storage, and fault tolerance

The IEC 61508 safety standard is somewhat ambiguous on this issue; it generally recommends separation but does not mandate it. In contrast, the IEC 61513 standard for the nuclear industries mandates physical separation of control and safety functions.

Today, a number of users in process industries are finding logical reasons for using the same systems for control and safety functions, because that will reduce the integration problems caused by the different programming procedures, languages, installation requirements, and maintenance. There is always the risk caused by communication problems resulting from integration of different software languages and procedures. It is preferred by the users to have the same service and support requirements for both the control and safety functions.

Control system and SIS suppliers now offer similar systems for both functions, including their HMI, configuration procedures, programming languages, and maintenance procedures. The systems may be physically separate but are similar and are provided with a common operator interface. They may communicate with each other, but are provided with adequate protection from corruption of one by the other (Table 5.8q).

Flexibility and Scalability

The installed base of systems for critical control or safety shutdown is largely either TMR (2oo3) or duplex (1oo2D) systems. However, other architectures, such as quad (2oo4D), hybrid 2oo4/1oo2, and other combinations, are also available. Increasingly, suppliers are offering configuration flexibility, where the user has the choice of combining two or more safety controllers to reduce failure rate and increase availability.

TABLE 5.8q
Advantages of Closer Integration with Control Systems

- Common data mapping
- Increased security
- Similar engineering tools
- Significant reduction in integration efforts

Safety controllers are also becoming more scalable. They are getting smaller, where one controller handles a limited number of I/O, but a number of these controllers working together can handle a much larger application. This is a saving for the users, as they no longer have to purchase large and expensive systems for their smaller applications.

Function Blocks

A state-of-the-art safety system provides facilities for simple sequencing (usually without looping) to allow orderly shutdown of a process on detection of a failure condition. Although orderly shutdown in case of some alarm conditions is often controlled by the basic control system, also incorporating it in the safety system reduces the risks.

Rich function blocks make it easier to configure functions, such as trip levels, deviation percentage, pretrip alarm, and degradation behavior. They make it easier to bypass specified alarms during start-up. Traditionally, cause-and-effect matrices (CEMs) are implemented in ladder logic. Rich functions blocks make their applications easier and support more intuitive presentations at run time.

SAFETY SYSTEM SELECTION

Before embarking on the selection procedure, a user should do a study to determine the safety protection requirements for the application. The requirements should conform to safety standards and meet the guidelines of regulatory agencies such as OSHA and EPA. Outside help from a third-party consultant or an established safety system supplier is strongly recommended if the necessary experience in safety study and risk reduction procedures does not exist in-house.

Conventional control systems provide the first line of defense against hazardous conditions. Incorporating alarming and effective shutdown procedures in the basic control systems can considerably reduce the risks. Further reduction or elimination of a number of risks is possible by adding protection measures such as hardware interlocks, relief valves, and improved operator access for manual intervention. In many situations, containment systems such as dikes and firewalls can considerably reduce the effects of an accident. Programmable safety systems are needed if these actions do not reduce the risks to an acceptable level.

Once the need for a programmable safety system is established, the user should also take into account the requirements of the certifying organization. These requirements specify the allowable operating conditions and requirements for redundant I/O and for initiating shutdowns in case of a processor failure.

TABLE 5.8r
System Selection Criteria

- TUV Certification and Restrictions
- Required Speed of Response
- Product Maturity and Installed Base
- Ease of Integration to Your Control System

TABLE 5.8s
Supplier Selection Criteria

- Knowledge & Experience in HAZOP Analysis
- Knowledge of Regulatory Requirements
- Industry Application Expertise
- Local Support

Speed of response can also be an important consideration for some applications, and it must be considered along with the response time of all the field instruments (Table 5.8r).

In making a safety system evaluation, the user should take into account its installed base and product maturity. The ease of integration of the safety system with the control system is also an important consideration. Providing the same interface and the same configuration software for the two systems is also recommended, because it will help reduce the learning efforts of engineers and operators. The control system should be able to monitor the status of the safety system and to monitor real-time data in order for the two systems to work in a unified fashion.

When selecting suppliers, their knowledge and experience in safety analysis and regulatory requirements should be considered, including experience in applying safety systems to similar applications. Finally, one should also consider the availability of local support and of spare parts (Table 5.8s).

ACKNOWLEDGMENTS

The help and suggestions made by Dr. William M. Goble of Exida are acknowledged.

ACRONYMS

ALARP	As Low As Reasonably Practicable
ANSI	American National Standards Institute
CEM	Cause-and-Effect Matrix
DCS	Distributed Control System
DIN	Deutsches Institut Normung (German Standards Institute)
EPA	Environmental Protection Agency
HAZOP	Hazard and Operability
IEC	International Electrotechnical Commission
I/O	Input/output
ISA	Instrumentation, Systems, and Automation Society
OSHA	Occupational Safety and Health Administration
SIL	Safety Integrity Level
SIS	Safety Instrumented System
TMR	Triple Modular Redundant
TUV	Technischer Uberwachungs Verin (Technical Inspection Association)

References

1. IEC 61508 Parts 1 to 7, "International Standard, Functional Safety of Electrical/Electronic/Programmable Electronic Safety-Related Systems," Geneva, Switzerland: International Electrotechnical Commission, 2000.
2. IEC 61511 Parts 1, 2, and 3, "International Standard, Functional Safety: Safety Instrumented Systems for the Process Industry Sector," Geneva, Switzerland: International Electrotechnical Commission, 2003.
3. ANSI/ISA-84.00.01—2004, "Standard, Functional Safety: Safety Instrumented System for the Process Industry Sector," Research Triangle Park, NC: ISA, 2003.

Bibliography

ARC Report, *Critical Control and Safety Shutdown System Strategies,* Dedham, MA: ARC Advisory Group, 1999.

Goble, W. M., *Control System Safety Evaluation and Reliability,* Research Triangle Park, NC: ISA, 1998.

Gruhn, P. and Cheddie, H. L., *Safety Shutdown Systems: Design, Analysis, and Justification,* Research Triangle Park, NC: ISA, 1998.

Marzal, E. M. and Schrapf, E. W., *Safety Integrity Level Selection,* Research Triangle Park, NC: ISA, 2002.

5.9 Relays

H. C. ROBERTS (1970, 1985) **B. G. LIPTÁK** (1995) **W. P. DURDEN** (2005)

Available Capacities:	Relays can handle microvolts to kilovolts, microamperes to kiloamperes, with response speeds from milliseconds to any longer period
Power and Frequency:	Relays can be actuated by microwatts but usually use milliwatts. They will operate from DC to RF
Environmental Limits:	Some relays can operate from –50° to 400°F (–45° to 204°C), from vacuum to 400 PSIG (2.8 MPa), and to 100 G shock or vibration
Sizes:	From 0.03 in.^{3} (0.47 cc) up
Cost:	Depending on size and capability, from about $10 to $100
Partial List of Suppliers:	ABB (www.ssac.com) Agastat (www.agastat.com) Allen-Bradley Co. (www.ab.com) Allied Controls (www.alliedcontrols.com) American Relays (www.americanrelays.com) American Zettler (www.azettler.com) Astralux (www.astralux.co.uk) Axicom (www.relays.tycoelectronis.com/axicom) Baocheng (www.qunli-relay.com.cn) Barnbrook Systems Ltd. (www.barnbrook.co.uk) BLP (www.blpcomp.com) CCI (www.completecontrols.com) Celduc Relais (www.celduc-relais.com) CII (www.ciitech.com) CIT Relay (www.citrelay.com) Continental Industries (www.ciicontrols.com) CP Clare/SRC (www.clare.com) Crompton (www.cromptoninstruments.com) Coca Enterprise Co. Ltd. (www.demax.com) Crydom (www.crydom.com) Cutler Hammer (www.cutler-hammer.eaton.com) Danfoss (http://uk.ic.danfoss.com) Daquan Relay Factory (www.chinadaquan.com) Deltrol Corp. (www.deltrol.com/controls) Dold Industries, Ltd. (www.dold.co.uk) Eagle Signal Controls (www.dancon.com) EAO Switch Corp. (www.eaoswitch.com) Furnas Electric Co. (www.sea.siemens.com) Greenwich Electronic (www.geirelays.com) Goodsky UK (www.goodsky.co.uk) Honeywell Inc. (www.honeywellsensor.com) Idec (www.idec.com) IMO (www.imopc.com) International Rectifier (www.irf.com) Inventa Electro Systems Nasik Private Limited (http://business.vsnl.com/chaitnya) Jennings (www.jenningstech.com)

KOYO (www.automationdirect.com)
Magnecraft & Struthers Dunn (www.magnecraft.com)
Meder (www.meder.com)
NAIS Matshushita (Aromat) (www.aromat.com)
Nuova Hi-G Italia; OEN (www.oenindia.com)
Olympic Controls (www.occorp.com)
Omron (www.omron.com)
Pickering (www.pickeringrelay.com)
Pioden (www.precisionrelays.co.uk)
Reliance Electric Controls (www.reliance.com)
Siemens (www.siemens.com)
Square D Co. (www.squared.com)
SRC Devices (www.srcdevices.com)
Teledyne Relays (www.teledynerelays.com)
Tianbo (www.tianbo-relay.com)
Tyco (http://relays.tycoelectronics.com)

INTRODUCTION

An electrical relay is a device that changes the state of a circuit in response to it being energized or de-energized by that or another circuit. In addition to electrical relays, there are pneumatic, fluidic, and other varieties. The relays described here are electromechanical or solid-state designs. Solid-state relays are only briefly mentioned in this section, because they are discussed in detail in Section 5.10.

The function of a relay is to open or close an electrical contact or a group of contacts in response to a “signal,” which is a change in some electrical condition. The contact closures of the relay can be used to select other circuits or to turn on or off various devices or operations.

More generally, a relay may be considered as an amplifier and a controller. It has a power gain that is defined as the ratio of output power handled to input power required. Thus, a relay may require a coil current of 0.005 A at 50 V but can control 2500 W of power—a gain of 10,000. As it will be discussed in this section, among the many forms of relays a variety of specific functions are available.

RELAY TYPES AND FEATURES

Electromechanical and electrothermal relays are discussed in this section. They are available in a variety of forms, but the basic characteristics that are described here are common to all of them.

Relays control the flow of electric current by means of contacts, which may be open or closed. When these contacts are open, they create a very high resistance (megaohms), and when they are closed, their resistance is low (milliohms). They may have multiple contacts. As many as eight SPDT contact assemblies are available on relays that are in stock. Each contact assembly is electrically isolated from all the others, and the contacts are actuated in some definite sequence.

The actuating coil can be (and usually is) completely isolated from the controlled circuit. It may be actuated by an electric energy that is of an entirely different character from that of the controlled circuit; e.g., milliamperes of DC may control kilowatts of RF.

Each of the various mechanical relay structures has certain advantages and certain limitations. Some respond rapidly—in less than 1 ms—but cannot safely handle large amounts of power. Others handle large amounts of power but at a slower speed, and so forth. Nearly all forms can be obtained with open contacts, can be enclosed in dust covers, or can be hermetically sealed in glass tubes or in vacuum-degassed plastic. Some glass tubes are evacuated for handling extremely high voltages. Some have contact assemblies suited for handling radio frequency voltages and to protect against capacitive cross-coupling.

There are also “solid-state relays,” which use transistors, SCRs, triacs, and similar components as current-control elements. In these devices the controlled circuit is usually isolated from the controlling circuit by a transformer (DC only), by an optical element, or perhaps by an electromechanical relay. These devices are covered in Section 5.10.

Special Relays

Some special relays that might be of interest to instrument engineers include sequence counters with transmitting contacts, opto-electronic relays in which coupling between actuator and circuit closure is a light beam, permitting elaborate interlocking of functions. These are often part of solid-state relays.

Also of interest are meter relays with microwatt sensitivity, ultrasensitive polarized DC relays (20 μW sensitivity), sensitive meter relays with two or more actuating values, resonant-reed relays for remote-control switching, coaxial RF relays, vacuum relays for extremely high voltages, multipole relays that can actuate 50 or more circuits simultaneously, polarized telegraphic and pulse-forming relays, voltage-sensitive relays for power system protection, phase-monitoring relays for protection in polyphase systems, impulse-actuated relays, instrumentation relays with negligible capacity between circuits, and crossbar switches that can select any one of 100 or more circuits. Solid-state devices with many of the same features are also available.

Relay Characteristics

The available relay designs include DC relays with contact capacities from microamperes to kiloamperes and coil energies from a few microwatts to several watts. In case of meter relays and amplifier relays, the actuating energy can be a small fraction of a microwatt. There are also AC relays handling from a few watts of power to many kilowatts. Relays can control both AC and DC potentials in the thousands of volts and frequencies from DC to RF. They can be made to respond to specific frequencies only or indiscriminately to all frequencies.

Many of the relay designs are for special applications. This discussion will be limited to the forms most often used in typical instrumentation work—principally small DC and AC relays, sensitive relays, miniature and sealed relays, and some small power relays. In the accompanying References, information may be found describing a wide variety of relay applications.

In selecting relays for particular applications, it is convenient to first consider the actuating energy that is available or needed to operate a relay. A rough classification, associating coil requirement with power-handling ability, is given in Table 5.9a. The values in this table are only approximate.

Relays can also be classified by operating function. Available energy often dictates relay choice. Examples are given in the following paragraphs.

Rating, Size, and Other Selection Criteria

Meter relays and ultrasensitive relays are actuated by very small energies (perhaps as small as a fraction of a microwatt). They are used in applications where only a very weak signal is available, such as at the output of a transducer or bridge.

General-purpose, or small control, relays are used in cases in which no great amount of power need be handled but flexibility in application and reliability in operation are both essential. Depending on the number of contacts, they usually require coil power of 200 to 800 mW, which may be either AC or DC. Typically they will control perhaps 5 A at 120 V per contact.

TABLE 5.9a
The Range of Relay Characteristics

Type of Relay	*Coil Power Required (Approximate)*	*Contact Capacity (Typical)*
Meter relay	As low as 1 mW	Low energy only
Ultrasensitive relay	Less than 10 mW	0.5–1.0 A, noninductive
Sensitive relay	From 10–60 mW	1.0 A noninductive typical
Typical crystal-can relay	100 mW per form C contact	0.5–1.0 A, noninductive
Transistor-can (TO-5) or miniature relay	150 mW per form C contact	0.5 A, noninductive
Reed relay	200 mW per form C contact	0.5–1.0 A, noninductive
General-purpose relay	200 mW per form C contact, up to 3 VA AC	10 A, 120 VAC
Small power relay	Usually AC coils, from 1–10 VA	30 A, 240 VAC

Small power relays are used in cases in which more power must be controlled. These usually have coils operating on 120 or 240 VAC, which can be controlled by sensitive relays. These small power relays have contact capacities up to 30 A, at voltages up to 600 V. Much larger relays are of course available (often called magnetic contactors or controllers).

Relays can also be classified by size. Subminiature relays may be as small as 0.13 in.^{3} (2.03 cc), and TO-5-size relays are the size of a TO-5 transistor case. In contrast, miniature plug-in relays measure approximately 1 in.^{3} (16 cc), and general-purpose plug-in relays are about 2 in.^{3} (31 cc) in size.

The type of *mounting* may be important (screw-mount, plug-in, or printed-circuit mounts); or the number of poles available in a single unit (as many as 48) may be the determining factor. Alternatively, extremely low (or perhaps extremely constant) contact resistance may be needed—mercury relays, or mercury-wetted contacts, have this characteristic. Reed relays and other subminiature units are suitable when many relays must be mounted on a single circuit board.

Relays may have appliance-grade, general-purpose, aerospace, or military-standard applications. There may be little difference in their reliability when they are used for their prescribed purposes, but the more expensive types will operate more satisfactorily under adverse conditions. Relays also vary in terms of mode of construction: There are clapper-type relays, telephone relays, solenoid-actuated types, reed relays, and many other forms.

Contact Configurations

Common terminology to describe the status of relay contacts is NO (Normally Open) and NC (Normally Closed). These refer to the relay contacts, when the relay is de-energized, as it is on the shelf. Relay contacts are also referred to as SPDT (Single-Pole-Double-Throw) or SPST (Single-Pole-Single-Throw). The term SPST (A and B in Figure 5.9b) means that the relay is provided with a single pair of contacts, which can be open or closed, similarly to a light switch. A Single-Throw relay would have either a Form A, Normally Open, contact or a Form B, Normally Closed, contact.

The term SPDT (C in Figure 5.9b) means that a pair of contacts (one NO, the other NC) is sharing a common input contact and will reverse its state when the relay is energized and will return to “shelf” state when the relay is de-energized. A Double-Throw relay is also referred to as having a Form C contact consisting of a NO and a NC contact. DPDT (Double-Pole-Double-Throw) relays have a pair of isolated contacts.

Electromechanical relays are produced with a wide range of contact arrangements, in various combinations. A number of these have become standard forms, and the ones that are used most frequently are shown in Figure 5.9b with their

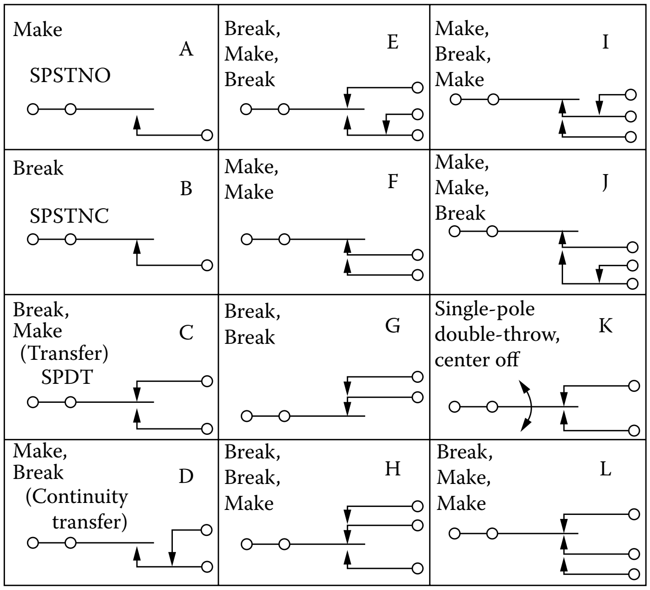


FIG. 5.9b
Standard relay contact configurations.

identifying code letters. The same relay contact assembly can be described as Single-Pole Double-Throw or SPDT; as make, break (continuity transfer); or has having form C contacts—all these terms mean the same thing. Four form C contacts means 4PDT, and so on.

Certain mechanical structures tolerate some contact pile-ups (the term for an assembly of leaf contacts) better than others. Small clapper-type relays seldom carry more than three form C contacts (although a modified clapper structure is used for high-density relays carrying as many as 48 contacts). Telephone-type relays may handle as many as eight form Cs, or perhaps four form Cs and one or two form Es or form Fs.

These contact assemblies permit the actuation of some circuits only after others have been actuated. This ensures the proper sequence of operation; there can be no "race of contacts." Solid-state relays seldom offer any wide selection of contact configurations; usually only form A or form B contacts are provided, unless the solid-state relay serves as a pilot for an electromechanical relay.

Electrical insulation between contacts can be very high, so that there is little leakage even at high voltages. Contact spacing can be wide, to protect from arc-over at high voltages. Capacity coupling can be kept low.

Mechanical Structures

Electromagnetic relays are actuated by magnetic forces that are produced by electric currents flowing through wire coils. In most relays of this type, the magnetic force moves an iron armature, while in a few other designs, especially in meter relays, a coil is moved in a magnetic field.

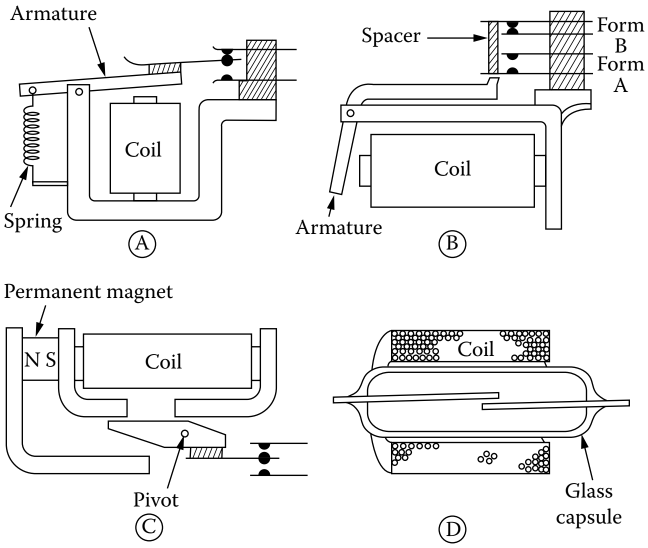


FIG. 5.9c
Typical relay structures.

The most widely used mechanical structures are shown in Figure 5.9c. The elements of a clapper-type relay are shown in part A of the figure, and the elements of a telephone-type relay are illustrated in part B. When there is no current flow in the coil, the relay armature is held away from the core by a spring. When current flows through the coil, the magnetic field that is produced pulls the armature toward the core, decreasing the air gap. As the air gap decreases, the pull usually increases, providing appreciably more contact pressure when the relay is energized than when the contact force is only that of the return spring.

Both the type depicted in part A and the type shown in part B can be used with either direct or alternating current. When AC is used, a shading coil is often added to smooth out the magnetic pull and to eliminate chattering. When AC current is used, the smaller air gap when the relay is actuated increases the impedance of the coil circuit and reduces the current flowing in that condition. Telephone-type relays are often fitted with shading coils (called "slugs") when used with DC current. This controls the speed of response of the relay.

Two other configurations in wide use are shown in Figure 5.9c, in parts C and D. The "balanced-force" mechanism is illustrated in part C. This structure has two mechanically stable positions. One of them is controlled by the permanent magnet, and the other is controlled by the magnetic force from the coil, which must be strong enough to overcome the pull of the permanent magnet. These relays can be made quite small in size, yet can be positive in action and affected only slightly by vibration.

Reed Relays The reed relay is shown in the partial sectional drawing of Figure 5.9c, in part D. Two reeds of magnetic metal are mounted in a glass capsule, which is itself installed within a coil. Current flowing through the coil produces a magnetic field, magnetizing the reeds and causing them to attract each other, to bend toward each other, and to make contact. The contacting surfaces are usually plated with precious metal contact alloy. The required spring action is provided by the reeds themselves.

Reed relays are among the fastest electromechanical relays. Some operate in less than 500 μsec. They are available in several contact configurations: They can be polarized, or they can be made into latching relays by the addition of small, permanent magnetic elements. Reed relays are available either with dry or with mercury-wetted contacts.

More than one reed assembly can be used in a single capsule, or several capsules can be operated by a single coil assembly. Reed relays are widely used in transistorized driver systems and in printed-circuit work, because of their small size, high reliability, and long life—a life of 100 million operations is not unusual.

Specialized Relay Structures The elementary magnetic structures described earlier can be combined to produce special-purpose relays.

If small permanent magnets are added to the relay, a relay armature may be forced in one direction for receiving a signal of one polarity and in the opposite direction when a signal of the opposite polarity is received. The elements of such a polarized relay are shown in Figure 5.9d, part A.

Mechanical or magnetic latching devices can be applied to produce a relay. This relay can permit either one of two circuits to be actuated, but not both at the same time. A simple form of this is shown in Figure 5.9d, part B. When one of the actuators is energized, its arm slides past the other arm, where it locks and is restrained from returning to its normal position even when de-energized. Energizing the other actuator reverses the procedure and latches the unit to the opposite side.

For certain applications, it is useful to be able to actuate one of two different circuits, depending on which of two signals is the larger—with little regard to the absolute magnitude of either signal. Two coils, with a single tilting armature, provide this ability (Figure 5.9d, part C). Usually, a weak spring is added to ensure that when no signal at all is present the relay will go to its neutral position.

Multiposition rotary switches can be driven by magnetic ratcheting devices, permitting the selection of any desired contact positions. Such stepping switches (or stepping relays, or “steppers”) are available in a wide range of positions and number of circuits. They can also be secured with electrical zero-reset or for continuous rotary operation. Stepping switches are used in many communications systems, in data-handling and materials-handling equipment, and in a large number of miscellaneous applications.

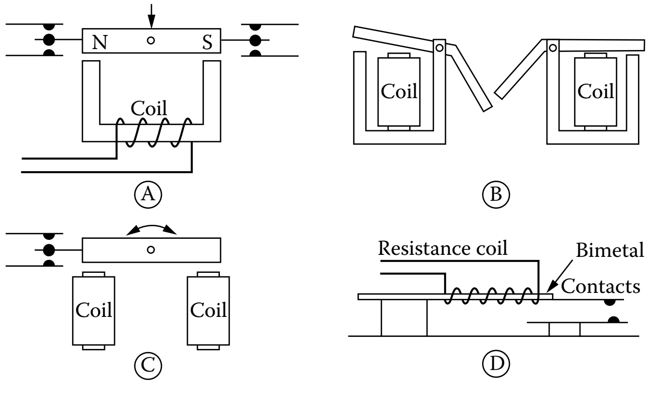


FIG. 5.9d
Some specific relay structures.

The lowest-current relays are the moving-coil meter-type relays. These utilize a d’Arsonval meter movement carrying delicate contacts. Sometimes the meter pointer is retained, and sometimes contact force is supplemented by magnetic contacts or by auxiliary coils. Such relays are susceptible to vibration and shock and to overloading. They are, however, capable of close adjustment for overvoltage or undervoltage applications and the like.

Time Delay Relays Thermal time delay relays can delay an action for a brief period after another action, in cases in which accuracy of timing is not critical. One form consists of a thermal bimetal strip wound with a resistance coil. The coil heats on the application of current, causing the bimetal strip to bend, which closes the electrical contacts (Figure 5.9d, part D).

In another popular form, the electric current flows through a stretched resistance wire, which expands and causes movement. These devices are somewhat affected by ambient temperature, and they cannot be recycled (returned to their initial cold position) instantly.

Another popular, low-cost time delay relay uses a small dashpot to delay the armature movement. Still another form (often called “slugged” relays) uses a heavy shading coil around a portion of the magnetic structure. This can produce a delay of a fraction of a second on opening or closing or on both actions. It is especially useful to protect against a “race-of-contacts.” These are typically low-cost, low-accuracy time delay devices. Thermal types can provide higher accuracy and repeatability, at a higher cost. Time delay relays are described in detail in Section 5.12.

Contact Materials

A variety of contact materials are available, with characteristics suitable for the various applications. For very low-current, low-voltage applications (dry circuit), it is essential to select contact materials that do not oxidize, develop insulating coatings, or erode mechanically. Some precious metals (such as gold and platinum) and some proprietary alloys satisfy these requirements. Such contacts are used in choppers and in meter relays, in which contact-sticking can be a serious problem.

Silver and silver-cadmium contacts withstand fairly high currents without overheating, but they tend to form coatings (oxide and sulfide coatings) that, while conductive, do have appreciable resistance. Tungsten-alloy contacts usually resist pitting and erosion when used at high voltages.

Mercury-wetted contacts can be expected to have higher current ratings and lower contact resistance than do dry metal contacts of the same size. Similarly to mercury-pool contacts, they are usually less noisy and display far less bounce. Dry metal contacts may fail by welding together; mercury contacts seldom do this. Relays with mercury contacts must usually be mounted in a nearly vertical position, and they are often vibration-sensitive.

Contact Shape and Mounting

The shapes of contacts depend on their use. Heavy-current contacts are usually dome-shaped. Low-resistance, small-current contacts are often crossed cylinders and are so placed that they wipe against each other. In wire-spring relays the round wires themselves, plated locally with contact material, form the contacts. They thus are long-lived and inexpensive. In reed relays, the flat strips, which are the reeds, also form the contacts. They may be shaped at the ends and usually are plated for good contact.

Contact mounting is an important part of relay design. In multicontact relays it is essential that all contacts be able to bear properly on their mating contacts, without interference. In low-voltage applications it is usually desirable that a wiping contact be provided. The higher voltages can break through thin surface films. In nearly all relays it is desirable that contact bounce and chatter be minimized. Some forms of reed relays, in particular, are very fast yet do not bounce.

SELECTION AND APPLICATION

Among the many factors affecting the selection of relays are cost, physical size, speed, and required energy. In addition, more restrictive parameters, such as mounting limitations and open or sealed contacts, are sometimes required for safety and for protection against unfavorable ambient conditions. The catalogs of relay manufacturers usually list dozens of types, forms, and sensitivities. Although relays have been in use for more than a century, new forms are still appearing. Tables 5.9e and 5.9f list the features of some representative designs of open-contact and sealed-contact relays, respectively.

TABLE 5.9e
Typical Open-Contact Relays

Relay Type	*Coil Description*	*Contact Description*	*Mounting*
Sensitive DC relay	1000–10,000 Ω, 20 mW per form C	Up to 4 form C, 1/2 A	Screw
Plate-circuit relay	2000–20,000 Ω, 50 mW per form C	Up to 2 form C, 2 A	Screw
Small utility relay	AC or DC, 6–120 V, about 250 mW	Up to 2 form C, 2 A, 120 V	Screw
General-purpose relay	AC or DC, about 200 mW per form C	Up to 3 form C, 10 A, 120 V	Screw or plug-in
Small power relay	Usually 120 VAC 2–10 VA	Up to 3 form C, 30 A, 600 V	Screw

Selection

In order for relays to be applied satisfactorily, the relay functions must be clearly understood and relay characteristics must be established. The relay must be selected to fit the need, and the circuitry must be designed to properly couple the relay with the rest of the system. Thus, it is usual to begin the selection process by determining how much energy must be controlled and how much energy is available in the signal that operates the relay.

One must also consider the number of contacts needed. It may be necessary to use two cascaded relays, if signal energy is too small, or two paralleled relays to provide enough contacts. Ambient conditions must also be considered. One must answer such questions as: Are sealed relays needed? Is there a space problem? Is there a vibration, shock, or temperature problem?

Relay Circuitry If a suitable relay has been selected and is available for the projected use, the circuit problem must be considered. In general, the same criteria may be used in designing relay circuits as in designing other circuits, yet because of some basic relay characteristics, relay circuitry involves some special problems of its own. Among the most important of these

TABLE 5.9f
Typical Hermetically Sealed Relays

Relay Type	*Contact Description*	*Mechanical Size (inches)*	*Speed of Action (milliseconds)*
Midget plug-in	1 form C, 0.5–1.0 A	$^3/_4 \times 2$	2
Mercury-wetted plug-in	1 form C, 2 A	$1^1/_2 \times 3^1/_2$	5
Mercury-wetted plug-in	4 form C, 2 A	$1^3/_4 \times 3^1/_2$	5
Balanced-armature DC, sensitive	2 form C, 2 A	1 × 1 × 2	3 (shock-resistant)
Crystal-can relay	2 form C, 1 A, noninductive	0.8 × 0.8 × 0.4	1.5 (shock-resistant)
Reed relay, dry contact	1 form C, 12 VA	0.5 × 0.5 × 3	1
Transistor-can relay	2 form C, 1 A, noninductive	0.3 × 0.3 × 0.3	1.5
Mercury-pool power relay	1 form A or B	2 × 2 × 6	Slow, vertical mount, affected by vibration

are the problems of transients across relay coils and the problem of protecting relay contacts from sparking, arcing, and welding.

In low-current, low-voltage relay circuit applications, as in most communication or logic circuits, serious problems with contacts usually do not arise, and often the most difficult problem is simply to keep the contact surfaces clean under low-current (dry-circuit) conditions. On the other hand, when larger currents must be handled (especially if the load is inductive or if it experiences an appreciable inrush for any reason), steps should be taken to protect the contacts from the effects of arcing, sparking, or welding. Under certain conditions, arcs tend to develop between contact and case or between contact and mounting. This can be prevented by proper circuit design.

One can minimize welding of contacts from high inrush currents by using sufficiently large contact areas that are made of suitable materials. Occasionally, more drastic measures—such as parallel contacts—must be considered. Lamp loads and the starting of single-phase motors—both drawing high inrush currents—are examples of troublesome conditions.

Sparking at contacts or arcing resulting from interruption of an inductive load can be minimized by use of spark suppressors (surge protectors, contact protectors, and so forth). These often are simple resistor-capacitor (RC) circuits but may be special devices—discharge tubes, diodes, or other solid-state instruments. Mercury-pool relay contacts are often used for heavy currents because the circulation of mercury provides a clean contact surface for each operation. Large power relays may use double-break contacts, blowout devices, or other means that are not within the scope of this discussion.

Transient Suppressors Most relay coils possess enough inductance to produce large transients when their currents are interrupted. A 28 V DC coil can produce a 1000 V transient, which can endanger insulation. Such transients can be reduced by use of semiconductor devices, neon lamps, or (on DC relays) a short-circuited winding as an absorption device. The problem can be a serious one and should not be ignored.

In many relay applications, it is advantageous to momentarily apply an overvoltage to ensure fast and positive action and then to reduce the coil current to a lower value to avoid overheating. There are many methods for doing this. Two of these methods are shown in Figure 5.9g. Shown in part A is a circuit useful only for DC relays. A 24 V relay is fed from a 50 V source; the capacitor (C) is initially charged to 50 V, so that when the controlling contacts are closed, the inrush current is nearly double the normal coil current. After the capacitor has discharged to its normal value, the resistor (R) limits the coil current. If R is equal to the coil resistance, only the normal coil current flows. In this diagram, a resistor-capacitor transient suppression circuit is connected in parallel with the coil. Other types of suppressors might be used with equally good results.

Circuit B in Figure 5.9g can be used with either AC or DC coil voltage. It involves a positive-coefficient resistor (often a small tungsten lamp, or a barretter) in series with the relay coil. When the barretter is cold, its resistance is rather low, but after current has passed through it for a few seconds, its temperature (and, hence, its resistance) increases, and the coil current is reduced to its normal value. This diagram shows two Zener diodes connected across the coil to serve as transient suppressors.

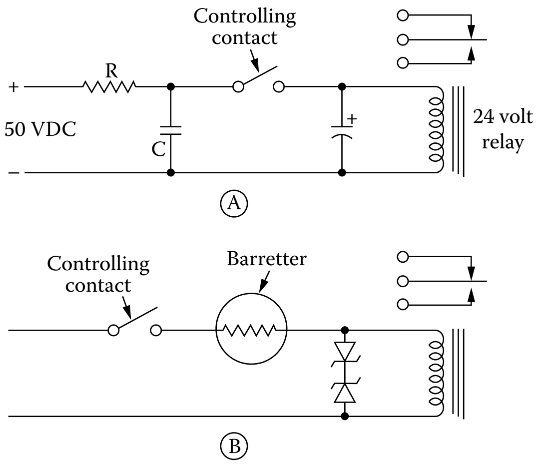


FIG. 5.9g
Typical coil circuits with transient suppressors.

Sometimes an unused relay contact is connected just to insert more resistance after a relay has been energized. It should be remembered that the inductance of an AC relay coil usually increases when the relay is closed, and this will reduce the coil current to some extent.

Relay users should not hesitate to consult the relay manufacturer for assistance in selecting and applying any of these devices. No catalog can possibly contain all the available data, and most manufacturers are happy to supply details on relay use.

RELATIVE COSTS

Small general-purpose relays can usually be purchased for about $10 each, with dust covers and simple contact configurations. More complicated contact assemblies, with hermetic seals, will cost more. If plug-in relays are used—and many general-purpose industrial relays are plug in—the socket will add several dollars to the cost and transient protection will cost a little more. The cost of high-sensitivity relays can reach $100. Installation charges are difficult to estimate, because they vary with locality, type of mounting, and type of base used.

Relays vs. Solid-State Devices

Relays compete primarily with silicon-controlled rectifiers, silicon switches, and transistors, which are discussed in more detail in Section 5.10.

Electromechanical Advantages

Electromechanical relays outperform such solid-state devices in some respects. They offer both extremely low resistances

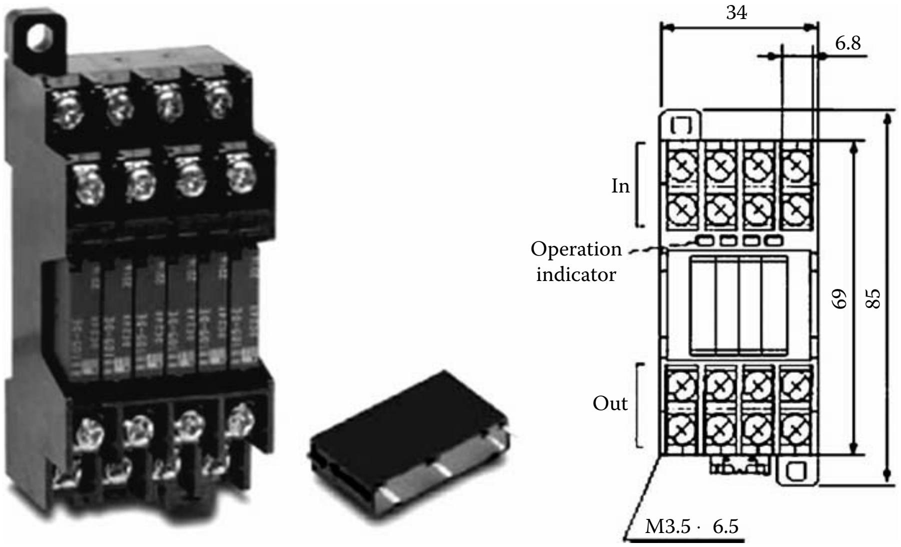


FIG. 5.9h
Card relay package and schematic. (Courtesy of KOYO Automation Direct.)

(and thus low voltage drop—about 1 μV with gold plated contacts) when closed and extremely high resistance when open.

They also provide essentially complete isolation of the controlled circuit from the controlling circuit without auxiliary links, and they can provide essentially simultaneous actuation of several circuits or actuation of several circuits at short intervals, without added cost. Some relays can handle extremely high voltages at a range of frequencies, and some can handle high currents in more compact form than can most solid-state devices. Relays provide positive circuit closures and can tolerate much abuse with only some shortening of life. Modern relays can also withstand the extremes of ambient conditions.

Some relays (in particular, high-density relays) are as small or nearly as small as the competing solid-state devices when installed. In some cases, they are cheaper to procure and to install than the competing solid state devices, and some—such as latching relays or steppers—can provide other advantages, such as a functional memory, which is not destroyed by power interruptions.

One form of high density design is the "card relay." This is a DIN rail-mounted package of four, six, or more individual relays. The relays are individually replaceable and can handle up to 5 A. An example of such a design is shown in Figure 5.9h.

Solid-State Advantages

For extremely high-speed operation, solid-state devices are superior to relays, because only a few relay forms will operate in less than a millisecond. Solid-state devices are more resistant to vibration than are relays but can be less resistant to some other environmental conditions. For the ultimate in compactness if cost is not a factor, and for logic rather than power-handling applications, integrated solid-state devices will usually be preferred.

Many relays operate on ordinary line power, in contrast with solid-state devices, which often require some other form—although this is frequently hidden in a packaged assembly. If only a few relays of different kinds and capacities are needed, electromechanical units may be more convenient for this reason alone.

CONCLUSIONS

Relay reliability depends upon their quality but even more on their correct selection. Relays carrying only small currents may be expected to operate millions of times before failure. Relays that control currents in the order of amperes necessarily wear faster. With proper contact protection, however, relay lives of 100,000–20,000,000 operations are common.

Reliability does depend upon selection of the proper relay; contact protection and transient suppression; the use of dust covers or hermetic seals when needed; proper circuit design; and observance of the usual rules of good practice as regards voltages, ambient conditions, and the like.

Bibliography

Asfahl, C. R., *Robots and Manufacturing Automation,* New York: Wiley, 1992.

Blake, B. M., "Host of Relays for Control Jobs," *Control Engineering,* January 1967.

Deeg, W. L., "Reed Switches for Fast Relay Logic," *Control Engineering,* 1967.

Fink, D. G., *Standard Handbook for Electrical Engineers,* 10th edition, New York: McGraw-Hill, 1968.

Gaylord, M. E., *Modern Relay Techniques,* New York: Transatlantic, 1975.

Kalpakjian, S., *Manufacturing Engineering and Technology,* Reading, MA: Addison-Wesley, 1992.

Kamm, L. J., *Understanding Electro-Mechanical Engineering, An Introduction to Mechatronics,* Piscataway, NJ: IEEE Press, 1996.

Kuphaldt, T. R., *Lessons in Electric Circuits, Volume IV (Digital), Chapter 5*, http://www.faqs.org/docs/electric/Digital/DIGI_5.html, 2000–2003.

Mason, C. R., *Art and Practice of Protective Relaying,* New York: Wiley-Interscience, 1956.

National Association of Relay Manufacturers, *Engineers' Relay Handbook,* 5th edition, Elkhart, IN, 2005.

Oliver, F. J., *Practical Relay Circuits,* New York: Hayden Book Co., 1971.

Roberts, H. C., "Microwatt Relays," *ISA Journal,* May 1966.

Stoffels, R. E., "Shift Registers and Stepping Switches for Sequence Control," *Control Engineering,* January 1967.

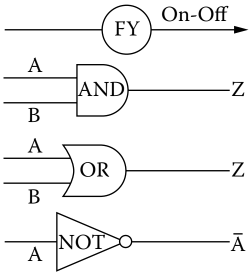
Strong, J. A., *Basic Digital Electronics,* London: Chapman and Hall, 1991.

Warrington, A. R., *Protective Relays: Their Theory and Practice,* Vols. I and II, London: Methuen, Inc., 1968 and 1978.

Yudewitz, N., "Back to Relay Basics," *Instruments and Control Systems,* February 1979.

5.10 Solid-State Logic Elements

P. M. KINTNER (1970) **R. A. GILBERT** (1985, 1995, 2005)



Flow sheet symbol

Components:	Size: 0.5 in. 1 in. 0.5 in. Type: TTL, CMOS, NMOS Power: ±12 V DC, 5 V DC Terminals: Solderable pins
Chassis Mounts:	Size: 2 in. × 1 in. × 1.8 in. Type: CMOS, thick film hybrid circuits Power: 12 V DC, 24 V DC, 48 V DC, 12 V AC, 24 V AC, 120 V AC Terminals: Plug-in modules
Single Boards:	Size: 8 in. × 12 in. × 1 in. Power: Input—120 V AC; output—5 V DC, 12 V DC, 24 V DC, 120 V AC Terminals: Stand-alone unit (RS-232 option)
Costs:	Components: $0.50 to $40; chassis mounts; $45 to $150; single boards: $60 to $560.
Partial List of Suppliers:	(Note: *A* = components, *B* = chassis mounts, *C* = single boards) Amtex Electronics (www.amtex.com.au) Action Instruments (www.actionio.com) (C) Analog Devices (www.analog.com) Omron (www.omron.com) Opton22 (www.opto22.com) Omega Engineering (www.omega.com) Magnecraft/Struthers-Dunn (www.magnecraft.com) International Rectifier (www.irf.com) (B,C) Carlo Gavazzi (www.carlogavazzi.com) Crydom (www.crydom.com) (B) Texas Instruments (www.ti.com) ON Semiconductor (www.onsemi.com) Intersil (www.intersil.com) Fairchild Semiconductor (www.fairchildsemi.com) Toshiba America (www.toshiba.com/taec/) STMicroElectronics (www.st.com) National Semiconductor (www.national.com) (A)

INTRODUCTION

With the advent of semiconductor technology and subsequent hybrid device fabrication techniques, solid-state logic devices have undergone dynamic application changes in the past 10 years. Today, it is common to have a sensing element directly coupled, if not integrated, to input logic operations and output logic process condition requirements. It is also now possible to employ "smart" solid-state logic elements that function as stand-alone control loops.

These integrated circuits (ICs) perform control switching operations without moving parts. Their advantages are: (1) high switching speeds, (2) life independent of the number of operations, (3) operation unaffected by dirt and certain corrosive atmospheres, (4) small size, and (5) low power consumption. They make the process variable measurement, examine that sensor data according to specific logic expectations, and then decide what, if any, alteration in the device output signal should be provided. Their disadvantages are: (1) sensitivity to electrical interference, (2) lack of input/output isolation, and (3) low power output driving capabilities.

These mini distributed control systems reduce the information noise that is forwarded to the process controller, provide an average or logic filtered response to a process variable disturbance, and greatly enhance the performance and speed of the entire process control loop. This section will briefly discuss the fundamental engineering science principles required to apply the analog, digital, and logic components of solid-state logic technology.

ANALOG CIRCUIT ELEMENTS

Although modern solid-state logic devices completely incorporate any required analog, digital, and logic circuits elements, instrumentation and controls personnel are still required to interpret circuit and function diagrams that contain isolated circuit elements as they relate to solid-state device implementation. These elements include resistors, capacitors, inductors, diodes, transistors, and operational amplifiers. Many common solid-state logic device control applications include external resistors, capacitors, and diodes. These elements as well as switching transistors are discussed below.

The resistor represents the simplest and most basic auxiliary component in a solid-state control circuit application. It is used for its ability to lower the potential, voltage, as electrons pass through it. Although the reader is encouraged to explore resistance, resistivity, and Ohm's law, from a digital circuit control's application perspective, an appreciation of the voltage drop concept is particularly useful. Figure 5.10c illustrates this point.

A P-N junction is formed by joining a small amount of *p*-type semiconductor material to a small amount of *n*-type material. Transistors (Figures 5.10a and 5.10b) are formed by the combination of two P-N diodes. PNP or NPN transistors differ from each other in the positions of the *p* and *n* layers and the polarity of the applied voltages and in the directions of the resulting current flows.

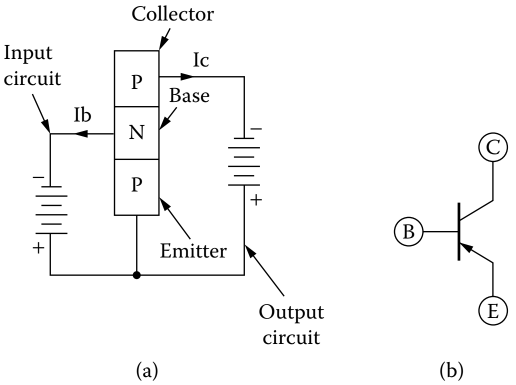


FIG. 5.10a
Combination of P-N junctions into a transistor. (a) Arrangement for the PNP transistor. (b) Commonly used circuit symbol for PNP transistor.

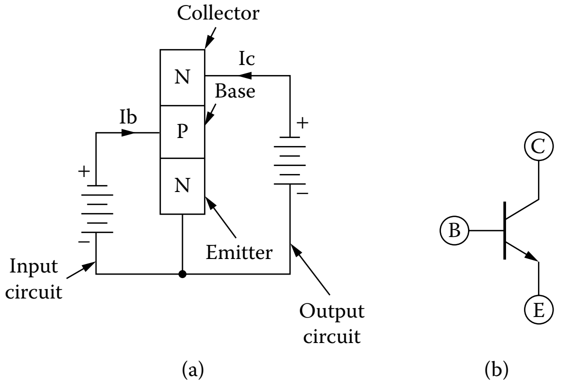


FIG. 5.10b
Combination of P-N junctions into a transistor. (a) Arrangement for the NPN transistor. (b) Commonly used circuit symbol for NPN transistor.

Transistor Switch

Figure 5.10c depicts the basic structure of a PNP transistor switch. The figure exemplifies several points to be presented in this section. However, initially the focus is on the voltage drop concept. The left side of the diagram shows a serial connection between two resistors, R_1 and R_2, and a parallel connection between R_2 and a set of normally open contacts, PSL 21. The figure indicates that there is a total voltage drop, V_T, when current flows between the node point above R_1 to the node point below R_2.

The discussion of this serial resistor and parallel contact often includes the terms voltage divide and pull up resistor. Both terms reflect the Ohm's law relationship that the sum of the voltage drops across R_1 and R_2 equals V_T but each term reflects a different perspective of the node point, (a), located between R_1 and R_2.

The term voltage divide brings the reader's attention to node point (a). The voltage at this point is determined using the voltage divide equation shown here as

$$V_{(a)} = V_T\,[R_1/(R_1 + R_2)] \qquad \mathbf{5.10(1)}$$

where the resistance ratio in the brackets is known as the voltage divide ratio. If, for example, the two resistors have the same value, then the voltage at the node point (a) would be one-half of the total voltage drop, V_T.

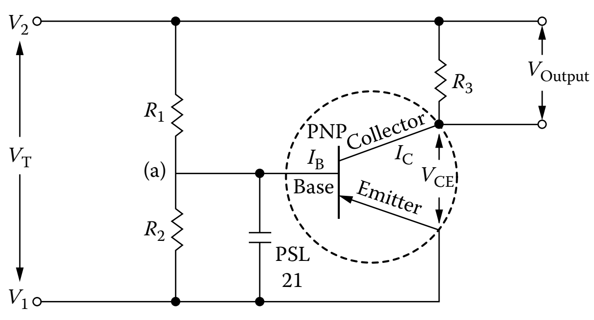


FIG. 5.10c
Voltage divide example in a PNP switching circuit.

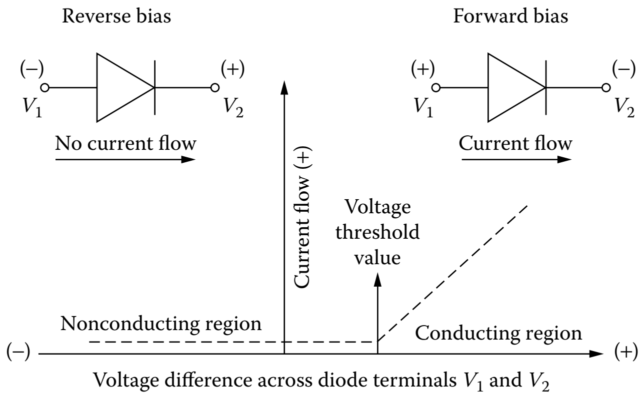


FIG. 5.10d
Diode voltage bias versus current characteristic curve.

The term pull up resistor draws attention to the fact that the voltage at node point (a) is higher than the voltage at the node point below R_2 as long as contact PSL 21 remains open. R_2 is now identified as a pull up resistor. The use of pull up resistors is very common in solid-state control circuit applications, and their role is to ensure a desired reference voltage at a specific input or output in the digital circuit.

Diodes and Their Switching

The diode represents the simplest semiconductor device digital control circuit element. It is formed when a small amount of *p*-type material and a similarly small amount of *n*-type material are joined to produce a single interface known as a P-N junction. Such a junction provides a comparatively low resistance for one direction of current flow and very high resistance for the flow of current in the reverse direction.

The diode circuit symbol representations are illustrated in Figure 5.10d. The figure shows the voltage assignments that define the reverse and forward bias conditions for a diode. If the voltage polarities depicted on the diode on the left side of the diagram are imposed, the diode is reverse biased and in a nonconduction mode. No current flows when a diode is reverse biased. The figure also presents the current versus voltage across the diode plot. This graph indicates that the current increases rapidly after a small threshold of voltage is exceeded.

A simplified diode switching application is presented in Figure 5.10e. The circuit translates the digital sensor voltage input signals to a control output voltage delivered across R_6. In this case, the voltage divide provided by R_4 and R_5 creates a reference voltage, V_{node}, that defines the bias on the two diodes. If a voltage lower than V_{node} is applied at Input *a* or Input *b*, the respective diode becomes forward biased, current flows, and a voltage drop across R_6, V_{output}, results. Thus, any control action generated by V_{output} occurs when either or both of the input sensors present the low signal to the circuit's respective input port.

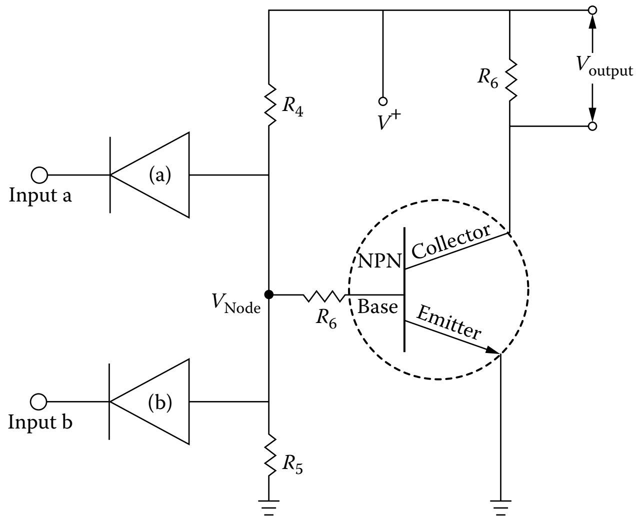


FIG. 5.10e
Circuit elements for a diode-transistor logic device.

Transistors

The transistor represents the simplest multiple junction semiconductor device employed in control logic element applications. Transistors are formed by the combination of two P-N junctions and contain two circuits: emitter-to-base and emitter-to-collector. The emitter-to-base circuit is generally referred to as the base circuit, and the emitter-to-collector circuit is known as the power, or output, circuit. The right section of Figure 5.10c shows a PNP transistor with its terminals appropriately labeled.

The manufacture and internal composition and action of the transistor is interesting but not of immediate value to the instrument engineer. However, three general characteristics (which can be observed by external measurements) are significant for logic applications or switching actions:

1. The emitter-to-base circuit controlling I_B in Figure 5.10c closely resembles a diode.
2. The emitter-to-collector circuit functions as an open/close switch.
3. The switch is open and the transistor is not conducting, I_C is zero, whenever the voltage difference between the emitter and the base is zero or whenever the emitter-to-base circuit is reverse biased.

Figure 5.10c illustrates a PNP transistor circuit as an interface to a low pressure-sensitive switch, PSL 21. Because of various design considerations it is usually desirable for the transistor to be conducting when no signal is applied and to stop conducting when a signal is applied.

Resistors R_1 and R_2 together with the normally open contact PSL 21 provide this function for the circuit. If properly selected, R_1 and R_2 generate the appropriate potential value at the base to yield full conduction in the emitter-to-collector circuit. In this state, I_C is at its maximum value.

The closing of contact PSL 21 alters the voltage at node point (a) enough to open the transistor, drastically reduce the value of I_C, and thus bring V_{output} to 0 V.

The proper selection of R_2 and R_3 depends on an understanding of the concept of saturation voltage. Examination of Figure 5.10c easily confirms that the voltage drop across the transistor, the collector voltage, V_{CE}, is defined as

$$V_{CE} = V_T - I_C R_3 \tag{5.10(2)}$$

where $I_C R_3$ is the voltage drop, V_{output}, generated when the PNP collector current, I_C, is not 0. Since I_C is related to the base current, I_B, by

$$I_C = \beta I_B \tag{5.10(3)}$$

with β defined as the amplification factor, V_{CE} can now be expressed as

$$V_{CE} = VT - \beta I_B R_1 \tag{5.10(4)}$$

The base current changes as a function of the voltage divide network that contains R_1 and R_2. As the potential at the transistor base is made more negative, I_B increases and V_C decreases to a minimum value. This minimum value, called the saturation voltage, is less than 0.2 V for typical switching transistors, and the effective resistance between collector and emitter at saturation is accordingly small (on the order of a few ohms).

By contrast, when I_B is reduced to 0, the collector-to-emitter resistance is that of a reverse-biased or nonconducting diode and is in the region of millions of ohms. The collector-to-emitter circuit operates to open and shut under the control of I_B with the output voltage, V_{output}, switching off and on accordingly.

Figure 5.10e also presents a transistor that interfaces input signals to a control output voltage. In this situation, the transistor is an NPN type that fundamentally operates the same way as the PNP transistor in Figure 5.10c. In both cases, the emitter-to-collector power circuit is turned on or off, and this maximum or minimum V_{output} value switching action depends on the flow of current in the emitter-to-base circuit within the transistor.

For the diode application in Figure 5.10e, the emitter-to-collector power circuit conducts and develops the desired V_{output} value when the process variable input signal forward biases either or both of the diodes. In this particular application, the process variable signal can be analog or digital, because the bias threshold voltage of the diodes determines the switching action. In fact, the diodes perform a rudimentary but effective analog-to-digital signal conditioning that allows an analog sensor to also operate as a limit alarm.

INTEGRATED CIRCUIT ELEMENTS

The term *integrated circuit* reflects the fabrication process that puts the functions of transistors, diodes, and resistors into a layer-constructed device. Before considering switching and logic applications for integrated circuits, the nature of the switching operations themselves will be defined. The general requirement for logic devices is that they generate on–off output control signals from on–off input signals in a way dictated by a logic operation. From a positive logic perspective, the five elementary and basic logic operations are:

1. AND: The output is on if all inputs are on.
2. OR: The output is on if one or more inputs are on.
3. NOT: The output is off if the input is on and vice versa.
4. NOR: The output is off if one or more inputs are on.
5. NAND: The output is off if all inputs are on.

Figure 5.10e represents a simple example of an individual component circuit that can be fabricated as an integrated circuit. For this specific example, the *IC* package would have five terminals available to the user with no access to the transistor or any of the individual resistors or diodes. Two of these terminal are reserved for the power and ground for the package, the third terminal provides access to V_{output}, while the remaining two terminals are assigned to Input *a* and Input *b*.

The reader is encouraged to review the overall function of the circuit in Figure 5.10e to confirm that it would perform an OR logic operation if the input signals were inverted before they were presented to the diodes.

Families of IC Switching

Although the fabrication of integrated circuits is not of immediate interest, it is important to appreciate that the technology options influence the applications. There are four main families of integrated circuit switching in common use for control:

1. RTL (resistance-transistor logic)
2. DTL (diode-transistor logic)
3. TTL (transistor-transistor logic)
4. HLL (high-level logic)

RTL was the first type of switching logic produced in integrated form. It basically replaced component circuits that were simple combinations of resistors and transistors. The primary advantage of this technology over its separate component equivalent is the size and reliability bonus that accompanies the integration process.

If the OR circuit in Figure 5.10e was manufactured as an *IC*, it would be an example of DTL technology. The major impact of integrated circuits accompanied the development of TTL technology. TTL devices have high switching speeds and are now common to most digital instrumentation and control applications. Table 5.10f is a summary of technology characteristics.

TABLE 5.10f
Integrated Circuit Characteristics

Type	*Resistance to Noise*	*Switching Speed*	*Cost*
RTL	Poor	Fair	Low
DTL	Fair	Good	Medium
TTL	Good	High	Medium
HLL	Excellent	Low	High

Some examples of TTL packages are provided in Figure 5.10g, which shows the pin configuration for the 7400, 7404, 7408, and the 7474 TTL packages. The 7400 and 7408 are know as 2 input quad NAND and 2 input quad AND packages, respectively. The 7404 is a hex inverter package while the 7474 is a dual D-type positive edge triggered flip-flop.

The four NAND devices in the 7400 and the four AND devices in the 7408 operate independently of each other. For example, if the proper operating voltages are applied to pins 14 and 7, a high input signal, logic "1," on pin 1 and pin 2 of the 7408 will generate the logic "1" signal on pin 3 independent of the signals that may be present on pins 4, 5, 9, 10, 12, or 13.

A similar statement is true for the 7400, with the exception that the logic "0" signal is generated on pin 3 of the 7400. Each inverter in the 7404 also operates independently. The logic "1" signal on pin 1 produces a logic "0" signal on pin 2 even though a logic "0" signal on pin 3 will simultaneously produce a logic "1" signal on pin 4.

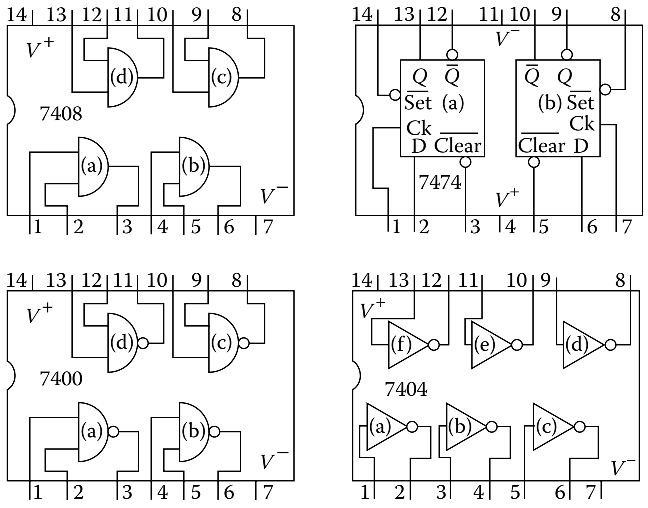
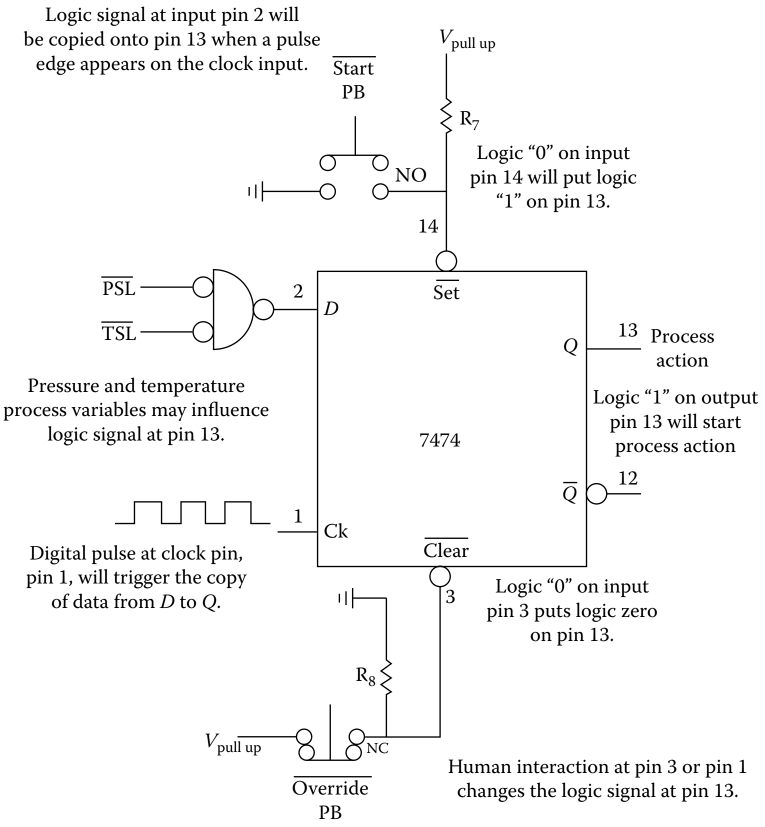


FIG. 5.10g
Pin configurations for the 7400, 7404, 7408, and 7474 TTL logic devices.


Logic signal at input pin 2 will be copied onto pin 13 when a pulse edge appears on the clock input.

Logic "0" on input pin 14 will put logic "1" on pin 13.

Pressure and temperature process variables may influence logic signal at pin 13.

Logic "1" on output pin 13 will start process action

Digital pulse at clock pin, pin 1, will trigger the copy of data from D to Q.

Logic "0" on input pin 3 puts logic zero on pin 13.

Human interaction at pin 3 or pin 1 changes the logic signal at pin 13.

FIG. 5.10h
Process variable sensors and pushbutton interface to a 7474 positive edge flip-flop.

APPLICATIONS

To illustrate the utility of integrated circuit devices, two simplified applications will be discussed.

Merging Human and Instrument Inputs

Figure 5.10h presents a function diagram that shows how logic devices can merge the human and instrument input requirements of a control scheme.

The 7474 solid-state logic flip-flop element shown is an integrated circuit that combines NAND and AND circuits to perform a specific switching operation at its outputs. The logic on pin 13, the *Q* output, will be at logic "1" or logic "0" depending on the signals at inputs 1, 2, 3, and 14. The logic signal on pin 12 is always the opposite of that presented to pin 13. The rules of operation for the 7474 are also provided in Figure 5.10h.

As suggested, the utility of the 7474 resides in its ability to combine the logic of human and process sensor inputs to direct a switching control action. The function diagram suggests an application that involves a process action that depends on the condition of two process variables, temperature and pressure, as well as human interaction with pushbutton inputs. The diagram indicates that a human override decision has priority over the process variable information. If the override pushbutton is engaged, a logic "0" is generated at the *Q* output and the process action will not be allowed to continue.

For this simple example, it is predicated that the temperature and pressure will not remain above their alarm values if the process action is discontinued. If that were not the case, an AND device could be added to the control scheme as a gate for the clock pin. The logic state at pin 13 would be fed back to serve as the gate pin signal on the AND. The AND rather than a NAND would be selected as the gate device, because a negative edge will be generated by the AND when the gate pin goes to logic "0."

The negative edge delivered to pin 1 ensures that the logic state at pin 2 will not be transferred to pin 13. This removes the possibility of a logic "1" signal on D immediately replacing the override generated logic "0" signal on *Q*.

The process action is started or restarted when the active low start pushbutton is engaged. At this point, the *Q* output is at logic "1" and will remain in that state when the start button is released. If the override pushbutton is not engaged, the process action will continue until both of the active low sensors detect that the respective process variables have dropped below their low-level limit values.

Again for simplicity, in this example, it is assumed that upon start-up the temperature and pressure immediately rise above alarm levels. In most applications this will not be the situation. For this example, one way to adjust to such a temperature or pressure lag is to use the AND gate discussed above and extend the time the gate pin detects the logic "0" signal.

Time Synchronization

Figure 5.10i presents an example of a data acquisition logic control circuit (sheet I) that deals with a designed time-dependent presentation of a logic signal for a control action. The simplified circuit depicts one of several sheets that describes the complete control scheme.

Sheet I suggests a complicated scheme that uses four sensor input signals to direct two output control signals, M-2,I-5 and P-2,I-6. The control scheme uses devices in single NAND and OR packages and a single 555. The 555 provides a pulse duration feature that is initiated by the proper combinations of the input signals. The 555 integrated circuit is a popular device (see Section 5.13) when adjusted pulse duration situations are required. The 555 is not a TTL device but is TTL compatible and is often used with TTL control circuits.

For this application, pin 3 provides, on demand, a logic "1" signal for a time determined by the R_{10} C_{10} circuit associated with pins 6 and 7. The output signal on M-2,I-5 is directed to sheet *M* as input 2, where it will activate an analog-to-digital conversion of a process variable measurement as shown in sheet *M*. The time needed to accomplish that conversion is provided by the extended logic "1" signal delivered to pin 12 on NAND device (d). To ensure proper transfer of the digitized variable value to its computer-based data logging system, a delayed signal to the data logger is provided to sheet P-2 via the R_{11} C_{11} circuit associated with output P-2,I-6.

The presentation of the data acquisition logic signal at pin 12 of NAND (d) depends on the signals that are delivered to the other 7400 NAND devices. All of these devices are operating as gates, with pins 2, 4, 13, and 9 serving as gate pins.

The reader may wish to review Figure 5.10i to confirm that the proper signals will not be presented to M-2,I-5 and P-2,I-6 unless input I-4,C-3 is at logic "1." In addition, inputs I-1,B-4 and I-2,D-2 as well as I-4,C-3 must be at logic "0." The diodes and inverter in line I-2,D-2 are needed to condition the signal arriving from an input circuit as documented in sheet D output 2, and the pull up resistor, R_9, together with the flip-flop action of AND device (a) and AND device (b) ensures that the output of the 555, pin 3, is off unless I-1,B-4 is really at logic "0."

The use of the 555 to control the duration of the logic signal presented to pin 12 of NAND device (d) emphasizes the range of skills instrumentation and controls personnel may have to master. In this specific situation, a 74121 is a TTL device that, on first glance, could be a substitute for the 555. However, the pulse duration requirement for this application is better suited when the 555 is used.

In general, Section 5.10 is presented from an application user perspective. The section provides information that introduces solid-state elements, reviews fundamental electronics concepts, and facilitates the interpretation of application circuits but does not cover the specific electrical and electronic considerations and details if the task is to actually design the circuit, as compared to designing or interpreting the control task.

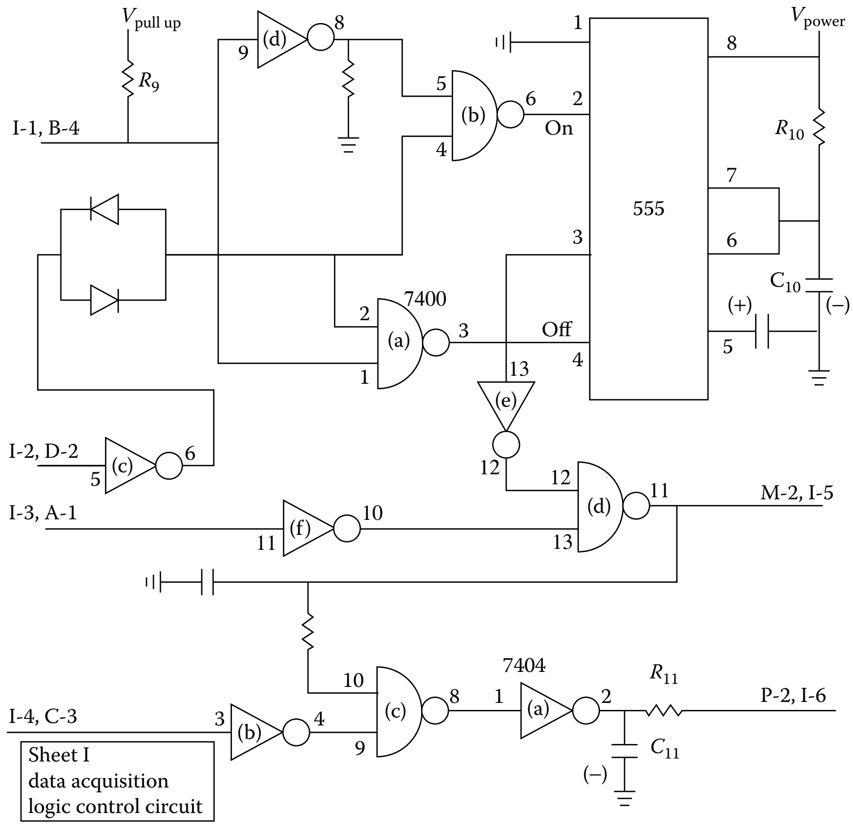


FIG. 5.10i
7400, 7404, and 555 interface for time synchronization for data acquisition.

For example, although the function diagram in Figure 5.10i does not violate any fan-in or fan-out device restrictions, voltage and current input and output characteristics are not provided, and the fan-in and fan-out requirements associated with the inclusion of the 555 into a TTL application are not discussed. There is no attempt to discuss the resistor and capacitor values needed to build the circuit for this application. Nor is there any focus beyond awareness of the specific diode signal conditioning requirements for this circuit.

IMPLEMENTATION OPTIONS

Although it is clear that integrated technology takes less space and is considerably faster than its original relay logic counterparts, there are several choices in implementing this relay replacement technology. The initial choice of solid-state logic elements over mechanical relay elements is a matter of economics and requirements. In general, solid-state elements are indicated when any of the following factors are important:

1. The switching speed requirements are high (above 100 per sec). At such speeds, the life of an ordinary relay is limited to a few hundred hours.
2. The system requirements are complex, involving a large number of elements for implementation.
3. The system receives or transmits signals from or to a digital computer.
4. Space is limited, such as in portable equipment.
5. Process downtime is costly and must be minimized.

If the amount of logic to be implemented is large and it involves a batch or continuous process application, the implementation selection is easy: The application should be installed on a programmable logic controller (see Section 5.4). However, for instrumentation applications or process applications similar to or smaller than the applications reviewed above, the use of dedicated solid-state circuits is a reasonable option.

Solid-State Logic Options

There are three popular methods of implementing dedicated solid-state logic control circuits. In the first option, the circuit is constructed on a plastic card, called a printed circuit board (PCB). The conducting connections shown between pins on the device packages are etched on the PCB, with the input and output leads brought to the edge of the board. The logic packages are attached to the appropriate places on the board, and the entire circuit board is plugged into a protective chassis. Circuit repair usually encompasses replacement of the entire board.

The second method of solid-state logic circuit implementation is the use of a plug-in module. This technique is popular because it closely mimics the construction philosophy in the relay logic circuits, which it is replacing. A rack is mounted near the application, and small, individual, often cubic solid-state packages are plugged into the rack. Wires are connected between cubes where necessary. This technique offers more versatility than the dedicated circuit board approach but also requires more space and is subject to more failures generated by human error during or after installation.

The third way to implement dedicated pure logic control involves using a single board controller card that is driven by programmable logic device (PLD). These cards allow the user to program various combinations of logic operations into the PLD. In operation, the inputs to this board generate output responses that are functions of the logic scenario currently resident in the PLD. These cards may or may not come with the appropriate industrial signal input/output interface. However, such interfaces are available and the entire setup, i.e., the PLD card and the interface cards, provides a compact stand-alone pure logic control.

Bibliography

Almaini, A. E. A., *Electronic Logic Systems,* Englewood Cliffs, NJ: Prentice Hall, 1986.

Amos, S. W., *Dictionary of Electronics,* 2nd edition, London: Butterworths, 1987.

Bishop, T., "The Logic of SSRs," *Instruments and Control Systems,* December 1977.

Carr, J. J., *Elements of Electronic Instrumentation and Measurement,* Englewood Cliffs, NJ: Prentice Hall, 1986.

Gilbert, R. A. and Llewellyn, J. A., *Principles and Applications of Digital Devices*, Research Triangle Park, NC: Instrument Society of America, 1982.

Patrick, D. R. and Fardo, S. W., *Electrical Motor Control Systems: Electronic and Digital Controls Fundamentals and Applications*, Goodheart-Willcox, 2000.

Paynter, R. T., *Introductory Electronic Devices and Circuits: Conventional Flow Version*, 6th edition, Englewood Cliffs, NJ: Prentice Hall, 2002.

Porat, D. L. and Arpad, B., *Introduction to Digital Techniques,* 2nd edition, New York: John Wiley & Sons, 1987.

Reilley, C. M., *Transistor Engineering and Introduction to Integrated Semiconductor Circuits,* New York: McGraw-Hill, 1961.

Roth, C. H., *Fundamentals of Logic Design,* Brooks Cole, 2003.

Smith, S. and Smith, K. C., *KC's Problems and Solutions for Microelectronic Circuits,* 4th edition, Oxford: Oxford University Press, 1998.

Strong, J. A., *Basic Digital Electronics,* London: Chapman and Hall, 1991.

Tocci, R.J. and Widmer, N. S., *Digital Systems: Principles and Applications*, 8th edition, Englewood Cliffs, NJ: Prentice Hall, 2000.

5.11 System Integration: Computers with PLCs

S. VITTURI (2005)

Data Transfer Speeds:	9.6 to 38.4 Kbps via serial point-to-point links 9.6 Kbps to 12 Mbps depending on which fieldbus is used 100 Mbps to 1 Gbps or more if Ethernet LANs are employed
Transmission Distance:	*Serial links* are used to transfer small amounts of data over short (<100 m) distances. *Ethernet*: The maximum coverable distance between two stations connected to a switch is 100 m, but the use of rooters and bridges can extend that limit to several thousand meters. *Fieldbuses* can transmit over several hundred meters.
Costs:	*Serial links*: Port available free of charge on PC. The serial interface board for the PLC ranges from \$200 to \$800. Fieldbuses require one interface board at the PC and another one at the PLC, both with the same protocol and each costing about \$500. *Ethernet* connections also require two interface boards, the widespread board at the PC is about \$20, while the PLC Ethernet board is about \$500.
Partial List of Interface Board Suppliers:	Omron (www.omron.com) Rockwell Automation (www.rockwell.com) Siemens (www.ad.siemens.de)
Serial Link Standards:	Electronic Industries Association (www.eia.org) American National Standard Institute (www.ansi.org)
LAN Standard:	Institute of Electrical and Electronic Engineers (www.ieee.org)
Fieldbus Standard:	International Electrotechnical Commission (www.iec.ch)
Fieldbus Organizations:	ControlNet (www.controlnet.org) DeviceNet (www.odva.org) Foundation Fieldbus (www.fieldbus.org) Profibus (www.profibus.com)

INTRODUCTION

The integration of DCS systems with various communication networks is discussed in Section 4.9, and the integration of DCS systems with programmable logic controllers (PLCs) or personal computers (PCs) is discussed in Section 4.10. In this section, the integration of PCs and PLCs is discussed, which can be achieved by the use of three types of connections: (1) serial links, (2) Ethernet connections, and (3) fieldbus-type communication networks. The transfer speeds and costs of each of these links been listed above.

The integration will only be successful if suitable communication protocols are adopted. When serial links are used, it is important to correctly match their electrical features; therefore, proprietary communication protocols have to be used. Conversely, communication networks and buses are equipped with their own protocols as defined in the international standards to which they belong.

The data transfer speed required by a PC–PLC connection is not only determined by the number of bits that have to be transferred through that link. The response time of the system is also influenced by factors such as the time required to acquire the data that is to be transmitted and the time required to execute the program in either the PC or the PLC.

The requirement for connecting computers with programmable logic controllers is becoming more and more frequent. In the past such connections could only be made with mainframes, workstations, and personal computers. Currently, due to both the improvement in PC performance and because of their increased use in industrial environments, the majority of PCs and PLCs can be connected. Therefore, in this section, the only computers that will be discussed will be PCs.

A PC can be connected to a PLC for two different purposes: programming and monitoring/supervision.

After a short description of the connection that serves programming, the rest of this section will be devoted to the discussion of the monitoring/supervision connection.

PCs FOR PROGRAMMING THE PLC

In this case, the PC is used to both implement and test the program of the PLC. Each PLC supplier provides software packages to implement and test their programs. In practice, these packages assist the programmers to completely develop their control systems. Thus, in general the software packages should include:

- Compiler/interpreter programs suited for the PLC programming languages
- Tools to transfer program blocks to/from the PLC
- Tools to force or observe the status of the internal variables of the PLC
- Tools to implement the cross-checking of variables throughout the PLC program

PCs FOR MONITORING AND SUPERVISING THE PLC

When the PC is connected to the PLC for purposes of monitoring and supervision, the PLC is usually designed to control the industrial plant or machine while communicating with a PC, which is an integral part of the plantwide automation system. Such a connection usually allows the PLC to handle the operation of process, while the PC is there as an overseer.

Usually, the control tasks (whose execution may be time critical) are completely implemented by the PLC, while the monitoring and supervisory tasks are resident in the PC. In this way there is a clear separation between the tasks, and the PC–PLC connection can be interrupted without any harmful consequences on the plant or machine that is being controlled.

In order to implement the monitoring and supervisory functions, the PC has to be equipped with a suitable software package. However, while the PLC suppliers provided the software for supporting the programming connection, they do not supply this software package. This software can be obtained by the user, either from the customization of a general-purpose Supervisory Control and Data Acquisition (SCADA) system or developed ad hoc for the application.

SCADA System

In a SCADA system, a group of tasks are running on the PC, each of them performing specific functions. The database stores the tag numbers of devices and loops, which are exchanged for the various tasks. Because many of these tags are mapped onto the corresponding PLC variables by one or more of the tasks, the handling of the plants or machines may be implemented by means of read/write actions on the relevant tags.

Figure 5.11a illustrates how the SCADA system notifies a human operator through a PC of an alarm condition that was detected by the PLC. The alarm condition is detected by a sensor, which is located in the field. The field device sends the alarm to the PLC through its input boards. This initiates a suitable alarm handling procedure in the PLC, which results in the transmission of an alarm event, via the PC–PLC connection, to the PC.

Such an event is handled by a PLC task (action) of the SCADA and results in associating the event with the tag number of the alarm, which resides in the SCADA database. Then, the alarm handling task of the SCADA notifies the human operator of the existence of the alarm condition by

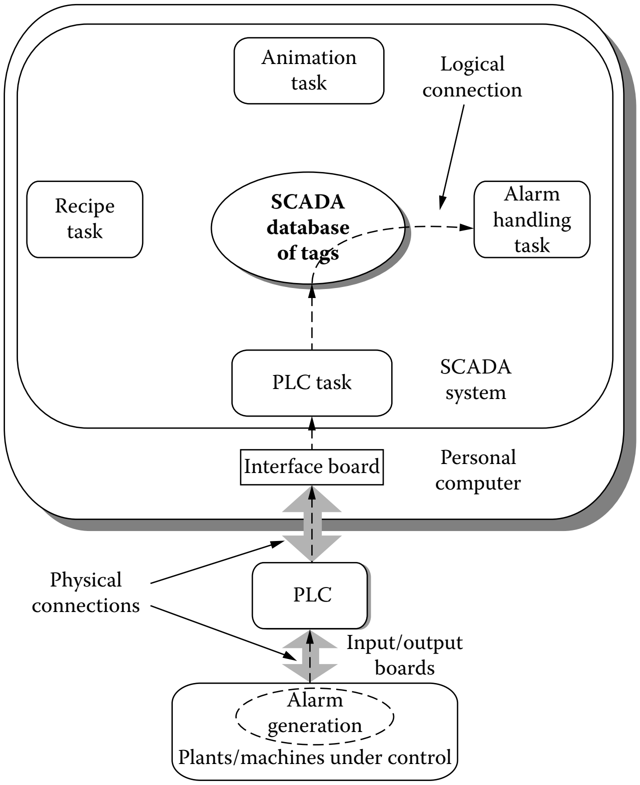


FIG. 5.11a
Example of transmitting an alarm event by SCADA.

sending a suitable text message to the screen of the PC. If necessary, an audible signal can also be initiated.

A software package that is developed to serve only the monitoring and supervision of a PLC is usually simpler than a complete SCADA system and can be implemented using one of the high-level programming languages.

Handling of Tasks

No matter what software implementation is used, the ability to handle some actions (tasks) is indispensable for the monitoring and supervision by the PC. In particular, it is necessary to support the direct data exchange between the PC and the PLC. In Figure 5.11a this data exchange operation is identified as "PLC task."

"Animation task" in Figure 5.11a refers to the capability of performing graphics and dynamic tasks. These provide the human operator with mimic diagrams, which can dynamically represent the status of the plant. The operator can also issue commands from these mimic diagrams to change set points or to make other changes in the control system of the process.

Other tasks that are also available relate to trend recording of process variables, process data logging, and recipe handling.

IMPLEMENTATION

The physical connection between the PC and the PLC can be implemented in one of two ways:

1. Via a serial point-to-point link
2. Using a communication network

Serial Link

Point-to-point serial link connections have been often used in the past, because of their simplicity and low cost. From a hardware point of view, such a connection can be made by using two interface boards and one connecting cable. In this configuration, because the PC is typically provided with at least one serial interface board, it is only necessary to equip the PLC with a corresponding board.

The serial point-to-point links are normally based on the RS-232[1] and RS-422[2] standards. With reference to the International Standards Organization (ISO) model for Open System Interconnection (OSI),[3] these standards are concerned only with the features of the physical layers used by the links; that is, the levels of electrical signals, their timing, and so on.

Consequently, in order to transfer process data over the connection, it is necessary to develop a suitable protocol, which has to be implemented on both the PC and the PLC sides. Since these protocols are often designed by the users, they tend to strictly reflect the particular application and are not flexible or general in their nature. As a consequence, these protocols are completely proprietary and are not recognized by any standard.

The transmission speeds available with such connections are relatively low; typical values are 9.6 Kbps, 19.2 Kbps, and 38.4 Kbps. For these reasons such links are only used to transfer small amounts of data, because otherwise the overall transmission system would become very slow.

Finally, the distances covered with serial connections are limited to some tens of meters.

Communication Networks

The use of communication networks to connect "intelligent" devices is dramatically increasing in the processing industries. The communication networks used can be grouped into two categories:

1. Local area networks (LANs)
2. Fieldbuses

Considering the PC–PLC connection, both of these solutions require the use of interface boards in order to implement the particular protocol used by the networks.

LANs and the Ethernet Connection LANs have been standardized by the IEEE 802 committee, and the results have been subsequently recognized by the International Standards Organization in its 802 family of standards.

The IEEE 802 committee was concerned with only the first two layers of the OSI reference model, namely the physical layer and the data link layer. As shown in Figure 5.11b, the data link layer has been subdivided into two sublayers: the medium access control (MAC) and the logical link control (LLC) sublayers. The LLC provides a common interface

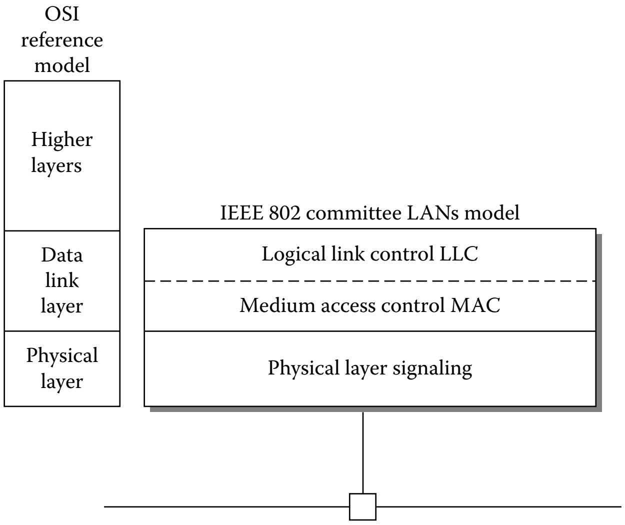


FIG. 5.11b
The model of the local area network (LAN) as specified by IEEE 802 committee.

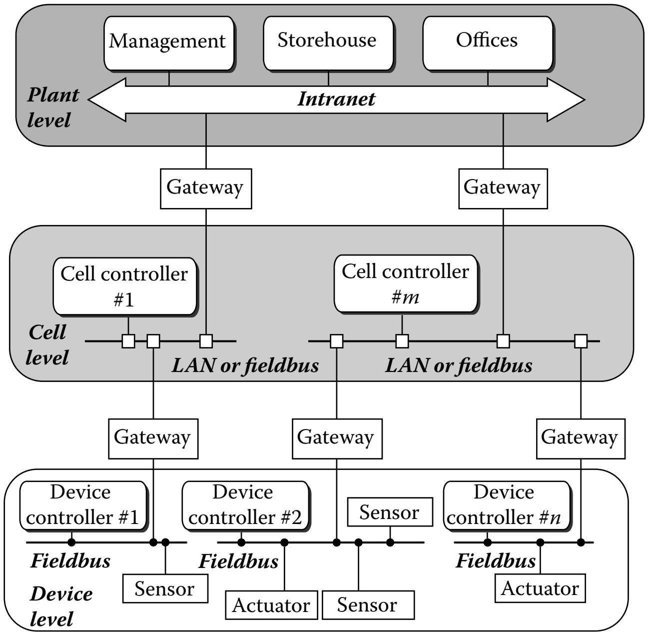


FIG. 5.11c

The three levels of plantwide communication networks used in process control and automation of complete plants.

toward the higher layers, while different MAC layers are used by the different types of networks.

Although the IEEE committee offers several choices, almost all of the LAN systems are based on the MAC specified by the IEEE 802.3 subcommittee. This MAC is universally known as Ethernet,[4] which is also used for the PC–PLC connections. Ethernet interface boards for both PCs and PLCs are available off the shelf and are widely used.

Fieldbuses Fieldbuses are networks that are designed to operate at the device and cell levels of plantwide automation systems.[5] As it is illustrated in Figure 5.11c, fieldbuses may implement control loops by providing the connection between a controller and its sensor (transmitter) and its final control element (control valve or actuator) at the “device” (the lowest) level of the plantwide automation systems. At the cell or intermediate level, fieldbuses provide the connection between intelligent devices.

In contrast with LANs, where Ethernet is the most widely used PC–PLC connection, among the various fieldbuses, there is no leading product. This is probably due to the slowness of the process of international standardization. Therefore, today there are a number of proprietary fieldbus products on the market, including DeviceNet, ControlNet, Fieldbus Foundation, Interbus, Profibus and WorldFIP. All of these products are now included in the various European and international standards: EN50170,[6] EN50254,[7] EN50325,[8] and IEC61158.[9]

Because each fieldbus product requires its specific interface board, and because all the interface boards for all these systems are not always available, the implementation of PC–PLC connections is easier using LANs than with fieldbuses. Nevertheless, if the same fieldbus is used throughout the plant, it ensures the compatibility of the connections at the device level of factory automation systems. For this reason, and because of their good performance, fieldbuses are often used.

Transmission Speeds and Distances Communication networks have several advantages over serial links. In particular, the communication networks are faster in terms of their transmission speeds and can transmit over longer distances. Moreover, they can provide high-layer protocols, which can make the data transfer sessions more efficient.

The typical Ethernet transmission speed is 100 Mbps, and the maximum distance between a station and the switch is 100 m; however, many available network components allow for the implementation of very complex configurations, extending the coverable distance between PC and PLC to several thousands of meters. Moreover, Ethernet components working at 1 Gbps are already available, and the migration to 10 Gbps is expected very soon.

The fieldbus’s transmission speed depends on the product used. For example, Profibus may work at 12 Mbps, Interbus at 2 Mbps, ControlNet at 5 Mbps, and DeviceNet at 500 Mbps. Although the above speeds are lower than that of Ethernet, it has to be considered that fieldbus protocols are in general more efficient, so that the performance of fieldbus-based PC–PLC connections can be equally satisfactory.

The transmission distance of a fieldbus is a function of its supplier; however, almost all of the available standards are suitable for distances of several hundreds of meters.

Communication networks make possible the use of standard high-layer protocols, which were explicitly designed for industrial applications. These permit a fast and efficient data transfer. For a long time, at the device and cell levels, the manufacturing message specification (MMS)[10] has been the sole application layer protocol, defined by an international standard. MMS is based on the definition of “communication objects” representing the unique data that may be exchanged among stations of the network.

The most important communication objects are variables, domains, and program invocations. Variables are used to carry statuses and commands of the plant and typically can be single values, arrays, or structures.

Domains are memory areas that may be used to transfer considerable amounts of data.

Program invocations are executable programs constituted by groups of domains that can be remotely invoked.

Interface Choices If the PC–PLC connection is implemented via a LAN, the MMS protocol (or an equivalent application layer protocol) has to be installed on both of the interface boards used. This is usually provided by means of software

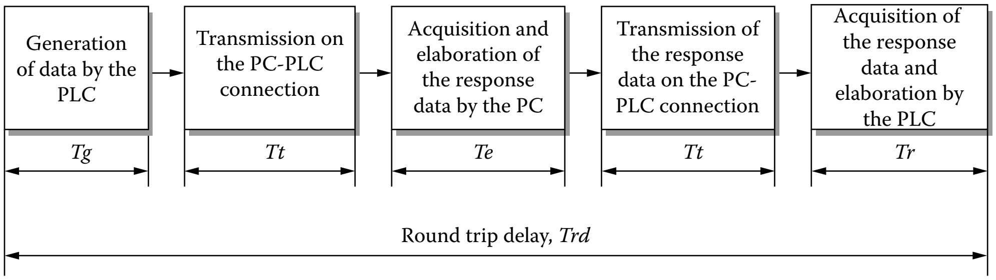


FIG. 5.11d
The steps and their associated time requirements, which add up to the total round trip time delay of a PC–PLC connection.

libraries, such as the product described in Reference 11, which is the MMS library for a specific Ethernet interface board.

The situation is slightly different for PC–PLC connections based on fieldbuses. In this case, the application layer protocol it is already included in the native fieldbus protocol tack. For example, the first version of Profibus had an application layer protocol named fieldbus message specification (FMS),[12] which is completely based on MMS. The same choice was adopted for the application layer of another important fieldbus, World-FIP,[13] included in the EN50170 standard.

Another technique that may be used for PC–PLC connections is the OLE for Process Control (OPC) interface, which is based on Microsoft's COM/DCOM™ technology. OPC[14] is not concerned with the communication protocols used, because it specifies an independent and high-level common interface between the applications located on the connected systems. The adoption of OPC makes use of client/server relationships. It introduces a unique way to exchange data between applications that are distributed on different systems, because it allows the client applications to access the data stored in the servers in a consistent manner.

It is also possible to use the type of protocols that are used for the Internet to connect PCs and PLCs with communication networks. The industrial use of such protocols is increasing. In particular, the file transfer protocol (FTP)[15] is well suited for PC–PLC connections, because it can exchange large amounts of structured data, which might be related to remote configuration procedures, trend analysis, and so on.

In addition, this type of configuration provides a possible link to the Internet and therefore the possibility for remote access to the connection via the Internet, because both the PC and the PLC are using suitable protocols.

PERFORMANCE

In installations where all control is executed in the PLC, and the PC serves only monitoring and supervisory purposes, the transmission time for passing the data over the PLC–PC connection is usually not critical. However, considering that the plant status monitoring information on the PC should also be up to date, it is still important to keep the "round trip" time delay of the connection as short as possible.

The steps and the associated time components of the total round trip time delay of a PC–PLC connection are shown in Figure 5.11d. Such communication can take place either through a serial link or over a communication network. In Figure 5.11d it has been assumed that the session was initiated by the PLC, but the same results would be obtained if the initiator of the transmission was the PC.

Generation Time (*Tg*)

As can be seen in Figure 5.11d, the generation time of the data is identified by the letter Tg and can be determined by the use of Equation 5.11(1):

$$Tg = Ta + \alpha \cdot Tc \qquad \textbf{5.11(1)}$$

where

Ta is the time required to acquire the transmitted signal from the field

Tc is the PLC's cycle time,

$\alpha \in [0,1]$ represents the percentage of the cycle time that has to elapse before the data is in fact acquired by the PLC

The PLC reads the input data at the beginning of each program cycle. The term $\alpha \cdot Tc$ in Equation 5.11(1) takes into account the fact that the signal from the field instrument can be received while the PLC is executing some instruction of its program. Thus, if the signal from the field is received immediately after that signal has been read by the PLC, it will be necessary to wait for one whole program cycle, hence $\alpha \cong 1$. On the other hand, if the generation occurs immediately before the signal is read, $\alpha \cong 0$.

Transmission Time (*Tt*)

Tt is the time necessary to transmit the data over the communication link between the PLC and the PC. If N is the number of bits to be transmitted and B is the bit rate of the link, then Tt is

$$Tt = \frac{N}{B} \qquad \mathbf{5.11(2)}$$

It is important to note that N has to include not only the data bits, but also the bits used by the transmission protocol, such as, for example, the bits used for error control and addressing.

Te, *Tr*, and *Trd*

The elaboration time Te cannot be expressed by an equation, because it depends on many factors. It comprises a first delay component, which is caused by the transmission of the data inside the PC from the network interface board to the task that will perform the elaboration. Next is the time it takes for the PC operating system to schedule the execution of the task, which depends on all the activity and occurrences in all the other tasks. Finally, the data are transferred to the network interface board for transmission over the connecting link.

The return transmission time Tr (similarly to Tg) is determined by the arrival time of the data at the PLC in relation to the program cycle. Thus, Tr can be expressed as

$$Tr = \beta \cdot Tc \qquad \mathbf{5.11(3)}$$

where Tc is the cycle time and $\beta \in [0,1]$.

In conclusion, the round trip delay (Trd), which is the time interval that has to elapse from the generation of the data by the PLC and the arrival of the corresponding response from the PC, may range between two values, which can be calculated by Equations 5.11(4) and 5.11(5):

$$Trd^{\mathrm{MIN}} = Ta + 2Tt + Te \qquad \mathbf{5.11(4)}$$

and

$$Trd^{\mathrm{MAX}} = Ta + 2Tc + 2Tt + Te \qquad \mathbf{5.11(5)}$$

Equation 5.11(4) assumes that $\alpha = \beta = 0$ while Equation 5.11(5) assumes that $\alpha = \beta = 1$.

CONCLUSIONS

The integration of DCS systems with various communication networks was discussed in Section 4.9, and the integration of DCS systems with PLCs or PCs was discussed in Section 4.10. In this section, the integration of PCs and PLCs was discussed.

"System integration" is a nebulous term. It can mean just about anything the integrator desires. It is up to the system owner to define the level of integration expected and to justify the cost of integration. The current trend is away from simple point to point connections and toward Ethernet and fieldbus connections using either standards-based open protocols, or industry-accepted proprietary protocols. The high speeds associated with these connections support not only massive information transfer, but also, in many cases, shared control.

References

1. Engineering Publications Office, Electronic Industries Association, "Interface between Data Terminal Equipment and Data Circuit-Terminating Equipment Employing Serial Binary Data Interchange," ANSI/TIA/EIA-232-F-1997, Washington, D.C., 1997.
2. Engineering Publications Office, Electronic Industries Association, "Electrical Characteristics of Balanced Voltage Digital Interface Circuits," ANSI/TIA/EIA-422-B-94, R2000, Washington, D.C., 2000.
3. International Standards Organization, "Open Systems Interconnection – Basic Reference Model," ISO/IEC 7498, Parts 1 to 4.
4. IEEE 802-3: 2000 standard, Part 3: "Carrier Sense Multiple Access with Collision Detection (CSMA/CD) and Physical Layer Specifications," Institute of Electrical and Electronics Engineers, 2000.
5. Vitturi, S., "Overall Fieldbus Trends," in *Instrument Engineers' Handbook: Process Software and Digital Networks,* pp. 495–504, Boca Raton, FL: CRC Press-ISA, July 2002.
6. European Committee for Electrotechnical Standardization, "General-Purpose Field Communication System, " Volumes 1 to 5, EN50170, July 2000.
7. European Committee for Electrotechnical Standardization, "High Efficiency Communication Subsystem for Small Data Packages," Volumes 1 to 3, EN50254, 1998.
8. European Committee for Electrotechnical Standardization, "Industrial Communication Subsystem Based on ISO 11898 (CAN) for Controller–Device Interface," Parts 1 to 3, EN50325, 2000.
9. International Electrotechnical Commission, "Digital Data Communications for Measurement and Control–Fieldbus for Use in Industrial Control Systems," Parts 1 to 6, IEC 61158 standard, 2000.
10. International Standards Organization, "Manufacturing Message Specification Parts 1 and 2," ISO IS 9506, 1990.
11. Siemens AG, "Simatic Net IE CP1613 Communication Processor with Technological Functions Protocol," Automation and Drives Division, Technical Support, 1998.
12. European Committee for Electrotechnical Standardization, "General-Purpose Field Communication System, Volume 2, Profibus," EN50170/2, 1996.
13. European Committee for Electrotechnical Standardization, "General-Purpose Field Communication System, Volume 1, WorldFIP," EN50170/3, 1996.
14. OPC Foundation, "OPC Common 1.10 Specification," 2002. Available online at www.opcfoundation.org.
15. Postel, J. and Reynolds, J., "RFC 959, File Transfer Protocol," October 1985.

Bibliography

Bray, O. H., *CIM—The Data Management Strategy,* Bedford, MA: Digital Press, 1990.

Hagar, M. et al., *MMS Beyond MAP—An Integration Protocol for Factory Networks,* Ann Arbor, MI: National Center for Manufacturing Sciences, 1990.

Hughes, T. A., *Programmable Controllers,* Research Triangle Park, NC: Instrument Society of America, 1989.

International Standards Organization Standard 7498, www.iso.org/iso/en, 1984.

Jones, V. C., *MAP/TOP Networking,* New York: McGraw-Hill, 1988.

Martin, J., *Local Area Networks—Architectures and Implementations,* Englewood Cliffs, NJ: Prentice-Hall, 1989.

Miklovic, D. T., *Real-Time Control Networks for Batch and Process Industries,* Research Triangle Park, NC: Instrument Society of America, 1992.

Postel, J. and Reynolds, J., "RFC 959, File Transfer Protocol," October 1985.

Savage, C. M., *Fifth Generation Management,* Bedford, MA: Digital Press, 1990.

Siemens AG, "Simatic Net IE CP1613 Communication Processor with Technological Functions Protocol," Automation and Drives Division, Technical Support, 1998.

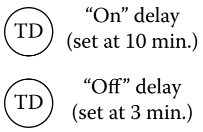
Williams, T. J., *A Reference Model for Computer Integrated Manufacturing,* Research Triangle Park, NC: Instrument Society of America, 1989.

Vitturi, S., "Overall Fieldbus Trends," in *Instrument Engineers' Handbook: Process Software and Digital Networks*, pp. 495–504, Boca Raton, FL: CRC Press-ISA, July 2002.

Zuboff, S., *In the Age of the Smart Machine,* Philadelphia: Basic Books, 1988.

5.12 Time Delay Relays

J. FRYDMAN (1985) **B. G. LIPTÁK** (1995) **W. P. DURDEN** (2005)



Flow sheet symbol

Type of Design:	A. Thermal devices, bimetallic strip springs B. Dashpots C. Pneumatic delays D. Inertial devices, weights, flywheels E. Solid state F. Motor-driven delays G. Software delay
Type of Delay:	On-delay, off-delay, interval, single-shot, pulse detection, and repeat cycle
Delay Settings:	A. 0.2 to 3 sec B. 1 to 120 sec C. 0.1 to 300 sec D. 0.05 to 0.2 sec F. Can be many hours
Timing Inaccuracy:	Generally 5 to10%. Type A is 5 to20%; for better accuracy use interval timers instead of time delays. Crystal-controlled oscillators can provide 0.1%
Reset Time:	About 100 ms
Life Expectancy:	Electrical types at full load usually operate a minimum of 100,000 times; mechanical types are good for 20 million operations or more
Load Switching:	Relay contacts, snap-action switches, SCRs, transistors, triacs
Environmental Limits:	14 to 140°F (–10 to 60°C); 5 to 95% relative humidity
Mounting Configurations:	Plug-in, surface-mount, panel-mount, in-line module
Costs:	From \$20 to \$200
Partial List of Suppliers:	ABB SSAC Inc. (www.ssac.com) Agastat (www.agastat.com) Allen-Bradley (www.ab.com) Amperite Co. (www.amperite.com) Artisan Controls Co. (www.artisancontrols.com) Automatic Timing & Control (ATC) (www.flw.com/atc/atc.htm) Bircher (www.bircher.com) Bright Toward Industrial Ltd. (www.relays.com.tw) Castel (www.castel.com) Comadan Producktion a/s (www.comadan.com) Crouzet (www.crouzet-usa.com) Dold (www.dold.com) Eaton-Durant Products (www.durant.com) Emerson Climate Technologies (www.emerersonclimate.com) Fiber (www.fiber.it) Fuji (www.automationdirect.com)

General Electric (www.geindustrial.com)
Hiquel (www.hiquel.com)
Idec (www.idec.com)
Instrumentation & Control Systems (www.ics-timers.com)
Kessler-Ellis Products (www.kep.com)
KOYO (www.automationdirect.com)
Kübler (www.kubler.com)
Kuhnke (www.kuhnke.de)
Lovatao Electric (www.lovatoelectric.com)
Macromatic Controls (www.marcromatic.com)
Omron (www.oeiweb.omron.com)
Pilz (www.pilz.com)
Phoenix Contact (www.phoenixcontact.com)
Schleicher (www.schleicher-de.com)
Siemens Energy & Automation (www.siemens.com)
Square D Co. (www.squared.com)
Tele Hause Steuergeräte (www.tele-power-net.com)
Telemecanique (www.telemecanique.com)
Texas Instruments (www.ti.com)

Time delay devices are used for delaying the start-up, shutdown, recycling, or continuation of processing operations until the desirable requirements have been satisfied or the required conditions have been obtained.

GENERAL CHARACTERISTICS

Time delay relays are special-purpose relays or logic components that have some characteristics of both relays and timers. Time is one of the variables of any process, and it often need to be monitored or controlled. Some operating steps in a process might need to coincide or may need to be separated from some other steps by a specific interval of time. Time delay relays serve to satisfy such timing requirements.

The two main components of time delay relays are the timing circuit mechanism that produces the required time interval and the load switching contacts that are actuated at the end of that time interval. For a time interval to be determined, the following prerequisites must be present:

1. A power source (if needed in addition to signal power)
2. Signal power
3. A change of state of the time-determining device
4. An indication that the time-determining device has changed to the desired state

The power or signal power can be thermal, pneumatic, AC, or DC electric current. The change in the state of the timer can be mechanical (e.g., caused by the rotation of a motor, the motion of a plunger in a restricting fluid, or the bending of a bimetal strip caused by a change in temperature), electrical (e.g., caused by an accumulation of charge of a capacitor or a count of oscillations or pulses), or software-based (e.g., based on the number of program scans or interrupt-driven).

The load switching can occur via snap-action switches, snap-action valves, relay contacts (e.g., SPDT, DPDT, reed [hermetically sealed], or mercury [hermetically sealed]), or solid-state devices (e.g., transistors, SCRs, or triacs).

Timer Modes and Characteristics

The mode of the time delay determines the relationship between the time when the signal power is applied to the time of load switching. This relationship may also depend on the continuous presence of the power source. The four most prevalent modes are on-delay, off-delay, interval, and single-shot. A special combination of the on-delay and the off-delay is called a repeat cycle timer.

In addition to the aforementioned combinations, one must consider the following characteristics when choosing time delay relays from among the vast number available:

1. Timing range
2. Fixed versus adjustable timing
3. Accuracy
   a. Dial setting accuracy
   b. Tolerance
   c. Repeat accuracy
   d. Time between operations
4. Environmental factors
   a. Temperature range
   b. Vibration, shock
5. Loading switching
   a. Duty cycle
   b. Type of load: resistive, inductive, lamp
   c. Life expectancy with load
6. Mounting considerations
   a. Size limitations
   b. Mounting style: surface-mount, panel-mount, plug-in, in-line module
   c. Terminals: plug-in, quick connect, solder lug, screw terminal

7. Housing
   a. Open chassis
   b. Dustproof
   c. Weatherproof
   d. Explosionproof
   e. Totally encapsulated

Types of Time Delays

Recovery Time *Recovery time* is the minimum amount of time between removal of the signal power and its reapplication, which is necessary so that the subsequent operation will have the desired repeatability. The analog or digital solid-state timing circuits have the shortest recovery time.

On-Delay An *on-delay* is alternatively called delay on make, delay on operate, delay on pull-in, delay on energize, slow-acting, or slow-operating. The load switching occurs a certain time after the application of the signal power (Figure 5.12a). If a power source is normally required and the power is interrupted before load switching takes place, the timing cycle has to be repeated from 0 to effect a delayed load switching (except in certain synchronous motor-driven delay relays).

Off-Delay An *off-delay* is alternatively called delay on break, delay on de-energize, delay on drop-out, delay on release, slow release, drop-out delay, or delayed drop-out. The load switching occurs a certain time after the removal of the signal power (Figure 5.12a). If the power source is normally required and the power is interrupted before the minimum amount of time or after removal of the signal power prior to load switching, the amount of time delay may be inaccurate.

Interval An *interval* timer is alternatively called interval on, on interval, pulse shaping, bypass timing, interval delay, and delay on energization with instantaneous transfer. After application of the signal power and while the signal power is maintained, the load switching occurs for a certain time only, and then the load switching is de-energized (Figure 5.12a). If the signal power is interrupted or the power source (if normally required) is removed before the completion of the time delay, the load switching is de-energized instantaneously.

Single-Shot A *single-shot* is alternatively called latched interval, latching off delay and latching delay on de-energization, momentary actuation, or one-shot. Load switching occurs for a certain time only, and then the load switching de-energizes after a momentary application of the signal power (Figure 5.12a). The signal power may be applied longer without altering the load switching interval. If the power source is normally required and it is interrupted while the load switching is energized, it de-energizes instantaneously.

Pulse Detection A *pulse detection* relay monitors a contact or control input to activate an interval time period. The output contact may already be activated or on with the application of power or it may wait until the control input is active. Once the control input is on the control input must cycle within the interval time period to either maintain the output or to drop the output as selected.

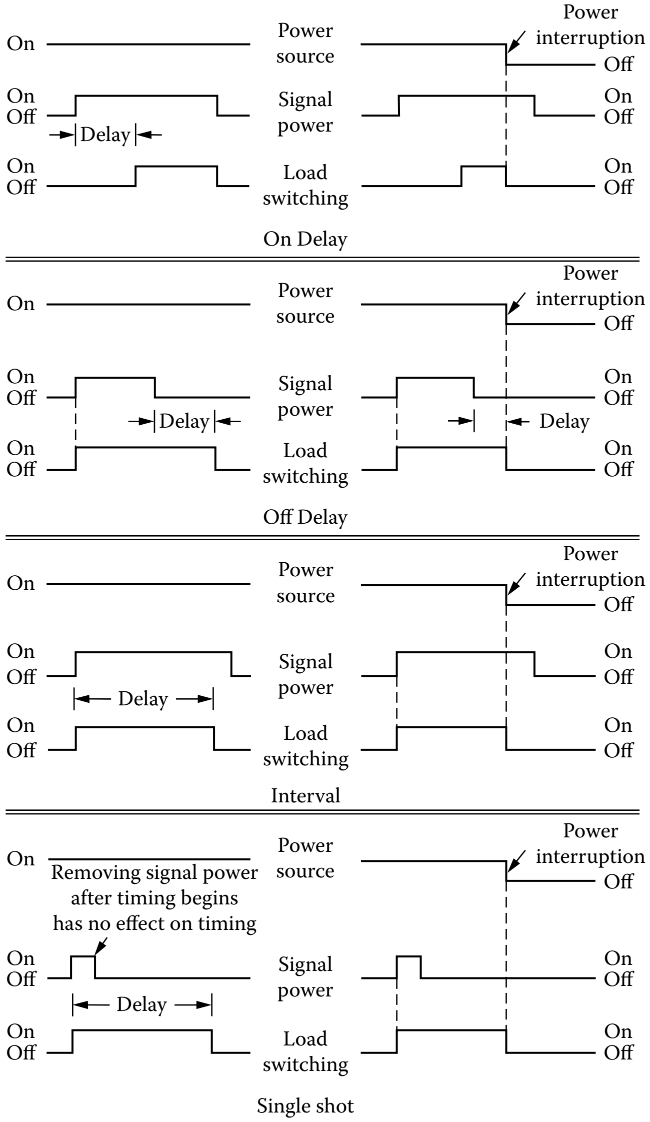


FIG. 5.12a
Timing charts of the various types of time delays.

Repeat Cycle A *repeat cycle* relay is alternatively called dual delay; combination delays; on-delay, off-delay; or slow acting-slow releases; or asymmetric pulser. The load switching occurs a certain time delay after the application of the signal power, and remains energized for another time delay after the removal of the signal power. If the power source is normally required and it is interrupted while the load switching is energized, it may be de-energized prematurely. The asymmetric pulser has the added feature of two time periods, one for the on period and one for the off period.

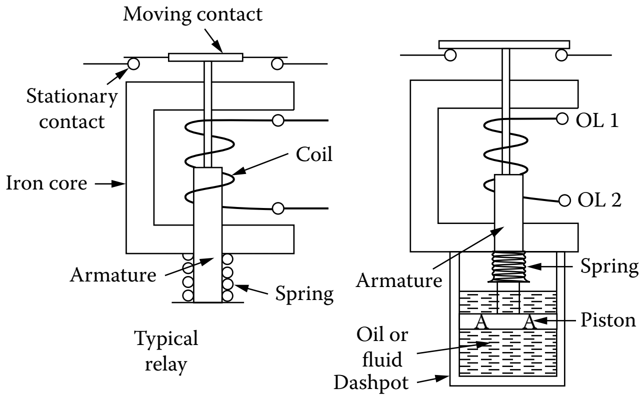


FIG. 5.12b
Dashpot time delay.

Types of Designs

Dashpot or Pneumatic In a *dashpot*-type time delay the armature of a relay is augmented with a piston that travels in an oil or fluid. The piston contains several holes through which the fluid passes. The current of the coil must be applied continuously to exert a pull on the armature. The armature travels slowly because of the resistance of the fluid, if it has the required viscosity to produce the necessary delay. The size of the aperture holes may be modified to vary the time delay. Once the armature travels the required distance, the moving contacts engage the stationary contacts. A dashpot time delay is shown in Figure 5.12b.

In a *pneumatic* time delay, the time it takes for a certain volume of air to pass through a preset size of opening is used for timing. This design is made dustproof by a diaphragm and a cap, which encase the head in which the air is recirculated.

Thermostatic In this version of a time delay design, one of the contacts is a bimetallic element that is wrapped with an insulated heating coil. The signal power is applied to the heating coil. The delay time is the time required to raise the temperature of the contact high enough to produce warping, which will finally cause the contact to transfer.

The timing tolerance (accuracy) of standard thermostatic time delay relays is typically only ±20% but special designs are available with better tolerances. These units are also only repeatable to approximately ±5%, but on the other hand, they are of low cost and inherently resistant to transients. A thermostatic time delay along with two forms of packaging is depicted in Figure 5.12c.

Motor-Driven This type of relay is powered by a synchronous motor. A tripping arm is driven through a simple train of machine-cut spur gears. The arm actuates a snap-action switch, and the motor speed and the gear reduction determine the time delay. Resetting is instantaneous by means of an electromagnetic coil that clutches and declutches the trip locking gear as required. The clutch coil is often equipped with a set of auxiliary contacts.

The power source is normally applied to the motor circuit, while the signal power is normally applied to the clutch coil. In this manner, there are instantaneous contacts actuated by the clutch coil and time-delayed contacts actuated by the tripping arm. When the power to the clutch is maintained, the time delay is increased by the delay in power interruption. Viewed otherwise, the relay accumulates all of the time that the power source has applied to the motor until it equals the time that is preset on the time delay relay.

The use of a synchronous motor ensures great accuracy and, with appropriate gearing, allows for very long delays, up to and including hundreds of hours. Sometimes these relays are provided with a progress pointer. A motor-driven delay is illustrated in Figure 5.12d.

Analog Solid State Analog time delays are based on the RC circuit, in which a capacitor is charged via a resistor until it reaches a certain voltage. When the predetermined voltage is reached, the load switching electromechanical relay or a solid-state switching device turns on or off, depending on the function.

This type of timing is much simpler, more reliable, and less expensive than the mechanically complex synchronous motor-driven delay relay. However, there are some shortcomings, which include the fact that the time-constant curve for the capacitor charging is exponential, making it difficult to set a potentiometer (the variable R) accurately.

Another limitation is that the RC circuit is sensitive to variations in supply voltage. Also, the setability and repeat accuracy of this design are affected by variations in the resistance because of variations in temperature, and in addition the

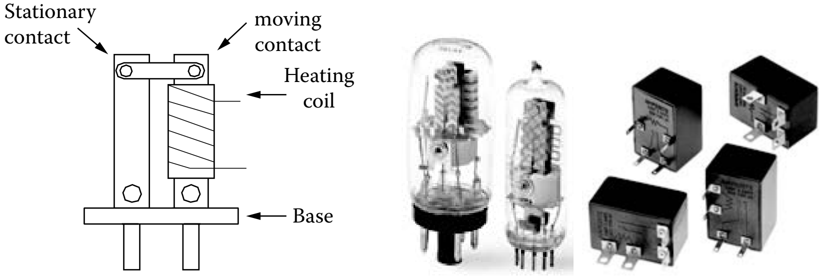


FIG. 5.12c
Components of thermostatic time delay relay in hermetically sealed, glass tube and plastic case. (Courtesy of Amperite Co.)

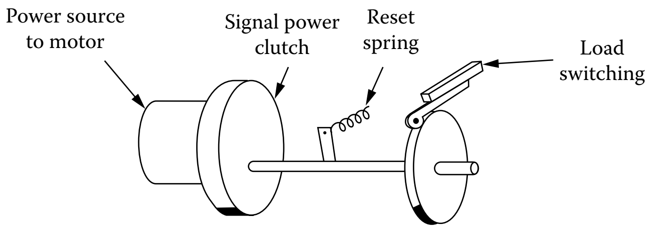


FIG. 5.12d
Motor-driven delay.

capacitor shifts its value as it ages. Steps can be taken to offset or lessen these adverse effects of voltage or temperature variations, but these steps add to the cost of the analog time delay relay. An analog time delay is shown in Figure 5.12e.

Digital Solid State The digital method for creating a time delay uses a frequency counting or dividing circuit. This approach, although more expensive than the analog RC circuit, provides better accuracy and repeatability and lends itself to medium- or large-scale (chip) integration. In this type of circuit, pulses from an oscillator or from the frequency of the power line are applied to a counter or a divider chain, which is preset to output a pulse at a specific count. This output pulse turns on or off the load switching electromechanical relay or solid-state switching device.

The power line frequency, 50 or 60 Hz, is used because of its stability. Using a higher frequency free-running oscillator provides finer time increments. For ultimate accuracy, a crystal-controlled oscillator is used, providing absolute accuracy and repeatability, setable to the fourth significant figure. This type of delay relay would most often be supplied with a thumbwheel switch or dip switches to program the dividing counter. A digital time delay is shown in Figure 5.12f.

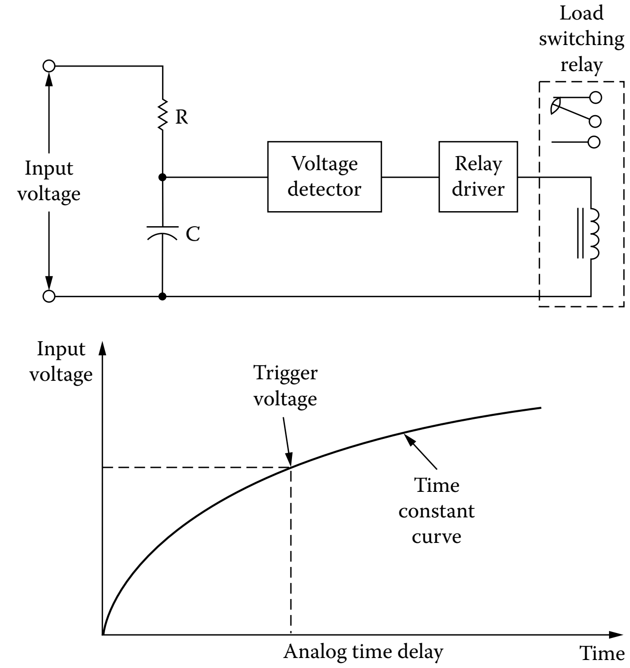


FIG. 5.12e
Analog time delay.

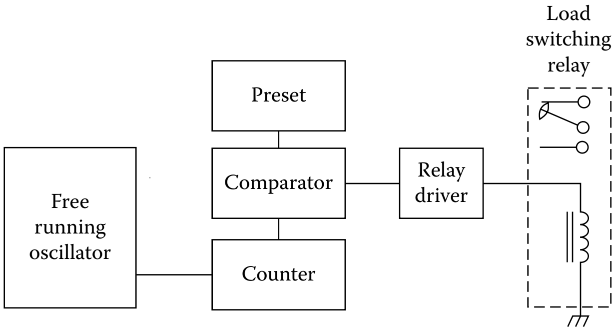


FIG. 5.12f
Digital time delay.

Software-Based Time Delays Modern control systems make increasing use of programmable controllers or microprocessors for all the logic, sequencing, and timing functions. All these functions are emulated in software by a program stored in memory. The program is merely a series of binary instructions written by the applications designer, upon which the central processing unit interprets and acts.

The software-based delays use the digital method of generating delays. Internal clock pulses are accumulated in a memory register or in a counter, which subsequently transfers its count to a memory register. The program regularly compares the accumulated count with a preset value that is stored in another memory register. When the accumulated count equals or exceeds the preset value, the program branches to another set of instructions that require the delay for further logic sequence.

An error may be present in the aforementioned method, affecting accuracy and repeatability. Because the software program takes varying amounts of time for acting on a set of instructions, successive scans through the whole program involve differing amounts of time. Therefore, when the software program returns each scan to compare the actual count with the preset count, the actual may exceed the preset by an unacceptable amount for the accuracy desired.

Some microprocessor-based devices have a modified digital method that corrects this problem. The actual count takes

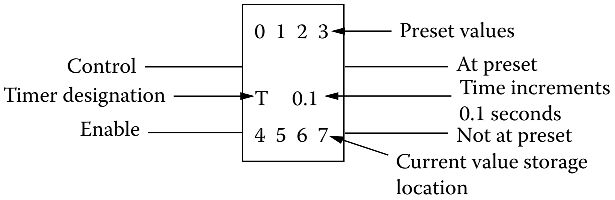


FIG. 5.12g
Software-based time delay. Timer block general form as viewed on a CRT of a typical programming panel of a PLC.

place in an integrated circuit, which creates an interrupt when it reaches the preset count. This interrupts the main software program immediately, and load switching can be executed without incurring additional time delay. A software-based delay is depicted in Figure 5.12g.

Bibliography

"All About Circuits," http://www.allaboutcircuits.com/vol_4/chpt_5/3 .html, December 2003.

Asfahl, C. R., *Robots and Manufacturing Automation,* New York: Wiley, 1992.

Auger, R. W., *The Relay Guide,* New York: Reinhold Publishing Co.

Benoit, C. N., "Specifying Solid-State Timing Devices," *Automation,* November 1972.

Collin, M., "TDRs Are Alive and Well," *Instruments and Control Systems,* December 1977.

DeVries, W. R., *Microcomputer Applications in Manufacturing,* New York: Harper and Row, 1984.

FAST Applications Handbook 1987, South Portland, ME: National Semiconductor Corporation, 1988.

Global Spec, http://electronic-components.globalspec.com/LearnMore/Electrical_Electronic_Components/Relays_Timers/Solid_State_Timers_Time_Delay_Relays, December 2003.

Kalpakjian, S., *Manufacturing Engineering and Technology,* Reading, MA: Addison-Wesley, 1992.

Kaufman, M. and Seidman, A. H., *Handbook of Electronic Calculations for Engineers and Technicians,* New York: McGraw-Hill, 1988.

Kaufman, M. and Seidman, A. H., *Handbook for Electronic Engineering Technicians,* New York: McGraw-Hill, 1984.

Kosow, I. L., *Control of Electric Machines,* Englewood Cliffs, NJ: Prentice Hall, 1974.

Margolin, R., "Time Delay Relays," *Electronics Products Magazine,* May 1979.

"Meter Relays, Analog Controllers," *Measurements and Control,* February 1993.

Mitra, S. K., *Introduction to Digital and Analog Integrated Circuits and Applications,* New York: Harper and Row, 1984.

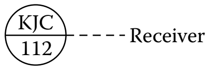
Schwartz, J. C., "Understanding Timers for Better Control," *Instruments and Control Systems,* December 1975.

Slater, N., "Solid-State Timing Shows Gains," *Design Engineering,* July 1980.

"Selecting and Specifying Time Delay Relays," *Instruments and Control Systems,* March 1979.

Strong, J. A., *Basic Digital Electronics,* London: Chapman and Hall, 1991.

5.13 Timers and Programming Timers



Multipoint on–off
time–sequencing
programmer

Flow sheet symbol

H. C. ROBERTS (1970) R. A. GILBERT (1985, 1995, 2005)

Types:	A. Mechanical B. Digital C. Electronic
Timing Ranges:	Typical ranges are 0.001 to 1 ms, 0.1 to 1 sec, 1 to 300 sec, 1 to 300 min, 1 to 24 h
Output Signals:	Voltage: 5 V DC, 12 V DC, 24 V DC, 24 V AC, 120 V AC
Currents:	From 20 mA to 500 mA
Sizes:	PC board assembly is 2 in. × 4 in. × 1 in. (50 mm × 100 mm × 25 mm) Plug-in component for single timer is about 2.5 in. × by 4 in. high (16 cm × by 100 mm high) Stand-alone chassis-based units are 10 in. × 12 in. × 16 in. (250 mm × 300 mm × 400 mm)
Costs:	B. $25 to $500 C. $10 to $300
Partial List of Suppliers:	Amperite (www.amperite.com) Crouzet (www.crouzet-usa.com) Magnecraft & Struthers-Dunn (www.magnecraft.com) Macromatic (www.macromatic.com) (A, B, C) Midtex (www.ciitech.com) and (www.tycoelectronics.com) NCC (www.natcon.com) (B, C) Potter & Brumfield (www.tycoelectronics.com)

INTRODUCTION

Time and the sequence of events create and characterize all processes. The human control of both defines a process as repeatable and inevitably determines if the process is industrially viable.

Timers and programming timers are tools that mark time and direct timed events. Timers and sequencers provide a method of controlling a process that has a prescribed set of states with expected time periods in each state. This section reviews the properties of digital and electronic times and includes a discussion of the fundamental aspects of monostable and a-stable multivibrators. Sequencers are also presented as synchronous and asynchronous devices.

This section begins with a brief review of such mechanical sequencers as the cam-operated timers, band-programming timers, and punched-card programmers.

MECHANICAL TIMERS AND SEQUENCERS

Historically, mechanical programming timers represented a landmark event in the industrial revolution. Today, these devices are still employed in manufacturing processes that have some degrees of freedom with respect to manufacturing tolerance or time required per manufacturing step. They are the devices of choice when the manufacturing environment is unfriendly to humans or to electronics, but they display a duel maintenance personality.

On the one hand, they are straightforward mechanical devices that can be repaired by most shop personnel. On the other hand, they are mechanical devices with moving parts that are subject to mechanical wear and sometimes bizarre mechanical failures.

Although most modern instrumentation and controls applications use digital or electronic timers and sequencers, they are in principle quite similar to cam-operated programming multiple-circuit time switches, drum programmers, band-programming timers, and punched-card programmers.

Cam Timers

Illustrations of both cam- and band-type mechanical devices are presented in Figure 5.13a. Cam-operated timers or time switches are provided with mechanical switches, which are arranged in line with an actuating cam attached to a shaft. A clock or a synchronous motor turns this cam shaft at the desired speed—one revolution corresponding to the desired

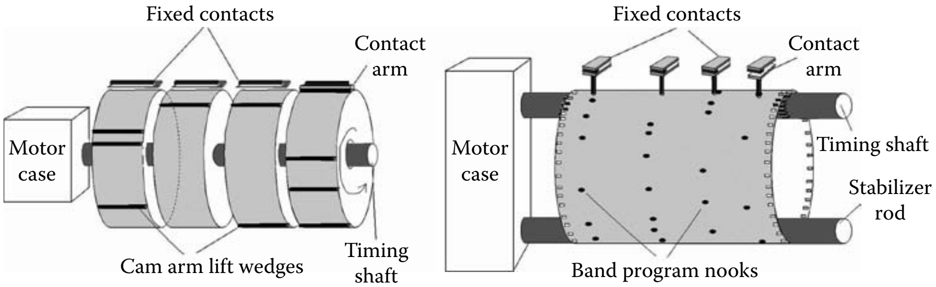


FIG. 5.13a
Illustration of cam-operated and band-operated timers.

cycle duration. During the cycle, each of the switches is closed and opened once (occasionally, two or three times are permissible) for the duration established by the cam settings.

Such a timer can control as many as 60 circuits, with a cycle as short as a few seconds or as long as weeks. Usually about 95% of the cycle can be programmed either “on” or “off” for each circuit; the contacts can handle currents as large as 10 A at 120 V. These timers are available in a variety of sizes and speeds.

Cam-operated timers are often used in conjunction with (either in cascade or in parallel) interval timers, which permits much greater flexibility of action. They are moderately priced and reliable, and their timing accuracy is high, limited by facility in adjustment. A notable limitation of cam-operated timers is that each switch can be actuated only a few times during each timing cycle. This limitation is entirely overcome in drum programmers, band-programming timers, and punched-card programmers.

Band or Drum Programmers

Band-programming timers consist of a belt, or band, of flexible metal or plastic material, several inches wide and as long as is needed to provided the required cycle length. An array of switches is provided, each with an actuating finger or contact arm resting on the band.

Although normally open contact band-programming timers are also available, the right side of Figure 5.13a illustrates the design where, when the band is intact, the switch is pressed closed, but where it is perforated the switch can open. As many as 80 such switches can be used; the band can be of any reasonable length. Many hours of operation can be controlled with many distinct on/off cycles for each switch during the total cycle.

Quite complicated switching patterns can be provided. In a band or drum programmer, the patterned band is often made to move against the switch array by a pulsed stepping motor rather than a simple synchronous motor. Timed pulses drive the stepping motor at varying rates according to a subsidiary pattern.

Punched-Card Programmers

Punched-card programmers operate analogously to band-programming devices. The cards are usually handled singly instead of in a continuous band. Such programmers are in fact quite similar to the Jacquard loom principle. Drum programmers are somewhat comparable in that a drum in which many movable studs are placed rotates in a time cycle under an array of switches.

TIMERS

Although the time of day is a significant factor when people are part of an industrial process, the process itself seldom depends on “the time of day.” The mechanical devices discussed earlier testify to this fact. Only the relative time in or between process steps and stages is important, with the real time not being a feature of any of these devices. The various time delay relays discussed in Section 5.12 are examples of control elements where time is important, used, but not logged.

One-Shot Timers

Figure 5.13b illustrates the “one shot” background concept for timers. The 74121 is representative of a mono-stable device that can produce a single square wave signal, a “one shot,” on command. Although this is a transistor-transistor logic (TTL) device, other technologies including a combination of discrete resistor and capacitor components can be used to create this pulse feature.

Figure 5.13b shows the 74121 pin on the left and an abbreviated time event log on the right. The log indicates that the two outputs, pin 1 and pin 6, respond to a rising edge event at pin 5 when pin 3 and pin 4 are at a low, logic “0,” signal level. Pin 5 is the input for a Schmitt trigger included in the 74121 package. This feature is useful when the triggering event has a slow rising edge.

The duration of the opposite output signals at pins 1 and 6 is the same and dependent on the value of the RC circuit connected to pins 14, 11, and 10. It is the ability to produce a “one shot” event with a desired and determined pulse length and duty cycle that defines the range and accuracy of a timer. A required increased value in both of these timing parameters is what dictates the increase in the cost of a timer.

Monostable and AStable Designs

A monostable device can produce a variety of timed event signals other than the “one shot” described above. Figure 5.13c

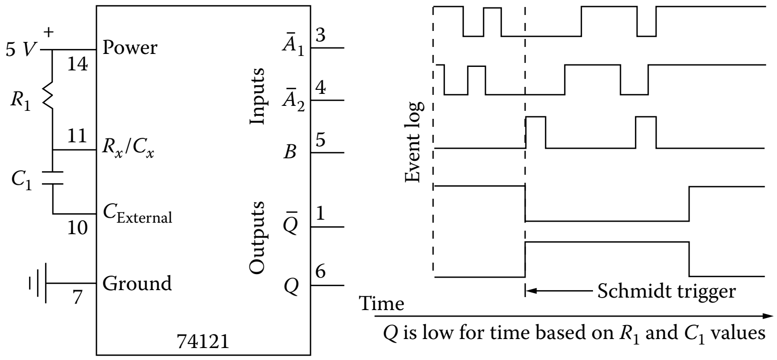


FIG. 5.13b
Pinout and event log for an example one shot timing circuit.

uses an event log to summarize the most popular types, but the corresponding 74121 circuits are not provided. Six different pulse event profiles are introduced: the One Shot, the Delay on Break, the True Delay on Break, the Delay on Make, the Interval On, and the Repeat Cycle. Drawing panels (i) through (v) are single pulse timing signal profiles, while panel (vi) represents a multiple pulse application discussed later.

In practice, the power is continuously supplied to the input voltage terminals of the one shot design. Upon closure of an active high external contact set, a trigger pulse is generated and the load is energized for the time dictated by an RC circuit. A maintained high signal on the external contact set will not affect the time the load is in the ON state.

Delay on Break

Panel (ii) of Figure 5.13c shows the operation of a Delay on Break. Power is also continuously applied to the input terminals, and upon closure of a normally open external initiate switch, the load transfers immediately and remains ON as long as the external contacts are closed. The load will remain ON after the external trigger contacts return to the normal state. The time delay is started instantly by opening these contacts. At the end of this delay period the load de-energizes and the timer is ready for another cycle.

In the True Delay on Break application, the supply voltage energizes the load. The timing period begins upon removal of this input voltage. At the conclusion of the fixed time period, the load is de-energized. In the Delay on Make, the timing period follows the application of the input voltage. Once the fixed timing period has expired, the output is energized. The load stays energized until it is reset by interrupting the input voltage.

Finally, the Interval On energizes the load immediately upon the application of the input voltage. At the end of this fixed time delay, the timer transfers back to its prepower position. Like the other timers summarized in Figure 5.13c, the removal and reapplication of the input voltage resets the timer.

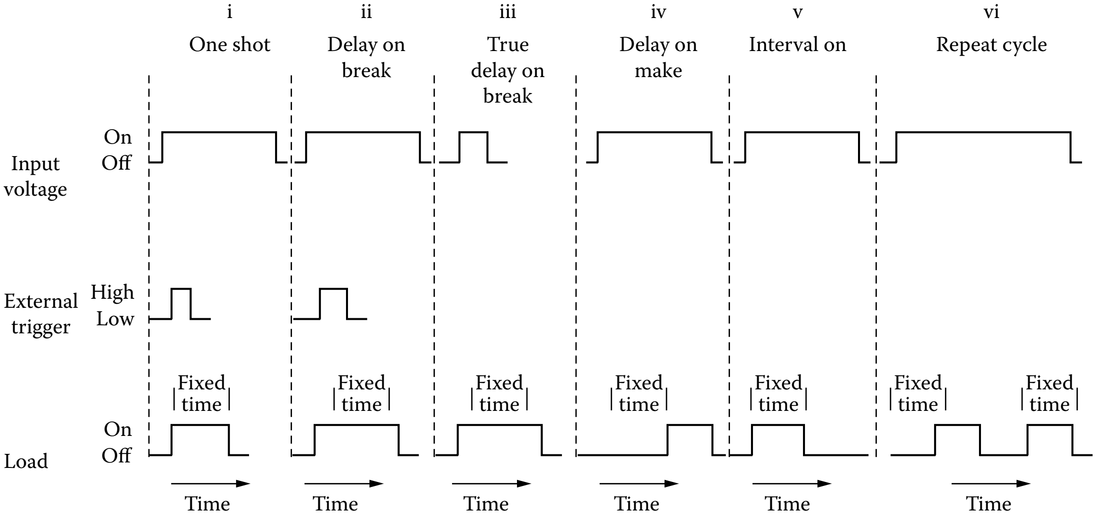


FIG 5.13c
Monostable and astable mode configurations for a 555 device.

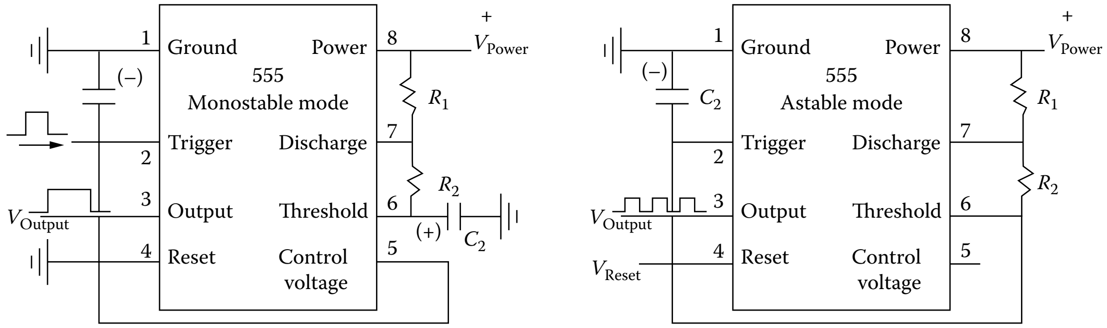


FIG. 5.13d
Example of a serial configuration of two 555 devices.

555 Devices

As suggested above, the "one shot" can be created in different technologies. Figure 5.13d shows two configurations for the 555. The 555 is a semiconductor-based device that can be configured as a "one shot." The output signal from this monostable mode shown on the left in the figure functions identically as the 74121 but differs from the 74121 "one shot" output with respect to current and pulse duration. The 555 output pin can deal with currents at least 10 times higher than the 74121's nominal 16 mA, as well as extended pulse duration times.

The astable 555 configuration is illustrated on the right side of Figure 5.13d. The diagram emphasizes that the difference between the monostable and astable modes is the feedback of the trigger, pin 2, signal to the threshold pin, pin 6, in the a-stable mode. In this mode, the 555 will adopt the oscillator role in a timer.

The 555 astable multivibrator circuit configuration shown is a typical arrangement for a digital sequencer time-based oscillator. The circuit generates a continuous train of clock pulses. A review of some of the circuit connections emphasizes the function of the 555. When the circuit is first engaged, the capacitor, C_2, connected between pin 2 and ground discharges and pin 2, is activated as the voltage drops below 1.4 V.

Once activated, the trigger sets the output line high and brings the discharge terminal, pin 7, to an open circuit state. Because the connection from the 555 to the junction of R_2 and R_3 is now essentially open, the capacitor recharges through the two resistors until the threshold voltages reach approximately 3.6 V. Once this voltage level has been reached, pin 6 is activated, pin 2 is deactivated, and the output and discharge pins move to 0 V. The 0 output on pin 7 provides a path for the capacitor to discharge through R_3.

When the capacitor discharge drops the voltage to one-third the supply voltage, the trigger input is reactivated, and the process starts over. It should be noted that all of the figures in this section omit some circuit details for the sake of concept illustration. In this case, pin 5 of the 555 astable configuration is shown with no connection. Often this pin is connected to ground by means of a 0.01 mF capacitor. Pin 5 is the control voltage input for the threshold voltage.

It is good practice to use the capacitor connection to ground because the capacitor and R_2 form a decoupling element that prevents power supply noise from altering the threshold-input voltage. The fundamental clock frequency for a 555 as an astable multivibrator can be calculated according to the formula:

$$1.44/(R_2 + R_3)C_z \qquad \textbf{5.13(1)}$$

The duty cycle, which in this case is the ratio of the time spent in the logic 0 state to the sum of the times spent in the logic 0 and 1 state, has a calculation formula:

$$D = R_3/(R_2 + 2R_3) \qquad \textbf{5.13(2)}$$

Thus, a duty cycle of 0.5 produces a square wave and is available when $R_3 >> R_2$.

Hybrid Timing Circuits

Figure 5.13e presents two a-stable devices connected in series. This arrangement allows a cluster of pulses, three shown in this diagram, to be delivered within a single pulse, self-triggered by the 555 on the left side of the diagram. Additional versatility is possible because the duty cycle of either a-stable device is adjustable. This concept is also commercially available with the 555 or equivalent on the left of the figure configured in the monostable mode. With this monostable circuit adjustment, the example suggests the fundamental characteristic of the Repeat Cycle timer outlined in panel (vi) of Figure 5.13c.

The assimilation of an a-stable oscillator into an electronic timer is straightforward. The illustrative example in Figure 5.13f merges the 555 in its a-stable mode to an operational amplifier. This combination emphasizes the manipulation of the output signal for the combination circuit at pin 6 of the arbitrarily selected AD 645. The input signal, $V(t)_{in}$, influences the shape and magnitude of the output signal. Although a ramp-input signal is portrayed, a constant signal is a common option reflecting the power needs of the intended application for the timer.

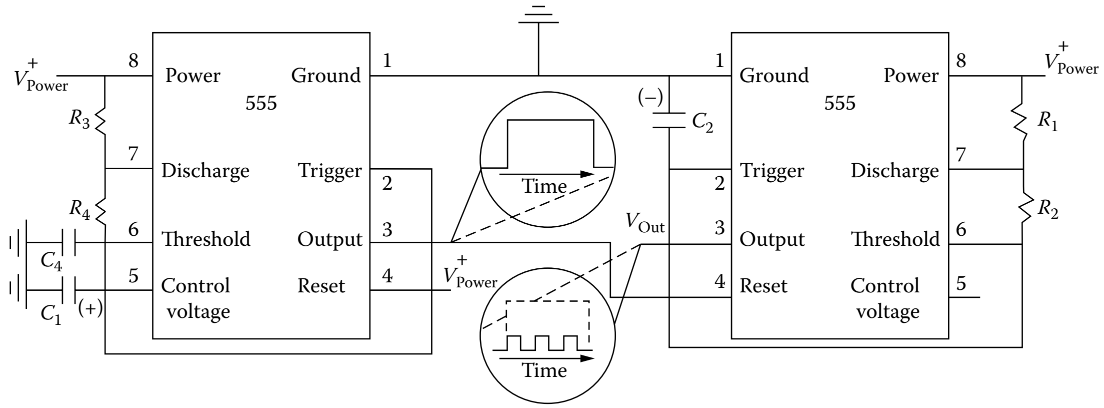


FIG. 5.13e
Elements of a digital-to-analog hybrid electronic timing circuit.

DIGITAL SEQUENCERS

The digital sequencer is a relatively simple digital circuit that mimics the behavior of mechanical sequencers without the mechanical mass and movement inherent in such devices. A simplified synchronous digital sequencer that summarizes the rudimentary idea is illustrated in Figure 5.13g. An understanding of these components and their role in the sequencer will provide a sound basis for selecting and using commercially available digital or electronic programmable sequencers.

The heart of any programmable sequencer is the element that provides the mechanism to generate the timing sequence. For mechanical sequencers it is the motor with the desired gear ratio. For the digital and electronic sequencers it is the a-stable multivibrator. As suggested above, an astable multivibrator is a device that repeatedly changes output states from one extreme to another. The duration of time in one of these extreme states is determined by resistance and capacitance elements in the astable circuit. The 555 in Figure 5.13g fills this role.

The figure illustrates another critical element of a digital sequencer. The 4017 has eight outputs that will go low in turn as a string of pulses is delivered to its clock pin, pin 14. The event log for these outputs and the synchronous pin, pin 3, is provided to the right of the 4017 device pin out. Once the 4017 is enabled via a low signal placed on pin 13, the 4017 responds to the next clock pulse by bringing pin 3 and pin 2 to logic “0.”

There are countless applications for synchronous timers and sequencers. The operation is straightforward and the arrangement of events is usually fixed and easy to determine.

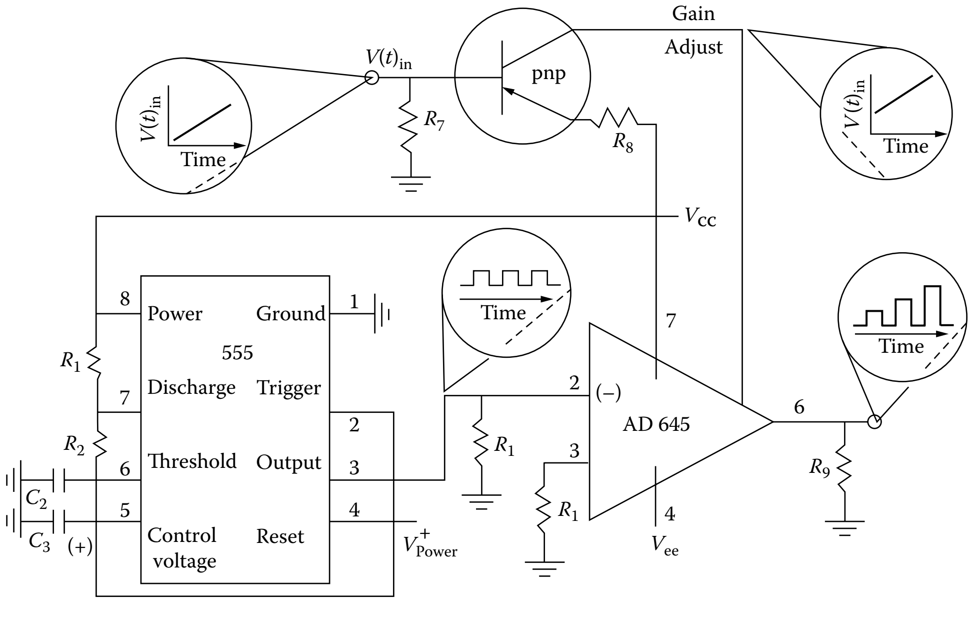


FIG. 5.13f
The assimilation of an astable oscillator into an electronic timer.

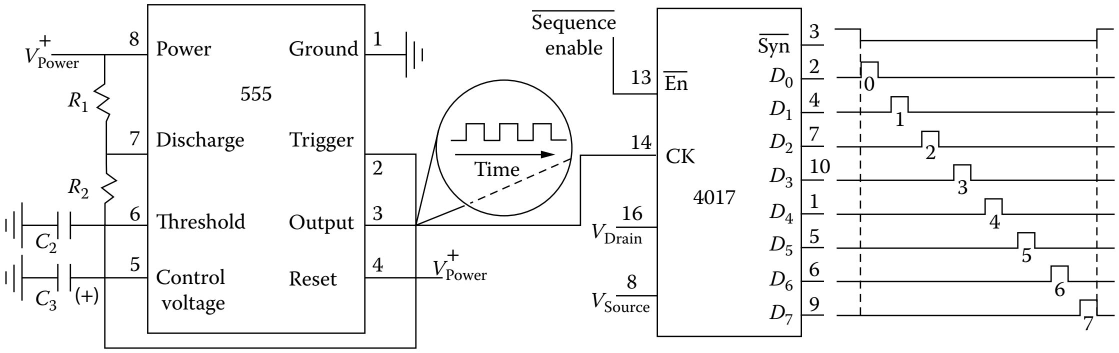


FIG. 5.13g
Elements of a simple synchronous sequencer circuit.

However, synchronous control applications all have the same potential drawback. The time for each step in the process will be fixed by the task that requires the most time.

This downside can be overcome by the use of timers that permit programmed time intervals between process events. Electromechanical timers are very popular in this situation, because they usually use a set of pegs or wedges that mark units of time in which the timer is in the ON or the OFF position.

Asynchronous Sequencers

The idea that real time is not a major issue in a sequential process timer application was introduced above. Upon closer consideration, it is also true that time itself is not the focus of attention during many multiple-step processes. The issue usually reduces down to the fact that each step in a process sequence requires a characteristic time, and then it is time to move to the next process step. In addition, many sequential processes may require other asynchronous process actions to be accomplished before the current step is completed. Figure 5.13h outlines the characteristics of an asynchronous sequencer for such applications.

Figure 5.13h introduces several digital components not employed in previous figures. The event mark is provided by the 555 configured as a monostable oscillator. This device triggers the next process event in the sequence, and that pulse is generated when the 9324 comparator device indicates that the binary pattern on its B_4, B_3, B_2, and B_1 inputs are identical to the pattern presented to its A_4, A_3, A_2, and A_1 inputs, respectively. The single output pulse from the 555 will generate a new counter value on the Q_C, Q_B, and Q_A outputs of the 74192 synchronous counter.

This three-bit output pattern will have a value one bit higher than the previous count. If, for example, the pattern 010, the number 2_8, were displayed, then 3_8, 011, would be the next pattern to be presented at the 74192 Q outputs. This new pattern, 011, is delivered to the inputs of the decoder, in this example the 74138, which proceeds to simultaneously

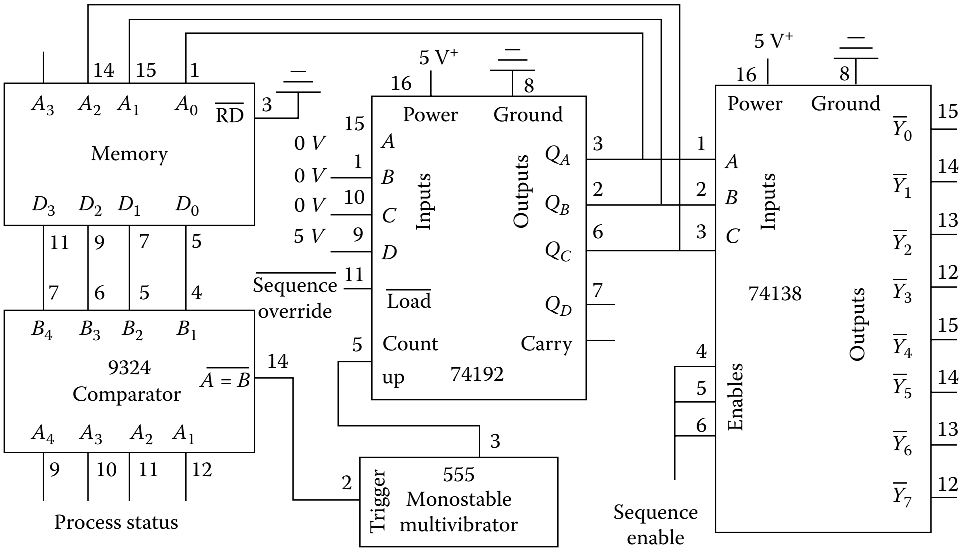


FIG. 5.13h
Illustrative example of a simplified asynchronous sequencer.

return output Y_2 to logic “1” and bring output Y_3 to logic “0.” All of the other 74138 *Y* outputs remain at logic “1,” and the process sequence moves from the process associated with the low signal at Y_2 to the process associated with the low signal at Y_3.

As this stepping from one process in the sequence to the next process occurs, the 74192 also delivers its count value to the address pins, A_2, A_1, and A_0 of a memory device. At this point in the circuit operation, the contents of this memory location are presented to the *B* inputs of the 9324, which are in turn compared to the bit pattern presented to the *A* inputs of the same 9324.

Although the abbreviated example in Figure 5.13h still has several contiguous logic steps, the operational scheme is quite simple. The content of each memory location corresponds to the expected status of the sequential process at the end of each particular step. The current status of the process step in question is delivered to port *A* of the 9324.

When the step is completed, the current status will match the expected status from the memory location; the 9324’s $A = B$ pin, pin 14, goes low to trigger the 555; and the next decoder channel, in this case Y_4, goes low. As long as the correct end status condition for each sequential step of the process is known and stored in the memory device, a digital asynchronous sequencing circuit will move the process from step to step without any need for timed clock events.

Finally, Figure 5.13h does not show the necessary additional circuitry needed to interface the output signals from the eight decoder output channels to the actual process switching circuit elements. Nor does it present the three-state or open collector nature of the current process status input circuit associated with port *A* of the 9324. These interfacing circuit requirements differ for various applications. However, most common industrial interface requirements are compatible with available digital sequencers.

ELECTRONIC SEQUENCERS

Electronic timers and sequencers are hybrid circuits that use active and passive circuit elements. They have a distinct size advantage over mechanical sequencers and provide some logic control functions well beyond the capabilities of their mechanical counterparts. Solid-state augmented electronic sequencers can function as a timed mechanical stepping switch.

When the circuit is powered up it starts out in the first operation step. When an event external to the circuit—typically, a start signal—activates the enable switch, the unit steps through a number of active positions preset at predetermined ON then OFF time intervals. After the completion of the last operation time cycle, the circuit returns to a passive state to await a new external start command.

Asynchronous electronic sequencers are also available. In these circuits, the current process step remains active until a new enable pulse is generated by the system under control. Once triggered, the electronic sequencer de-energizes the current solenoid and then activates the next.

Both synchronous and asynchronous electronic timers and sequencers are popular in dedicated and OEM applications. Usually available as small single circuit board units, they easily interface with existing electronic and hardware components. Although maintenance at the component level is possible, these electronic circuits can be purchased as modular assemblies and thus are suited for a modular replacement-based maintenance philosophy.

CONCLUSIONS

Electronic and digital sequencers have major advantages over mechanical sequencers with respect to versatility, performance, and size; however, when compared to each other, performance is not usually an issue. Electronic sequencers can be miniature and have thick conformal coatings to protect them from the environment.

Larger digital sequencers, however, have more versatility. In addition, most commercially available digital timers and sequencers can be computer interfaced via an RS-232 or other specified industry-related standard interface.

Both types of sequencer technologies provide a variety of timing actions (e.g., on delay, off delay, interval delay), timing ranges, and output relay actions (e.g., normally open, normally closed, DPDT contacts). Both are adaptable to various operating voltages and currents, and both have performance and environment parameters that exceed most industrial application requirements.

Bibliography

Asfahl, C. R., *Robots and Manufacturing Automation,* New York: Wiley 1992.

Carr, J., *Electronic Circuit Guidebook: IC Timers,* Volume 2, New York: Delmar Learning, 1997.

Cutts, D., “Timers and Counters Gaining Functionality,” *Control and Instrumentation,* Vol. 18, No. 6, 1986.

Dazahine, D., Pouxe, J., and Rossetto, O., *Design and performances of a compensated mean-timer,* New York: IEEE Transactions on Nuclear Science, 2000.

DeVries, W. R., *Microcomputer Applications in Manufacturing,* New York: Harper & Row, 1984.

FAST Applications Handbook 1987, South Portland, ME: National Semiconductor Corporation, 1988.

Gilbert, R. A. and Llewellyn, J. A., *Principles and Applications of Digital Devices,* Research Triangle Park, NC: Instrument Society of America, 1982.

Kalpakjian, S., *Manufacturing Engineering and Technology,* Reading, MA: Addison-Wesley, 1992.

Kaufman, M. and Seidman, A. H., *Handbook of Electronic Calculations for Engineers and Technicians,* New York: McGraw-Hill, 1988.

Kaufman, M. and Seidman, A. H., *Handbook for Electronic Engineering Technicians,* New York: McGraw-Hill, 1984.

Madden, D.G., “Drum Programme,” *Instrumentation Technology,* August 1970.

"Meter Relays, Analog Controllers," *Measurements and Control,* February 1994.

Mims, F. and Helms, H., *Mims Circuit Scrapbook, Volume 2,* New York: Elsevier Science & Technology Books, 2000.

Mitra, S. K., *Introduction to Digital and Analog Integrated Circuits and Applications,* New York: Harper & Row, 1984.

Schwartz, J. C., "Understanding Timers for Better Control," *Instruments and Control Systems,* December 1975.

Strong, J. A., *Basic Digital Electronics,* London: Chapman and Hall, 1991.

Tocci, R. J. and Widmer, N. S., *Digital Systems: Principles and Applications*, 8th edition, New York: Prentice Hall, 2000.

Xiao, X., Yi-Chium, C., Gonen, E., and Peijuan, L., ‘‘Simulation Analysis of RLC Timers in UMTS Systems," *Proceeding of the Winter Simulation Conference,* Vol. 1, 2002.

Control Valve Selection and Sizing

6

6.1
APPLICATION AND SELECTION OF CONTROL VALVES 1050

6.2
ACCESSORIES AND POSITIONERS 1087

6.8
DIAGNOSTICS AND PREDICTIVE VALVE MAINTENANCE 1161

6.9
DYNAMIC PERFORMANCE OF CONTROL VALVES 1165

6.10
EMERGENCY PARTIAL-STROKE TESTING OF BLOCK VALVES 1172

6.11
FIELDBUS AND SMART VALVES 1182

6.12
INTELLIGENT VALVES, POSITIONERS, ACCESSORIES 1193

6.13
MISCELLANEOUS VALVE AND TRIM DESIGNS 1199

6.14
VALVES: NOISE CALCULATION, PREDICTION, AND REDUCTION 1213

6.15
SIZING 1234

6.16
VALVE TYPES: BALL VALVES 1262

6.17
VALVE TYPES: BUTTERFLY VALVES 1273

6.18
VALVE TYPES: DIGITAL VALVES 1284

6.19
VALVE TYPES: GLOBE VALVES 1290

6.20
VALVE TYPES: PINCH VALVES 1323

6.21
VALVE TYPES: PLUG VALVES 1341

6.22
VALVE TYPES: SAUNDERS DIAPHRAGM VALVES 1348

6.23
VALVE TYPES: SLIDING GATE VALVES 1353

6.1 Application and Selection of Control Valves

B. G. LIPTÁK (1970, 1985, 1995, 2005) **A. BÁLINT** (2005)

Subjects Covered in this Section:

Actuators; see also Sections 6.3 and 6.4
Cavitation; see also Section 6.15
Characteristics; see also Section 6.7
Corrosion; see also Section 6.19
Erosion; see also Section 6.19
Fire safety; see also Sections 6.16 and 6.17
Flashing and erosion; see also Section 6.15
Gain; see also Section 6.7
High pressure drop applications; see also Section 6.15
High-pressure services
High-temperature services
Installation considerations
Intelligent valve features; see also Section 6.12
Jacketed valves
Leakage; see also Sections 6.16–6.23
Low-temperature services (cryogenics); see also Section 6.23
Noise abatement; see also Section 6.14 and 6.17
Packing designs; see also Section 6.19
Positioners; see also Section 6.2
Process data
Rangeability; see also Section 6.7
Selection chart for control valves
Sequencing, split-ranging
Sizing; see also Section 6.15
Small flow applications; see also Section 6.23
Specification forms for control valves
Toxic applications; see also Section 6.23
Vacuum services
Viscous and slurry services

INTRODUCTION

In the field of control valve design, the most important developments of the last decade occurred in the areas of electric and digital actuators (Section 6.3), in valve diagnostics (Section 6.8), dynamic performance evaluation (Section 6.9), safety shutdown systems (Section 6.10), fieldbus interaction (Section 6.11), intelligent positioners (Section 6.12), valve status detection and use for control (Section 6.13), and in the increased availability of special valve designs (Section 6.12). Because each of these topics are covered in the noted separate sections, they are not treated in detail in this section.

While this section attempts to discuss all basic aspects of control valve selection and application, in this area too, there exists some overlap with other sections. For example, as noted in the alphabetic listing above, the topics of valve characteristics and rangeability are discussed in more detail in Section 6.7; noise and its reduction in Section 6.14; sizing in Section 6.15; valve actuators and accessories including positioners in Sections 6.2 and 6.4; and the features of the particular valve designs in Sections 6.16 to 6.23. Therefore, the reader is advised to treat this section only as an overview of the subject of control valve applications and refer to the individual sections of this chapter for the detailed discussion of its many specific aspects.

It should also be noted that in most of the sections in this chapter English units are used with their SI equivalents given in parenthesis. An exception is Section 6.14 on valve noise calculation, which follows the general practice of acoustics and uses SI units. Appendices 1 and 2 (at the end of this handbook) give all the conversion factors that are required to go from one to the other system of units.

So, for example, 1 lb/in.^{2} equals 6.89 kPa (kilo Pascals) or 0.0689 bars. Hence, a 3–15 PSIG range is the approximate equivalent of 20–100 kPa. The valve capacity coefficient used in this chapter is the C_v, which is unity, if the valve passes 1.0 gpm of cold water at a specific gravity of 1.0 at a pressure drop of 1.0 psid. The metric equivalent is the K_v, which corresponds to a valve passing 1 m^3 of cold water per hour at a pressure drop of 1 bar. Therefore, if the reader wishes to convert any *Cv* value into K_v, the multiplier is 1.17, and therefore $C_v = 1.17\ K_v$.

Orientation Table

In order to provide some overall orientation about the relative merits of the different valve designs, an orientation table, Table 6.1a, was prepared. In this table some of the more common applications are described, together with some indications of the suitability of the various valve designs. At the end of this section, a standard control valve specification form, prepared by Instrumentation, Systems, and Automation Society, is provided together with an explanation of each of the entries in that form. The discussion of the various topics related to valve selection follows the approximate order of the entries in that form. Therefore, after some introductory remarks, the discussion begins with topics related to the service conditions (process data) and then continues with topics related to the features and accessories of the valve and its installation.

CONTROL VALVE TRENDS

When this handbook was first published some 35 years ago, the overwhelming majority of throttling control valves were the globe types, characterized by linear plug movements and actuated by spring-and-diaphragm operators. At that time, the rotary valves were considered to be on/off shut-off devices. Today, globe valves are still widely used, but their dominance is being challenged by the less expensive rotary (ball, butterfly, plug) valves, which are usually actuated by cylinder operators. This trend represents a mixed blessing and, therefore, is worth further discussion.

Globe vs. Rotary Valves

The main advantages of the traditional globe design include the simplicity of the spring-and-diaphragm actuator; the availability of a wide range of valve characteristics; the relatively low likelihood of cavitation and noise; the availability of a wide variety of specialized designs for corrosive, abrasive, and high/low temperature and pressure applications; the linear relationship between control signal and valve stem movement; and the relatively small amounts of dead band and hysteresis in its operation. These features make the globe valve usable without positioners, which on fast processes is an advantage.

The main reason why rotary valves have been increasing their market share is their lower manufacturing cost and higher relative flow capacity ($C_d = 20$–40 instead of $C_d = 10$–15, as for globe). They also weigh less, can act as both control and shut-off valves, and are easier to seal at the stem to meet OSHA and EPA requirements.

The limitations of globe valves, in addition to their higher cost per unit C_v (in Europe, the equivalent term K_v is used), include their relatively slow speed and low “stiffness” (plug position is affected by dynamic forces in the process fluid), both of which can be improved by using hydraulic cylinder actuators operating at higher pressures.

Major disadvantages of rotary valves are their higher tendency to cavitate and to produce excessive amounts of noise. They are also more likely, due to their smaller size per unit C_v (K_v), to have larger pipe reducers with the associated waste of pressure drop and distortion of characteristics. Their control quality can suffer from the nonlinear relationship between actuator linear movement and valve rotation, plus from the linkages, which can introduce substantial hysteresis and dead play. These characteristics result, in most cases, in a definite requirement for using a positioner, which on fast processes can cause the deterioration of control quality.

Valves vs. Other Final Control Elements

Before proceeding through the steps of selecting a control valve, one should evaluate if a control valve is truly needed in the first place, or if a simpler and more elegant system will result through some other means. For example, an overflow weir can suffice to keep levels below maximum limits, and choke or restriction fittings can serve the function of pressure letdown at constant loads. In other locations, it might be possible to reduce the investment by using regulators instead of control valves.

The advantages of regulators include their high speed (high gain) and their self-contained nature, which eliminates the need for power supplies or utilities. If remote set point adjustment is needed, regulators can be provided with air-loaded pilots to accommodate that requirement. While all regulators (being proportional-only controllers) will display some offset as the load changes, the amount of offset can be minimized by maximizing the regulator gain.

In still other applications, it is prudent to replace whole flow control loops with positive-displacement metering pumps or to replace the control valve with variable-speed centrifugal pumps. The cost-effectiveness of the approach is usually found to be in lowered pumping costs, because the pumping energy that was “burned up” in the form of pressure drop through the control valve is not being introduced, and therefore it is saved.

CONTROL VALVE SIZING

Control valve sizing is discussed in depth in Section 6.15, and therefore only a few general recommendations are made here.

One should first determine both the minimum and maximum C_v (K_v in Europe) requirements for the valve, considering not only normal but also start-up and emergency conditions. The selected valve should perform adequately over

TABLE 6.1a

Orientation Table for Selecting the Right Control Valves for Various Applications

Features & Applications	*Control Valve Types*															
	Ball: Conventional	*Ball: Characterized*	*Butterfly: Conventional*	*Butterfly: High-performance*	*Digital*	*Globe: Single-ported*	*Globe: Double-ported*	*Globe: Angle*	*Globe: Eccentric disc*	*Pinch*	*Plug: Conventional*	*Plug: Characterized*	*Saunders*	*Sliding gate: V-Insert*	*Sliding gate: Positioned disc*	*Special: Dynamically balanced*
Features:																
ANSI class pressure rating (max.)	2500	600	300	600	2500	2500	2500	2500	600	150	2500	300	150	150	2500	1500
Max. capacity (C_d)	45	25	40	25	14	12	15	12	13	60	35	25	20	30	10	30
Characteristics	F	G	P	F, G	E	E	E	E	G	P	P	F, G	P, F	F	F	F, G
Corrosive Service	E	E	G	G	F, G	G, E	G, E	G, E	F, G	G	G, E	G	G	F, G	G	G, E
Cost (relative to single-port globe)	0.7	0.9	0.6	0.9	3.0	1.0	1.2	1.1	1.0	0.5	0.7	0.9	0.6	1.0	2.0	1.5
Cryogenic service	A	S	A	A	A	A	A	A	A	NA	A	S	NA	A	NA	NA
High pressure drop (over 200 PSI)	A	A	NA	A	E	G	G	E	A	NA	A	A	NA	NA	E	E
High temperature (over 500°F)	Y	S	E	G	Y	Y	Y	Y	Y	NA	S	S	NA	NA	S	NA
Leakage (ANSI class)	V	IV	I	IV	V	IV	II	IV	IV	IV	IV	IV	V	I	IV	II
Liquids:																
Abrasive service	C	C	NA	NA	P	G	G	E	G	G, E	F, G	F, G	F, G	NA	E	G
Cavitation resistance	L	L	L	L	M	H	H	H	M	NA	L	L	NA	L	H	M
Dirty service	G	G	F	G	NA	F, G	F	G	F, G	E	G	G	G, E	G	F	F
Flashing applications	P	P	P	F	F	G	G	E	G	F	P	P	F	P	G	P
Slurry including fibrous service	G	G	F	F	NA	F, G	F, G	G, E	F, G	E	G	G	E	G	P	F
Viscous service	G	G	G	G	F	G	F, G	G, E	F, G	G, E	G	G	G, E	F	F	F
Gas/Vapor:																
Abrasive, erosive	C	C	F	F	P	G	G	E	F, G	G, E	F, G	F, G	G	NA	E	E
Dirty	G	G	G	G	NA	G	F, G	G	F, G	G	G	G	G	G	F	G

Abbreviations:

A = Available
C = All-ceramic design available
F = Fair
G = Good
E = Excellent
H = High

L = Low
M = Medium
NA = Not available
P = Poor
S = Special designs only
Y = Yes

a range of 0.8 $C_{v\min}$ to 1.2 $C_{v\max}$. If this results in a rangeability requirement that exceeds the capabilities on one valve, use two or more valves.

Control valves should not be operated outside their rangeability. Driskell (see Bibliography) properly points to the fact that all "fat" settles in the control valve. In constant speed pumping systems, each design engineer will add their own safety margin in calculating pressure drops through pipes and exchangers, and finally in selecting the pump.

Therefore, the control valve will end up with all these safety margins as added pressure drops, resulting in a much-oversized valve. A highly oversized valve will operate in a nearly closed state, which is an unstable and undesirable operating condition. In variable-speed pumping systems, this problem does not exist, because there the pump speed is adjusted to meet the load, and therefore the effect of accumulated safety margins is eliminated.

Collecting the Process Data

In order to select the right control valve, one must fully understand the process that the valve controls. Fully understanding the process means not only understanding normal operating conditions, but also the requirements that the valve must live up to during start-up, shutdown, and emergency conditions. Therefore, all anticipated values of flow rates, pressures, vapor pressures, densities, temperatures, and viscosities must be identified in the process of collecting the data for sizing.

In addition, it is desirable to identify the sources and natures of potential disturbances and process upsets. One should also determine the control quality requirements, so as to identify the tolerances that are acceptable in controlling the particular variable. The process data should also state if the valve needs to give tight shut-off, if the valve noise needs to be limited, or any other factors that might not be known to the instrument engineer. These can include subjective factors, such as user preferences, or objective ones, such as spare parts availability, delivery, life expectancy, or maintenance history.

Lines 1–12 in the "Specification Form for Control Valves" (at the end of this section) describe the service conditions (process data) that must be provided for the control valve. It is important to carefully determine not only the "normal" values for this data but also the "minimum" and "maximum" values, because the valve must operate properly throughout its range — not just under normal conditions.

Determining the Valve Pressure Drop

Assigning the sizing pressure drop for the valve is more complex than picking a number like 10% or 25% of the total system drop or a number like 10 or 25 psi (0.69 or 1.72 bar). It requires an understanding of the interrelationships that exist in pumping, fan, or compressor systems. If a system consists of nothing else but a pump, a control valve, piping, vessels, and an elevated destination, the energy profile through the system will be as illustrated in Figure 6.1b.

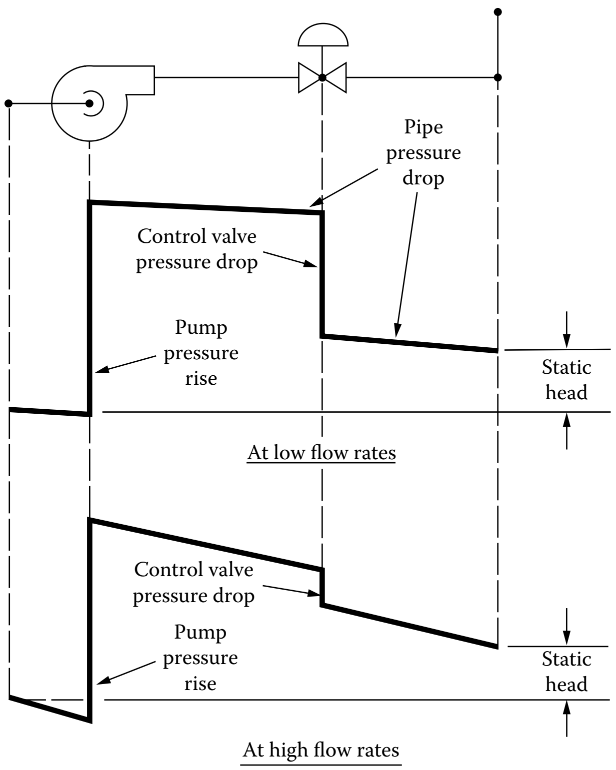


FIG. 6.1b
Pressure profiles of a pumping system at high and low flow rates.

Note that the pressure drop available for the control valve drops as the flow rate rises. This is because at higher flows, the pump discharge pressure will be lower, while the pressure drop through the piping will rise. In other words, the control valve does not work with a fixed pressure drop, nor with a fixed percentage of the total system drop, but it simply takes whatever is left over from what is available and what is required by the rest of the system. Therefore, as shown in Figure 6.1c, the valve energy loss at any particular flow rate (load) is the difference between the corresponding points on the pump and system curves.

As can be seen in Figure 6.1c, the available pressure drop increases as the load (flow rate) drops, and it becomes the minimum when the process flow rate (load) is the maximum. This means that the actual control valve rangeability (in terms of C_v) must be much larger than the ratio of maximum and minimum flows. There is a similar effect on the valve characteristics, which will be discussed (together with rangeability) in both the paragraphs that follow and also in more detail in Section 6.7.

Because of the complexity of the problem, no simple rule of thumb can be used in assigning the valve pressure drop; thus, it is important for the engineer to have a good understanding of the pump and system curves together with the rangeability and characteristics (gain) requirements of the valve. In the process of deciding what pressure drop to assign, you must acknowledge the conflict between control quality and energy conservation. The higher the pressure drop through the valve (relative to the rest of the system), the more

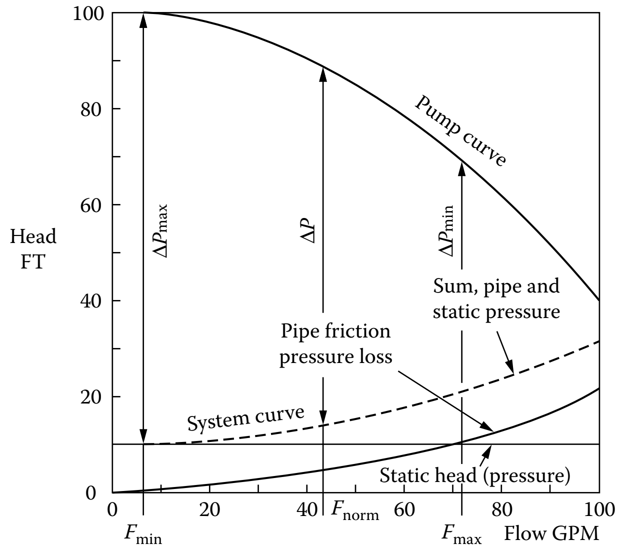


FIG. 6.1c
The difference between the pump discharge pressure curve and the system curve (which is the sum of the static head and the pipe friction loss) is the available valve differential.

impact it will have on the process while, at the same time, the more pumping energy it will waste.

Some might argue that, in fact, no pressure drop needs to be assigned to the control valve during the design phase because as safety margins accumulate, the pump will be oversized anyway. There is some practical wisdom in this attitude, because it is true that by the time the pump service conditions pass from the process engineer to the mechanical engineer, then are sent out for bidding to the manufacturers, and finally a pump is selected, its flow and pressure capabilities will always much exceed the originally specified requirements, and there will be plenty of pressure drop for the valve. While this argument sounds convincing and convenient, it is wrong. It is wrong because it results in unpredictable performance (possibly high noise) and usually also results in oversized control valves, which tend not only to be unstable, but also to have low rangeability.

Therefore, the proper approach to the selection of valve pressure drop is to first determine the total *friction energy loss* (excluding static energy) of the system at *normal* load (flow) and assign 50% of that to valve pressure energy drop. Based on that assignment, one should next determine the resulting valve drop at minimum and maximum loads (flows) and select a valve that can handle the required C_v rangeability. As will be discussed later, one should also select a valve characteristic that, after being "distorted" by the change in valve drop as the load varies, will give acceptable (stable) loop performance.

CHARACTERISTICS, GAIN, AND RANGEABILITY

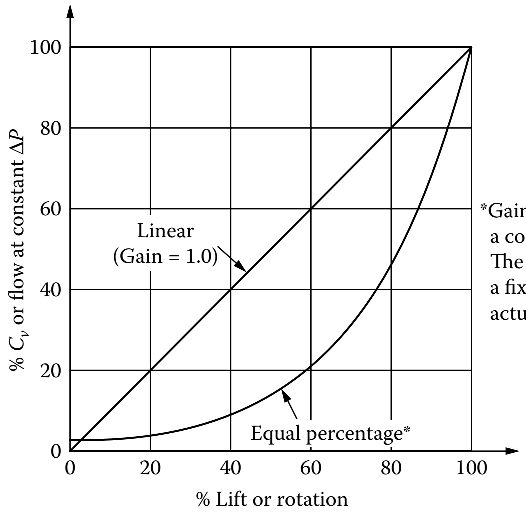
Good control valve performance usually means that the valve is stable across its full operating range, it is not operating near to one of its extreme positions, it is fast enough to correct for process upsets or disturbances, and it will not be necessary to retune the controller every time the process load changes. In order to meet the above goals, one must consider such factors as valve characteristics, rangeability, installed gain, and actuator response. These topics will be separately addressed here and in Section 6.7.

Characteristics and Gain

Characteristics and gain are discussed in more detail in Section 6.7, but they are also briefly covered here. The reason process control engineers must be concerned with "selecting" the right valve characteristics (Figure 6.1d) is because the valve is part of the control loop, and the loop will be stable only if the products of all its gain components (the gains of the process, sensor, controller, and valve gains) is constant. Usually, the controller is tuned so that this gain product is 0.5, in order to give quarter-amplitude damping. This was discussed in some detail in connection with Figure 2.1x in Chapter 2.

If the gains of the loop components do not vary with load, but are constant, the desirable choice is to use a constant gain control valve. A constant gain valve is a linear valve whose theoretical gain (change of flow per unit change of lift) is 1. If the gain of any of the loop components (such as the process) decreases with load (flow through the valve), the proper choice of control valve gain is 1, which increases with load (equal percentage), because this combination will keep the gain product of the loop relatively constant.

Installation Causes Distortion As was pointed out in connection with Figure 6.1c, in mostly friction systems, such as pumping through long pipes (where only a small portion of the total pump energy is used to overcome constant static pressure),



*Gain increases at a constant slope. The rate of rise is a fixed % of the actual flow.

FIG. 6.1d
Inherent characteristics of control valves.

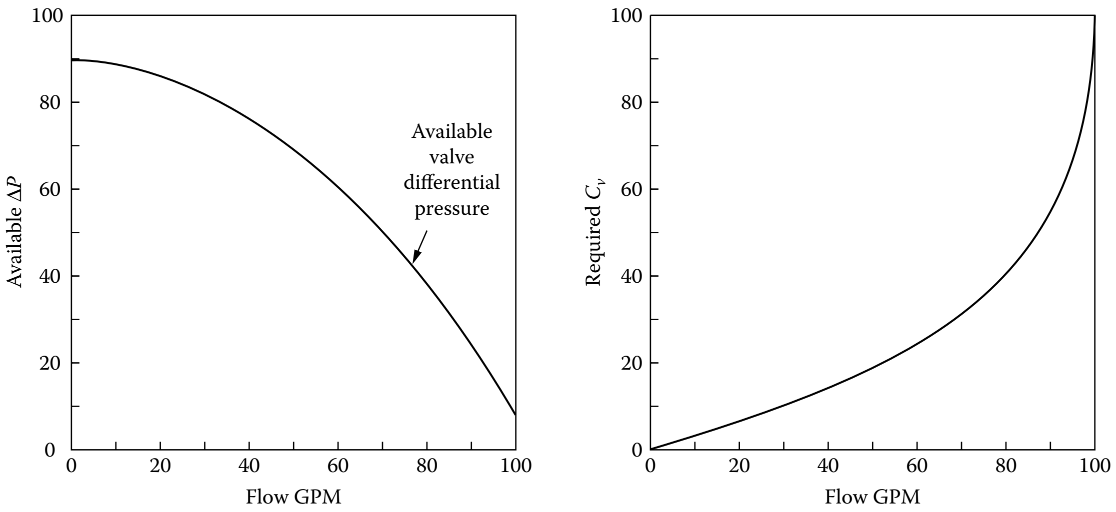


FIG. 6.1e

In mostly friction systems, an increase in load (flow rate) results in a drop in the pressure drop, which is available for the control valve. Therefore, the same amount of increase in flow rate requires a larger increase in the valve capacity coefficient C_v (K_v). For such application, an equal-percentage valve is needed.

the pressure drop available for the control valve is dropping as the load (flow rate) is increasing, and therefore more C_v is needed for unit flow increases as the flow rises (Figure 6.1e).

This is a different condition from the condition at which the valve characteristics were established in the testing facility of the manufacturer, where the flow rate through the valve was measured under *constant* pressure drop conditions. Therefore, when such valves are installed, their gain (characteristics) shift, as will be discussed in Section 6.7. One way to correct for such distortion is to obtain "near-linear" installed characteristics by installing a valve with "ideal inherent" equal-percentage characteristics.

As will be discussed in connection with positioners (Section 6.2), it is also possible to use a cam in the positioner to modify the installed characteristic of the valve, but this can dramatically both change the loop gain of the positioner and limit its dynamic response. Therefore, it is preferred to change the inherent characteristic of the valve trim than to install cams in the valve positioner.

The installed characteristic of the valve can also be modified by characterizing the control signal going to the positioner. This characterization occurs outside the positioner feedback loop, and therefore it has the advantage of not changing the loop gain of the positioner. This method also has its dynamic limitations. For example, a 1% change in the controller output signal may be electronically narrowed to a change in the valve signal of only 0.1% (in the flat regions of the valve characteristic), but such change is too small for some valves to respond to at all.

Therefore, the best solution to obtaining a constant loop gain is to select the inherent characteristics of the valve trim to compensate for the nonlinearity of the process and, thereby, arrive at an installed flow characteristics, which is nearly linear over the operating range of the valve.

Selecting the Valve Characteristics Different engineers have approached the problems caused by process nonlinearity (drop in process gain) in different ways. One approach, that of the old school, was to oversize the pump so that the ratio between the minimum and maximum energy loss in Figure 6.1c will not be large, and therefore the gain of the process will not change much with load. This approach works, but it wastes pumping energy.

Different engineers began to develop different rules of thumb to be used in selecting valve characteristics for the various types of control loops. These recommendations vary in complexity. Shinskey, for example, recommends equal percentage for temperature control and the use of linear valves for all flow, level, and pressure control applications (except vapor pressure, for which he recommends equal percentage). According to Driskell, one can avoid a detailed dynamic analysis by just considering the ratio of the maximum and minimum valve pressure drops ($\Delta p_{\max}/\Delta p_{\min}$) and follow the rule of thumbs listed for the most common applications in Table 6.1f.

Lytle's recommendations are summarized in Table 6.1g; they are more involved, as they take more variables into account.

TABLE 6.1f

Valve Characteristics Selection Guide

Service	*Valve ($\Delta p_{max}/\Delta p_{min}$) Under 2:1*	*Valve ($\Delta p_{max}/\Delta p_{min}$) Over 2:1 but Under 5:1*
Orifice-type flow	Quick-opening	Linear
Flow	Linear	Equal %
Level	Linear	Equal %
Gas pressure	Linear	Equal %
Liquid pressure	Equal %	Equal %

TABLE 6.1g

Recommendations on Selecting Control Valve Characteristics for Flow, Level, and Pressure Control Loops‡

LIQUID LEVEL SYSTEMS

Control Valve Pressure Drop	*Best Inherent Characteristic*
Constant Δ*P*	Linear
Decreasing Δ*P* with increasing load, Δ*P* at maximum load > 20% of minimum load Δ*P*	Linear†
Decreasing Δ*P* with increasing load, Δ*P* at maximum load < 20% of minimum load Δ*P*	Equal-percentage
Increasing Δ*P* with increasing load, Δ*P* at maximum load < 200% of minimum load Δ*P*	Linear
Increasing Δ*P* with increasing load, Δ*P* at maximum load > 200% of minimum load Δ*P*	Quick-opening

PRESSURE CONTROL SYSTEMS

Application	*Best Inherent Characteristic*
Liquid process	Equal-percentage†
Gas process, small volume, less than 10 ft of pipe between control valve and load valve	Equal-percentage
Gas process, large volume (process has a receiver, distribution system, or transmission line exceeding 100 ft of nominal pipe volume), decreasing Δ*P* with increasing load, Δ*P* at maximum load > 20% of minimum load Δ*P*	Linear†
Gas process, large volume, decreasing Δ*P* with increasing load Δ*P* at maximum load < 20% of minimum load Δ*P*	Equal-percentage

FLOW CONTROL PROCESSES

		Best Inherent Characteristic	
Flow Measurement Signal to Controller	*Location of Control Valve Relation to Measuring Element*	*Wide Range of Flow Setpoint*	*Small Range of Flow but Large ΔP Change at Valve with Increasing Load*
Proportional to *Q*	In series	Linear	Equal-percentage†
Proportional to *Q*	In bypass*	Linear	Equal-percentage
Proportional to Q^2 (orifice)	In series	Linear†	Equal-percentage
Proportional to Q^2 (orifice)	In bypass*	Equal-percentage	Equal-percentage

* When control valve closes, flow rate increases in measuring element.

† Most common.

‡ From Reference 1.

Process Nonlinearity Yet another approach in overcoming the process nonlinearity caused by the variation in valve pressure drop is to modify the controller output signal to eliminate that nonlinearity (discussed in Section 6.7) or to replace the control valve with a complete flow control loop (Figure 6.1h). By selecting a linear flow transmitter, the characteristics of the slave loop will also be linear, and therefore its gain will be 1.0.

This approach works fine, but it also increases the system cost. In addition, while it eliminates a nonlinearity, it introduces a slave loop, which on fast processes can degrade the quality of control quality. By using intelligent control valves (Section 6.12), one can obtain flexibility in implementing one or the other approach.

Process nonlinearity also exists for reasons other than variations in the available valve pressure drop. The gain of heat-transfer processes, for example, always drops as the load increases, because the heat-transfer surface is constant, and therefore the heat transfer is more efficient when small amounts of heat need to be transferred. Consequently, in order to keep the gain product of the control loop constant, it is necessary to compensate the dropping process gain with an increasing valve

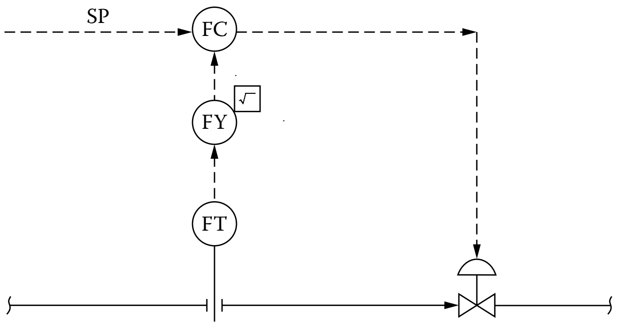


FIG. 6.1h

Valves can be linearized by replacing them with a complete slave control loop.

gain. For this reason, all temperature control valves are always equal percentage.

Composition processes can also be nonlinear, but their nonlinearity is usually more complex, such as the titration curve of a pH control system. In these situations, the likely solution is to use a linear valve (constant gain) and a nonlinear controller (Figure 2.19d), one whose gain varies as the mirror image of the process gain.

Valve Rangeability

Control valve rangeability is also discussed in more detail in Section 6.7. Here, it should suffice to state that the required rangeability should be calculated as the ratio of the C_v (K_v) required at maximum flow (and minimum pressure drop) and the C_v (K_v) required at minimum flow (and maximum pressure drop). The decision on whether a particular control valve is capable of providing the required rangeability should be evaluated on the basis of a plot of valve gain vs. valve C_v. If the actual valve gain is within 25% of the theoretical valve gain between the minimum and maximum C_v, the rangeability is acceptable. As will be discussed in Section 6.7, the rangeability definitions used by manufacturers are usually not based on valve gain.

One way to increase the rangeability is to have the controller operate more than one control valve. As will be discussed below, such multiple valves can be split-ranged, sequenced, or operated in a floating mode.

Control Valve Sequencing When the rangeability requirements of the process exceed the capabilities of a single valve, control valve sequencing loops must be designed that will keep the loop gain constant while switching valves. This requires careful thought.

Assuming that the task is to sequence two linear valves, with sizes of 1 and 3 in. (25 and 75 mm) having C_vs (K_vs) of 10 (8.62) and 100 (86.2), respectively, Shinskey's recommendation is the following: If the large valve were to operate from 9 to 15 PSIG (0.6 to 1.0 bar) and the small one from 3 to 9 PSIG (0.2 to 0.6 bar), the loop gain would change by 10 when passing through 9 PSIG (0.6 bar).

The only way to keep the loop gain constant in this example would be to operate the small valve from 0 to 10% and the large valve from 10 to 100% of controller output. This would result in a 3–4.2 PSIG (0.2–0.28 bar) range for the 1 in. (25 mm) and a 4.2–15 PSIG (0.28–1.0 bar) range for the 3 in. (75 mm) valve. Therefore, if more than one valve is required to increase rangeability, in most cases equal-percentage valve characteristics are needed.

Sequencing Equal-Percentage Valves The sequencing of equal-percentage valves is done as follows: If the small valve had a C_v of 10 and a rangeability of 50:1, its minimum C_v would be 10/50 = 0.2. A line drawn on semilogarithmic coordinates connecting C_v (K_v) 100 (86.2) and 0.2 (0.172) appears in Figure 6.1i. Observe that the C_v of 10 of the small valve falls slightly above the mid-scale of the controller output (to about 65%), providing a much more favorable span for the calibration of the positioner.

In order to have the two valves act as one without disturbing the smooth equal-percentage characteristics at the points of switching, only one valve must be open at any one time. Therefore, the large valve must be prevented from operating at low flows, because in its nearly closed position its characteristics are not equal-percentage. For these reasons, only one valve must be open at any one time.

In the scheme shown in Figure 6.1i, the small valve alone is manipulated until the controller output reaches the value corresponding to its full opening. At this point the pressure switch energizes both three-way solenoid valves, venting the small valve and opening the large to the same flow that the small had been delivering. Switching takes place in 1 sec or less, adequate for all but the fastest control loops.

When the controller output falls to the point of minimum flow from the larger valve (35%), the solenoids return to their original position. Thus, the switch has a differential gap adjusted to equal the overlap between valve positioners (30%). The range of the positioner for the large valve is found by locating its minimum C_v in Figure 6.1i. A rangeability of 50 would give a minimum C_v of 2.

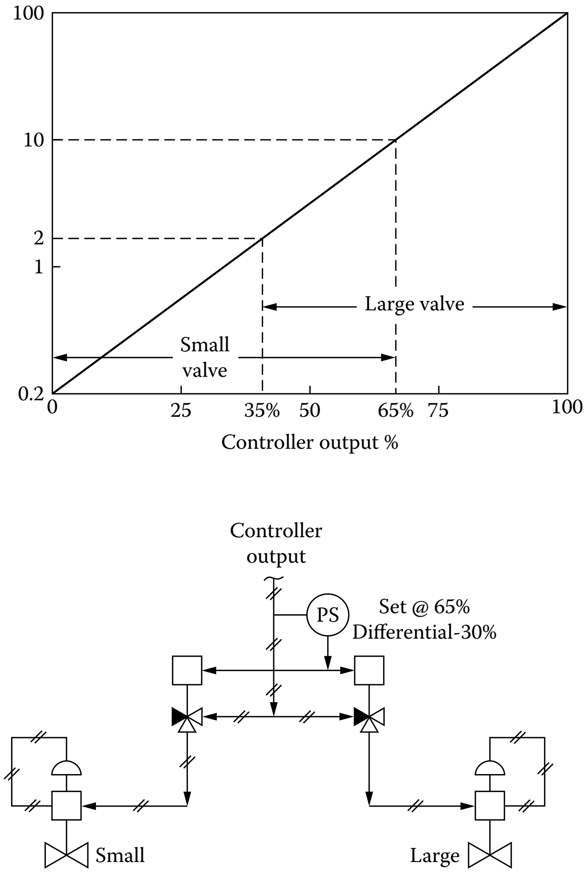
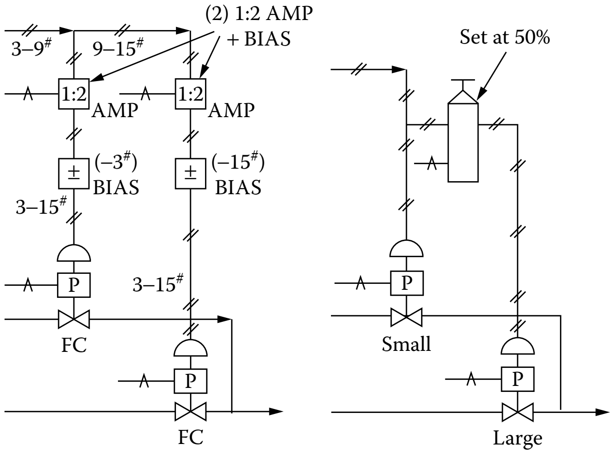
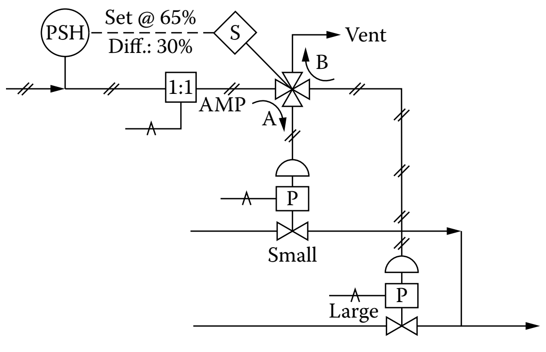


FIG. 6.1i
Sequencing two equal-percentage valves, while minimizing the upset caused by switching valves.



Split ranging
(Gain plus BIAS)

"Floating" control of
large valve



Small/Large valve select

FIG. 6.1j
Alternate methods of obtaining high turndown through the use of multiple control valves.

This same approach can be used to sequence three or more valves. If linear characteristics are required, one should insert a 10:1 multiplier relay in the controller signal to the small valve, so that a 0–10% controller output will result in a 0–100% signal to the small valve.

Split-Ranging or Floating Some process control engineers feel that the switching scheme in Figure 6.1i is too abrupt or could cause a maintenance problem. For these reasons, they prefer the methods of valve sequencing illustrated in Figure 6.1j.

The split-ranging loop shown in the upper left of the figure contains both gain and bias relays to provide the added rangeability. Even better response can be provided by the use of a "floating" valve position controller, shown on the top right of Figure 6.1j. This controller slowly moves the larger valve so as to keep the smaller one near its 50% opening. This way, the small valve provides sensitivity and fast response to the loop within its capacity. The large valve is a sort of an automatic bypass, which sets the capacity and has a limited frequency response.

The "large/small valve selection" scheme shown at the bottom of Figure 6.1j differs from the one in Figure 6.1i only in that it uses only one four-way solenoid pilot instead of two three-way units. The purpose of the 1:1 amplifier is to eliminate the bounce when the solenoid switches.

As the purpose of valve sequencing is to increase the rangeability of the loop without upsetting its stability, the existence of two or more valves should not be noticeable by the controller. In other words, in a well-designed sequencing system, the controller would operate as if its final control element were a single valve, having the desired gain characteristics and a very wide rangeability.

In order to keep the gain characteristics of the valve pair correct, there should be no "bumps" when the larger valve is opened. This requires that only one valve be throttled at a time and the other be closed. From this perspective, the performance of the small/large valve selection scheme in Figure 6.1j is superior to that of the split-ranging or the floating methods.

ACTUATOR SELECTION

Sections 6.3 and 6.4 discuss the applications and relative advantages of the different pneumatic, electric, digital, and hydraulic actuator designs.

The popularity of the spring-and-diaphragm actuator is due to its low cost, its relatively high thrust at low air supply pressure, and its availability with "fail-safe" springs. By trapping the pressure in the diaphragm case, it can also be locked in its last position. It is available in various designs: springless, double diaphragm (for higher pressures), rolling diaphragm (for longer strokes), and tandem, which provides more thrust. One of the limitations of this design is the lack of actuator "stiffness" (resistance to rapidly varying hydraulic forces caused, for example, by flashing). For such applications, hydraulic or electromechanical (motor gear) actuators are preferred, although a stiffer spring (6–30 PSIG, which corresponds to 0.41 to 2.59 bars) in a spring-and-diaphragm unit is sometimes sufficient to correct the problem.

Piston Actuators

Linear piston actuators provide longer strokes and can operate at higher air pressures than the spring-and-diaphragm actuators. When used to operate rotary valves, the linear piston or spring-and-diaphragm actuator does not provide a constant ratio of rotation per unit change in air signal pressure. Therefore, the use of positioners is always advisable.

Rotary piston actuators operate at higher air pressures and can provide higher torque, suitable for throttling large ball or butterfly valves. The double-acting version of this actuator does not have a positive failure position, but such a position can be added by extending the piston case and inserting a helical spring.

For higher torque (over 1000 ft lb_f, which corresponds to 1356 Nm), heavy-duty transfer linkages are required (Scotch yoke or rack and pinion); such units cannot be easily disassembled and maintained in the field. These actuators

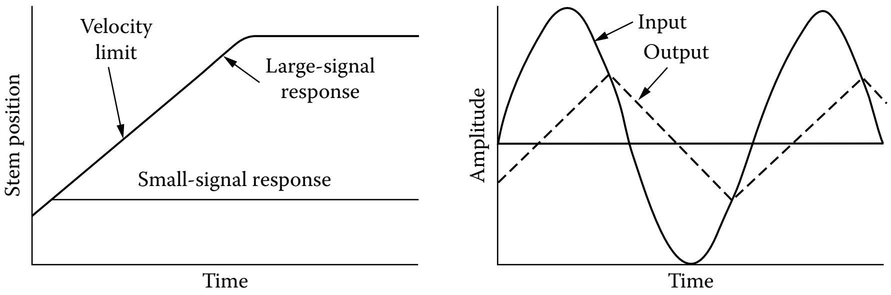


FIG. 6.1k
Response of velocity-limited actuators to a step change (left) and to a high-amplitude sine wave (right) in the control signal, according to Shinskey.

also require positioners, because the relationship between air signal change and resulting rotation is not linear. The use of positioners on fast loops can deteriorate the loop's performance, as it will need to be detuned.

Actuator Speeds of Response

In the family of pneumatic actuators, the spring-and-diaphragm actuators are the slowest (1–30 sec/stroke), spring-returned pistons with supply/exhaust at both ends of the piston are the fastest. Dual pistons can stroke valves in 0.5 sec.

In the family of electric actuators, the slowest are the electromechanical motor-driven valves (5–300 sec/stroke); hydroelectric actuators can move at 0.25 in./sec (6 mm/sec) or faster if hydraulic accumulators are used. The fastest are the small, on/off solenoids, which can close in 8–12 ms, while throttling solenoids require about 1 sec to stroke.

Valve actuators can be velocity limited because they cannot move faster than their maximum design speed. This is true of both electric and pneumatic motors or actuators. In case of the latter, maximum speed is set by the maximum rate at which air can be supplied or vented. If the full stroking (100%) of a valve takes 4 sec, then its velocity limit is 25% per second.

Valve signal changes usually occur in small steps, and therefore the velocity limit does not represent a serious limitation because, for example, the time required to respond to a 5% change is only 0.2 sec. This is fast enough for most loops. Figure 6.1k illustrates the response of velocity-limited actuators to various types of control signals.

Actuator speeds can be increased by enlarging the air flow ports and by installing booster relays. On on/off valves, the addition of a quick-dump valve (Figure 6.2o) will dramatically increase the venting rate. The dynamic performance of the actuator can also be affected by modifying the tare volume, pressure range, or dead band. In order to reduce the dead band, one usually needs to modify the piston seals, linkages, or rack and pinion connections.

Most valve actuators display some dead band or hysteresis band due to packing friction (Figure 6.1l). This can cause instability if the change in the control signal is small enough to fall within the hysteresis band width.

Actuator Power

The actuator is sized on the basis of the power or thrust required to overcome the unbalanced forces in the valve body and the seating force, and on the basis of the "stiffness" necessary for stability. In pneumatic actuators, the thrust is a function of piston or diaphragm area times air pressure. While the control signal is usually 3–15 PSIG (0.2 to 1.0 bar), the actuating pressure can be as high as the air supply pressure, if positioners or amplifier relays are installed. If their costs can be justified, electrohydraulic actuators will give the highest power and speed of response.

Valve Failure Position

In pneumatic actuators, the fail-safe action can be provided by a spring or from an air reservoir, but the latter represents added expense, complexity, and space requirements.

Electric actuators are usually more expensive, except for such designs as the spring-loaded or modulating solenoid valves discussed in Section 6.3. Electric actuators are therefore most often used where air is not available, where the thrust required is less than 1000 lb_f (4448 N), where it is acceptable to have the valve fail in its last position, and where slow response is not a drawback.

Like most generalizations, however, these are not completely true: Electromechanical motor and hydraulic actuators are available with high thrusts and can be provided with positive failure. Hydraulic actuators are also available with

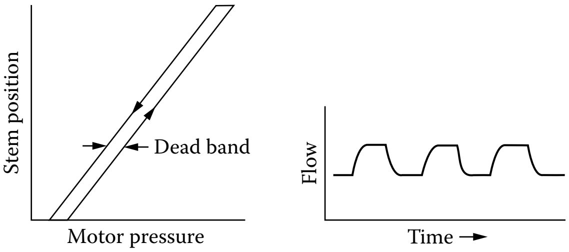


FIG. 6.1l
Dead band in the valve (left) can result in limit cycling (right) when the loop is closed.

fast response speeds. To gain an in-depth understanding of the capabilities of actuators, refer to Sections 6.3 and 6.4.

It is the responsibility of the process control engineer to specify the valve failure position. It is the general practice to fail energy supply valves (steam, hot oil, and so on) closed and energy-removing valves (cold or chilled water) open. The flow sheet abbreviations can be FC (fail closed), FI (fail indetermined), FL (fail in last position), and FO (fail open). Spring-loaded actuators are the most convenient means of providing FC or FO action, while two-directional air or electric motors will naturally tend to fail in their last positions.

In addition to anticipating the consequences of actuator power failure, one should also consider the results of other component failures, such as the spring, diaphragm, piston, and so on. When such failures occur, the ultimate valve position will *not* be a function of the actuator design, but of the process fluid forces acting upon the valve itself. The choices are FTO (flow to open), FTC (flow to close), or FB (friction bound—tends to stay in last position). FTO action is available with globe valves. FTC action can be obtained from butterfly, globe, and conventional ball valves. Rotary plug, floating ball, and segmented ball valves tend to be friction bound, with the flow direction possibly affecting the torque required to open the valve.

POSITIONERS

The positioner is a high-gain plain proportional controller that measures the valve stem position (to within 0.1 mm), compares that measurement to its set point (the controller output signal), and, if there is a difference, corrects the error. The open-loop gain of positioners ranges from 10 to 200 (proportional band of 10–0.5%), and their periods of oscillation range between 0.3 and 10 sec (frequency response of 3–0.1 Hz). In other words, the positioner is a very sensitively tuned, proportional-only controller.

Positioners that are electronically and digitally controlled, or are intelligent and are capable of self-diagnostics, communication on fieldbuses, and other advanced features, are not discussed here, because they are described in detail in Section 6.12.

When to Use Positioners

The main purpose of having a positioner is to guarantee that the valve does, in fact, move to the position where the controller wants it to be. The addition of a positioner can correct for many variations, including changes in packing friction due to dirt, corrosion, or lack of lubrication; variations in the dynamic forces of the process; sloppy linkages (dead band); or nonlinearities in the valve actuator. The dead band of a valve/actuator combination can be as much as 5%; when a positioner is added, it can be reduced to less than 0.5%. It is the job of the positioner to protect the controlled variable from being upset by any of the above variations.

In addition, the positioner can also allow for split-ranging the controller signal between more than one valve, can increase the actuator speed or thrust by increasing the pressure or volume or the actuator air signal, and can modify the valve characteristics by cams or electronic function generators.

While the above positioner capabilities can be convenient, they can also be obtained without the use of positioners. For example, split-ranging can also be done by the use of split-ranged valve springs or by multiple/biasing relays in the air signal line to the valve. Similarly, increasing the speed/thrust of the valve can be achieved by booster relays, and changes to the control valve characteristics can be obtained, not only by replacing the plug, but also by pneumatic or electronic characterizing of the controller signal. Therefore, these reasons do not necessitate the use of positioners.

Actuators without springs always require positioners. When a valve is in remote manual (open loop) operation, it will always benefit from the addition of a positioner, because a positioner will reduce the valve's hysteresis and dead band while increasing its response. When the valve is under automatic (closed loop) control, the positioner will be helpful in most slow loops, which control analytical properties, temperature, liquid level, blending, slow flow, and large volume gas flow.

A controlled process can be considered "slow" if its period of oscillation is three times the period at which the positioned valve oscillates. In such installations as B in Figure 6.2b, the addition of a positioner increases the open-loop gain and, therefore, the loop response. As a consequence, the tuning of the controller could also be improved by increasing the gain (making the proportional band narrower) and adding more repeats per minute in the integral setting.

When Not to Use Positioners

In the case of fast loops, positioners are likely to degrade loop response, contribute to proportional offsets, and cause limit cycling (fast flow, liquid pressure, small volume gas pressure).

The positioner in effect is the cascade slave of the loop controller. In order for a cascade slave to be effective, it must be faster than the speed at which its set point, the master output signal, can change. The rules of thumb used in this respect suggest that the time constant of the slave should be ten times shorter (open-loop gain ten times higher) than that of the master and the period of oscillation of the slave should be three times shorter (frequency response three times higher) than that of the primary. The criteria for positioners need not be this stringent, but still, it is recommended not to use positioners if the positioned valve is slower than the process variable it is assigned to control.

A controlled process can be considered "fast" if its period of oscillation is less than three times that of the positioned valve. In such situations, the positioned valve is one of the slowest components in the loop and, therefore, slows down the load (by limiting the open-loop gain of the loop and lengthening the period of oscillation). Part A in Figure 6.2b illustrates such a situation, where the loop can be tuned more tightly (higher gain, more repeats/minute) and, therefore, responds better without a positioner. It might also be noted that after a new steady state is reached, the positioned installation gives more noisy control because of the hunting and limit cycling of the positioner, which cannot keep up with the process.

Some will argue that all loops can be controlled using positioned valves if they are sufficiently detuned. This is true, but "detuning" means that the controller is made less effective (the amount of proportional and integral correction is reduced), which is undesirable. A 0.2 gain (500% proportional) setting on a level controller means that the tank can be flooded or drained before the controller fully strokes the valve. Also, there are cases where the gain must be so low (the proportional band so wide) that it is outside the available PB setting range of the controller.

Positioners to Eliminate Dead Band

All valves and dampers will display some dead band because of friction in their packing, unless positioners are used. Whenever the direction of the control signal is reversed, the stem remains in its last position until the dead band is exceeded, as shown in Figure 6.1l.

This figure on the right shows that if a sine wave control signal is driving the valve actuator (motor), it produces a stem motion that is distorted and shifted in phase. This phase shift, when combined with the integrating characteristic of certain processes and with the reset action of a controller, causes the development of a limit cycle. According to Shinskey, widening the proportional band will not dampen the oscillation, but only make it slower.

The limit cycle will not appear if a proportional-only controller is used, and if the process has no integrating element. Processes that are prone to limit cycling in this way are liquid level, volume (as in digital blending), weight (not weight-rate), and gas pressure—all of which are related to the integrals of flow. Whenever one intends to control such a process with a proportional and integral (PI) controller, the use of positioners should be considered. In case of level control, one can accomplish the same goal by using a plain proportional controller and a booster or amplifier instead of a positioner.

Positioners in general will eliminate the limit cycle by closing a loop around the valve actuator. Positioners will also improve the performance of valves on slow processes, such as pH or temperature. On the other hand, dead band caused by stem friction should not be corrected by the use of positioners on fast loops, such as flow or "fast" pressure.

The positioner's function as a cascade slave, as was explained earlier, can cause oscillation and cycling on fast loops if the controller cannot be sufficiently detuned (minimum gain is not low enough). Similarly, negative force reactions on the plug require an increase in actuator stiffness and not the addition of a positioner. Actuator stiffness can be improved by increasing the operating air pressure or by using hydraulic actuators.

Split-Range Operation

The use of positioners for split-range applications is usually accepted regardless of the speed of the process. This is not entirely logical, because on fast loops the control performance can be degraded by the use of positioners. In such cases, some instrument engineers do discourage the use of positioners to implement split-ranging. Instead, they recommend gain-plus-bias relays so that the positioner (the less-accurate device) will operate over its full range (Figure 6.1j). This also eliminates the need for a special calibration.

One can also consider accomplishing the split-range operation through the use of different spring ranges in the valve actuators. In addition to the standard 3 to 15 PSIG (0.2 to 1 bar) range spring, valves can also be obtained with other spring ranges. These include 3–7 PSIG (0.2–0.5 bar), 4–8 PSIG (0.28–0.55 bar), 5–10 PSIG (0.34–0.68 bar), 7–11 PSIG (0.5–0.75 bar), 8–13 PSIG (0.55–0.9 bar), and 9–13 PSIG (0.62–0.9 bar).

Lastly, if split-range positioners are installed on fast processes, the resulting degradation of control quality can be limited by adding a restrictor or an inverse derivative relay in the control signal to it and, thereby, artificially making the controller appear to be slower than it really is. This technique is not highly recommended (except to reduce wear and tear on the valve in noisy loops), because restrictors are prone to plugging or maladjustment and because they both degrade the loop performance.

Accessories

If the need is to increase the speed or the thrust of the actuator, it is sufficient to install an air volume booster or a pressure amplifier relay, instead of using a positioner. Boosters will give better performance than positioners on fast processes, such as flow, liquid pressure, or small-volume gas pressure control, and they will not be detrimental (nor will offer advantages) if used on slow processes.

If the reason for adding a positioner is to alter or modify the control valve characteristics, this is not a valid justification on fast processes, because this aim can be satisfied by the use of dividing or multiplying relays in the controller output, which will not degrade the quality of control (see Section 6.7).

PROCESS APPLICATION CONSIDERATIONS

In selecting control valves, the properties of the process fluid must be fully considered. The process data should be carefully and accurately determined because even small variations in temperature or pressure can cause flashing or cavitation. Considerations include such obvious variables as pressure, temperature, viscosity, slurry, or corrosive nature, or the less obvious factors of flashing, cavitation, erosion, leakage, sterilization, and low flow rates. These are discussed in the paragraphs that follow below.

Pressure Considerations

The available design pressures for each valve type are listed in the feature summaries in the front of Sections 6.16–6.24.

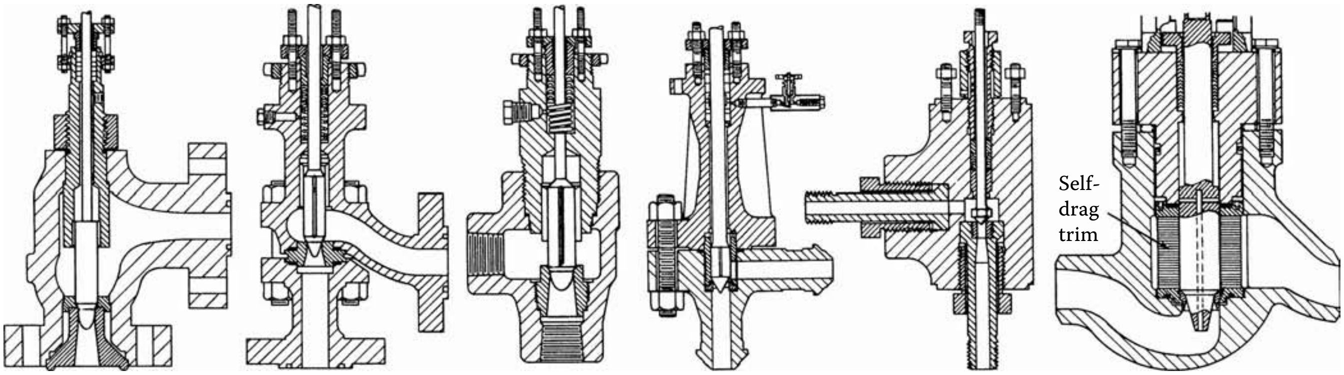


FIG. 6.1m
High-pressure valve designs.

In selecting the control valve for a particular application, one should pay particular attention to high pressure, high differential pressure, and vacuum services. These will be discussed in the paragraphs below.

High-Pressure Services When designing valves for high-pressure services, the following features are of particular importance:

1. Increased physical strength
2. Selection of erosion-resistant material
3. Use of special seals

Valve bodies can usually withstand higher pressures than can the piping. Valve bodies for high-pressure services are usually forged to provide homogeneous materials free of voids and with good mechanical properties. The loads and stresses on the valve stem are also high. For this reason, higher strength materials are used with increased stem diameters. As shown in Figure 6.1m, the stems are usually kept short and are well-guided.

High-pressure services will also increase the probability of noise, vibration, and cavitation, which will be discussed in later paragraphs.

High Differential Pressure With high flow rates and high pressure drops, a large amount of energy is dissipated in turbulence. A fraction of this energy is radiated as noise (see Section 6.14). For most (but not all) gases and conditions, one result of the high pressure drop can be a very low outlet temperature. It is not always proper to use the gas laws to predict these temperatures, and instead, actual thermodynamic properties should be used. The very low temperature will cause some valve materials to become brittle, and careful selection of alloys and other materials is always required.

Depending on the gas involved and other materials in the flowing fluid, hydrates or other solids may form in the valve. Liquid droplets may develop and cause erosion. It is necessary to investigate for any peculiarity of the flowing fluid. A high-velocity jet leaving the valve can erode downstream piping. Very high forces are developed on the valve body and internal parts and can cause valve instability. With the change in magnitude and, often, direction of the fluid, substantial reaction forces are developed. Serious damage may result if the valve and piping are not properly restrained.

High operating pressure frequently involves high pressure drops. This usually means erosion, abrasion, or cavitation at the trim. These will be discussed in detail in the coming paragraphs, but it should also be mentioned here that cavitation and erosion resistance are usually not properties of the same metal. Materials resistant to erosion and abrasion include 440C stainless steel, flame-sprayed aluminum oxide coatings (Al_2O_3), and tungsten carbide.

On the stem, where the unit pressure between it and the packing is high, it is usually sufficient to chrome-plate the stem surface to prevent galling. Special "self-energizing" seals are used with higher pressure valves (above 10,000 PSIG, or 69 MPa, service) so that the seal becomes tighter as pressure rises. Popular body seal designs for such service include the delta ring closure and the Bingham closure (Figure 6.1m).

As discussed in more detail in Section 6.19, the self-energizing seals are used in connecting the high-pressure valves into the pipeline. These designs depend on the elastic or plastic deformation of the seal ring at high pressures for self-energization.

Special packing designs and materials are also required in high-pressure service, because conventional packing would be extruded through the clearances. To prevent this, the clearance between stem and packing box bore is minimized, and extrusion-resistant material, such as glass-impregnated Teflon, is used for packing.

Some of the likely causes of valve failure in high pressure drop services include:

1. Elastomer elements in Saunders or pinch valves (particularly if they fail to open) can be ruptured.
2. The stem thrust can be excessive for globe valves. If globe valves are flow to close, high Δp can damage the seat or prevent the actuator from opening the valve. If they are flow to open, they might open against the actuator.

3. High Δp can exceed the capabilities of plug or floating ball valves, and it can bend the shafts of butterfly valves, damage the bearings of trunion-type ball valves, or damage the seat of floating ball valves.
4. Pressure cycling generated by positive-displacement pumps can cause bolt fatigue if the number of cycles is excessive.

Vacuum Service Low pressures can prevent some pressure-energized seals from properly operating, or they can cause leakage. In some processes, the in-leakage from the atmosphere results in overloading the vacuum source; in others, it represents a contamination that cannot be tolerated. Potential leakage sources include all gasketed areas and, to an even greater extent, the locations where packing boxes are used to isolate the process from the surroundings.

For vacuum service, valves that do not depend on stuffing boxes to seal the valve stem generally give superior performance. Such designs include Saunders valves and pinch valves. These designs, unfortunately, are limited in their application by their susceptibility to corrosion and their temperature and control characteristics. Their applicability to vacuum service is further limited by their design.

The jacketed pinch valve versions, for example, require a vacuum source on the jacket side for proper operation, and the mechanically operated pinch and Saunders designs are limited in their capability to open the larger-size units against high vacuum on the process side. The vacuum process tends to keep the valve closed, and this can result in the diaphragm's breaking off the stem and rendering the valve inoperative.

For services requiring high temperatures and corrosion-resistant materials, in addition to good flow characteristics and vacuum compatibility, conventional globe valves can be considered, with special attention given to the type of packing and seal used.

One approach to consider is the use of double packing, as shown in Figure 6.1n. The space between the two sets of packing is evacuated so that air leakage across the upper packing is eliminated. The vacuum pressures on the two sides of the lower packing are approximately equal, and therefore there is no pressure differential to cause leakage across it. Usually, the space between the two packings is exposed to a slightly higher vacuum than the process, so that no in-leakage is possible.

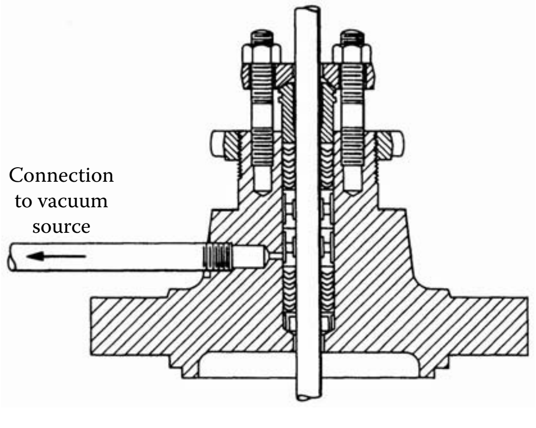


FIG. 6.1n
Double packing is used to seal the valve stem if the process is under vacuum.

Double packing provides reasonable protection against in-leakage under vacuum, but it does not relieve the problems associated with corrosion and high temperatures. When all three conditions exist (vacuum, corrosive flow, and high temperature), the use of bellows seals (discussed in Section 6.19) can be considered. The bellows are usually made of 316 stainless steel and are tested by mass spectrometers for leakage. They not only prevent air infiltration but also can protect some parts of the bonnet and top-works from high temperature and corrosion. Like all metallic bellows, these too have a finite life, and therefore it is recommended that a secondary stuffing box and a safety chamber be added after the bellows seal. A pressure gauge or switch can be connected to this chamber between the bellows and the packing to indicate or warn when the bellows seal begins to leak and replacement is necessary.

Considerations similar to those noted for high-vacuum service would also apply when the process fluid is toxic, explosive, or flammable.

High-Temperature Service

The temperature limitations of each valve design are listed in their feature summaries in Sections 6.16–6.24.

All process conditions involving operating temperatures in excess of 450°F (232°C) are considered high temperature. The maximum temperatures at which control valves have been successfully installed are up to 2500°F (1371°C).

High operating temperatures necessitate the review of at least three aspects of valve design:

1. Temperature limitations of metallic parts
2. Packing temperature limitations
3. Use of jacketed valves

Metallic Parts High temperatures can cause galling, can affect clearances, and can soften hardened trims. Temperature cycling can cause thermal ratcheting and stress, resulting in body or bolting rupture if the rate or frequency of temperature cycles is high.

The high operating temperatures are considered in selecting materials for both the valve body and trim. For the body, it is suggested that bronze and iron be limited to services under 400°F (204°C), steel to operation below 850°F (454°C), and the various grades of stainless steel, Monel, nickel, or Hastelloy alloys to temperatures up to 1200°F (649°C).

For the valve trim, 316 stainless steel is the most popular material, and it can be used up to 750°F (399°C). For higher temperatures, the following trim materials can be considered: 17–4 pH stainless steel (up to 900°F, or 482°C), tungsten carbide (up to 1200°F, or 649°C), and Stellite or aluminum oxide (up to 1800°F, or 982°C).

At high temperatures, the guide bushings and guideposts tend to wear excessively, and this can be offset by the selection of proper materials. Up to 600°F (316°C), 316 stainless steel guideposts in combination with 17–4 pH stainless steel guide bushings give acceptable performance. If the guideposts are surfaced with Stellite, the above combination can be extended up to 750°F (399°C) service. At operation over 750°F (399°C), both the posts and the bushings require Stellite.

Packing Designs The ideal packing provides a tight seal while contributing little friction resistance to stem movement. With TFE packing, which is industry standard, the required stem finish is between 6 and 8 μ in. RMS (0.15 and 0.2 μm).

A common packing design might consist of Teflon V-rings, which are discussed in more detail in Section 6.19. Double packing with leak-off connection in between can be used on toxic or vacuum service (Figure 6.1n). On toxic services, an added seal can be provided by the injection of high-viscosity silicone or plastic packing, but this is rarely done.

Solid rings or ribbons of pure graphite (Graphoil), while more expensive than Teflon, are also popular because they are suited for higher temperatures. On the other hand, they require more loading to energize the packing than does Teflon, and the resulting friction can cause stem lockup. On less demanding services, O-rings are also used, but not frequently because under pressure these elastomers will absorb gases, which can destroy the O-ring when depressurized rapidly.

Metallic bellows-type seals are seldom used because of their pressure limitations and unpredictable lives. On toxic services, they should be provided with automatically monitored guard packing for security.

The bonnets are usually flanged and are extended on hot or cold services so as to bring the operating temperature of the packing closer to the ambient. Screwed bonnets are not recommended for severe duty, and welded bonnets are not used at all, except as an extreme precaution on hazardous services. The sliding stems can sometimes drag atmospheric contaminants or process materials into the packing, but this can be overcome by close tolerance guide bushings or wiper rings. Packing contamination is less likely with rotary valves. In case of LPG, the packing should be isolated from outboard roller bearings and the intervening space vented to protect the lubricant.

Packing Limitations Packing and bonnet designs in general were discussed in the previous paragraph and will also be covered in Section 6.19. Here only their suitability for high-temperature service is reviewed.

The packing temperature limitation for most nonmetallic materials is in the range of 400–550°F (204–288°C), the maximum temperature for metallic packing is around 900°F (482°C), and Teflon should not be exposed to temperatures above 450°F (232°C). Pure graphite (Graphoil) can be used from –400 to 750°F (–240 to 399°C) in oxidizing service and up to 1200°F (649°C) in nonoxidizing service, with an ultimate potential of 3000°F (1649°C).

Bonnets can be screwed, welded, or flanged. Screwed bonnets are not recommended for high-temperature service. Finned bonnet extensions were used in the past on high-temperature services, when packing material capabilities were more limited. These finned designs were not effective, and therefore with the introduction of Graphoil, their use was largely discontinued on rotary valves. For sliding stem valves, Teflon V-rings within extension bonnets are frequently selected and used up to 850°F (454°C).

On high-temperature services, it can be effective to mount the bonnet below the valve. In liquid service, with the bonnet above the valve, the packing is exposed to the full process temperature due to the natural convection of heat in the bonnet cavity.

If the bonnet is mounted below the valve, no convection occurs, and the heat from the process fluid is transferred by conduction in the bonnet wall only. Therefore, by this method of mounting, the allowable process temperature can be substantially increased in some processes. This is not the case for all applications, because some processes do not generate effective condensate seals, and in other services the liquid-vapor interface line can cause metallurgy problems.

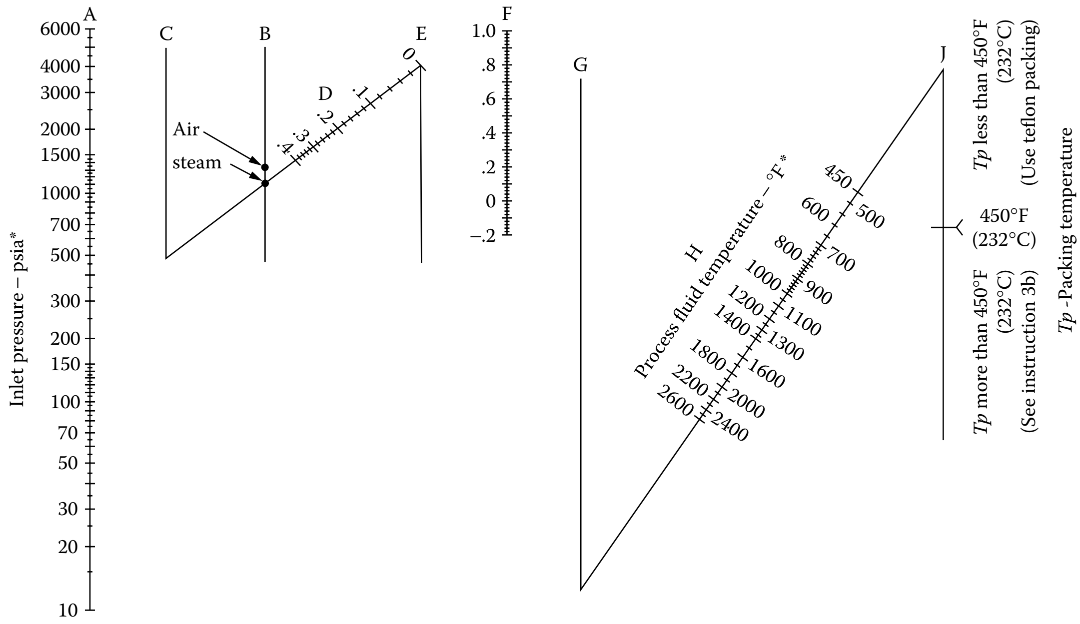
Figure 6.1o provides a method for determining packing temperature, in gas or vapor service, with the bonnet above the valve for one particular valve design. With vapor service, it is likely that vapors will initially condense on the wall of the bonnet, lowering the temperature to the saturation temperature of the process fluid, but in some cases the heat conducted by the metallic bonnet wall will be sufficient to prevent this condensation from occurring.

In short, the packing temperature will be at or above saturation temperature (T_s) in vapor service, but if the bonnet is mounted below the valve, the packing temperature is substantially reduced, due to the accumulated condensate. If the bonnet is below the valve, the relationship between process and packing temperature is not affected by the phase of the process fluid.

In case of ball or plug valves with double-sealing, it is important to vent the space between the seals to the line, so that damage will not be caused by thermal expansion.

Jacketed Valves A number of control valve designs are available with heat-transfer jackets. Others can be traced or jacketed by the user. Jacketed valves can be installed for either cooling or heating. When a cooling medium is circulated in the jackets, this is usually done to lower the operating temperature of the heat-sensitive working parts.

Such jacketing is particularly concentrated on the bonnet, so that the packing temperature is reduced relative to the process. For certain operations at very high temperatures, intermittent valve operation is recommended, such that when the valve is closed it is cooled by the jacket, and when it is opened, it is kept open only long enough to prevent temperature equalization between the valve and the process.



*See Section A.1 for SI units

Instructions

1. Determine constants from table at right corresponding to bonnet selection to locate points on scales "D" and "F."

2. Solve nomograph Key:

Line Up Straight Edge On	*Locate Intersection On*
A TO B	C
C TO D	E
E TO F	G
G TO H	J

3a. If Tp ≤ 450°F (232°C), use Teflon packing.

3b. If Tp > 450°F (232°C), either use high-temperature packing or select another bonnet with smaller "F" value and recheck packing temperature.

Bonnet Characteristics

Valve Size	*Bonnet Material*[*]	*Bonnet Factors*	*Standard Bonnet*	*Extension Bonnet*	*Radiation Fin Bonnet*
1"	CS	D	0.07	0.29	
		F	.83	.25	
	SS	D	.13	.33	
		F	.65	.15	
1½"	CS	D	.07	.24	0.25
		F	.83	.39	.35
	SS	D	.12	.30	.31
		F	.69	.22	.21
2"	CS	D	.11	.29	
		F	.72	.25	
	SS	D	.17	.33	
		F	.55	.15	
3"	CS	D	.10	.24	.31
		F	.74	.39	.20
	SS	D	.17	.31	.34
		F	.57	.21	.11
4"	CS	D	.07	.25	
		F	.83	.36	
	SS	D	.14	.31	
		F	.65	.20	
6"	CS	D	.04	.23	
		F	.90	.40	
	SS	D	.08	.30	
		F	.80	.22	
8"	CS	D	.06	.23	.33
		F	.86	.40	.14
	SS	D	.11	.30	.35
		F	.72	.22	.10

[*] *CS: carbon steel, SS: stainless steel.*

FIG. 6.1o

Nomograph for packing temperature determination on hot gas or vapor services. (From "Determination of Proper Bonnet and Packing for High-Temperature Processes," R.F. Lytle, Fisher Controls, Emerson Process Management.)

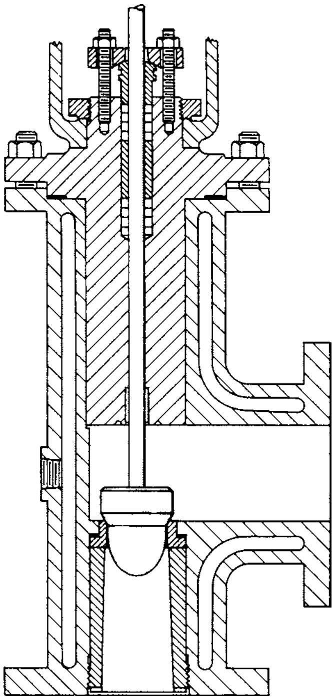


FIG. 6.1p
Jacketed control valve for high-temperature service.

Heating jackets with steam or hot oil circulation are used to prevent the formation of cold spots in the more stagnant areas of the valve or where the process fluid otherwise would be exposed to relatively large masses of cold metal. Figure 6.1p shows one of these valves, designed to prevent localized freezing or decomposition of the process fluid due to cold spots.

Many standard globe pattern valves can be fitted with a jacket to allow heating or cooling as required (Figure 6.1q). Normally, these jackets are for services requiring steam.

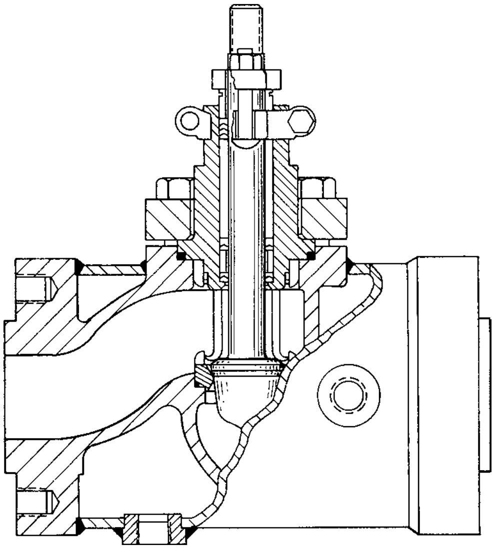


FIG. 6.1q
Steam-jacketed valve. (Courtesy of Flowserve Corp.)

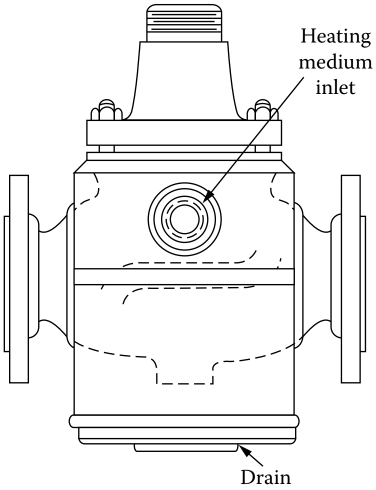


FIG. 6.1r
Special steam jacket for retrofit installation on valve.

Dowtherm or similar heating fluids prevent solidification or crystallization of certain fluids. Often, the manufacturer can provide these jackets, but where this is not available, there are firms that specialize in designing and installing such jackets on valves and other equipment.

These special jackets can be designed either to weld to the valve as a permanent fixture (Figure 6.1r) or as separate devices bolted or clamped to the valve body. In the latter case, it may be necessary to use a heat-transfer paste between jacket and valve body to give efficient transfer by eliminating the air gap.

Low-Temperature Service

Cryogenic service is usually defined as temperatures below –150°F (–101°C). Properties of some cryogenic fluids are listed in Table 6.1s. Valve materials for operation at temperatures down to –450°F (–268°C) include copper, brass, bronze, aluminum, 300 series stainless steel alloys, nickel, Monel, Durimet, and Hastelloy. The limitation on the various steels falls between 0 and –150°F (–17 and –101°C), with cast carbon steel representing 0°F (–17°C) and $3^1/_2$% nickel steel being applicable to –150°F (–101°C). Iron should not be used below 0°F (–17°C).

Conventional valve designs can be used for cryogenic service with the proper selection of construction materials and with an extension bonnet (as described in detail in Section 6.19) to protect the packing from becoming too cold. The extension bonnet is usually installed vertically so that the boiled-off vapors are trapped in the upper part of the extension, which provides additional heat insulation between the process and the packing.

If the valve is installed in a horizontal plane, a seal must be provided to prevent the cryogenic liquid from entering the extension cavity. When the valve and associated piping are installed in a large box filled with insulation (“cold box”), this requires an unusually long extension in order to keep the packing box in a warm area.

TABLE 6.1s
Properties of Cryogenic Fluids

	Methane	*Oxygen*	*Fluorine*	*Nitrogen*	*Hydrogen*	*Helium*
Boiling point (°K) *(°C)*	–259 (–162)	–297 (–183)	–307 (–188)	–320 (–196)	–423 (–253)	–452 (–269)
Critical temperature (°F) *(°C)*	–117 (–83)	–181 (–118)	–200 (–129)	–233 (–147)	–400 (–240)	–450 (–268)
Critical pressure (psia) *[bar(A)]*	673 46.1	737 50.5	808 55.3	492 33.7	188 12.9	33 2.26
Heat of vaporization at boiling point (BTU/lbm) *(J/kg)*	219 ($5.09 \cdot 10^5$)	92 ($2.14 \cdot 10^5$)	74 ($1.72 \cdot 10^5$)	85 ($1.98 \cdot 10^5$)	193 ($4.49 \cdot 10^5$)	9 ($2.09 \cdot 10^4$)
Density (lbm/ft^3) gas at ambient conditions *(kg/m^3)*	0.042 (0.673)	0.083 (1.33)	0.098 (1.57)	0.072 (1.153)	0.005 (0.080)	0.010 (0.16)
Vapor density at boiling point	0.111 (1.778)	0.296 (4.74)	—	0.288 (4.614)	0.084 (1.346)	1.06 (16.98)
Liquid density at boiling point	26.5 (424.5)	71.3 (1142)	94.2 (1509)	50.4 (807.4)	4.4 (70.5)	7.8 (125)

Cryogenic Valves A special design variation on the globe valve is the cryogenic valve. Section 6.19 shows a number of cryogenic service designs, including the Y-valve design. The most common design for this service is shown in Figure 6.1t.

This design is specific to cryogenic (down to –454°F, or –270°C) service and no other. Body configurations are straight through, as shown, or angle body. Because of the need for Charpy impact for the extremely cold service, the materials are limited to bronze and austenitic stainless steels such as 304, 316, and 316L. Normally, the valves are welded into the piping or soldered in some cases with bronze. Small valves, 1 in. (25 mm) and 2 in. (50 mm), can be socket weld or butt weld, but butt weld in larger sizes up to the 10 in. (250 mm) maximum are available.

Body ratings through ANSI 600 and flange ends, either integral or separable, are available depending upon manufacturer. Seat rings may be integral hard-faced with Stellite, screwed-in metal, or soft seat for tight shut-off. The metal shut-offs will be ANSI Class III or IV, depending upon manufacturer, and the soft seat using Teflon or Kel-F will provide Class VI.

Cold Box Valves Cold box valves are designed to have a low body mass for fast cool-down and reduced heat transfer. The long extended bonnet is provided with a plug stem seal to minimize liquid "refluxing" into the bonnet and packing area, thereby minimizing the heat loss due to conduction and convection. Actually, the small amount of liquified gas passing into the bonnet vaporizes and provides a vapor barrier between the liquified gas and the packing area. In addition, the pressure resulting from the vaporization of the liquid prevents additional liquid from passing into the bonnet area. Excess pressure vents back into the body. It is possible to fit these valves with vacuum jackets where the application requires this additional insulation.

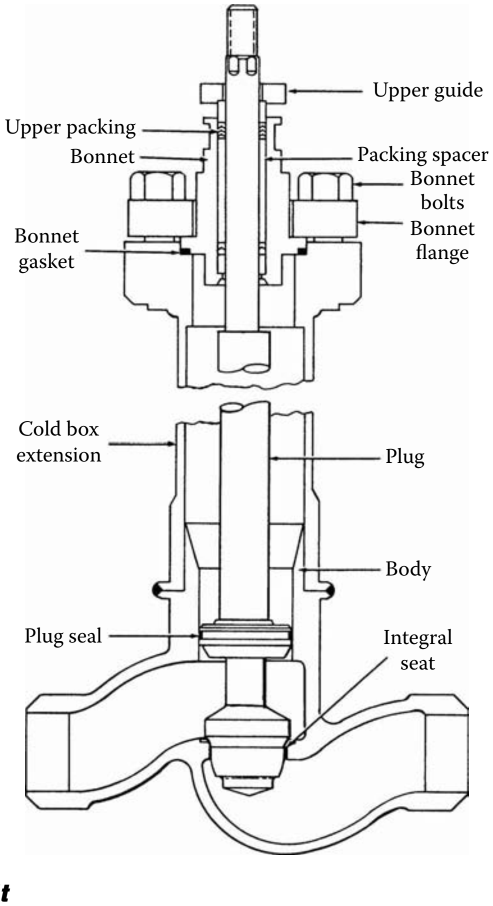


FIG. 6.1t
Cold box valve with weld ends and welded bonnet. (Courtesy of Flowserve Corp.)

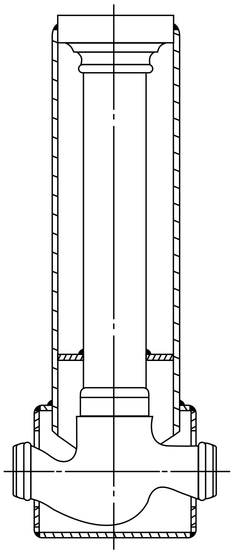


FIG. 6.1u
Vacuum jacketing of cryogenic valve.

Features that are desirable for cryogenic valves include small body mass, which ensures a small heat capacity and, therefore, a short cool-down period. In addition, the inner parts of the valve should be removable without removing the body from the pipeline, and if the valve is installed in a cold box, no leakage can occur inside this box because there are no gasketed parts.

The most effective method of preventing heat transfer from the environment into the process is by vacuum jacketing the valve and piping (Figure 6.1u). The potential leakage problems are eliminated by the fact that there are no gasketed areas inside the jacket.

For cryogenic services where tight shut-off is required, Kel-F has been found satisfactory as a soft seat material because cold can cause many other elastomer materials to harden, set, or shrink. If used as seals, this can cause leaking.

Cavitation and Erosion

The cavitation and erosion phenomena in connection with globe valves are discussed in detail in Section 6.19. The choking effect of cavitation and its influence on valve sizing is covered in Section 6.15. Sections 6.14 and 6.15 describe how the liquid pressure recovery factor (F_L) is related to the ratio between the valve pressure drop and the difference between the inlet and the vena contracta pressure.

As was shown in Section 6.15, the cavitation coefficient K_c is the ratio between the valve pressure drop at which cavitation starts and the difference between the inlet and the vapor pressure of the application. Section 6.15 also shows how the F_{LP} factor can be calculated if the valve is placed within reducers, and it also shows the valve pressure differential at which choking starts can be calculated.

The allowable maximum Δp before cavitation begins is $\Delta p = K_c\,(p_1 - p_v)$. As the F_L and K_c values of the different valve designs drop, the probability of cavitation increases. F_L and K_c values for fully open valves are also given in Section 6.15, and F_L values for throttled valves are given in Figure 6.1v.

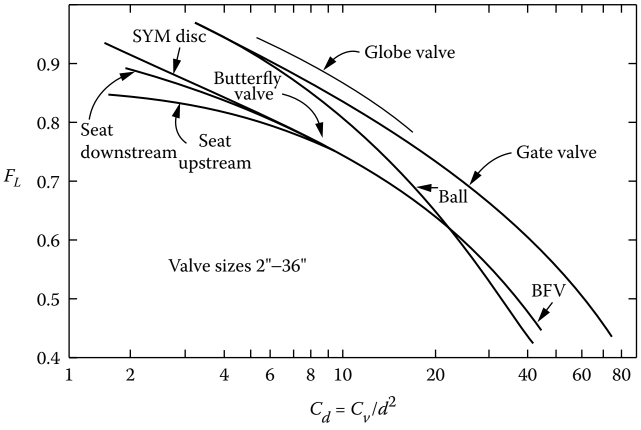


FIG. 6.1v
In this figure, the pressure recovery factor (F_L) of different valve designs is shown as a function of their discharge coefficients (C_d values, which in the metric system are defined as $\varepsilon = K_v / DN^2$) as these valves are throttled from their full open positions.[2]

The high flow velocity at the vena contracta of the valve is reached by obtaining its energy to accelerate from the pressure energy of the stream. This causes a localized pressure reduction that, if it drops below the fluid's vapor pressure, results in temporary vaporization. (Fluids form cavities when exposed to tensions equal to their vapor pressure.) Cavitation only occurs when the pressure in the vena contracta region drops below the vapor pressure of the flowing fluid. The vapor pressure is a function of fluid temperature and chemical structure.

Cavitation damage always occurs downstream of the vena contracta when pressure recovery in the valve causes the temporary voids to collapse. Destruction is due to the implosions that generate the extremely high-pressure shock waves in the substantially noncompressible stream. When these waves strike the solid metal surface of the valve or downstream piping, the damage gives a cinder-like appearance. Cavitation is usually coupled with vibration and a sound like rock fragments or gravel flowing through the valve.

Cavitation damage always occurs downstream of the vena contracta at the point where the temporarily formed voids implode. In case of flow-to-open valves, the destruction is almost always to the plug and seldom to the seat.

Methods to Eliminate Cavitation

Because no known material can remain indefinitely undamaged by severe cavitation, the only sure solution is to eliminate cavitation completely. Even mild cavitation over an extended time will attack the metal parts upon which the bubbles impinge. Hard materials survive longer, but they are not an economical solution except for services with mild intermittent cavitation. Cavitation damage also varies greatly with the type of liquid flowing.

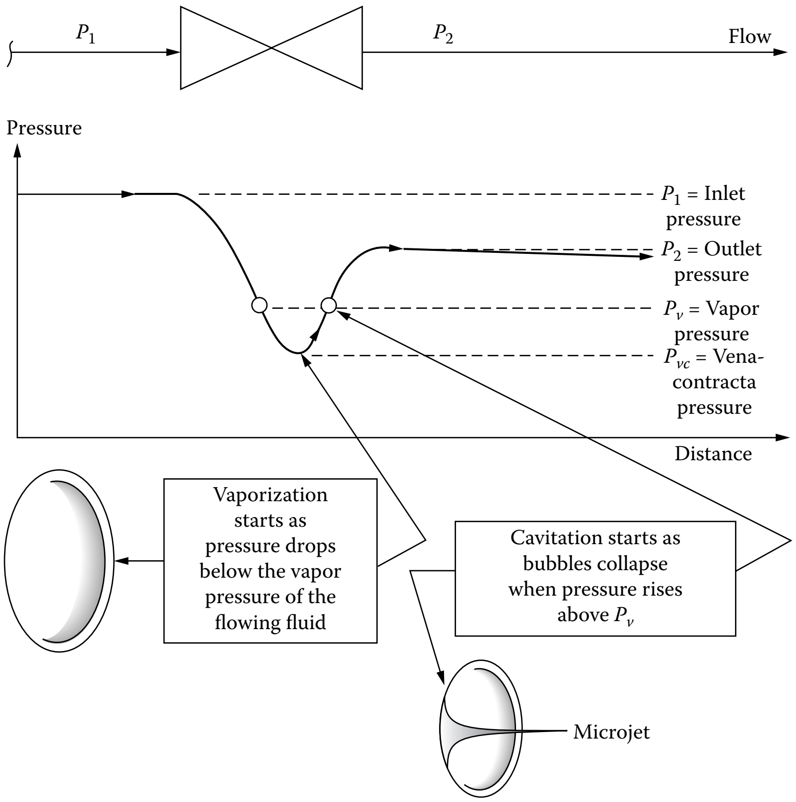


FIG. 6.1w

Cavitation occurs when downstream of the vena contracta the pressure rises. When it reaches the vapor pressure of the process fluid, the vapor bubbles implode and release powerful microjets that will damage any metallic surface in the area.

The greatest damage is caused by a dense pure liquid with high surface tension (e.g., water or mercury). Density governs the mass of the microjet stream, illustrated in Figure 6.1w, and surface tension governs the more important jet velocity. Mixtures are least damaging, because the bubble cannot collapse as suddenly. As the pressure increases, partial condensation in the bubble changes the vapor composition, leaving some vapor to slow the collapse. Some applications of cavitating mixed hydrocarbons show no mechanical damage or high noise level. Cavitation can be reduced or eliminated by several methods, listed in the following paragraphs.

Revising the Process Conditions A reduction of operating temperature can lower the vapor pressure sufficiently to eliminate cavitation. Similarly, increased upstream and downstream pressures, with Δp unaffected, or a reduction in the Δp can both relieve cavitation. Therefore, control valves that are likely to cavitate should be installed at the lowest possible elevation in the piping system and operated at minimum Δp. Moving the valve closer to the pump will also serve to elevate both the up- and downstream pressures.

If cavitating conditions are unavoidable, then it is preferred to have not only cavitation but also some permanent vaporization (flashing) through the valve. This can usually be accomplished by a slight increase in operating temperature or by decreasing the outlet pressure. Flashing eliminates cavitation by converting the incompressible liquid into a compressible mixture.

Revising the Valve Design Where the operating conditions cannot be changed, it is logical to review the type of the valve in terms of its pressure recovery characteristics. The more treacherous the flow path through a particular valve, the less likelihood exists for cavitation. Inversely, the valves most likely to cavitate are the high recovery valves (ball, butterfly, gate) having low F_L and F_c coefficients (Section 6.15). Figure 6.1x illustrates some of the ways available to eliminate cavitation.

Figure 6.1y shows a number of anticavitation valve designs that combine multiple-port and multiple-flow-path features.

If cavitation is anticipated, the engineer should select valves with low recovery and, therefore, high F_c and F_L coefficients. Different valve designs react differently to the effects of cavitation, depending upon where the bubbles collapse. If the focus is in midstream, materials may be unaffected. For example, in the "Swiss cheese"-type design, small holes in the skirt or cage are arranged in pairs on opposite sides of the centerline of the valve. Streams from opposing holes impinge on each other, causing the cavities to collapse in the liquid pool (theoretically). This method, illustrated in Figure 6.1z has been used successfully for mild cavitation.

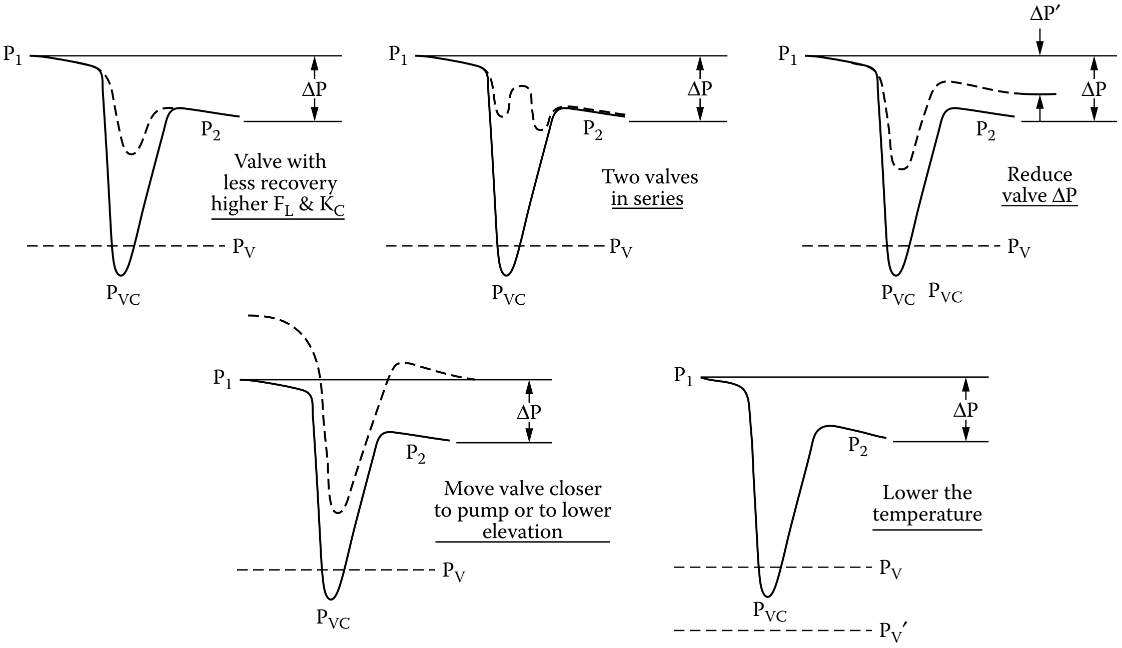


FIG. 6.1x
The pressure profiles shown in dotted lines illustrate some of the options available to the process control engineer to eliminate cavitation.

Labyrinth-type valves avoid cavitation by a very large series of right-angle turns with negligible pressure recovery at each turn, but the narrow channels are subject to plugging if particulate matter is in the stream (Figure 6.1aa).

The multistep valves at the bottom of Figure 6.1aa can avoid cavitation by replacing a single and deep vena contracta, as would occur in a single-port valve, with several small vena contracta points as the pressure drop is distributed between several ports working in series. If the vapor pressure of the process fluid is below the outlet pressure of the valve (Condition A), this valve is likely to work.

On the other hand, if P_v is greater than P_2 (Condition B), this valve is likely to cavitate in its noted intermediate port. One might note that if Condition B occurred in the conventional valve (noted by the dotted line), no cavitation would occur, because some of the vapor formed at the vena contracta would never recondense but would stay in the vapor state (flashing). Therefore, these multistep valves are not recommended for flashing applications.

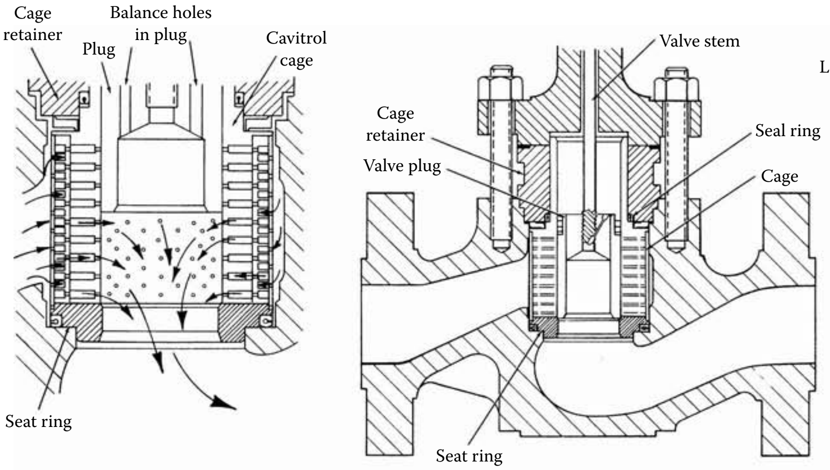
Gas Injection Another valve design variation that can alleviate cavitation is based on the introduction of noncondensible gases or air into the region where cavitation is anticipated. The presence of this noncompressible gas prevents the sudden collapse of the vapor bubbles as the pressure recovers to values exceeding the vapor pressure, and instead of implosions, a more gradual condensation process occurs. As shown in Figure 6.1bb, the gas may be admitted through the valve shaft or through downstream taps on either side of the pipe, in line with the shaft and as close to the valve as possible. Because the fluid vapor pressure is usually less than atmospheric, the air or gas need not be under pressure.

Revising the Installation In order to eliminate cavitation, it is possible to install two or more control valves in series. Cavitation problems can also be alleviated by absorbing some of the pressure drop in restriction orifices, chokes, or in partially open block valves upstream or downstream to the valve. The amount of cavitation damage is related to the sixth power of flow velocity or to the third power of pressure drop. This is the reason why reducing Δp by a factor of two, for example, will result in an eightfold reduction in cavitation destruction.

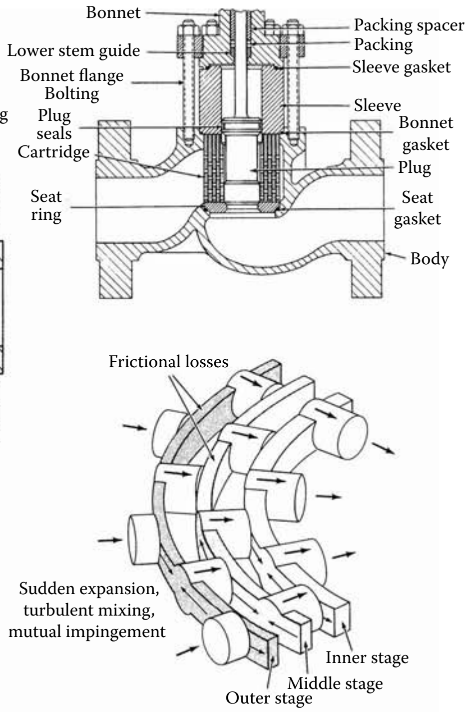
In some high-pressure let-down stations, it might not be possible to completely eliminate cavitation accompanied by erosion or corrosion. In such installations, one might consider the use of inexpensive choke fittings (shown in Figure 6.1cc) instead of (or downstream to) control valves.

A single, fixed-opening choke fitting is applicable only when the process flow rate is relatively constant. For variable-flow applications, one can provide several choke fittings of different capacities isolated by several full bore on/off valves, providing a means of matching the process flow with the opening of the required number of chokes. If the chokes discharge into the vapor space of a tank, this will minimize cavitation damage because the bubbles will not be collapsing near to any metallic surfaces.

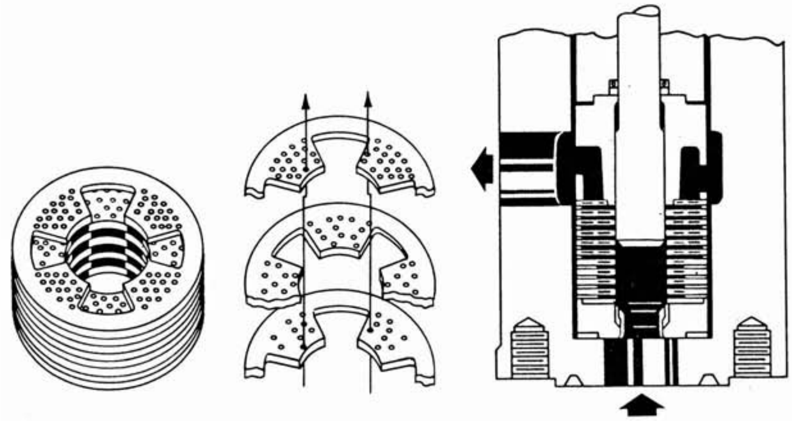
Material Selection for Cavitation While no material known today will stand up to cavitation, some will last longer than others. Table 6.1dd shows that the best overall selection for cavitation resistance is Stellite 6B (28% chromium, 4% tungsten, 1% carbon, 67% cobalt). This is a wrought material and can be welded to form valve trims in sizes up to 3 in. (75 mm). Stellite 6 is used for hard-facing of trims and has the same



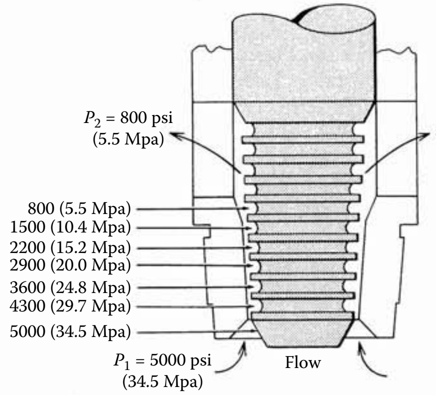
Cavitrol® trim element and valve installation. (*Courtesy of Fisher Controls, Emerson Process Management.*)



Channel stream® trim detail and valve installation. (*Courtesy of Flowserve Corp.*)

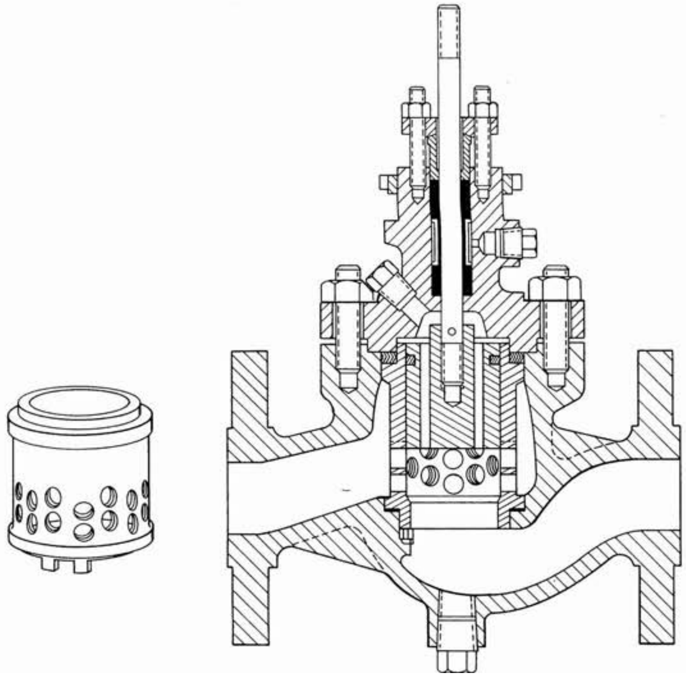


VRT® (Variable Resistance Trim) element, plates, and assembly in a valve. (*Courtesy of Masoneilan, Division of Dresser Flow Control.*)

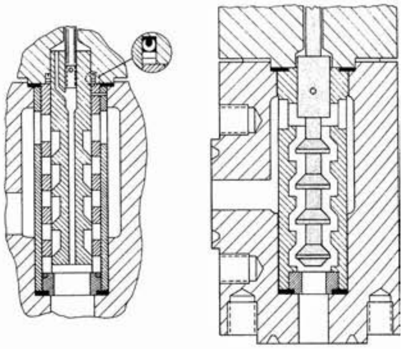


Typical pressure drop chart

Cascade-Trim® step type plug for high pressure breakdown. (*Courtesy of Copes Vulcan, Inc., SPS Process Equipment.*)

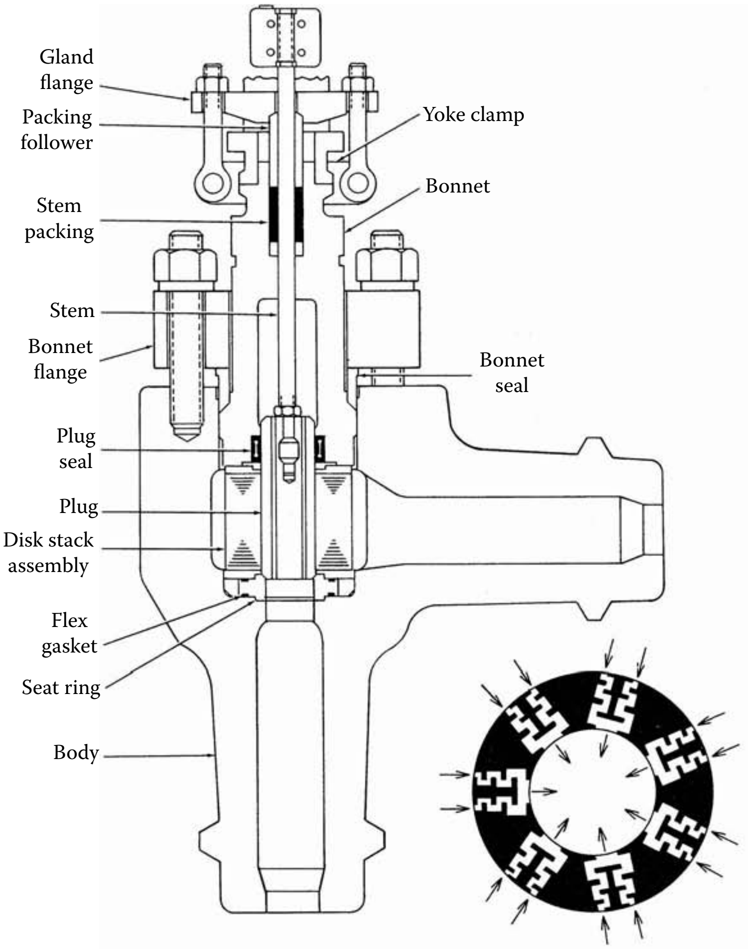


Flash-Flow® trim element and valve installation. (*Courtesy of Hammel Dahl ConoFlow.*)

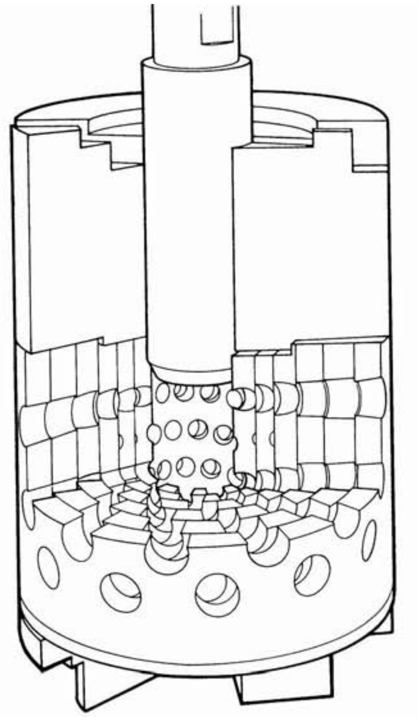


Step plug and orifice trim for liquid service. (*Courtesy of Masoneilan, Division of Dresser Flow Control.*)

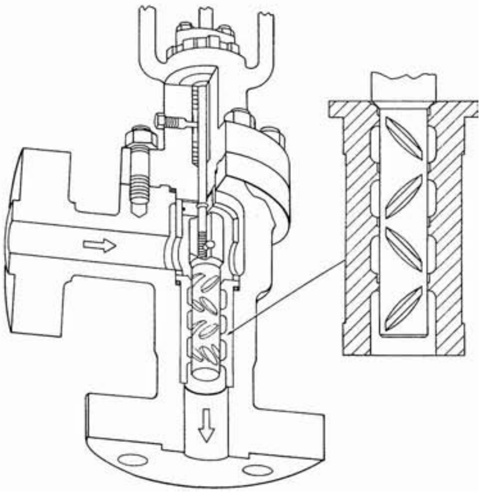
FIG. 6.1y
Control valve designs that are less likely to cavitate due to their multipath and multiturn flow paths.



Self-drag® valve with example of disk element. (*Courtesy of CCI-Control Components, Inc.*)



Hush® trim element and plug flow is into plug bore and out. (*Courtesy of Copes Vulcan, Inc., SPC Process Equipment.*)



Turbo-cascade® trim element and valve installation. (*Courtesy of Yarway , Tyco Valves & Controls.*)

FIG. 6.1y
(Continued).

chemical composition but less impact resistance. Correspondingly, its cost is lower.

In summary, the applications engineer should first review the potential methods of eliminating cavitation. These would include adjustment of process conditions, revision of valve type, or change of installation layout. If none of these techniques can guarantee the complete elimination of cavitating conditions, the design engineer should install chokes or special anticavitation valves that can last for some reasonable period, even if some cavitation is occurring.

Control Valve Noise

The calculation of noise levels generated by control valves and the methods of lowering these noise levels are both covered in Section 6.14 and, therefore, will not be repeated

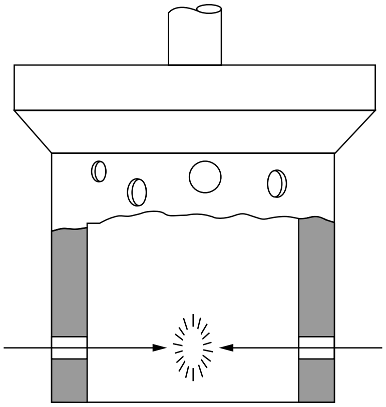
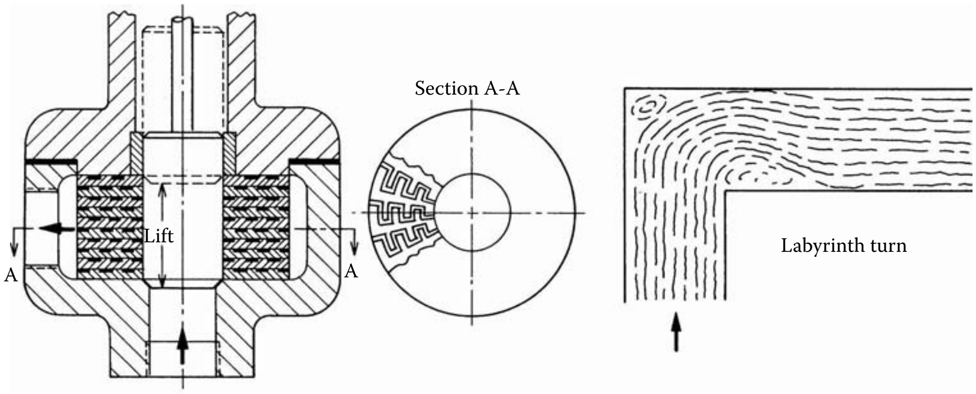


FIG. 6.1z
The "Swiss cheese" design can withstand mild cavitation.

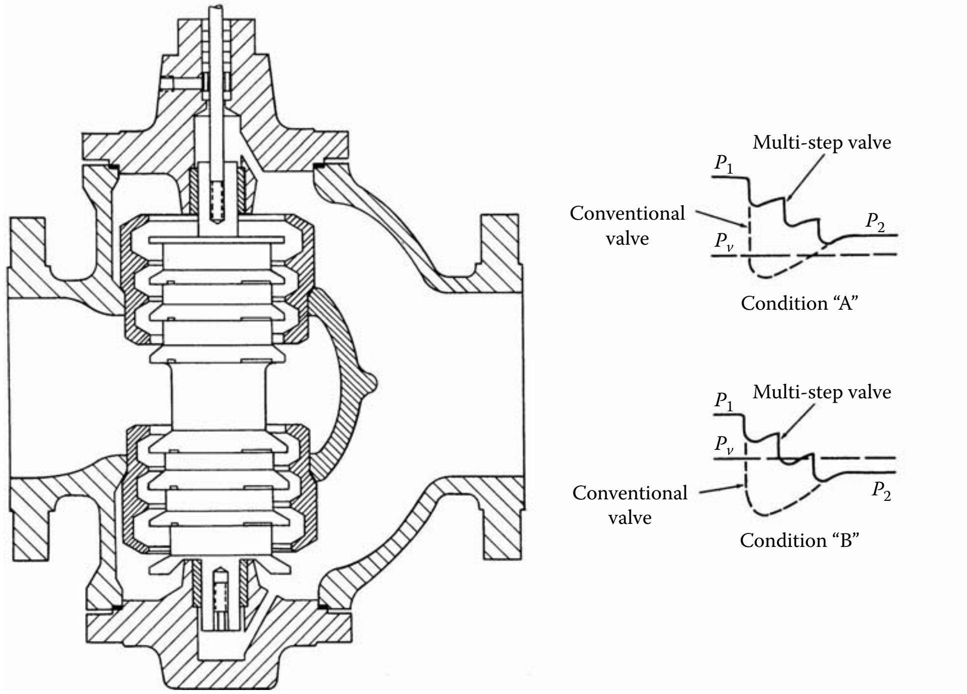
here. As can be seen in that section, many of the features of low-noise control valves are similar to the features of the anticavitation valves shown in Figure 6.1y.

Flashing and Erosion

Cavitation occurs when (in Figure 6.1w) $p_2 > p_v$, while flashing takes place when $p_2 < p_v$. When a liquid flashes into vapor, there is a large increase in volume. In this circumstance, the piping downstream of a valve needs to be much larger than the inlet piping in order to keep the velocity of the two-phase stream low enough to prevent erosion. The ideal valve to use for such applications is an angle valve with an oversized outlet connection. In Section 6.15, the method for calculating the exit velocity in such two-phase flashing applications is illustrated



CCI-Courtesy Control Components, Inc.



Courtesy Masoneilan, Division of Dresser Flow Control.

FIG. 6.1aa
The labyrinth (top) and multistep (bottom) valve designs help to reduce the probability of cavitation.

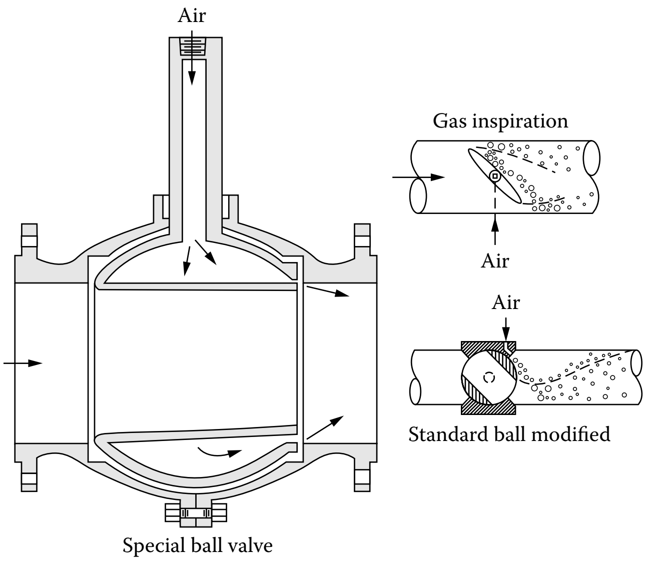


FIG. 6.1bb
Cavitation can also be alleviated by the admission of air into the flowing stream.

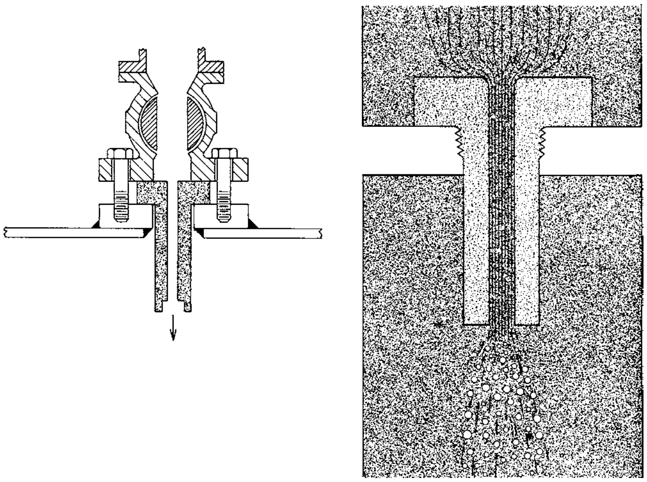


FIG. 6.1cc
The probability of cavitation can also be reduced by installing a choke fitting downstream of the valve.

by an example. In addition, the piping must be designed so that it is not damaged by slug flow.

The impingement of liquid droplets can be erosive if the velocity is great enough (> 200 ft/s, or 60 m/s, across the orifice), such as in applications involving high-pressure let-down of gas or vapor with suspended droplets. On high-pressure let-down applications, the ideal valve to use is the dynamically balanced plug valve provided with a hard-faced plug.

Section 6.15 gives an example on how the exit velocity from a steam let-down valve can be calculated. Erosion caused by high exit velocities can also cause corrosion problems.

TABLE 6.1dd
Relative Resistance of Various Materials to Cavitation

Trim or Valve Body Material	*Relative Cavitation Resistance Index*	*Approximate Rockwell C Hardness Values*	*Corrosion Resistance*	*Cost*
Aluminum	1	0	Fair	Low
Synthetic sapphire	5	Very high	Excellent	High
Brass	12	2	Poor	Low
Carbon steel, AISI C1213	28	30	Fair	Low
Carbon steel, WCB	60	40	Fair	Low
Nodular iron	70	3	Fair	Low
Cast iron	120	25	Poor	Low
Tungsten carbide	140	72	Good	High
Stellite #1	150	54	Good	Medium
Stainless steel, type 316	160	35	Excellent	Medium
Stainless steel, type 410	200	40	Good	Medium
Aluminum oxide	200	72	Fair	High
K-Monel	300	32	Excellent	High
Stainless steel, type 17-4 pH	340	44	Excellent	Medium
Stellite #12	350	47	Excellent	Medium
Stainless steel, type 440C	400	55	Fair	High
Stainless steel, type 329, annealed	1000	45	Excellent	Medium
Stellite #6	3500	44	Excellent	Medium
Stellite #6B	3500	44	Excellent	High

Some metals do not corrode due to a self-regenerating protective surface film; however, if this film is removed by erosion faster than it is formed, the metal corrodes rapidly. For lining, either nobler metals or ceramics should be considered in such situations. The preferred arrangement for flashing service is to use a reduced port angle valve discharging directly into a vessel or flash tank.

Corrosion

Some information data on corrosion is contained in Table 6.1dd. A detailed tabulation of the chemical resistance of materials is given in Appendix A.3 of this volume. In evaluating a particular application, one should also consider the facts that most process fluids are not pure and that the corrosion rate is much influenced by flow velocity and the presence of dissolved oxygen.

On corrosive services, one can also consider the use of lined valves (tantalum, glass, plastics, and elastomers), but one should consider the consequences of lining failure. Damage can be caused by accidentally exposing the lining to high concentrations of inhibitors or line cleaning fluids. Gas absorption in elastomer linings can also cause blistering.

Viscous and Slurry Service

When the process stream is highly viscous or when it contains solids in suspension, the control valve is selected to provide an unobstructed streamline flow path.

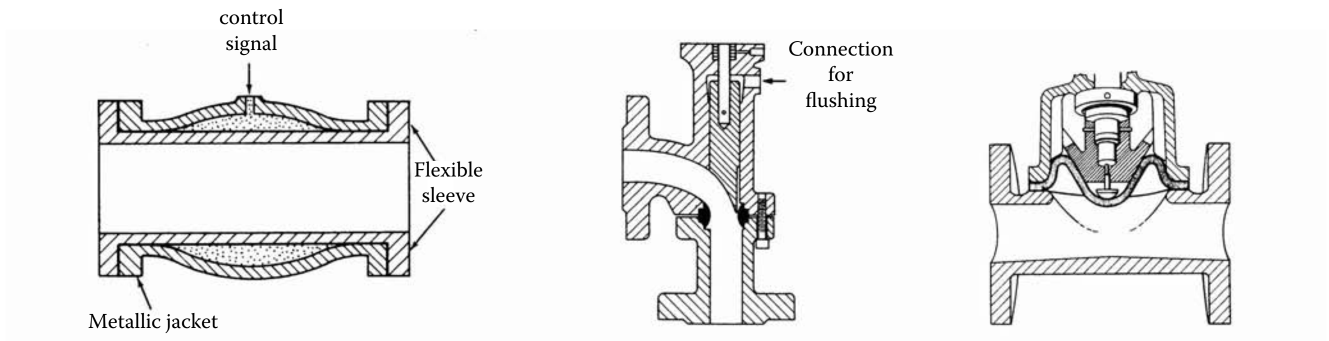
The chief difficulty encountered with heavy slurry streams is plugging. Conditions that can contribute to this include a difficult flow path through the valve, shoulders, pockets, or dead-ended cavities in contact with the process stream. Valves with these characteristics must be avoided because they represent potential areas in which the slurry can accumulate, settle out, and gel, freeze, solidify, decompose, or as most frequently occurs, plug the valve completely.

The ideal slurry valve is one that

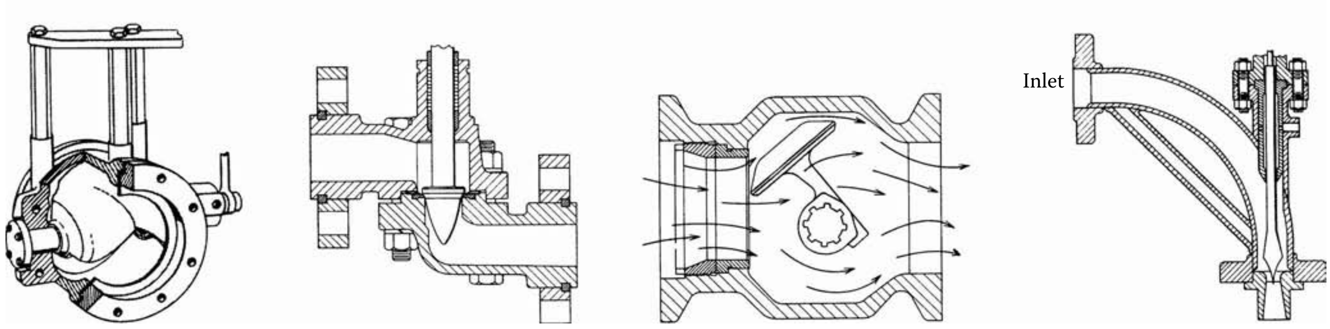
1. Provides full pipeline opening in its open position
2. Provides for unobstructed and streamlined flow in its throttling position
3. Has high pressure and temperature ratings
4. Is available in corrosion-resistant materials
5. Is self-draining and has a smooth contoured flow path
6. Will fail safe
7. Has acceptable characteristics and rangeability
8. Has top works that are positively sealed from the process

Unfortunately, no one valve meets all of these requirements, and the instrument engineer has to judge which features are essential and which can be compromised.

If, for example, it is essential to provide a full pipe opening when the valve is open, there are several valves that can satisfy this requirement. They include the various pinch valves (A in Figure 6.1ee), the full opening angle valves (B



A Jacketed pinch valve B Full opening angle valve C Saunders valve



D Characterized ball valve E Self-draining valve F Eccentric rotating plug valve G Sweep angle valve

FIG. 6.1ee
Valves for viscous and slurry services.

in Figure 6.1ee), some of the Saunders valve designs (C in Figure 6.1ee), and the full-ported ball valves (Section 6.16). While these units all satisfy the requirement for a fully open pipeline when open, they differ in their limitations.

The pinch valves, for example, are limited in their materials of construction, pressure, temperature ratings, flow characteristic, speed of response, and rangeability, but they do provide self-cleaning streamlined flow, which in some designs resembles the characteristics of variable venturi.

The Saunders valves and pinch valves have similar features, including the important consideration that the sealing of the process fluid does not depend on stuffing boxes. They are superior in the availability of corrosion-resistant materials, but they are inferior if completely unobstructed streamline flow is desired. Pinch valves are suitable for very low pressure drop services only, while Saunders and wedge plug valves can operate at slightly higher pressures. Lined butterfly valves are a good choice if the process pressure is high while the valve drop is low.

The angle valve with a scooped-out plug satisfies most requirements except that its flow characteristics are not the best, and it is necessary to purge it above the plug in order to prevent solids from migrating into that area.

Full-ported ball valves in their open position are as good as an open pipe section, but in their throttling positions both their flow paths and their pressure recovery characteristic are less desirable.

Valves that do not open to the full pipe diameter but still merit consideration in slurry service include the following designs: characterized ball valves (D in Figure 6.1ee), various self-draining valve types (E in Figure 6.1ee), the eccentric disc rotating globe designs (F in Figure 6.1ee), and the sweep angle valves (G in Figure 6.1ee).

Each of these has some features that represent an improvement over some other design. The characterized ball valve, for example, exhibits an improved flow characteristic in comparison with the full-ported ball type. It is well-suited for process fluids containing fibers or larger particles. The self-draining valve allows slurries to be flushed out of the system periodically. Complete drainage is guaranteed by the fact that all surfaces are sloping downstream.

The sweep angle valve, with its wide-radius inlet bend and its venturi outlet, is in many ways like the angle slurry valve. Its streamlined nonclogging inner contour minimizes erosion and reduces turbulence. In order to prevent the process fluid from entering the stuffing box, a scraper can be furnished, which if necessary can also be flushed with some purge fluid. The orifice located at the very outlet is built like a choke fitting. Both orifice and plug may be made of abrasion-resistant ceramic or hard metals.

For slurries with large solid particles, the ideal orifice shape is a circle, such as that of an iris valve or a jacketed pinch valve (A in Figure 6.1ee).

Orifice size, particle size, and rangeability are interrelated. For any particle size and orifice shape, there is a minimum opening below which plugging can be expected. To get good rangeability (control at low flow rates) the valve Δp should be made small. One way to accomplish this is by use of a "head box" (Figure 6.1ff). The selection of the valve style and the piping configuration around the valve inlet must be guided by the intractability of the particular slurry.

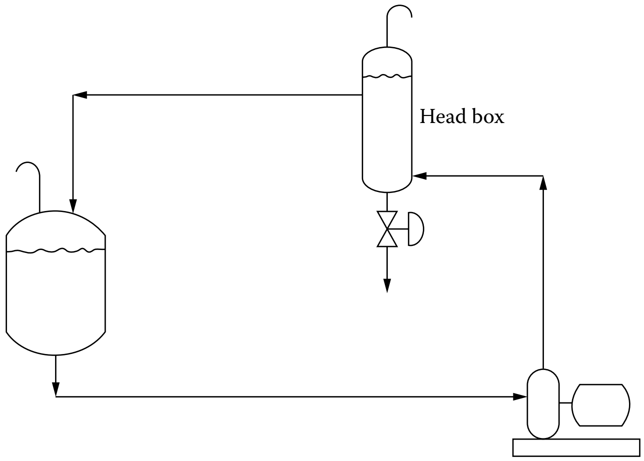


FIG. 6.1ff
In slurry service, the use of a head box can provide a small and constant pressure drop across the control valve.

Valves That Can Be Sterilized

In food processing applications, in addition to the above, valves must not contain pockets where process material can be retained, and they should be constructed so that they can be easily sterilized and disassembled for cleaning. Materials of construction should not contain compounds that are prohibited by the FDA, such as some of the elastomer compounding materials.

Valve Leakage

Any flow through a fully closed control valve when exposed to the operating pressure differentials and temperatures is referred to as leakage. It is expressed as a cumulative quantity over a specified time period for tight shut-off designs and as a percentage of full capacity for conventional control valves.

According to ANSI B16.104, valves are categorized according to their allowable leakage into six classes. These leakage limits are applicable to unused valves only:

Class I valves are neither tested nor guaranteed for leakage.
Class II valves are rated to have less than 0.5% leakage.
Class III valves are allowed up to 0.1% leakage.
Class IV valves must not leak more than 0.01% of their capacity.

TABLE 6.1gg
Valve Seat Leakage Classifications per ANSI B16.104-1976 (FCI 70–2)

Class	*Maximum Leakage*
I	No test required
II	0.5% of rated valve capacity
III	0.1% of rated valve capacity
IV	0.01% of rated valve capacity
V	5×10^{-4} ml/minute of water per inch of orifice diameter per psi differential
VI	ml per minute of air or nitrogen vs. port diameter per the following tabulation

Nominal Port Diameter		*Maximum Seat Leakage, ml/minute*
in.	*mm*	
1	25	0.15
$1^1/_2$	38	0.30
2	50	0.45
3	75	0.90
4	100	1.70
6	150	4.00
8	200	6.75

Class V valves are specified to have a leakage of 5×10^{-4} ml/min water flow per inch (25.4 mm) of seat diameter, per 1 psi (0,0685 bar) differential pressure.

Class VI is for soft-seated valves, and leakage is expressed as volumetric air flow at rated Δp up to 50 psi (3.45 bar).

Generally the functions of tight shut-off and those of control should not be assigned to the same valve. The best shut-off valves are rotary on/off valves, which are not necessarily the best choices for control.

Table 6.1gg gives a summary of the ANSI leakage classes, and Table 6.1a lists the leakage capabilities of the various control valve designs. For added details on soft-seated designs and other special designs, refer to the feature summaries at the front of each section.

Some valve manufacturers list in their catalogs the valve coefficients applicable to the fully closed valve. For example, a butterfly valve supplier might list a C_v of 13.2 (K_v = 11.4) for a fully closed, metal-to-metal seated 24 in. (600 mm) valve. It should be realized that such figures apply only to new, clean valves operating at ambient conditions. After a few years of service, valve leakage can vary drastically from installation to installation as affected by some of the factors to be discussed.

It should also be noted that some fluids are more difficult to hold than others. Low-viscosity fluids such as Dowtherm, refrigerants, or hydrogen are examples of such fluids.

Soft Seats One of the most widely applied techniques for providing tight shut-off over reasonable periods of time is the use of soft seats. Standard materials used for such services include Teflon and Buna-N. Teflon is superior in its corrosion resistance and in its compatibility to high-temperature services up to 450°F (232°C). Buna-N is softer than Teflon but is limited to services at 200°F (93°C) or below.

Neither should be considered for operating conditions such as static pressures of 500 PSIG (34.5 bar) or greater, for use with fluid containing abrasive particles, or if critical flow is expected at the valve seat.

The leakage of double-ported valves is much greater than that of single-ported ones, and it can be as high as 2–3% of full capacity in metal-to-metal seated designs.

Temperature and Pipe Strain It is frequently the case that either the valve body is at a different temperature than the trim or the thermal expansion factor for the valve plug is different from the coefficient for the body material. It is usual practice in some valve designs (such as the butterfly) to provide additional clearance to accommodate the expansion of the trim when designing for hot fluid service. The leakage will, therefore, be substantially greater if such a valve is used at temperatures below those for which it was designed.

Temperature gradients across the valve can also generate strains that promote leakage. Such gradients are particularly likely to exist in three-way valves when they are in combining service and when the two fluids involved are at different temperatures. This is not to imply that three-way valves are inferior from a leakage point of view. Actually, their shut-off tightness is comparable to that of single-seated globe valves.

Pipe strains on a control valve will also promote leakage. For this reason, it is important not to expose the valve to excessive bolting strains when placing it in the pipeline and to isolate it from external pipe forces by providing sufficient supports for the piping.

Seating Forces and Materials The higher the seating force in a globe valve, the less leakage is likely to occur. An average valve has a seating force of 50 lb_f per linear inch (8750 N/m) of seat circumference. Where necessary, a much increased seating force will create better surface contact by actually yielding the seat material. Seating forces of this magnitude (about ten times the normal) are practical only when the port is small.

Seating materials are selected for compatibility with service conditions, and Stellite or hardened stainless steel is an appropriate choice for nonlubricating, abrasive, high-temperature, and high pressure drop services. These hard surface materials also reduce the probability of nicks or cuts occurring in the seating surface, which might necessitate maintenance or replacement.

Small-Flow Valves Valves with small flow rates are found in laboratory and pilot plant applications. Even in industrial installations, the injection of small quantities of neutralizers, catalysts, inhibitors, or coloring agents can involve flows in the range of cubic centimeter per minute.

Valves are usually considered miniature if their C_v is less than 1 ($K_v = 0.862$). This generally means a $^1/_4$ in. (6.25 mm) body connection and a $^1/_4$ in. (6.25 mm) or smaller trim. The top works are selected to protect against oversizing, which could damage the precise plug.

The field of small-flow control valves is a highly specialized area, unlike any other application. The mechanical design and fluid flow constraints encountered essentially make these valves custom applied for the service. Small-flow valves are used in laboratories, process pilot plants, and some areas of full-scale plants.

The design and building of these valves test the ingenuity of any manufacturer. There are several design types available, and great care must be taken to match the application with the valve design. While the manufacturers publish C_v (K_v) ratings for their valves and trims, these should be treated with extreme caution and used for reference purposes only (below 0.01 C_v [0.00862 K_v]).

In many cases, the flow may shift from laminar to turbulent with flowing conditions and valve stroke changes. It is quite common for a laminar flow pattern to predominate, particularly with viscous fluids or in low-pressure applications. Laminar flow means the flowing quantity will vary directly with pressure drop instead of with the square root of pressure drop.

It is wise to devise a test procedure simulating the service application to evaluate performance of a specific valve before using it in actual service. It is not uncommon for two "identical" small valves to exhibit somewhat different $C_v(K_v)$ capacities and flow curves under the same test conditions.

Needle Valves There are at least three approaches to the design of miniature control valves: (1) the use of smooth-surfaced needle plugs, (2) the use of cylindrical plugs with a flute or flutes milled on it, and (3) positioning the plug by rotating the stem.

One of the most common designs looks like a miniature version of the standard globe control valve (Figure 6.1hh). The trim consists of a precision honed and close-fitting plug fitted into an orifice made of a hard alloy. The control area consists of a fine taper slot milled into the outer surface of the piston-shaped plug or a long shallow taper plug. In smaller C_v (K_v) trims, this slot may be a calibrated scratch in the surface.

It is not uncommon to find up to 30 trims available in a given body to cover the C range of 0.000002 to 0.1 (K from 0.00000172 to 0.0862 K_v). The rangeability, i.e., the ratio between maximum and minimum controllable flow, can be limited for this type of valve due to the inherent leakage flow between piston and orifice. The smaller the trim size, the lower the rangeability.

Needle plugs give more dependable results than the ones with grooves, scratches, or notches because the flow is distributed around the entire periphery of the profile. This results in even wear of the seating surfaces and eliminates side thrusts against the seat. The trim is machined for very small clearances, and hard materials or facings are recommended to minimize wear and erosion.

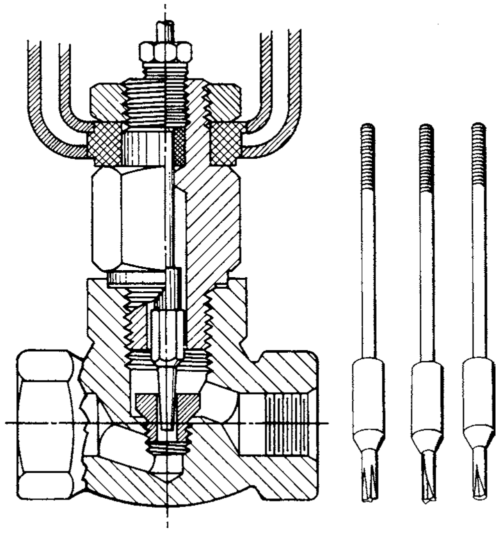


FIG. 6.1hh
Needle-type small-flow valve and plugs with flutes milled on cylindrical surface.

Needle plugs are available with equal-percentage (down to $C_v = 0.05$ [$K_v = 0.043$]), linear, and quick-opening characteristics (Figure 6.1hh). Some manufacturers claim the availability of valves with coefficients of $C_v = 0.0001$ ($K_v = 0.000086$) or less. At these extremely small sizes it is very difficult to characterize the plugs (equal-percentage is not available), and the valve rangeability also suffers.

It is easier to manufacture the smaller cylindrical plugs with one or more grooves (Figure 6.1hh) and obtain the desired flow characteristic by varying the milling depth. Both the needle and the flute plugs are economical, but it is difficult to reproduce their characteristics and capacity accurately.

Ball Valves In contrast to the relatively long stroke piston and orifice valve discussed, there is another valve design for small-flow applications that has a very short and variable adjusted stroke (Figure 6.1ii). Here, a synthetic sapphire ball is allowed to lift off a metal orifice and throttle the flow. The particular advantage of this valve is that the diaphragm stroke can be adjusted to produce various stem lifts with a standard 3–15 PSIG (0.2–1.0 bar) signal.

Two versions are available. One can be adjusted to cover 0.07 to 0.00007 C_v (0.06 to 0.00006 K_v) and the other can be adjusted from 1 to 0.001 C_v (0.862 to 0.000862 K_v). These can be used for high-pressure, high-drop applications ranging from 3000 to 30,000 PSIG (207 to 2068 bar), depending upon the model.

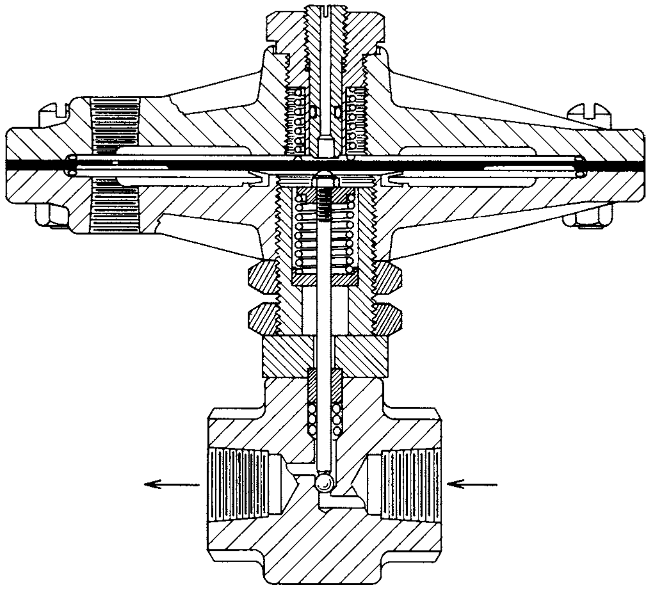


FIG. 6.1ii
Low-flow ball valve with adjustable short-stroke actuator. (Courtesy of A.W. Cash Co.)

Stem Rotation Type Another short-stroke valve especially suitable for high-pressure service (up to 50,000 PSIG, or 3447 bar) is shown in Figure 6.1jj. Variation in C_v (K_v) rating of identical plugs and seats is achieved by mechanical adjustment of a toggle arrangement. The toggle can change the valve stroke between 0.010 and 0.150 in. (0.25 to 3.75 mm). With different trim inserts, a C_v range of 1 to 0.000001 (K_v range of 0.862 to 0.00000086) can be covered.

In this design, the lateral motion of the plug is achieved by rotating the stem through a lead screw. The linear diaphragm motion is transferred into rotation by the use of a slip ball joint. Valve capacity is a function of orifice diameter (down to 0.02 in., or 0.5 mm), number of threads per inch in the lead screw (from 11 to 32), amount of stem rotation (from 15 to 60°), and the resulting total lift, which generally varies from 0.005 to 0.02 in. (0.125 to 0.5 mm).

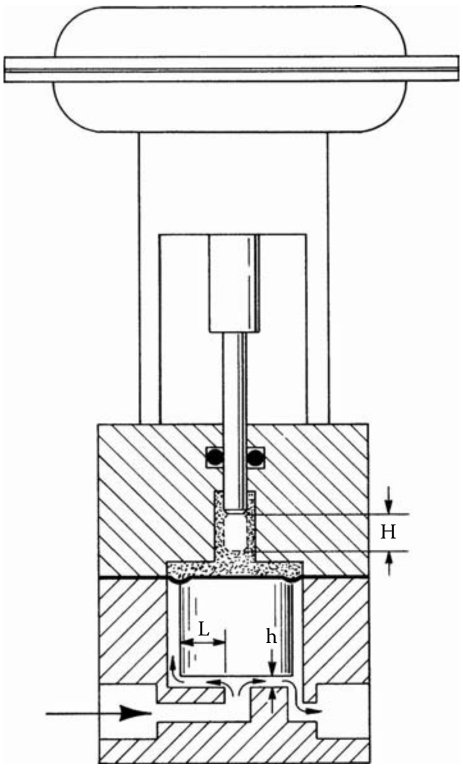


FIG. 6.1kk
Low-flow valve using laminar flow element. (Courtesy of Emerson Process Management.)

The extremely short distance of valve travel makes accurate positioning of the plug essential, and this necessitates a positioner. The combination of a long stem and short plug travel makes this valve sensitive to stem load and temperature effects. Because this differential thermal expansion can cause substantial errors in plug position, this valve is limited to operating temperatures below 300°F (149°C).

Laminar Valve Finally, the latest addition to the low-flow valve family is shown in Figure 6.1kk. This valve is designed

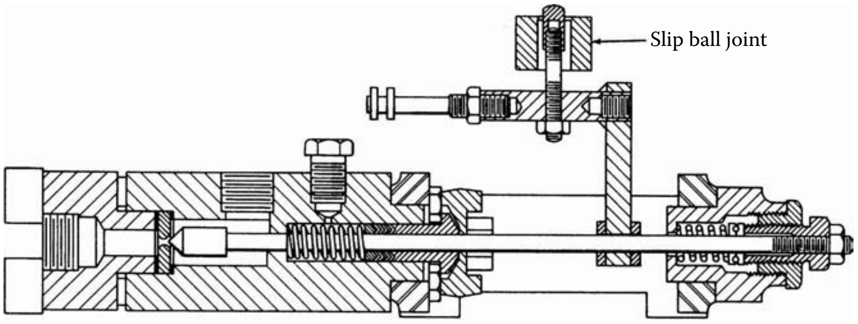


FIG. 6.1jj
Small-flow plug positioned by stem rotation.

to operate on the laminar flow principle. The flow is controlled by forcing the fluid through a long and narrow path, formed between two parallel surfaces. The actuator varies the laminar gap through a diaphragm and O-ring-sealed hydraulic ram. Because, in the laminar regime, flow varies linearly with pressure drop across the valve and with the third power of the gap width between the two surfaces (valve travel), this design has extremely high control rangeability along with a wide C_v range of 0.02 to 0.000001 (K_v from 0.017 to 0.00000086). The laminar principle also eliminates cavitation effects with liquids and sonic choking velocities with gases. This design can be used with inlet pressures and pressure drops up to 2675 PSIG (185 bar).

While it is not commonly thought of for low-flow valve application, the labyrinth disc valve design (see Figure 6.1aa) can be manufactured for this purpose. Its most common application is for reducing high-pressure fluid samples for analyzers, but obviously it can be used in other services. One common factor for this design, as well as for all of the others discussed, is the need for the fluid to be extremely clean. These valves are not tolerant of dirt or sediment due to the small passage and close clearances. Unless the fluid is known to be clean, it is necessary to provide for a high level of filtration upstream.

INSTALLATION

When a valve is larger than 4 in. (100 mm) or sometimes when it is more than one size smaller than the pipe, it is advisable to use pipe anchors to minimize force concentrations at the reducers and more frequently to relieve flange stress loading due to valve weight. The end connections on the valve should match the pipe specifications. If welded valves are specified, the nipples should be factory-welded and the welds should be stress-relieved. If lined valves are specified, their inside diameters should match that of the pipe to avoid extrusion. On flangeless valves, the bolting and the tightness of the gaskets can be a problem if the valve body is long.

If valves are fast closing (or fail) in long liquid lines, water hammer can result in the upstream pipe or vacuum can develop in the downstream line. Fast-opening steam valves can thermally shock the downstream piping. Steam traps should be provided at all low points in a steam piping network. Anchors should be provided in all locations where sudden valve repositioning can cause reaction forces to develop.

Flow-to-close single-seated valves should not be used because if operated close to the seat, hydraulic hammer can occur. If the damping effect of the actuator alone will not overcome the vertical plug oscillation, then either the actuator should be made "stiffer" (higher air pressure operation) or hydraulic snubbers should be installed between the yoke and the diaphragm casing.

Climate and Atmospheric Corrosion

In humid environments such as the tropics, moisture will collect in all enclosures, and therefore drains should be provided. Electrical parts should all be encapsulated where possible or be provided with suitable moistureproof coating.

Vent openings should be provided with storage plugs and insect screens. Even with such precautions, the vents and seals will require preventive maintenance and antifungus treatment in some extreme cases.

Cold climates can produce high breakaway torque of elastomers, and in general, metals and plastics will become more brittle. Electrohydraulic actuators will require heating because oils and greases can become very viscous.

In high-temperature environments, the weak link is usually the actuator, but liners, plastic parts, and electric components are also vulnerable. The damage is not only a function of the temperatures but also of the lengths of time periods of exposure. Diaphragm temperature limits are a function of their materials: Neoprene — 200°F (93.3°C), Nordel — 300°F (148.7°C), Viton — 450°F (232.2°C), silicone glass—500°F (260°C). For higher temperatures, one can replace the diaphragms with pistons (or with metallic bellows in some extreme cases) or add heat shields.

In power plant applications, valves frequently must be designed to withstand anticipated seismic forces. If the atmosphere contains corrosive gases or dusts, it is desirable to enclose, purge, or otherwise protect the more sensitive parts. The stem, for example, can be protected by a boot.

In hazardous areas, all electrical devices should either be replaced by pneumatic ones or be made intrinsically safe or explosionproof.

CONTROL VALVE SPECIFICATION FORM

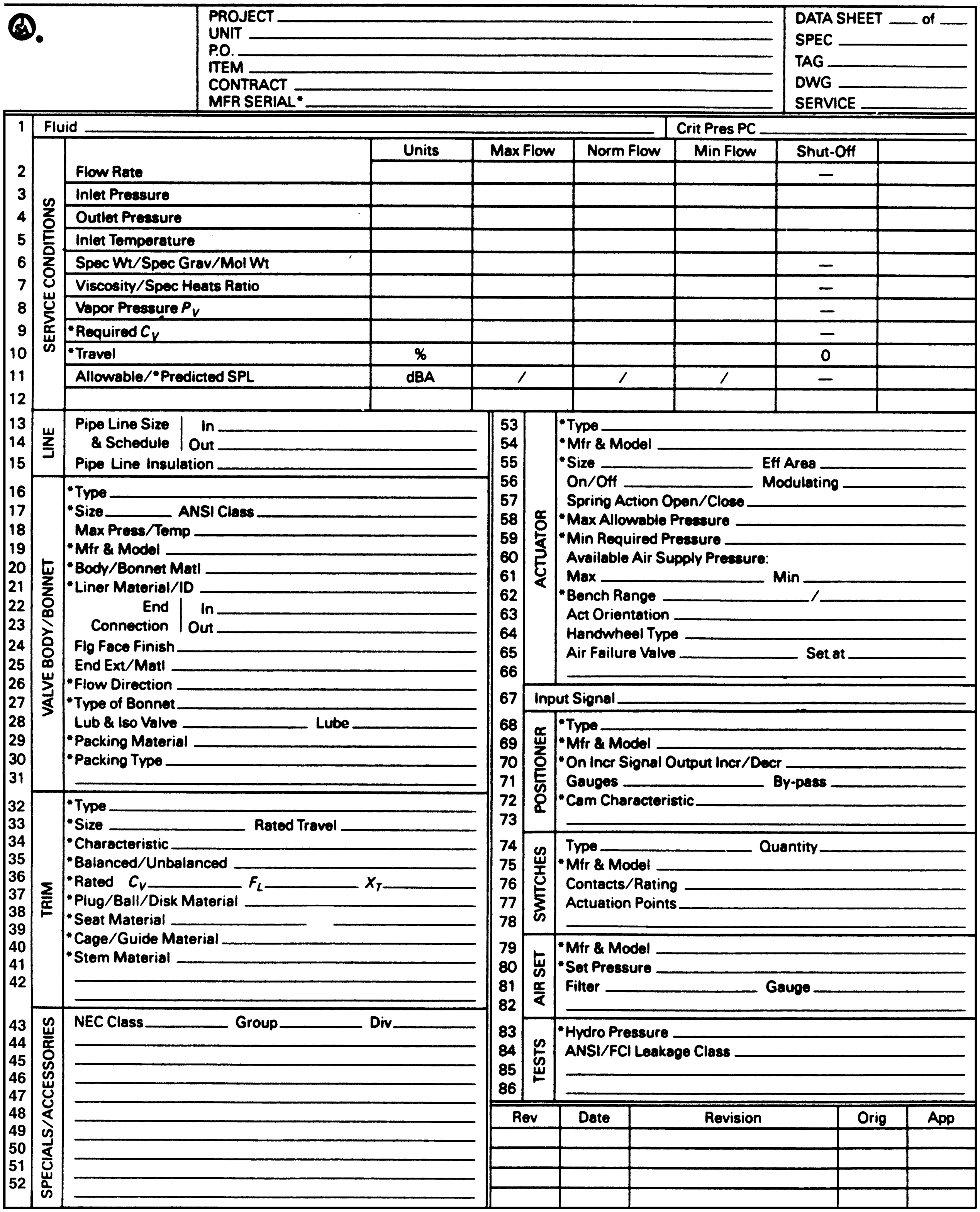
Compiling the information necessary to specify a control valve is best done with the aid of a tabulation sheet. Many large companies have their own customized forms. Figure 6.1ll shows a general-purpose form (ISA Form S20.50 Rev. 1) standardized by the Instrumentation, Systems, and Automation Society. After the form, general instructions are provided to assist in completing the form.

A similar data sheet standard has been published by IEC as publication IEC 60534-2-3, "Industrial Process Control Valves, Part 7: Control Valve Data Sheet."

References

1. Lytle, R. F., "Equipment Selection for Control System Performance," paper presented at PUPID-SECON Conference in Birmingham, U.K., 1974.
2. Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

FIG. 6.1ll
ISA S20 Specification Form For Control Valve



PROJECT ______
UNIT ______
P.O. ______
ITEM ______
CONTRACT ______
MFR SERIAL* ______

DATA SHEET ___ of ___
SPEC ______
TAG ______
DWG ______
SERVICE ______

1 Fluid ______ Crit Pres PC ______

			Units	Max Flow	Norm Flow	Min Flow	Shut-Off	
2	SERVICE CONDITIONS	Flow Rate					—	
3		Inlet Pressure						
4		Outlet Pressure						
5		Inlet Temperature						
6		Spec Wt/Spec Grav/Mol Wt					—	
7		Viscosity/Spec Heats Ratio					—	
8		Vapor Pressure P_V					—	
9		*Required C_V					—	
10		*Travel	%				0	
11		Allowable/*Predicted SPL	dBA	/	/	/	—	
12								

LINE
13 Pipe Line Size | In ______
14 & Schedule | Out ______
15 Pipe Line Insulation ______

VALVE BODY/BONNET
16 *Type ______
17 *Size ______ ANSI Class ______
18 Max Press/Temp ______
19 *Mfr & Model ______
20 *Body/Bonnet Matl ______
21 *Liner Material/ID ______
22 End | In ______
23 Connection | Out ______
24 Flg Face Finish ______
25 End Ext/Matl ______
26 *Flow Direction ______
27 *Type of Bonnet ______
28 Lub & Iso Valve ______ Lube ______
29 *Packing Material ______
30 *Packing Type ______
31 ______

TRIM
32 *Type ______
33 *Size ______ Rated Travel ______
34 *Characteristic ______
35 *Balanced/Unbalanced ______
36 *Rated C_V ______ F_L ______ X_T ______
37 *Plug/Ball/Disk Material ______
38 *Seat Material ______ ______
39 *Cage/Guide Material ______
40 *Stem Material ______
41 ______
42 ______

SPECIALS/ACCESSORIES
43 NEC Class ______ Group ______ Div ______
44 ______
45 ______
46 ______
47 ______
48 ______
49 ______
50 ______
51 ______
52 ______

ACTUATOR
53 *Type ______
54 *Mfr & Model ______
55 *Size ______ Eff Area ______
56 On/Off ______ Modulating ______
57 Spring Action Open/Close ______
58 *Max Allowable Pressure ______
59 *Min Required Pressure ______
60 Available Air Supply Pressure:
61 Max ______ Min ______
62 *Bench Range ______ / ______
63 Act Orientation ______
64 Handwheel Type ______
65 Air Failure Valve ______ Set at ______
66 ______

67 Input Signal ______

POSITIONER
68 *Type ______
69 *Mfr & Model ______
70 *On Incr Signal Output Incr/Decr ______
71 Gauges ______ By-pass ______
72 *Cam Characteristic ______
73 ______

SWITCHES
74 Type ______ Quantity ______
75 *Mfr & Model ______
76 Contacts/Rating ______
77 Actuation Points ______
78 ______

AIR SET
79 *Mfr & Model ______
80 *Set Pressure ______
81 Filter ______ Gauge ______
82 ______

TESTS
83 *Hydro Pressure ______
84 ANSI/FCI Leakage Class ______
85 ______
86 ______

Rev	Date	Revision	Orig	App

---

[*] Copyright ISA © 1981, reprinted with permission of the Instrumentation, Systems, and Automation Society (ISA).

Instructions for Filling Out the Valve Specification Sheet

Line	*Explanation of Terms and Definitions*	*Examples*
PROJECT	Specify project name for which control valve is intended.	XYZ Nuclear PS
UNIT	Specify unit within project.	#1
P.O.	Specify purchase order number from purchaser to control valve manufacturer.	P.O. 12345
ITEM	Specify item number of purchase order.	3
CONTRACT	Specific contract number of project for purchaser's reference.	56-V-32510
MFR SERIAL	This line may show the valve manufacturer's serial number(s) and is normally filled in at the time of shipment of the valve. Serial numbers often contain the manufacturer's shop order number.	C12650-3

Line	*Explanation of Terms and Definitions*	*Examples*
DATA SHEET	Specify data sheet number. Normally assigned by purchaser.	3 of 12
SPEC	Specify number of technical specification on which valve selection is based.	FL-13265-A
TAG	Specify tag number, if any, used to designate location of valve.	FV-103
DWG	Specify piping and instrumentation diagram number, loop diagram number, engineering flow diagram number, etc.	17-453
SERVICE	Describe service of control valve and/or pipe line number.	Feedwater control Reheat spray 2" MA 1051 WA7

NOTE: The above lines are suggestions only and may be modified to fit the individual company's needs. If the provided space is insufficient, add an additional sheet and refer to it.

Line No.	*Explanation of Terms and Definitions*	*Examples*
1	Describe fluid flowing into valve and its state. Indicate corrosive or erosive service and the corrosive or erosive agents.	Superheated stream Saturated water Crude oil and natural gas
	Specify thermodynamic critical pressure of the fluid.	3206 psia
2	Specify volumetric or mass flow rate at inlet or standard conditions. Maximum flow condition, if greater than normal flow condition, is the condition for which the valve is sized.	3000 gpm 10 000 bpd 600 std. m^3/s 7500 scfm 300 kg/h
3	Specify inlet pressure (gauge or absolute).	5000 psig 2000 kPa abs.
4	Specify outlet pressure (gauge or absolute).	1000 psig 400 kPa gauge
5	Specify inlet temperature in °F, °R, °C, or K. Must agree with state of fluid and its inlet pressure.	750°F 200°C 815 K
6	Specify specific weight (in lb/ft^3 or kg/m^3), specific gravity, or molecular weight of fluid. Identify the appropriate term.	61.9 lb/ft^3 1.03 44.01
7	Specify viscosity in appropriate units for liquids or specific heats ratio for gases.	20 centipoise 17.8 centistokes 1.27
8	Specify vapor (saturation) pressure at inlet temperature in absolute units. Only required for liquid flow.	680 psia 46.9 bar abs.
9	Specify required C_v as calculated for each condition per ANSI/ISA S75.01-1986. No additional safety (oversize) factor should be included at this point.	260

Line No.	*Explanation of Terms and Definitions*	*Examples*
10	Specify travel of the valve in percent of rated travel calculated from required C_v, rated C_v of the valve, trim selected, and characteristic (see lines 33, 34, and 36). 0% is full closed, 100% is full open.	78%
11	Specify laboratory-measured allowable and predicted sound pressure levels, both normally in dBA as measured per ISA-S75.07-1987.	90/87 dBA
12	Extra line for information not covered in lines 1 through 11.	Compressibility factor Z Ambient temperature Base pressure and temperature
13 & 14	Specify size and schedule (or wall thickness if nonstandard) of pipe line into which valve is installed.	8" SCH 40 15" OD × 0.500" wall DN200, PN100
15	Specify pipe line insulation. This information is required for predicted sound pressure level calculations.	2" thermal None
16	Specify type of valve body.	Globe (through, angle) Split body Double port Butterfly Ball Pinch
17	Specify nominal size of valve body. Specify ANSI class in accordance with ANSI B16.34-81.	4" 600 2500 SPECIAL

Instructions for Filling Out the Valve Specification Sheet Continued

Line No.	*Explanation of Terms and Definitions*	*Examples*
18	Specify maximum pressure and temperature of the valve.	2500 psig, 650°F
19	Specify manufacturer and model number.	XYZ Controls Model 719-2
20	Specify body and bonnet material.	Steel, ASTM A216, WCB
21	Specify body liner material, if any, and its inside diameter.	Polyurethane, 3.9″
22 & 23	Specify end connection. May be integral or welded onto body.	6″ RTJ Class 1500 flange Buttweld end 2″ FNPT
24	Specify flange face finish per ANSI B16.5-81 or special finish as required.	ANSI B16.5-81 Special finish: 32 RMS
25	Specify end extensions, if any. Normally, refers to sections of pipe or reducers welded to the body by the valve manufacturer.	6″ long, SCH 80, A106, GR. B
26	Specify direction of the flow through the body. FTO = flow-to-open, FTC = flow-to-close valve. NOTE: The descriptors "FTO" and "FTC" refer to the direction of fluid forces on the closure member. If immaterial, leave blank. When FTO and FTC are not applicable, specify direction as appropriate.	FTO FTC
27	Specify type of bonnet.	Standard Cooling fin Extended
28	Specify whether a lubricator and isolation valve are required. Specify lubricant.	Yes Silicone
29	Specify packing material.	Graphite impreg. asbestos TPE Non-asbestos
30	Specify type of packing.	Braided Molded V-ring Laminated filament Pressure/vacuum
31	Extra line for special body or bonnet not covered in lines 16 through 30.	Body drain Separable flanges Flangeless
32	Specify type of trim.	Single seat cage-guided Multi-stage Multi-hole Top- and bottom-guided Double seat
33	Specify nominal size and rated travel of installed trim.	2″ 50 mm
34	Specify inherent flow characteristic of installed trim.	Linear Equal % Modified parabolic Quick-opening
35	Specify whether trim is balanced or unbalanced. Semi-balanced trim should be considered as balanced.	Balanced Unbalanced
36	Specify rated C_v, F_L, and X_T of installed trim. Refer to ANSI/ISA-S75.01-1986.	260 0.9 0.68
37	Specify closure member, i.e., plug, ball, or disk material as applicable.	17-4 PH, H-1150 316
38	Specify seat material.	420 hardened 316 hardfaced
39	Specify cage, bearing, or guide material.	410 hardened
40	Specify stem material.	17-4 PH H-1150 316
41 & 42	Extra lines for additional trim requirements not covered in lines 32 through 40.	Chrome-plate Pilot-operated
43	Specify hazardous location classification per the *National Electrical Code®*, ANSI/NFPA 70-1987.	NEC® Class 1, Div. 1, Group C
44–52	Specify special requirements and/or accessories not covered elsewhere.	Solenoid valves E/P transducer NACE MR-01-75 Seismic Net weight = 275 lb
53	Specify type of actuator.	Diaphragm, pneumatic Hydr. piston, double-acting Pneumatic rotary vane
54	Specify manufacturer and model number.	XYZ Controls P-100-160
55	Specify nominal size and effective diaphragm/piston area.	8″ 160 square inch 0.2 m^2
56	Specify whether actuator is for on/off or modulating service.	Modulating On/off
57	Specify whether spring, if any, acts to open or to close valve.	Open Close None
58	Specify maximum pressure for which the actuator is designed.	100 psig 60 kPa
59	Specify minimum pressure required to fully stroke the installed valve under specified conditions.	65 psig
60 & 61	Specify limits of available air or hydraulic supply pressure. If upper limit is greater than line 58, a reducing valve (air set) should be furnished. Lower limit or reducing valve setting must be higher than pressure shown on line 59.	90 psig/70 psig

Instructions for Filling Out the Valve Specification Sheet Continued

Line No.	*Explanation of Terms and Definitions*	*Examples*
62	Specify the pressures in the actuator when valve starts travel and at its rated travel position without fluid forces acting on the valve.	8/32 psig 10/22 psig 1.2/2.1 kPa
63	Specify orientation of actuator as "VERT.UP" or "VERT.DOWN" (vertical) or "HORIZ." (horizontal). For rotary valves, also specify whether mounting is "RH" (right-hand) or "LH" (left-hand) as viewed from valve inlet, if appropriate. Specify additional information as appropriate or provide sketch.	VERT.UP HORIZ. RH LH
64	Specify type and orientation of handwheel (manual override), if any.	Top-mounted Side-mounted/LH
65	Specify if air failure valve (actuator air lock-up valve) is required and at what supply pressure it shuts.	Yes 40 psig
66	Extra line for additional actuator requirements not covered in lines 53 through 65.	Hydraulic damper Stroking speed 1"/sec. Stainless steel tubing
67	Specify input signal range for full travel.	3–15 psig 200–100 kPa 4–20 mA
68	Specify type of positioner.	None Single acting Double acting
69	Specify manufacturer and model number.	XYZ Control Co. Model AB
70	Specify whether an increasing signal increases or decreases output pressure to actuator.	Incr. Decr.
71	Specify whether air pressure gauges and positioner are required.	No Yes
72	Specify cam characteristic, if positioner has a cam. Normally linear.	Linear Square root
73	Extra line for positioner requirements.	Aluminum-free
74	Specify type and quantity of limit switches.	Mech. (lever arm) Proximity Pneumatic 2
75	Specify manufacturer and model number.	ABC Electric Co. Model A20Z
76	Specify electrical rating and number of contacts and action.	10A, 600 V AC/ DPDT
77	Specify valve travel at which switches are to actuate.	Full open/full closed
78	Extra line for additional limit switch requirements not covered in lines 74 through 77.	NEMA 4 IP 65
79	Specify manufacturer and model number of air set (pressure regulator).	RBJ Co. Model R-70
80	Specify output pressure setting.	70 psig 20 psig
81	Specify whether filter and/or output pressure gauge is required.	Yes No
82	Extra line for additional air set requirements not covered in lines 79 through 81.	Mount separate from valve
83	Specify pressure of hydrostatic test. Normally per ANSI B16.37-80 or API 6A-83.	3350 psig
84	Specify leakage class per ANSI/ FCI 70-2-76.	Class IV
85 & 86	Extra lines for additional test requirements not covered in lines 83 and 84.	Hydro for 30 minutes Helium leak test Stroking time test Dead band test

Bibliography

Adams, M., "Control Valve Dynamics," *InTech,* July 1977.

ANSI/ISA-TR75.02-2000, "Control Valve Response Measurement from Step Inputs," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 2000.

Aubrun, C., Robert, M., and Cecchin, T., "Fault Detection in a Control Loop," *Control Engineering Practice,* Vol. 3, No. 10, pp. 1441–1446, October 1995.

Arant, J. B., "Fire-Safe Valves," *InTech,* December 1981.

Arant, J. B., "Positioner Use Is Myth-Directed," *InTech,* November 1992.

Ball, K. E., "Final Elements: Final Frontier," *InTech,* November 1986.

Baumann, H. D., "A Case for Butterfly Valves in Throttling Applications," *Instruments and Control Systems,* May 1979.

Baumann, H. D., "How to Assign Pressure Drop across Control Valves for Liquid Pumping Services," *Proceedings of the 29th Symposium on Instrumentation for the Process Industries,* Texas A&M University, 1974.

Baumann, H. D., "Effect of Pipe Reducers on Valve Capacity," *Instruments and Control Systems,* December 1968.

Baumann, H. D., "The Introduction of a Critical Flow Factor for Valve Sizing," *ISA Transactions,* Vol. 2, No. 2, April 1963.

Baumann, H. D., "Control-Valve Sizing Improved," *InTech,* 46(6): pp. 54–57, June 1999.

Beard, C. S., *Final Control Elements,* Philadelphia, PA: Chilton, 1969.

Biegacki, S. and VanGompel, D., "The Application of DeviceNet in Process Control," *ISA Transactions,* Vol. 35, No. 2, pp. 169–176, 1996.

Bishop, T., Chapeaux, M., Jaffer, L., et al., "Ease Control Valve Selection," *Chem. Eng. Prog. 98,* (11): 52–56 November 2002.

Boger, H. W., "Recent Trends in Sizing Control Valves," 23rd Annual Symposium on Instrumentation, Texas A&M University, 1968.

Boger, H. W., "Flow Characteristics for Control Valve Installations," *ISA Journal,* November 1966.

Bolek, W., Sasiadek, J., and Wisniewski, T., "Two-Valve Control of a Large Steam Turbine," *Control Engineering Practice,* Vol. 10, No. 4, pp. 365–377, April 2002.

Borden, G. and Zinck, L., "Control Valve Seat Leakage," *InTech,* November 1986.

Brunone, B. and Morelli, L., "Automatic Control Valve-Induced Transients in Operative Pipe System," *Journal of Hydraulic Engineering,* Vol. 125, No. 5, p. 534, May 1999.

Buckley, P. S., "Design of Pneumatic Flow Controls," *Proceedings of the 31st Annual Symposium on Instrumentation for the Process Industries,* Texas A&M University, 1976.

Buckley, P. S., "A Control Engineer Looks at Control Valves," *Proceedings of 1st ISA Final Control Elements Symposium,* Wilmington, DE, 1970.

Buresh, J. F. and Schuder, C. B., "The Development of a Universal Gas Sizing Equation for Control Valves," *ISA Transactions,* Vol. 3, No. 4, October 1964.

Cain, F. M., "Solving the Problem of Cavitation in Control Valves," *Advances in Instrumentation and Control—Proceedings of the ISA/91 International Conference,* Vol. 46, Part 2, pp. 1133–49, Paper #91-0462, 1991.

Cain, F. M. and Barnes, R. W., "Testing for Cavitation in Low Pressure Recovery Control Valves," *ISA Transactions,* Vol. 25, No. 2, pp. 61–67, 1986.

"Control Valves — Globe, Plug, Pinch, Needle, Gate," *Measurements and Control,* February 1993.

"Control Valves, Regulators," *Measurements and Control,* June 1992.

"Control Valve Market to Reach nearly $3.5 Billion by 2008," *Control Eng.,* April 2003.

Coughlin, J. L., "Control Valves and Pumps: Partners in Control," *Instruments and Control Systems,* January 1983.

"Control Valves Deserve Your Respect," *Control Engineering,* Vol. 49, No. 4, p. 64, April 2002.

Cunningham, E. R., "Solutions to Valve Operating Problems," *Plant Engineering,* September 4, 1980.

"Digital Hydraulic Valves Gaining Momentum," *Design News,* Vol. 58, March 3, 2003.

Dobrowolski, M., "Guide to Selecting Rotary Control Valves," *InTech,* December 1981.

Driskell, L. R., "Sizing Control Valves," *ISA Handbook of Control Valves*, ISA, 1976.

Fagerlund, A. C., "Recommended Maximum Valve Noise Levels," *InTech,* November 1986.

Gassmann, G. W., "When to Use a Control Valve Positioner," *Control,* September 1989.

George, J. A., "Sizing and Selection of Low Flow Control Valves," *InTech,* November 1989.

Gibson, H. I., "Variable-Speed Drives as Flow Control Elements," *ISA Transactions,* Vol. 33, No. 2, pp. 165–169, July 1994.

"Growth Forecasted for the Control Valve Market," *Control Solutions* 74 (12):8, December 2001.

Hammitt, D., "Key Points about Rotary Valves for Throttling Control," *Instruments and Control System,* July 1977.

Hammitt, D., "How to Select a Valve Actuator," *Instruments and Control Systems,* February 1977.

Hanssen, A. J., "Accurate Valve Sizing for Flashing Liquids," *Control Engineering,* February 1961.

Hanson, C. L. and Clark, J. C., "Fast-Closing Vacuum Valve for High-Current Particle Accelerators," *Review of Scientific Instruments,* January 1981.

Hegberg, M. C., "Control Valve Selection for Hydronic Systems," *ASHRAE J.,* 42 (11): 33, November 2000.

Hegberg, M. C., "Control Valve Selection Response," *ASHRAE J.,* 43(3): 24, March 2001.

Hill, A. G. and Lau, K. H., "Artificial Intelligence in Control Valve Selection," *ISA Transactions,* Vol. 28, No. 1, pp. 37–44, 1989.

Horch, A., "A Simple Method for Detection of Stiction in Control Valves," *Control Engineering Practice,* October 1999, pp. 1221–1231.

Hutchison, J. W. (ed.), *ISA Handbook of Control Valves,* 2nd ed., Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1976.

"Installing Smart Positioners–A Wise Move," *Chemical Engineering Magazine,* December 1, 2000.

"Intelligent Actuator Allows Programmable Valve Control," *Water World,* Vol. 15, No. 5, p. 101, May 1999.

Jämsä-Jounela, S.-L., Dietrich, M., Halmevaare, K., and Tiili, O., "Control of Pulp Levels in Flotation Cells," *Control Engineering Practice,* Vol. ll, No. 1, pp. 73–81, January 2003.

Jury, F. D., "Positioners and Boosters," *Instruments and Control Systems,* October 1977.

Kam, W. Ng., "Control Valve Noise," *ISA Transactions,* Volume 33, No. 3, pp. 275–286, September 1994.

Karpenko, M., Sepehri, N., and Scuse, D., "Diagnosis of Process Valve Actuator Faults Using a Multilayer Neural Network," *Control Engineering Practice,* Vol. 11, No. 11, pp. 1289–1299, November 2003.

Kayihan, A. and Doyle, F. J., III, "Friction Compensation for a Process Control Valve," *Control Engineering Practice,* Vol. 8, No. 7, pp. 799–812, July 2000.

Keagle, J., "It's 18 mA, Do You Know Where Your Positioner Is?" *InTech,* May 1992.

Keles, O. and Ercan, Y., "Theoretical and Experimental Investigation of a Pulse-Width Modulated Digital Hydraulic Position Control System," *Control Engineering Practice,* Vol. 10, No. 6, pp. 645–654, June 2002.

Keskar, P. Y., "Analysis of Lightning-Related Damages to Instrumentation and Control System for Water and Wastewater Plants," *ISA Transactions,* Vol. 35, No. 1, pp. 9–15, May 1996.

Kimura, T., Hara, S., Fujita, T., and Kagawa, T., "Feedback Linearization for Pneumatic Actuator Systems with Static Friction," *Control Engineering Practice,* Vol. 5, No. 10, pp. 1385–1394, October 1997.

Kirsner, W., "Control Valve Selection," *ASHRAE J.,* 43(3): 24, March 2001.

Lee, J. G. and Kim, O. H., "Development of a New Hydraulic Servo Cylinder with Mechanical Feedback," *Control Engineering Practice,* Vol. 7, No. 3, pp. 327–334, March 1999.

Lipták, B. G., "Control Valves in Optimized Systems," *Chemical Engineering,* September 5, 1983.

Lipták, B. G., "Control Valves for Slurry and Viscous Services," *Chemical Engineering,* April 13, 1964.

Lipták, B. G., "How to Size Control Valves for High Viscosities," *Chemical Engineering,* December 24, 1962.

Lipták, B. G., "Valve Sizing for Flashing Liquids," *ISA Journal,* January 1963.

Louleh, Z., Cabassud, M., and Le Lann, M.-V., "A New Strategy for Temperature Control of Batch Reactors: Experimental Application," *Chemical Engineering Journal,* Vol. 75, No. 1, pp. 11–20, August 1999.

Monsen, J. F., "Spreadsheet Sizes Control Valves for Liquids/Gas Mixtures," *InTech,* December 1990.

Moore, R. L., "Flow Characteristics of Valves," in *ISA Handbook of Control Valves,* 2nd ed., Pittsburgh, PA: Instrumentation, Systems, and Automation Society, 1976.

Morgenroth, J., "Quarter-Turn Plug, Ball, and Butterfly Valves," *Plant Engineering,* July 24, 1980.

"1992–94 Catalog of U.S. & Canadian Valves & Actuators," Value Manufacturers Association of America.

O'Keefe, W., "Learn Fluid-Handling Lessons from Nuclear Isolation Valves and Actuator Systems," *Power,* January 1981.

Page, G. W., "Predict Control Valve Noise," *Chem. Eng., New York,* 107 (9): 23–26, August 2000.

Pham, D. T., Jennings, N. R., and Ross, I., "Intelligent Visual Inspection of Valve-Stem Seals," *Control Engineering Practice,* Vol. 3, No. 9, pp. 1237–1245, September 1995.

Price, V. E., "Smart Valve Intelligence Takes Many Forms," *InTech,* August 1992.

Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 ISA Conference, Houston, TX, October 1992.

Rahmeyer, W., "Cavitation Testing of Control Valves," Instrument Society of America, paper no. C.I. 83-R931, presented at the International Conference in Houston, TX, October 1983.

Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

"Recommended Voluntary Standard Formulas for Sizing Control Valves," Fluid Controls Institute, Inc., FCI 62-1, May 1962.

"Revised Control Valve Standard," *Hydrocarb. Process.,* 82(9): 114, September 2003.

Restrepo, A., González, A., and Orduz, S., "Cost-Effective Control Strategy for Small Applications and Pilot Plants: On/Off valves with Temporized PID Controller," *Chemical Engineering Journal,* Vol. 89, No. 1–3, pp. 101–107, October 2002.

Riveland, M. L., "The Industrial Detection and Evaluation of Control Valve Cavitation," Instrument Society of America, paper no. C.I. 82-909, presented at International Conference in Philadelphia, PA, October 1982.

Roth, K. W. and Stares, J. A., "Avoid Control Valve Application Problems with Physics-Based Models: Kinetic Energy Criteria have Many Limitations," *Hydrocarb. Process.,* 80(8):37, August 2001.

Ruel, M., "Control Valve Health Certificate," *Chem. Eng., New York,* 108(12):62–65, November 2001.

Sanderson, R. C., "Elastomer Coatings: Hope for Cavitation Resistance," *In Tech,* April 1983.

Scott, A.B., "Control Valve Actuators: Types and Application," *InTech,* January 1988.

Sharif, M.A. and Grosvenor, R.I., "The Development of Novel Control Valve Diagnostic Software Based on the Visual Basic Programming Language," *P. I. Mech. Eng. I.-J. Sys.,* 214(12):99–127, 2000.

Shinskey, F. G., "Control Valves and Motors," Foxboro Publication No. 413–8.

Singleton, E. W., "Control Valve Sizing for Liquid Viscous Flow," Engineering Report No 7, Introl Limited.

Spreter, R.,, "Valve Application Benefit from Technology Upgrades," *Power Engineering,* Vol. 103, No. 9, p. 46, September 1999.

"Standard Control Valve Sizing Equations," ANSI/ISA-S75.01-1977.

Thornhill, N. F., Cox, J. W., and Paulonis, M. A., "Diagnosis of Plantwide Oscillation through Data-Driven Analysis and Process Understanding," *Control Engineering Practice,* Vol. 11, No. 12, pp. 1481–1490, December 2003.

Tullis, J. P., *Hydraulics of Pipelines: Pumps, Valves, Cavitation, Transients,* New York: John Wiley and Sons, 1989.

"Valve Actuator Roundup," *InTech,* January 1988.

Wendy, E. J. and Stanton, C. C., "Industry Corner: A World View of Industrial Valves," *Business Economics,* Vol. 32, No. 2, p. 56, April 1997.

Weir, W., "Control Valve Market Analyzed in New Study," *Hydrocarbon Processing,* Vol. 75, No. 9, p. 27, September 1996.

Weirauch, W., "Electric Actuators 'Leading the Way' in Device Network Use," *Hydrocarbon Processing,* Vol. 75, No. 10, p. 31, October 1996.

Weirauch, W., "Valve Automatization Gains Renewed Interest," *Hydrocarbon Processing,* Vol. 80, No. 7, p. 25, July 2001.

"What's New in Valves and Valve Operators," *Pipe Line & Gas Journal,* Vol. 227, No. 11, p. 43, November 2000.

"What's New in Valves and Valve Operators," *Pipelines & Gas Journal,* Vol. 229, No. 11, p. 61, November 2002.

Wilton, S. R., "Control Valves and Process Variability," *ISA Transactions,* Vol. 39, No. 2, pp. 265–271, April 2000.

Wolter, D. G., "Control Valve Selection," *InTech,* October 1977.

Yang, J. C. and Clarke, D. W., "The Self-Validating Actuator," *Control Engineering Practice,* Vol. 7, No. 2, pp. 249–260, February 1999.

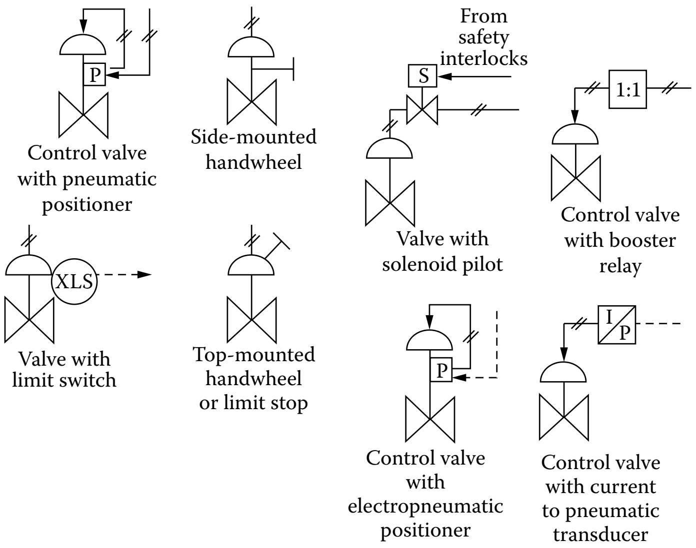
Yu, W. X., Lin, J., Wang, C.W., et al., "A Thermo-Control Valve Fabricated from Shape-Memory Alloy for Use in the Oil Field," *Rare Metal Mat. Eng.,* 31(5): 393–396, October 2002.

Yu, F. C., "Easy way to estimate realistic control valve pressure drops," *Hydrocarb. Process.,* 79(8): 45–48, August 2000.

6.2 Accessories and Positioners

H. D. BAUMANN (1970) **C. G. LANGFORD** (1985, 2005)

B. G. LIPTÁK (1995)



Flow sheet symbols

Types:	Air accumulators, air sets, handwheels, I/P transducers, limit stops, limit switches, positioners, relays (biasing, booster, lock-up, quick-exhaust, reversing), smart (microprocessor-based) valve electronics, solenoid pilots, stem position transmitters, servo solenoid valve.
Materials of Construction:	Bodies and cases: die-cast zinc, "white metal," aluminum. Bellows: copper alloys. Diaphragms: elastomer-coated fabric, thin metal. Misc. parts: steel, brass, aluminum.
Supply Pressure (Gauge):	Varies from 20 to 25 PSIG (140 to 170 kPa) for signal relays to 60 to 100 PSIG (400 to 700 kPa) for positioners or boosters, some to 150 PSIG (1000 kPa).
Inaccuracy:	Positioners: repeatable to ± 0.1 to 1% and accurate to ± 0.5 to 2% of span. Transducers: ±0.5 to 1% of span. Boosters: ±0.1 to 1% of span.
Signal Ranges (Gauge Pressure):	Pneumatic: 3–9, 3–15 (preferred), 6–30, and 9–15 PSIG (20–60, 20–100, 40–200, and 60–100 kPa). Electronic: Digital Fieldbus and other, 1–5, 4–20 (preferred), and 10–50 mA DC. Others are used such as 0–10 and 1–10 VDC.
Costs:	Air sets, \$40; handwheels, \$500 for larger size; I/P transducers, \$300; limit stops, \$100; limit switches, \$50 to \$150; positioners, \$400 (pneumatic), \$600 (electronic); \$1500 (special) solenoid pilots, \$80; stem position transmitter, \$250 to \$500.
Partial List of Suppliers:	*Air Sets:* Adams Valve Inc. (www.adamsvalves-usa.com/page.cfm?name=welcome) ControlAir Inc. (www.controlair.com/news/index.htm) Cashco Inc. (www.cashco.com/) Fisher Controls (www.emersonprocess.com/fisher) ITT Conoflow (www.conoflow.com/) Masoneilan Dresser (www.masoneilan.com/internet/businessunits/measurement/subunits/masoneilan/index.cfm) Moore Products Co. (www.sea.siemens.com/default.asp)

Handwheels and Limit Stops:
Anchor/Darling Valve Co. (see Flowserve)
Auma Actuators Inc. (www.auma-usa.com/)
Copes-Vulcan Inc. (www.dezurik.com/)
Daniel Valve Co. (www. danielvalve.com/)
Duriron Co. (www.flowserve.com/valves/index.htm)
Fisher Controls (www.emersonprocess.com/fisher/)
Grinnell Supply Sales Co. (www. grinnell.com/)
Keystone Controls Inc. (www.tycovalves.com/)
Leslie Controls Inc. (www.lesliecontrols.com/)
Limitorque Corp. (www.flowserve.com/valves/index.htm)
Pacific Valves (www.cranevalves.com/pv.htm)
Mastergear Div. (www.regal-beloit.com/)
Neles-Jamesbury Inc. (www.metsoautomation.com/)
PBM Inc. (www.pbmvalve.com/)
Rotork Controls Inc. (www.tycovalves.com/)
Velan Valve Corp. (www.velan.com/)
Valtek (www.flowserve.com/valves/index.htm)
Xomox (www.xomox.com/Xomox/htdocs/)
I/P Transducers, Positioners, Stem Position Transmitters:
ABB Kent Inc. (www.abb.com/)
Bailey Controls Co. (see ABB)
Bray Valve and Controls (www.bray.com/)
DeZurik (www.dezurik.com/)
Fisher Controls International Inc. (www.emersonprocess.com/fisher/)
Foxboro Co. (www.foxboro.com/)
Honeywell Industrial Controls (www.honeywell.com/)
ITT Conoflow (www.conoflow.com/)
Jordan Controls Inc. (www.jordancontrols.com/home_default.asp)
Kammer Valves Inc. (www.flowserve.com/valves/index.htm)
Leslie Controls Inc. (www.lesliecontrols.com/)
Masoneilan Dresser, (www.masoneilan.com/internet/businessunits/measurement/subunits/masoneilan/index.cfm)
Moore Products Co. (www.sea.siemens.com/default.asp)
Neles (METSO) Controls (www.metsoautomation.com/)
Valtek Inc. (www. flowserve.com/valves/index.htm)
Limit Switches:
Allen Bradley (www.ab.com/)
Go Switch Inc. (Topworx) (www.topworx.com/)
Masoneilan, Dresser Valve & Controls Div. (www.masoneilan.com/internet/businessunits/measurement/subunits/masoneilan/index.cfm)
Micro Switch/Honeywell (content.honeywell.com/sensing/)
Pepperl + Fuchs Inc. (www. pepperl-fuchs.com/pa/welcome_e.html;)
Proximity Controls Inc. (www.proximitycontrols.com/products.htm)
Relays (Biasing, Booster, Reversing):
ABB Kent-Taylor, http://www.abb.com/
Bailey Controls Co. (www.abb.com/)
Fairchild Industrial Products Co. (flw.com/fairchild/)
Fisher Controls International (www.emersonprocess.com/fisher/)
Foxboro Co. (www.foxboro.com/)
Moore Products Co. (www.sea.siemens.com/default.asp)
Robertshaw Controls Co. (www.robertshawindustrial.com/)
Smart (Microprocessor-Based) Valve Electronics:
EIM Controls (www.eim-co.com/eimweb.nsf/pages/company)
Kaye & MacDonald (www.kayemacdonald.com)
Limitorque (www.flowserve. com/valves/index.htm)
Rotork Controls (www.tycovalves.com/)
Valtek (www.flowserve.com/valves/index.htm)
Solenoid Pilot:
Automatic Switch Co. (www.asco.com/)

Crosby Valve and Gauge Co. (www.tycovalves.com/)
Gilmore Valve Co. (www.gilmorevalve.com/products)
Keystone International Inc. (www.tycovalves.com/)
Leslie Controls Inc. (www.lesliecontrols.com/)
Parker Hannifin Corp. (Skinner) (www.parker.com/)
Richards Industries—Valve Group Inc. (www.richardsind.com/)
Spirax Sarco Inc. (www.spiraxsarco.com/us/)
Tom Wheatley Valve Co. (www.masoneilan.com/internet/businessunits/measurement/subunits/masoneilan/index.cfm*)*
Servo Solenoid Valve:
Target Rock Valve (http://www.cwfc.com/enertech/F13TRHome.htm)
Hydraulic Relays(Water):
GA: Industries (www.gaindustries.com)

INTRODUCTION

This section describes some of the traditional valve accessories, including positioners. For the valve accessories that provide self-diagnostic capability or can communicate on the fieldbuses, the reader is referred to Section 6.12, which is fully devoted to an in-depth discussion of intelligent valves, positioners, and accessories.

The word "accessories" applies to the many devices added to control valves. Many of these are more necessities than accessories, depending on the application. This section lists and discusses some of the many accessory devices that can be attached to both throttling and on/off valves. The purpose is to improve their performance or to obtain remote feedback on their status. The positioner is the single most important valve accessory, and this section will emphasize the design and the application of positioners. Before starting that discussion, a brief summary will consider the contribution of microprocessors to the development of smart valve accessories. See also Section 6.9, Dynamic Performance of Control Valves, for a more detailed look at valve dynamics.

SMART VALVES

Section 6.3 discusses some of the more common applications of intelligent actuator systems, and Section 6.12 is fully devoted to an in-depth discussion of intelligent valves, positioners, and accessories.

The microprocessor can provide logic to improve the positioning and the performance of the valve. Additional services may include operational and maintenance communications with the valve.

Microprocessor-based systems are available in watertight or explosionproof (NEMA 6 or 7) construction and can tolerate ambient temperatures from –40 to 185°F (–40 to 85°C). They can also be used where any or all of high humidity, fungus, or dust are present. As shown in Figure 6.2a, these devices have become quite complex, and any proposed application for "smart" positioners should review the operating conditions and the application requirements with the manufacturer. Some of the positioners emphasize maintenance services; some make improved control their main goal.

Microprocessor-based systems can incorporate self-tuning and self-calibrating electronic positioners or electronic proportional, integral, derivative (PID) controllers, which respond to a digital or analog external set point and commands. Locating the controller logic at the valve and dedicating it to full-time control is an advantage on critical loops, providing fast response for applications such as compressor surge protection, where speed of response is a major consideration.

There may also be an improvement in control robustness and reliability if the system can continue to control without the central control system. This does not mean that positive control is lost, because the system is still "under orders" from the central system, only the continuous cycle of measure-compute-act-feedback is local. Commands for set point, mode of operation, and data collection reside in the central system. Local control eliminates the time lost during communications and response of the central control system like a programmable logic controller (PLC) or distributed control system (DCS).

Smart positioners can modify the characterization of the valve to reduce control nonlinearities in the valve and actuator and to provide special action. It is possible to match the valve to the process requirements to reduce the change in process gain over the range of operation or to provide special response or action as required.

Smart actuator circuitry can also protect electric motor operators from electrical overload when the valve is jammed or from reverse phase operation. Smart systems can also take advantage of the lower cost of digital communications over a single loop of two-conductor or fiber-optic cable. This method of communications can substantially reduce the wiring cost of installations that can control a number of actuators, pumps, and other devices.

The smart valves can have calibration changed and reconfigured and can detect such performance changes as pressure drop changes resulting from a fouled pump or pipeline. They can also limit the valve travel to stay within the range where the required characteristics (gain) are guaranteed or to limit maximum or minimum flows.

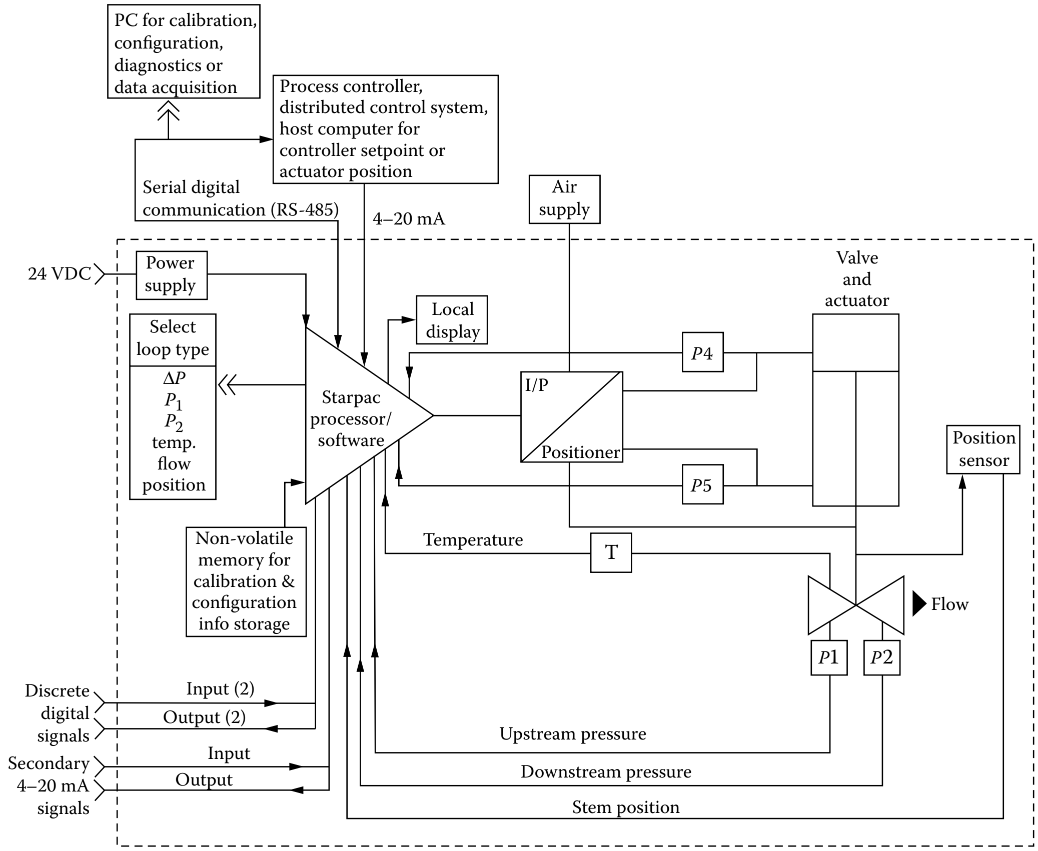


FIG. 6.2a

Smart valve packages may have local display and sensors for temperature, flow, pressures, pressure differential, and stem position. (Courtesy of Valtek, Flowserve Corp.)

Some of the smart valve packages incorporate additional sensors installed on or in the valve. These measurements may include upstream and downstream process pressure, pressure difference, flow, process fluid temperature, stem position, and actuator operating pressures (Figure 6.2a). This allows both for total distributed control and for the flexibility of manipulating the valve to control any one of the variables.

These valves may also include local indicators and the capability of calibration or reconfiguration of the valve, transmitters, or controller. Diagnostic capabilities include self-tuning and the ability to evaluate the valve and process responses to step or to ramp changes in the valve position. Upon failure of the electrical power supply the valve functions can be supported by battery backup.

The user is cautioned that some of these more sophisticated accessories may require more than the usual two-wire 4–20 mA DC power/signal wiring. Some are "four-wire" with separate power required. Some are two-wire, but may present a higher impedance load to the system than the simpler field instruments. This higher source voltage requirement may prevent wiring two positioners in series for split ranging.

Primarily used with on/off valves, on-line partial stroke testing may be provided for safety applications to provide verification of the valve actuation system. The user is advised to check the current versions of applicable standards because there is significant interest and activity in this area.

POSITIONERS

The positioner along with the associated actuator constitutes a simple position control loop. The valve begins to act only after the error is detected and the controller output has changed. Inherently, this is a lagging action.

With the typical air supply pressure used, 15–100 PSIG (100–700 kPa), any large change in valve signal will result in a critical pressure drop within the positioner. When the internal flow restriction has a pressure drop (ΔP) greater than 50% of the absolute inlet pressure, then sonic flow velocities will exist. This sets a constant and limiting flow rate through the pilot valve.

This maximum flow rate capacity is probably the one listed in the product specifications. As the valve stem approaches the desired position, the ΔP decreases and the flow decreases to be more nearly proportional to the signal error. This results in some slowing of the response at the end of the stroke change and may help to stabilize response.

When to Use Positioners

The purpose of a positioner is to improve the accuracy of control valve response. This means that the valve position will more closely approach the position commanded by the control system. A positioner can reduce the effects of many dynamic variations. These include changes in packing friction due to dirt, corrosion, lubrication, or lack of lubrication; variations in the dynamic forces of the process; sloppy linkages (causing dead band); and nonlinearities in the valve actuator.

The dead band of a good valve/actuator is 2% (Section 6.4), but it has been measured at up to 5%. Large plug valves and ball valves with less than perfect linkages and inadequate actuators may be far worse. A better positioner with the proper actuator can often have a dead band of less than 0.5% of stroke.

The positioner increases the actuator speed or thrust by increasing the actuator pressure or airflow volume, and can modify the valve characteristics through the use of mechanical links and cams or electronic function generators. While these positioner capabilities are very important, some of these capabilities can also be obtained or approximated with other accessories.

For example, split-ranging is possible using pneumatic relays. Multiplier/biasing relays in the air signal line to the valve modify the relationship between controller output and actuator air pressure. The response speed/thrust of the valve can be increased with the use of airflow booster relays, and changes to the control valve characteristics can be obtained, not only with a different plug characteristic shape, but also by pneumatic or electronic characterizing of the controller signal. Consider that the computing relays may have a lower flow capacity than a positioner and the speed of response may be affected.

When the valve is in remote manual (open-loop) operation, the positioner will reduce the effects of the valve hysteresis and dead band and improve the accuracy of the response. When the valve is under automatic (closed-loop) control, the positioner will normally improve control, but will do that only if the loop response is slow when it is compared to the control valve response (analysis, temperature, liquid level and blending, slow flow, large volume gas flow).

An imperfect positioner may degrade loop response, contribute to proportional offsets, and cause limit cycling in fast loops where the valve cannot keep up (fast flow, liquid pressure, small volume gas pressure). Older and simpler positioners did limit response and create problems.

The issue is more of the speed of the valve relative to the controlled process rather than absolute speed (Figure 6.2b). It is not uncommon for the valve to be the slowest part of the loop and limit process response. This may be a problem, or it may be perfectly acceptable. The needs of the process should be the basis for determining the performance requirements of the control system. The very worst situation occurs when the valve system and the process have equal or similar time constants.

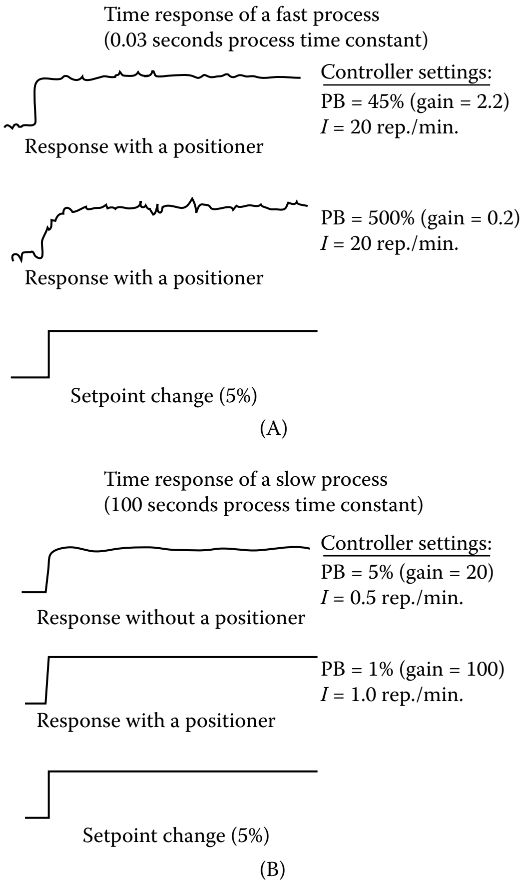


FIG. 6.2b
The response of a "fast" process (A, on top) is better without a positioner, while the response of "slow" processes (B, on bottom) is better with a positioner.[2]

Actuators without springs or equivalent padding pressure to provide the spring return function will usually require positioners with dual outputs to drive both sides of the diaphragm or piston. The spool valve or relay valve will have a second output, with its action the reverse of the first one. Because the opposing air pressure below the diaphragm decreases when the air pressure above the diaphragm increases, the actuator has a greater net force available without the opposing spring force.

It is necessary to have a packing gland or O-ring seal around the valve stem for the lower air pressure. A positioner with only a single output can be used with a reversing relay to provide the second output.

A number of different control design requirements can be accomplished with positioners. A reverse-acting positioner (increase in input causes a decrease in output) makes an air-to-open, spring-to-close valve function as an air-to-close, spring-to-close valve.

This combination may be specified for systems with interlocks or batch operations. The use of reverse acting positioners is discouraged by some users for maintenance reasons. Beware of terminology here: "Fail Open" has been

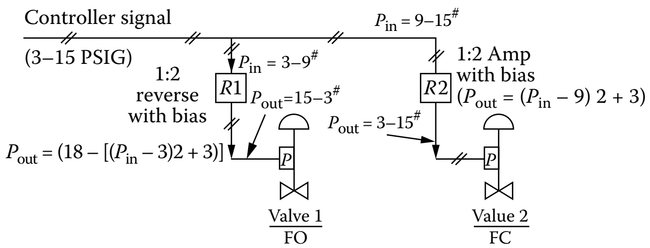


FIG. 6.2c

Split-range operation of two valves can be obtained by the use of amplifying and reversing relays. They can be either be pneumatic (shown here) or electronic.

confused with "Flow to Open," and "Fail Closed" has been confused with "Flow to Close." These are different and must be defined and understood especially in safety applications.

Heating, ventilation, and air-conditioning (HVAC) systems with low-pressure pneumatic thermostats may use some sort of simple positioner to increase actuator pressures. The alternative is electric signals and electric motor-activated valves. These services rarely need fast response or tight control.

Split-Range Operation Split-ranging is the use of two valves controlled from one controller output signal. With positioners, each positioner moves its valve over only a part of the controller output range.

One typical application is for temperature control. Here, the cooling valve is full open at 0% controller signal and closed at 50% signal and beyond, while the heating valve is closed over the 0–50% range and begins to open at 50% controller signal and fully opens at 100%. See Figure 6.2c and Table 6.2d.

The example is a simple one and has many possible difficulties. The specific application may require a different sort of calibration. A dead zone with both valves closed at a controller output of 50% will have no total flow over that dead band. This may be used to reduce costs of heating and cooling.

Other applications may require that both valves be partly opened at the 50% controller output point. For a system where the valves add the temperature control fluid into a circulating system, it might work well. But, as an example, if the system is "once through" where the heat transfer rate is proportional to the heating/cooling fluid velocity, the valves are usually set up with some overlap in valve position. That is, the cooling flow might not shut off until the Controller Output (CO) is 75%, and the heating flow might start at 25% CO. The total flow and thus heat transfer rate will remain more nearly constant and stabilize the operation.

With this approach, maintenance is simplified because standard positioner calibrations are used and the full positioner accuracy is retained. Note that the specifications for the positioner are defined for the full range of the positioner and that if only half the output range is used, the effective dead band for each valve is doubled. For a maximum in flexibility, control each valve from a separate control system output.

TABLE 6.2d

As it is Shown in Figure 6.2c, it is Possible to Use Relays instead of Split-Range Positioners

Signals in %

Controller Output	*Output Relay 1*	*Output Relay 2*	*Control Valve 1*	*Control Valve 2*
0	100	0	100	0
25	50	0	50	0
50	0	0	0	0
75	0	50	0	50
100	0	100	0	100

Signals in PSIG (bar)

Output Controller	*Output Relay 1*	*Output Relay 2*	*Control Valve 1*	*Control Valve 2*
3 (0.2)	15 (1.0)	3 (0.2)	100%	0%
9 (0.6)	3 (0.2)	3 (0.2)	0%	0%
15 (1.0)	3 (0.2)	15 (1.0)	0%	100%

Signals in % (with overlapping flows)

Controller Output	*Output Relay 1*	*Output Relay 2*	*Control Valve 1*	*Control Valve 2*
0	100	0	100	0
25	75	25	75	25
50	50	50	50	50
75	25	75	25	75
100	0	100	0	100

Tight Shut-Off In some applications, there is a need for special tight shut-off. A positioner calibration can provide the full actuator pressure applied to the valve stem when the valve is commanded to be closed.

High Rangeability A special application might require accurate flow control for a small percentage valve opening. The positioner calibration can provide that the valve moves very little for a large fraction of the low end of the signal range but still opens fully at 100% of signal. Examples include a drastic difference between operating conditions and start-up or emergency shutdown, or a large butterfly valve.

An application as simple as a level control may be very poor because the valve is oversized for special operating problems. A very simple flow control loop will add considerable accuracy at a modest cost. The primary requirement for the flow sensor is to be reliable; even poor accuracy is adequate. A simple field controller, even pneumatic, will constitute an inexpensive secondary control loop requiring little attention.

Gap Action If there is a great problem with packing wear, it is possible to set up the controller or the positioner for "gap action." Here, the valve signal changes only when the control error exceeds a certain limit.

Condensate pots that accumulate steam condensate are an example of this. A partly opened valve will experience damaging cavitation. A valve that quickly switches between fully open at a high condensate level and fully closed at a low level will avoid cavitation. Boiler blowdown valves are a similar application.

When Not to Use a Positioner

The positioner/actuator can be defined as the secondary control loop in the cascade loop, working with the process controller output providing the control signal. In order for a cascade secondary to improve control, it must respond more quickly than the primary loop. The ideal situation would be if the time constant of the secondary was one tenth (open-loop speed of response ten times as fast) of that of the primary and if the period of oscillation of the secondary loop was three times that of the primary.

Even if the valve response is not as fast as these ideals, there still may be very helpful isolation of the process from process fluid changes. No process control response time can be faster than the slowest element in the control loop. The advantages of a cascade system are achieved with good secondary response.

A general principle is to use a cascade system when the advantages of isolating the primary control loop from the nonlinearity, and variations in the secondary system, justify the added cost and complexity. In such installations (B in Figure 6.2b shows a gas pressure control process), the addition of a positioner increases the open-loop (steady state) gain and improves the loop response. For this situation, the process response may be further improved because the primary controller can be tuned with increased gain (narrower proportional band) and increased integral action (reset) to speed up the integral action.

It is clear that poor valve response reduces the quality of control much more than the poor valve response would imply. In this situation the controller tuning must be modified (gain reduced, integral lengthened) to avoid oscillation due to dead band and response delay. In typical real applications controller tuning is conservative and avoids any hint of oscillation, and the response is even worse than it might be. Operators usually object to cycling processes.

Part A in Figure 6.2b illustrates a fast liquid flow process, where the loop without a positioner can be tuned more tightly (for higher gain and more repeats/minute); such a loop responds better without a poor positioner. It might also be noted that after a new state is reached, the positioned installation gives better and noisier control because of increased speed of response.

Some argue that all loops can be (and should be) controlled with positioned valves "because they provide improved response." This is true, but there are situations where the positioner will add unjustified cost. The very fast flow loop of Figure 6.2b is one example.

Another is a small valve with a relatively large actuator and good available force; it has a small valve stem and little friction and a small plug with low fluid forces. Situations where cost is primary and quality of control is not highly important may also omit positioners. Examples are HVAC where forces are low and signal changes are slow, but there is a desire to control the cost. Positioners are not normally required for on/off service.

Positioner Performance

The valve positioner is a servo-amplifier acting with the valve actuator to control the position of the valve stem. Without a positioner, the stem position may be changed or motion restrained by varying fluid pressures on the plug and by unpredictable friction forces. See Section 6.9 for more details on control valve response.

The pneumatic input signal is typically 3–15 PSIG (20–100 kPa), and the typical stroke is between $^1/_2$ and 3 in. (12–75 mm). Errors of 0.05 in. (0.1 mm) are typical. The usual input air signal is the output of a pneumatic controller in the field or from a control panel or control system. The controller signal is the result of comparing a set point pressure to a measured variable as presented by a 3–15 PSIG value and the output is intended to reduce the error.

Most positioners use air both as the operating fluid and as the source of power. Hydraulic actuator/positioners use high-pressure oil and are used on applications requiring large valves for high differential pressure services. These are units requiring actuator pressures higher than the normally available 60–100 PSIG (0.4–0.7 MPa) instrument air supply pressure. Hydraulic actuator/positioners are considerably more expensive; and they may be acoustically noisy and require specialized maintenance.

Electric motor operators are also available (Section 6.3). They are used where air is not available and where their typically slow operating speed is acceptable. There are some specialized electric valve actuators that can provide very high position precision; others can provide very fast response.

The analog electropneumatic positioner input section responds to a standard milliampere signal. Such units usually consist of a pneumatic positioner plus an integral I/P. Some users prefer to use a pneumatic positioner with a separate I/P (current to pressure converter), usually located near the valve. A few systems provide for packaging all of these accessories into one unit.

An all-electric server solenoid motor-type positioner/actuator valve is also available. At least one control valve design uses a servo solenoid valve (SV) to achieve very fast and accurate response using the process fluid pressure for actuation. This approach requires a relatively clean fluid.

The positioner normally provides a substantial improvement in valve and control loop performance, with the greatest improvement realized on slow control loops with low controller gains typically level, temperature, or analytical control (Figure 6.26).

Positioners have typical open-loop gain (change in output pressure per change in input signal, with the valve stem locked in position) of 10:1 up to 200:1. Dead band (the minimum input change for a detectable output change) is claimed as 0.1–0.5% of span. Vendors claim positioning accuracy of 0.002–0.005 in. (0.05–0.13 mm) under bench test conditions.

Positioner Designs

The control valve positioner as used with a globe-style valve is typically mounted on the valve yoke and has mechanical linkage, which is connected to the valve stem to sense position. The pneumatic positioner is powered by compressed air at pressures between 25 and 150 PSIG (170 and 1000 KPa).

Tubing conveys the positioner output air to and from the actuator. There are some standards defining standardized mounting details for positioners, solenoid valves, and other accessories.

Other designs mount the positioner directly in line with the valve stem, and installed on the end of the actuator. In this design the linkage to the valve stem is direct in a straight line. In this scheme, the positioner provides air directly to the top of the actuator. Figure 6.2e illustrates such a force-balance positioner design.

For rotary valves, the positioner is mounted above the valve body. In every case, the mountings must be solid and not allow any relative motion, which adds directly to the dead band. The installation should be so arranged that there is little temptation for workers to damage the connecting tubing and linkages by climbing on the valve. The exterior of the positioner should have a finish compatible with the chemical environment.

The interior is (mostly) protected from corrosion by the steady bleed of dry clean air from the pilot valve. If the valve will be exposed to a very high level of vibration or acoustic noise, the purchase specifications should note this. Some positioner designs will withstand vibration better than others.

Pipeline applications may use the pipeline gas to operate the pneumatic instruments. For these applications, the instrument vents are purchased with tapped connections to carry the vented gas to a safe location. Because the gas used may not be totally dry, there may be a need for filters and separators to protect the instruments. Consult the valve and instrument manufacturers for requirements. Also consider the area electrical classification for electropneumatic positioners.

Because all positioners use small restrictions to reduce the air consumption at steady state, it is vital that the air be clean. Some positioner designs include a small integral filter. Satisfactory long-term operation of the instruments is considerably improved if all the dirt and moisture are removed in a good external supply filter before it reaches the positioner.

Liquids are efficient transporters of dirt and solids in supply lines to places where they will cause trouble. Liquids also can lead to ice, which can block flow in restrictors resulting in positioner malfunction, and then it may vanish without a trace after causing an unexplained upset. A simple mechanical separator will normally not remove enough moisture to properly protect the instruments.

Force-Balance Positioners The force-balance positioner shown in Figure 6.2e has an element that compares the force generated by the input signal with the force generated by the feedback spring connected to the valve stem. Figure 6.2f shows the electropneumatic force-balance positioner.

Motion-Balance Positioners The motion-balance positioner in Figure 6.2g compares the motion of an input bellows or diaphragm with linkage attached to the valve stem. Either can be very accurate. Bellows-type input elements are generally thought to be more accurate than diaphragms, and although slightly more likely to fail in fatigue, both types are used successfully. New and mostly electronic positioners differ widely in design and performance.

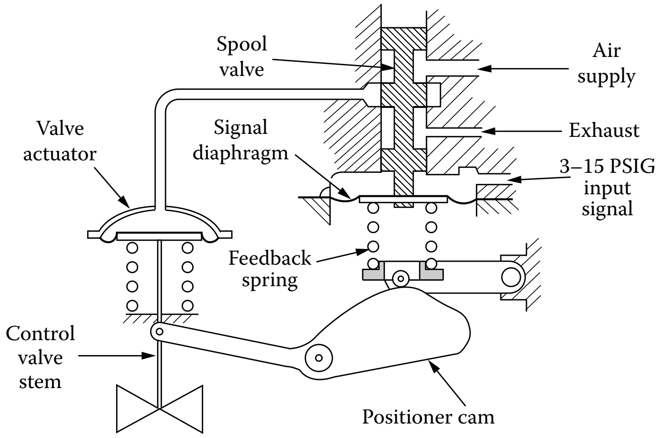


FIG. 6.2e
Force-balance positioner.

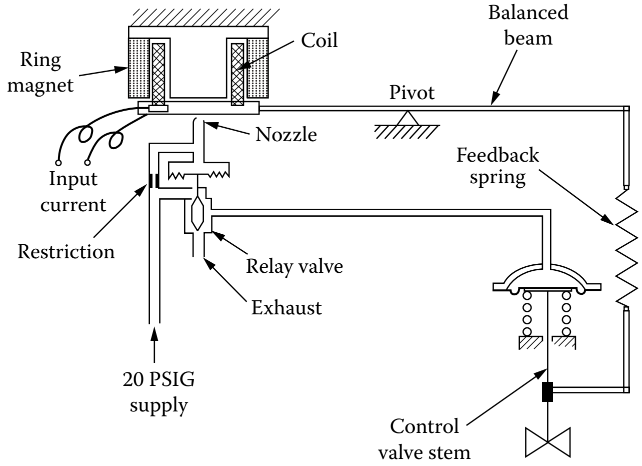


FIG. 6.2f
Electropneumatic force-balance positioner.

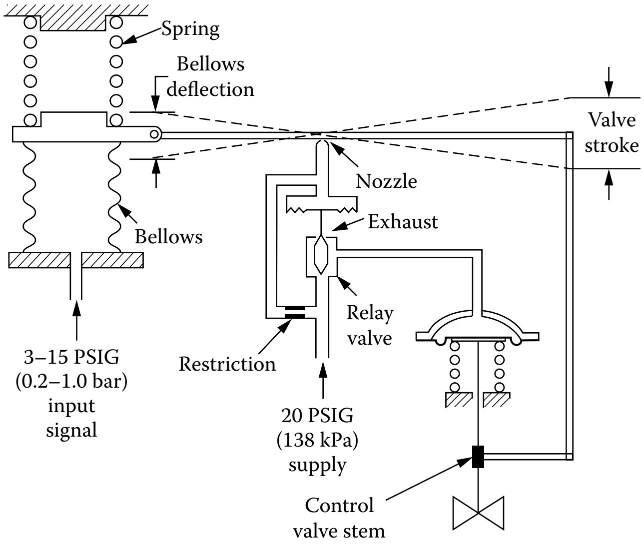


FIG. 6.2g
Motion-balance positioner.

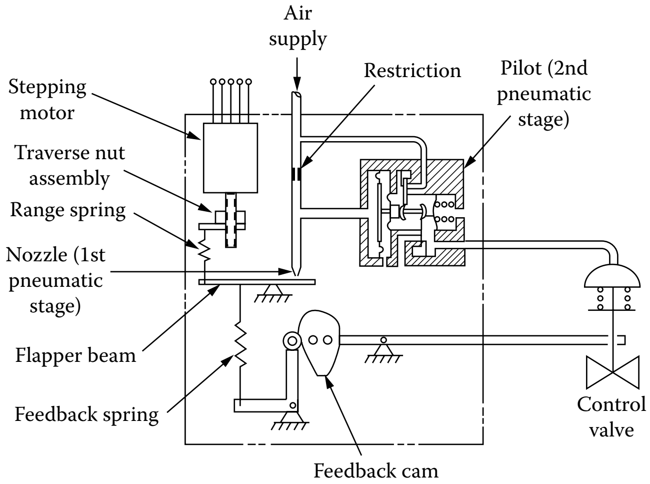


FIG. 6.2h
Digital-to-pneumatic valve positioner.

Electrohydraulic Positioners The electrohydraulic positioner addresses the need to control actuators on very large valves and for high-pressure differential applications. These systems typically require an external motor-driven pump, reservoir, pressure control regulators, high-pressure accumulators, and filters. The positioners may use high pressure oil jet tubes that are directed by linkages to regulate the oil pressure in the actuator.

Hydraulic pressures are typically 750–1,500 PSIG (5,000–10,000 kPa). A hydraulic pressure accumulator is often provided to store the energy for peak requirements. The valve vendors can suggest or supply the hydraulic system packages. Specification issues to be determined include the use of flammable or nonflammable oils, electrical classification, peak needs, noise, space required, costs, and maintenance. If the area has several large valves, a larger central hydraulic system may be used.

Digital to Pneumatic Positioners The digital valve actuators are described in Section 6.3. Some of the digital to pneumatic positioner designs have used rotary motors to control the pilot system. These tend to feature lock-in-place on loss of input signal. Others use a small fast solenoid valve that switches rapidly between open and closed to create an average air pressure for the actuator.

Still others may use a piezoelectric valve (electrical signals cause a deflection in a special crystal structure), either proportional or pulsating. A pulse stepping motor may rotate a shaft to set a follower, which develops an input signal force or position (Figure 6.2h). The remainder of the positioner is pneumatic. Stroke speed may be limited by stepping motor response. The only way to make decisions during the selection process with the many different designs with their many subtle differences may be to rely on proven performance.

Positioner Accessories

A number of options are available for most positioner designs, such as gauges to display the supply, signal, and output pressures. These add little cost but are a great help during checkout and maintenance.

Bypass air switches will connect the controller output directly to the actuator. These were once popular when 15 PSIG (100 kPa) pressure was adequate to move the spring-diaphragm actuator if the positioner was direct acting (increase input–increase output). The bypass is rarely specified now because this feature permits only very limited maintenance on the positioner and most modern high-performance positioners do not offer this option.

There are also safety implications in providing the possibility of changing the valve response. At one time, the usual justification for the bypass valve was to provide a way to get a poor performing positioner out of the loop for field troubleshooting. In some corporations the only control valves now purchased without positioners are small (1 in. [25 mm] and less), for low-pressure differential (0–20 PSIG [0–140 kPa]), and less critical service (comfort heating and ventilation), and where absolutely minimum cost is required.

Mechanical switches are used to provide a signal to the operator or control system that the valve stem is at or beyond a specified position. The specifications for these switches should consider the voltage and current surges caused by lightning strikes on long wire runs. These are available to suit the local electrical classification requirements. For some manufacturers of piston-type actuators, springs to provide for fail-safe action on loss of air are considered an accessory and are not supplied except as specified.

Position Indicators

Globe-style valves normally have a simple stem position pointer mounted on the stem with a simple scale on the yoke. A local valve position indicator may be mounted on rotary valves to allow the operator to see the state of the valve. Stem position transmitters, usually derived from positioner designs, are used to provide independent remote indication of valve stem position.

TRANSDUCERS

I/P (Electropneumatic) Transducers

A large variety of converters (transducers) are discussed in detail in Section 3.3. The electropneumatic transducer (I/P) converts electrical signals (usually 4–20 mA DC) into a pneumatic signal, usually 3–15 PSIG (0.2–1.0 barg). The most common application is the interface between an electronic controller output and the pneumatic control valves. It is also used between digital control systems and control valves. A few designs use signal feedback to improve accuracy (Figure 6.2i).

Most I/P transducer designs are of the "motion-balance" type (Figure 6.2j), where the small force developed by the milliampere current through a coil in a magnetic field causes motion in a nozzle-baffle assembly, resulting in a changing pneumatic pressure. The nozzle-baffle system is an amazingly accurate mechanism for measuring small distances.

When an I/P is used within a control loop between the controller and the control valve, its error is combined with the valve error, which is detected by the loop controller and then driven towards zero error. Repeatability and reasonable linearity are required, and most I/Ps have advertised accuracies of 0.35–1.0% of full scale. Most I/Ps have relatively low air capacity, and a booster relay may be needed to drive a pneumatic actuator unless a positioner is used.

Digital Electropneumatic Transducers

A variety of devices are used to convert digital signals in addition to the electronic digital-to-analog converter. One device uses a stepper motor, as in the digital positioner mentioned above.

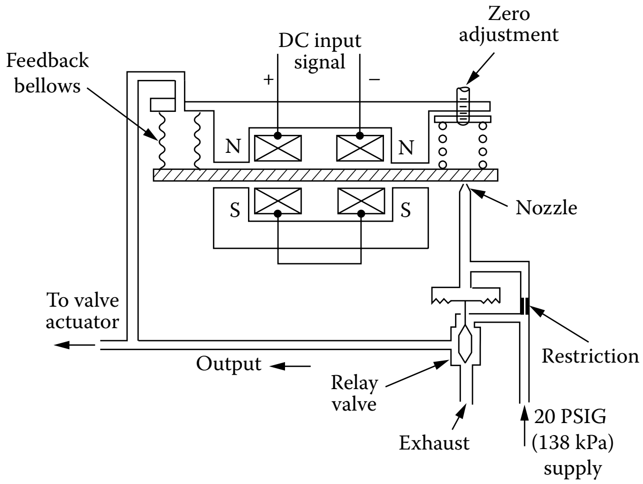


FIG. 6.2i
Electropneumatic force-balance transducer design and components.

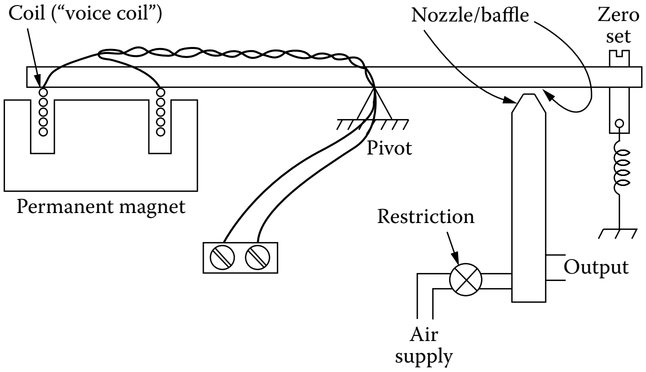


FIG. 6.2j
In a motion-balance electropneumatic transducer, the coil is of similar construction to the voice coil of a loudspeaker. Leads are arranged to go through a pivot point to reduce their effect on output.

A second device has a number of wires connected to it and receives data on all these wires simultaneously. This "parallel" data signal is converted to an analog current, and this controls the output pressure. Another device responds to a string of pulses ("serial" data) to set the output pressure.

RELAYS

Booster Relays

A "booster" relay, Figure 6.2k, is a device that amplifies pneumatic signals in volume (capacity) or in pressure, or both. Most booster designs were derived from a pressure regulator, with the input signal providing the loading force in place of the regulator spring. Internal pressure feedback

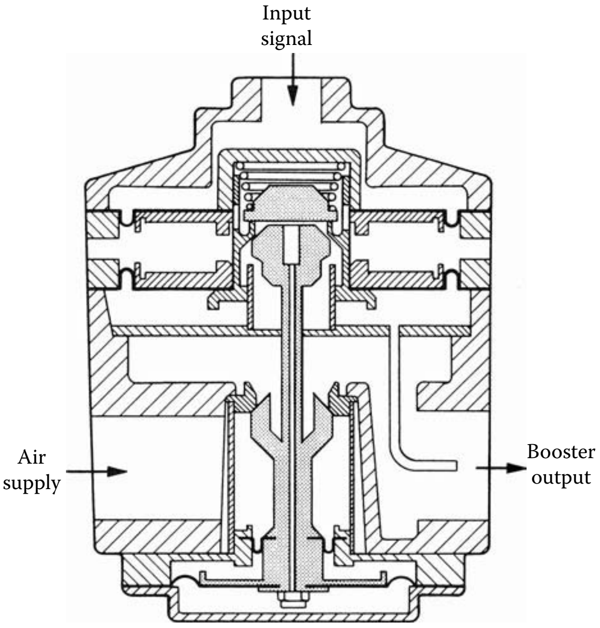


FIG. 6.3k
Volume booster relay.

may improve accuracy. The downstream-facing pitot tube compensates for flow-related pressure drop.

The feedback diaphragm has a smaller area than the input diaphragm so that greater pressure is required to achieve force balance. Volume boosters with a pressure ratio of 1:1 (gain of 1) are sometimes used to speed up a valve actuator response if a positioner is not used. The booster cannot overcome inaccuracies due to friction or forces on the valve plug, but it will reduce their effects because of faster response.

The stroking times of pneumatic piston actuators (Section 6.4) are a function of their size and of the size of the connecting tubing. A size 300 actuator with $^3/_8$ in. (9.5 mm) tubing has a stroking time of 27.7 sec. If the output of the positioner operating that actuator is piped to one booster on each side of the piston, the stroking time is reduced to 5.4 sec, and if two boosters in parallel are used on each side, the striking time is reduced to 2.7 sec.

Large control valves and large actuators may require flow or pressure boosters installed between the positioner and the actuator to achieve the required speed. One special "dead band booster" design, Figure 6.2l, does not respond until an (approximately) 1 PSIG (6.9 kPa) difference between input and output is exceeded.

A built-in needle valve allows a limited airflow to bypass the booster gain portion and provides adjustable damping. Springs provide the 1 PSIG dead band with this relay. Very large volume amplification occurs for fast input signal change, but it does not amplify a slow change.

One common application for this relay is the centrifugal compressor antisurge control, where the control valve must open very quickly (1–3 sec) but then is required to throttle smoothly. The needle valve is adjusted with the complete control system assembled and operating. Careful tuning is vital to proper operation. With the booster needle valve fully closed, the positioner-valve system may be unstable.

Best operation occurs when the needle valve is open only enough to smoothly dampen the oscillations. If more than one of these relays is used to add capacity for one side of the piston actuator, it is absolutely necessary to set the needle valves as identically as possible to avoid a complex and probably unstable interaction between the relays.

It is worth repeating here that no air device will operate properly without an adequate air supply. With their large air-handling capacities, valve booster relays require 0.5–0.75 in. (12–20 mm) air supply and output tubing to the valve operator. Large filters are used to minimize any air supply restriction. In one special case, a large air tank 10 ft^3 (0.3 m^3) was installed near the valve in order to provide adequate local air surge capacity.

Reversing and Other Relays

Besides the booster function, pneumatic relays can meet other control valve requirements. The fixed-gain-plus-adjustable-bias relay for split-ranging was mentioned in connection with Figure 6.2c. Another 1:2 reversing relay is shown in Figure 6.2m. This design has a reverse gain of two (0% input results in 100% output, 50% input results in 0% output). It will reverse the operation of a valve and provide the gain and bias needed for split-ranging, while retaining the native failure action on loss of the supply air.

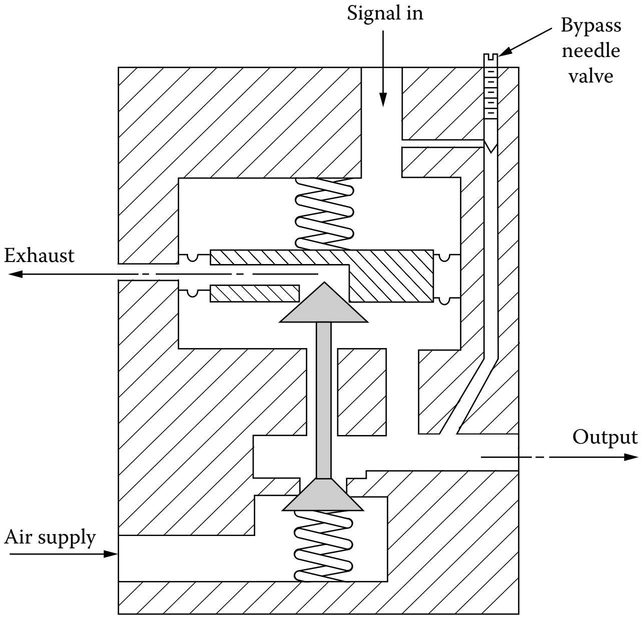


FIG. 6.2l
The design of a dead band booster in which the springs provide the 1 PSIG (6.9 kPa) dead band.

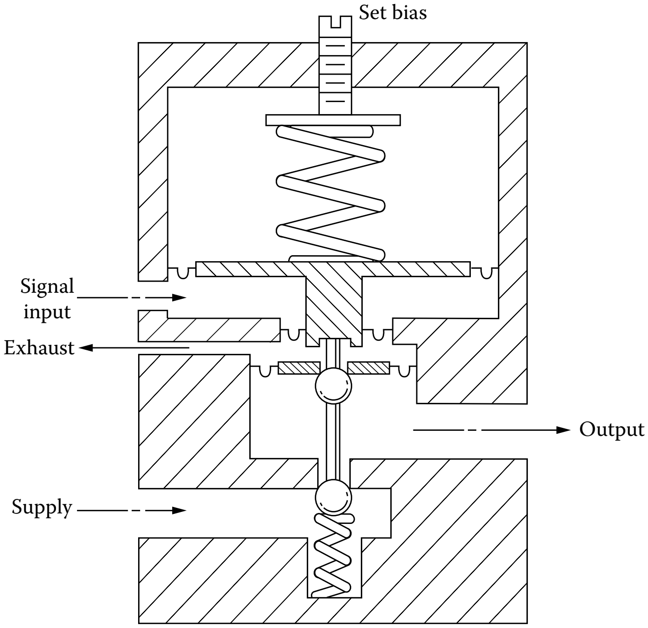


FIG. 6.2m
2:1 reversing relay.

The relay is a force-balance device. The output pressure generates a force on a diaphragm with one half the area of the input diaphragm so that the output pressure must be twice the input pressure. A wide variety of relays are available, amplifying, "reducing" (gains of less than one), computing (averaging, square root, and so on.), selecting (highest, lowest), some with provision for biasing the output.

Complex control concepts have been implemented entirely with pneumatic logic. This functionality is more commonly accomplished today with electronic circuits or in the control system. Very successful override control systems have been designed with one valve position controlled by several process variables such as pressure, level, and temperature.

Table 6.2n lists the air capacity for a few selected pneumatic components to show the range of typical devices. The data was found in various places and some is old. If the exact capacity of a relay or the stroking speed of a valve is important, it would be wise to check with the manufacturer.

Quick-Exhaust Relays

The quick-release valve or quick-dump valve, Figure 6.2o, is a pilot valve that opens a high-capacity vent when the input pressure drops below a set pressure. This is used where it is needed to quickly open a valve that vents a vessel or system during a shut-down.

It is not unusual for a positioner to bleed excess air from the actuator for some seconds before a valve begins to move, because the actuator is completely filled (saturated) with supply pressure air. The quick exhaust helps reduce the lost time. The controller continues to call for valve action after it reaches its mechanical stop.

TABLE 6.2n
Air Capacity of Selected Pneumatic Components

Model	Air Capacity	PSI Supply	Function	Data Source
Moore 750E	7.5 SCFM	60 PSIG	Positioner	Catalog
Moore 71	4 sec, 3–15, 350 in^3	18 PSIG	Positioner (1)	Catalog
Moore I/P	0.1 SCFM	20 PSIG	I/P converter	Catalog
Moore 50M	3.7/2.1 SCFM	20 PSIG	Controller	Catalog
Moore 61H	10.5 SCFM	9 PSIG	Booster relay	Catalog
Moore 61L	4.5 SCFM	9 PSIG	Booster relay	Catalog
Moore 61F	2.4 SCFM	9 PSIG	Booster relay	Catalog
Moore 610F	2.4 SCFM	9 PSIG	Booster relay	Catalog
Moore 612	0.12 SCFM	9 PSIG	Multi-input low-signal select	Catalog
Moore 66BA2	2.2 SCFM	9 PSIG	1:2 A relay (2)	Catalog
Moore GC 661	2.2 SCFM	9 PSIG	Amp + bias relay	Catalog
Moore 68	2.4 SCFM	9 PSIG	Computing relays (1)	Catalog
Moore GC 77	NA (low)		E/P transverter	Catalog
Moore 750 I/P	7.5 SCFM	60 PSIG	I/P positioner	Catalog
Moore 750	7.5 SCFM	60 PSIG	Positioner	Catalog
Moore 72	0.7 SCFM (3)	25 PSIG	Positioner	Moore paper
ASCO 8320	24 SCFM(4)	30 PSIG	Solenoid valve	Catalog
Valtek Booster	2300/1350(5)	60 PSIG	High capacity	Catalog
Fisher 3710	8 SCFM	60 PSIG	Standard capacity	Catalog
Fisher 3710	14 SCFM	60 PSIG	High capacity	Catalog
Valtek Beta	11 SCFM	60 PSIG	Positioner	Catalog

Notes:

1. Typical
2. Typical for family of amplifying and reducing relays
3. Higher available as option
4. Typical three-way, "universal," no minimum pressure difference requirement, used for valve interlock
5. Supply/exhaust

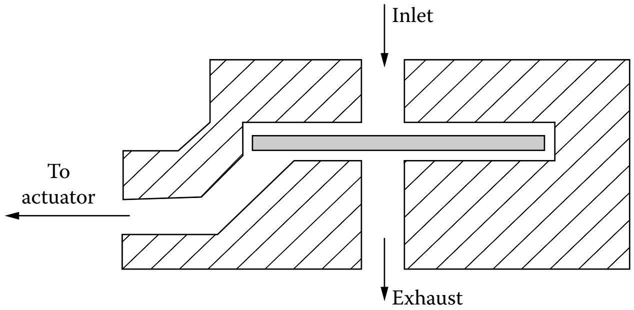


FIG. 6.2o
Quick-exhaust valve in which the elastomer flapper normally allows air to flow to the actuator while sealing the exhaust; if the inlet pressure is below that in the actuator, it allows the air from the actuator to be exhausted.

Quick-exhaust three-way solenoid valves having a large exhaust port for a similar effect can also be used.

Relays to Lock-up Valve Position

A lock-up relay or "fail in last position" relay, Figure 6.2p, is available to seal in the existing actuator pressures if the air supply is lost. It is intended to hold the valve in the last controlled position. If the valve is partly open at the time of air failure, this "frozen" position is not absolutely predictable because the servo-action is lost, variable plug forces may move the stem, and air may leak out in an unpredictable manner.

When a piston actuator is to be locked in its last position when the air supply fails, it is necessary to install two lock-up valves at the two ends of the piston and use a three-way switching valve to vent their actuators when the air supply pressure drops to some preset limit (Figure 6.2q).

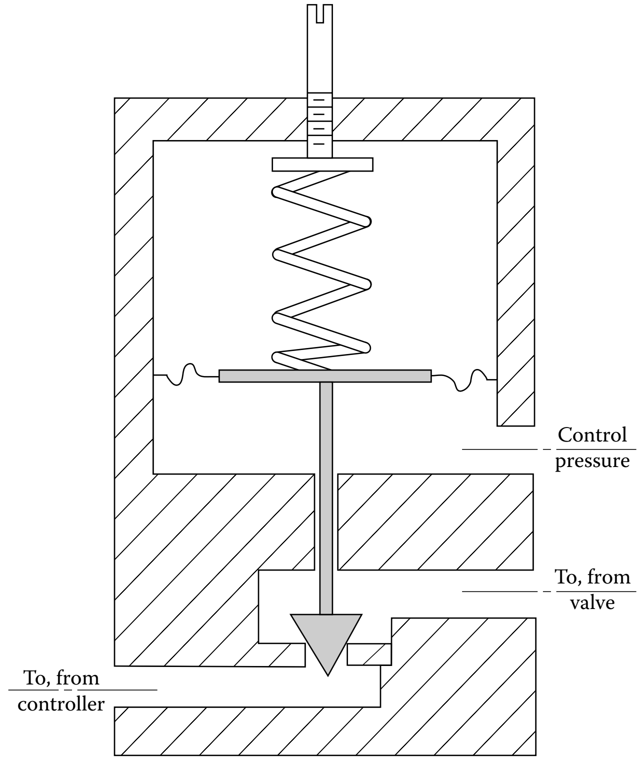


FIG. 6.2p
Lock-up relay: If control pressure drops below the spring setting of this relay, then the relay valve closes and seals in the last air pressure in the actuator.

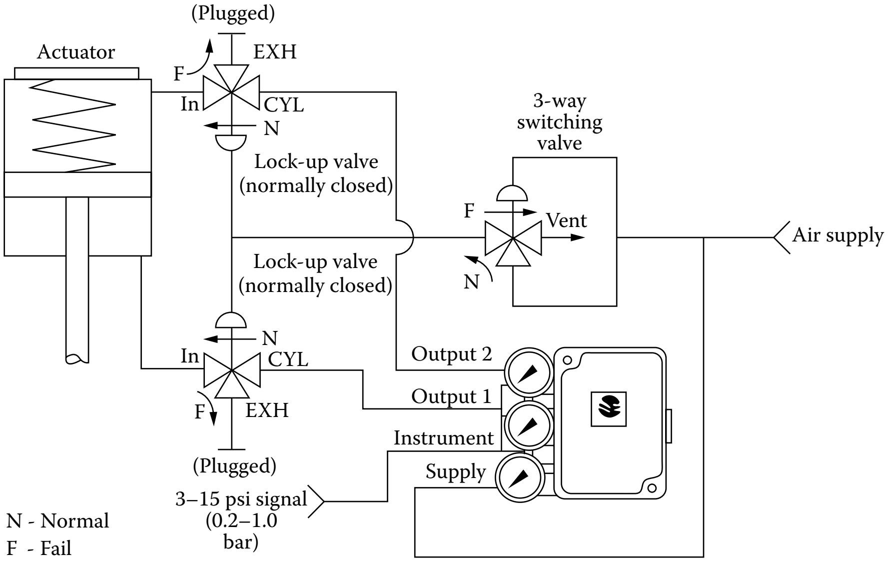


FIG. 6.2q
A valve with piston actuator, which is provided with fail-in-place lock-up controls. (Courtesy of Valtek, Flowserve Corp.)

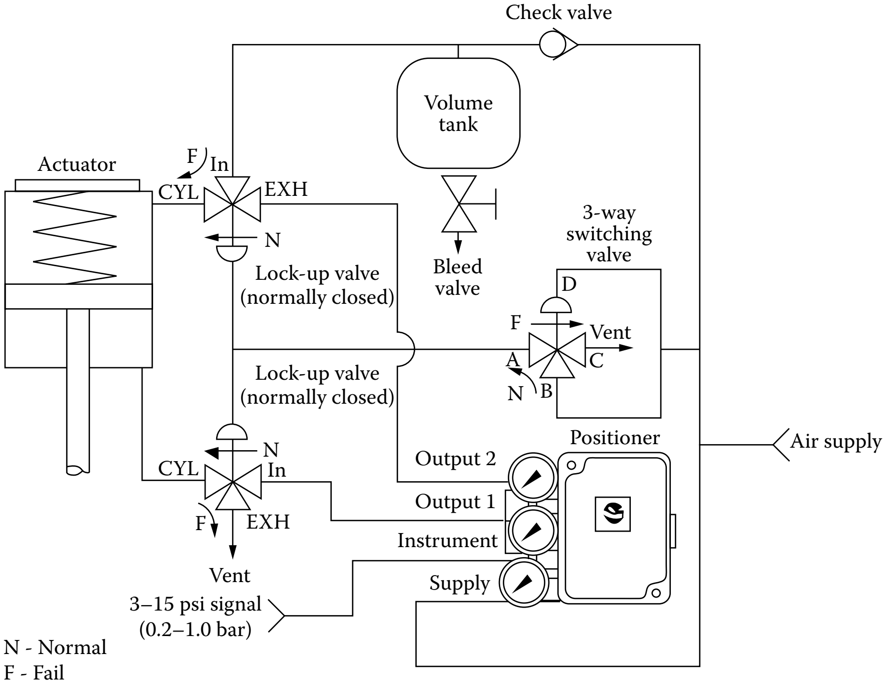


FIG. 6.2r

External air tank can provide the energy source to guarantee a positive failure position for piston-actuated valves. (Courtesy of Valtek, Flowserve Corp.)

Failure Position Guaranteed by Stored Air

In piston-operated valves, the air stored within the piston can also be used as the energy source to move the valve to a safe position when the air supply is lost. This can be achieved by locking the air in on one side of the piston while venting the other, when the air supply pressure drops to a predetermined value.

If there is concern that the piston volume is insufficient to fully stroke the valve under all conditions when the air supply might fail, an external tank can be used as the source of driving energy. Figure 6.2r illustrates such an installation for a fail-closed (air-to-open) valve.

Some piston-type actuators have no spring, and others have a spring too weak to guarantee valve action on loss of air supply. For these, an air tank (typically 0.5 ft^3, or 0.014 m^3) with a check valve on the inlet may be used to allow limited operation or orderly shutdown (Figure 6.2s). Because everything leaks, this will only be reliable for a reasonable time. Longer term requirements may be satisfied if the process pressure tends to hold the valve plug in the desired direction.

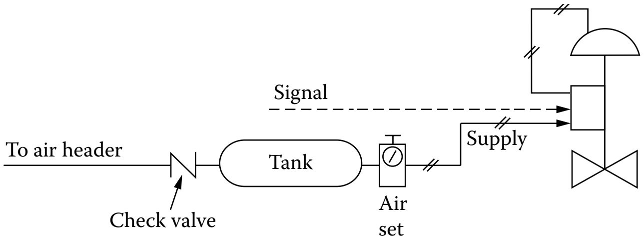


FIG. 6.3s

A temporary air supply can be provided by an air accumulator tank, which thereby will guarantee orderly shutdown.

ENERGY SUPPLIES

Air Sets

The "air set" is the air regulator with filter and drip pot used to supply air to the positioner or other instrument. It is often purchased with the valve, mounted, and piped. The regulator must have the pressure range to cover the spring range of the actuator, and it should have a built-in overpressure vent to protect the actuator; a gauge is valuable to aid in setting the output pressure and for discovering a failed or plugged supply.

Flow capacity is a problem only with very large valves. Some piston actuators and positioners will operate with up to 150 PSIG (1 MPa) supply, and there is a trend toward eliminating the regulator, but the filter is retained for reliable valve operation. Most manufacturers recommend that each control valve have its own individual air set.

The risk in providing air sets for high-pressure actuators such as air cylinders is that field operators can thereby limit the supply pressure, which in turn reduces actuator stiffness (resistance to the dynamic forces of the process). An air set must be used when the pressure rating of the actuator or positioner is lower than the air supply pressure.

The use of air filters is always recommended. They should be installed in the air supply serving the positioner and should be designed for the maximum air supply pressure. Their purpose is to remove moisture, oil, and all particles that are 5 μ or larger. In industry it is axiomatic that dirt is everywhere, and it will always find small restrictions to plug.

Hydraulic (High-Pressure) Operation

High-pressure, high-performance, hydraulic oil-filled systems are used for special needs, such as very large valves or those that require forces beyond 150 PSIG air supply. These function much like larger versions of the hydraulic power steering used on automobiles.

Very high speed performance and power is available, all at added cost. Most of the accessories and functions commonly found in pneumatic systems are available. These include: positioners, flow boosters, and pressure regulators.

Hydraulic (Water) Operation

There are two sorts of "hydraulic" operations. The first was addressed above. The second "hydraulic" system is the ubiquitous but rarely considered municipal water supply and wastewater systems, which have their own special set of requirements and solutions.

Actuators, pilot valves, regulators, and control relays are available that operate using water pressure just like the pneumatic systems used elsewhere. These control the water to fill elevated storage tanks and stop flow when the tanks are filled, and then implement the desired operating logic when water pressure is lost. Even fluid flow rate flow control is possible.

Self-contained water operation eliminates the need for electrical or air power and signals. Relays and valves to control water pressure surges can be provided. Users should contact the appropriate specialized suppliers for further information.

LIMIT SWITCHES

Switches are installed on electric motor-driven valves to open the circuit and stop driving the motor when the valve is at its limit (fully open or closed) or on motor over-torque. The name "limit switch" is also used to describe switches installed to signal when a valve is at or beyond a predetermined position. These switches are used for operator information, interlock inputs, or computer feedback.

It is necessary to consider the mounting problems, electrical classification of the area, the electrical characteristics of circuit, overtravel of actuating arm, and corrosive nature of the area. Usually, it is easiest to purchase the valve complete with the required switches already installed.

Because of environmental problems, some users have been using sealed magnetically actuated or proximity switches. Note that it is difficult to adjust limit switches closer than ±5–10 % and that dead bands of 2–5% can occur. Problems with failure of small contacts have been traced to voltage surges picked up on long field wiring.

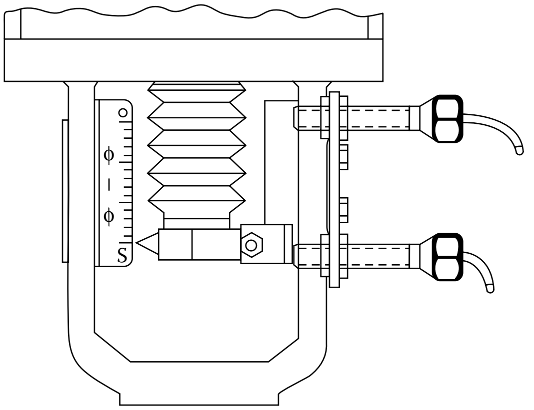


FIG. 6.2t
Installation of proximity switches to detect the open and closed position of a valve or a linear valve actuator. (Courtesy of Valtek, Flowserve Corp.)

Plantwide standardization on limit switch specifications for makes and model number will reduce spare part storage and simplify maintenance. When specifying the limit switches, one should specify the required contact ratings, the contact configurations (SPDT, DPDT, and so on), and the type of housing required. Typical choices include weatherproof, explosionproof, or hermetically sealed explosionproof.

Figure 6.2t illustrates a proximity switch installed on a linear actuator. The maximum spacing allowable between the switch and the sensed surface is 0.11 in. (3 mm). The switch is available in UL- and CSA-approved explosionproof designs.

SOLENOID VALVES

The solenoid valve (see Section 6.3 for more details) as a control valve accessory is used (1) to operate on/off pneumatic actuators or (2) to interrupt the action of modulating valves by switching air or hydraulic pressures. It is common practice to use a solenoid valve as the pilot for a pneumatically operated on/off valve because of the wide choice of features and capabilities available in the solenoid valve.

Solenoid valves are primarily used as parts of start-up or shutdown, interlock, or batch systems to cause the control valve to take some predetermined action under certain conditions. Three philosophies are in common use.

Three-Way Solenoids

In the first, the 3–15 PSIG (0.2–1.0 bar) signal to the positioner is blocked and the downstream tubing either vented or connected to some other preset pressure (Figure 6.2u).

This approach is reliable because the solenoid valve is lightly stressed, the positioner and valve have been in continuous use, and any failure or poor operation should have been detected during normal operation. With this scheme, the substituted signal can be any value over the operating range, the

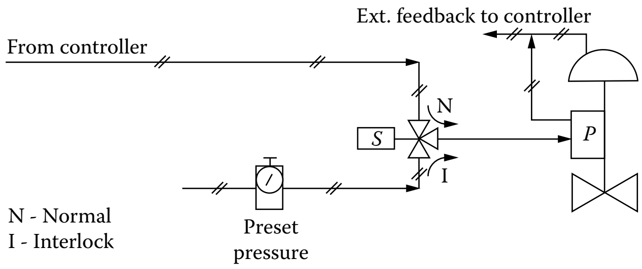


FIG. 6.2u

Under normal conditions the solenoid passes the controller signal to the positioner. Under abnormal conditions the interlock solenoid valve blocks the controller signal and opens the path for a manually preset air signal to reach the positioner.

valve will go to the desired opening, and the advantages of the positioner are retained. If the actual positioner input is fed back to the controller external feedback connection, then a smooth return to normal control may be expected when the control loop is returned to normal operation.

In the second philosophy, the solenoid valve is installed in the tubing between the positioner and the actuator (Figure 6.2v). Solenoid valves with adequate pressure rating and flow capacity are required. Only three control valve actions are possible: fully closed, fully open, or lock-up existing pressures.

A third philosophy is to shut off the air supply to the positioner and let the valve act as designed on loss of air supply. This is one scheme used in batch operations where the valves are switched into service when a process unit or pump is turned on.

In order for the system to shut down (valve to close) in case of loss of power or emergence of an unsafe condition, it is desirable for the solenoid valve to be continuously energized during normal operation. This will guarantee that any failure, loss of power, or a broken wire will cause a fail-safe action.

As a control valve accessory, usually a three-way (three ports) solenoid valve is required. Some designs require that pressure be always applied to one certain port and that another certain port always be used as the vent. This does not always suit the required logic, but valves can be found designed for "universal" operation where there is more freedom in assigning port function.

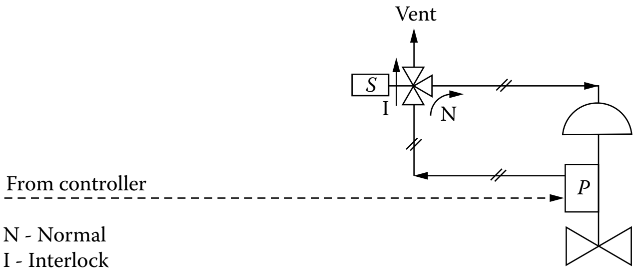


FIG. 6.2v

Interlock solenoid valve controlling actuator directly.

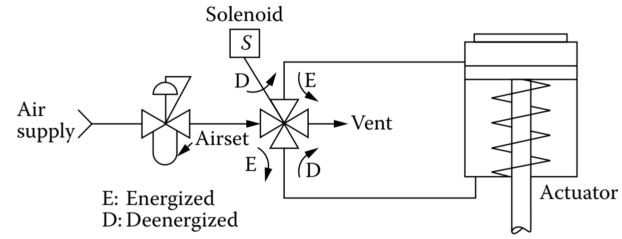


FIG. 6.2w

On/off cylinder actuator operated by four-way solenoid.

Also, note that clean, dry, oil-free instrument air provides no lubrication, and some types of solenoid valves (spool-type) will have a short life or become unreliable without lubrication. See also the discussion of the impact of solenoid valve operation on valve response in Section 6.9.

Four-Way Solenoids

For on/off cylinder-operated valve actuators, four-way solenoids are often used (Figure 6.2w). They are fast, provide positive operation, and are available for a variety of AC or DC voltage services and with Class F coils for up to 310°F (154°C) temperature services.

Solenoid Capacity

Each approach must consider the flow capacity of these solenoid valves. The desired solenoid valve C_v must be greater than the C_v of the positioner to avoid a reduction in stroke speed. If they are equal, then the valve speed will be roughly half of that without the solenoid.

It is typical of solenoid valves that small valves are directly operated and the larger ones are pilot operated. In pilot-operated valves, a small direct-operated valve uses air pressure to switch the larger main valve. Pilot-operated valves require a certain minimum air pressure differential in order to operate the main valve.

If the solenoid valve is tripped while the air pressure is less than this pressure, the main valve will not change state and tripping the pilot will have no effect. The valve will trip later when the pressure differential becomes high enough to operate it. Where full pressure is always present this is not an issue.

HANDWHEELS

Handwheels are used to provide for partial or complete manual control of the valve and to override the pneumatic actuator. Some, mounted on top of the actuator, Figure 6.2x, can only push on the valve stem to close (or open with inverted trim).

Others can be configured for continuous, bidirectional operation with force amplification ratios from 40:1 up to over 100:1. These top-mounted designs are illustrated in Figure 6.2y.

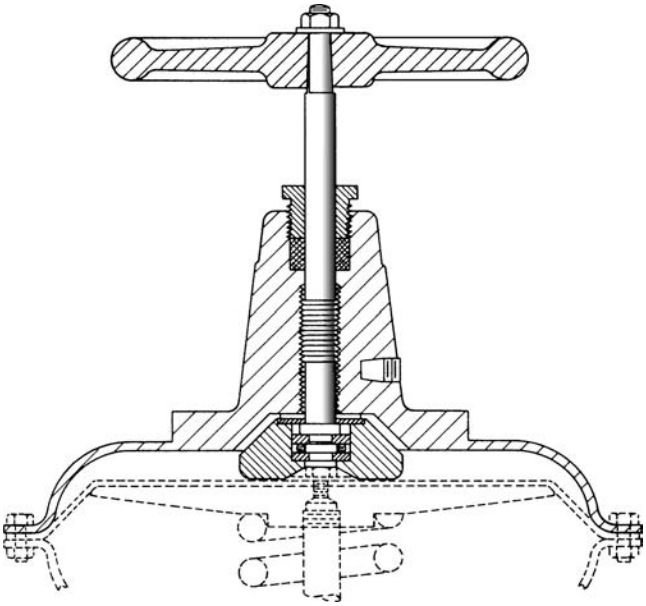


FIG. 6.2x
Top-mounted handwheel can be used to manually operate a spring-opposed pneumatic actuator, but only in one direction.

Side-mounted handwheels have engagement clutches to allow the handwheel to fully stroke the valve open and closed, Figure 6.2z. In the continuously connected handwheel design, the handwheel is provided with a "neutral" position. When placed in that position, the handwheel does not interfere with automatic operation of the valve.

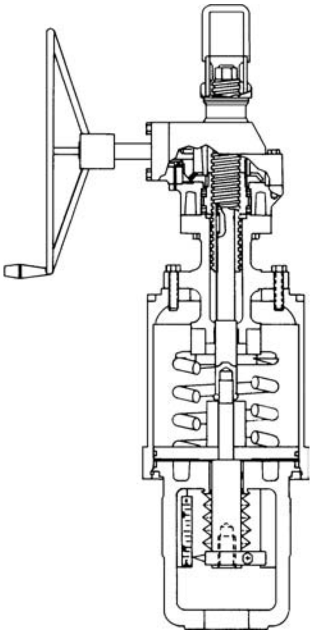


FIG. 6.2y
Top-mounted, continuously operated bidirectional handwheel. (Courtesy of Valtek, Flowserve Corp.)

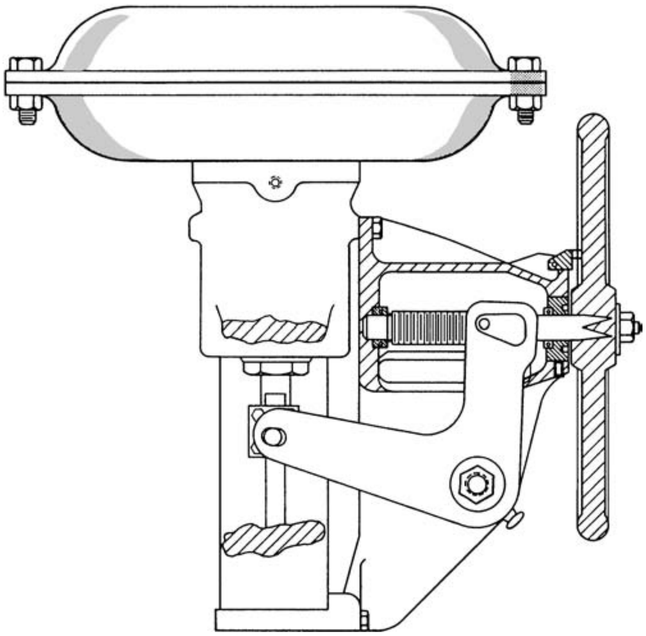


FIG. 6.2z
The side-mounted handwheel, if engaged, can fully stroke a spring-opposed pneumatic actuator in both directions.

Turning the handwheel one way forces the stem to extend, and turning it the other way forces the stem to retract. Adjusting the handwheel screw away from the "normal" position introduces a "limit stop" on the valve travel in one direction or the other, but not both. Consider that some handwheels may interfere with interlock shutdown operation of the valve.

For the manual operation of rotary valves, either hand levers 15–22 in. (38–56 cm) in length or clutch-equipped gearbox-type handwheels are used.

If manual throttling control is intended, the plant design must consider how the human operator will know how to set the valve and if the process can be safely controlled manually. This is not always the case, and handwheels have limited application in modern continuous process plants.

LIMIT STOPS

It is possible to install fixed limit stops to limit valve stem motion to either ensure a minimum opening or limit a maximum opening (Figure 6.2aa). These are usually purchased with the valve. Consider how to document the purpose and settings for these for maintenance purposes.

BYPASS VALVE

Perhaps not thought of as an accessory, the manual bypass for steam shut-off valves are critical to start-ups. The manual valves are "cracked open" to pressurize and heat up the steam

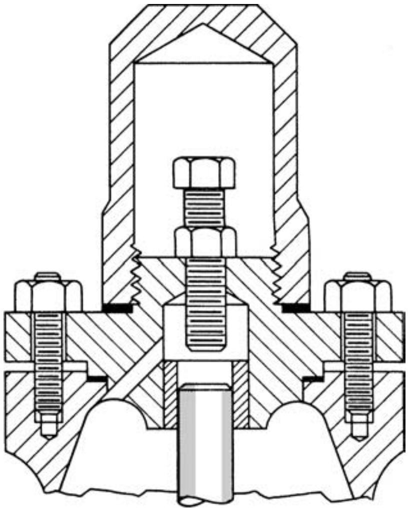


FIG. 6.2aa

Externally adjustable limit stop, shown in a valve body subassembly.

header. Opening the large valve could result in serious damage downstream as slugs of water are propelled into equipment. Other applications may have similar issues.

References

1. Coughran, M. T., "Measuring the Installed Dead Band of Control Valves," ISA, 1997.
2. Gassmann, G. W., "When to Use a Control Valve Positioner," *Control,* September 1989.
3. Langford, C. G., "A User's View of Process Control and Control Valve Positioners," ISA, 1996.
4. Langford, C. G., "A Method to Determine Control Valve Dynamic Requirements," ISA, 2002.
5. ANSI/TSA-75.25.01, "Test Procedure for Control Valve Response Measurement from Step Inputs."
6. ANSI/ISA-TR75.25.02-2000, "Control Valve Response Measurement from Step Inputs."
7. ISA-75.13-1996, "Method of Evaluating the Performance of Positioners with Analog Input Signals and Preumatic Output."

Bibliography

Arant, J. B., "Positioner Use Is Myth-Directed," *InTech,* November 1992.

Carey, J. A., "Control Valve Update," *Instruments and Control Systems,* January 1981.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems,* January 1980.

Gassman, G. W., "When To Use a Control Valve Positioner," *Control,* September 1989.

Brayton, P., "Installing Smart Positioners–A Wise Move," *Chemical Engineering Magazine,* December 1, 2000.

Jury, F. D., "Positioners and Boosters," *Instruments and Control Systems,* October 1977.

Keagle, J., "It's 128 mA, Do You Know Where Your Positioner Is?" *InTech,* May 1992.

Lipták, B. G., "Control Valves in Optimized Systems," *Chemical Engineering,* September 5, 1983.

Lloyd, S. G., "Guidelines for the Use of Positioners and Boosters," *Instrumentation Technology,* December 1968.

Miller, T. J., "Pneumatic Valve Actuators Continue to Dominate as Electronics Move into Process Control," *Control Engineering,* Vol. 28, No. 10, p. 88, 1981.

Price, V. E., "Smart Valve Intelligence Takes Many Forms," *InTech,* August 1992.

Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 ISA Conference, Houston, TX, October 1992.

Schuder, C. B., "Fluid Forces in Control Valves," *Instrumentation Technology,* May 1971.

Scott, A. B., "Control Valve Actuators: Types and Application," *InTech,* January 1988.

Shinskey, F. G., "Dynamic Response of Valve Motors," *Instruments and Control Systems,* July 1974.

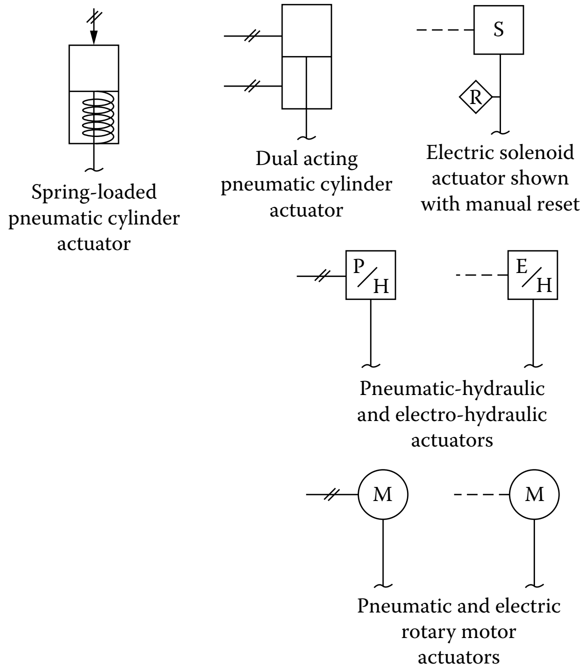
Parkinson, G., "'Smart' Actuator Incorporates All the Controls in One Package," *Chemical Engineering Magazine,* January 1, 2003.

6.3 Actuators: Digital, Electric, Hydraulic, Solenoid

C. S. BEARD (1970)

B. G. LIPTÁK (1985, 1995)

P. M. B. SILVA GIRÃO (2005)



Flow sheet symbols

Types:	1. Digital 2. Electromechanical (linear and rotary) 2a. Stepping motors in smaller sizes 2b. Reversible motor gears for larger sizes 3. Hydraulic and electrohydraulic (the pump can be driven by stepping or servomotors) 4. Solenoid
Energy Sources:	Electric, hydraulic or both
Speed Reduction Techniques:	Worm gear, spur gear, or gearless
Torque Ranges:	0.5–30 ft · lb$_f$ (0.7–41 N · m) for type 2a, and 1–250,000 ft · lb$_f$ (1.4–338,954 N · m) for type 2b actuators with gearboxes
Linear Thrust Ranges:	The maximum of about 500 lb$_f$ (2224 N) output force can be obtained from type 2a actuators; 100–10,000 lb$_f$ (445–44,500 N) can be obtained from type 2b ones. The thrust capability of type 3 actuators can exceed 100,000 lb$_f$ (445,000 N).
Speeds of Full Stroke:	Small solenoids can close in 8–12 msec. Throttling solenoids can stroke in about 1 sec. Electromechanical motor-driven valves stroke in 5–300 sec. Electrohydraulic actuators generally move at 0.25 in./sec (6.35 mm/sec) but can be speeded up by the use of hydraulic accumulators.
Costs:	0.25 in. (6.35 mm) solenoid pilot in stainless steel costs about \$50; with a two-way design in explosionproof plastic construction it costs about \$250, and when built for high-temperature service it costs about \$300. An electromechanical, type 2a actuator for 0.5–1 in. (12.7–25.4 mm) small ball valve can be obtained in explosionproof construction with limit switches for on/off service for \$750; with positioner and feedback potentiometer such an actuator costs about \$1500.

A typical rotary actuator with 300 ft · lb_f (407 N· m) torque rating costs about $2500.

The cost of larger type 2b or type 3 actuators can exceed $10,000.

Partial List of Suppliers:

ABB Group (1, 2) (www.abb.com)
Aeroflex Incorp. (2) (www.aeroflex.com)
Allenair Corp. (2, 3, 4) (www.allenair.com)
Auma Actuators Inc. (1, 2) (www. auma-usa.com)
ASCO-Automatic Switch Co. (Division of Emerson) (4) (www. asco.com)
Bafco Inc. (3) (www.bafcoinc.com)
Barksdale, Inc. (3) (www.barksdale. com)
Bettis-Emerson Process Management (2, 3) (www.emersonprocess.com/valveautomation/bettis)
Bodine Electric Co. (2) (www.bodine-electric.com)
Bosch-Rexroth Corp. (2, 3, 4) (www.boschrexroth.com/BoschRexroth/corporate/en/index.jsp)
Bray Controls-Bray International, Inc. (2, 4) (www.bray.com)
Bürkert Fluid Control Systems (1, 4) (www.buerkert.com)
Butler Automatic, Inc. (3) (www.butlerautomatic.com)
Center Line-Crane Co. (2) (www.cranevalve.com/products.htm)
Circle Seal Controls Inc. (4) (www.circle-seal.com)
Clark-Cooper Corp. (4) (www.clark-cooper.com)
Curtiss-Wright Flow Control Corp. (3, 4) (www.cwfc.com)
Danaher Corp. (2, 3) (www.danaher.com)
Danfoss (4) (http://us.ic.danfoss.com)
Detroit Coil Co. (4) (www.detroitcoil.com/PAGES/mainfram.html)
DeZurik/Copes-Vulcan (1, 2, 3) (www.dezurik.com)
Eaton Corp. (2) (www.eaton.com)
El-O-Matic-Emerson Process Management (2) (www.emersonprocess.com/home/products/index.html)
Engineering Measurements Comp. (Emco) (1) (www.emcoflow.com)
Emerson Process Management (4) (www. emersonprocess.com)
ETI Systems, Inc. (1, 2) (www.etisystems.com)
Exlar Corporation (1, 2) (www.exlar.com)
Flo-Tork Inc. (3) (www.flo-tork.com)
GE Water Technologies (1, 2) (www.gewater.com)
Hoke, Inc. (2) (www.hoke.com)
Honeywell Automation and Control (1, 2) (www.honeywell.com/acs/index.jsp)
Humphrey Products Co.(4) (www.humphrey-products.com/home.htm)
Invensys-Eurotherm (2, 3) (www.eurotherm.com)
Jordan Valve (2) (www.jordanvalve.com)
Kammer Valves-Flowserve Corp. (1, 2) (www.flowserve.com/valves/control/index.stm)
Keane Controls Corp. (4) (www.fluidprocess.com/Keane/Keane.htm)
Keystone Valve USA Inc. (2, 3) (www.keystonevalve.com)
Leslie Controls, Inc. (1, 2) (www.lesliecontrols.com)
Limitorque Corp.-Flowserve Corp. (1, 2) (www.limitorque.com)
Metso Automation Inc. (1, 2, 3) (www.metsoautomation.com)
McCanna Inc.-Flowserve Corp. (2) (www.mccannainc.com)
Micro Mo Electronics Inc. (2) (www.micromo.com)
Nihon Koso Co. Ltd; (1, 2, 3) (www.koso.co.jp/e/seihin.html)
Norgren-Herion (4) (www.herionusa.com)
OCV Control Valves (1, 2, 3, 4) (www.controlvalves.com)
Oil City Valve Automation (2, 3, 4) (www.ocvauto.com)
Oilgear Co. (3) (www.oilgear.com)
Oriental Motor USA Corp. (2) (www.orientalmotor.com)
Parker Hannifin Corp. (1, 2, 3, 4) (www.parker.com)
Plast-O-Matic Valves Inc. (2, 4) (www.plastomatic.com)
Regin Hvac Products Inc. (2) (www.regin.com)
Rotork Controls Inc. (2, 3) (www.rotork.com)

Saint-Gobain Performance Plastics-Furon Fluid Handling Div. (4) (www.furon.com)
Samson Controls, Inc. (1, 2, 3) (www.samson-usa.com)
Servo Systems Co. (2) (www.servosystems.com)
Servotronics Inc. (4) (www.servotronics.com)
Shafer-Emerson Process Management (3) (www.emersonprocess.com/valveautomation/shafer)
Shore Western Mfg. Inc. (3) (www.shorewestern. com)
Siemens AG (1, 2, 3) (www.siemens.com)
SMAR International Corp. (1, 2) (www.smar.com)
Snap-Tite Inc. (4) (www.snap-tite.com)
Sonceboz Corp. (2) (www.sonceboz.com)
Spirax Sarco (2) (www.spiraxsarco.com)
Superior Electric Co. (2) (www.superiorelectric.com)
Thunderco Inc. (3) (www.thunderco.com)
Tyco Valves (2, 3) (www.tycovalves.com)
Valcor Scientific (4) (www.valcor.com)
Valtek-Flowserve Corp. (3) (www.flowserve.com/Valves/control/index.stm)
Vetec Ventiltechnik GmbH (2, 3) (www.vetec.de)
Worchester Controls-Flowserve Corp.(2, 3) (www.worcestercc.com)

Note: *More information may be obtained for instance at* www.thomasregisterdirectory.com, *particularly at* www.thomasregisterdirectory.com/actuators/actuators_categories.html, *at* www.medibix.com, *or at* www.globalspec.com.

INTRODUCTION

This section covers a variety of valve actuators, except the pneumatic ones, which are discussed in Section 6.4. The various accessory items such as positioners, handwheels, limit switches, potentiometers, and the like are also discussed in Section 6.2. The discussion of digital actuators in this section is intended to complement Section 6.12 on intelligent valves and Section 6.18 covering digital control valves.

The discussion in this section begins with some selection and application guidelines, which is followed by the description of the five actuator categories: 1) digital, 2) electromechanical, 3) electrohydraulic, 4) motors and pumps, and 5) solenoids. The section ends with a brief discussion of the trends in valve actuation, including the features of microprocessor-based “smart” and intelligent actuators.

SELECTION AND APPLICATION

The following are some of the characteristics to consider in the application and selection of all types of actuators. Table 6.3a gives a summary of advantages, disadvantages, and applications for some of the designs.

Actuator Types

Valve actuator types discussed in this section belong to one of the following categories: (a) electric actuators, which use a motor to drive a combination of gears to generate the desired torque or thrust level. This category includes (i) rod linear actuators, whose output rod provides linear motion via a motor-driven ball or ACME screw assembly. In this design, the actuator’s load is attached to the end of a screw, or rod, and is often unsupported; (ii) rodless linear actuators, whose load is attached to a fully supported carriage. Rodless linear actuators provide linear motion via a motor-driven ball screw, ACME screw, or belt drive assembly; and (iii) electric rotary actuators that use a motor to drive components rotationally.

The next category (b) are the hydraulic and electrohydraulic valve actuators, which convert fluid pressure into motion. They include (i) linear actuators or hydraulic cylinders, which use a cylinder and hydraulic fluid to produce linear motion and force, and (ii) hydraulic rotary actuators that use pressurized hydraulic oil to rotate mechanical components.

The third category (c) are the linear solenoids that convert electrical energy into mechanical work via a plunger with an axial stroke in either a push or pull action. They can be rated for continuous duty (100% duty cycle operation, continuous duty solenoids) or for off-on applications, less than 100% duty cycle (intermittent duty solenoids).

The last category (d) are digital actuators, which now include all types of valve actuation solutions where the valve is digitally controlled.

Pneumatic actuators are still the technology favored by valve actuators buyers. Nevertheless, the market share of electric, hydraulic, and electrohydraulic actuators is increasing, while the overall use of solenoid values has dropped.[1]

Actuator Features

Speed and Torque Ranges Speed requirements vary from less than 1 rpm to about 160 rpm. The upper limit of available torque is about 250,000 ft · lb_f (339,000 N · m) with gearboxes.

TABLE 6.3a
Applications, Advantages, and Disadvantages of Various Actuator Designs[2]

Actuator Types	*Advantages*	*Disadvantages*	*Applications*
Electromechanical	High thrust High stiffness coefficient Powered by electricity or pneumatics	Complex design No mechanical fail safe Large, heavy structure	Linear or rotary valves 2–36 in. body size
Electrohydraulic	High thrust High stiffness Fast speeds Powered by electricity; no pneumatic source required	Complex design Large, heavy structure Hydraulic temperature sensitive	Linear or rotary valves 2 in. to unlimited
Electric (servomotor or stepping motor)	Direct interface with computer system	Large structure Low thrust No mechanical fail safe Slow speed	Linear valves $^1/_2$–2 in. body size

Stem thrusts and rotational drive torques are limited only by the size of the motor used and the ability of the gear, bearings, shafts, and so on to carry the load. Speed of operation depends on the gear ratios, adequate prime move power, and means of overcoming the inertia of the moving system for rapid stopping. This is most important for proportional control uses. Some actuators have a limited selection of drive speeds, while others are furnished in gear ratios in discrete steps of 8–20% between speeds.

Manual Operation Manual operation is sometimes necessary for normal operational procedures, such as start-up, or under emergency conditions. Only units that are rotating very slowly or those with low output should have continuously connected handwheels. Most units have the handwheel on the actuator with a clutch for demobilizing the handwheel during powered operation. Clutches are manual engage and manual disengage, or manual engage and automatic disengage upon release of the handwheel. Others are manual engage when the handwheel is rotated, with the motor re-engaging when the handwheel is not being rotated; or power re-engage, which takes the drive away from the operator upon energization, leaving the handwheel freewheeling.

Electrical Equipment Most of these actuators include much of the electric gear within the housings of the unit. Components such as limit, auxiliary, and torque switches and position or feedback potentiometers are run by gearing to stem rotation, so they must be housed on the unit. Installation is optional concerning pushbuttons, reversing starters, lights, control circuit transformers, or line-disconnect devices in an integral housing on the unit. Any or all of these components may be located externally, such as in a transformer, switch, or control house. The enclosures must be designed to satisfy NEMA requirements for the area.

High Breakaway Force Resistance to opening requires a method of allowing the motor and gear system to develop speed to impart a “hammer blow,” which starts motion of the valve gate or plug. Selection of motors with high starting torque is not always sufficient. The dogs of a dog-clutch rotate before picking up the load, or a pin on the drive may move within a slot before picking up the load at the end of the slot. Systems are used that delay contact for a preselected time or until the tachometer indicates the desired speed of rotation.

Torque Control for Shutdown Torque control for shutdown at closure or due to an obstruction in the valve body is accomplished in numerous ways, but each one uses a reaction spring to set the torque. When rotation of the drive sleeve is impeded, the spring will collapse, moving sufficiently to operate a shut-off switch.

Position Indication An indicator can be geared to the stem rotating gear, but it becomes a problem when actuator rotation varies from 90 to as much as 240 revolutions. Gearing of a cam shaft operating the position indicator or auxiliary switches must be calculated to obtain a fairly uniform angle. Upon correct gearing, an indicating arrow or transmitting potentiometer can be rotated.

Maintenance of Last Position This is no problem when the actuator includes a worm gear or stem thread. Use of spur gears can cause instability when positioning a butterfly or ball valve. Status quo is obtained by use of a motor brake or insertion of a worm gear into the system.

Protection against Stem Expansion The status quo ability of a thread or worm gear is detrimental when the valve itself is subjected to temperatures high enough to expand the stem. This expansion, when restrained, can damage the seat or plug, bend the stem, or damage the actuator thrust bearings. One of the original patents for this type of actuator included Belleville springs to allow the drive sleeve to move with the thermal expansion and relieve the linear force.

Mounting Methods Industry has dictated a set of dimensions for the mounting flanges and bolt holes for newly manufactured valves. Retrofit mounting requires adaptation to existing valves. Mounting requires a plate to match the existing valve, which is screwed into the yoke upon removal of the manual drive sleeve, welded or brazed to the yoke, or, for a split yoke, bolted to the yoke.

Adaptability to Control This feature includes adaptability to many voltages and to single- or polyphase supplies. Polyphase motors of 240–480 V and 60 Hz predominate. Single-phase motors up to about 2 horsepower (hp) (1,492 W) are used. Reversing starters with mechanical interlocks are used for both proportional and on-off service.

The coils that open and close the contacts are energized by an open-center double-pole switch, which can be incorporated in the automatic control circuit. For manual control, the starter may be of the type that maintains contact, requiring the open or close button to be held in position. Some units are also wired for momentary depression so that the actuator runs until it reaches the limit switch, or until a stop button is depressed.

Proportional Control Proportional control of these large units can be accomplished by including the coils of the reversing starter in a proportional control circuit. This requires a position feedback, which may be a potentiometer. For this type of control, a Wheatstone bridge circuit would be used.

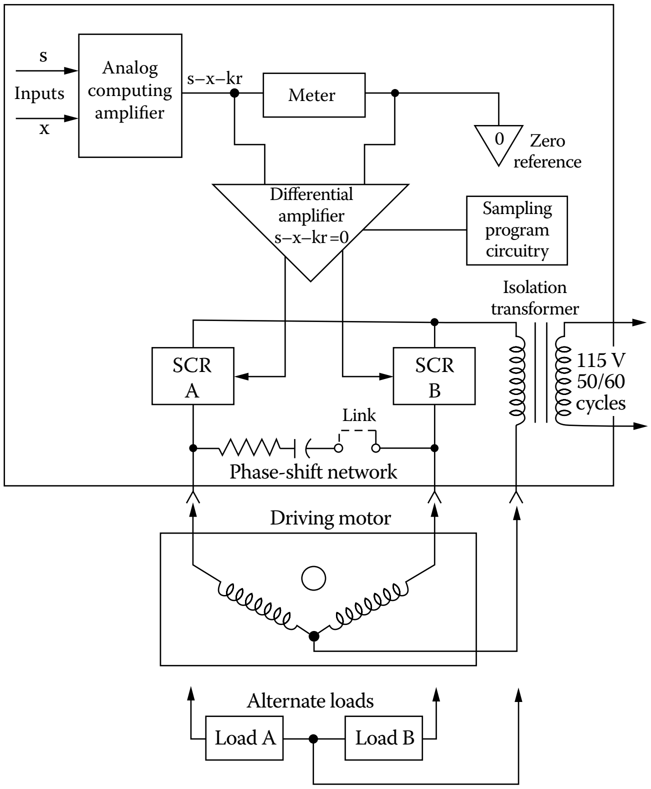
The reversing starter controls any voltage or phase required by the motor. A transducer to transform any of the accepted electronic controller outputs (e.g., 1–5 mA) to a resistance relative to controller output permits use of the actuator in these systems.

A smaller unit, with a force output of 1460 lb_f (6494 N) at stall and 500 lb_f (2224 N) at a rate 0.23 in./sec (5.84 mm/s), uses solid-state control of the motor. At the same time it eliminates stem position feedback into the controller.

A DC signal "*x*" (Figure 6.3b) from a process transmitter provides the loop with its measurement, such as temperature, as well as such high-responds systems as flow. A differential amplifier responds to the magnitude and the polarity of an internally modified error signal, which triggers silicon-controlled rectifiers (SCRs) to obtain bidirectional drive. The synchronous motor can be driven in either direction, depending upon the relation to the set point.

DIGITAL VALVE ACTUATORS

Digital actuators can accept the output of digital computers directly without digital-to-analog converters. Only simple on-off elements are needed for their operation. The number of output positions that can be achieved is equal to 2^n, where n is the number of inputs. Accuracy of any position is a function of the manufacturing tolerances. Resolution is established by the number of inputs and by the operating code selected for a given requirement. The smallest move achievable is called a 1-bit move.



s = instantaneous set point
x = instantaneous transmitter signal
k = adjustable transmitter compensation
r = instantaneous rate of change of transmitter signal

FIG. 6.3b
Proportional motor control circuit with position feedback.

The code may be binary, complementary binary, pulse, or special purpose. A three-input piston adder assembly produces eight discrete bit positions. The adders in Figure 6.3c are shown in the 6-bit extended position. The interlocking pistons and sleeves will move when vented or filled through their selector valves. This same adder can be used to position a four-way spool valve, with a mechanical bias to sense position. The spool valve controls the position of a large-diameter piston actuator or force amplifier.

Use of a DC motor featuring a disc-armature with low moment of inertia has created another valve actuator particularly adaptable to a digital input. Brushes contacting the flat armature conduct current to the armature segments. Incremental movement is caused by half-waves at line frequency for rotation in either direction. The rotation of the armature is converted to linear stem motion by use of a hollow shaft internally threaded to match the valve stem.

Actuator output is 5,000 lb_f (22,240 N) maximum for noncontinuous service at a rate of 0.4 in./sec (10 mm/s) through a valve stroke of 3 in. (76.4 mm). The actuator is de-energized at stroke limits or at power overloads by thermal overload relays.

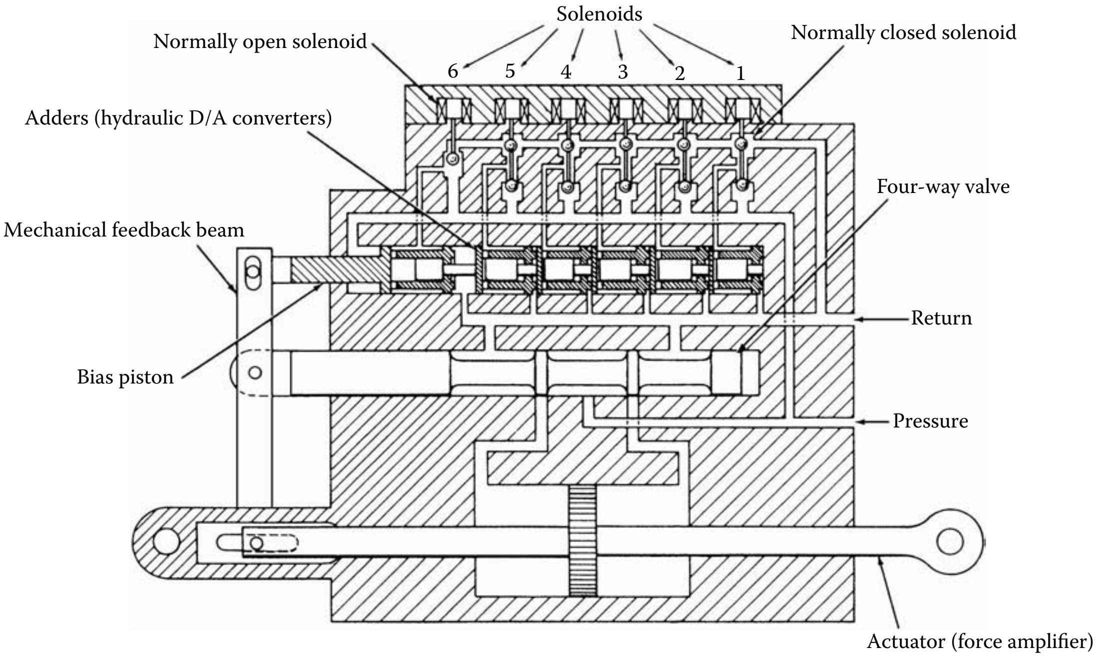


FIG. 6.3c
Digital valve actuator.

Operation of the actuator requires application of a thyristor (SCR) unit designed for this purpose. This unit accepts pulses from a computer or pulse generator. The thyristor unit consists of two SCRs with transformer, triggers with pulse shift circuit, facility for manual actuator operation, and the previously mentioned thermal overload relays. The output consists of half-waves to pulse the armature of the actuator.

Modules are also available for process control with the necessary stem position feedback, slow pulsing for accurate manual positioning, and full-speed emergency operation.

The digital-to-pneumatic transducer shown in Figure 6.3d is used to convert a controller output to a pressure signal for operating a pneumatic valve.

The concept of digital valve actuator has been extended to other valve actuating systems such as stepping motors, which are discussed in the following paragraph dedicated to electromechanical actuators, and to the so-called smart or intelligent actuators briefly introduced below in Smart and Intelligent Actuators. The discussion of digital actuators in this section is intended to complement Section 6.12 on intelligent valves and Section 6.18 covering digital control valves.

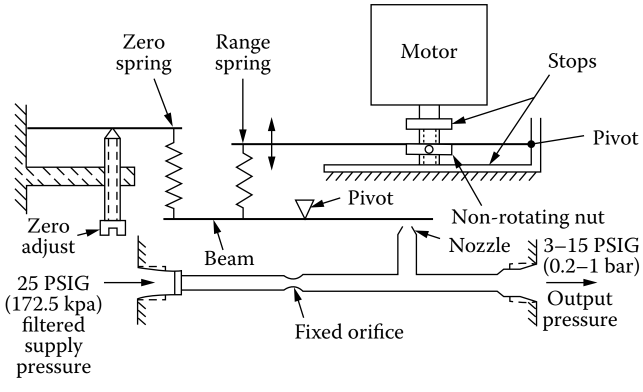


FIG. 6.3d
Digital-to-pneumatic transducer.

It is in this context that some of the suppliers in the partial list of suppliers at the beginning of this section identified as providing digital valve actuators indeed provide modern smart valve actuating solutions.

ELECTROMECHANICAL ACTUATORS

Electric actuators can utilize a reversible electric motor provided with an internal worm gear to prevent drive direction reversal (back-drives) by unbalanced loads. These units can operate both linear and rotary valves. Servomotor drives can position valves in response to feedback signals from linear or rotary encoders. Two-phase AC servomotors are available with up to 1 HP rating, while direct current servomotors can meet higher loads.

Stepping motors, which rotate the shaft by a discrete step angle when energized electrically, can also be used in valve actuators. The electromechanical device that rotates the shaft can be a solenoid used to operate a star-wheel or ratchet device; it can be the stepping movement of a permanent magnet, the flux of which causes poles or teeth to align and thereby affect rotation; or it can be a variable reluctance unit, where the rotor-stator poles or teeth are aligned by electric fields.

When used as valve actuators, the stepping motors are well suited for direct digital control, and pulse feedback can be provided for accurate closed-loop positioning.

Of the above designs the actuators using reversible motors are the most often used variety and therefore they will be discussed in more detail in the following paragraphs.

Reversible Motor Gear Actuators

Lever-operated valves and dampers are available with torques of 5–500 ft · lb_f (6.8–678 N · m) and with full stroke speeds of 10–60 seconds. For very large valves or dampers, this type of actuator can deliver torques of 150–10,000 ft·lb_f (203–13,558 N · m) and a full stroke speed of 30–300 seconds.

The power consumption of these larger units ranges from 0.5–5 kVA operating on 460 V three-phase power supply. Some of these designs can be obtained with springs to return the valve to a safe condition upon electric failure.

The motor gear actuator consists of an electric motor connected to a gear train; the gear train rotates a stationary drive nut, which in turn drives the threaded or keyed valve stem up or down. As the valve plug contacts the seat, the resistance is transmitted to a Belleville spring, which at a preset limit interrupts the motor power circuit.

For throttling applications, the actuator can be provided with a positioner that compares the external control signal (analog or digital) with an internal position feedback signal and keeps turning the motor until the error between the two signals is eliminated.

Rotary Output Actuators

Worm Gear Reduction The actuator shown in Figure 6.3e is an example of the use of a double worm gear reduction to obtain output speeds of around 1 rpm with an input motor speed of 1800 rpm. The worm gear is self-locking, so it prevents the load from moving downward by back-driving the motor. However, the worm gear is less than 50% efficient, so more power is used compared to spur gears. Also shown is a handwheel for manual operation during power loss.

Spur Gear Reduction The spur gear actuator in Figure 6.3f has very low power loss through the relatively efficient spur gears. Ordinarily the load could back-drive this system, but a friction member at the motor end of the gear train minimizes this undesirable action. Of course, if the load has a frictional characteristic, it will not impose a back-driving torque on the actuator.

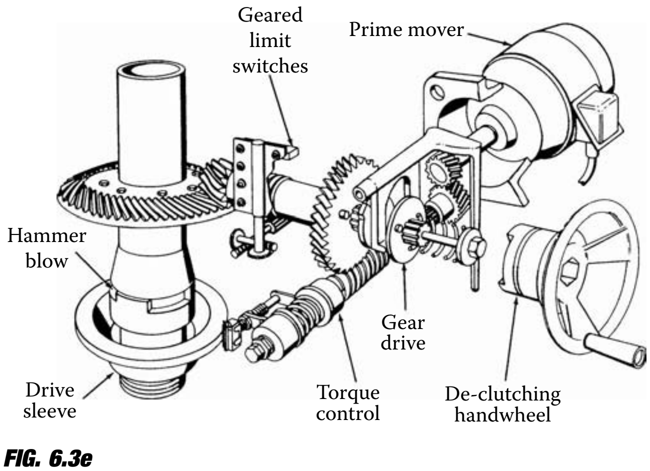


FIG. 6.3e
Electric actuator with worm gear reduction.

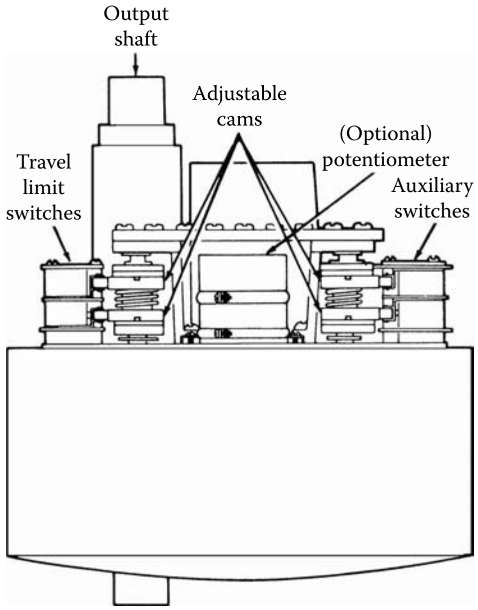


FIG. 6.3f
Spur gear reduction.

Actuators for larger outputs are externally mounted motors. Two motors operating a single gear train have been used to obtain 3,750 ft · lb_f (5084 N · m) of torque through 90° rotation in 75 sec. An actuator has been designed for accurate rotary positioning that develops 5000 ft·lb_f (6800 N · m) at stall and will rotate 90° in 10 sec. SCRs energize the motor as commanded by a servotrigger assembly housed separately.

Electrical gear may include adjustable cams to operate limit and auxiliary switches. A potentiometric feedback calibrated to the rotation of the actuator is required for use with a control circuit. The unit shown in Figure 6.3f was developed to slide over and be keyed to the shaft of a boiler damper. Actuators of this type must be adaptable to mounting on and operating a variety of quarter-turn valves. Because of the difficulty in setting limit switches to accurately stop the valve in the shut position, it is advisable to incorporate a torque-limiting device to sense closure against a stop.

Opening can be controlled by a limit switch. One compact unit contains the features noted with an output of 750 ft · lb_f (1017 N · m) of torque. The unit is powered by a motor with a high-torque capacitor and includes a mechanical brake, feedback potentiometer, limit switches, and a de-clutchable hand wheel.

Flex-Spline Reduction Figure 6.3g is an example of a unique single-stage, high-reduction system. Instant breakaway and efficient transfer of prime mover power is obtained

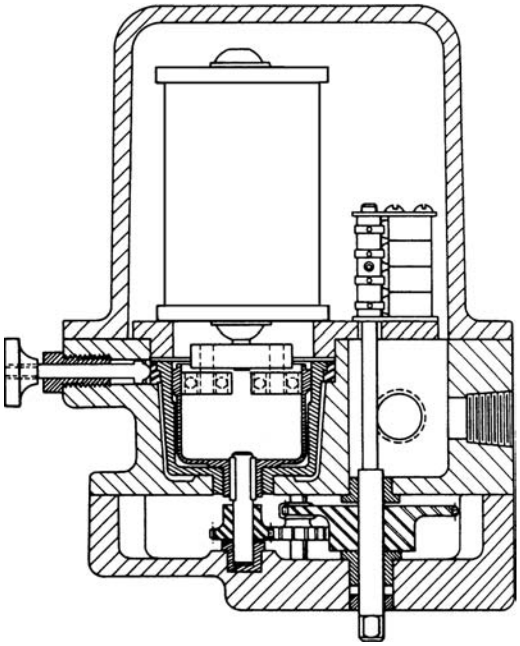


FIG. 6.3g
Flex-spline gear reduction.

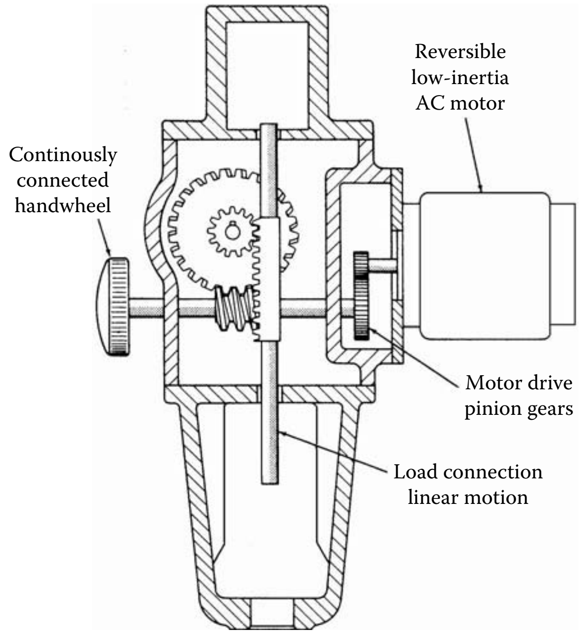


FIG. 6.3h
Rack and pinion assembly converts the rotation of a horizontal, motor-driven shaft into vertical, linear motion.

with a modified concentric planetary system consisting of a semiflexible gear within a rigid gear. A three-lobed bearing assembly transmits power to the gears by creating a deflection wave transmission that causes a three-point mesh of the gearing teeth on 30% or more of the external gearing surface.

The semiflexible geared spline has fewer teeth than the nonrotating internal gear it meshes with. The spline slowly rotates as it is pressed into the larger gear by the bearings on the motor shaft.

Linear Output Actuators

Motors and Rack Figure 6.3h uses a worm and a rack and pinion to translate horizontal shaft motor output to vertical linear motion. Maximum force output is approximately 1500 lb_f (6672 N) at about 0.1 in./min (2.5 mm/min). A continuously connected handwheel, which must rotate the rotor of the motor, can be used when there is short stem travel and relatively low force output.

The actuator is designed with a conventional globe valve bonnet for ease of mounting. Units operate on 110 V and have been adapted to proportional use with a 135 ohm Wheatstone bridge or any of the standard electronic controller outputs.

Motor and Travelling Nut Linear unit consists of the motor, gears, and a lead screw that moves the drive shaft (Figure 6.3i). A secondary gear system rotates cams to operate limit and auxiliary switches.

The unit may have a brake motor for accurate positioning and a manual handwheel. The bracket on the rear end allows the actuator to rotate on the pin of a saddle mount, so that the drive shaft can be pinned directly to the lever arm of a valve or a lever arm fixture that fits the shaft configuration of a plug cock or ball valve. Maximum output is 1600 lb_f (7117 N) at 5 in./min (127 mm/min).

Rotating Armature An internally threaded drive sleeve in the armature of the motor is used to obtain a linear thrust up to 6,600 lb_f (29,360 N) at a rate of 10 in./min (254 mm/min). Bearings in the end cap support the drive assembly (Figure 6.3j). The drive stem is threaded to match the drive sleeve and is kept from rotating by a guide key.

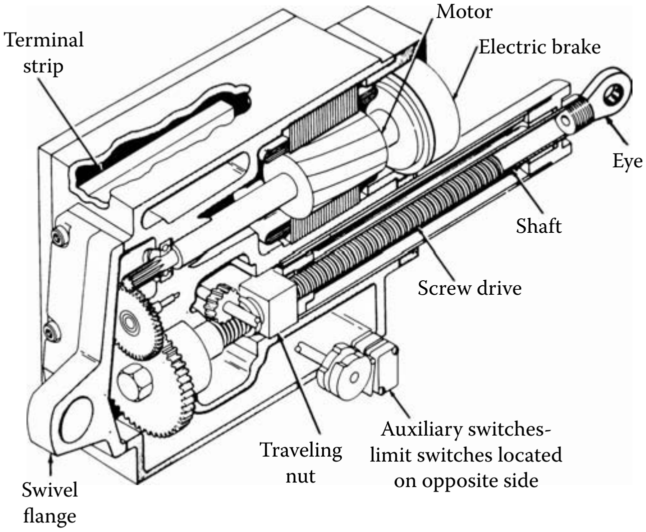


FIG. 6.3i
Electric quarter-turn actuator with linear output and with limit switches.

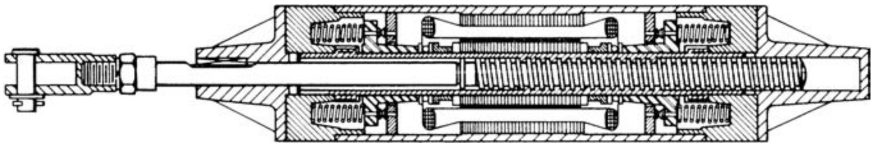


FIG. 6.3j
Electric actuator with rotating armature has been adapted for proportional control.

Thrust-limit switch assemblies are mounted in each end of the housing to locate the hollow shaft in mid-position. When the linear movement of the drive stem is restricted in either direction, the limit switch involved will operate to shut down the unit. Thermal cut-outs in the motor windings offer additional overload protection. Strokes are available from 2–48 in. (51–1219 mm).

The unit has been adapted for proportional control by use of an external sensing position for feedback. For use as a valve actuator, it must be mounted so that the drive stem can be attached to the valve stem, or a suitably threaded valve stem must be supplied.

Rotary to Linear Motion An electric proportional actuator (Figure 6.3k) is designed for continuous rotation of a drive-sleeve on a ball-screw thread. 3,000 lb_f of thrust (13,335 N) is obtained at a stem speed of 1 in./min (25 mm/min).

One or two DC signals are used separately or numerically added or subtracted. Triacs operate on position error to control a DC permanent magnet motor that positions a stem within an adjustable dead band. Degree of error and rate of return are sensed by a lead network to determine the direction and time that the motor must run.

Stem reversals are almost instantaneous. The back emf of the motor is used as a velocity sensor and is fed into a circuit that allows adjustment of the speed of drive sleeve rotation. Gain can be adjusted to control oscillation of the stem.

Stem position feedback is by a linear variable differential transformer (LVDT). Use of the DC motor allows for torque control through sensing of motor current. A manual handwheel is furnished that can only be used when the unit is de-energized by a manual/automatic switch.

Application of the linear requirement of a valve necessitates a linkage for translation from rotary to linear motion. Use of a linkage (Figure 6.3l) provides a thrust for operating the valve.

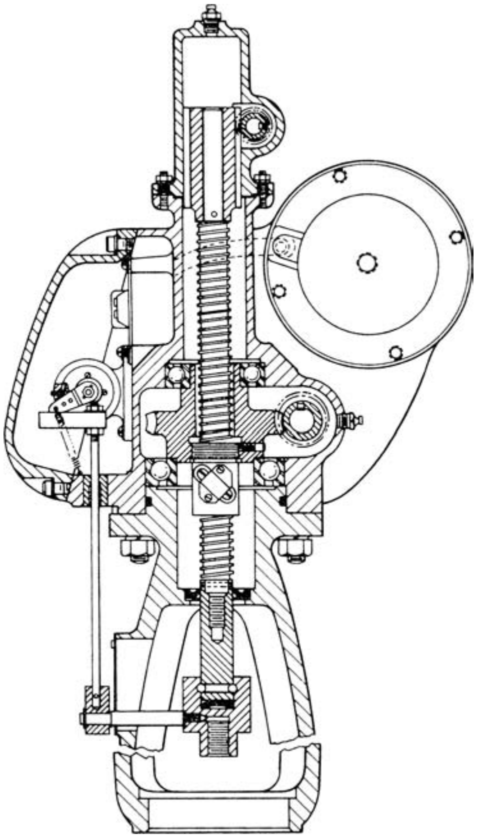


FIG. 6.3k
Proportional electrical drive converts rotary to linear motion.

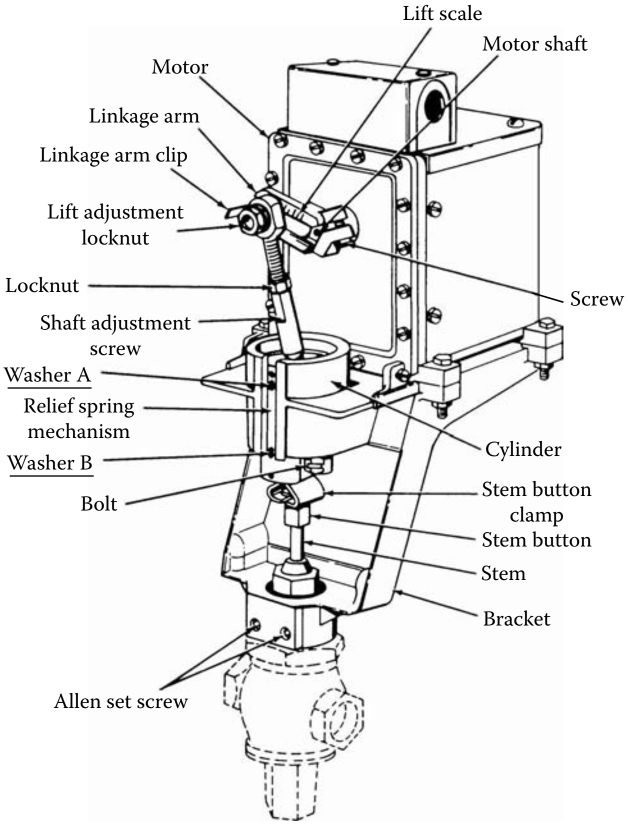


FIG. 6.3l
Electric actuator with linkage to convert rotary motion to linear.

ELECTROHYDRAULIC ACTUATORS

The main features of electrohydraulic actuators have been summarized in Table 6.3a. They are often used where instrument air is unavailable or where the required actuator stiffness or thrust cannot be obtained from pneumatic actuators. They are heavier than pneumatic actuators; therefore, in order to avoid straining the bonnet, vertical upright installations are recommended.

Electrohydraulic actuators are superior to electromechanical ones in the areas of speed, positioning without overshoot, actuator stiffness, and complexity. Electrohydraulic actuators can be stepping motor or servovalve driven. In the servovalve designs the pumps are running continuously, while in the stepping-motor configuration they run only when the valve needs repositioning.

Electrohydraulic actuators can be mounted directly onto the stems of valves that are 6 in. (152 mm) or smaller and can be provided with springs to guarantee a fail-safe position in case the hydraulic fluid is lost. The actuator usually consists of a hydraulic cylinder, a pump with motor, some feedback linkage on valve position, and a balancing arrangement. The balancing or positioning mechanism compares the external control signal with the valve position and actuates the hydraulic system if repositioning is required.

In spring-loaded designs the hydraulic fluid might drive the valve in one direction (opening it, for example), while the spring drives it in the other direction (closing it, for example). Open, closed, or last-position failure positions are available. When higher thrusts, longer strokes, or rotary valves are involved, the fail-safe position can be provided by storing energy in a hydraulic accumulator instead of using a spring.

Hydraulic actuators are sensitive to viscosity changes caused by ambient temperature variations and, therefore, in subfreezing temperatures must be provided with heaters, which can require substantial heat-up periods before the valve can be operated. Because the hydraulic fluid is incompressible, the actuator is "stiff" and provides stable and accurate positioning with hysteresis and dead band within 5% of span.

Their speed of operation is similar to pneumatic actuators—0.125–0.25 in. (3.2–6.4 mm) per second of stem movement—but can be speeded up substantially by the use of high-pressure accumulators. Intelligent, programmable electrohydraulic actuators can be provided with bidirectional hydraulic gear pumps, which are driven by microprocessor-controlled stepping motors.

External Hydraulic Source

The term "electrohydraulic" has been applied to actuator systems in which the hydraulic pressure to one or more actuators is supplied by a hydraulic mule. The hydraulic power is supplied to the actuator by electrical control means.

In the broad sense, the use of two three-way solenoids or one four-way solenoid externally mounted to the actuator constitutes an electrohydraulic system.

More extensively, "electrohydraulic" applies to a proportionally positioned cylinder actuator. This requires a servo-system, which is a closed loop within itself. A servo-system requires one of the standard command signals, which is usually electrical but can be pneumatic. This small signal, which often requires amplification, controls a torque motor or voice coil to position a flapper or other form of variable nozzle. This positions a spool valve or comparable device to control the hydraulic positioning of a high-pressure second-stage valve.

The second-stage valve directs operating pressure to the cylinder for very accurate positioning. Closing the loop requires mechanical (Figure 6.3m) or electrical feedback to compare the piston position with the controller output signal.

The electrical feedback can be a servoamplifier, illustrated in connection with a linear hydraulic actuator shown in Figure 6.3n.

Hermetically Sealed Power Pack

A much more compact electrohydraulic actuator combines the electrohydraulic power pack with the cylinder in one package. Many of these actuators are designed as a truly integral unit. An electric motor pump supplies high-pressure oil through internal ports to move the piston connected to the stem (Figure 6.3o).

The small magnetic relief valve is held closed during the power stroke until de-energized by an external control or

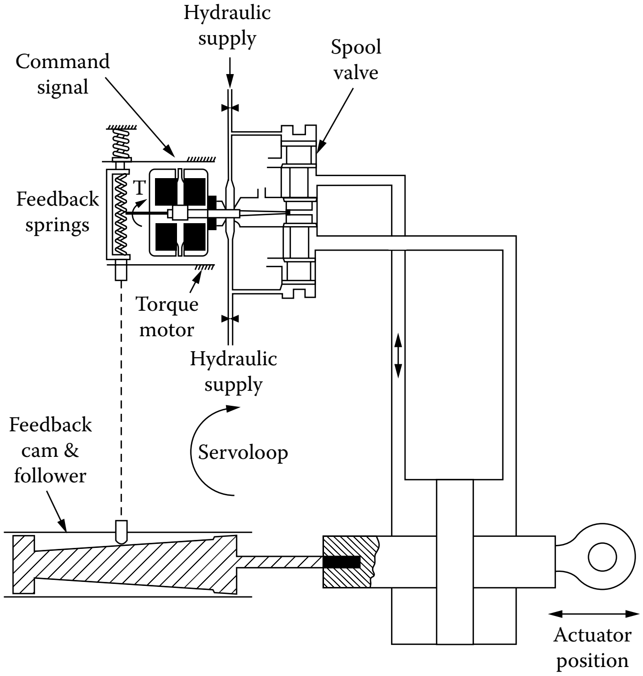


FIG. 6.3m
Two-stage servovalve with mechanical feedback.

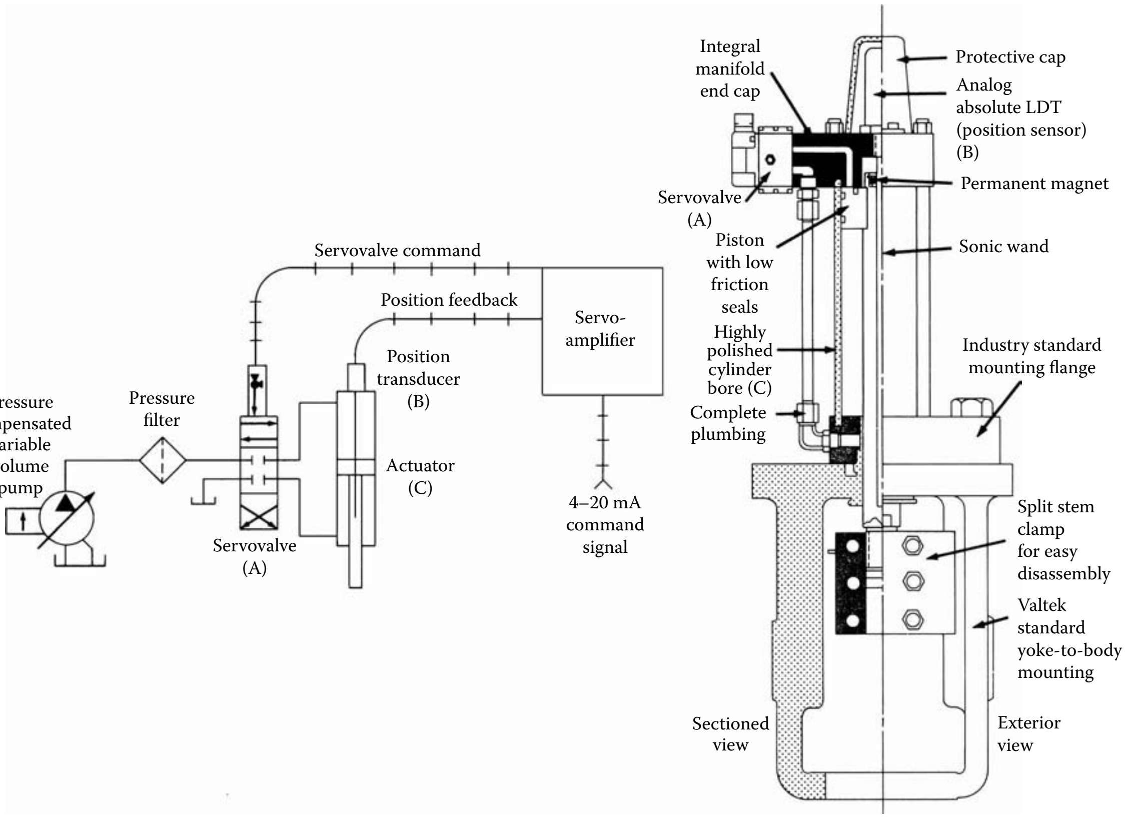


FIG. 6.3n
Servovalve-operated electrohydraulic linear valve actuator. (Courtesy of Valtek, Flowserve Corp.)

emergency circuit to allow the spring to cause a "down" stroke. The same unit can be used to cause the spring to return to the up position using a Bourdon switch to produce force limit.

Motor and Pump Combinations

Reversible Motor and Pump A reversible motor can be used to drive a gear pump in a system to remove oil from one side of the piston and deliver it to the other side (Figure 6.3p).

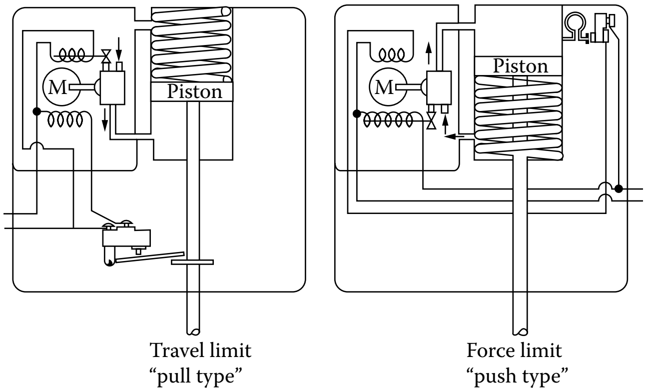


FIG. 6.3o
Hermetically sealed electrohydraulic power pack.

The check valves allow the pump to withdraw oil from the reservoir and position the directional control valve in order to pressurize the cylinder. Reversing the motor (and pump) reverses the direction.

When the motor is de-energized, the system is "locked up." For proportional control, feedback is necessary from stem position to obtain a balance with the control signal.

Jet Pipe System A very old control system for a cylinder, the jet pipe, is employed in an electrohydraulic actuator. An electromechanical moving coil in the field of a permanent magnet is used to position a jet that can direct oil to one end or the other of the cylinder actuator (Figure 6.3q). Force-balance feedback from stem position creates the balance with the controller signal.

Hydraulic Control of Pinch Valves Controlled hydraulic positioning of a sleeve valve is obtained with a moving coil and magnet to position the pilot (Figure 6.3r), which controls pressure to the annular space of the valve. Feedback is in the form of a Bourdon tube, which senses the pressure supplied to the valve and moves the pilot valve to lock in that pressure.

Multiple Pump The multiple pump system consists of three pumps running on the shaft of one prime mover (Figure 6.3s). There is one pump for each side of the piston and one for

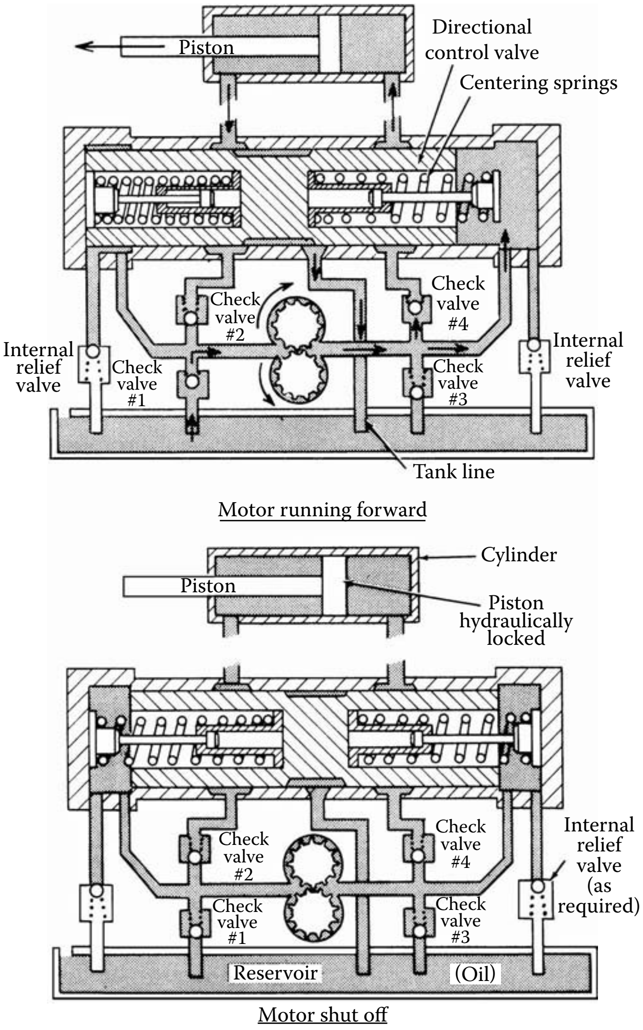


FIG. 6.3p
Pump with reversible motor combination, where the actuator is "locked" when the pump is de-energized.

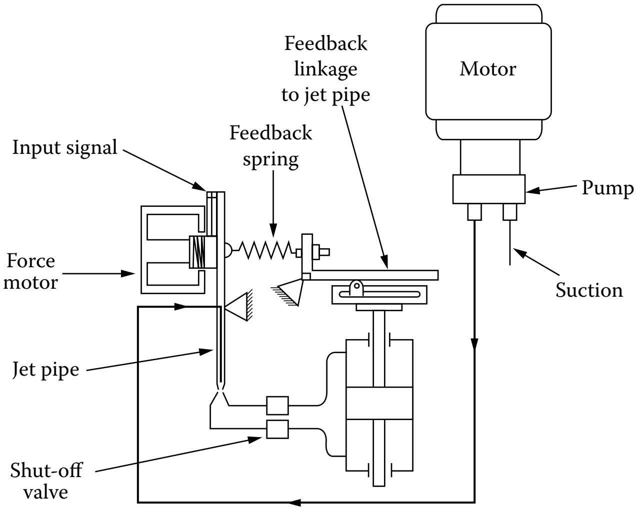


FIG. 6.3q
Electrohydraulic actuator with jet pipe control.

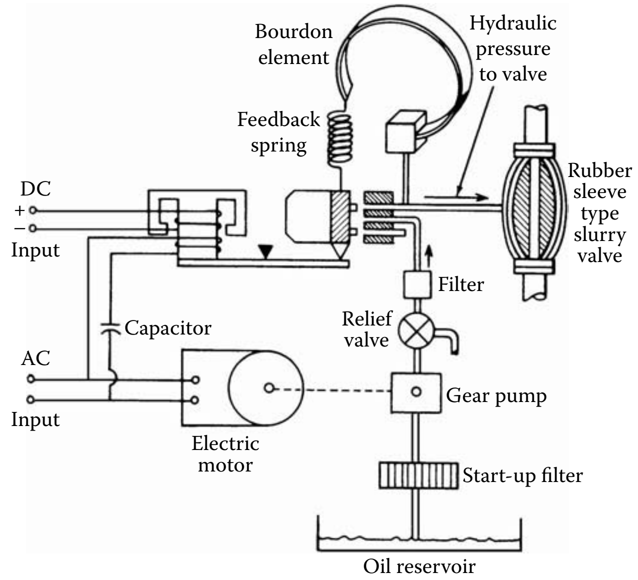


FIG. 6.3r
Electrohydraulic control of a jacketed pinch valve.

the control circuit. The force motor tilts a flapper to expose or cover one of two control nozzles. The flow through a restricting nozzle allows pressure to be transmitted to one side of the piston or the other.

Force-balance feedback is created by a ramp attached to the piston shaft, which positions a cam attached to the feedback spring. Upon loss of electric power, the cylinder shut-off valves close to lock up the pressure in the cylinder and assume a status quo. A bypass valve between the cylinder chambers allows pressure equalization to make use of the manual handwheel.

Vibrating Pump A vibratory pump is used in a power pack mounted on a cylinder actuator (Figure 6.3t). A 60 Hz alternating source causes a plunger to move toward the core, and, upon de-energizing, a spring returns it. The pump operates on this cycle and continues until the piston reaches the end of its stroke, when a pressure switch shuts it off.

The solenoid that retains the pressure is de-energized by an external control circuit. Maximum stem force is 2,500 lb_f (11,120 N) at a rate of about 1 in./min (25 mm/min), or 5,000 lb_f (22,240 N) at 0.3 in./min (7.6 mm/min).

Two-Cylinder Pump A two-cylinder pump, driven by a unidirectional motor, injects pressure into one end of a cylinder or the other, depending upon the positions of two solenoid relief valves (Figure 6.3u). The solenoid on the left is closed to move the piston to the right, with hydraulic pressure relieved through the other relief valve. Motion continues until the valve it is operating is seated.

The build-up of cylinder pressure operates the pressure switch at a predetermined setting to de-energize the motor and both solenoid relief valves. This locks the hydraulic

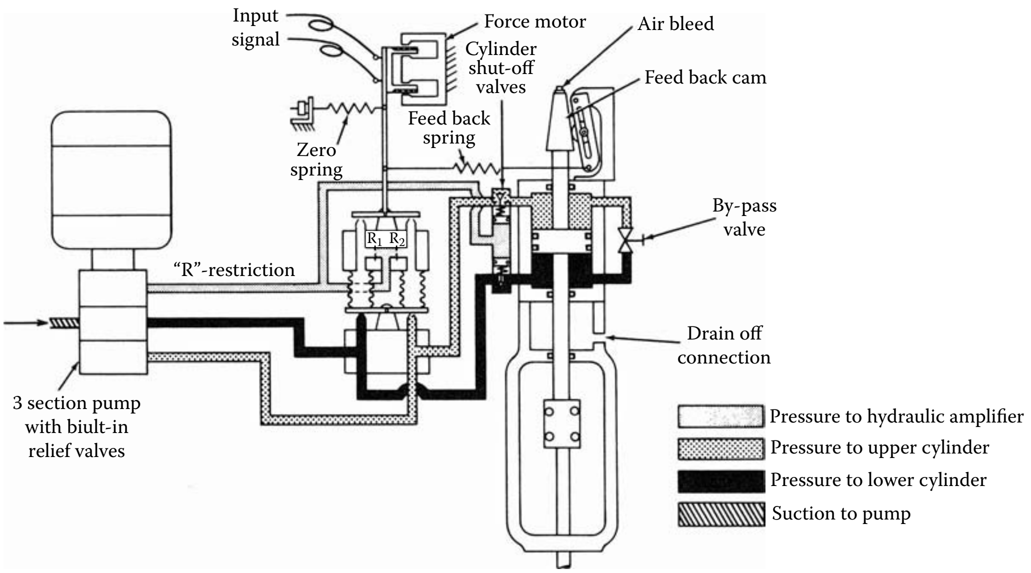


FIG. 6.3s
Hydroelectric valve actuator with three pumps running on the shaft of one mover.

pressure in the cylinder. Switching of the three-way switch will start the motor and reverse the sequence of the relief valves to move the piston to the left. At full travel (which is the up position of a valve stem), a limit switch shuts down the unit. Open-centering the three-way switch at any piston position will lock the piston (and valve stem) at that point.

Remote control is accomplished by manipulating the open-center switch. Automatic control is acknowledged by including this open-center function, which may be solid state, in the control circuit. A potentiometer or LVDT that senses the stem position is required for feedback.

Stem output is 6,000 lb_f (26,690 N) at a rate of 3 sec/in. (0.12 sec/mm). Stem travels up to 7 in. (178 mm) are available,

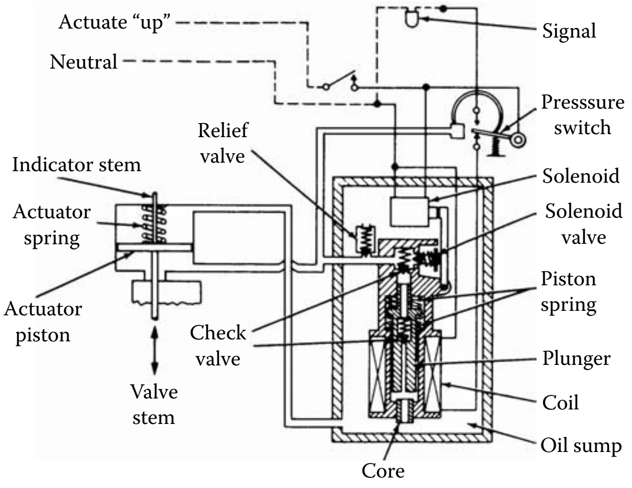


FIG. 6.3t
Electrohydraulic actuator with vibratory pump.

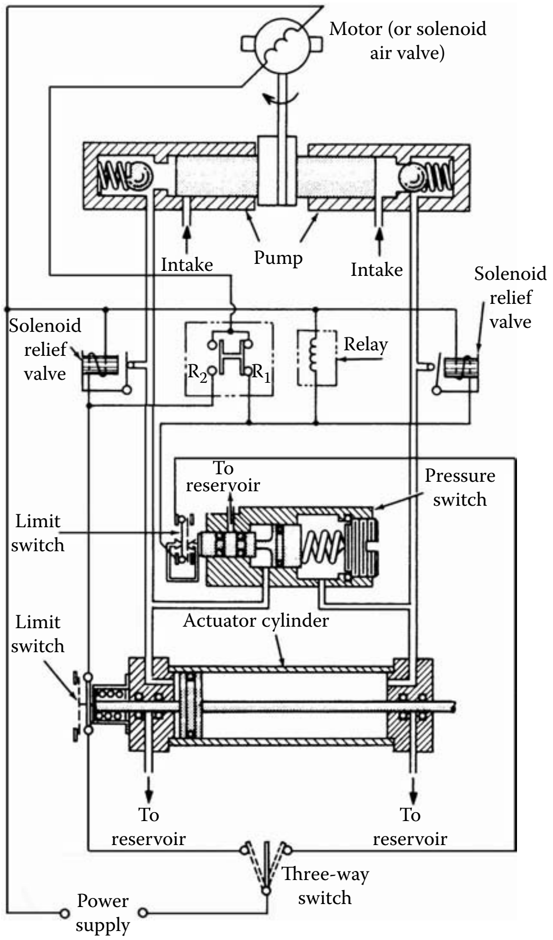


FIG. 6.3u
Two-cylinder pump-type electrohydraulic actuator.

although longer travels are feasible. The entire system is designed in a very compact explosionproof package that can be mounted on a variety of valve bonnets. The speed of response to energizing and de-energizing the control circuit makes it feasible to adapt the unit to digital impulses.

SOLENOID VALVES

Solenoids move in a straight line and therefore require a cam or other mechanical converter to operate rotary valves. These actuators are best suited for small, short-stroke on-off valves, requiring high speeds of response. Solenoid-actuated valves can open or close in 8–12 ms.

Their fast closure is not always an advantage: In water systems, it can cause water hammer. They are limited to pressure drops below 300 PSIG (20.7 bars), although when provided with pilots, levers, or double seats, they can handle higher pressure drops. They are available in two- or three-way designs, with power requirements ranging from 10–30 W with 6–440 VAC or 6–115 VDC power supplies. Solenoids are reliable devices, and they can provide multimillion cycles on liquid service.

Solenoids (consisting of a soft iron core that can move within the field set up by a surrounding coil) are used extensively for moving valve stems. Although the force output of solenoids may not have many electrical or mechanical limitations, their use as valve actuators has economic and core (or stem) travel limitations, and they are expensive.

A solenoid valve consists of the valve body, a magnetic core attached to the stem and disc, and a solenoid coil (Figure 6.3v). The magnetic core moves in a tube that is closed at the top and is sealed at the bottom; this design eliminates the need for packing. A small spring assists the release and initial closing of the valve. The valve is electrically energized to open.

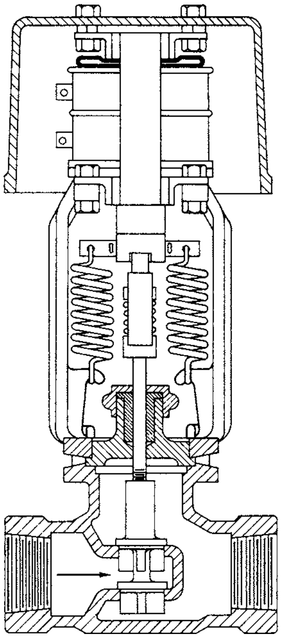


FIG. 6.3w
When valve packing is required, the increased friction can be overcome by a solenoid valve design with two or more strong return springs.

Stronger springs are used to overcome the friction of packing when it is required (Figure 6.3w). Reversing the valve plug results in reverse action (open when de-energized).

Even stronger stroking force can be obtained by using the force amplification effect of a mechanical lever, in combination with a strong solenoid (Figure 6.3x).

Using a solenoid to open a small pilot valve (Figure 6.3y) increases the port size and allowable pressure drop of solenoid-operated valves.

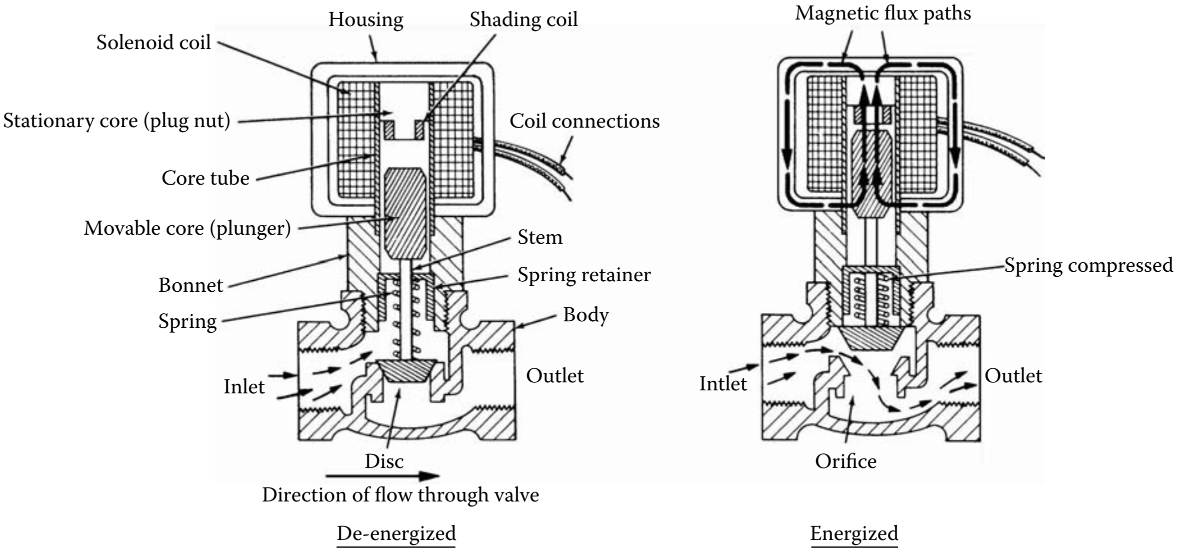


FIG. 6.3v
The operation of a direct-acting solenoid valve involves the lifting of the plunger when energized.

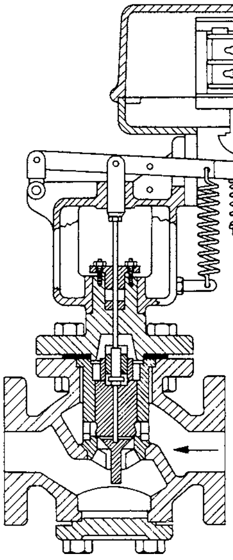


FIG. 6.3x
The force generated by the solenoid is amplified in the lever-type actuator.

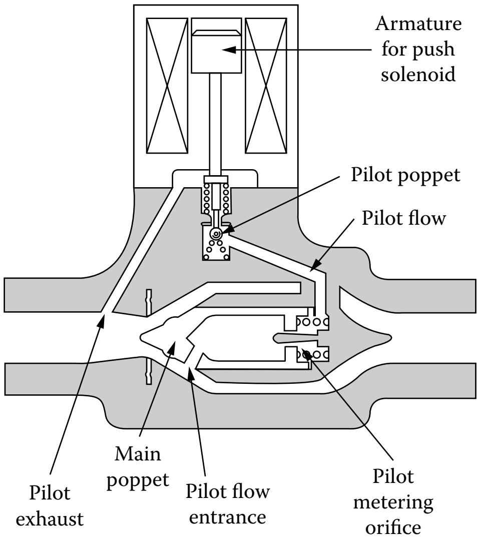


FIG. 6.3z
Pilot-operated in-line solenoid valve, where the pilot serves to apply pressure to the main poppet piston.

Small solenoid pilot valves are widely employed to supply pressure to diaphragms or pistons for a wide range of output forces. Pilot operation applies pressure to a diaphragm or piston or may release pressure, allowing the higher upstream pressure to open the valve. A good example is the in-line valve (Figure 6.3z).

Most solenoid valves are designed to be continually energized, particularly for emergency shutdown service. Thus the power output is limited to the current whose I^2R-developed heat can be readily dissipated.

Using a high source voltage and a latch-in plunge overcomes the need for continuous current. The single-pulse valve-closing solenoid is disconnected from the voltage source by a single-pulse delatch solenoid and hence does not heat up after it is closed. A pulse to the delatch solenoid permits the valve to be opened by a spring.

Three-way solenoid valves with three pipe connections and two ports are used to load or unload cylinders or diaphragm actuators (Figure 6.3aa). Four-way solenoid pilot

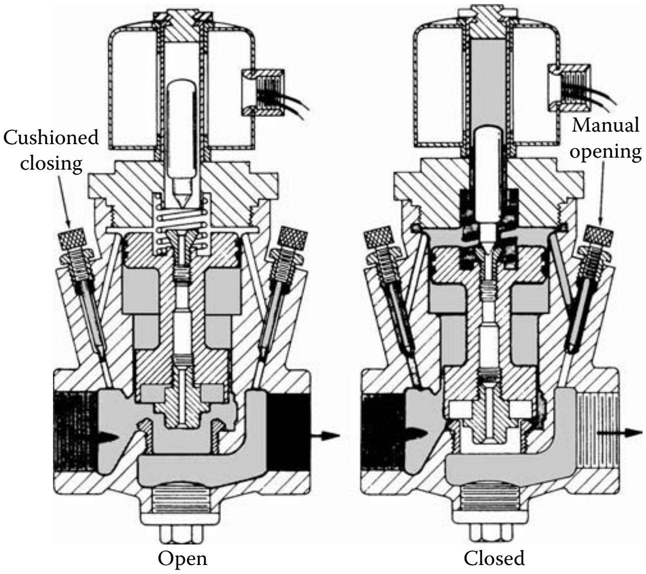


FIG. 6.3y
The solenoid valve can close against higher pressure drop, if the solenoid operates a small pilot valve.

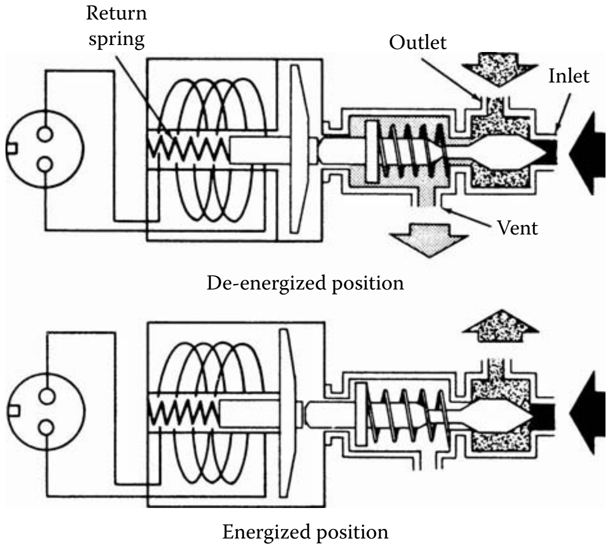


FIG. 6.3aa
The operation of a normally closed three-way solenoid valve.

valves are used principally for controlling double-acting cylinders.

Modulating Solenoid Valves

Modulating magnetic valves (Figure 6.3bb) utilize spring-loaded low-power solenoids to provide throttling action. The only moving part in this design is the valve stem, which has the valve plug attached to one end and an iron core attached to the other.

The valve opening is thereby a function of the voltage applied across the solenoid. (Figure 6.3cc illustrates the change in plunger force as a function of applied voltage and plunger air gap.)

Such throttling solenoids are frequently used in the HVAC industry and in other applications where the valve actuators do not need to be very powerful. The actuator thrust requirements are lowered by balancing the inner valve through the use of pressure equalization bellows or floating pistons. The positive failure position of these valves is provided by spring action.

Throttling solenoids are typically available in $^1/_2$–8 in. (127–203 mm) sizes and are limited by the pressure difference against which they can close. They are available in two- or three-way designs and can handle water flows up to 5000 gpm (19 m^3/m).

The use of throttling solenoids can completely eliminate the need for instrument air in the control loop. These valves

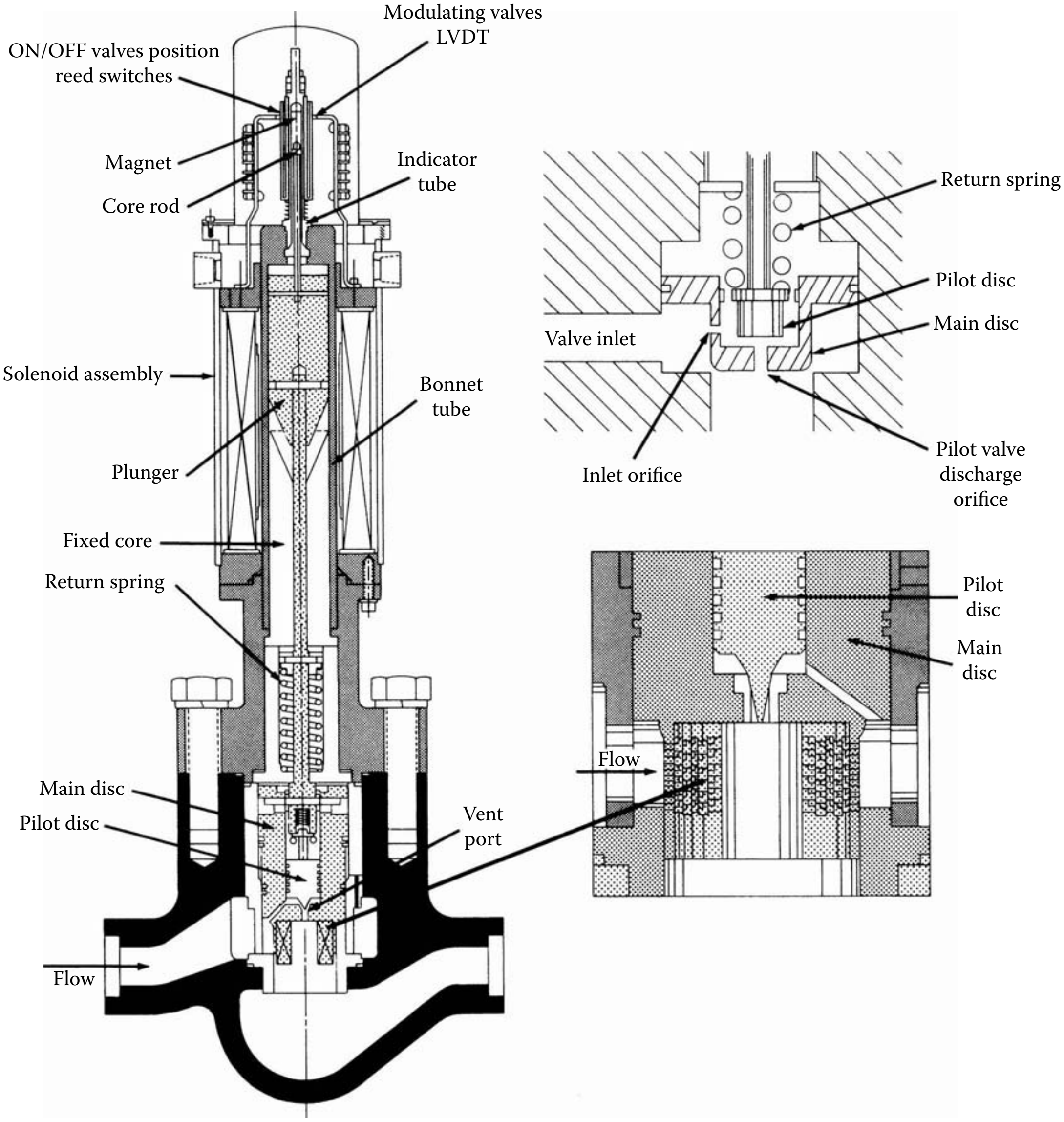


FIG. 6.3bb
Throttling solenoid valve design with LVDT used to measure the stem position for feedback.

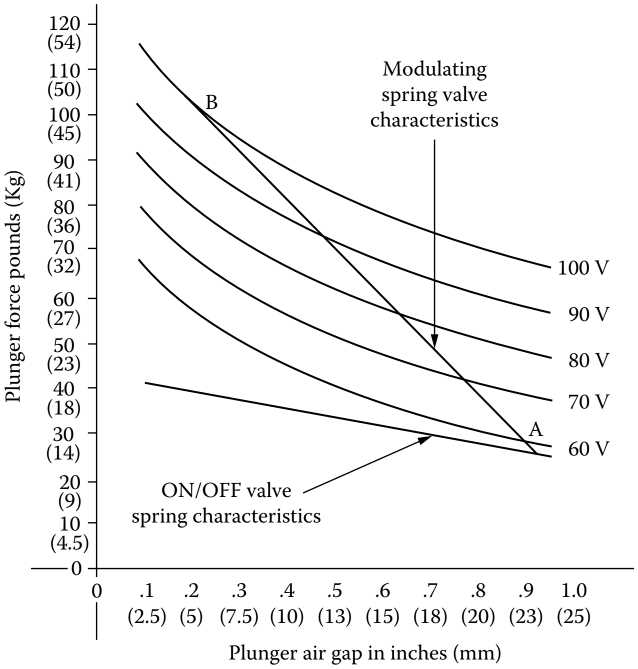


FIG. 6.3cc

The force available to the pilot plunger of a throttling solenoid depends on the voltage and the air gap.

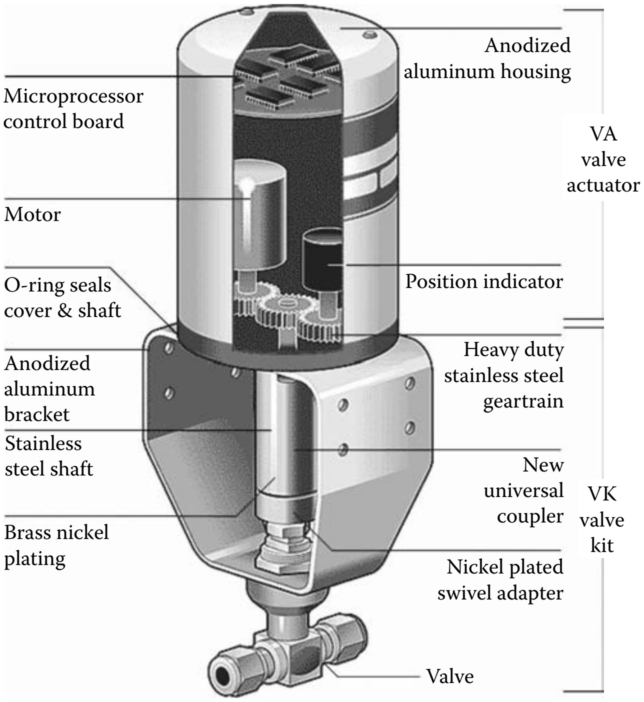


FIG. 6.3dd

Microprocessor-based electric valve actuator. (Courtesy of ETI Systems, Inc.)

offer a higher speed and better range than their pneumatic counterparts. Some manufacturers claim a range of 500:1 and a stroking time of 1 sec. The design illustrated in Figure 6.3bb can be driven directly from microprocessor-based building automation systems.

The design shown in Figure 6.3bb also includes a separate positioner, which accepts a 4–20 mA DC input from the controller and delivers a DC output signal to the throttling solenoid. Valve position feedback is obtained through the use of a linear variable differential transformer mounted directly on the valve.

SMART ACTUATORS

In addition to specifying the thrust, travel, failure position, control signal, power supply, speed, electrical area classification, and ambient conditions that an actuator must meet, one can also consider the use of microprocessor-based smart systems and of some other types of intelligent systems.

One of the trends in process automation is the use of distributed control and, in some cases, to locate the controller as close as possible to the final control elements (control valves). Often the control signal is digital (digital valve control, or DVC[3]). It is in this respect that an increasing number of features have been added to valves, valve actuators, and positioners through the integration of microprocessors with them.

Applications

Microprocessors have been included in valve actuators and valve positioners for some time (Figure 6.3dd), but the degree of exploitation of the possibilities of the smart and intelligent actuators for the purposes of process control is highly dependent on data access and thus on digital networking capabilities.

The tendency is to provide new actuators with some sort of network connection.[4] As discussed in more detail in Sections 4.16 and 6.11, Foundation fieldbus, HART, and Profibus are some of the leading suppliers. Open architectures, such as Foundation fieldbus, have made positive contributions, because by broadening the market they allowed more users to take advantage of better performing actuating systems at lower prices. Moreover, the compatibility problems between different fieldbus systems are in the process of being resolved, and it is hoped that through international standardization "plug and play" capabilities[3] will soon be available for such devices.

The intelligence provided by microprocessors can and is used for valve tuning, semiautomatic calibration, and data collection for maintenance and diagnostic purposes. It is also used for standalone control when that might improve the positioning, the protection, or the communication of the valve. Table 6.3ee lists some of the more common applications of smart and intelligent actuator systems.

TABLE 6.3ee
Applications for Integrated Intelligent Systems[5]

- Under-instrumented loops where additional information is required for optimization or improved control.
- Loops requiring very tight control with a fast update for the PID algorithm.
- Applications requiring a large turndown in flow measurement and control capabilities.
- Processes that need local supervisory and control capabilities to provide continuous operation or controlled shutdown upon loss of control signal.
- New or retrofit installations where there is insufficient space for conventional instrumentation systems.
- Critical systems requiring continuous monitoring and predictive diagnostics on the valve or process.
- Applications where changing conditions require the loop to be reconfigured to control diverse variables for best control.
- Standalone remote applications requiring programmable operation and remote interface for monitoring.
- New applications where an intelligent system can be used as a fully integrated system, reducing engineering, cost, and maintenance requirements.
- Specialized control functions such as gap control for very large turndown, or self-contained minimum pump recirculation systems for pump or compressor protection.

Microprocessor-based systems are available in watertight or explosionproof (NEMA 6 to 7) construction and can tolerate ambient temperatures from –40 to 185°F (–40 to 85°C), as well as the presence of moisture, fungus, or dust. They can incorporate sensors, an electronic positioner, or an electronic proportional, integral, derivative (PID) controller, which can operate off a digital or analog external set point.

Locating the controller card at the valve and dedicating it to full-time control is an advantage on very fast critical loops, such as compressor surge protection, where otherwise surge could evolve while the central DCS control system is scanning/updating other loops.

Smart systems can also take advantage of the lower cost of digital communications over a single loop of two-conductor or fiber-optic cable (Figure 6.3ff). This method of communication substantially reduces the wiring cost of installations that can control up to 250 actuators, pumps, or solenoids. The smart valves can be remotely calibrated and reconfigured and can be used to detect such performance changes as pressure changes resulting from a fouled pump or pipeline. They can also limit the valve travel to stay within the range where the required characteristics (gain) are available.

It is perhaps at the maintenance level that the most benefits can be obtained through the use of smart and intelligent systems.[6] Smart actuator circuitry can protect electric motor operators from burning out when the valve is jammed or from reverse phasing. More importantly, by close monitoring of the valve, the required information can be obtained to evaluate the valve and actuator performance and thereby replace preventive and corrective maintenance by predictive maintenance with the benefits of improved process performance and products quality at a lower cost.

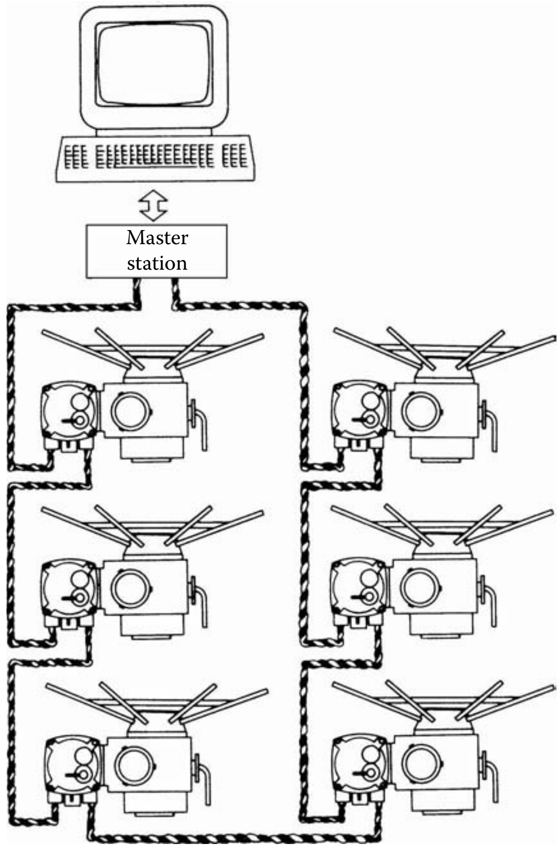


FIG. 6.3ff
Multiple intelligent actuators can be monitored and controlled by a two-wire multiplexer loop. (Courtesy of Rotork Controls Ltd.)

References

1. Harrold, D., “A Changing Landscape,” Control Engineering, December 2003, www.manufacturing.net/ctl/article/CA339684? text = valve + actuators.
2. Scott, A. B., “Control Valve Actuators: Types and Application, *InTech,* January 1988.
3. Harrold, D., “Making Valve Controllers/Positioners Smarter is Smart Business,” *Control Engineering,* January 2003, www.manufacturing.net/ctl/index.asp?layout=articlePrint&articleID=CA269795.
4. Hoske, M. T., “Implementing Industrial Networks,” *Control Engineering,* July 1998, www.manufacturing.net/ctl/article/CA197636?text=valve+actuators.
5. Price, V. E., “Smart Valve Intelligence Takes Many Forms,” *InTech,* August 1992.
6. Merritt, R., “A Real (Valve) Turn-On” *Control Design,* August 2002, www.controldesign.com/Web_First/CD.nsf/ArticleID/JFEY-5BMRJD.

Bibliography

Barnes, P. L., "Protect Valves with Fire-Tested Actuators," *Instrument and Control Systems,* October 1982.

Baumann, H. D., "Trends in Control Valves and Actuators," *Instrument and Control Systems,* November 1975.

Berris, R., Hazony, D., and Resch, R., "Discrete Pulses Put Induction Motors into the Stepping Mode," *Control Engineering,* Vol. 29, No. 1, 1982, p. 85.

Colaneri, M. R., "Solenoid Valve Basics," *Instruments and Control Systems,* August 1979.

Cottell, N., "Electrohydraulic Actuation-Still in Control?", IEE Colloquium on Actuator Technology: Current Practice and New Developments, May 1996, London, U.K.

Fitzgerald, W. V., "Loop Tuning and Control Valve Diagnostics," Paper #91–0500, 1991 ISA Conference, Anaheim, CA, October 1991.

Hammitt, D., "How to Select Valve Actuators," *Instruments and Control Systems,* February 1997.

Hoerr, D., "Low-Power DC-Solenoid Valves," *M&C News,* September 1992.

IEC 60534-6 Ed. 1.0 b:1985, "Industrial Process Control Valves. Part 6: Mounting Details for Attachment of Positioners to Control Valve Actuators," 1985.

IEC 60534-6-1 Ed. 1.0 b:1997, "Industrial Process Control Valves. Part 6: Mounting Details for Attachment of Positioners to Control Valves Section 1: Positioner Mounting on Linear Actuators," 1997.

IEC 60534-6-2 Ed. 1.0 b:2000, "Industrial Process Control Valves. Part 6-2: Mounting Details for Attachment of Positioners to Control Valves. Positioner Mounting on Rotary Actuators," 2000.

Kervin, D., "Zero-Crossing Triac Drivers Simplify Circuit Design," *Control Engineering,* Vol. 29, No. 3, p. 76, 1982.

Koechner, Q. V., "Characterized Valve Actuators," *Instrumentation Technology,* March 1977.

Liantonio, V., "High Pressure Modulating Solenoid Valve for Steam/Gas Service," *InTech*, January 1988.

Lipták, B. G., "Control Valves in Optimized Systems," *Chemical Engineering,* September 5, 1983.

Lockert, C. A., "Low Load Technique Controls Motor Energy Losses," *Control Engineering,* Vol. 29, No. 2, 1982, p. 142.

MSS SP-86-2002, "Guidelines for Metric Data in Standards for Valves, Flanges, Fittings and Actuators," 2002.

Price, V. E., "Smart Valve Intelligence Takes Many Forms," *InTech,* August 1992.

Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 ISA Conference, Houston, TX, October 1992.

Valve & Actuator User's Manual, British Valve & Actuator Association, 2002.

Scott, A. B., "Control Valve Actuators: Types and Application," *InTech,* January 1988.

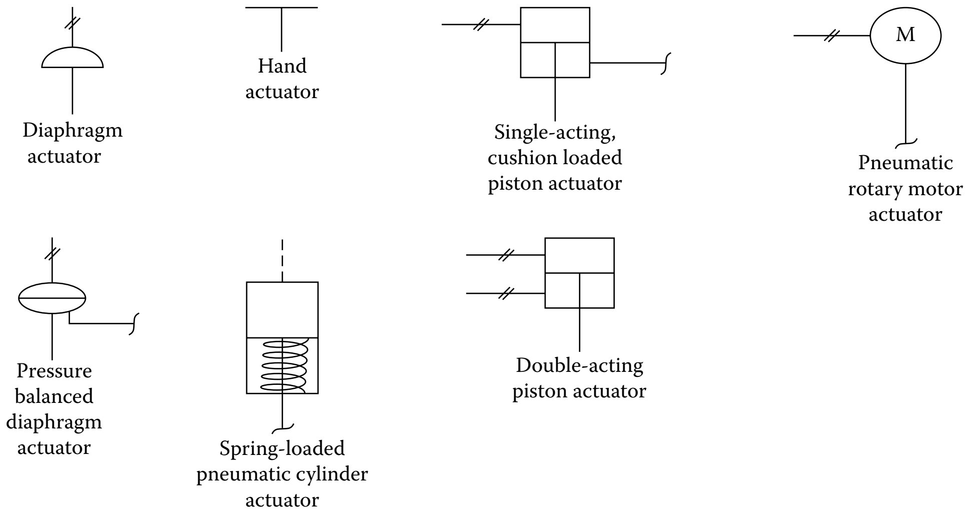
Shinskey, F. G., "Dynamic Response of Valve Motors," *Instruments and Control Systems,* July 1974.

Usry, J. D., "Stepping Motors for Valve Actuators," *Instrumentation Technology,* March 1977.

VDI/VDE 3845, "Industrial Process Control Valves. Interfaces between Valves, Actuators, and Auxiliary Equipment, 1998."

6.4 Actuators: Pneumatic

C. S. BEARD (1970) **O. P. LOVETT** (1985)
B. G. LIPTÁK (1995) **H. L. MILLER** (2005)



Flow sheet symbols

Types:	A. Linear A1. Spring-and-diaphragm A2. Piston B. Rotary B1. Cylinder with Scotch yoke B2. Cylinder with rack and pinion B3. Dual cylinders B4. Spline or helix B5. Vane B6. Pneumohydraulic B7. Air motor B8. Electropneumatic
Applicable to Valve Sizes:	A1. 0.5 to 8 in. (12 to 200 mm) A2. 0.5 to 30 in. (12 to 750 mm) B. 2 to 30 in. (50 to 750 mm)
Standard Spring Ranges:	A1. 3–9, 3–15, 9–15, 6–30 PSIG (20–60, 20–100, 60–100, 40–200 kPa) A1. 60 PSIG (414 kPa); some higher
Max. Actuator Pressure Ratings:	A2. 150 PSIG (1035 kPa); accessories may lower ratings B. 250 PSIG (1725 kPa)
Actuator Temperature Ratings:	A1 and A2. –20 to 150°F (–30 to 66°C); some higher B. −40 to 200°F (–40 to 95°C); special up to 350°F (177°C)
Actuator Areas:	A1. 25 to 500 in.^{2} (0.016 to 0.323 m^2) A2 and B. 10 to 600 in.^{2} (0.006 to 0.38 m^2); bore diameters from 2 to 44 in. (50 mm to 1.1 m) and strokes up to 24 in. (0.61 m)

Linear Thrust (Stem Force) Ranges:	A1. 200 to 45,000 lb_f (22 to 10,100 N) A2. 200 to 32,000 lb_f (22 to 7,200 N); specials up to 186,000 lb_f (41,800 N)
Speeds of Full Stroke:	A1. Small (25 to 100 in.^2) actuators, 1 to 5 sec. A2. 0.33 to 24.0 in./sec (8 to 600 mm/sec) or less than 1 to 30 sec/stroke B. 0.33 to 6.0 in./sec (8 to 150 mm/sec) or 1 to 30 sec/stroke. Dual pistons (Figure 6.4aa) can stroke large rotary valves in 0.5 sec. For the same size units, spring-and-diaphragm units are the slowest, spring-returned pistons are faster, and double-acting pistons with supply/exhaust at both ends are the fastest
Torque Ranges:	(Values are given for double-acting actuators; for spring-return designs torque is about half.) B. 10 to 100,000 $\text{ft} \cdot \text{lb}_\text{f}$ (0.69 to 6850 $\text{N} \cdot \text{m}$); special units for 5 million $\text{ft} \cdot \text{lb}_\text{f}$ (343,000 $\text{N} \cdot \text{m}$) have been built
Cost:	Included in control valve cost
Hysteresis, Dead Band, and Linearity:	A and B. Generally within 2%, but when spring-and-diaphragm or rotary piston actuators are operating rotary valves, the linearity is worse
Partial List of Suppliers:	ABB Kent-Introl (www.abb.com) Actuation Valve & Control Ltd. (www.actuation.co.uk) Allenair Corp. (www.allenair.com) Aluma Actuators Inc. (www.aluma-usa.com) Arca Regler Gmbh (www.arca-valve.com) Bardiani Valvole SpA (www.bardiani.com) Bray Valve & Controls (www.bray.com) Cashco Control Valves & Regulators (www.cashco.com) China Zhejiang Chaoda Valve Co. Ltd (www.chinavalve.com) Circor International, Inc. (Leslie Controls) (www.circor.com) Combraco Industries Inc. (www.combraco.com) Control Components Inc. (Bailey, CCI, BTG, STI, Sulzer, Shin Woo) (www.ccivalve. com) Controlmatics Industrial Products (www.contromatics.com) Dresser (Leeden, Masoneilan) (www.dresser.com) Emerson Process Management (Contex, Fisher, El-O-Matic, Shafer Valve) (www. Emersonprocess.com) Flo-Tork, Inc. (www.flo-tork.com) Flowserve Corp. (Anchor/Darling, Kammer, Valtek, Worcester, McCanna, Limitorque) (www.flowserve.com) Jordan Valve (www.jordanvalve.com) Koso America Inc. (Hammeldahl) (www.rexa.com) K-Tork International Inc. (www.ktork.com) Larox Flowsys Inc. (www.larox.fi) Nihon Koso (www.koso.co.jp) Metso Automation USA, Inc. (Neles-Jamesbury, Valmet Automation) (www.metsoautomation.com) Mumatics Inc. (www.mumatics.com) Norriseal Controls (www.norriseal.com) Parcol SpA (www.parcol.com) Red Valve Co. Inc. (www.redvalve.com) Rotork Controls Inc. (www.rotork.com) Samson Regeltechnick.bv (www.samsom-regeltechnick.nl) Severn Glocon Ltd. (www.severnglocon.com) Spirax Sarco Inc. www.spiraxsarco-usa.com SPX Valves & Controls (Copes Vulcan, Dezurik) (www.spxvalves.com) Tyco Flow Control Div. (Keystone, Morin, Biffi, Descoti, Sempell, Yarway, Grinnell, MCF) (www.tycovalves-na.com) Welland & Tuxhorn Gmbh (www.welland-tuxhorn.de) Wier Group plc (Atwood & Morrell, Batley Valve, Blakeborough Controls, Hopkinsons, Sebim, Flowguard) (www.weirvalve.com) Xomax Corp. (www.xomax.com) Yamatake Corporation (www.yamatake.co.jp)

INTRODUCTION

Pneumatic valve actuators respond to an air signal by moving the valve trim into a corresponding throttling position. This section covers the two basic designs most frequently utilized: the diaphragm and the piston actuator. The discussion of diaphragm- and piston-type actuators is followed by the treatment of pneumatic-rotary and pneumatic-hydraulic actuators.

In connection with the performance of these actuators, an analysis is presented of the various forces positioning the plug, including diaphragm, spring, and dynamic forces generated by the process fluid. An understanding of the interrelationships among these forces will allow the reader to properly size these actuators and make the correct spring selection.

The failure safety of valve actuators and the relative merits of diaphragm vs. piston actuators and the topic of high-speed actuation using pneumatics is also discussed. This section is concluded with a summary of the results of a long-term evaluation of the performance of pneumatically actuated control valves in the field.

DEFINITIONS

An actuator is that portion of a valve that responds to the applied signal and causes the motion resulting in modification of fluid flow. Thus, an actuator is any device that causes the valve stem to move. It may be a manually positioned device, such as a handwheel or lever. The manual actuator may be open-closed, or it may be manually positioned at any position between fully open and fully closed. Other actuators are operated by compressed air, hydraulics, and electricity.

The actuators discussed here are those capable of moving the valve to any position from fully closed to fully open and those using compressed air for power. Of such there are two general types: the spring-and-diaphragm actuator and the piston actuator.

In a spring-and-diaphragm actuator, variable air pressure is applied to a flexible diaphragm to oppose a spring. The combination of diaphragm and spring forces acts to balance the fluid forces on the valve.

In a piston actuator, a combination of fixed and variable air pressures is applied to a piston in a cylinder to balance the fluid forces on the valve. Sometimes springs are used, usually to assist valve closure. Excluding springs, there are two variations of piston actuators: cushion loaded and double acting.

In the cushion-loaded type, a fixed air pressure, known as the cushion pressure, is opposed by a variable air pressure and is used to balance the fluid forces on the valve. In the double-acting type, two opposing variable air pressures are used to balance the fluid forces on the valve.

An actuator can be said to have two basic functions: (1) to respond to the external signal of a controller and cause an inner valve to move accordingly (with the proper selection and assembly of components, other functions can also be obtained, such as a desired fail-safe action) and (2) to provide a convenient support for valve accessory items, such as positioners, limit switches, solenoid valves, and local controllers.

ACTUATOR FEATURES AND SELECTION

Table 6.4a describes the applications and relative advantages of a variety of actuator designs. The table lists both the advantages and the limitations of the various designs. The popularity of the spring/diaphragm actuator is due to its low cost, its relatively high thrust at low air supply pressures, and its availability with “fail-safe” springs. By trapping the pressure in the diaphragm case, it can also be locked in its last position. It is available in springless designs, double diaphragm designs (for higher pressures), rolling diaphragm designs (for longer strokes), and tandem designs (for more thrust).

One of the limitations of this design is the lack of actuator “stiffness” (resistance to rapidly varying hydraulic forces - for example, those caused by flashing). For such applications double-acting piston actuators are used and for extraordinary requirements hydraulic or electromechanical (motor gear) actuators may be preferred. A stiffer spring, 6–30 PSIG (41–210 kPa), in a spring/diaphragm unit is sometimes sufficient to correct the problem.

Linear piston actuators provide longer strokes and can operate at higher air pressures than can the spring/diaphragm actuators. When used to operate rotary valves, the linear piston or spring/diaphragm actuator does not provide a constant ratio of rotation per unit change in air signal pressure; therefore, the use of positioners is always a requirement.

Rotary piston actuators operate at higher air pressure and can provide higher torques, suitable for throttling large ball or butterfly valves. The double-acting version of this actuator does not have a positive failure position, but this can be corrected by extending the piston case and inserting a helical spring. For higher torques (over 1000 ft · $\mathrm{lb_f}$ [68 Nm]), heavy-duty transfer linkages are required (Scotch yoke or rack and pinion); such units cannot be disassembled and maintained in the field. These actuators also require positioners because the relationship between air signal change and resulting rotation is not linear.

SPRING/DIAPHRAGM ACTUATORS

This discussion is restricted to pneumatic actuators. The external signal, therefore, is an air signal of varying pressure. The air signal range from a pneumatic controller is commonly 0–18 PSIG (0–124 kPa). Signal or actuator input pressure starts at 0 PSIG, not 3 PSIG (21 kPa).

A common mistake is to confuse the 3–15 PSIG (21–104 kPa) range of transmitter output pressure with the signal to a valve. The higher value of 18 PSIG (124 kPa) is fixed only by the air supply to the controller (or positioner),

TABLE 6.4a
Features of Pneumatic Actuators

Type of Actuator	*Advantage*	*Disadvantage*	*Application*
Linear spring-and-diaphragm	Low cost Mechanical fail-safe Moderate thrust Small package Simple design Excellent control with or without control devices	Slow speed Poor stiffness Instability	Linear valves 1/2–8 in. (12–200 mm) body size
Linear piston	Moderate cost Moderate thrust Small package Simple design Excellent control with control device Long stroke High-speed options Moderate stiffness	Large spring compression when used for failure	Linear valves 1/2–30 in. (12–760 mm) body size
Rotary spring-and-diaphragm	Moderate cost Mechanical fail-safe Small package Simple design Easily reversible Excellent control with control device	Low thrust in spring cycle Instability	Rotary valves 1–6 inch (25–150 mm) body size
Rotary pistons	Low cost Moderate thrust Small or large package Good control with control device Mechanical fail-safe option	Slow speed Large spring Compression	Rotary valves 1–24 inch (25–600 mm) body size

See Section 4.3 for a detailed treatment of the features of hydraulic and electric actuators.

and it can easily be set to 20 PSIG (138 kPa) or higher. A variety of other input pressures are sometimes used, such as 0–30 or 0–60 PSIG (0–207 or 0–414 kPa).

Both the spring-and-diaphragm and the piston actuator produce linear motion to move the valve. These actuators are ideal for use on valves requiring linear travel, such as globe valves. A linkage or other form of linear-to-rotary motion conversion is required to adapt these actuators to rotary valves, such as the butterfly type.

Steady-State Force Balance

In spring-and-diaphragm actuators the stem positioning is achieved by a balance of forces acting on the stem. These forces are caused by the pressure on the diaphragm, spring travel, rubbing friction, and fluid forces on the valve plug (Figure 6.4b).

Equation 6.4(1) can be derived from a summation of forces on the valve plug adopting the positive direction downward.

$$PA - KX - P_v A_v = 0 \qquad \textbf{6.4(1)}$$

where A is the effective diaphragm area, A_v is the effective inner valve area, K is the spring rate, P is the diaphragm pressure, P_v is the valve pressure drop, and X is stem travel. Equation 6.4(1) applies to a push-down-to-close actuator and valve combination with flow under the plug. This type of actuator is commonly referred to as direct acting.

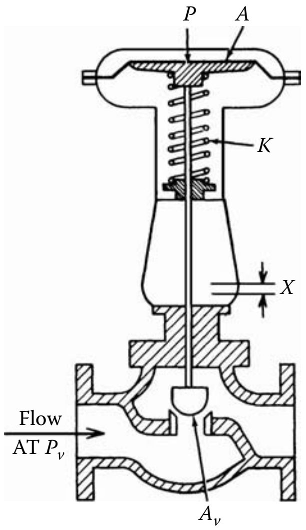


FIG. 6.4b
Forces acting on a spring-and-diaphragm actuated, fail open (FO) control valve.

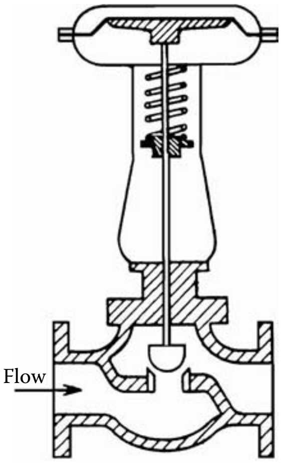


FIG. 6.4c
Reverse-acting spring-and-diaphragm actuator, attached to a flow to open, spring to close (fail closed = FC) control valve.

Another popular actuator configuration is one causing the stem to rise on an increase of air pressure. It is commonly called a reverse acting actuator (Figure 6.4c).

By using the same sign convention, the force balance equation for this valve configuration is given in Equation 6.4(2).

$$-PA + KX - P_vA_v = 0 \qquad \textbf{6.4(2)}$$

If the flow direction is reversed in Figure 6.4b, the equation becomes:

$$PA - KX + P_vA_v = 0 \qquad \textbf{6.4 (3)}$$

Likewise, reversing flow direction in Figure 6.4c results in Equation 6.4(4):

$$-PA + KX + P_vA_v = 0 \qquad \textbf{6.4(4)}$$

These equations are simplified because they do not consider friction and inertia. Friction occurs in the valve stem packing, in the actuator stem guide, and in the valve plug guide or guides. Usually, for static valve actuator sizing problems, negligible error is introduced by ignoring the friction terms.

If Equation 6.4(1) is plotted as signal pressure vs. stem travel and if the case of no fluid forces on the plug (bench test) is assumed, then the curve shown in Figure 6.4d is obtained.

Next, consider the case of plug forces due to fluid flow, assuming that the term P_v is constant for all travel positions. This has the effect of shifting the straight line to the right to some position depending on the magnitude of P_v. Curves similar to those in Figure 6.4d can readily be drawn for the other valve configurations represented by Equations 6.4(2), 6.4(3), and 6.4(4). The distance between the lines is the force resulting from A_v is the effective inner valve area, K is the spring rate, P is the diaphragm pressure, P_v is the valve pressure drop, intersecting the abscissa is the pressure needed to move the valve plug.

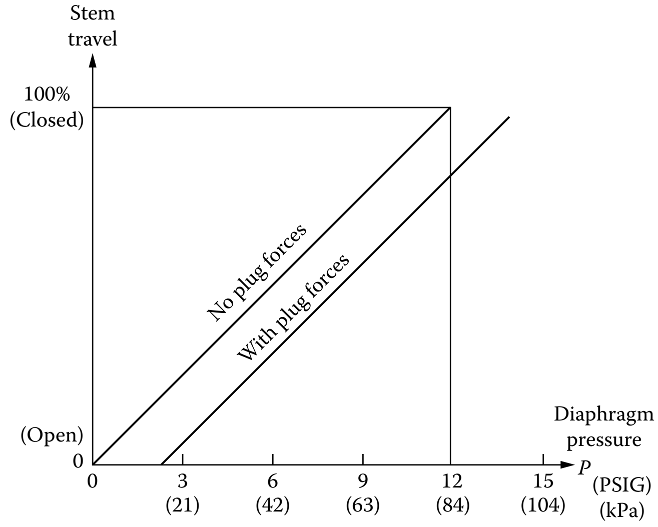


FIG. 6.4d
In a fail open (FO) valve, as the air pressure on the diaphragm rises, the valve closes. The distance between the two sloping lines corresponds to the force acting on the plug (A_v), which force is generated by the differential between the up and downstream pressures (P_v), when these forces tend to keep the valve open, as in Figure 6.4b.

Actuator Sizing Example

Let us assume that the forces acting on a 1 in. (25 mm) single-ported globe valve are to be evaluated. In that case:

A = 46 in.^{2} (0.03 m^2)
X = 5/8 in. (15.9 mm) full travel
K = 885 ft lb$_f$/in. (7.83 N/mm)

If no plug forces exist, Equation 6.4(1) reduces to $PA = KX$. Solving for the pressure change required to obtain full travel from open to closed:

$$P = KX/A = (885)(5/8)/(46) = 12.03 \text{ PSIG } (83 \text{ kPa})$$

This is reasonably close to the 12 PSIG (0.83 kPa) desired operating span. Practical considerations of variations in spring constants and in actuator-effective areas usually prevent such a close approach to the desired span, and frequently a ±10% leeway is permitted.

When there are plug forces, it is seen from Equation 6.4(1) that an additional actuator force is required to maintain balance. The actuator pressure required to begin stem motion can be calculated for the case of a 1 in. diameter (25 mm) plug ($A_v = \pi/4$) and 100 PSIG (690 kPa) pressure drop. Equation 6.4(1) can be used to solve for P as follows (stem travel is zero, thus there are no spring forces):

$$\begin{aligned} P &= [(K = 885)(X = 0) + (P_v = 100)(A_v = \pi/4)]/(A = 46) \\ &= 1.7 \text{ PSIG } (11.8 \text{ kPa}) \qquad \textbf{6.4(5)} \end{aligned}$$

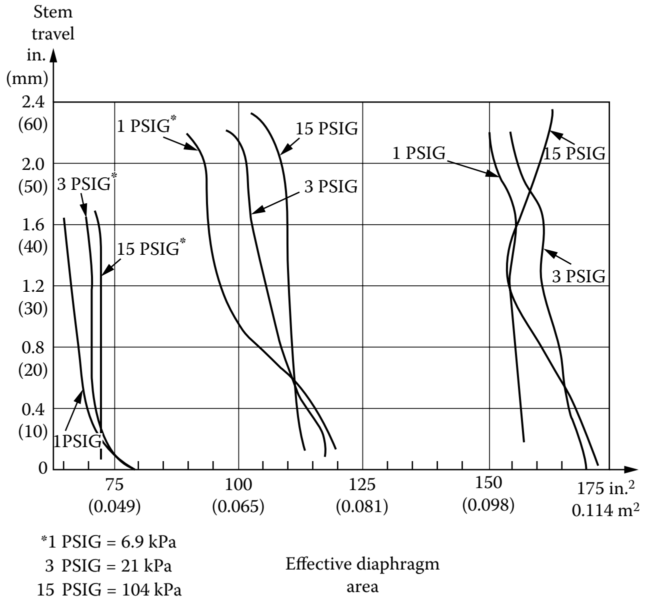


FIG. 6.4e
The effective diaphragm areas vary with both the stem travel and with the pressure acting on the diaphragm.

This means that the diaphragm pressure must increase to 1.7 PSIG (11.8 kPa) before stem travel begins. This is the distance between the two sloping straight lines in Figure 6.4d.

Actuator Nonlinearities

In practice we encounter many nonlinearities, and the ideal curves in Figure 6.4d are not obtained. These nonlinearities are due to several factors, such as the variable effective diaphragm areas. The effective diaphragm area varies with travel and with the pressure level on the diaphragm. Figure 6.4e illustrates this for three different sizes of diaphragms.

Another source of nonlinearity is in the variation of the valve plug forces (P_vA_v). Figure 6.4f illustrates the variations in these plug forces for two 4 in. (100 mm), single- and double-ported valves. The figure also shows the effects of flow over and under the plugs of a single-ported valve.

Springs are also nonlinear in that the spring rates vary with travel. By judicious selection of springs, considering their spring rate and travel, the effects of their nonlinearity on the valve assembly can be minimized.

When all of these nonlinearities are considered, a plot of actuator travel vs. diaphragm pressure would not be a straight line as shown in Figure 6.4d, but might be a curve such as the one shown in Figure 6.4g.

A nonlinear curve, such as the one labeled “actual” in Figure 6.4g, is not necessarily objectionable. When used in an automatic control loop, the static nonlinearities are compensated for by the controller. This curve is actually a part of the gain term in the valve’s transfer function, and the other part is the flow characteristic. When a valve positioner is used, the positioner overcomes these nonlinearities, and the result is similar to the ideal curve shown in Figure 6.4g.

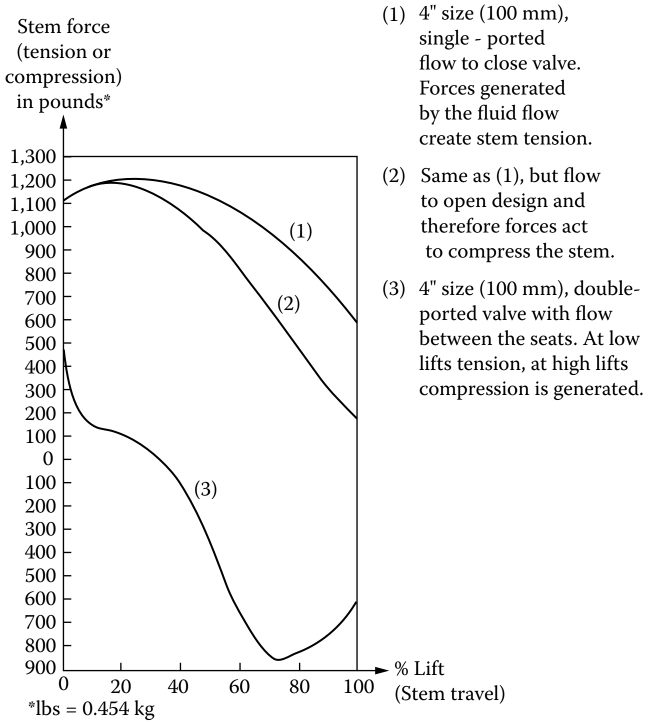


FIG. 6.4f
The forces acting on the valve plug are also nonlinear, because the pressure difference across the valve varies as a nonlinear function of stem travel.

Dynamic Performance of Actuators

Several control valve subsystems must be analyzed in order to thoroughly evaluate their dynamic performance. The separate systems include:

1. The spring-mass system of the valve’s moving parts.
2. The pneumatic system from controller output to valve diaphragm chamber. If a valve positioner is used, there are two separate pneumatic systems: one from the controller output to the positioner and another from the

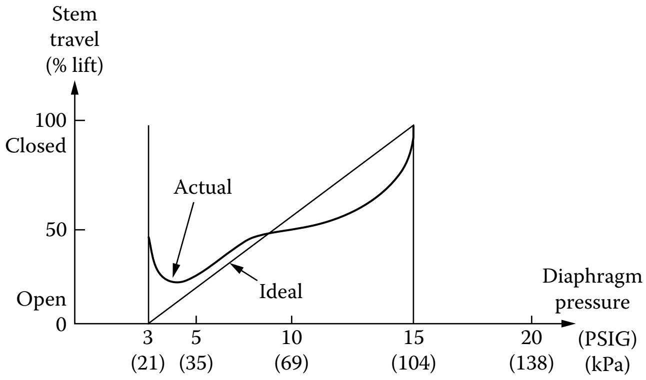


FIG. 6.4g
Ideal and actual relationship between diaphragm pressure and stem travel.

positioner output to the diaphragm chamber. The interconnecting tubing is consideration in all of the pneumatic systems.

Spring-Mass System Dynamics Analysis of the spring-and-mass system is only valid for linear systems. It is necessary either to neglect consideration of the nonlinear elements or have a system wherein the nonlinear effects are minor. In the case of control valves with sufficient power in the actuator, the latter case is approached. With such an understanding of the nonlinear effects, we proceed as though valve actuators were linear devices.

The spring-mass system is represented by the following differential equation:

$$M\frac{d^2X}{dt^2}+b\frac{dX}{dt}+KX=PA-P_vV_v \qquad \textbf{6.4(6)}$$

where b is the net friction force and M is mass. The net friction force would include friction due to seals, mechanical rubbing, and viscous friction on the plug.

The static, time-independent terms of Equation 6.4(6) are identical with Equation 6.4(1). The transfer function of the valve actuator is the LaPlace transform of differential Equation 6.4(6):

$$\frac{X(s)}{P(s)}=\frac{A/K}{(w/gK)s^2+(b/K)s+1} \qquad \textbf{6.4(7)}$$

where g is the gravitation constant, s is the LaPlace operator, and w is the weight of moving parts. This can be written in terminology more useful to instrument engineers using the time constant τ (tau) and damping factor ζ (zeta).

$$\frac{x(s)}{P(s)}=\frac{1}{\tau^2s^2+2\tau\varsigma s+1} \qquad \textbf{6.4(8)}$$

The coefficient of the s^2 term in Equation 6.4(7) is the square of the reciprocal of the undamped natural frequency of the spring-mass system. It is a useful number in understanding the relative importance of a control valve's dynamic components.

Table 6.4h is a list of the natural frequencies of different size valves of the "average" design. It should be noted that even the largest valve with its un-damped natural frequency of Hz = 9 is ten times faster than the typical pneumatic performance of a control valve.

For a more detailed discussion of the dynamics of diaphragm actuators and of the effect of standing pressure waves in the piping on that dynamics, the reader is referred to the discussion by Lynch.[1]

TABLE 6.4h
Natural Frequency of Control Valves

Valve Size (in.)	*Valve Size (mm)*	*Undamped Natural Frequency (Hz)*
1	25	32
$1^1/_2$	37.5	27
2	50	22
3	75	16
4	100	14
6	150	10
8	200	9

Resolution and Valve Oscillation The minimum change in the stem position of a control valve is called its resolution, which limits the ability of the control signal to position the valve exactly at a specific point of its travel. Because of this limitation, valves can continuously oscillate, as the increment of the stem position that can be delivered is larger than what is required. The result is a continuous sequence of overshooting and undershooting the stem position target. The valve's resolution, this minimum change in position, cannot be very accurately calculated, because of the variations between the valve designs.

The definition of resolution for all pneumatic actuators is the ratio of the change in friction force to the spring rate. The relative resolution (R) is calculated by taking the difference between the static and dynamic friction forces ($F_s - F_d$) and dividing that with the difference between the mechanical spring rate and the spring rate of the trapped air ($K_a - K_s$). The resolution is calculated or percent of full stroke is then obtained by dividing by the stroke. In equation form the resolution, R, in dimensions of length is:

$$R = (F_s - F_d)/(K_a - K_s) \qquad \textbf{6.4(9)}$$

where:

R is the resolution in units of either length or percentage of full stroke
F_s is the static spring friction
F_d is the dynamic (running) spring friction
K_a is the spring rate of the trapped air
K_s is the spring rate of the spring

The spring rate of the trapped air (K_a) is shown by Equation 6.4(10):

$$K_a=\frac{1.4PA^2}{V} \qquad \textbf{6.4(10)}$$

where:

A is the effective area of the actuator
V is the volume of trapped air in each actuator cavity
P is the absolute pressure in the actuator

The main contributors to the friction forces are the packing friction on the actuator stem, the valve stem and the effect of any valve internal balancing seals. The motion of the actuator is faster than the time it would take for the air to be exchanged on both sides of the diaphragm or piston. Therefore, the air pressure conditions at each point of travel can be evaluated in a static manner, and the air exchange dynamics can be ignored.

A number of observations can be made in connection with Equation 6.4(9). One such observation is that the dynamic or running friction (F_d) is always less than the static or breakaway friction (F_s). The difference usually is 25–35%. Another observation is that the friction forces for a PTFE (Teflon) seal are less than the friction forces generated by higher temperature seals made of fibrous graphite.

The size of these friction forces is much affected by the amount of extra torque applied to tightening packing box seals during installation. Too much compression of the seals will result in high friction forces and in stem travel oscillations.

The mechanical spring rate (K_s) is essentially constant. The air spring rate (K_a) can be increased by selecting a large effective area (A) in the actuator in combination with a small air volume (V). Because the air spring rate (K_a) is a function of the air volume and because the mechanical spring force changes with the stem movement, the resolution (R) will also vary with valve travel.

The relationship between resolution and stem travel for a variety of actuator designs is shown Figure 6.4i. In most designs the use of a spring tends to reduce the resolution. In spring-diaphragm combinations, the resolution is improved (reduced) when the spring opposes the direction of valve stem travel.

Good maintenance is essential to minimize the frictional forces in valve actuators. The positioner, which delivers or exhausts air to/from the actuator, is slower than the speed at which small changes in stem position occur.

Once the valve is installed the only means available to the user to change the resolution is to modify the supply air pressure, up to the limit of the design pressures. Because the air spring rate (K_a), which can be calculated by Equation 6.4(10), is small in comparison to the mechanical spring rate (K_s), an increase in the air supply pressure is not likely to have much impact on the resolution.

For reasons of competition, the valve manufacturers usually provide the smallest actuator they can for the particular application. Yet, for a small additional expense the user can usually obtain an actuator with a larger effective area and obtain a noticeable improvement in resolution and, therefore, in the controllability of the loop. By so doing, the continuous oscillation of the valve can often be stopped.

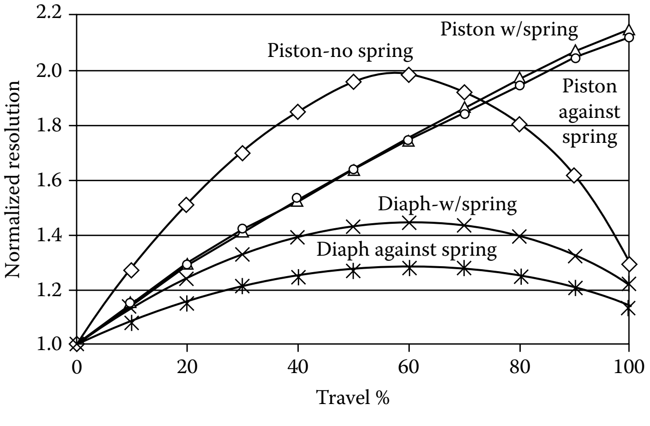


FIG. 6.4i
Actuator resolution curve vs. travel for a number of actuator designs.

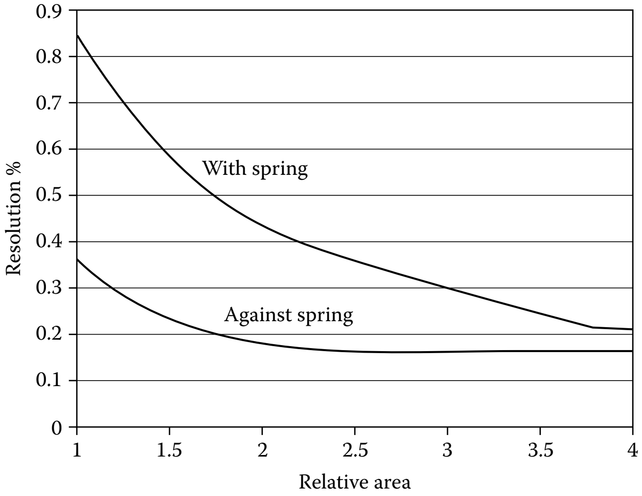


FIG. 6.4j
Resolution vs. diaphragm effective area.

As it is shown in Figure 6.4j, the valve resolution can be much reduced (improved) by using a larger actuator. As can be noted from the figure, a doubling of the effective area of a diaphragm actuator cut the actuator resolution nearly in half.

Safe Failure Position

The valve application engineer must choose between the two readily available fail-safe schemes for control valves, either fail open or fail closed. The choice will be based upon process safety considerations in the event of control valve air failure. Complete plant air failure, controller signal failure, and local air supply failure must all be considered. Local failure is significant when a valve positioner is being used and when piston actuators with cushion loading are used.

The choice must be based on detailed knowledge of the valve application in the overall process or system. Two generalizations are that in a heating application, the valve should fail closed, and in a cooling application it should fail open. There are certainly applications where either failure mode is equally safe; then, considerations of standardization may be used.

Fail-safe involves the selection of actions of actuator and inner valve. Both actuator and inner valve usually offer a choice of increasing air pressure to push the stem down or up, and pushing the stem down may open or close the inner valve. The proper choice of combinations may be made by fail-safe considerations. The process application of the valve must be investigated to determine whether, on instrument air failure, it would be better to have the valve go fully open, fully closed, or remain in its last position.

There may not be much flexibility in the inner valve action. For example, a single-seated top-guided valve must push down to close the plug. There is freedom of choice, however, in either single- or double-seated top- and bottom-guided valves. Other valve bodies, such as the Saunders and pinch valve styles, must be of the push-down-to-close type.

Rotary types, such as butterfly and ball valves, may be arranged either way.

The inner-valve flexibility leads to two cases: one in which either inner-valve action is permissible and one in which the inner-valve must be push-down-to-close.

When there is a choice of inner-valve action, overall valve action may be obtained by selecting the suitable inner-valve action and always using increasing air to push down the actuator. This is known as a direct actuator. A direct actuator is preferred because of economy reasons in spring-and-diaphragm actuators. The savings may be in purchase cost. It is also realized in maintenance costs, because there is no actuator stem seal to cause possible leakage and maintenance costs.

When the inner valve must be push-down-to-close, it is necessary to use both direct and reverse actuators to accomplish the desired fail-safe actions. Figure 6.4k summarizes the available diaphragm failure options.

The piston-type actuator is equally suitable for direct or reverse action. If it is the actuator to be used, the application engineer has complete freedom of the choice of selecting the valve action.

The Role of the Positioner The above description provides a "baseline" for safe valve failure if the actuators are not provided with positioners. By the addition of a positioner, the topic of safe valve failure becomes quite complex. This is because in this case not only the pneumatic signal to the actuator can fail, but also the air supply to the positioner. In order to satisfy these requirements and also to make available the "fail in place" configurations, it is necessary to provide various accessories to either exhaust or trap the actuator air pressure. These accessories include pneumatic pilot valves (see Section 6.2) that, if air pressure is lost, will trip to provide a safe valve action. As far as digital systems are concerned, as of this writing only one digital positioner manufacturer provides a tight shut-off positioner.

Valve failure (Overall)	*Fail Open*		*Fail Closed*	
Actuator	Direct	Reverse	Reverse	Direct
Inner Valve	Direct	Reverse	Direct	Reverse

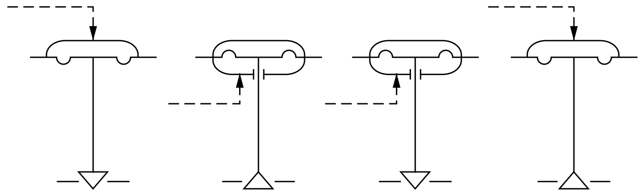


FIG. 6.4k

Overall valve failure positions, which can be achieved by various combinations of direct or reverse actuators and inner valves.

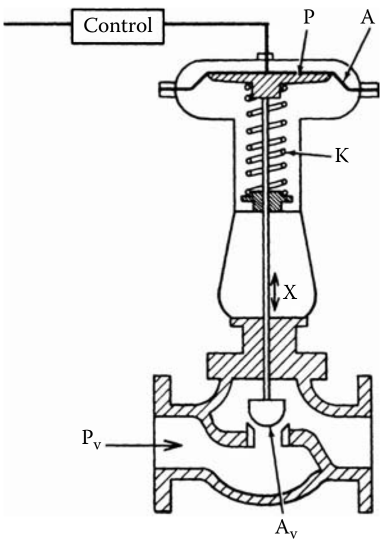


FIG. 6.4l

Forces acting on control valve with a spring-and-diaphragm type actuator.

In the cases where only the positioner can send an air signal to the valve actuator, the valve failure position is usually unaffected by the addition of the positioner. An exception to this statement is if a nonbleeding digital positioner is used.

Actually, it is questionable if a bleed can be considered as a means of providing positive failure position. This is because the time to bleed the air out through the positioner can be quite long. This long time that is required to reach a failure position is often unacceptable, and in such cases, one should provide pneumatic pilot valves to quickly trap or exhaust the air.

Pneumatic Response Times

An earlier discussion considered the transfer functions of the spring-and-diaphragm actuator between the actuator pressure to the resulting stem travel. Next we will consider the pneumatic transfer function of the air signal from the controller to the diaphragm actuator (Figure 6.4l).

A short tube behaves linearly as a pure resistance, and the air volume in the actuator above the diaphragm behaves as a capacitance. So the combination is a resistance-capacitance time constant. Some time constant values obtained from tests with very short tubing are given in Table 6.4m.

TABLE 6.4m

Time Constants for Short Tube Sections

Valve Size		*Time Constant*
(in.)	*(mm)*	*(sec)*
1	25	0.03
2	50	0.05
4	100	0.8

Performance is usually limited by the controller's or positioner's ability to supply the required air fast enough. The time constant values in Table 6.4l were obtained from tests in which the air supply was not limiting. These figures show that the valves are capable of fast response. Section 3.1 contains a more detailed discussion of transmission lags and methods of boosting.

PISTON ACTUATORS

Piston actuators are either single or double acting. The single-acting actuator, shown in Figure 6.4n, utilizes a fixed air pressure, known as the cushion, to oppose the controller signal. This valve does not have spring or diaphragm area nonlinearities, but it is of course subject to the same plug force nonlinearities (Figure 6.4f) as the spring-and-diaphragm actuator.

In order to use such an actuator for throttling purposes, it is necessary to have a positioner. The positioner senses the actuator motion and causes the valve to move accordingly. It cannot be used as a proportioning travel device without the positioner; consequently, its performance is that of the "ideal" curve in Figure 6.4g.

A double-acting piston actuator is one that eliminates the cushion regulator and uses a positioner with a built-in reversing relay. Thus, the positioner has two air pressure outputs, one connected above the piston and the other below. The positioner receives its signal and senses travel in the same manner as a single-acting positioner. The difference is in the outputs; one pressure increases and the other decreases to cause piston travel.

Table 6.4o provides typical actuator stroking times as a function of actuator sizes, stroking distances, and connecting tube sizes. For closed-loop control applications, the speed of response that is critical is not the full stroking time but rather the time it takes to move the valve about 5% of its full stroke, which is much faster than the values given in Table 6.4o. Therefore, velocity limiting of the loop usually occurs only when the valve is "slow" relative to the controlled variable and when the change in the controller signal to the valve is large.

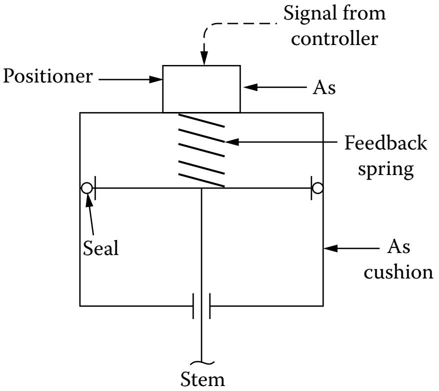


FIG. 6.4n
Single-acting piston actuator.

TABLE 6.4o
Spring-Loaded Piston Actuator Stroking Times for Linear Control Valves[*]

	Time (seconds) for Maximum Stroke[**]		
Actuator Size	*$^1/_4$ in. Tubing*	*$^3/_8$ in. Tubing*	*Stroke (inches)*
25	1.2	1.0	1.5
50	3.5	3.1	3
100	9.6	8.6	4
200	20.8	18.4	4
300	31.3	27.7	4

Actuation pressure: 60 psi

[*] Courtesy of Valtek Flowserve Corp.

[**] Stroking time only (does not include time from receipt of signal and beginning of stem motion).

One advantage of the piston actuator is that higher pressures can be used for motive power. The higher pressure provides better stiffness and resolution. It also provides more force to keep the valve closed. The higher force between the valve plug and the seating surface ensures a tighter shutoff and helps to meet the leakage specifications of the original design.

The piston actuators can also generate longer strokes, because their stroke is limited only by the length of the cylinder used. Such longer strokes are required for many special valve designs that are used in services where cavitation, noise, or pipe vibration is a problem (see Section 6.14).

All of the failure modes are available with the piston actuator. A spring may be used inside or outside of the cylinder to cause the valve to fail in the desired position. The use of outside springs necessitates the use of additional seals and increases mechanical complexity. The use of an inside spring requires a larger cylinder volume that could impact the resolution of the valve.

To obtain a positive failure position without a spring requires a standby air tank to provide the needed power for moving the valve into the failure mode. It also requires control accessories to ensure proper exhaust or lock-up of the air pressures.

HIGH-SPEED ACTUATORS

In the past, high-speed actuation was usually provided by hydraulic actuators. Applications that require fast control valve motion include compressor recycle, turbine bypass, and pressure relief applications. Control valves used in starting up or shutting down a process may also require fast actuation

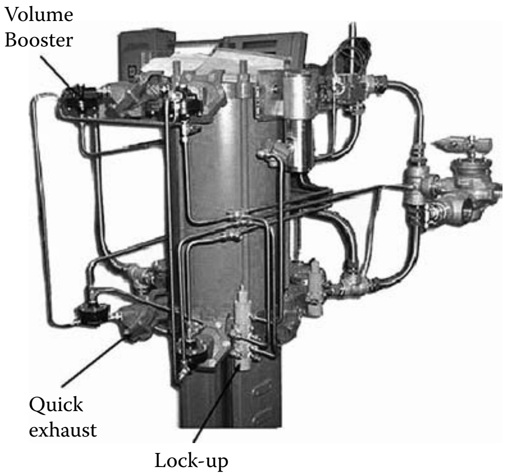


FIG. 6.4p
High speed pneumatic circuitry. (Courtesy of Control Components Inc.)

to protect equipment. Fast speed in these cases usually means full valve stroke in 1 to 2 sec.

In case of pneumatic actuators, the factor that is limiting the speed is the speed at which air can be fed or exhausted from the cylinder. This is accomplished by accessories that direct the air to the cylinder while bypassing the positioner and by increasing the air feed line sizes and the air supply pressure. Because of the increased pressure, the piston actuator is usually preferred in comparison with the diaphragm design, because of its higher pressure rating.

Diaphragm actuators are usually limited to 40 PSIG (275 kPa), with a few designs permitting 60 PSIG (410 kPa) operation. The high-pressure designs can operate with the usual plant air systems of 150 PSIG (1035 kPa) with normal operating pressures of 100 PSIG (690 kPa). This range of air supply pressures is readily available in most plants.

The high-speed pneumatic actuation can be achieved by the use of boosters to feed and exhaust air from the cylinders. These boosters are actuated when the control signal calls for a 10% or so change in valve opening. An actuator that is being fed by two boosters located above and below the piston is shown in Figure 6.4p.

In addition to the boosters, other accessories are also needed to guarantee the required failure positions. There are many disadvantages to these designs, including the difficulty of tuning and calibrating all of the component devices individually so as to maintain a stable operation. In addition, the tubing configuration requires a substantial amount of space around the actuator, and the accessories and tubing provide a lot of handhold and support points for operators, which can lead to damage. These systems can also be sensitive to vibration and to high temperatures radiating from a hot valve.

A newer development is shown in Figure 6.4q. This design required the development of high-capacity servo-valves that can be positioned to very close tolerance and of a programmable electronic controller. The controller can receive the HART or fieldbus protocol of the control system, can maintain the travel of the actuator so that there is no visible overshoot even during the fastest transients, and can calibrate the actuation system within seconds. The higher pressure and stiffness of the actuator allows positioning within 0.25% of the total travel. Dead time and hysteresis are also considerably lower than for the more conventional pneumatic actuators.

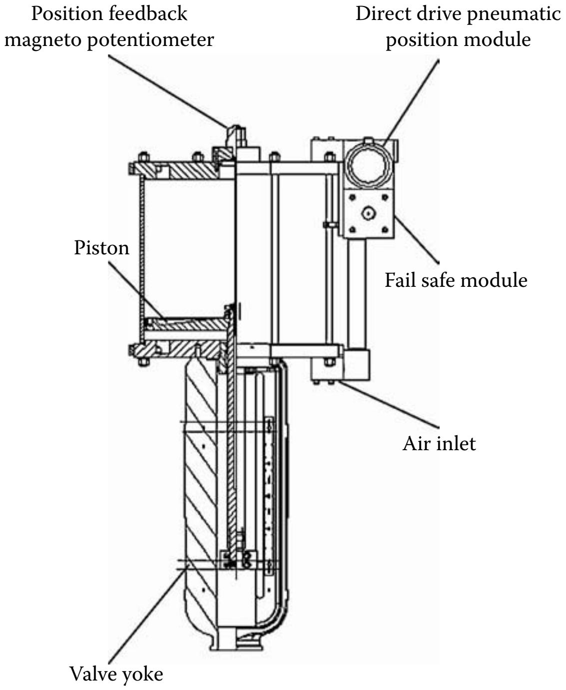


FIG. 6.4q
High-speed piston actuator without boosters. (Courtesy of Control Components Inc.)

Figure 6.4r shows a response curve for an actuator with a 14 in. (355 mm) travel in which full stroke is achieved in 1.3 sec. The dead time is less than 0.140 sec, and there is no visible overshoot when operating a relatively high friction valve. There are no mechanical linkages, tubing, or accessories

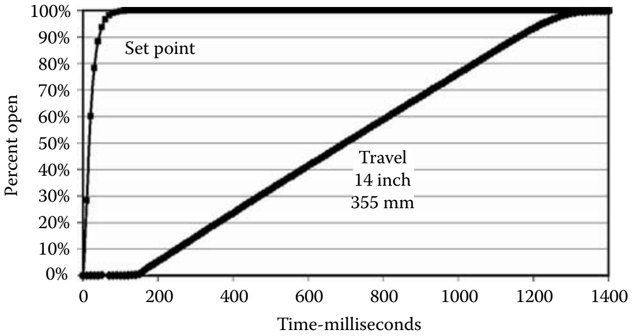


FIG. 6.4r
High speed actuator step change response. (Courtesy of Control Components Inc.)

to fail from vibration or from other environmental conditions. According to the supplier, all failure positions can be provided and diagnostic software is also available. (For more details, see Sections 6.8 on valve diagnostics, 6.11 on fieldbus interaction, and 6.12 on intelligent valves.)

The high-speed pneumatic actuators reduce the need to use hydraulic actuators, when the stroking time of a second or two is sufficient. One pneumatic actuator supplier feels that both electric and hydraulic high-speed actuators are more expensive than the pneumatic ones. (For a detailed discussion of hydraulic and electric actuators refer to Section 6.3.)

RELATIVE MERITS OF DIAPHRAGM AND PISTON ACTUATORS

Table 6.4a compared the features of diaphragm- and piston-type pneumatic actuators. When choosing between piston actuators and the spring-and-diaphragm type, the fail-safe consideration may be the reason for the final selection. If properly designed, the spring is the best way of achieving fail-closed action. Fail-open action is less critical.

Piston actuators may depend upon air lock systems to force the valve closed on air failure. Such systems may work well initially, but there are possibilities for leaks to develop in the interconnecting tubes, fittings, and check valves. Therefore, such piston actuator systems are not considered reliable by many. However, according to one manufacturer, field data on reliability shows the opposite to be true. Air lock systems also add to the actuator's cost. Piston actuators may also be specified with closure springs to provide positive failure positions.

Valve installation in the line is also a factor to consider. Flow over the plug assists in maintaining valve closure after air failure, but the considerations involving dynamic stability are more important. Therefore, the use of "flow-to-open" valves is recommended for most diaphragm actuators, as the actuator force is usually marginal.

Piston actuators are larger and require more space than do diaphragm actuators. This is particularly the case when the piston is provided with a spring to provide a positive failure position.

Both the diaphragm and piston actuators use manifolds and have become available in modular designs. This has somewhat reduced their costs and lowered their potential for leakage by reducing the number of connections that could leak. An example of the modular design of a linear actuator is shown in Figure 6.4s. The linear actuator includes a positioner, boosters, and adjustable quick exhaust.

Figure 6.4t illustrates the modular design of a rotary piston actuator. The rotary design is the balanced pinion type that can be provided with a plug-in positioner, which can be either the conventional or the smart design. The smart positioners also are available with bus communication options.

The modular designs provide flexibility, as they may be changed out in the field to accommodate process changes or upgrades or to correct errors in the original design. These modular designs are usually provided with position indicators, which are enclosed within the actuator housing. Because of these housings, reliability is better than in designs where the linkages are exposed. Most control modules will eventually include bus control communication and self-diagnostics capability.

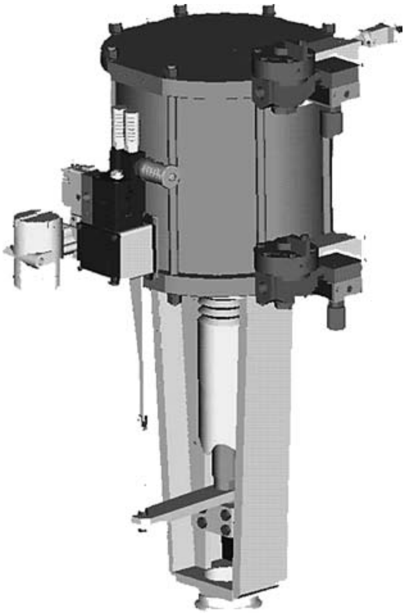


FIG. 6.4s
Piston actuator with modular manifolds. (Courtesy of Control Components Inc.)

Pneumatic vs. Hydraulic Actuators While the purpose of this section is to describe the features of pneumatic actuators and while a detailed discussion of hydraulic and electric actuator features is given in Section 4.3, a few comments

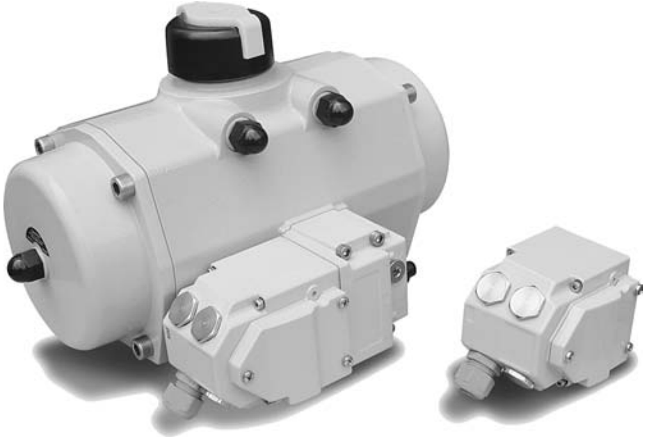


FIG. 6.4t
Rotary actuator provided with interchangeable control modules. (Courtesy of Emerson Process Management.)

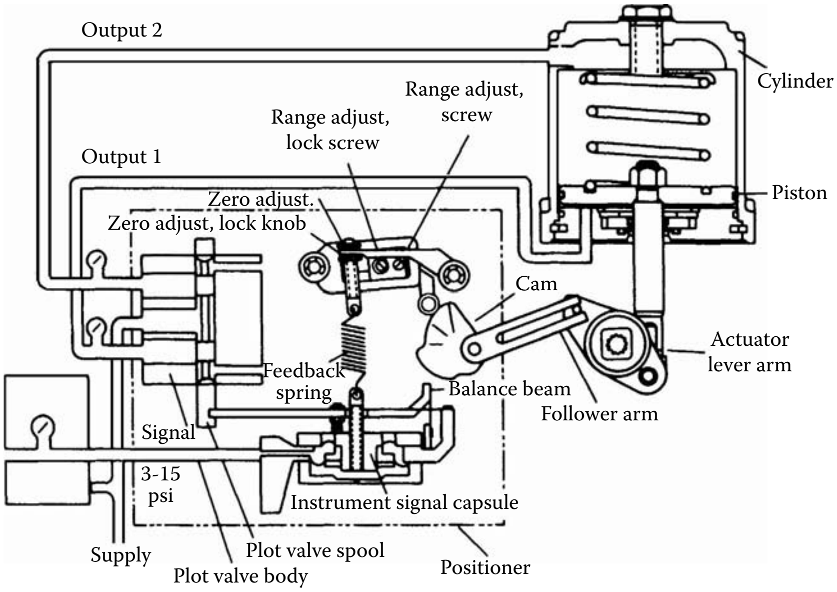


FIG. 6.4u
When linear actuators are used to operate rotary valves, a unit change in controller signal will not result in a unit change in rotation unless a positioner is used. (Courtesy of Flowserve Inc.)

from the perspective of a pneumatic piston actuator manufacturer is included here.

The main advantages of hydraulic actuators are speed and stiffness. This is the case because of the high density and the incompressibility of liquids in comparison to air. The speed difference between pneumatic and hydraulic actuators has been narrowed, as the stroking time of some pneumatic piston operators (particularly with dual pistons) is about 1 sec.

With the use of 100 PSIG (690 kPa) or higher air pressures, the piston actuator stiffness and stability have also improved and approach that of the hydraulic actuators. The stem forces provided by a pneumatic piston can equal or exceed those of the hydraulic cylinders, because the area of an oversized air piston can provide higher stem forces than a standard hydraulic actuator. For example, if the hydraulic cylinder area is one 50th of the air piston, while the air pressure is one 20th of the oil pressure, the stem force produced by the hydraulic actuator is actually less than that of the pneumatic actuator.

ROTARY VALVE ACTUATORS

When linear spring-and-diaphragm, piston, or cylinder actuators are used on rotary valves, their performance will not be linear unless a positioner is used (Figure 6.4u). By the addition of a positioner, one can guarantee that the ratio between a unit change in controller signal and the resulting rotation will be uniform. On the other hand, loose-fitting actuator lever and follower arms can still create dead play in the actuator, which will lower the responsiveness of the loop when the direction of change in the control signal reverses.

As can be seen in Figure 6.4v, the torques required to rotate ball, butterfly, or plug valves are not linear. These valves are usually used only for on/off applications. When considered for closed-loop throttling control, the nonlinear relationship between the air signal and the resulting rotation makes the use of a positioner essential.

Figure 6.4w shows the nonlinearity in the torque characteristics of double-acting and spring-loaded cylinder actuators.

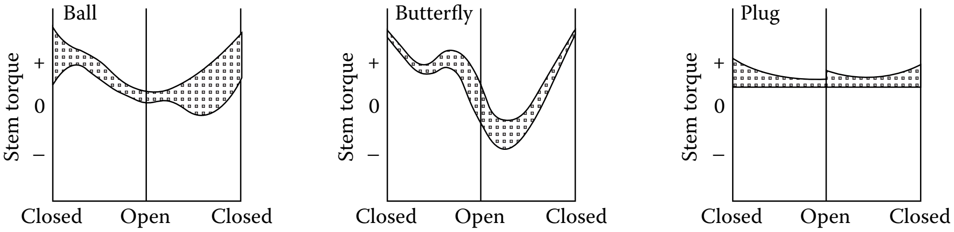


FIG. 6.4v
The torque characteristics of such rotary valves as ball, butterfly, and plug valves are not linear.

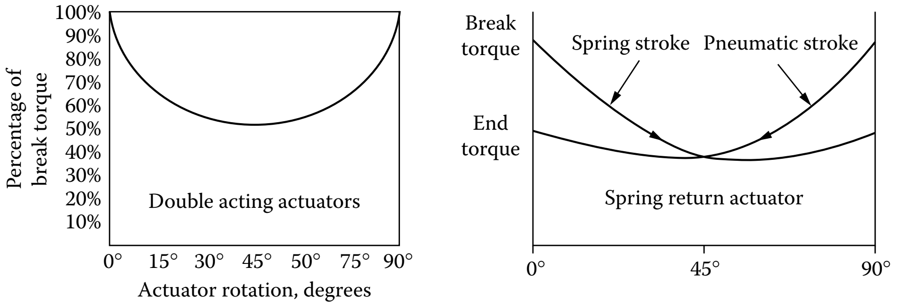


FIG. 6.4w
Torque characteristics of spring-return and double-acting cylinder actuators. (Courtesy of Rotork Controls Ltd.)

Naturally, the piston actuator should be so selected that its break torque exceeds the peak torque requirement of the valve, which occurs at the beginning and the end of the stroke (break torque).

The torque characteristics of the double-acting piston actuator of Figure 6.4x shows two maximums. One maximum occurs at the closed position and one near the 60–80° position, which corresponds to the peak of the dynamic torque for butterfly valves. Naturally, this actuator too should be so selected that its break torque exceeds the peak torque requirement of the valve, which occurs at the beginning (break torque) and near the end of the stroke.

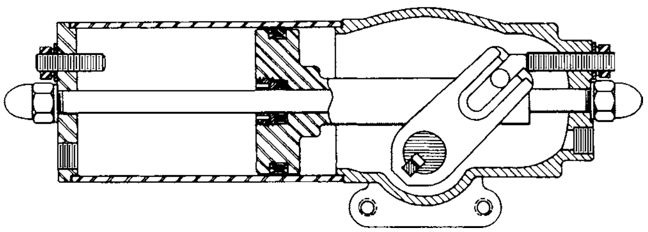


FIG. 6.4y
Pneumatic cylinder actuator with Scotch yoke.

Cylinder Type

Increased use of ball and butterfly valves or plug cocks for control has bred a variety of actuators and applications of existing actuators for powering these designs.

Positioning a quarter-turn valve with a linear output actuator using a lever arm on the valve resolves itself into a problem of mounting and linkages. The actuator can be stationary, with a bushing to restrain lateral movement of the stem. This requires a joint between the stem and a link pinned to the lever arm. The actuator can be mounted on a gimbal mechanism to allow required movement. The actuator can be hinged to allow free rotation to allow for the arc of the lever arm.

Various Scotch yoke designs, such as the one shown in Figure 6.4y, can be used with one, two, or four cylinders. Use of rollers in the slot of the lever arm utilizes the length of the lever arm of the valve opening or closing points.

A rack and pinion can be housed with the pinion on the valve shaft and the rack positioned by almost any linear valve actuator. The rack (Figure 6.4z) can be carried by a double-ended piston or by two separate pistons (Figure 6.4aa) in the same cylinder, where they move toward each other for counterclockwise rotation and away from each other for clockwise rotation.

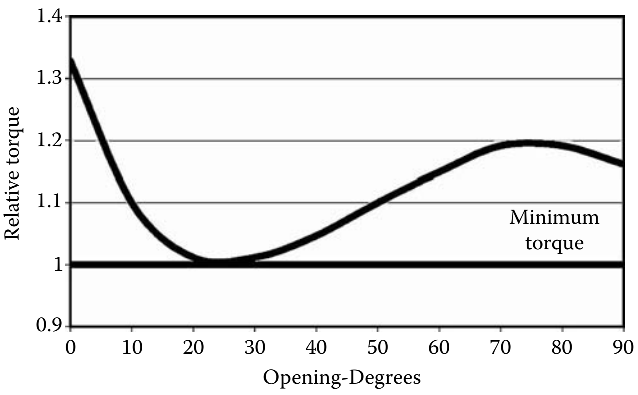


FIG. 6.4x
Double-acting piston actuators for rotary motion. (Courtesy of Metso Automation.)

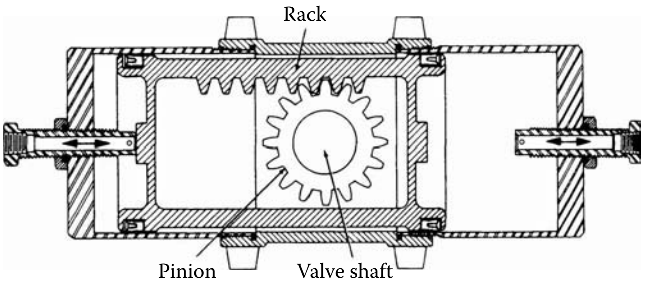


FIG. 6.4z
Rack and pinion actuator with dual-acting cylinder.

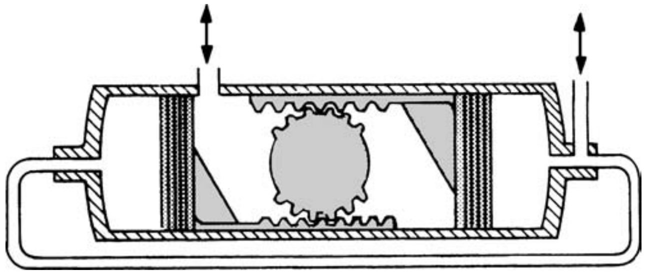


FIG. 6.4aa
Rack and pinion actuator operated by two separate pistons.

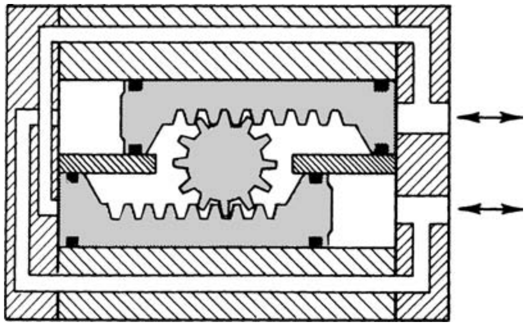


FIG. 6.4bb
Rack and pinion actuator operated by two parallel pistons.

Similar action is obtained (Figure 6.4bb) by two parallel pistons in separate cylinder bores. Dual cylinders are used in high-pressure actuators used to rotate ball valves as large as 16 in. (400 mm) in less than 0.5 sec.

An actuator similar to the one shown in Figure 6.4cc can be spring-loaded for emergency or positioning operation.

On/off operation of cylinders for quarter-turn valves requires solenoid or pneumatic pilots to inject pneumatic or hydraulic pressure into the cylinders. Open or closed position must be set by stops that limit shaft rotation or piston travel. Thereby the valve rotation is stopped and held in position until reverse action is initiated. Positioning action requires a calibrated spring in the piston or diaphragm actuator, a valve positioner, or a positioning valve system that loads and unloads each end of the cylinder.

Rotation of the valve must be translated to the positioner by gears, direct connection, cam, or linkage. The valve positioner must be the type that includes the four-way valve. The positioning valve system can be a four-way valve with a positioner for use with a pneumatic controller. The piston is sometimes positioned by a servo-system consisting of a servovalve that accepts an electronic signal, a four-way valve to amplify and control the pressure to the cylinder, and a feedback signal from a potentiometer or LVDT.

An electrohydraulic power pack or pneumatic pressure source may be used to furnish pressure to a pair of cylinders, one for each direction of rotation, as shown in Figure 6.4dd.

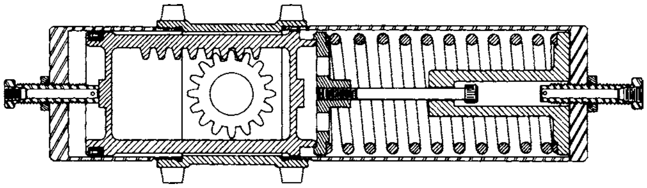


FIG. 6.4cc
Rack and pinion actuator with spring-loaded (fail-safe) cylinder.

Rotation by Spline or Helix

A multiple helical spline rotates through 90° as pressure below the piston (Figure 6.4ee) moves the assembly upward. A straight spline on the inside of this piston extension sleeve rotates the valve closure member through a mating spline. The cylinder is rated at 1500 PSIG (10 MPa). The actuator will rotate 1 in. (25 mm) and $1^1/_2$ in. (37.5 mm) valve stems. The mounting configuration is designed to adapt to many quarter-turn valves.

A nonrotating cylinder (Figure 6.4ff), with an internal helix to mate with a helix on a rotatable shaft, creates a form of rotating actuator. Hydraulic or pneumatic pressure in the drive end port (left) causes counterclockwise rotation; clockwise rotation is caused by pressure on the opposite side of the piston. There is a patented seal between the internal and the external bores of the cylinder and the external surface of the shaft. The unit is totally enclosed by seals to protect it from contaminated atmospheres. A hydraulic pump, reservoir, and necessary controls can be mounted integrally.

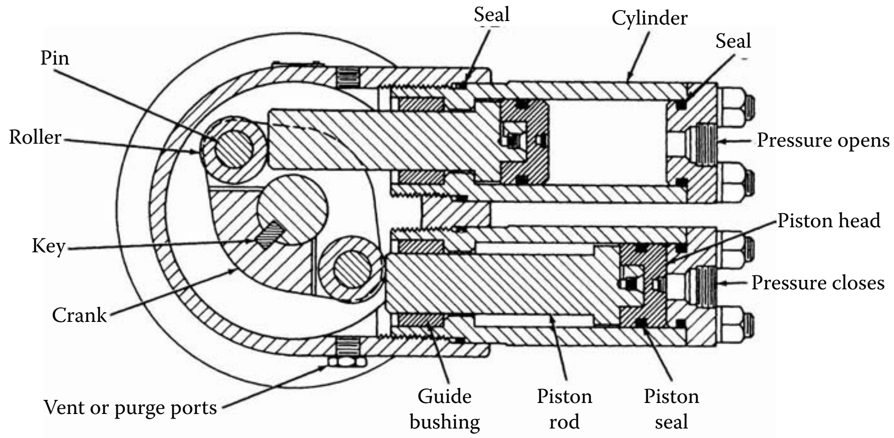
Vane Type

Injection of pressure on one side of a vane to obtain quarter-turn actuation is straightforward and obtainable with a minimum number of parts. A single vane (Figure 6.4gg) can be used for 400 to 30,000 in. lb_f (27 to 2060 N · m).

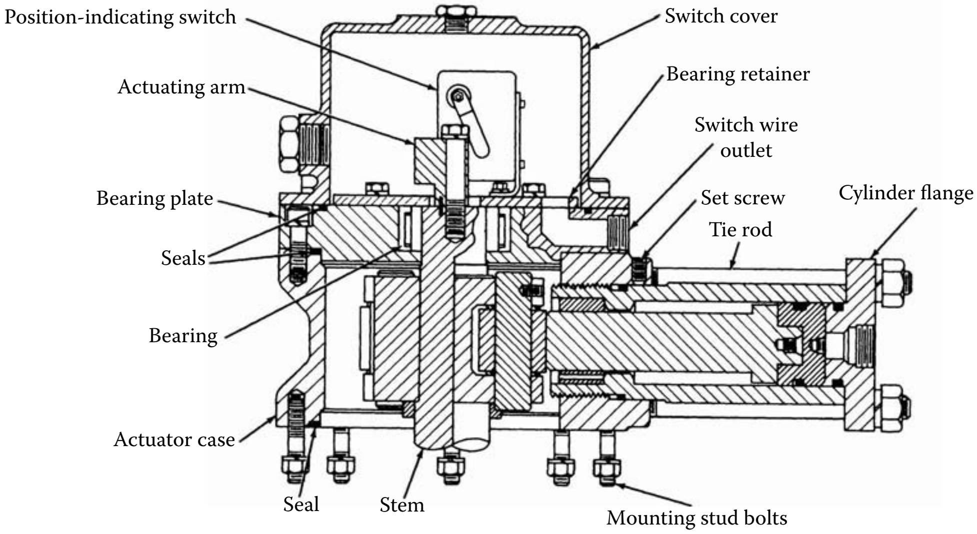
Units can be mounted together for double output, or a double-vane design (Figure 6.4hh) can also be used. The success of the vane actuator as a control device is dependent upon the control systems. By use of an auxiliary pneumatic pressure source, all types of fail-safe actions are possible, although not as positively as with spring loading. Use of line pressure to create hydraulic pressure on the vane is piloted by both manual and automatic methods. Use of a rotary potentiometer to sense position and complete a bridge circuit is necessary for proportional control.

Rotary Pneumatic Actuators

Pneumatic pressure is used to power a rotary motor to drive any of the large gear actuators. Control is by a four-way valve. The motor shown in Figure 6.4ii is running in one position. This will continue until the valve is repositioned or until a cam operates a shut-off valve at one end of the stroke. Reversal of the four-way valve causes reverse operation. An intermediate position causes the motor to stop. The four-way valve can be operated by pneumatic or electric actuators for remote automatic control. A position transmitter will allow adaptation to closed-loop proportional control.



Top section



Side section

FIG. 6.4dd
Crank and roller actuator operated by a pair of cylinders.

OTHER PNEUMATIC ACTUATORS

Pneumohydraulic Actuators

An actuator with two double-acting cylinders uses an integrally designed pneumatic or electric powerpack (Figure 6.4jj). This has the advantage of furnishing a constant hydraulic pressure to the cylinders regardless of the power source to the prime mover. Few cylinder sizes are needed to cover a wide range of torque outputs. The prime movers are sized and selected to obtain the actuator speeds desired with the pneumatic pressures or electric voltages available. Multiple auxiliary switches, position transmitter, and positioning devices are adapted to the unit.

Gas pressure is used to create hydraulic pressure using two bottles (Figure 6.4kk). The stability of a hydraulically operated cylinder is utilized in this manner, using line gas pressure and the bottle size for amplification. The manual control valve can be replaced by a variety of electric or pneumatic pilot valves for automatic control. A hand pump is furnished that can take over the hydraulic operation in the absence of gas pressure or malfunction of the pilot controls. This self-sustaining approach to cylinder operation finds wide application for line break shut-off and for the various diverting and bypass operations of a compressor station.

A hydrostatic system consisting of a pneumatic prime mover (Figure 6.4ll) on the shaft of a hydraulic pump to run a hydraulic motor has interesting features. This actuator incorporates many of the features of other high-force geared actuators that rotate a drive sleeve. Torque control consists of a relief valve in the hydraulic line to the motor. This eliminates the reactive force of spring-loaded torque controls. Starting torque occurs

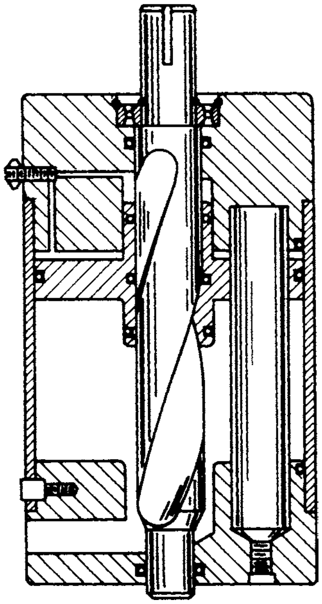


FIG. 6.4ee
Helical spline actuator.

because hydraulic slippage of the pump allows the motor to reach maximum speed.

Direction and deactivating control is attained with a four-way valve. As many as 16 auxiliary switches, settable at any position, are housed in the unit. Limit switches can be pneumatic or electric. A wide range of torque outputs and speeds is obtained by selection of prime mover, pump, and motor combinations. Initial success of the unit was partially due to its adaptability for retrofit to existing valves. The unit can be manually operated if required.

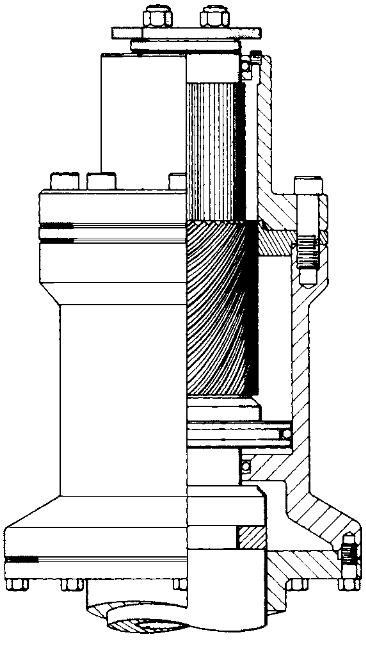


FIG. 6.4ff
Rotating helix actuator.

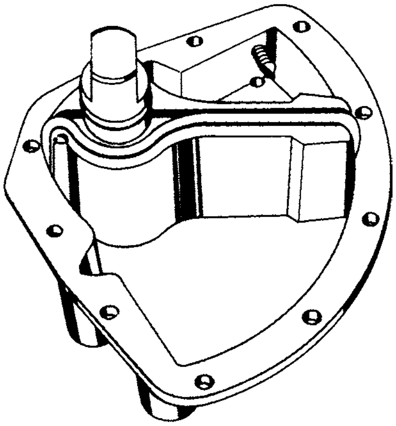


FIG. 6.4gg
Single-vane quarter-turn actuator.

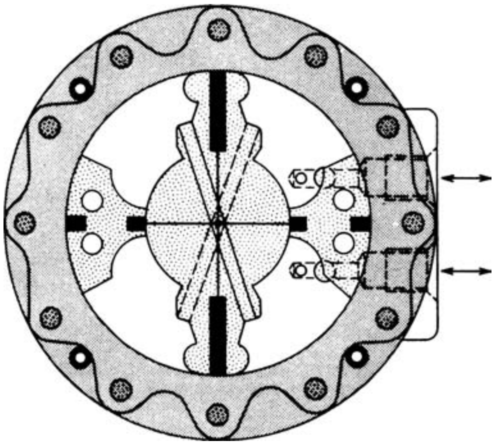


FIG. 6.4hh
Double-vane quarter-turn actuator.

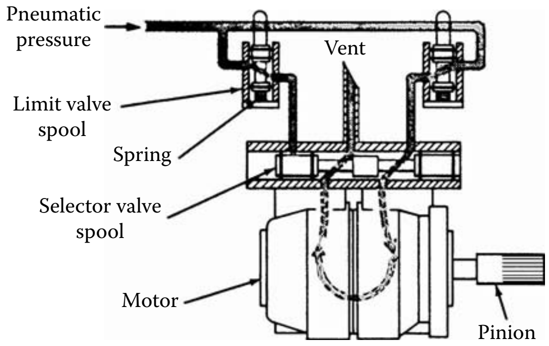


FIG. 6.4ii
Rotary air motor actuator.

Electropneumatic Actuators

An actuator that defies classification, except that it is pneumatically powered and electrically controlled for proportional

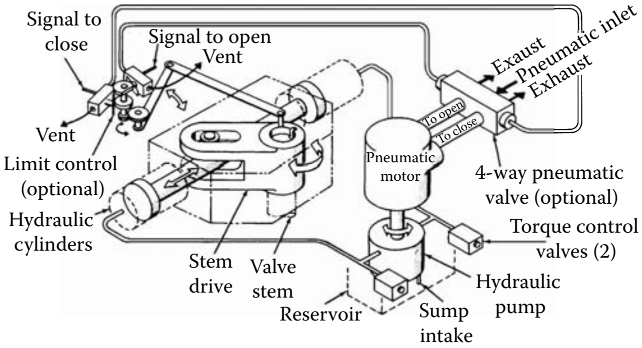


FIG. 6.4jj
Actuator with two double-acting cylinders and power pack.

application, is described at this point. Operation of a threaded drive sleeve occurs when spring-loaded pawls create a jogging action on a drive gear. Pressure introduced through one of the external lines selects the pawl to become active when the rocker arm is repetitively rocked by the pneumatic motor. A lead screw positions a sliding block to operate control switches, potentiometer, and position indicators. The lead screw is driven from a small spur gear and bevel gears.

Air is supplied from 60–140 PSIG (414–966 kPa) to give torque outputs up to 360 or 720 ft lb_f (25 or 50 Nm) using two actuators. Air consumption is from 0.75 SCF–1.70 SCF (0.021–0.048 m^3) per revolution at 140 PSIG (966 kPa). Maximum valve stem diameter that can be rotated is 2 in. (50 mm).

Numerous motivating combinations are used for control, including electropneumatic, electric, and fully pneumatic. A wide variety of components can be used to build up these systems, as shown in Figure 6.4mm.

RELIABILITY

Some of the reliability studies on the performance and reliability of pneumatically operated valves were done in the offshore oil industry[2] and in the power industry.[3,5] These studies did not evaluate only the actuator but the complete valve and actuator system combined.

The offshore study for pneumatically operated control valves was made in 1984 and reported the highest failure rate of 1 failure every 10,000 hours; or a little less than 1 failure per year. A 1990 study made for pneumatically operated control and on/off valves by the Institute of Nuclear Power Operation[4] (INPO) showed a failure rate of 116,000 hours.

In 1997 a summary on the reliability of air-operated valves by the electric power industry was published.[3] The sample consisted of 525 failures from 4,726 component years of operation. This would equate to a failure every 9 years or 79,000 hours of operation. The data covered all types of valves with both piston and diaphragm actuators from eight manufacturers. The study looked at what was called "high duty" (e.g., control valves) vs. "low duty" (e.g., shut-off valves) as well as the impact of intrusive maintenance. A summary of the results

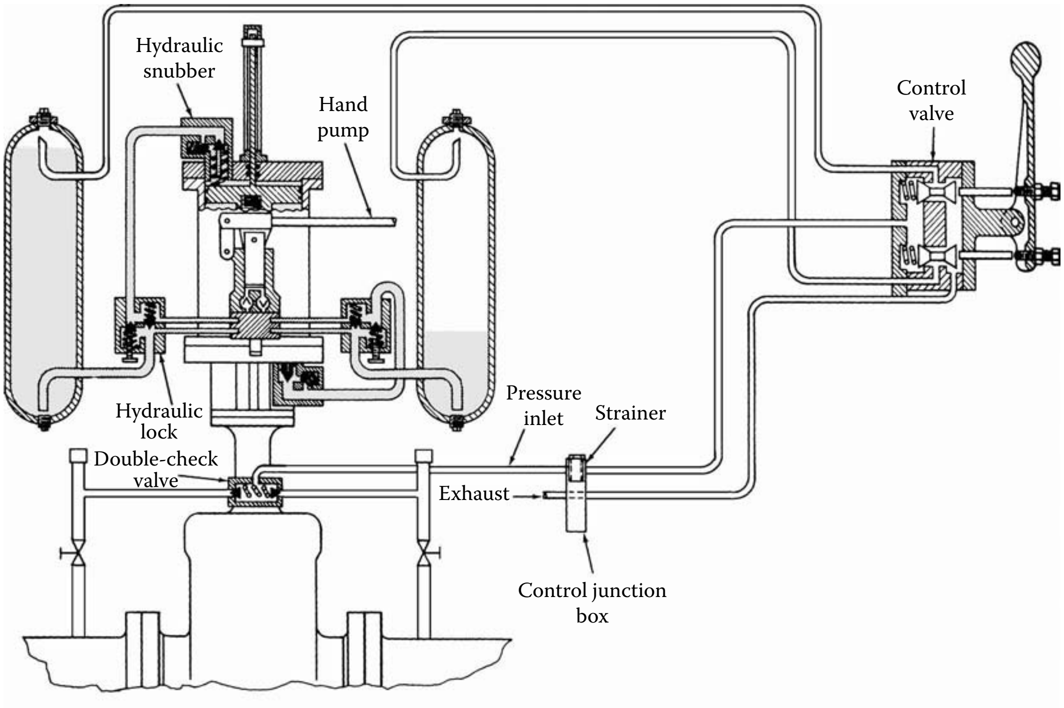


FIG. 6.4kk
Pneumohydraulic actuator powered by line pressure. (Courtesy of Shafer Value Co.)

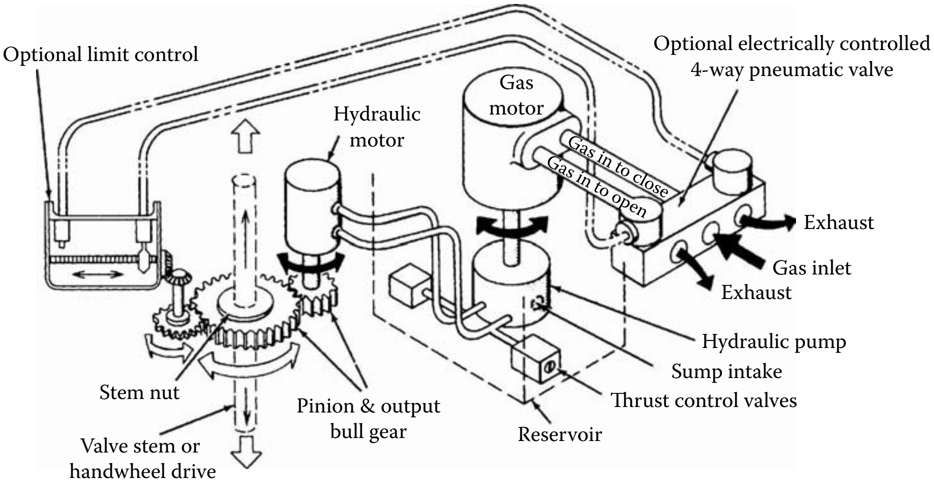


FIG. 6.4ll
Actuator system with pneumatically powered hydrostatic valve. (Courtesy of Ledeen Div. Textron.)

from this study is given below:

- High duty cycle valves exhibit up to 13 times lower reliability than low duty cycle valves.
- Intrusive preventive maintenance causes reliability to decrease by a factor of 6 times.
- The reliability of piston-type actuators was 4 times better than that of diaphragm-type actuators.
- The reliability of diaphragm-actuated valves was 6 times more likely to be impacted by duty cycle than were piston actuators.

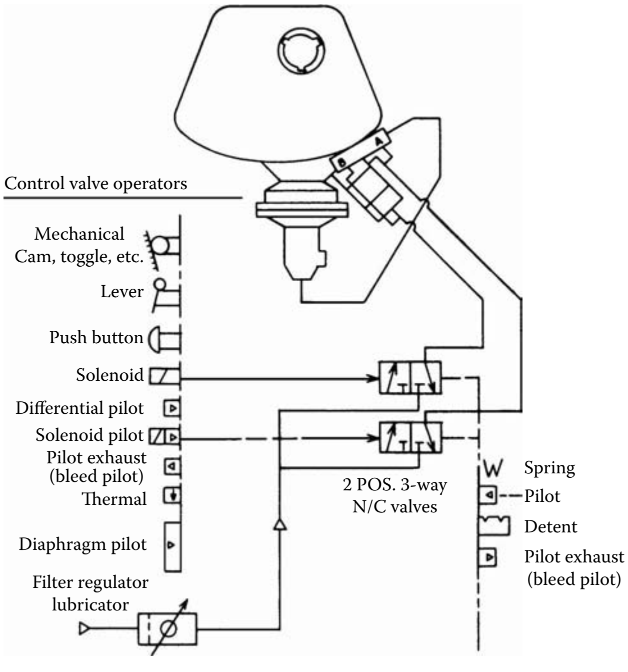


FIG. 6.4mm
Electropneumatic valve actuator. (Courtesy of OIC Corp.)

These results are a bit surprising because the common perception is that the diaphragm actuator is much more reliable that the piston actuator. It is also possible that by combining all types of valves (on/off and throttling) and by placing emphasis on duty cycle (opening and closing of the valves), the results of this evaluation would differ with one that evaluated only throttling control valves.

CONCLUSIONS

Before deciding on the type of control valve actuators to be used in a particular application, the reader is advised to also read Section 6.3, which discusses hydraulic, electric, and digital valve actuators, and to study Table 6.3a, which lists the advantages and disadvantages of electromechanical, electrohydraulic, and servo- or stepping-motor-operated electric actuators. Those who want to learn even more about the features and performance of available valve actuators are advised to study the test reports, books, and articles listed in the Bibliographies of both Sections 6.3 and 6.4.

As far as actuator trends are concerned, pneumatic actuators are still the technology favored by the users of valve actuators. Nevertheless, from 1998 to 2003, the market share of electric actuators increased from 11% to 45% and the market share of hydraulic and electrohydraulic actuators increased from near 0% to 28%. During the same period, use of solenoid valves has dropped by 42%.[6]

References

1. Lynch, J., “Impedance-Coupled Valve and Fluid System Instability,” American Nuclear Society, Pittsburg Meeting, May 1991.
2. “Offshore Reliability Data Handbook,” OREDA participants, Hovik, Norway, 1984.

3. Hinchcliffe, G. and Worledge, D., "Preventive Maintenance Program Basics: Air Operated Valves," Report TR-106857-V1, Electric Power Research Institute, Palo Alto, CA, July 1997.
4. Private communication between PLG, Inc. and Control Components Inc., 1994.
5. "IEEE Guide to the Collection and Presentation of Electrical, Electronic, Sensing Component, and Mechanical Equipment Reliability Data for Nuclear-Power Generating Stations," IEEE Standard 500-1984, The Institute of Electrical and Electronics Engineers, Inc., 1983.
6. Harrold, D., "A Changing Landscape," *Control Engineering,* December 2003 (www.manufacturing.net/ctl/article/CA339684?text=valve+actuators).

Bibliography

ANSI/ISA-75.25.01, "Test Procedure for Control Valve Response Measurement from Step Inputs."

ANSI/ISA-TR75.25.02-2000, "Control Valve Response Measurement from Step Inputs."

Barnes, P. L., "Protect Valves with Fire-Tested Actuators," *Instruments and Control Systems,* October 1982.

Baumann, H. D., "The Need for Pneumatic Power Signals for Control Valves," *InTech,* January 1988.

Baumann, H. D., "Trends in Control Valves and Actuators," *Instruments and Control Systems,* November 1975.

Carey, J. A., "Control Valve Update," *Instruments and Control Systems,* January 1981.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems,* January 1980.

Ritz, G., "Control Valve Actuator Options–Part 1," *Control,* June 1994.

"Control Valve Dynamic Specification," *Entech,* 1998.

Coughran, M. T., "Measuring the Installed Dead Band of Control Valves" *ISA,* 1997.

Fitzgerald, W. V., "Loop Tuning and Control Valve Diagnostics," Paper #91-0500, 1991 ISA Conference, Anaheim, CA, October 1991.

Grabau, T., "Superior Actuator Diaphragm Material for Control Valves in Low Temperature Applications," *Valve Magazine,* Fall 2002.

Hammitt, D., "How to Select Valve Actuators," *Instruments and Control Systems,* February 1977.

Heney, P., "Rotary Actuators: Turnabout Is Fair Play," *Hydrolics and Pneumatics,* April 2005.

Holtgraver, E., "Quarter-Turn Pneumatic Actuators: Torque 101," *Valve Magazine,* Summer 2004.

Howleski, M. F., "Adapting Electric Actuators to Digital Control," *Instrumentation Technology,* March 1977.

"Installing Smart Positioners–A Wise Move," *Chemical Engineering Magazine,* December 1, 2000.

Kalsi, M. S., Eldiwany, B., Sharma, V., and Somogyi, D., "Dynamic Torque Models for Quarter-Turn Air," "Operated Valves," Air Operated Valve Users' Group, 20th Conference, Clearwater Beach, FL: Kalsi Engineering, Inc., 2001.

Koechner, Q. V., "Characterized Valve Actuators," *Instrumentation Technology,* March 1977.

Langford, C. G., "A Method to Determine Control Valve Dynamic Requirements," *ISA,* 2002.

Lipták, B. G., "Control Valves in Optimized Systems," *Chemical Engineering,* September 5, 1983.

Ludwig, A. "User-Friendly Control Valve," Valve World 98 Conference, KCI Publishing BV, the Netherlands, 1998.

Merrick, R. C., *Valve Selection and Specification Guide,* Van Nostrand Reinhold, 1991.

Miller, S.F., "Electronic Valve Controller Replaces Conventional Pneumatic System," ISA Expo 2004, Paper ISA04-p, Instrumentation, Systems, and Automation Society, Research Triangle Park, North Carolina, Houston, Oct. 5–7.

Mowrey, A., "Advancements in Valve and Actuator Technologies," *Valve Magazine,* Fall 2001.

Patel, A., "Air Operated Actuators for High Cyclic Duty," Air Operated Valve users' Group, Clearwater Beach, FL, Jan. 12–14, 2005.

Price, V. E., "Smart Valve Intelligence Takes Many Forms," *InTech,* August 1992.

Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 ISA Conference, Houston, October 1992.

Scott, A. B., "Control Valve Actuators: Types and Applications," *InTech,* January 1988.

Samy, S. and Stemler, D., "Linear Piston Actuators," Air Operated Valve Users' Group, 21st Conference, Clearwater Beach, FL, Control Components Inc., 2002.

Shinskey, F. G., "Dynamic Response of Valve Motors," *Instruments and Control Systems,* July 1974.

"Valve Actuator Roundup," *InTech*, January 1988.

"'Smart' actuator incorporates all the controls in one package," *Chemical Engineering Magazine,* January 1, 2003.

6.5 Advanced Stem Packing Designs

J. B. ARANT (2005)

Designs and Sizes:	Chevron or V-rings, and rectangular rings. All valve and stem sizes available commercially. Special packing can be designed for any and all rotary or linear valve sizes.
Design Pressures:	Up to 2500 PSIG ANSI
Design Temperatures:	–450 to 1000°F (–268 to 538°C), depending upon the materials of construction and configuration of the packing system, and valve design. Teflon with a cooling bonnet can be used up to 850°F (454°C).
Packing Materials:	Teflon®, Kalrez®, Zymaxx®, and Expanded Graphite®
Cost:	The costs of Teflon and Expanded Graphite packing sets and rings are moderate. The KVSP packing is more costly on first costs. However, they are more economical overall due to their greatly reduced maintenance needs, longevity, and overall performance, with regard to "fugitive emissions" and the resulting valve dynamics.
Partial List of Suppliers:	Teflon, Kalrez, Zymaxx, and KVSP are all registered trademarks of E.I. Du Pont Co. (now, Du Pont- Dow Elastomers); Expanded or Flexible Graphite is a registered trademark of Union Carbide Co.

INTRODUCTION

Similarly to most industrial products, control valve packing has also evolved over time. The two main requirements of control valve stem packing are that 1) it should seal the valve stem to eliminate or minimize leakage, and 2) it should not influence the performance or dynamics of the control valve.

The ability of a control valve to respond to small control signal changes without overshooting and cycling is inversely proportional to packing friction. Therefore, the ideal packing would be one that is as tight (as leakage free) as a bellows seal and yet completely friction free. Real packings do have some emissions (Figure 6.5g) and do have some friction. The lower the friction, the smaller the dead band (resolution) and the shorter the response time of the packing (Table 6.5h).

Prior to 1950, stem leakage was a constant problem with these older packing designs. Today, both of the above requirements can be met. Modern seals can be extremely tight, providing performance approaching that of a bellows seal, and they can also be low friction. This guarantees not only the good dynamic response of the control valve, but also low maintenance and a long service life.

Today's packing is no longer just a packing, but it is a well-engineered part of the control valve and its sealing systems. They can be installed to operate under wide temperature and pressure ranges. They can be used in applications involving temperatures from cryogenic to very high temperatures. They can also operate under process pressures from vacuum to several thousands of PSIG.

History

The early control valves were provided with only one type of packing, which consisted of braided asbestos rings. These rings were formulated with various braiding designs, utilizing various types of asbestos, binders, and lubricants. For control valves, the best versions were used, such as Blue African asbestos with premium binders and braid lubricants.

These rings were installed in a very smooth, finished well inside a valve bonnet. The valve stem was well polished. The valve bonnet was fitted with a lubricator assembly that included a shut-off isolation valve and a "ram screw." Depending on the process fluid, the appropriate choice of "grease sticks" was inserted and the lubricant was forced into the packing area by the ram. After this, the isolating valve was closed to prevent the pressurized process fluid from escaping.

By today's standards, this was system, yet it was very workable. All instrument maintenance mechanics, and many of the instrument engineers and supervisors, routinely carried

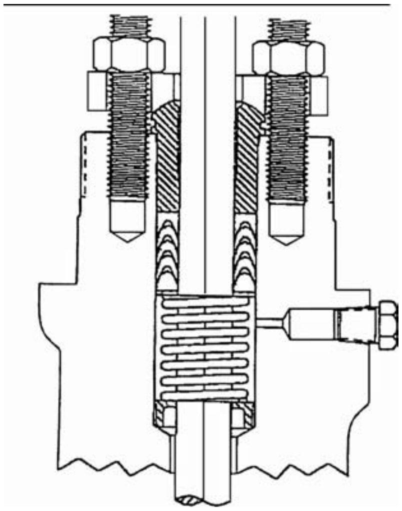


FIG. 6.5a
Valve stem packing made out of PTFE V-rings with internal coil spring located below the packing.

boxes of lubrication sticks in their pockets so as to be able to control stem leakage while guaranteeing a well-lubricated packing. It was a never-ending chore.

Teflon V-Rings In the mid- to late 1940s, the Du Pont Co. developed a fluorocarbon polymer, Teflon®, and it occurred to the control valve manufacturers that this material could be made into packing rings. These rings not only had the potential of providing a good seal, but were also inherently self-lubricating and resistant to essentially all process fluids. So the Teflon V-rings, or chevron rings, as they came to be known, were born (Figure 6.5a).

Because of its corrosion resistance to almost any process fluid, its self-lubrication, easy manufacturing via molding, and good temperature and pressure ratings, Teflon rapidly replaced asbestos as a packing material in many industries.

Essentially the only problem that arose was that this packing material was so "slick" and had such low friction that the control valve had a tendency to overshoot and oscillate at times. This was partially because these early valves were designed for use with the higher friction asbestos packing. The solution to this problem was to add some friction back into the Teflon packing. So, the packing gland was tightened a little, and the instrument mechanics learned when to do it.

Graphite Later, Union Carbide developed the Expanded Graphite®, or flexible carbon, materials that could be formed into shapes such as packing rings (Figure 6.5b). This increased the choice of packing.

Graphite can be used at higher operating pressures if the process fluids are the nonoxidizing-type fluids, although Teflon, by far, provided a superior control valve packing. Teflon could be used at temperatures up to 850°F, if the valve was provided with an extended cooling bonnet.

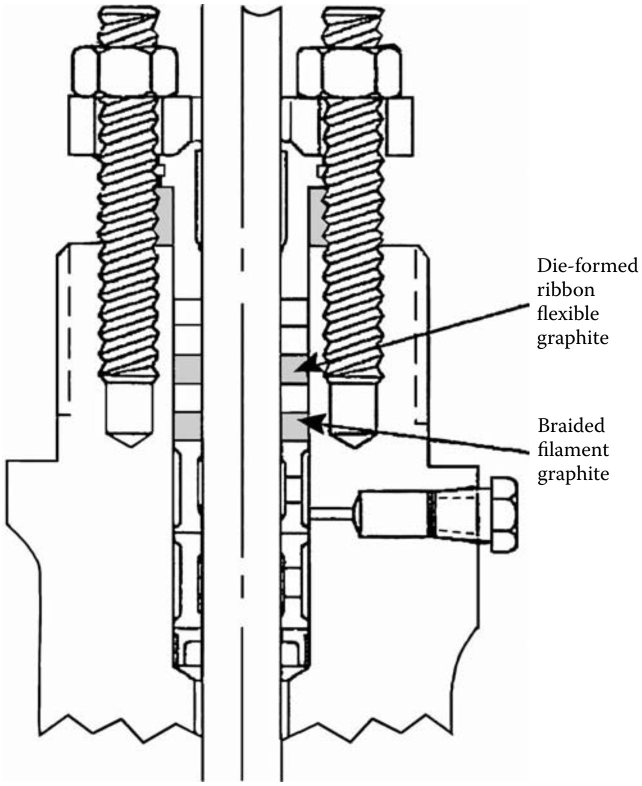


FIG. 6.5b
Graphite packing consisting of die-formed ribbons and braided filament graphite rings.

Graphite packing works very well with manually operated valves, but it gives poor performance on control valves, because it affects their dynamics and resolution. This happens because it tends to "plate out" on the valve stem. This further increases stem friction and necessitates increased loading of the packing in order to seal the stem. In addition, there can also be a fugitive emission problem with graphite packing.

The Enviro-Seal and the Dual Packing Another early development in the history of leak-resistant packing systems was one developed by Fisher Controls, using its Enviro-Seal® stem packing system (see Figures 6.5c and 6.5d).

While the development of the Enviro-Seal was a big step forward in valve stem sealing and in coping with leakage and fugitive emissions, it required about the same amount of packing seal maintenance as did the older packings. In addition its friction was much higher than that of Teflon, which negatively impacted its performance. This was reported in ISA (Instrumentation, Systems, and Automation Society) 75.25 Appendix A, which evaluated the link between packing friction and valve performance.

Nevertheless, the Enviro-Seal helped to "hold the fort" until better packing designs eventually came onto the marketplace.

Another interim and early improvement in the field of packing designs was the use of dual Teflon packing sets, with a lantern ring between the sets. The lantern ring could be

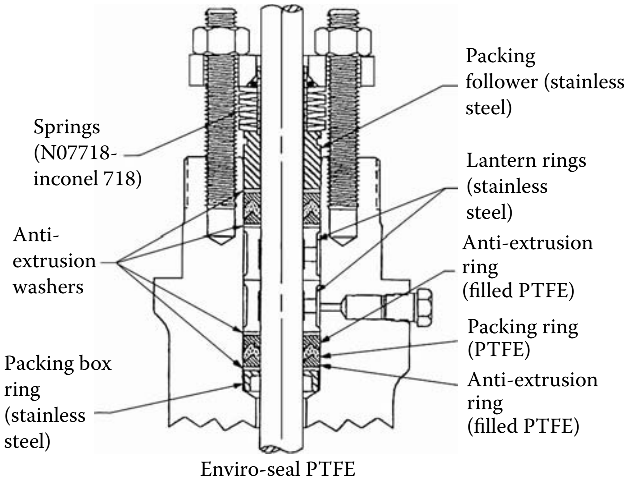


FIG. 6.5c
The Enviro-Seal was developed by Fisher Controls and included a PTFE packing ring and external Belleville disc springs.

used for sealing greases, or to allow piping to a leakage disposal system. While this design worked to some degree, it was also a somewhat clumsy and high maintenance system.

Industrial Practices In general, chemical and most other industries favored the Teflon packings, while the petroleum industry continued to favor asbestos and Expanded Graphite, because of their suitability for high-temperature and "fire safe" services, which are common in that industry. Their rationale was that "fire safety" should be their main concern, because they handled flammable fluids at high temperatures.

This logic was questionable, as the chemical industries also handled highly flammable fluids along with high temperatures and were just as vulnerable to fires. In the chemical industry a woven graphite ring was added at the top of the Teflon stack, with which these packings passed the API 607 valve fire test. With such packing designs, the chemical industry had years of experience with no particular problems.

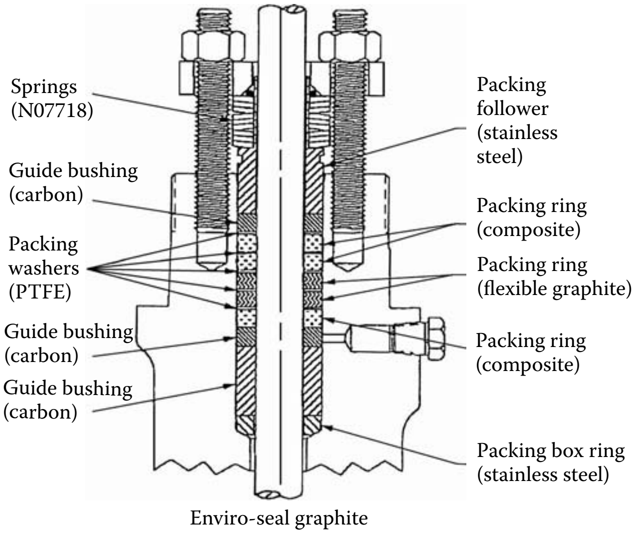


FIG. 6.5d
Fisher's Enviro-Seal provided with flexible graphite packing rings and external Belleville disc springs.

This difference in philosophies between the two industries persisted until recent years when the Federal Clean Air Act changed everything. Asbestos was phased out by the federal government as a "hazardous" material, and the petroleum industries were left with few packing options other than Expanded Graphite, plus some unsuccessful attempts at developing synthetic fibers as "asbestos substitutes."

Development of the KVSP Packing In the late 1980s, the Du Pont Co. and industry in general, came under pressure from the Federal Clean Air Act. The response was to again develop a new valve stem packing based on the use of forming chevron rings out of Kalrez®. This was a fluroelastomer material that had all of the corrosion-resistance properties of Teflon, plus a higher temperature rating along with the elastomeric characteristics. At that time, the primary use of Kalrez was in O-rings and other sealing shapes.

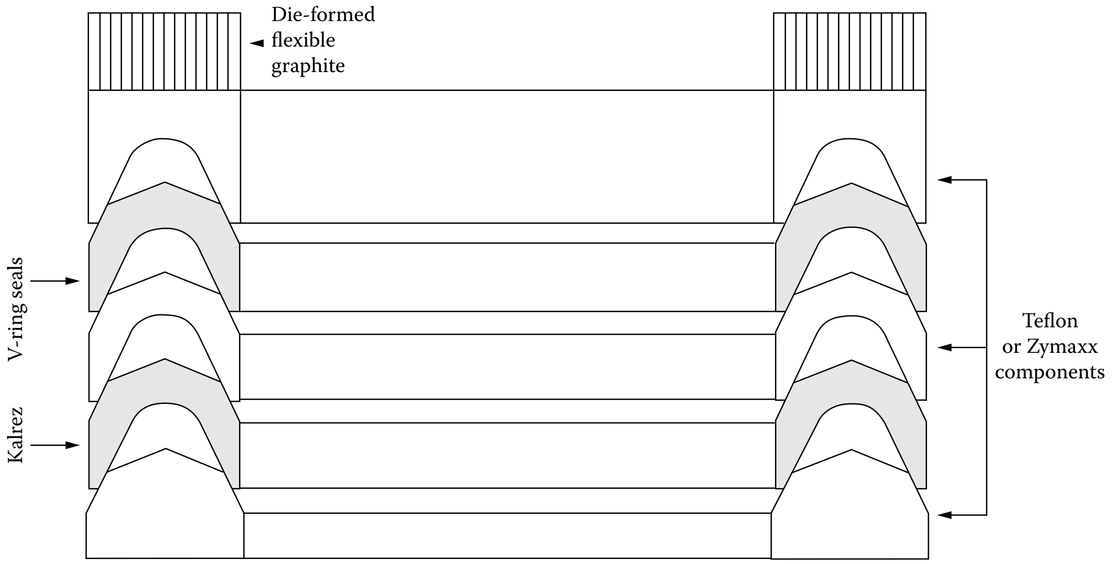
During these years, because of air pollution concerns, the federal government was trying to eliminate or to minimize more of the "fugitive emissions" from valve and pipe joint seals. Therefore, extensive tests were conducted and it was found that Kalrez Valve Stem Packing (KVSP), this new valve stem packing, did an excellent job in reducing emissions far below the levels of the federal and state (California) clean air standards. According to the testing per the Instrumentation, Systems, and Automation Society 75.25.01 and 75.25.02 KVSP performed essentially as well as Teflon. Thus was born the KVSP (Figure 6.5e).

After very extensive testing, using helium to simulate the most severe process plant leakage conditions, both the process industries and the control valve manufacturers started to use KVSP packing as a premium packing for solving their valve stem fugitive emission problems while still maintaining the dynamic response of Teflon packing. By replacing non-Teflon valve stem packing with KVSP, process plants and even California refineries in populated areas were able to meet EPA and Federal Clean Air Act fugitive emission standards.

Figure 6.5f illustrates the improvement that KVSP represents in terms of packing friction relative to graphite. As can be seen from the figure, the variability has been more than cut in half by using KVSP.

Further testing, development, and field experience with the KVSP packing led to the "fire-safe" design, per the API 607 Standard, in addition to the high-temperature (KVSP 500) and the creep- and extrusion-resistant (KVSP 400/500) variations. Specific combinations of Kalrez, Teflon, Zymaxx® (a high strength Teflon composite reinforced with long carbon fiber structural composites), and expanded woven graphite rings provided the solutions to many specific applications.

While Teflon and Expanded Graphite rings are still extensively used for control valve packing, KVSP packing is making



• KVSP 400 Series	Kalrez/Teflon
• KVSP 500 Series	Kalrez/Zymaxx
• KVSP Fire-Safe	Kalrez/Zymaxx/graphite

Note: consists of Kalrez and Teflon PTFE/Zymaxx
Maximum continuous temperature use 204°C (400°F)/288°C (550°F)

FIG. 6.5e
The KVSP packing system consists of Teflon or Zymaxx components and Kalrez V-ring seals below a flexible graphite ring.

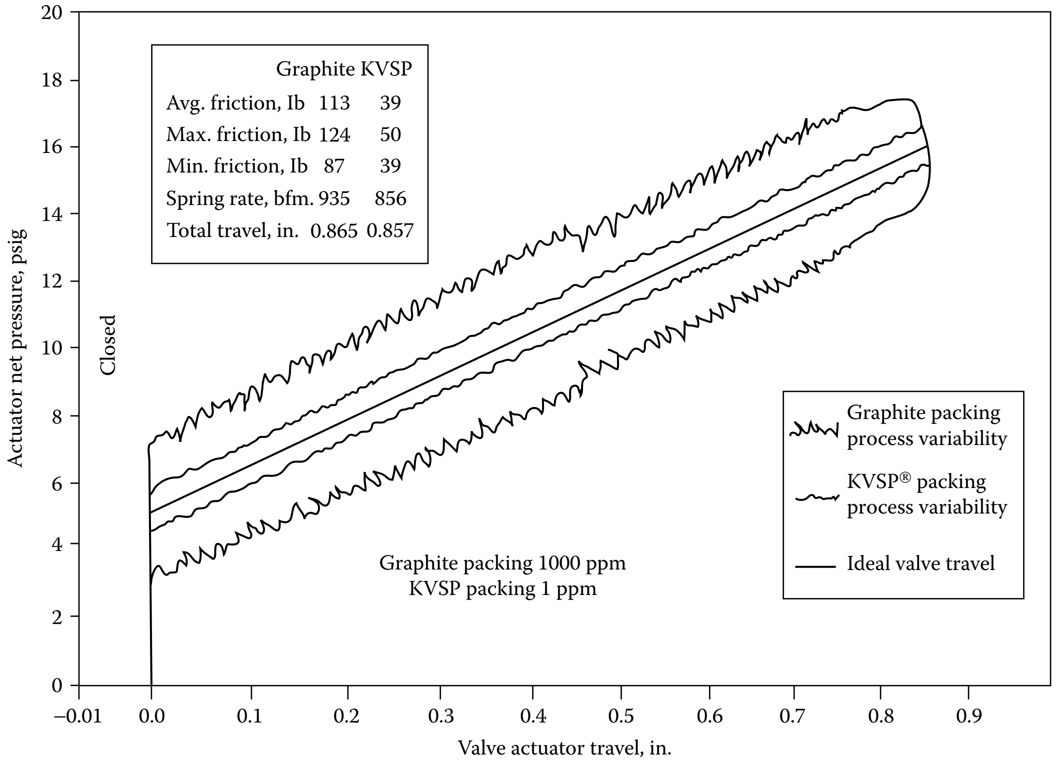


FIG. 6.5f
In addition to better controlling fugitive emissions, KVSP also reduces the packing friction effects on process control (variability). The curves compare the variability of the valve travel in response to increasing actuator pressure, when graphite and KVSP packing is used.

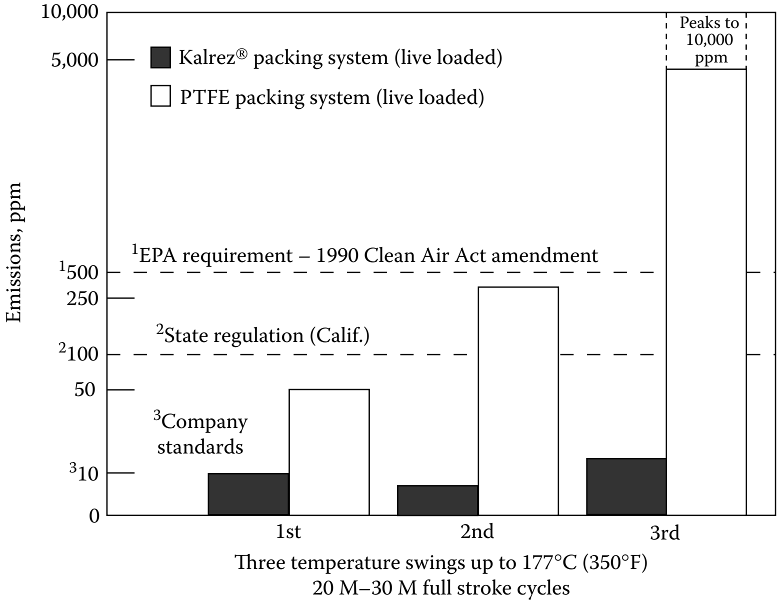


FIG. 6.5g
Emissions caused by creep and cold flow during temperature swings are much reduced when Kalrez packing is used instead of PTFE. These emissions satisfy both EPA's Clean Air Act and the California state regulations.

rapid inroads in many areas due to its superior capabilities and lower overall "ownership costs." One could say that KVSP is the nearest thing to a "universal" valve stem packing material today.

This packing works well on both linear and on rotary stem control valves, and is suitable to practically all process fluids. Some users have adopted it as their primary packing, which is used in all applications except for the more benign and innocuous ones such as water, where Teflon is more economical. Field retrofitting kits are available for valves to upgrade spring-loaded Teflon packings in existing valves with KVSP.

KVSP® Performance The major attributes of this packing, patented by Du Pont-Dow Elastomers, are as follows:

- When using it, the emissions are not measurable and exceed both the federal and California State Clean Air emission standards.
- Its resistance to HAPS and toxins is equivalent to that of Teflon.
- Its leakage containment is almost as good as that of a bellows.
- Its packing friction is very low.
- Its dynamic response is equivalent to that of Teflon.
- Its process variability of 0.1% is also equivalent to that of Teflon.
- When using it, the plant yield and throughput increases, because it has as little impact on process control dynamics and resolution as does Teflon.

A major performance advantage of KVSP over Teflon is that it is far less sensitive to leakage caused by thermal cycling at high temperatures. Figure 6.5g shows the emissions resulting from temperature swings when using KVSP and Teflon. It can be noted that because of creep and cold flow, the emissions using KVSP (Kalrez) are much reduced.

Table 6.5h provides a summary of the performance testing results that was published in Instrumentation, Systems, and Automation Society 75.25. From the table one can see that the ability of a control system to respond to small control signal changes is inversely proportional to packing friction, and the lower it is, the better control dynamics can be expected. This is because with lower packing friction, the valve's step response is faster and its dead band (resolution) is smaller.

Kalrez and PTFE V-rings are effective not only on control valves, but also for the stem packing of automatic on/off and manual valves, if fugitive emissions are to be minimized. As

TABLE 6.5h
Performance of Leading Packing Designs

Packing Type	*Dead Band or Resolution*	*Step Response Time*
Fisher's Graphite Enviro-Seal	0.6–1.2%	0.5–7.7 sec
Fisher's PTFE Enviro-Seal	0.2–0.4%	0.5–4.4 sec
PTFE V-Packing	0.0–0.1%	0.5–4.4 sec
KVSP Packing	0.1–0.2%	0.4–2.9 sec

Data Based on ISA 75.25 Control Valve Performance Testing Summary to the number of packing rings, five-ring packing sets are the norm. Three-ring packing can be used in many rotary stem valve applications, because the rotary stem movement is easier to seal against leakage than are the linear stems. Yet, for reasons of convenience and standardization, the five-ring design is the most often used.

Bibliography

ANSI/ISA 75.25.01-2000, "Test Procedure for Control Valve Response Measurement from Step Inputs," 2000.

ANSI/ISA-TR75.25.02-2000, "Control Valve Response Measurement from Step Inputs," 2000.

Brestal, R. et al., "Control Valve Packing Systems," technical monograph 38, Marshalltown, IA: Fisher Controls, 1971.

Control Valve Dynamic Specification, EnTech Control Inc., Toronto, Canada.

Coughran, M. T., "Measuring the Installed Dead Band of Control Valves," ISA TECH/97–1114.

ISA S26, 1968, and ANSI MC 4.1, 1975, "Dynamic Response Testing of Process Control Instrumentation,"

Langford, C. G., "A User's View of Process Control and Control Valve Positioners," ISA 1996–Paper #96-076.

"Packing Selection Guidelines for Sliding Stem Valves," Bulletin 59.1:062, Marshalltown, IA: Fisher-Rosemount, March 2005.

Senior, K. A., "Technical Guidelines and Design Information, Using KVSP Packing Systems for Improving Process Control And Minimizing Fugitive Emissions," white paper, DuPont Dow LLC, 2000.

Senior, K. A., "Valve Packing Systems Improve Process Control," *Chemical Processing,* June 1997.

6.6 Capacity Testing

(Copyright ISA 1996, reprinted with permission of ISA, The Instrumentation, Systems, and Automation Society. Standard formatted for publication in this handbook by A. Bálint, 2005)

INTRODUCTION

The "Control Valve Capacity Test Procedure" is covered by Instrumentation, Systems, and Automation Society's (ISA's) Standard S75.02-1996. The international equivalent of these U.S. standards are IEC 60534-2-1, "Industrial Process Control Valves, Part 2-1: Flow Capacity Sizing Equations for Fluid Flow under Installed Conditions," and IEC 60534-2-3, "Industrial Process Control Valves, Part 2-3, Flow Capacity Test Procedure." The equations in the international editions are essentially equivalent to those in the U.S. standards, with the exceptions of using metric units.

A revision of the ISA standard is anticipated in the near future, and for this reason only the front part of S75.02-1996 is reprinted below. The reader is advised to refer the revised standard (when published) for the test procedures and calculations. Meanwhile, Section 6.15 can be used for all valve sizing applications, while Section 6.14 provides all the information needed for valve noise calculations.

The reader should also note that in the ISA test procedure the pressure taps are not located at the in- and outlets of the control valve, but two pipe diameters upstream and six pipe diameters downstream. Therefore, the pressure differential measured during the test is the sum of the valve pressure drop (ΔP_v) and the drop through eight pipe diameters' length of straight piping (ΔP_p). This can introduce errors, if the pressure drop in the pipe exceeds 5–10% ($\Delta P_p / (\Delta P_p + \Delta P_v) > 0.1$).

This section is reprinted from "Control Valve Capacity Test Procedure," ANSI/ISA-S75.02-1996, with permission of the copyright holder, ISA, The Instrumentation, Systems, and Automation Society, 67 Alexander Drive, P.O. Box 12277, Research Triangle Park, NC 27709.

1. SCOPE

This test standard utilizes the mathematical equations outlined in ANSI/ISA-S75.01, "Flow Equations for Sizing Control Valves," in providing a test procedure for obtaining the following:

a. Valve flow coefficient, C_v
b. Liquid pressure recovery factors, F_L and F_{LF}
c. Reynolds number factor, F_R
d. Liquid critical pressure ratio factor, F_F
e. Piping geometry factor, F_P
f. Pressure drop ratio factor, X_T and X_{TP}

This standard is intended for control valves used in flow control of process fluids and is not intended to apply to fluid power components as defined in the National Fluid Power Association (NFPA) Standard T.3.5.28-1977.

2. PURPOSE

The purpose of this standard is to provide a procedure for testing control valve capacity and related flow coefficients for both compressible and incompressible fluids. This standard also provides a procedure to evaluate the major data.

3. NOMENCLATURE

(For the actual control valve sizing equations for both compressible and incompressible fluids, the reader should refer to Section 6.15)

4. TEST SYSTEM

4.1 General Description

A basic flow test system as shown in Figure 6.6a includes

a. Test specimen
b. Test section
c. Throttling valves
d. Flow-measuring device
e. Pressure taps
f. Temperature sensor

4.2 Test Specimen

The test specimen is any valve or combination of valve, pipe reducer, and expander or other devices attached to the valve body for which test data are required (Figure 6.6a). Modeling of valves to a smaller scale is an acceptable practice in this standard, although testing of full-size valves or models is preferable. Good practice in modeling requires attention to significant relationships such as Reynolds number, the Mach number where compressibility is important, and geometric similarity.

Symbol	*Description*
C_v	Valve flow coefficient
d	Value inlet diameter
D	Internal diameter of the pipe
F_d	Valve style modifier
F_F	Liquid critical pressure ratio factor, dimensionless
F_k	Ratio of specific heats factor, dimensionless
F_L	Liquid pressure recovery factor of a valve without attached fittings, dimensionless
F_{LP}	Product of the liquid pressure recovery factor of a valve with attached fittings (no symbol has been identified) and the piping geometry factor, dimensionless
F_P	Piping geometry factor, dimensionless
F_R	Reynolds number factor, dimensionless
G_f	Liquid specific gravity at upstream conditions (ratio of density of liquid at flowing temperature to density of water at 15.6°C [60°F]), dimensionless
G_g	Gas specific gravity (ratio of flowing gas to density of air with both at standard conditions, which is equal to the ratio of the molecular weight of gas to the molecular weight of air), dimensionless
K	Ratio of specific heats, dimensionless
m	The number of similar flow paths (i.e., $m = 1$ for single-ported valves, $m = 2$ for double-ported, etc.)
N_1, N_2, etc.	Numerical constants for units of measurement used
p_1	Upstream absolute static pressure, measured two nominal pipe diameters upstream of valve-fitting assembly
p_2	Downstream absolute static pressure, measured six nominal pipe diameters downstream of valve-fitting assembly
Δp	Pressure differential, $p_1 - p_2$
p_v	Absolute vapor pressure of liquid at inlet temperature
q	Volumetric flow rate
$q_{\max}$	Maximum flow rate (choked flow conditions) at a given upstream condition
Re_v	Valve Reynolds number, dimensionless
T_1	Absolute upstream temperature (in K or degrees R)
x	Ratio of pressure drop to absolute inlet pressure ($\Delta p/p_1$), dimensionless
x_T	Pressure drop ratio factor of the valve without attached fittings, dimensionless
x_{TP}	Value of x_T for valve-fitting assembly, dimensionless
Y	Expansion factor, ratio of flow coefficient for a gas to that for a liquid at the same Reynolds number, dimensionless
ν(nu)	Kinematic viscosity, centistokes

Subscripts:

1. Upstream conditions
2. Downstream conditions

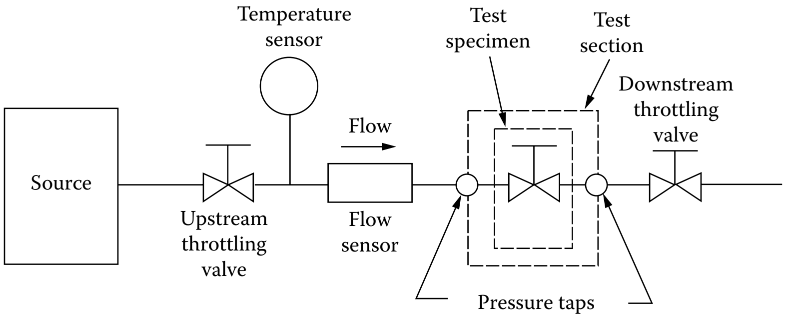
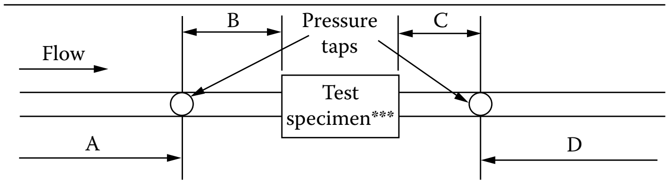


FIG. 6.6a
Basic flow test system.

TABLE 6.6b
Piping Requirements, Standard Test Section

A and ***	*B*	*C*	*D*	*Standard Test Section Configuration*
At least 18 nominal pipe diameters of straight pipe	2 nominal pipe diameters of straight pipe	6 nominal pipe diameters of straight pipe	At least 1 nominal pipe diameter of straight pipe	

*Dimension "A" may be reduced to 8 nominal diameters if straightening vanes are used. Information concerning the design of straightening vanes can be found in "ASME Performance Test Code PTC 19.5-1972, Applications, Part II of Fluid Meters, Interim Supplement on Instruments and Apparatus."

**If an upstream flow disturbance consists of two ells in series and they are in different planes Dimension "A" must exceed 18 nominal pipe diameters unless straightening vanes are used.

***See Section 4.2 for definition of the test specimen.

4.3 Test Section

The upstream and downstream piping adjacent to the test specimen shall conform to the nominal size of the test specimen connection and to the length requirements of Table 6.6b.

The piping on both sides of the test specimen shall be Schedule 40 pipe for valves through 250 mm (10 in.) size having a pressure rating up to and including ANSI Class 600. Pipe having 10 mm (0.375 in.) wall may be used for 300 mm (12 in.) through 600 mm (24 in.) sizes. An effort should be made to match the inside diameter at the inlet and outlet of the test specimen with the inside diameter of the adjacent piping for valves outside the above limits.

The inside surfaces shall be reasonably free of flaking rust or mill scale and without irregularities that could cause excessive fluid frictional losses.

4.4 Throttling Valves

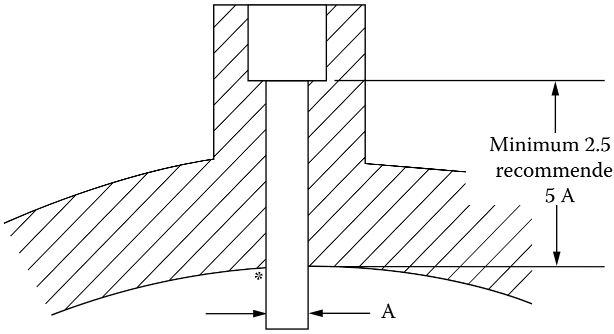
The upstream and downstream throttling valves are used to control the pressure differential across the test section pressure taps and to maintain a specific downstream pressure. There are no restrictions as to style of these valves. However, the downstream valve should be of sufficient capacity to ensure that choked flow can be achieved at the test specimen for both compressible and incompressible flow. Vaporization at the upstream valve must be avoided when testing with liquids.

4.5 Flow Measurement

The flow-measuring instrument may be any device that meets specified accuracy. This instrument will be used to determine the true time average flow rate within an error not exceeding ±2% of the actual value. The resolution and repeatability of the instrument shall be within ±0.5%. The measuring instrument shall be calibrated as frequently as necessary to maintain specified accuracy.

4.6 Pressure Taps

Pressure taps shall be provided on the test section piping in accordance with the requirements listed in Table 6.6b. These pressure taps shall conform to the construction illustrated in Figure 6.6c.



Size of pipe	A Not exceeding	A Not less than
Less than 50 mm (2 in.)	6 mm (1/4 in.)	3 mm (1/8 in.)
50 mm to 75 mm (2 to 3 in)	9 mm (3/8 in.)	3 mm (1/8 in.)
100 mm to 200 mm (4 to 8 in.)	13 mm (1/2 in.)	3 mm (1/8 in.)
250 mm and greater (10 in. and greater)	19 mm (3/4 in.)	3 mm (1/8 in.)

FIG. 6.6c
Recommended pressure connection.

* Edge of hole must be clean and sharp or slightly rounded, free from burrs, wire edges or other irregularities. In no case shall any fitting protrude inside the pipe. Any suitable method of making the physical connection is acceptable if above recommendations are adhered to. Reference: "ASME Performance Test Code PTC 19.5-1972. Applications, Part II of Fluid Meters, Interim Supplement on Instruments and Apparatus."

Orientation:

Incompressible fluids—Tap center lines shall be located horizontally to reduce the possibility of air entrapment or dirt collection in the pressure taps.

Compressible fluids—Tap center lines shall be oriented horizontally or vertically above pipe to reduce the possibility of dirt or condensate entrapment.

Multiple pressure taps can be used on each test section for averaging pressure measurements. Each tap must conform to the requirements in Figure 6.6c.

4.7 Pressure Measurement

All pressure and pressure differential measurements shall be made to an error not exceeding ±2% of actual value. Pressure-measuring devices shall be calibrated as frequently as necessary to maintain specified accuracy.

Pressure differential instruments are required in the measurement of the pressure differential across the test specimen to avoid additional inaccuracies resulting from taking the difference of two measurements. Exceptions to this are the procedures in Sections 6.2 and 8.2 for determining maximum flow rates for incompressible and compressible flow, respectively. (These sections are not included here, the reader is referred to ANSI/ISA-75.02-1996 for guidance in these procedures.)

4.8 Temperature Measurement

The fluid temperature shall be measured to an error not exceeding ±1°C (±2°F) of actual value.

The inlet fluid temperature shall remain constant within ±3°C (±5°F) during the test run to record data for each specific test point.

4.9 Installation of Test Specimen

The alignment between the centerline of the test section piping and the centerline of the inlet and outlet of the test specimen shall be as listed in Table 6.6d below:

When rotary valves are being tested, the valve shafts shall be aligned with test section pressure taps.

Each gasket shall be positioned so that it does not protrude in the flow stream.

4.10 Accuracy of Test

Valves having an N_3C_v/d^2 ratio of less than 30 will have a calculated flow coefficient, C_v, of the test specimen within a tolerance of ±5%.

TABLE 6.6d
Allowable Misalignment between the Centerlines of the Test Specimen and of the Test Section Piping

Pipe Size	*Allowable Misalignment*
15–25 mm	0.8 mm
($^1/_2$–1 in.)	($^1/_{32}$ in.)
32–150 mm	1.6 mm
($1^1/_4$–6 in.)	($^1/_{16}$ in.)
200 mm and larger 8 in. and larger)	1% of the diameter

5. TEST FLUIDS

5.1 Incompressible Fluids

Fresh water shall be the basic fluid used in this procedure. Inhibitors may be used to prevent or retard corrosion and to prevent the growth of organic matter. The effect of additives on density or viscosity shall be evaluated by computation using the equations in this standard. The sizing coefficient shall not be affected by more than 0.1%. Other test fluids may be required for obtaining F_R and F_F.

5.2 Compressible Fluids

Air or some other compressible fluid shall be used as the basic fluid in this test procedure. Vapors that may approach their condensation points at the vena contracta of the specimen are not acceptable as test fluids. Care shall be taken to avoid internal icing during the test.

SUMMARY

The international equivalent of this U.S. standard are IEC 60534-2-1, "Industrial Process Control Valves, Part 2-1: "Flow Capacity Sizing Equations for Fluid Flow under Installed Conditions," and IEC 60534-2-3, "Industrial Process Control Valves, Part 2-3, Flow Capacity Test Procedure."

A revision of the Instrumentation, Systems, and Automation Society standard is anticipated in the near future, and for this reason only the first five sections of S75.02-1996 were presented here. The reader is advised to refer the revised ANSI/ISA-75.02-1996 standard (when published) for Sections 6 to 10 for the test procedures and calculations. Meanwhile, Section 6.15 can be used for all valve sizing applications, while Section 6.14 provides all the information needed for valve noise calculations.

6.7 Characteristics and Rangeability

B. G. LIPTÁK (1995, 2005) **A. BÁLINT** (2005)

INTRODUCTION

The characteristics, gains, and rangeabilities of control valves are interrelated. The process control engineer should clearly understand these terms, because they describe the personality of the valve and as such play an important role in the closed-loop performance of the loop.

This section begins with a brief discussion of valve gains, followed by an explanation of the difference between theoretical and actual (installed) valve characteristics. The section ends with an explanation of valve rangeability.

VALVE GAIN AND LOOP GAIN

The gain of any device is its output divided by its input. For a linear (constant gain) valve, the valve gain (G_v) is the maximum flow divided by the valve stroke in percentage (Fmax/100%).

When a loop is tuned to provide quarter amplitude damping (Figure 6.7a), the controller gain (G_c = 100/%PB) is adjusted until the overall loop gain (the product of the gains of all the loop components) reaches 0.5. If a linear controller and a linear transmitter are used, their gains are constant. Therefore, if the process gain (G_p) is also constant, a linear (G_v = constant) valve is needed to keep the loop gain product constant at 0.5.

If the transmitter is nonlinear, such as a d/p cell without square root extraction, the transmitter gain will rise in proportion to flow, and therefore the loop will be unstable at high flows and sluggish at low flows. The usual solution is to install a square root extractor, which makes the transmitter linear (G_t = constant) with flow. One can also correct for the nonlinearity of the transmitter by using a nonlinear controller or a valve whose gain drops with flow (quick opening).

Nonlinear Processes

If the process is nonlinear (G_p varies with load) while the other loop gains are constant, a change in load will result in the loop gain's shifting away from 0.5. Therefore, if the total gain rises, the loop will become unstable; if it drops, the loop will become sluggish. Therefore, if the process gain changes with load (a nonlinear process), the loop can remain stable only if another gain component of the loop is also changing, and if that change is the inverse of the gain change of the process.

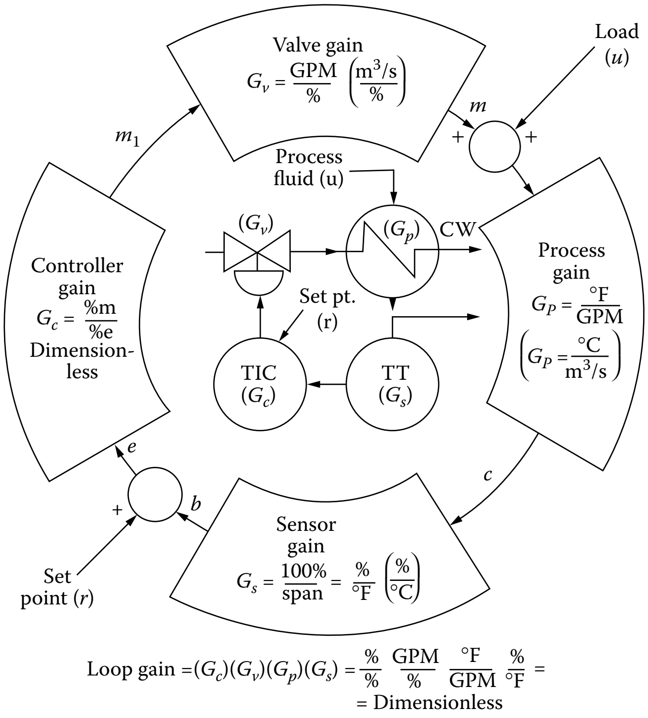


FIG. 6.7a
The loop gain is the product of the gains of the loop component. In a properly tuned loop (decay ratio of $^1/_4$), this gain product should be constant at 0.5.

Therefore, as the process gain (G_p) drops, this other gain must be rising, thereby keeping the loop gain constant at around 0.5. This other gain can either be the gain of the controller (G_c) or that of the control valve (G_v).

If the controller gain varies with load, the controller is called a nonlinear controller. When the control valve gain varies with load (flow), it is named according to the type of nonlinearity that exists between the flow through the valve and the valve stroke.

If, as the flow is increasing, the valve gain is also rising and that rate of rise is a constant rate with flow, the valve characteristic is called equal-percentage. If the valve gain is increasing at a variable rate with flow, it is named in accordance with the type of nonlinearity it provides (parabolic, hyperbolic, and so on). If the valve gain is dropping when the flow increases, it is called a quick-opening valve.

Some processes, such as the pH process, are highly nonlinear. In such applications as pH control, the variation in the process gain (G_p) is compensated by an inverse variation in the controller gain (G_c), which is selected to drop when G_p rises. Therefore, the controller of a pH process is a nonlinear controller.

In heat transfer processes, because the heat transfer area is constant, its efficiency of heat transfer drops as the load (the amount of heat to be transferred across that fixed area) rises. Therefore the process gain (G_p) drops with the heat load. To compensate for this drop, the valve gain (G_v) must rise with the load. Such a valve is called an equal-percentage valve, which should be the valve characteristic selected for all temperature control valve applications.

Installed Valve Gain

As will be discussed in more detail later, the inherent valve gain changes after the installation of the valve, if the valve pressure differential varies with load. This is the case in all mostly friction pumping systems, because as the load (flow) rises, the pressure drop in the piping system also increases, which leaves less pressure drop for the valve.

As the valve differential pressure drops with increasing flow rate, the valve gain (G_v) also drops. This tends to shift the installed gain of equal-percentage valves towards linear and the installed gain of linear valves towards quick opening. Therefore, on mostly friction pumping systems (Figure 6.1c), if it is desired to keep the valve gain relatively constant (linear characteristics), it is recommended to install an equal-percentage control valve.

An even more effective method of keeping the valve gain (G_v) constant is to replace the valve with a linear control loop (Figure 6.1h). The disadvantage of this cascade configuration (in addition to the higher cost) is that this will degrade the control quality if the controlled process is fast, because in each cascade system, the gain of the outer loop must be smaller than that of the inner loop. This will necessitate the detuning (lowering the gain, increasing the proportional band) of the master controller (outer loop), which generates the set point for the flow controller (FC) in Figure 6.1h.

THEORETICAL VALVE CHARACTERISTICS

Valve Testing

The standard methods of testing the capacity of valves are discussed in Section 6.6. It should be noted that the goal of this test is only to determine the valve C_v (K_v) within an error of 5%. What is important to note is that the valve characteristics (G_v characteristics) are neither tested nor defined by the standard.

It should also be noted that during the testing, the pressure drop through the valve itself is not measured, because the ΔP is detected across a pipe section, which includes the valve, plus a length of eight pipe diameters of straight pipe (Section 6.6).

The end result of such a test is a valve characteristic curve, which describes the flow through the valve as it is stroked from 0 to 100% of its stroke. The C_v (K_v) data provided by manufacturers is usually reliable within an error of about 10%, if the installation is identical to the test setup (usually it is not).

Valve Characteristics

The inherent characteristics of a control valve describes the relationship between the controller output signal received by the valve actuator and the flow through that valve, assuming that:

1. The actuator is linear (valve travel is proportional with controller output).
2. The pressure difference across the valve is constant.
3. The process fluid is not flashing, cavitating, or approaching sonic velocity (choked flow).

Some of the widely used inherent lift to flow rate relationships are illustrated in Figure 6.7b. For example, in a linear valve, travel is linearly proportional to capacity and therefore the theoretical gain is constant at all loads. (The actual gain is shown in Figure 6.7e.)

In equal-percentage valves, a unit change in lift will result in a change in flow rate, which is a fixed percentage of the flow rate at that lift. For example, in Figure 6.7b, each percentage increase in lift will increase the previous flow rate by about 3%. Therefore, the theoretical gain of equal-percentage valves is directly proportional to flow rate (actual gain shown in Figure 6.7e) and increases as the flow rate increases. On a logarithmic chart (left side of Figure 6.7b), the equal-percentage characteristics correspond to a straight line having a slope that corresponds to its fixed percentage.

In quick-opening valves, the gain decreases with increasing flow rates. Figure 6.7b shows the quick-opening valve characteristics with the same total lift as for the other plug types. If the travel of the quick-opening plug is restricted so that the distance of 100% lift travel corresponds to only $^1/_4$ of the seat diameter, then the valve characteristics will approach linear (if the hydraulic resistance is constant) with the gain being nearly constant.

Valve and Process Characteristics

Control loops are usually tuned at normal load levels (at normal flow rates through the control valve), and it is assumed that the total loop gain will not vary with process load. This assumption is seldom completely valid, because the process gain usually does change with load. Because one

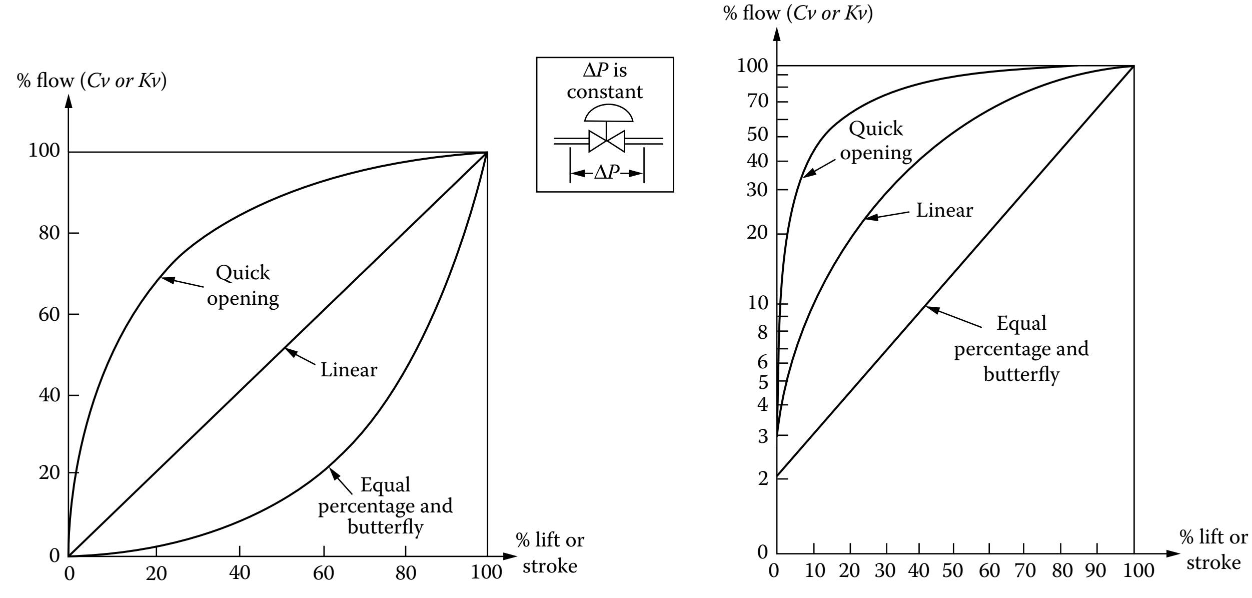


FIG. 6.7b
Inherent flow characteristics of quick-opening, linear, and equal-percentage control valves.

cannot afford to retune the controller for each new load, it is desirable to select control valves that will compensate for these load change effects.

For example, when a liquid-to-liquid heat exchanger is being controlled, the process gain and dead time (transportation lag) will both decrease as the load increases. Therefore, one should attempt to compensate for this inverse load-to-gain relationship by using a valve with a direct load-to-gain relationship, such as the equal-percentage valves. If one does that, as the heat-exchanger load increases and therefore the process gain drops, the valve gain will rise, thereby compensating for that effect and reducing the total change in the loop gain.

Equal-percentage valves are not ideal, though, if high turndown is required or if there are solids in the throttled process fluid.

An opposite example is a control loop whose sensor has an expanding scale, such as an orifice plate or a vapor-filled thermometer. With such detector, the loop gain is increasing with load, and therefore the gain of the selected valve should decrease with load. Therefore, a quick-opening control valve is often used.

In a fairly large number of cases, the choice of valve characteristics is of no serious consequence. Just about any characteristic is acceptable for the following applications:

1. Processes with short time constant, such as flow control, most pressure control loops, and temperature controls through mixing a cold and a hot stream
2. Control loops operated by narrow proportional band (high gain) controllers, such as most regulators
3. Processes with load variations of less than 2:1

In general it can be said that the quick-opening characteristics are used in regulators and on orifice-type measurements, if no square root extractors are provided. The equal-percentage characteristics are most often used on heat-transfer type temperature control applications and on pumping systems where the valve differential pressure varies more than 2:1 as the load (flow rate) changes. Linear characteristics are used in most other cases.

Selection Recommendations

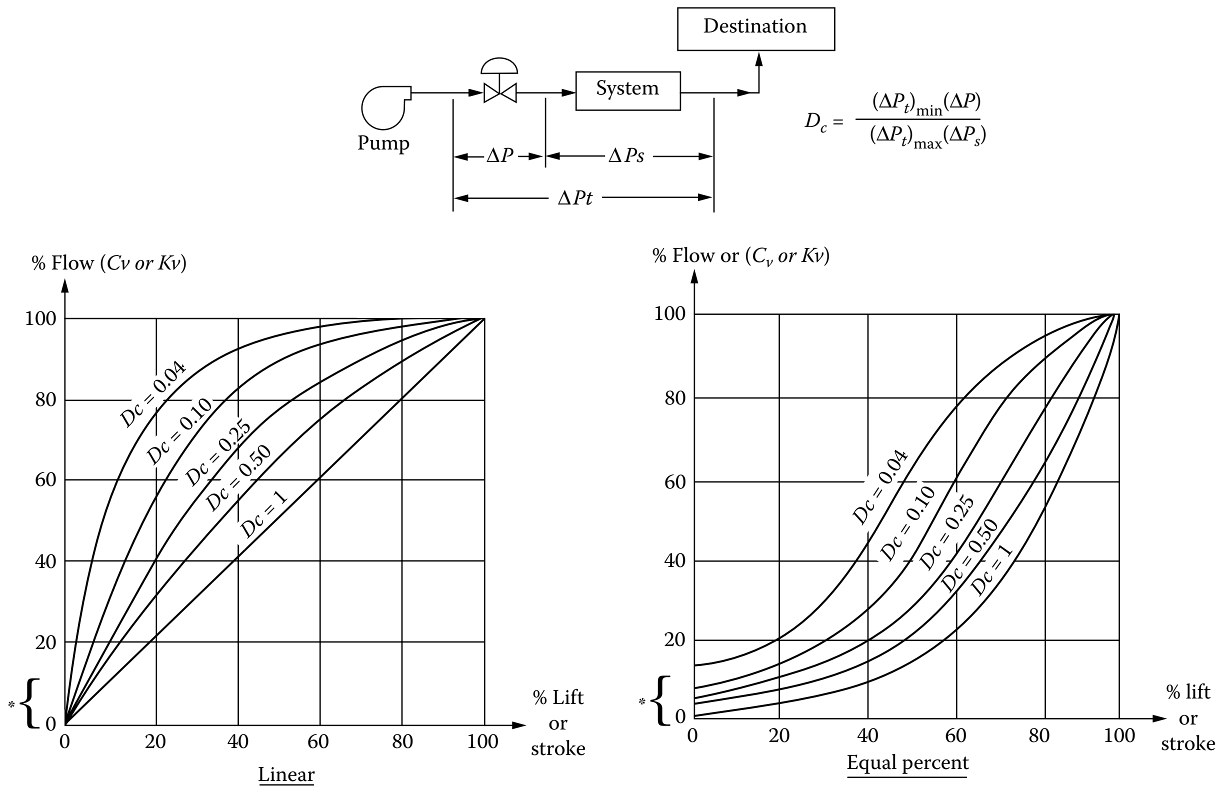
As was discussed in Section 6.1, different engineers have developed different rules of thumb in selecting valve characteristics for the various types of control loops. These recommendations vary in complexity.

Shinskey, for example, recommends the use of equal-percentage valves for heat-transfer control and the use of linear valves for all flow, level, and pressure control applications, except vapor pressure, for which he recommends equal-percentage valves.

Driskell suggests that for relatively constant valve differential pressures, quick-opening valves should be used for square root-type flow loops, equal-percentage valves for temperature and liquid pressure, and linear valves for all others. If the valve differential pressure varies with load, his quick-opening recommendation shifts to linear, and his linear recommendation shifts to equal-percentage.

Lytle's recommendations were summarized in Table 6.1g. They are even more involved, as they take into account more variables.

As will be shown in Sections 6.16 to 6.23, the inherent characteristics of ball and butterfly valves are similar to the



*Note that the minimum "controllable" flow increases as D_c drops.

FIG. 6.7c
These figures illustrates the effects of the distortion coefficient (D_c) on inherently linear (left) and on inherently equal-percentage valves (right), according to Boger.

equal-percentage, while the characteristics of Saunders and pinch valves are closer to the quick-opening.

Installation Causes Distortion

As was pointed out in connection with Figure 6.1c, in mostly friction-type pumping systems, the pressure drop available for the control valve is dropping, as the load (flow rate) is increasing (Figure 6.1e).

This is a different condition from the test conditions under which the valve C_v (K_v) was measured by the manufacturer, because on the manufacturers' test stand, the flow rate through a valve is measured under constant pressure drop conditions. Therefore, when such a valve is installed, its gain (characteristics) will shift as shown in Figure 6.7c. If that shift is substantial, one might obtain a "near linear" installed characteristic by installing a "theoretically" equal-percentage valve.

Different engineers have approached the shift between inherent and installed valve characteristics in different ways. One approach, that of the "old school," was to oversize the pump, so that the *ratio* between the "minimum" and "maximum" valve pressure drops (Figure 6.1c) will not be large, and therefore the gain of the process will not change much with load. This approach works, but it wastes pumping energy.

Distortion Coefficient

When the control valve is installed into the piping in a process plant, its flow characteristics are no longer independent of the rest of the system. This is because the flow through the valve will be subject to the frictional resistance, which is in series with the valve. The consequence is the type of distortion illustrated in Figure 6.7c.

From the curves in Figure 6.7c, one can conclude if a particular installation will have a very substantial effect on both flow characteristics and rangeability of the valve, or not. Under conditions of excessive distortion, clearance flow alone can increase as much as tenfold, and equal-percentage characteristics can be distorted toward linear or even quick opening.

It should be emphasized that Figure 6.7c assumes the use of a constant speed pump (Figure 6.1c). In variable-speed pumping systems, one might adjust the pump speed so as to keep the valve Δp constant, and therefore in such control systems the installed and theoretical valve characteristics are more similar, and less distortion is allowed to occur. Naturally, in variable-speed pumping systems one can completely eliminate the valve and just throttle the pump speed.

The predictability of installed valve behavior is reduced, not only because the inherent valve characteristics deviate

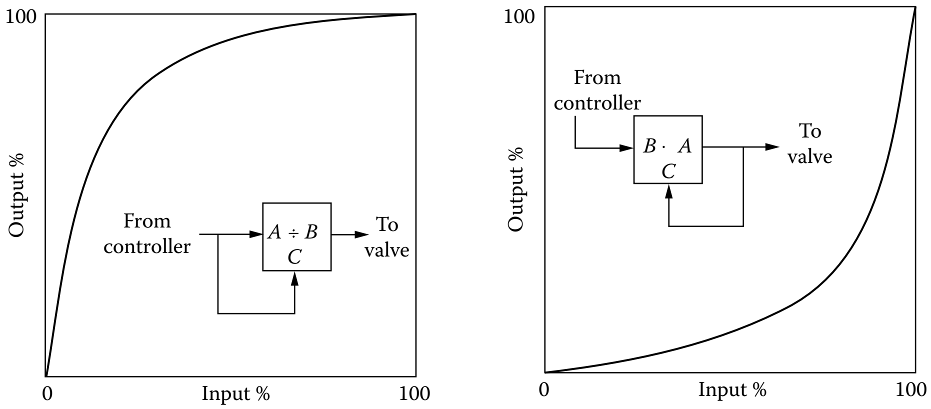


FIG. 6.7d
The valve characteristics can be modified by inserting a divider or multiplier relay into the controller output signal.

from their theoretically prescribed character, but also because:

1. Actuators without positioners will introduce nonlinearity.
2. Pump curves will also introduce nonlinearity.

It should also be recognized that in order to learn the true valve characteristics requirement, a full dynamic analysis of the process is required. Even if one took the trouble to perform such analysis, it would probably yield a valve characteristics requirement that is not commercially available in conventional air-operated control valves. For these reasons, one might consider any one of the following options:

1. Characteristics that are an intrinsic property of the valve construction, such as an equal-percentage ball or butterfly or a beveled (quick-opening) disc
2. Valves that are characterized by design, such as globe valves having linear or equal-percentage trims
3. Digital control valves that can be characterized by software
4. Characteristics that are superimposed through auxiliary hardware, such as function generators, characterized positioners, cams, and so on
5. Intelligent control valves (Section 6.12), which can electronically modify the control signal, which is received as a function of the inherent valve characteristics and of the desired valve gain

Correcting the Valve Characteristic

The linear valve has a constant gain at all flow rates, while the gain of the equal-percentage valve is directly proportional to flow. If the control loop tends to oscillate at low flow but is sluggish at high flow, one should switch the valve trim characteristics from linear to an equal-percentage.

If, on the other hand, oscillation is encountered at high flows and sluggishness at low flows, the equal-percentage valve trim should be replaced with a linear one.

Changing the valve characteristics can be done more easily by modifying the controller output signal or by inserting accessories into the control signal leading to the actuator than by replacing the valve.

One approach proposed by Fehérvári/Shinskey is to insert a divider or a multiplier into the control signal line, as illustrated in Figure 6.7d. By adjusting the zero and span at port C, a complete family of curves can be obtained. The divider is used to convert an air-to-open equal-percentage valve into a linear, or an air-to-close linear valve into an equal-percentage one.

The multiplier is used to convert an air-to-open linear valve into an equal-percentage, or an air-to-close equal-percentage valve into a linear.

According to Shinskey, both devices are perfectly standard, sensitive, stable, easy to calibrate, and “real lifesavers when one needs a linear butterfly valve.”

RANGEABILITY

The conventional definition of rangeability used by most manufacturers is the ratio between maximum and minimum “controllable” flow through the valve. Minimum controllable flow (Fmin) is defined as the flow below which the valve tends to close completely. In other words, this widely held definition of Fmin refers not to the leakage flow (which occurs when the valve is closed), but to the minimum flow that is controllable in the sense that it can be changed up or down as the valve stroke is changed.

Using this definition, manufacturers usually claim a 50:1 rangeability for equal-percentage valves, 33:1 for linear valves, and about 20:1 for quick-opening valves. These claims suggest that the flow through these valves can be controlled down to 2, 3, and 5% of the flow corresponding to their rated C_v (K_v).

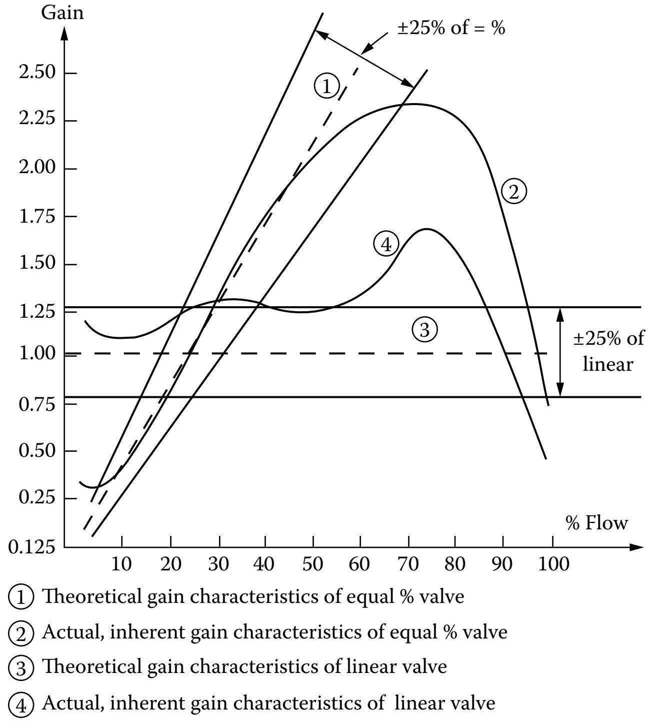


FIG. 6.7e

The theoretical vs. the actual characteristics of a 2 in. (50 mm) cage-guided globe valve, according to Driskell.

The above definition of rangeability is based on the inherent C_v (K_v) determined during testing. It can be seen in Figure 6.7c that the minimum controllable flow rises as the distortion coefficient (D_c) drops. At a D_c value of 0.1, for example, the 50:1 rangeability of an equal-percentage valve drops to close to 10:1.

This is because the valve pressure drop is much higher at low flows (Figure 6.1e), and therefore the minimum valve opening will pass much more flow. Thus, the required rangeability should be calculated as the ratio of the C_v (K_v) required at maximum flow (and minimum pressure drop) and the C_v (K_v) required at minimum flow (and maximum pressure drop).

Improved Definition of Rangeability

The decision on whether a particular control valve is capable of providing the required rangeability should not be evaluated on the basis described above. This is because a loop is uncontrollable not only when the valve cycles between closed and some minimum flow but also when the loop gain product shifts away from 0.5 (the tuning target for quarter amplitude damping).

Therefore, the acceptable flow range within which the valve can safely be used for closed-loop control must be based on a relationship between the theoretical and the actual valve gain. The valve rangeability should therefore be defined as the flow range over which the theoretical (inherent) valve gain and the actual installed valve gain will stay within preset limits.

Therefore, the rangeability of the valve can be defined as the ratio of the minimum and maximum C_vs (K_vs) bordering the region within which the actual valve gain is within ±25% of the theoretical valve gain.

This advanced definition of valve rangeability establishes the point at which the flow-lift characteristic starts to deviate from the expected by more than 25%. (Figure 6.7e shows the points where the actual gain's deviation from the theoretical starts to exceed 25%.)

If one defines "intrinsic rangeability" as that range of the ratios of C_v(max) to C_v(min) within which the values of the valve gain do not vary more than ±25% from the theoretical, then according to Figure 6.7e, the rangeability of a linear valve can be greater than that of an equal-percentage valve. Actually, if one uses this definition, the rangeability of equal-percentage valves is seldom more than 10:1.

The rangeability of some rotary valves can be higher because their clearance flow tends to be less than that of other valves, and their body losses near the wide open position tend to be lower than those of other valve designs.

Valve rangeability can also be limited by the sensitivity and accuracy of positioning.

Why Traditional Rangeability Is Wrong

The problem with traditional rangeability definitions is that they tend to overstate the range within which the control valve can be used. This results in poor control quality for the whole loop. The main reasons why the traditional definition for control valve rangeability is unacceptable are as follows:

1. The minimum "controllable" flow (F_{min}) is determined using a test during which the valve pressure differential is constant, while in most real-life installations the pressure differential is maximum when the flow is minimum. Therefore, the real value of F_{min} should be higher than the claimed values.
2. The valve should not only be "controllable" within its rangeability but should have a gain (G_v) that is close to its theoretical gain.
3. The traditional definition of rangeability is based on dividing the maximum flow (F_{max}) by F_{min}. This approach is also wrong, because while the flow through a nearly 100% open valve is "controllable," its gain is nowhere close to the theoretical value.

As was shown in Figure 6.7e, if the acceptable valve gain is defined as 1, which is within ±25% of the theoretical valve gain, F_{max} of the linear valve should not exceed 60% and F_{max} of the equal-percentage valve should not exceed 70% of the maximum flow through the valve. In terms of valve lift, these flow limits correspond to 85% lift for equal-percentage and 70% lift for linear valves.

CONCLUSIONS

It is time to realize that smart transmitters and sophisticated control algorithms alone cannot result in properly functioning control loops. Stable and responsive closed-loop control also requires that the gain of the final control element (the valve) be much more predictable and better controlled.

In many valve designs (digital valves excluded) this can only be achieved if the valve rangeabilities are redefined and thereby restricted (reduced). The manufacturers should contribute to the achievement of this improvement by publishing the characteristic curve for each valve showing its $F_{\min}$ and gain values at different distortion coefficients (D_c).

The user's contribution should be a better understanding of the role that the valve gain plays in process control and the realization that in most control valve designs the upper one third of the stroke is not usable for stable control.

Bibliography

Alcozer, S., Segovia, J. P., and Sbarbaro, D., "Using Higher Order CMAC to Improve the Performance of Control Valves," *ISA Transactions,* Volume 36, Issue 1, pp. 65–70, 1997.

Arant, J. B., "Positioner Use Is Myth-Directed," *InTech,* November 1992.

Bajsic, I. and Bobic, M., "Gain of control valve with polygonal flow characteristics," *STROJ VESTN-J MACH E,* 46(5), 264–275, 2000.

Ball, K. E., "Final Elements: Final Frontier," *InTech,* November 1986.

Baumann, H. D., "How to Assign Pressure Drop Across Control Valves for Liquid Pumping Services," *Proceedings of the 29th Symposium on Instrumentation for the Process Industries,* Texas A&M University, 1974.

Bialkowski, B., Coughran, M., and Beall, J., "Control Valve Performance Update—The Last 10 Years," *PULP PAP-CANADA,* 102 (): 21–22, November 2001.

Boger, H. W., "Flow Characteristics for Control Valve Installations," *ISA Journal,* November 1966.

Borden, G. and Zinck, L., "Control Valve Seat Leakage," *InTech,* November 1986.

Byeon, H. W. and Ulrey, J. L., " Designing a Standard Thermal Power Plant for Daily Startup/Shutdown: The HP Bypass Control and Safety Function," *ISA Transactions,* Volume 36, Issue 1, pp. 71–77, 1997.

"Calibrating Control Valve Accuracy," *MECH ENG,* 122 (6): 33–33, June 2000.

Campos, M. C. M., Satuf, E., and de Mesquita, M., "Intelligent System for Start-up of a Petroleum Offshore Platform," *ISA Transactions,* Vol. 40, Issue 3, July 2001.

Champagne, R. P. and Boyle, S. J., "Optimizing Valve Actuator Parameters to Enhance Control Valve Performance," *ISA Transactions,* Vol. 35, Issue 3, 1996, Pages 217–223.

"Control Valves, Regulators," *Measurements and Control*, June 1992.

Coughlin, J. L., "Control Valves and Pumps: Partners in Control," *Instruments and Control Systems,* January 1983.

Coughran, M. T., "Measuring the Installed Dead Band of Control Valves," *ISA Transactions,* Volume 37, Issue 3, pp. 147–154, July 1998.

Davis, J. A. and Stewart, M., " Predicting Globe Control Valve Performance — Part II: Experimental Verification," *J FLUID ENG-T ASME,* 124 (3): 778–783, September 2002.

"Exclusive Control Valve Performance Diagnostics," *HYDROCARB PROCESS,* 82 (12): 30–30, December 2003.

Dobrowolski, M., "Guide to Selecting Rotary Control Valves," *InTech,* December 1981.

Driskell, L. R., "Sizing Control Valves," *ISA Handbook of Control Valves,* 1985.

Gassmann, G. W., "When to Use a Control Valve Positioner," *Control,* September 1989.

George, J. A., "Sizing and Selection of Low Flow Control Valves," *InTech,* November 1989.

Harrold, D., "Calibrating Control Valves," *Control Engineering Europe,* Vol. 3, Issue 3, p. 43, June/July 2002.

Hägglund, T., "A Control-Loop Performance Monitor," *Control Engineering Practice,* Vol. 3, Issue 11, pp. 1543–1551, October 1995.

Kaya, A., "Training in Process Control by Distributed Computers: A 'Hands-On' Approach," *ISA Transactions,* Vol. 29, Issue 3, pp. 35–41, 1990.

Lipták, B. G., "Control Valves in Optimized Systems," *Chemical Engineering,* September 5, 1983.

Ogawa, K. and Kimura, T., "Hydrodynamic Characteristics of a Butterfly Valve—Prediction of Torque Characteristics," *ISA Transactions,* Vol. 34, Issue 4, December 1995.

Price, V. E., "Smart Valve Intelligence Takes Many Forms," *InTech,* November 1986.

Simula, M., "Improve Control with Smart Valve Technology; Benefits Include Reduced Maintenance and Better Stability," *Hydrocarbon Processing,* Vol. 78, Issue 8, p. 63, August 1999.

Thanomsat, C., Taft, C. W., and Annaswamy, A. M., "Level Control in Feedwater Heater Systems Using Nonlinear Strategies," *ISA Transactions,* Vol. 37, Issue 4, pp. 299–312, September 1998.

Tolliver, T. L., "Continuous Control Experiences," *ISA Transactions,* Vol. 30, Issue 2, pp. 63–68, 1991.

Tullis, J. P., *Hydraulics of Pipelines: Pumps, Valves, Cavitation, Transients,* New York: John Wiley and Sons, 1989.

Valenti, M., "Improving Hydraulic Performance with Intelligent Valve," *Mechanical Engineering-CIME,* Vol. 118, Issue 4, p. 56, April 1996.

6.8 Diagnostics and Predictive Valve Maintenance

K. BEATTY (2005)

Types of Predictive Maintenance:	a) Performance tests b) Valve characterization and signatures c) Valve response tests
Types of Diagnostic Systems:	a) Field testing using specially mounted sensors b) Smart positioners with integral sensors c) Host system tools that measure response d) Partial stroke test for safety-related valves
Special Features:	a) Actuator pressures, temperature, and position measurements b) Fully integrated with smart positioners, or field installed for individual tests c) Specially prepared valves for regular testing
Cost:	Premium for \$300–\$1000 for smart positioner, \$50–200 per test for field-mounted test systems
Partial List of Valve Diagnostics Suppliers:	ABB Inc. (www.abb.com/us/instrumentation) Dresser Flow Control, Masoneilan Operations (www.masoneilan.com) Fisher Controls International (www.fisher.com) Flowserve, Flow Control Div. (www.flowserve.com) Honeywell Industry Solutions (www.iac.honeywell.com) Invensys Flow Control (www.invensysflowcontrol.com) Metso Automation (www.jamesbury.com) Samson Controls Inc. (www.samson-usa.com) Siemens Energy and Automation (www.sea.siemens.com) Smar International Corp. (www.smar.com) Yamatke Corp (www.ycv.com) Yokogawa Corporation of America (www.yca.com)

INTRODUCTION

There is currently a major change in the approach taken to equipment maintenance due to longer times between turnarounds and a reduction in the available personnel. Valves are typically one of the higher maintenance pieces of equipment in a process. This is because, unlike measurement instrumentation that only monitors the process, valves are dynamic devices, and in order for them to function properly, their main components, such as their actuators, packings, positioners, and I/P converters, all need to be operational. If any of these components fail, the result is that the process can no longer be remotely controlled. For this reason, control valve testing should evaluate the condition of all of these components.

Valve diagnostics and predictive maintenance is relatively new. In the past, the approach to maintenance was to schedule the periodic rebuilding of critical valves based on the time that the valve has been in service and not on the basis of actual need. This type of preventative maintenance method has several drawbacks. One is that by using this approach, a lot of unnecessary maintenance is done. Secondly, if critical failures occur between maintenance intervals, they are overlooked. Finally, sometimes the act of maintenance itself causes problems in valves that had no problem before.

DIAGNOSTICS

Instrumentation Used

Diagnostic information can be obtained from smart positioners, local portable sensors, and the evaluation of process control signals. As discussed in more detail in Section 6.12, smart positioners are relatively inexpensive and easy to use. For these reasons, they are used most commonly.

The local sensors utilized in diagnosing the control valves are usually integrally mounted on the valve. In safety-related control valve systems (as discussed further in Section 6.10), specialized positioners are used to provide partial stroke tests, which can verify the operational status of the valve without interrupting the process.

When diagnosing existing installations, the testing is often performed by analyzing the past performance and records of the process instrumentation. When time permits, portable sensors can be installed in the field to monitor the performance of the valves and to provide the required data for diagnostic tests.

Intermediate signal control boxes can also be used, primarily in safety-related systems. These can partially stroke on/off safety and shut-off valves (Section 6.10) by briefly interrupting the signal to a solenoid. This way, the system can verify that the valve is operational, because it is beginning to operate. In order not to adversely affect the operating process, once the valve begins to respond, the solenoid is quickly re-energized. This way, the valve is tested without a need to shut down the process while testing.

Diagnostic Methods

Standards such as ISA-75.13-1996, “Method of Evaluating the Performance of Positioners with Analog Input Signals and Pneumatic Output,” or “ANSI/ISA-75.25.01-2000,” Test Procedure for Control Valve Response Measurement from Step Inputs,” are useful in evaluating the performance capabilities of control valves. These standards describe the performance tests required to determine HRL hysteresis, repeatability, and linearity (HRL) of control valves.

These tests provide very useful information on the performance of both the positioner and the valve actuator, but they do not provide much information that would be useful in predictive maintenance.

Many smart positioners have built-in diagnostics that can provide usage information and current status that is very useful in troubleshooting existing problems. Most of these positioners do not provide on-board predictive diagnostic information. In order to perform predictive diagnostic tests, the positioner usually has to be accessed through a digital interface and its data downloaded into a PC for analysis.

Many new host control systems can pass the information directly to a diagnostic PC over a network connection. Some positioners also have internal data historians that provide detailed information on the valve command, actual position, and actuator pressures that provide a level of detail not normally available in host historians for diagnosing transient problems.

Characteristics Tests

Another category of tests are the characteristic or signature tests, which are well suited for predictive diagnostics. They can be generated by a smart positioner or obtained by using a portable test rig.

Valve signatures are obtained by ramping the valve through a range of travel and capturing information concerning the position command, the valve travel, and the actuator pressure(s). This information can be analyzed independently, or over time to observe changes in the characteristics of the control valve.

For example, a signature can be the friction in a valve over a range of travel. Changes or abnormalities in the friction plot over the range of travel provide direct information related to seating problems or galling that may be occurring in the valve trim. A full travel signature can provide information on the travel calibration of pneumatic actuators or help to locate leaking or worn seals and fittings.

New software tools are also available for the historical analysis of valve command signals and of the corresponding process response. This method of analysis is nonintrusive and can be applied with both smart and traditional positioners.

Many of the analysis packages can access the required data directly from a historian in the DCS. This method of analysis involves the comparing of the controlled process measurement with the command signal received by the control valve. If the controlled process measurement is sensitive enough to respond to small changes in the valve opening, then this measurement can be correlated with the control signal received by the valve and, thereby, determine if there is sticking or lagging and quantify the magnitude of the problem. The main limitation with this type of analysis is that it provides an indication of the existence of a problem, but does not provide detailed information.

Some equipment suppliers now offer diagnostic services on a contractual basis. In providing these services, the suppliers utilize large databases that they have compiled over time, using various types of valve designs in a wide variety of process services. These services often include a review of the existing installation to see if other valve designs might not be better suited for the applications at hand.

Valve Signatures

A valve signature is obtained by running a reproducible test on a control valve. During such a test, a repeatable signal is sent to the control valve, and the response to that signal is captured and then analyzed to see if the control valve operates correctly.

The valve should be removed from the process and placed on a test stand while running the complete signature test. This is because during the test, the valve is forced to travel over a wide range, and if this was not done off-line, it could cause harm to equipment, personnel, or the process. For the signature of an emergency shutdown valve, refer to Figure 6.8a.

However, a partial stroke signature can sometimes be obtained on-line. If a valve signature is obtained near the end of the valve stem travel, or even beyond the end of travel, it still can be of some value. For example, if the behavior of a normally closed valve is tested with a control signal range of –5 to 0%, one can obtain diagnostic data on the response

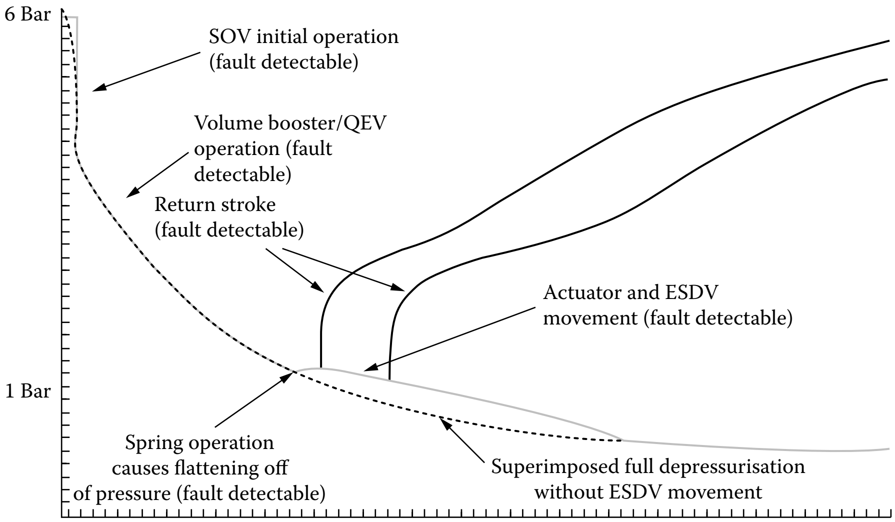


FIG. 6.8a
The signature of an emergency shutdown valve (ESDV). (Courtesy of Drallim Industries Ltd.)

of the pilot valve in the positioner without opening the valve (moving the valve stem) at all.

Similarly, limited signatures can be obtained, when control valves are in operation, by changing their control signals within a small range (1–2% of travel) around the required opening of the value.

Valve Signature Types There are four types of valve signatures. These are obtained by detecting the valve's response to steps, ramps, partial strokes, or internal operation. The step and ramp signatures can be performed in either the opening or closing directions. Figure 6.8b shows the *ramp test signature* of a control valve with a double-acting actuator, and Figure 6.8c shows the same signature for a valve provided with a spring-loaded, single-acting actuator.

When performing a ramp test, first the starting and ending valve positions are entered. The ramp rate and the sampling time is usually also entered. The data gathered at each sample point includes the value of the control command signal, the valve position, and actuator pressure.

In order to obtain a *step signature*, first the step size is defined by setting the starting and ending valve positions, and then the recording time period is set, where time zero is the time at which the step command is issued. The data gathered at each sample point include elapsed time and the values of the control command, valve position, and actuator

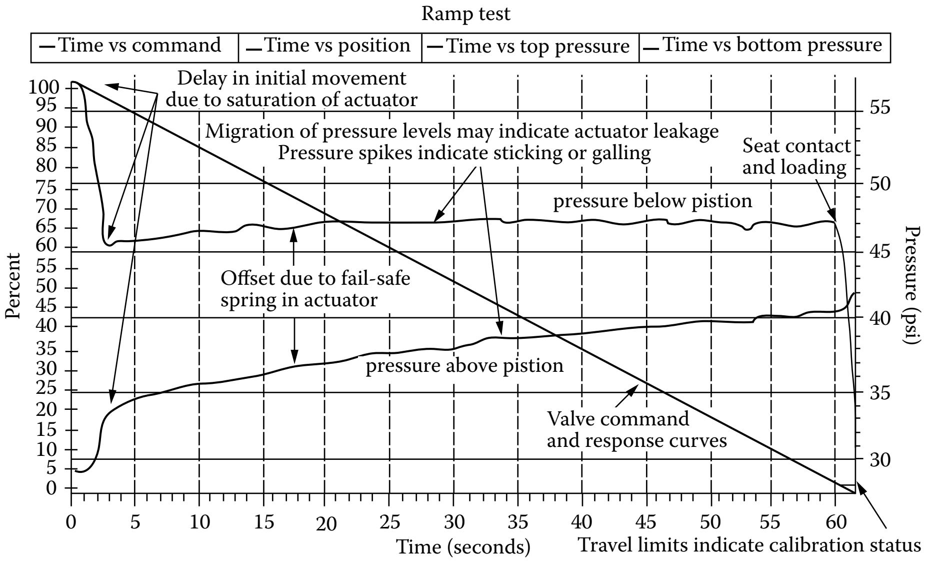


FIG. 6.8b
The ramp test signature of a control valve with a double-acting actuator. (Courtesy of Flowserve Corporation.)

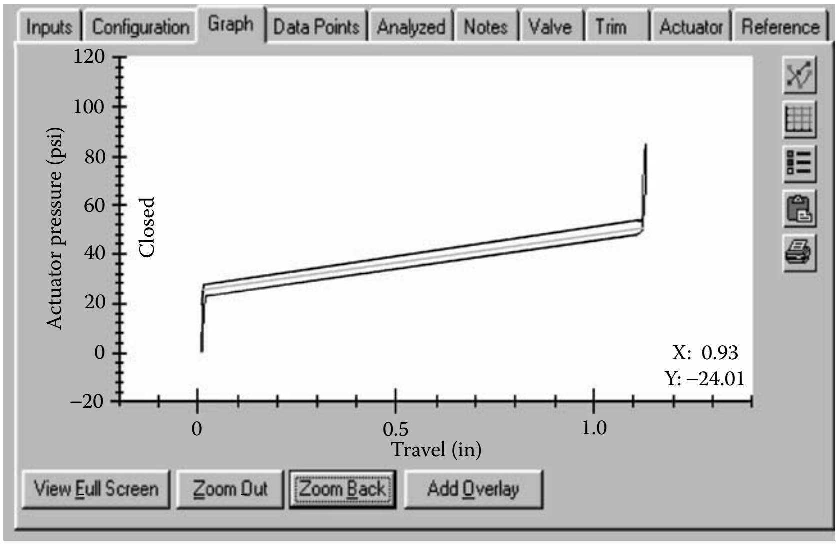


FIG. 6.8c
The ramp test signature of a control valve with a spring-loaded (single-acting) actuator. (Courtesy of Fisher Controls.)

pressure. The result is a graph that shows the response of the valve to a step in its control signal including the approach and stabilization of the valve at the new position.

A *partial stroke test* is similar to a step test, but it is based on a temporary and small change in the control signal. Instead of reaching and maintaining an end position, the test only lasts long enough for the valve to begin to move or sometimes not even that, but only begins to apply the necessary pressure to begin to move the valve stem before resuming normal operation.

Internal operation signatures are primarily positioner diagnostic tools, because they test the operation of the pneumatic relay.

Analyzing Valve Signatures

Most valve manufacturers provide information and guidance for interpreting the above-described signature tests. Interpretation of valve signatures considers two different factors: One is the characteristics of type of valve used, the second is the past history of valve performance.

In terms of valve characteristics, the control signal and valve travel relationship is a function of the type of valve used. Similarly, the minimum and maximum stem friction values depend on the type of packing used by the particular valve design (see Section 6.5 on packing designs). Other characteristics include the balancing pressure range of double-acting actuators and positioners, which, if it is outside the design range, is an indication of problems in the actuator. Also, there usually are inflection points in the valve signature, which indicate the particular seating and linkage characteristics of the valve.

The second basis for evaluating the signature of a valve is by making a historical comparison with past performance. The reference signatures can best be obtained when the valve is in good working order. Once a set of baseline signatures for each valve have been obtained, subsequent signatures can be directly compared against that baseline reference signature, which was taken under the same conditions as the existing ones. The differences between the reference and actual signatures can point to possible problem areas.

The Instrumentation, Systems, and Automation Society (ISA) committee "SP 75.26.01, Valve Diagnostic Data Acquisition and Reporting" is in the process of developing a standard for the methodology to be used in acquiring valve signatures. Their scope includes the defining of the test procedures to be used and the obtaining of reports that can be transferred from one diagnostic system to another.

CONCLUSIONS

The analyzing of valve diagnostic information is a new and evolving field. The diagnostic needs of a plant can depend on its age, on the type of control system being used, and also on the experience and attitude of the operating staff. Today, a variety of diagnostic services are offered by valve manufacturers, instrument suppliers, and service companies. With the passage of time, standardization is expected both in the methods of obtaining and in the interpretation of valve signature tests.

Bibliography

Essam, D., "Final Element Testing, A Way Forward," Drallim Industries, October 26, 2001, http://www.ISA.org/~safety/valve.htm.

Miller, L., "Valve Diagnostics Past, Present, and Future," *Fluid Handling Systems*, November 2002.

Merritt, R., "Some Process Plants are Out of Control," *Plant Services Magazine*, December 2003.

Fedenczuck, P., Fountan, P., and Miller, R., "Profitable Performance," *Hydrocarbon Engineering*, November 2001.

6.9 Dynamic Performance of Control Valves

C. G. LANGFORD (2005)

VALVE RESPONSE

This section first defines the terms associated with control valve mechanical response. It then explains how the speed and accuracy of the control valve response will determine the quality of control of the process. Then it provides suggestions for improving control valve response to achieve acceptable performance.

Improved data collection and analysis with modern control systems make the quality of process control more visible and quantifiable. The business requirements for profit and improved quality control require better control performance.

Modern control valve positioners, often incorporating computer technologies, can improve system performance (see Sections 6.2 and 6.12). See the Bibliography for test data from poorly behaving valves (see Figure 6.9a). The elimination of asbestos-based packings and the increased concerns over packing leakage had the result of improvements in packing design and materials. Developments in low-friction packing materials provide a powerful tool for improved control (see Section 6.5).

Some applications may have control response requirements that preclude the use of a conventional control valve. For these, a pressure or flow regulator may be the better choice. Regulators tend to be relatively fast and reasonably accurate. They lack the control options provided with a separate controller and the design and materials flexibility provided by a separate valve. Some regulators are available with a pressure receiver replacing the spring and screw top works. This provides for remote control of the loading pressure, and thus the set point. Many regulators use techniques that decrease the "droop," which is the normal pressure drop at high flows (see Section 7.7).

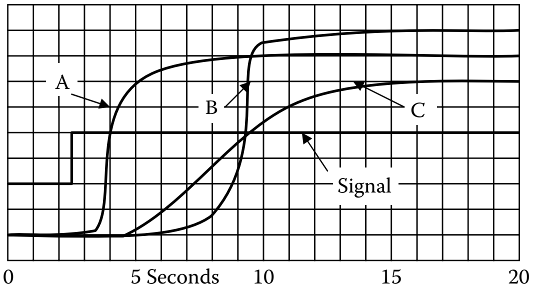


FIG. 6.9a
The responses to a step change in control signal by valves A, B and C.

Definitions

Valve: A valve is a device used for the control of fluid flow. It consists of a body that must satisfy the key requirement of containing the fluid with flow ports, and a moveable closure member, which opens, restricts, or closes the port(s) (see ANSI/ISA-75.05.01-2000, "Control Valve Terminology").

Actuator: An actuator is a device that supplies the force causing the movement of the valve closure member. Most commonly these are fluid or electrically powered (see ANSI/ISA-75.05.01-2000). Actuators most often use compressed air but may use electric, hydraulic, or electrohydraulic power.

Motion conversion mechanism: A mechanism installed between the valve and the power unit of the actuator to convert between linear and rotary motion where required. The conversion may be from linear actuator action to rotary valve operation or from rotary actuator action to linear valve operation.

Accessories: Additional devices used in the operation of the control valve. As described in ANSI/ISA-75.05.01-2000, typical examples include a positioner, transducer, signal booster relay, air set, and snubber (see also Section 6.2).

Dead band is the range through which an input signal may be varied, with reversal of direction, without initiating an observable change in output signal (ANSI/ISA-S51.1-1979 [R1993]). In standard ANSI/ISA-75.25.01-2000 it is defined in percentage of input span. Note that in some other literature this definition is used for dead zone. See Figures 6.9b and 6.9c. The input signal is shown here as a linear ramp for clarity.

Control valve response is the dynamic (considering time) relationship between the change in a control signal and the change in the valve stem position. The process operator is interested in the process response, not the valve stem position. It is important to remember the end purposes of the control valve functionality are business and safety. Process safety is easier if the process is under good control. Much of the discussion in this section uses the conventional globe style valve for the examples. However, the basic principles apply to all final control elements. The response of electric or air motors in a film windup, the performance of swirl vanes

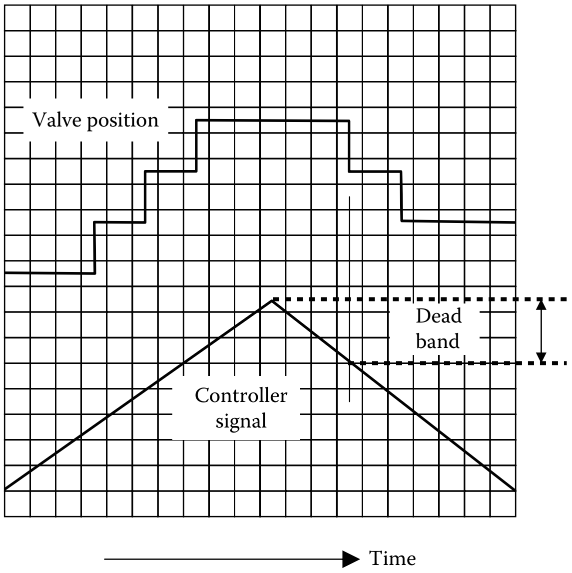


FIG. 6.9b
Dead band is defined as the amount by which the controller output signal has to drop, before the valve position changes. This figure illustrates the response of a valve with high packing friction.

used on the inlet of a blower, and variable speed pumps all have dynamic characteristics similar to valves.

DISCUSSION

The function of a control valve is to change the closure member ("trim") configuration in a way that the flow restriction coefficient, C_v, changes in a controlled and predicted manner. Bernoulli's law states that except for losses and energy taken from or added to the fluid, the sum of the pressure energy, the kinetic energy, and the internal energy within a system will remain constant.

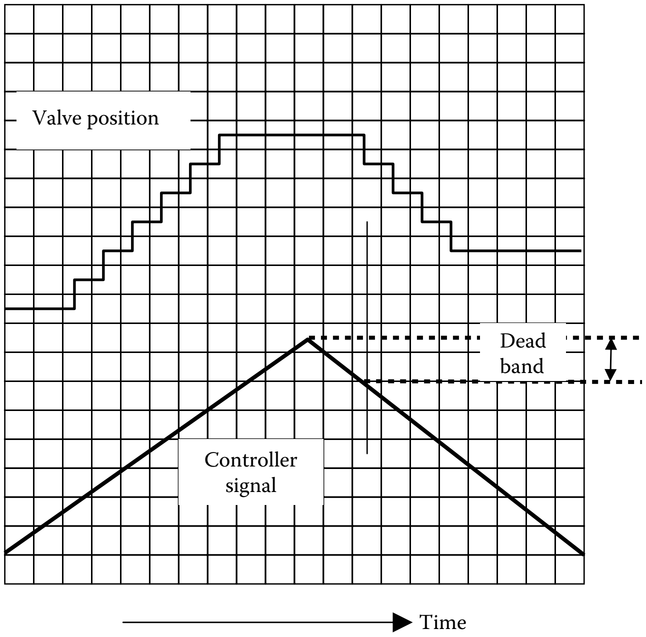


FIG. 6.9c
If the packing friction of the valve is reduced, the dead band required to initiate a change in valve position is also reduced.

The tortuous and varying fluid path in a control valve converts part of the pressure energy into kinetic energy, with a high-velocity jet and increased turbulence. Most of the turbulence dissipates into heat (internal) energy as the flow jet slows down into a more orderly velocity distribution downstream of the restriction. Some of the velocity energy will change back into pressure energy downstream of the flow jet at a lower pressure. The energy converted into heat will leave the valve with the flowing fluid or is lost through the piping wall. Except for inefficiencies, the quantity of energy lost is essentially equivalent to the energy required for a pump to increase fluid pressure by the same amount. The valve closure member connects to the actuator by way of a valve stem or shaft. The valve mechanical response is a function of the valve actuator and its various accessories and with the forces that act to cause or restrain the motion.

The response of the process is a function of the response of the valve and the relationship is normally not a simple one. For difficult or sensitive applications, the valve response requirements should be determined from actual or calculated knowledge of the process response requirements rather than from a simple specification based only on common practices or habit.

Valve System

The installed control valve is a system, consisting of the body and trim, the actuator, and all the accessories. Air pressure supplied by a pneumatic controller or current to pressure (I/P) relay causes flow in to, or out of, the actuator and increases or decreases the pressure on the diaphragm (see Figure 6.9d). Information about valves and accessories is available in this handbook (Sections 6.2 and 6.12) and from catalogs or from various computer programs (see Table 6.2n). The manufacturers' data can aid in developing practicable performance purchase specifications. There is no point in specifying an impossible valve or one that will require very special and expensive attention.

Typically, a spring provides the opposing force in an actuator to move the stem in the opposite direction when the actuator air pressure decreases. The force is proportional to the change in stem position; this is Hook's law. The diaphragm actuator in the typical diaphragm changes effective area with pressure and stroke position. This results in an imperfect conversion of pressure into force.

In the sliding stem valve, stem packing friction creates a force that opposes motion in either direction. This is often a large force, from hundreds to thousands of pounds. The result is to increase the pressure (mass of gas) required in the actuator to cause a change in stem position. Any source of friction will have a similar effect. With ball and plug valves, the friction of the liner and the seat is often much larger than stem packing

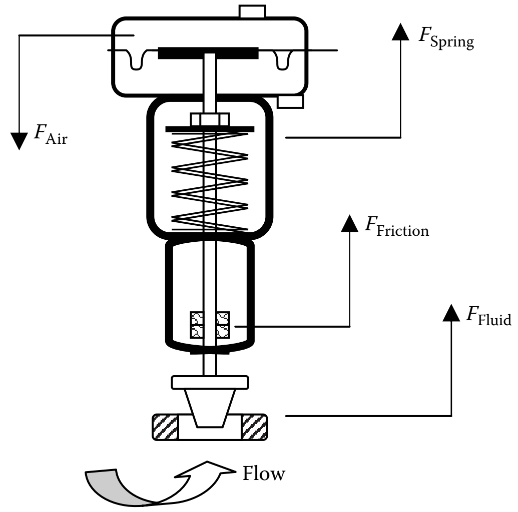


FIG. 6.9d
If both the process fluid and the valve spring are arranged to resist the closing of the valve, the air signal must generate a force, which will overcome these two, plus the force generated by the packing friction.

friction. In this case, appreciable valve shaft "windup" may occur with shaft distortion from the high torque. This windup might not be apparent from outside the valve. Any freedom of motion between the closure member and the shaft will increase the dead band. Valve design details all play a part in this.

Fluid forces on the closure member will vary with fluid conditions and the closure member design and position and may act in either direction. These forces are more significant in the larger valves. For valves smaller than $^1/_2$ in. (12 mm), fluid forces are small. Balanced trim reduces these forces.

Install Positioner

The valve closure member will move when the sum of the forces acting to move the stem exceeds the sum of the forces that act to oppose motion. Because static (nonmoving) friction is higher than dynamic (moving) friction, the valve stem will not move in a smooth manner but will jump to a new position and then stop, because the actuator force decreases when it moves and the pressure drops. This new position is only rarely perfectly, and exactly, the position represented by the control signal. This valve response will result in a flow response and, then after that, in a process response. The process value will result in a change in the control system output, and the cycle starts again. The word "cycle" used here makes the point that cycling is almost inevitable. It is almost certain that the stem will move in steps. This discussion is about the size and significance of those steps.

In the absence of other forces, the spring coefficient defines the stem position for a given actuator pressure. With other forces added in, the stem position will differ. Adding a positioner to the system reduces the effects of the other forces and raises the "stiffness" of the system as the positioner acts to oppose errors due to friction and fluid forces.

The flow of air into or out of the actuator is an integrating (summing) process and will always lag or follow the input signal change. The process under control will also have characteristics of lag and dead time. The control system adds more lag. Despite all the efforts of the finest minds, it is not yet possible to see accurately into the future. Sophisticated model-based control may help but nothing can be perfect. From this, it is clear that no real control system can control perfectly. The goal is to have something that is "good enough" to satisfy the system requirements. Valves, sensors, and the control system performance all together determine the quality of process control. Improvements from a base case will generally cost money, and the added costs must be justified and supported by the predicted desired improved performance. The relationship between these various times and magnitudes defines how well the process is controlled. Simply, a very fast process will require fast control response, and a process with high gain requires high accuracy.

The catalog listings for most positioners and pneumatic relays list the available air flow capacity in SCFM (SCMM) or other mass units at the specified supply pressure (see Table 6.2n), the air capacity of selected pneumatic components. The positioner is a narrow proportional (high gain) pneumatic relay. This sensor and amplifier with the actuator comprises a servomechanism, with the valve stem as the controlled device. The positioner compares the difference between an input signal defined in terms of the stroke of the valve and the actual stem position as measured with a mechanical linkage, and it controls air to the actuator. The early simple designs had the disadvantage of cycling if the gain was too high for the response time of the valve assembly. That is, because the system is an integrating controller, the inherent phase lag in the stem response requires that the gain be limited or a low-frequency cycle will result. If the valve is the slowest element in the control loop, or behaves poorly, then it is the limiting control device. If the positioner slows the valve response, then it becomes the limiting device. If the volume of the actuator is large compared to the air flow rate, then the air flow rate is the limiting element. During the 1960s, typical positioners were very simple, with few options, and they had limited air flow capacity.

In the early days of automatic control, special high-performance valves received special engineering attention. This led to papers that made the case that positioners could have a negative effect on control. The original positioners had a one-size-fits-all philosophy designed to work with any size of actuator. Price was a concern, and the possible improvement in control was not valuable enough to justify the extra costs, not only for purchase, but also for installation and operation. It is clear that in applications using larger valves, control improves with the use of positioners because of the servomechanism error correction. Concern with the speed of response led to higher air flow capacity. Later positioners were not linear in response, with low gain for a small position error and higher gain for larger errors. This also has the result that stem motion speed will differ with signal step size. Some modern positioners adjust to suit the specific application and may include adjustable gain and integral action. Digital devices can improve accuracy, be more repeatable, and include the flexibility of configuration for the specific application.

Positioners must have an air supply of 0.375 in. (10 mm) tubing as a minimum, and each positioner requires a separate adequate supply filter-regulator. Control valves use substantial amounts of air, and severe control interaction can occur if an inadequate air supply affects the air supply pressure to other valves and transmitters. The interactions can be very complex, and the oscillations are impossible to stop.

Increase Force

The second available significant improvement is to increase the actuator force. A larger actuator will provide a greater force but adds cost and increases the air consumption. Increasing air supply pressure, up to the safe actuator operating pressure, improves response. An undersized or marginal actuator will exhibit slow response because it requires a greater change in pressure to start motion, and as actuator pressure approaches supply pressure, the flow rate will decrease. Many diaphragm actuators commonly used at 25 PSIG (170 kPa) can safely operate at pressures to 30 or 60 PSIG (200 or 400 kPa). It is necessary only to increase the air supply pressure and modify the calibration or use a pressure amplifying relay. If the change in actuator pressure required to move the stem is reduced, then the jump in position is reduced. The analogy might be using a wrench with a longer handle to loosen a frozen bolt. Control of the wrench is much easier if it is not necessary to exert your full strength on the wrench handle. There are practical limits to actuator sizes for each valve body size and type. Some actuator schemes use stacked actuators in line to achieve greater forces.

Reduce Friction

Perfect alignment of the valve actuator system and stem will minimize friction. The very close tolerances between the valve stem and the stem guide can only tolerate a small misalignment without binding up (locking) the assembly and prevent any movement. Packing friction will increase with misalignment. Careful attention during maintenance is important. Any looseness or play in the assembly will cause unpredictable friction.

The proper tension on the packing follower bolts will prevent valve stem leakage without adding unnecessary friction. Careless overtightening may make the valve nearly inoperative and greatly reduce the life of the packing.

Packing selection and design are important (see Section 6.5). Graphite packing or packing with graphite content, typically specified for high-temperature operation, will substantially increase friction. Stem packing friction will vary over a range of 10:1 or even more, depending on packing materials and designs. It is necessary to remember that the packing temperature may differ considerably from the process temperature towards the ambient and will depend on the design and installation. Valve manufacturers have data to help in these decisions. Optional extended packing designs provide even greater thermal isolation from the process. The low-friction TFE packings popular in the chemical industry are limited to moderate temperatures. Modern elastomeric low-friction packings can serve at temperatures far higher than the temperature limits of fluorocarbon packings and have passed standard fire tests in some valves (see Section 6.5). Packing selection is no longer a simple task.

Defining Response

The ANSI/ISA-75.25.01 Standard, "Test Procedure for Control Valve Response Measurement from Step Inputs," defines step sizes for signal input step change in terms of the resulting response. It defines four step sizes. First is a signal step so small that there is no response. The next larger step size results in some response, but it does not match the required specifications in amplitude or time. The next larger step size results in a response satisfying the specifications used. The largest step size is large enough that the amplitude is within specifications but the time is longer than specified. In each case, the specification used is a function of the application.

As a useful definition for response time, the ISA (Instrumentation, Systems, and Automation Society) standard uses T_{86}, defined as the time for the valve to complete 86.5% of the final motion.

Measuring Response

Shop Test Valve testing "on the bench" at the factory or in a maintenance shop proves that the valve is mechanically functional, demonstrates valve response without fluid forces, and is much more convenient than field tests. A serious limitation is that there is no way to prove that the packing as installed will prevent stem leakage, and there is only a very limited ability to include fluid forces. (See Figure 6.9e.)

In Situ Test In situ (in place) testing has the advantage of providing the desired information about the valve and process response under actual service conditions. The packing is fully functional; the test fluid pressures and temperatures are the service conditions.

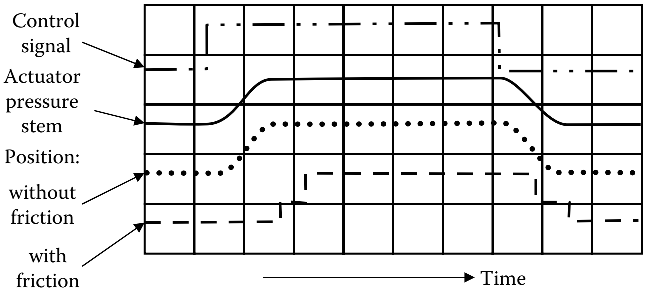


FIG. 6.9e

Response curves: On the top—step change in control air signal, second from top—air pressure inside the actuator, third from top—position of the (theoretical) valve stem w/o friction, fourth from top—position of an actual valve stem with friction.

There are two kinds of in situ tests. Normal operation with normal flows, temperatures, and pressures provides the "perfect model" and provides the required data. Operating constraints may limit the allowable range of valve opening and testing. The alternative is to take data from a shutdown unit. This can determine only the valve mechanical response and may not be completely representative of the operating situation. In either case, it may be necessary to add temporary or permanent instrumentation to gather all of the desired information. Valve manufacturers and service companies can provide add-on instrumentation. Data collection may be as simple as personal field observation of stem movement and positioner gauges. It is easy to identify large stem position jumps and slow action by observing the signal gauges and the valve. If this approach does not provide adequate information, it may be necessary to introduce precise signal step changes of various sizes as shown in the above referenced standard and to take measured data from the valve stem along with the related flow meter and other process sensors. There is a general opinion that the usual standard process instrumentation is rarely adequate to collect all of the desired data. The user will want to take advantage of all the available resources and to develop and justify a plan to quantify the problem and determine the solution.

Relationships

A 1% change in stem position does not translate into a 1% change in flow coefficient and result in a 1% change in the process variable. The engineering requirement is to determine the effect of flow changes on the process and then derive the required control valve specifications from that.

Determine Required Response Specifications

In situ testing should provide enough information to define the required valve response specifications. The valve manufacturer can provide guidance for sound decisions for improving the performance of a specific valve design.

The engineering analysis required to predict valve requirements is not difficult. For a number of reasons, the most common error in valve specifications is to oversize the valve. An oversized valve will increase the effective dead band by that same percentage. The definition of inherent rangeability of a control valve is: "Rangeability: The ratio of the largest flow coefficient (C_v) to the smallest flow coefficient (C_v) within which the deviation from the specified inherent flow characteristic does not exceed the stated limits" (Ref: ISA-75.05-1983). The ANSI/ISA-75.05.01-2000 Standard includes a definition for C_{vR} that is a calculated valve coefficient for the installed valve. This provides a way to quantify the real valve performance and to determine the effective rangeability.

The installed rangeability for an oversized valve may be much less than for a properly sized valve, because control is impossible at small valve openings due to the dead band and very high gain near the seat, and the excess capacity of the fully open valve has no effect on flow. Accurate service condition information is vital. It is far too common to add excessive contingency factors in sizing pumps and valves. A full process simulation will provide valuable data about process gain and valve gain and result in an understanding of how dead band will affect control. In the absence of the simulation, the user will need to estimate the valve gain at operating conditions and to estimate the response time needed. It is not economical to specify very tight control valve dynamics response specifications except where the process conditions justify this added cost.

Application Examples

Simple Tank Level Control Fluid enters the tank at a variable rate; an outlet valve controls level, and process considerations determine acceptable level limits. The analysis consists of considering the tank volume, the normal volume flow rates, and the allowable level deviation from set point. If the inlet flow can suddenly stop, calculate the required outlet valve response time from the rate of change in level. The required maximum allowable valve dead band is determined from the required precision for level. The allowable level dead band can be converted into flow dead band and then into the stem dead band at operating conditions in order to calculate the required valve dead band specification. A simple pressure control loop is a very similar problem.

Heat Exchanger One fluid exchanges its heat energy with another fluid in a heat exchanger. The process engineer should provide the range of flow rates and temperatures and the impact of a change in the heating/cooling flow rate. In most cases, the response time is not an issue. The sensitivity, or gain, of the temperature from the flow rate changes will determine the allowable dead band. Because this is a slow process, the dead band effects become very visible. The common attempt to conceal the dead band by reducing controller gain and lengthening reset time will only make the cycle time so long that everyone forgets it. Any reduction in dead band will improve control. Knowledge of the expected valve dead band will permit calculating the required controller settings.

Flow Control

Most flow measurements are inherently noisy; the associated controller has low gain and fast reset function to filter the higher frequency valve control signals and still bring the flow towards the set point. The valve speed of response must be adequate for the anticipated pressure upsets. Dead band is not an issue because the controller signal changes rapidly and often. Usually, the control is adequate with a standard valve. In compressible flow, the downstream volume provides some smoothing. Heavy filtering at the transmitter will make everything look nice but the system will respond slowly to upsets.

Reactor Mixing

A fast process reaction resulting when two fluids mix in a mixing volume is difficult to control. Downstream mixing will not make up for the errors. Past practice has been to make both flow measurements as similar as possible, and the control systems and valves as similar as possible, in order that the dynamics all are similar and, wishfully, all the dynamic errors will match and balance out. The process reaction is over and gone too quickly. The dead band must be very small, and the required precision of adjustment is very high. Special motor-operated small valves have satisfied these requirements. Process data is required to determine the allowable dead band.

Neutralizing Waste Water

This is another difficult control problem; flows and stream analysis may vary widely (see Section 8.32). The worst problem is that the neutralizing curve may have a very high gain over a narrow range, and this may vary with time. The region of very high gain is usually close to or at the set point value. The requirements are that the valve have enormously high precision and accuracy and then operate over a wide range. There are no real valves available to satisfy these specifications. The most popular solution is to use two valves in parallel. The small valve provides the precision and the larger bypass valve provides the bulk of the flow capacity. One reset-only controller adjusts the larger valve to cause the pH proportional-only controller to adjust the smaller valve to operate near 50% open. This solves both of the problems of precision and capacity. The natural dead band of the larger valve provides a gap action to allow the smaller valve to control over the small flow range. The price paid is a slow speed of response for large demand changes. Common practice is to use a downstream surge tank to mix and average the outflow pH. The design problem is to size the small valve large enough to be in control most of the time but to not oversize it and have too large a dead band. In real life, this means the study of the predicted neutralization curves and converting this information to the equivalent stem motion. For even greater rangeability, it is possible to extend this concept to a third valve. In each case, prediction of each dead band is valuable information.

Antisurge Valve

A centrifugal compressor requires a certain minimum mass flow through the machine in order to maintain a discharge pressure. If gas flow decreases below this value, then the compressor will lose flow and pressure and start to cycle internal flows and pressures. In larger machines, this surging puts excessive force on bearings and shafts and can quickly destroy the machine. Normal practice is to measure the gas mass flow into the machine and open an antisurge valve to discharge to atmosphere or to recycle enough gas around the machine to maintain the minimum flow. On a sudden decrease in gas demand, such as an interlock shutdown, the valve must open very quickly and then begin to throttle as needed in order to maintain stable operation. The compressor manufacturer can provide information on the required valve opening speed, or it is possible to estimate these values from the machine curve and knowledge of the pressurized volume. Adequate response may require the dead band high-volume booster relay described in Section 6.2. Other issues for these valves are associated with aerodynamic noise.

Delay or Slowdown Valve Action

There are a number of situations where it is desirable to delay or to slow valve action. For example, water hammer is the excessive pressure surge that is the result of a sudden change in fluid flow in a very long pipeline. Accurate control of the valve is still important; in some installations, the fluid flow will exert a high force and rapidly force the valve closed.

When flow suddenly changes, a pressure surge wave will travel through the fluid at the speed of sound (c) in the fluid. This pressure wave will reflect off the first solid surface (elbow or tee) and travel back and forth. This wave may shake or damage the valve and the piping. The situation is strongly dependent on the exact flow and piping situation. Even a few percentages of entrained air or gas bubbles in a liquid will increase the compressibility enough to make a considerable difference. Reported problems include pipe hangers broken in a 2-in. pipe line at a high flow rate and very fast shut-off valves. A 4-ft.-diameter fiberglass waste water pipe system collapsed when the upstream valve closed too quickly. Some hydraulic power plants use surge tanks to absorb the energy. Possibly the best scheme is to use a reset-only controller to get a predictable, linear delay. With a digital control system, a ramp function configuration will provide an accurate and dependable rate of closure or opening. If the valve is on/off, an equal percentage-type trim will provide some smoothing of flow. Some authors have suggested a fast-acting pressure control loop can override the control signal to slow the valve to limit the overpressure.

A very noisy process signal can lead to excessive valve or packing wear. The “inverse derivative” relay, which passes a fraction of a sudden input change immediately and the balance of the change later, will provide a low pass filtering effect (see Equation 1.2[14]). A needle valve, or restrictor, used with a volume chamber will provide a resistor/capacitor-type filter.

Safety Solenoid Valves

Emergency shutdown systems (ESDs) may use a solenoid valve piped between the positioner and the valve actuator installed to reduce the chance that the positioner could interfere with positive valve action. Consider that the air flow capacity (C_v) of the solenoid might limit air flow and valve response. See Section 6.2 for more details.

Troubleshoot Valve Response

Control valve response may be poor because of

- Oversize valve
- Incorrect trim
- Excessive packing friction
- Undersized actuator
- Loose linkage
- Inadequate mounting
- Inadequate air supply capacity or pressure, or plugged air filter
- Damaged or leaking tubing
- Leaking air lines or fittings
- Inadequate positioner flow capacity
- Broken actuator spring
- Incorrect calibration
- Leaking or damaged diaphragm or piston
- Positioner/actuator time constant equal to process time constant
- Manual valve handwheel or limit stops prevent free motion
- Bypass valve open
- Valve body plugged
- Broken valve stem
- Failed or plugged interlock solenoid valve

Some of these possibilities are unlikely depending on the situation history; that is, did it once work well? Did it have recent maintenance? Are there changes to the flow system? In all the above, once the cause is identified, the solution is usually clear. Identification may be as simple as from observing the positioner gauges and the valve action. As in any control situation, instability occurs when phase (time) lag causes the control loop to react too late. A valve actuator has capacitance; that is, the valve does not change position when air flow starts to enter but moves only when enough air accumulates to build up pressure and cause motion. The classic control oscillation problem occurs when the valve time constant is the same as the rest of the process. A positioner has a small input capacitance, but because it causes large flow amplification, it reduces the effective actuator capacitance. A control loop tuned with very little gain and a long reset time may have a cycle time of hours and hide the cycle. It is often very difficult to discover whether a given control loop is causing a cycle or is just responding to another loop.

Bibliography

ANSI/ISA-75.25.01-2000, "Test Procedure for Control Valve Response Measurement from Step Inputs."

ANSI/ISA-TR75.25.02-2000, "Control Valve Response Measurement from Step Inputs."

ANSI/ISA-S75.05-1983, "Control Valve Terminology."

ANSI/ISA-S51.1-1979 (Reaffirmed 1993), "Process Instrumentation Terminology."

ANSI/ISA-S75.13, "Method of Evaluating Performance of Positioners with Analog Input Signals and Pneumatic Output."

Control Valve Dynamic Specification, EnTech Control Inc., Toronto, Canada, 1998.

References

Rich, G., *Hydraulic Transients*, Mineola, NY: Dover Publications, Inc., 1963.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1994.

Coughran, M. T., "Measuring the Installed Dead Band of Control Valves," ISA TECH/97-1114.

Champagne, R. P. and Boyle, S. J., "Optimizing Valve Actuator Parameters to Enhance Control Valve Performance," *ISA Transactions* 35 (1996) 217-223.

Instrumentation, Systems, and Automation Society-S26 (1968) and ANSI MC4.1-1975, "Dynamic Response Testing of Process Control Instrumentation."

Senior, K. A., "Technical Guidelines and Design Information, Using KVSP Packing Systems for Improving Process Control and Minimizing Fugitive Emissions," white paper, DuPont Dow L.L.C., 2002.

Senior, K. A., "Valve Packing Systems Improve Process Control," *Chemical Processing,* June 1997.

Brestal, R. et al., "Control Valve Packing Systems," technical monograph 38, 1992, Marshalltown, IA: Fisher Controls.

"Packing Selection Guidelines for Sliding Stem Valves", Bulletin 59.0:062, Marshalltown, IA: Fisher-Rosemount.

Langford, C. G., "A User's View of Process Control and Control Valve Positioners," ISA, 1996, Paper #96-076 1054-0032/721-728.

Cho, C. H. "Using Control Element Characteristics to Compensate for Process Nonlinearities," *Control Engineering,* August 1974.

6.10 Emergency Partial-Stroke Testing of Block Valves

A. S. SUMMERS (2005)

Features:	Discussion of partial-stroke testing of block valves
Purpose:	Enhanced diagnostics can be used to achieve higher safety integrity level using partial-stroke testing.
Codes, Standards, and Recommended Practices:	Instrumentation, Systems, and Automation Society (ISA), ANSI/ISA 84.00.01-2004, "Functional Safety: Electrical/Electronic/Programmable Electronic Safety-related Systems," Research Triangle Park, NC International Electrotechnical Commission (IEC), IEC 61508, "Functional Safety: Safety-Instrumented Systems for the Process Industry Sector," Geneva, Switzerland International Electrotechnical Commission (IEC), IEC 61511, "Functional Safety: Safety Instrumented Systems for the Process Industry Sector," Geneva, Switzerland

INTRODUCTION

ANSI/ISA 84.00.01-2004 (ISA 84) and IEC 61511 are new functional safety standards, covering the design, implementation, operation, maintenance, and testing of safety instrumented systems (SISs). Successful implementation of the safety life cycle model associated with these standards hinges on an essential design constraint: the safety integrity level (SIL). The SIL is a numerical benchmark, related to the probability of failure on demand (PFD). SIL is affected by the design quality, e.g., device integrity, voting, and common cause faults, and by the operation and maintenance strategy, e.g., diagnostics and testing interval.

For many operating companies, one of the most difficult parts of complying with the standards is the testing interval often required for final elements, such as emergency isolation valves or emergency block valves. Traditionally, these valves have been tested at unit turnaround, using an off-line, full-stroke test to demonstrate performance.

Thirty years ago, turnarounds were relatively frequent, occurring on average every 2–3 years. Due to successful mechanical reliability and preventive maintenance programs, many operating companies have been able to extend unit turnarounds. In some industries, it is now common to have turnaround intervals of 5–6 years. Extended turnaround intervals yield great economic returns through increased production. However, extended turnaround intervals also mean that block valves are expected to go longer between function tests, yet still achieve the same performance. This is simply not possible.

When SIL 2 or SIL 3 performance is required, 5-year function tests are inadequate. Consequently, it is necessary to supplement the off-line full-stroke test. This involves implementation of valve diagnostics, such as partial-stroke testing, or alternate testing strategies, such as on-line full-stroke testing.

Many users consider partial-stroke testing (PST) as a cost-effective alternative to on-line full-stroke testing (FST). The use of PST often eliminates the need for full flow bypasses, reducing engineering, capital, and installation costs, as well as potentially removing a bypass that could be inadvertently left open. Partial-stroke testing improves the block valve performance, as measured by the Average Probability of Failure on Demand (PFD_{AVG}). The amount of the reduction is dependent on the valve and its application environment.

This section will discuss the method of determining the actual impact of the partial-stroke test on PFD_{AVG}. It will also present a discussion of the three partial-stroke testing methodologies that are currently being evaluated and used by industry.

THE PARTIAL-STROKE TEST

There are three basic types of partial-stroke test equipment: mechanical limiting, position control, and solenoid valves. Each type involves different levels of sophistication and risk.

Mechanical Limiting

Mechanical limiting methods involve the installation of a mechanical device to limit the degree of valve travel. When mechanical limiting methods are used, the valve is not available for process shutdown.

The mechanical devices used for partial-stroke testing include collars, valve jacks, and jammers.

- Valve collars are slotted pipes that are placed around the valve stem of a rising stem valve. The collar prevents the valve from traveling any farther than the top of the collar. Any fabrication shop can build a valve collar, suitable for test use.
- A valve jack is a screw that is turned until it reaches a set position. The valve jack limits the actuator movement to the screw set position. The valve jack is ordered from the valve manufacturer when the valve is purchased. Valve jacks work with both rising stem valves and rotary valves.
- Mechanical jammers are integrated into the rotary valve design. They are essentially slotted rods that limit valve rotation when placed in position using an external key switch. Since the jammer is integrated into the rotary valve, the jammer must be purchased from a valve manufacturer. A contact can be provided for the key switch to allow annunciation in the control room whenever the key is used.

Mechanical limiting methods are inexpensive in terms of capital and installation costs. These methods are manually initiated in the field and are personnel-intensive.

A limit switch or visual inspection is used to confirm block valve movement. Successful test implementation and return of the block valve to normal operational status are completely procedure driven. For valve collars and jacks, bypass notification to the control room is entirely procedural. For the jammer, automatic notification using the key switch contact can be provided.

Two of the biggest drawbacks to these methods are the loss of protection that occurs during the test and the lack of assurance that the valve is in or has been returned to normal status. There is no way to know for certain that the jack or jammer has been completely retracted without actuating the valve. Furthermore, unauthorized use of the valve jack or jammer cannot be determined by casual inspection. This means that the valve could potentially be out of service, with operations personnel unaware of the situation.

These methods do not add to the normal operating spurious trip rate. However, there is the potential for a spurious trip during the partial-stroke test. For valve collars, the main culprit of spurious trips is improper installation, causing the collar to pop off the stem when the valve begins to move. Jacks and jammers must be placed in service by the technician; so procedural mistakes can result in the valve closing completely rather than just partially. Therefore, these methods are really only as good as the written procedures and technician training.

Position Control

Position control uses a positioner to move the valve to a predetermined point. This method can be used on rising stem and rotary valves. Because most emergency block valves are not installed with a positioner, this method does require installation of additional hardware. Consequently, cost is a major drawback for the position control method.

A limit switch or position transmitter can be used to determine and document the successful completion of the tests. If a smart positioner is used for the position control, a HART maintenance station can collect the test information and generate test documentation. Of course, the use of a smart positioner and maintenance station further increases the capital cost.

Some manufacturers have promoted the use of the positioner in lieu of a solenoid valve for valve actuation. However, most positioners do not have a large enough vent port (C_v) for rapid valve closure. Consequently, a solenoid valve should still be used for valve actuation. This solenoid valve must be installed between the positioner and the actuator.

The positioner does contribute to the spurious trip rate during normal operation, because the positioner can fail and vent the air from the valve. When a solenoid valve is installed between the positioner and the actuator, the safety functionality is never lost during the partial-stroke test. De-energizing the solenoid valve will shut the valve, regardless of the positioner action.

Solenoid Valve

A partial-stroke test can be accomplished by pulsing a solenoid valve. The solenoid valve can be the same as the one used for valve actuation, resulting in lower capital and installation costs for this method than other methods. If the actuation solenoid valve is used, this method will also test the solenoid valve's capability to execute safe shutdown.

Single Solenoid Valves The Minerals Management Service, which oversees safety for oil and gas operations for U.S. offshore waters, provides one method for partial-stroke testing. This method relies on the operator to pulse a single solenoid valve by turning a field-mounted switch, which de-energizes the solenoid coil for as long as the field operator holds the switch.

The field operator monitors the valve position and releases the button when the operator confirms valve movement. When the valve moves, it can be inferred that the solenoid valve successfully vented. Of course, the main risk is that the operator may hold the switch too long, allowing the valve to close sufficiently to disrupt the process, resulting in unit shutdown.

It is also possible to automate the single solenoid valve test using a pulse timer adjusted to achieve the desired valve travel. Valve travel confirmation is accomplished using a limit switch or position transmitter, allowing automatic documentation of test status. Because a failure of the solenoid valve or block valve may result in excessive block valve travel, the pulse timer should be voted with the limit switch or position transmitter. If the valve reaches its desired travel point before the pulse timer is finished, the solenoid valve is reset. The test can be programmed in the SIS logic solver with the test implemented automatically on a programmed cycle time or initiated by the operator per a maintenance schedule.

Another method is to measure the block valve position as related to air pressure in the actuator during the time of the solenoid pulse. This results in a "fingerprint" of the breakaway pressure for block valve closure, which can be compared with the original valve fingerprint. In order for this method to be effective, maintenance must have a specific procedure for examining the fingerprint to identify that the block valve is degraded and to respond with corrective action.

When a simplex solenoid valve is being used to PST the block valve, the solenoid is de-energized and then re-energized. If the solenoid valve does not reset, the test becomes a trip. Using redundant solenoid valves can essentially eliminate this problem.

Redundant Solenoid Valves When arranged in a two-out-of-two (2oo2) configuration, redundant solenoid valves provide improved reliability during normal operation and reduce the probability of a spurious trip during the PST. For processes that are sensitive to spurious trips, the reliability improvement is typically sufficient to justify the additional capital and installation costs.

There are commercially available solenoid valve packages that provide on-line diagnostics of solenoid coil failure, facilitate on-line solenoid valve testing, and perform on-line, partial-stroke testing of the block valve. One particular package (patent pending) operates in a one-out-of-one hot standby (1oo1HS) configuration. During normal operation, the air signal passes through the package from the signal source to the valve actuator. When a trip occurs, the solenoid package vents the air from the valve actuator and allows the valve to move to its fail-safe position.

With the 1oo1HS, one solenoid valve is used as the primary actuation solenoid and is confirmed on-line using a pressure switch. A secondary solenoid valve is off-line and confirmed in the vented state (off-line state) by a pressure switch. The safety logic solver is programmed so that if the primary solenoid valve goes to the vent state without being commanded (as detected by the pressure switch), the secondary solenoid valve is energized, preventing the spurious trip.

Solenoid valve testing is performed by cycling the solenoid and by verifying that each solenoid valve successfully vents and resets using the pressure switches. The 1oo1HS can be used for PST by incorporating a PLC timer to pulse the power to the solenoids for just long enough to achieve the partial stroke. To verify the movement of the valve, a position transmitter or limit switch is used.

The position indication is also used to prevent overstroking of the block valve; i.e., if the valve moves too far during the timed stroke, the solenoids are re-energized. For preventative maintenance activities, overstroke or understroke alarms can be configured to let maintenance know if the valve is moving too quickly or too slowly during the test.

TABLE 6.10a

The Relationship among the Safety Integrity Level (SIL), the Average Probability of Failure on Demand (PFD_{AVG}), and Risk Reduction

SIL	PFD_{AVG}	Risk Reduction
4	0.0001 to 0.00001	10,000 to 100,000
3	0.001 to 0.0001	1,000 to 10,000
2	0.01 to 0.001	100 to 1,000
1	0.1 to 0.01	10 to 100

Impact of PST on SIL

Safety integrity level (SIL) and MTTF are defined by the average probability of failure on demand for demand mode SIS. In Instrumentation, Systems, and Automation Society 84 and IEC 61511, there are four SIL classes. Each SIL class provides an additional order of magnitude risk reduction, as shown in Table 6.10a.

The SIS standards require an examination of the PFD_{AVG} to ensure that the risk reduction assigned to the SIS is met. It is important to note that SIL is not a property of a specific device. It is a SIS property, encompassing the field sensors through the logic solver to the final elements (e.g., solenoid valve, block valve, or pump motor control circuit).

SIL is an important concept for safe operation, but plant management demands that the process plant operate at a high utilization rate. If a SIS is installed that has a low Mean Time to Failure Spurious ($MTTF^{spurious}$), the project will be considered a failure, regardless of the SIL that the SIS achieves. Consequently, any SIS assessment should include an analysis of the $MTTF^{spurious}$.

Let's examine a typical SIS, including transmitters, a redundant logic solver, solenoid valves, and block valves. In order to perform the PFD_{avg} and $MTTF^{spurious}$ calculations, failure rate data is required. This data can be selected from various industry-published databases. The values used in this analysis are shown in Table 6.10b.

For illustration purposes, the analysis will be presented in two parts: 1) the impact of PST on block valve performance, and 2) the impact of PST on the SIS performance. For the latter case, a dual solenoid valve package (1oo1HS) will be used for block valve actuation and PST. Other PST methods can be assessed using similar techniques.

The results for other PST methods may provide very different results for PFD_{AVG} and $MTTF^{spurious}$. Consequently, the results presented in this section should not be generalized for all PST equipment. The reader is further cautioned to ensure that any data used during SIL verification is appropriate for their application. In other words, the results presented here may not be directly applicable to the reader's application.

TABLE 6.10b
Failure Rates Used in the Analysis

Device	*MTTF^D (Years)*	*$\text{MTTF}^{spurious}$ (Years)*	*Data Sources$^{(1)}$*
Transmitters	100	100	OREDA data for various transmitters
Redundant Logic Solver	5000	Not included in analysis	Vendor data for 1oo2D, 2oo3D, or 2oo4D logic solvers
Solenoid Valve	60	15	TR84.00.02 data
Pressure Switch	75	60	CCPS data
Block Valve	40	455	OREDA data

(1) Due to the nature of the analysis presented in this chapter, values were selected to represent generic devices rather than specific devices. For the evaluation of an actual SIF, failure rate values representing the specific devices should be used.

Block Valve Analysis

Block valve components can be examined to determine which failures potentially result in the valve not operating when a process demand occurs. The causes of failures and associated modes of failure are shown in Table 6.10c.

The PFD_{AVG} can be calculated using the dangerous failure rate (λ^D) and the testing interval (TI). The mathematical relationship, assuming that the mean time to repair is small compared to the testing interval and that $\lambda^D \times \text{TI}$ is much small than 0.1, is as follows:

$$\text{PFD} = \lambda^D \times \text{TI}/2 \qquad \textbf{6.10(1)}$$

TABLE 6.10c
Causes of Block Valve Failures and the Associated Modes of Failure

Cause of Failures	*Modes*
Actuator sizing insufficient for valve actuation	Valve fails to close (or open)
Valve packing is seized	Valve fails to close (or open)
Valve packing is tight	Valve is slow to move to closed or open position
Air line to actuator crimped	Valve is slow to move to closed or open position
Air line to actuator blocked	Valve fails to move to closed or open position
Valve stem sticks	Valve fails to close (or open)
Valve seat is scarred	Valve fails to seal off
Valve seat contains debris	Valve fails to seal off
Valve seat plugged due to deposition or polymerization	Valve fails to seal off

TABLE 6.10d
Average Probability of Failure on Demand of a Typical Block Valve Given as a Function of Testing Interval (Note: Does Not Include the Contribution of the Solenoid Valve)

Testing Interval	*PFD_{avg}*
1 year	1.25E–02
2 years	2.50E–02
3 years	3.75E–02
4 years	5.00E–02
5 years	6.25E–02

Thus, the relationship between PFD and TI is linear. Longer test intervals yield a larger PFD_{AVG}. The Offshore Reliability Data Handbook (OREDA) database has data for various valve types and sizes. For the purposes of illustration, a MTTF^D of 40 years will be used. The failure rate λ^D can be calculated from the MTTF^D using the simplified equation:

$$\lambda^D = \frac{1}{\text{MTTF}^D} \qquad \textbf{6.10(2)}$$

For a MTTF^D of 40 years, this yields a dangerous failure rate of 2.5e–02 failures per year. The valve failure rate varies with type, size, and operating environment (e.g., process chemicals, deposition, polymerization, etc.). The reader should determine the appropriate failure rate for the specific application.

The PFD, based on 2.5e–02 failures per year, is shown in Table 6.10d for various testing intervals. As expected, the valve performance at a 5-year testing interval is not the same as the valve performance at a 2-year testing interval. Due to the degraded performance at longer testing intervals, many companies have found that they must test the block valves on-line. Once facilities for on-line testing are installed, full-stroke testing can easily be performed. However, since a full-stroke test involves full contact of the valve seating members, frequent stroking can cause excessive wear to the block valve seat. This is a serious concern for soft-seated valves. Increased testing may achieve a higher integrity, but cause damage to the valve seat, leading to earlier valve failure.

Another major concern is that the plant is unprotected while the block valve is in bypass for testing. The fraction of the time that the valve is in bypass must be considered in the PFD calculation. If the valve is bypassed every 6 months for testing and the test takes 1 hour, the PFD is increased by 2.28^4. For longer bypass periods or more frequent testing, the impact on the PFD is even more significant. To maintain safety, operating procedures must include a list of the actions to be taken when the valve is in bypass, such as continuous monitoring of critical process variables and when manual shutdown should be initiated.

Partial-Stroke Testing An alternative option to a full-stroke testing is a partial-stroke test. The test involves moving the valve from the fully open position. This tests a portion of the valve failure modes. The remaining failure modes are

TABLE 6.10e
Dangerous Failures, Failure Modes, and Test Strategy

Failures	*Failure Modes*	*Test Strategy*
Actuator sizing is insufficient to actuate valve in emergency conditions	Valve fails to close (or open)	Not tested
Valve packing is seized	Valve fails to close (or open)	Partial- or full-stroke
Valve packing is tight	Valve is slow to move to closed or open position	Partial- or full-stroke, if speed of closure or resistance to closure is monitored
Air line to actuator crimped	Valve is slow to move to closed or open position	Partial- or full-stroke, if speed of closure or resistance to closure is monitored. Physical inspection
Air line to actuator blocked	Valve fails to move to closed or open position	Partial- or full-stroke
Valve stem sticks	Valve fails to close (or open)	Partial- or full-stroke
Valve seat is scarred	Valve fails to seal off	Full-stroke test with leak test
Valve seat contains debris	Valve fails to seal off	Full-stroke test
Valve seat plugged due to deposition or polymerization	Valve fails to seal off	Full-stroke test

tested using a full-stroke test. The main purpose of the partial-stroke test is to reduce the required full-stroke testing interval.

How does partial-stroke testing affect the PFD? The valve failures are modeled in two parts: 1) those failures that can be tested using the partial-stroke (PS), and 2) those failures that can only be tested using a full-stroke (FS). For the calculation, the dangerous failure rate, λ^D, must be divided into what can be tested at the partial-stroke ($\lambda^D{}_{\mathrm{PS}}$) and what can only be tested with a full-stroke ($\lambda^D{}_{\mathrm{FS}}$).

To determine the percentage of failures that could be detected using PST, the failure mode distributions for various valve types and sizes contained in the Offshore Reliability Data Handbook were examined. This evaluation can be done for any valve type, based on the application environment and the shut-off requirements.

Table 6.10e provides a listing of typical dangerous failures and failure modes for block valves. The test strategy indicates whether the failure mode can be detected by partial-stroke testing or only by full-stroke testing. Based on the OREDA data, the typical percentage of the failures that can be detected by a PST is 70% for many valve types and services. Additional analysis can be performed to justify a higher percentage of detected failures. However, it is very difficult to substantiate a percentage greater than 85% for process industry applications. Those failures that are not detected during the PST are tested using an FST.

An imperfect testing model is used for calculating the $\mathrm{PFD}_{\mathrm{AVG}}$ of the block valve when PST is utilized. The percentage of detected (PD) failures is used with the dangerous failure rate of the block valve (λ^D), testing interval (TI), and the mean time to repair (MTTR), as follows:

$$\mathrm{PFD}_{\mathrm{AVG}} = \mathrm{PD} \times \lambda^D \times \mathrm{TI}_{\mathrm{PST}}/2 + (1 - \mathrm{PD}) \times \lambda^D \times \mathrm{TI}_{\mathrm{FST}}/2 + \lambda^D \times \mathrm{MTTR} \qquad \mathbf{6.10(3)}$$

FST Supplemented by PST Table 6.10f provides a simple comparison of the $\mathrm{PFD}_{\mathrm{AVG}}$ for block valves tested using FST

TABLE 6.10f
Comparison of Average Probability of Failure on Demand (PFD_{avg}) for Block Valve(s) Undergoing Full-Stroke Testing (FST) Only or FST with Monthly Partial Stroke Testing (PST) (Note: Does Not Include Common Cause Failures or the Contribution of the Solenoid Valve)

	Single Block Valve		*Dual Block Valves*	
FST Interval (Years)	*FST only PFD_{AVG}*	*FST and Monthly PST PFD_{AVG}*	*FST only PFD_{AVG}*	*FST and Monthly PST PFD_{AVG}*
1	1.25^{2}	4.55^{3}	1.58^{4}	2.07^{5}
2	2.50^{2}	8.30^{3}	6.28^{4}	6.89^{5}
3	3.75^{2}	1.21^{2}	1.41^{3}	1.45^{4}
4	5.00^{2}	1.58^{2}	2.51^{3}	2.50^{4}
5	6.25^{2}	1.96^{2}	3.92^{3}	3.82^{4}

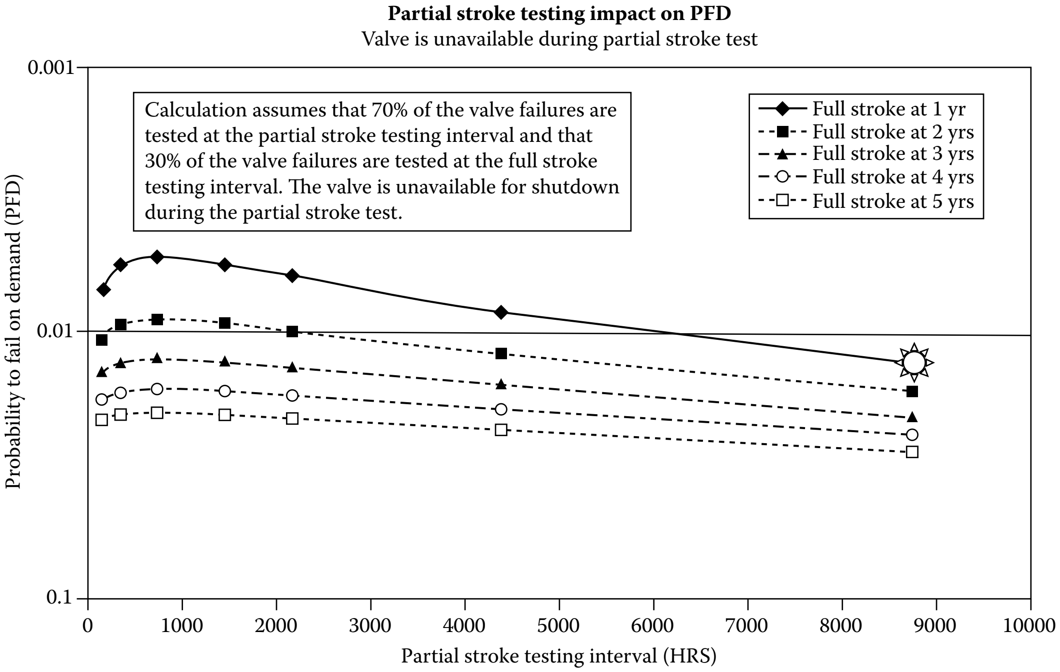


FIG. 6.10g
Relationship between the PST and PFD valves. Valve is unavailable during the test.

only and using FST supplemented with PST. The PFD_{AVG} is shown as a function of the full-stroke testing interval. For this illustration, the partial-stroke test is performed monthly. Similar calculations can be performed at other PST intervals. The results in Table 6.10f do not include the solenoid valve contribution to the PFD_{AVG}. The PFD_{AVG} for the single block valve is significantly reduced when PST is utilized. For double block valves, the PST lowers the PFD_{AVG} by nearly an order of magnitude.

The reader is cautioned that this breakdown is based on average valve performance and may not represent the breakdown for the reader's application. This evaluation should be done for each valve type, based on the application environment and the shut-off requirements.

If the service is erosive, corrosive, or plugging, the failure rate and failure mode breakdown will be different from that shown in this chapter. If the valve is specified as tight-shut-off, the contribution of minor seat deformation or scarring may be more significant than shown in this chapter.

Probability of Failure Using a dangerous failure rate of 2.5^2 failures per year, Figure 6.10g shows the PFD when the test procedure requires bypassing the valve during the test. As expected, the PST does improve the PFD_{AVG} performance of the valve. The star illustrates the point where the partial-stroke test and full-stroke test are both conducted at a 1-year interval. This corresponds to the result shown for a 1-year full-stroke test in Table 6.10a.

The downward trend of the curves for very frequent partial-stroke testing is due to the valve being bypassed during the test. This removal results in the valve not being available for the fraction of time that the valve is being tested. The calculation assumes that the total test time is 30 minutes. If the actual test time is longer, the effect will be more pronounced.

Figure 6.10h shows the PFD when the test procedure allows the valve to remain in service during the test. Very frequent partial-stroke tests improve the PFD substantially, because there is no loss of functionality during the test. Again, the star illustrates the point where the partial-stroke test and full-stroke test are both performed at a 1-year interval.

For both test procedures, partial-stroke testing does improve the valve performance. For example, 5-year full-stroke testing achieved a PFD of 6.25^2 (Table 6.10a). A 5-year full-stroke test supplemented with a 1-month (720 hours) partial stroke test achieved a PFD of 1.96^2, which is an approximately 30% reduction in PFD. In the cases of 1-year and 2-year full-stroke testing, a single block valve can potentially achieve SIL 2 performance when supplemented with frequent partial-stroke tests.

For longer full-stroke testing intervals, the valve performance can increase from low SIL 1 to high SIL 1, depending on the partial-stroke testing interval. From the graphs, it is easy to see that no amount of partial-stroke testing is going to allow a single valve to achieve high SIL 2 performance, let alone SIL 3 performance, at full-stroke testing intervals of 1 year or more.

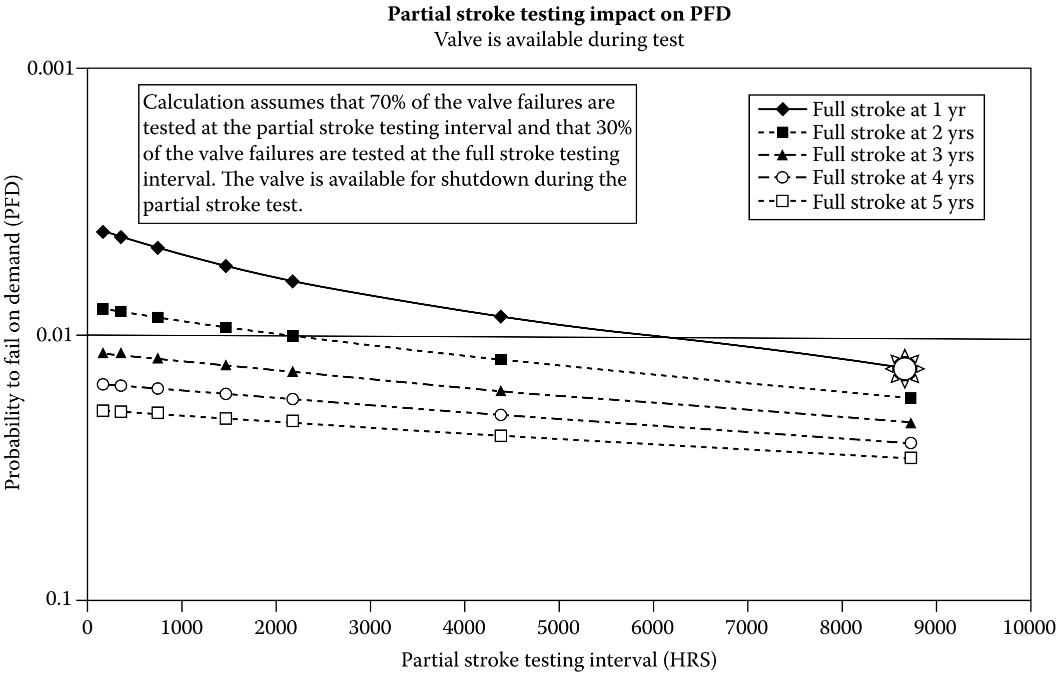


FIG. 6.10h

Relationship between PST interval and PFD. Valve is available during the test.

OVERALL SIS PERFORMANCE

Let's examine a typical SIS, including transmitters, a redundant logic solver, solenoid valves, and block valves. For the PFD_{avg} and $MTTF^{spurious}$ calculations, the failure rate data presented in Table 6.10b was used. In addition, the following assumptions were made:

- For redundant transmitters, it was assumed that analog signal comparison is performed and a fault alarm is initiated when the transmitters deviate unacceptably. "D" at the end of the voting architecture indicates the use of diagnostic coverage. The diagnostic coverage is assumed to be 80% for dual redundancy and 90% for triplicated redundancy.
- The PST can detect 70% of the valve dangerous failures.
- Common cause factor is assumed to be 2%.
- Mean time to repair is assumed to be 24 hours.

The results of the analysis were plotted using bar charts to illustrate the relative contribution of each device on the PFD_{AVG}.

Single Block Valve Case

For the simplex solenoid valve cases shown in Figure 6.10i, the SIS only achieved SIL 1 at annual function testing. When PST is used to supplement the full-stroke test, the SIS achieves very high SIL 1 to mid-range SIL 2. The individual bars can be examined to determine the major contributors to the PFD_{AVG}. For the 1oo1 and 2oo2D cases, the transmitter is the major contributing cause to the PFD_{AVG}. The impact of the transmitter can be decreased by changing the voting to 1oo2D or 2oo3D, which results in the single block valve being the major contributor.

For the spurious trip rate calculation, the logic solver was not included in the calculation, because the spurious failure rate can vary so much dependent on the specific architecture. Some logic solvers have high spurious trip rates, and their contribution to the overall $MTTF^{spurious}$ can overwhelm the other SIS components. For this illustration, the logic solver was not included in this calculation to allow the examination of the field devices. For actual installations, the logic solver must be included in the PFD_{AVG} calculation.

The benefit of using the 1oo1HS is evidenced in Figure 6.10j. In each case, the spurious trip rate was reduced significantly when the 1oo1HS was used instead of a simplex solenoid valve. The simplex solenoid valve cases had a spurious trip rate of 10 years for the 1oo2D transmitter case and 15 years for the other transmitter cases.

When the 1oo1HS is used, the $MTTF^{spurious}$ exceeds 45 years for the 1oo2D transmitter case and exceeds 420 years for the 2oo2D and 2oo3D transmitter cases. If the plant life is 20 years, the single solenoid valve cases yield an average of 1.5 trips during the life of the plant. Even if a spurious trip costs only $100,000, the simplex solenoid valve will result in the loss of $150,000 over the life of the plant. For the 1oo1HS cases, the $MTTF^{spurious}$ is greater than the plant life expectancy, yielding substantial savings.

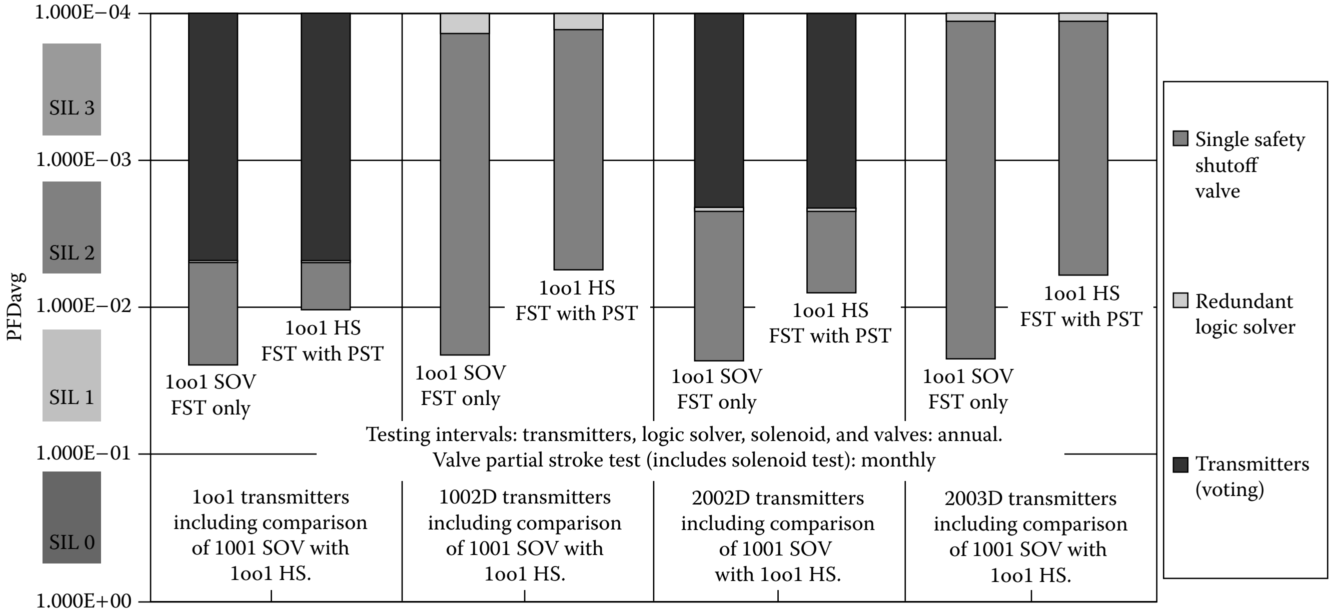
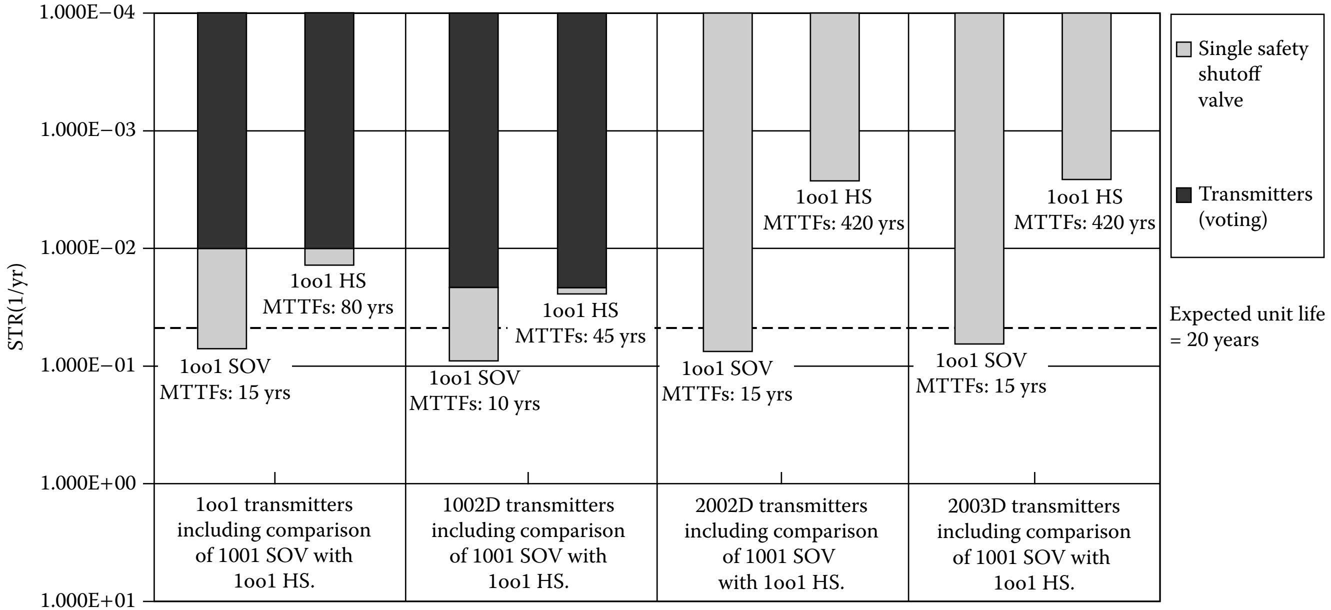


FIG. 6.10i
Plot of PFD_{AVG} showing the benefit of using PST of a single block valve.

Dual Block Valve Case

Dual block valves are used in many installations to provide greater assurance that the process is isolated. When dual block valves are used, it is common for test intervals to be extended to unit turnaround. Consequently, for the analysis, a 5-year full-stroke test was assumed for the logic solver, solenoid valves, and block valves, while annual testing was used for the transmitters.

For the simplex solenoid valve cases shown in Figure 6.10k, the SIS achieved high SIL 1 to the borderline between SIL 1 and SIL 2 at 1-year full-stroke testing. When PST is used to supplement the full-stroke test, the SIS achieves mid-range SIL 2 with 1oo1 and 2oo2D transmitter voting. For 1oo2D



This comparison includes field devices only. The contribution of the redundant logic solver to the spurious trip rate is not shown.

FIG. 6.10j
Plot of $MTTF^{spurious}$ showing the benefit of using PST of a single block valve.

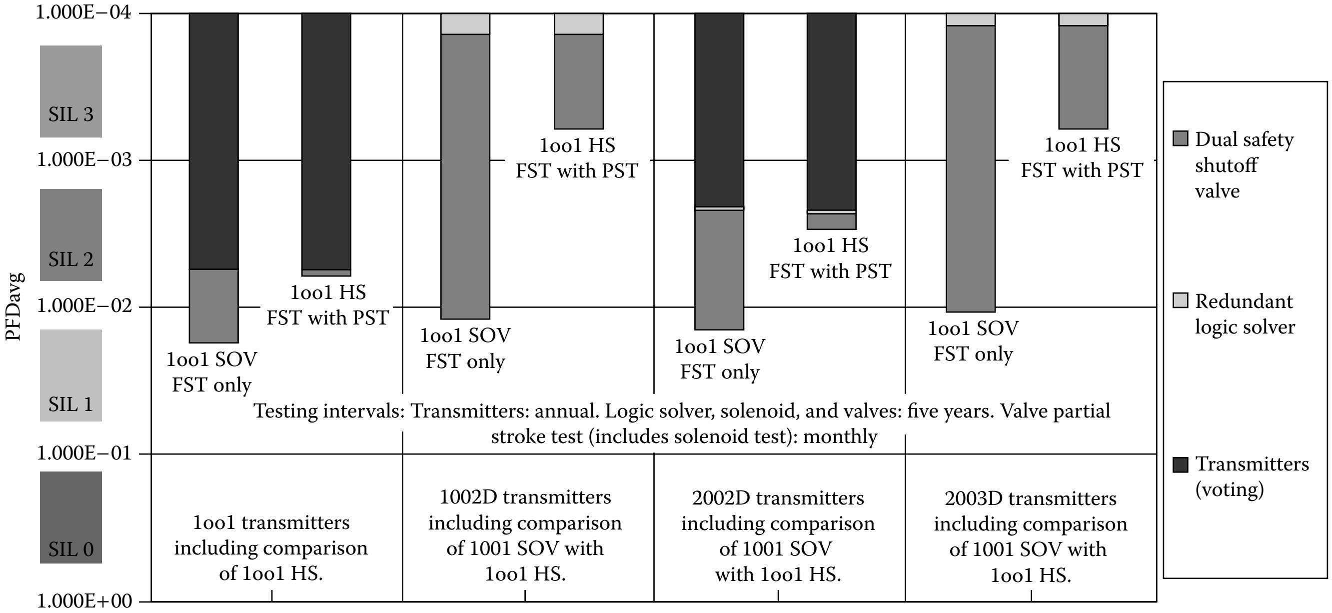
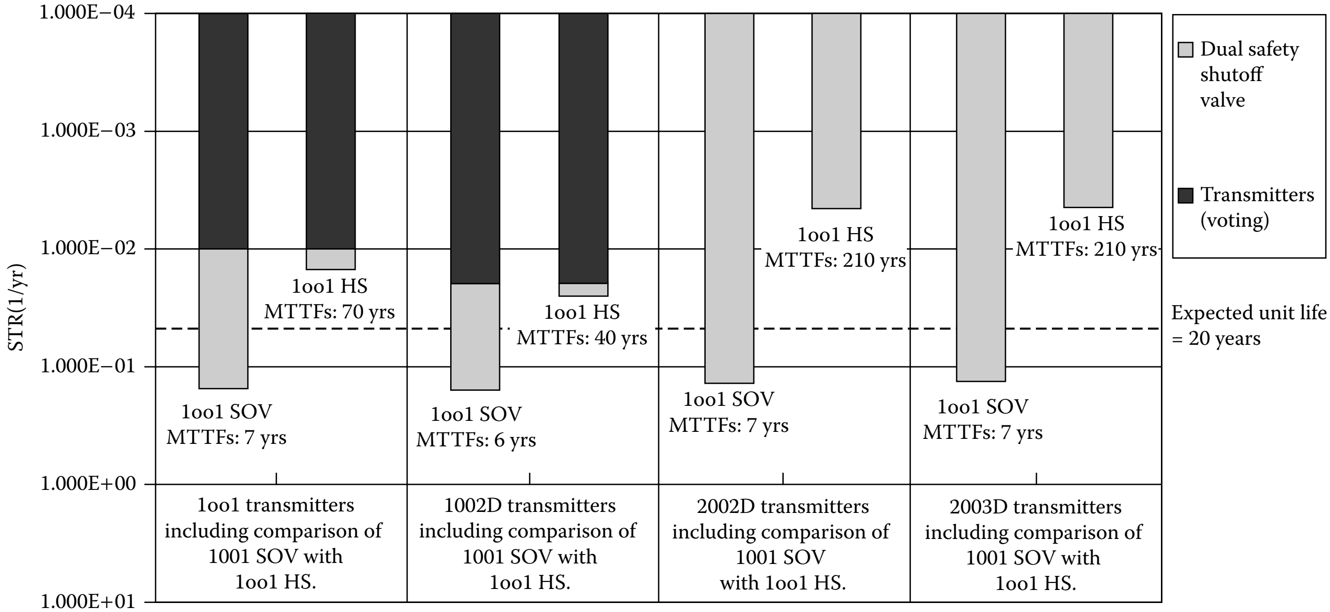


FIG. 6.10k
Plot of PFD$_{AVG}$ showing the benefit of using PST of a dual block valve.

and 2oo3D voting, the PST makes it possible to achieve SIL 3.

Again, the individual bars can be examined to determine the major contributors to the PFD$_{AVG}$. As seen in the simplex block valve SIS, the transmitter is the major contributing cause to the PFD$_{AVG}$ for the 1oo1 and 2oo2D transmitter cases. The contribution of the transmitter is seriously reduced when the voting is changed to either 1oo2D or 2oo3D. In these latter cases, the major contributors to the PFD$_{AVG}$ are the dual block valves.

The benefit of using the 1oo1HS is evidenced in Figure 6.10l. The MTTF^{spurious} for the simplex solenoid valve cases is an average of 7 years. When the 1oo1HS is used for valve actuation, the MTTF^{spurious} is more than 40 years for



This comparison includes field devices only. The contribution of the redundant logic solver to the spurious trip rate is not shown.

FIG. 6.10l
Plot of MTTF^{spurious} showing the benefit of using PST of a dual block valve.

the 1oo2D transmitter case and more than 210 years for 2oo2D and 2oo3D transmitter cases.

If the plant life is assumed to be 20 years, the simplex solenoid valve cases yield an average of 2.8 trips during the life of the plant. Again, if a spurious trip costs $100,000, the plant will lose more than $280,000 over the life of the plant due to spurious trips.

CONCLUSIONS

Partial-stroke testing does provide measurable improvement of the PFD over full-stroke testing alone. The amount of improvement is dependent on the specification, configuration, and application environment. The three partial-stroke testing methodologies offer choices between manual and automated testing.

It is important to remember that partial-stroke testing is used to achieve diagnostics on the valve in lieu of on-line, full-stroke testing. Some of these methods presented in this chapter have a higher potential for spurious block valve closure (e.g., on-line spurious actuation or unintentional actuation during test) than others. For processes that are sensitive to spurious trips, the selection of specific method should take into account not only the diagnostic capability, but also the reliability.

Whichever method is selected, procedures must be written to ensure that the block valve is not tripped during testing, the test is properly carried out, incorrect valve performance is documented, and maintenance is performed to return the valve to fully functional status. The primary issue is that partial stroke testing can reduce the full-stroke testing interval required to achieve the required SIL performance.

Bibliography

ANSI/ISA-ISA 84.00.01-2004 (IEC 61511), "Functional Safety: Safety Instrumented Systems for the Process Industry Sector," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 2004.

IEC 61508, "Functional Safety: Electrical/Electronic/Programmable Electronic Safety-Related Systems," Geneva, Switzerland: International Electrotechnical Commission, 1999.

"Guidelines for Process Equipment Reliability Data," Center for Chemical Process Safety, New York: American Institute of Chemical Engineers, 1989.

IEC 61511, "Functional Safety: Safety Instrumented Systems for the Process Industry Sector," Geneva, Switzerland: IEC, 2003.

"OREDA: Offshore Reliability Data Handbook," 4th edition, Norway: Det Norske Veritas Industri Norge as DNV Technica, 2002.

ISA TR84.00.02, "Safety Instrumented Functions (SIF)—Safety Integrity Level (SIL) Evaluation Techniques, Part 1: Introduction," Instrumentation, Systems, and Automation Society, 2002.

Summers, A. E., "Partial-Stroke Testing of Block Valves," *Control Engineering,* November 2000.

Summers, A. E. and Zachary, B. A., "Partial-Stroke Testing of Block Valves," ISA Expo 2002, Houston, TX, Instrumentation, Systems, and Automation Society, 2002.

Summers, A. E. and Zachary, B. A., "Improve Facility SIS Performance and Reliability," *Hydrocarbon Processing,* October 2002.

6.11 Fieldbus and Smart Valves

J. BERGE (2005)

Partial List of Suppliers: *(Intelligent Positioners):*	ABB (www.abb.com/) Biffi (www.tycovalves.com/) Dresser (www.dresser.com/) Emerson (www.emersonprocess.com/) Flowserve (www.flowserve.com) Foxboro (www.foxboro.com/) Metso (www.metso.com/) Samson (www.samson.de/) Siemens (www.siemens.com/) SMAR (www.smar.com/) Tyco (www.tycovalves.com/) Westlock (www.westlockcontrols.com/) Yamatake (www.yamatake.com/) Yokogawa (www.yokogawa.com/)
Electric Actuator:	EIM Controls (www.eim-co.com/) Emerson (www.emersonprocess.com/) Rotork (www.rotork.com/) Flowserve (www.flowserve.com) Ledeen (www.dresser.com/)
Discrete Valve Coupler:	Bürkert (www.buerkert.de/) Emerson (www.emersonprocess.com/) Flowserve (www.flowserve.com) Pepperl+Fuchs (www.pepperl-fuchs.com/) StoneL (www.stonel.com/) SMAR (www.smar.com/) TopWorx (www.topworx.com/)

INTRODUCTION

The impact of fieldbus and smart technology on the design of control valves is profound. It changes how valves are installed, commissioned, set up, maintained, and operated. Ultimately, many benefits and savings can be achieved from using digital communication for valves.

The brains of an intelligent control valve are in the digital valve positioner. An intelligent control valve is therefore primarily an issue of using digital valve positioners. For a basic introduction on positioners please refer to Section 6.2, for a discussion of intelligent valves and positioners see Section 6.12, and for a discussion of valve diagnostics and predictive maintenance, refer to Section 6.8.

The brains of an on/off valve are in the discrete valve coupler (DVC). Therefore, using an intelligent on/off valve is an issue of using a discrete valve coupler. A Foundation™ Fieldbus or PROFIBUS-PA valve receives the command signal over the digital data highway or network.

A smart control valve receives the signal via 4–20 mA but also provides digital communication for auxiliary functions. Actually, in the case of Foundation Fieldbus more often the control is done in the positioner itself, and the positioner receives the process variable from the transmitter over the fieldbus.

Valve instrumentation based on HART, Foundation Fieldbus, and PROFIBUS-PA includes

Control valve positioner
Valve coupler
Electrical actuator
Fieldbus to current converter

Fieldbus to pneumatic signal converter
Fieldbus remote I/O
Position transmitter

BENEFITS AND SAVINGS

For valve users who spend too much effort and resources on maintenance, and see too much variability in product quality, network-enabled valves can significantly reduce the cost of operation. Unlike traditional hardwired valves, networked valve packages have microprocessors, which also allow for digital communication.

Networking of valves and microprocessors has enabled several technological breakthroughs that in turn can result in savings. This is because all information can be accessed remotely and because all functions are performed in device firmware instead of hardware: flow characterization, soft limit switch, feedback, and automatic setup.

HART, Foundation Fieldbus, and PROFIBUS-PA

In Foundation Fieldbus-based systems all valves have a permanent network connection to the engineering tool and to the asset management tool (Figure 6.11a). The digital communication works across the same single pair of wires that is used for the desired valve position. "Always on" communications means that interrogation for monitoring, parameterization, and diagnostics can conveniently be done any time.

Unlike in the past when smart valves, using HART communication, used temporarily connected handhelds, smart valves should have permanent communication with the control system in order to leverage the smart capabilities. A permanent communication link between valves and software enables on-line asset management. Operators can get on-line feedback of the control valve's true position and can monitor all critical parameters in real time without having to go into the field.

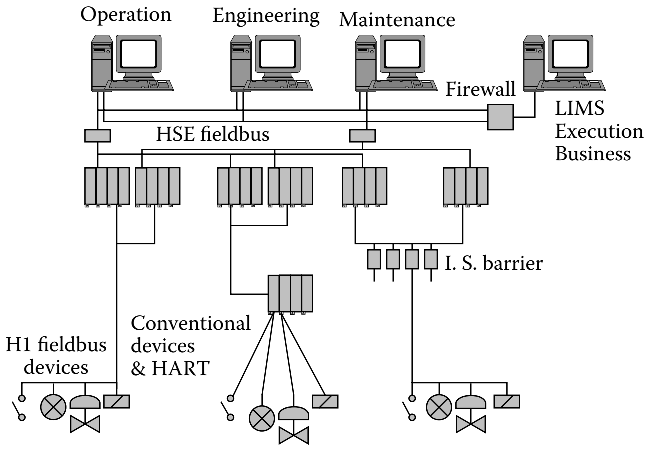


FIG. 6.11a
Illustration of a networked system architecture, which is "always on."

A fair amount of data can be accessed also from a smart HART-based valve positioner. However, fieldbus with its higher speed and standardized data formats enables positioners to provide more diagnostics, and faster. Fieldbus-based positioners therefore can bring more diagnostics to the user than their smart counterparts.

Digital valves can be configured remotely across the bus from a handheld or software tool. This makes it possible to adjust tuning, without having to leave the control room. A single type of digital valve positioner or discrete coupler can be configured for all control and on/off valves, respectively.

Self-Diagnostics Digital valves have internal self-diagnostics that can be interrogated remotely across the bus from a handheld or software tool without having to go into the field. Diagnostics to this level of detail was never possible before the introduction of microprocessors and digital communication. More sophisticated valve diagnostics can be done by using on-line asset management or by utilizing proprietary software tools using information tapped from the positioner.

In the case of Foundation Fieldbus the digital valve positioner or coupler can even issue alerts when failure is detected. The user can, from a single workstation, access all the positioners and other instruments using the engineering tools. The health of any device in the plant can thus be checked easily without having to venture into the field. In the past, valve diagnostics software was connected on a single unit basis. Diagnostics was performed after the valve had already failed, "post-mortem."

Additionally, digital valves internally collect operational statistics such as total travel and the number of reversals. The operational statistics are used for condition-based maintenance schemes.

Interoperability Digital valves can be monitored remotely to see variables such as actuator pressure or torque, and to feed back actual valve position. In the case of Foundation Fieldbus and PROFIBUS-PA, the actual position feedback provided by the digital valve positioner or coupler is fast enough to be used as part of the control loop. The feedback signal travels digitally on the same wire as the desired valve position from the controller or logic, i.e., without additional wiring or external proximity switches.

The HART, Foundation Fieldbus, and PROFIBUS-PA protocols are all developed and managed by independent organizations formed by several manufacturers. This means that many valve products, control systems, and other tools that communicate by using the same protocols are therefore compatible with them. It is possible to select valve positioners, couplers, and tools independent of the control system.

All three protocols support electronics device description (DD, a.k.a. EDDL) in some form. The DD files are loaded in the configuration tool, permitting the tool to communicate with the digital valve and make all its special features accessible.

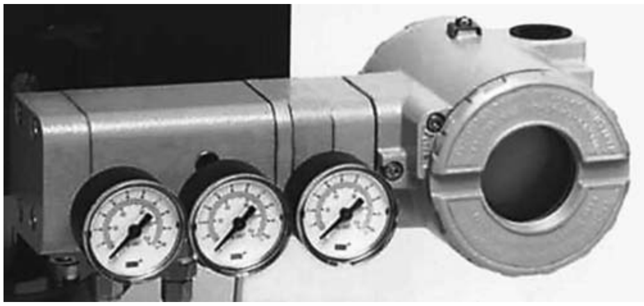


FIG. 6.11b
Digital valve positioner. (Courtesy of SMAR.)

DD for Foundation Fieldbus is very well supported in many control systems, while the DD for HART and PROFIBUS-PA is not so common. Thus better interoperability is achieved with Foundation Fieldbus, as the full set of information can be accessed from virtually any fieldbus system.

PROFIBUS-PA is also taking two other approaches to interoperability; one is device profile, which defines a common basic set of functionality providing some degree of interoperability, and the other is FDT/DTM (field device tool/device type manager). Although FDT/DTM is also not widely supported in control systems, it does have some potential advantages when it comes to sophisticated valve diagnostics. Future enhanced versions of DD may also address this capability.

Control Algorithm in Positioner A defining characteristic of Foundation Fieldbus is that it also is a standard function block diagram programming language to build the control strategy. In a modern network-based control system architecture control is often done in the field instrument (Figure 6.11b), and therefore it is possible to reduce or eliminate the number of control cards usually found in a DCS or PLC system.

Fieldbus positioners perform control in the field by executing the PID control function block. Other blocks perform functions such as split-range, flow characterization, and the positioning servo itself. It is possible to adopt decentralized control using Foundation Fieldbus programming language instead of central controllers and proprietary languages.

The failure of the electronics in the valve affects only a single loop. Diagnostics in the positioner and coupler detect loss of signal from the controller and subsequently bring the valve to its fail-safe position.

Foundation Fieldbus and PROFIBUS-PA completely do away with 4–20 mA. Valves, transmitters, and other devices are multidropped on a network. Typically up to 16 devices are connected to the same pair of wires. Theoretically it is also possible to multidrop HART devices, but the communication speed is too slow for valve applications.

Wiring Costs Using digital valves, the total installed cost can be reduced, achieving savings before the valve is even put into operation. Digital communications reduce installation cost in new construction and also in revamps.

Networked control valve positioners require less wiring than their hardwired counterparts. For example, if actual position feedback is required from an analog positioner, it is necessary to run a second pair of wires back into the control room, where the DCS or PLC will also require an analog input point (on the left side of Figure 6.11c).

For Foundation Fieldbus and PROFIBUS-PA no additional wiring is required because the feedback value is communicated on the bus together with the other values (at the center of Figure 6.11c). HART positioners also do not require an additional wire to run back into the control room. If the control system does not have native support for HART, it is possible to add a HART-to-current converter in the control room instead (Figure 6.11c right).

The converter connects to the same two wires used for the positioner to receive the desired valve position signal from the DCS or PLC, and then it uses HART to poll the positioner to get the actual position digitally (Figure 6.11d).

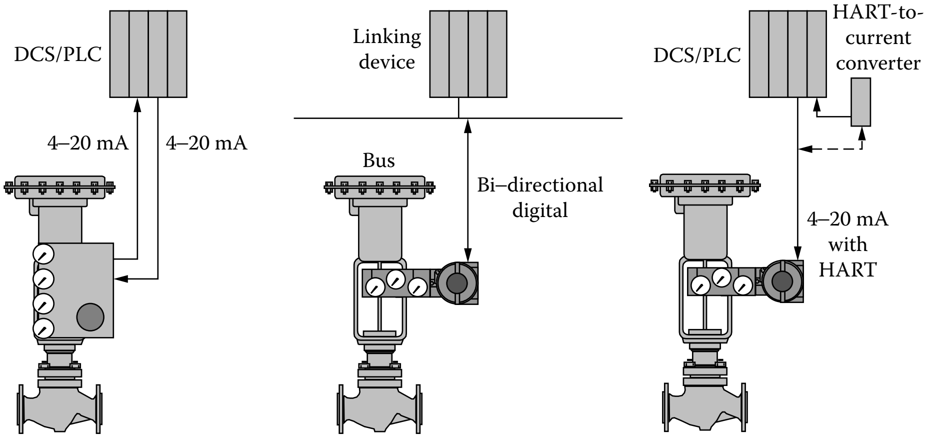


FIG. 6.11c
Wiring requirements of providing position feedback in analog (left), Fieldbus (center), and HART (right) systems.

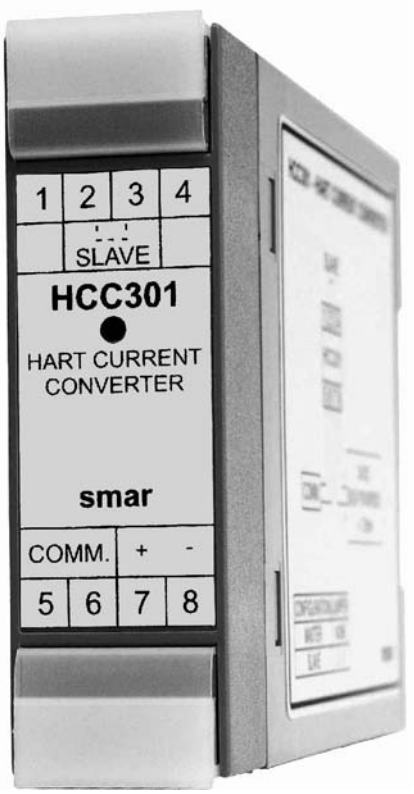


FIG. 6.11d
HART-to-current converter. (Courtesy of SMAR.)

The converter generates an analog signal accepted by the analog input card in any control system or single loop controller. Additional safety barriers are also eliminated. If discrete open/close statuses are required, these are simply detected as alarms in the control system using the actual position signal.

Similarly, networked discrete valve couplers require less wiring than their hardwired counterparts. For example, if open/close statuses are required from a hardwired valve, it is necessary to run two additional pairs of wires back into the control room, where the DCS or PLC will also require two discrete input points. However, the discrete valve coupler provides it on the same bus (Figure 6.11e).

An electric actuator may be even more extreme because the need for signals such as open/stop/close or desired valve position, shut-down, open/close statuses or actual position, local/remote, opening/closing, torque switch or torque value, and failures is not uncommon (Figure 6.11f). Using a bus, a lot of hardwire and I/O modules are eliminated.

In this case, using bus technology, it is possible to reduce by anywhere from two to ten pairs of wires per valve even if one is only considering the auxiliary functions. Needless to say, the savings in cable and controller I/O cards and all the associated costs of engineering and installation is significant.

Both Foundation Fieldbus and PROFIBUS-PA permit several devices to be connected on the same pair of wires together with other devices. Moreover, an interface module can handle as many as 64 devices, thus eliminating the need for a large number of conventional I/O modules. Significant savings in wiring and installation costs are therefore possible as compared to hardwired valves. In intrinsically safe installations the savings are even greater, because as many as four or eight devices share the same single safety barrier, depending on the scheme used. This too is a significant cost saver.

To really benefit from Foundation Fieldbus and PROFIBUS-PA multidrop wiring, the valve electronics shall draw as little power from the bus as possible. Pneumatic valve positioners that use piezo technology to drive the flapper have significantly lower power consumption than the solenoid type. Lower current draw means lower voltage drop along the wires, which translates into more devices, longer wires, and fewer barriers. Therefore, by using lower power consumption positioners, such as 12 mA, further savings are possible.

Positioners, couplers, and actuators based on HART, Foundation Fieldbus, and PROFIBUS-PA are available from multiple competing vendors. The competition among interoperable products brings prices down lower than for units using proprietary communication.

Valve Calibration and Configuration

Configuring and calibrating valves in the past required tedious mechanical adjustments. Using handheld terminal or computer software to access the valve, these tasks require less time and resources and thus result in savings (Figure 6.11g). For more detail on the subject of configuring intelligent devices, refer to Section 1.6 in Volume 1 of this handbook.

Valve instrumentation without microprocessor and communication requires mechanical and electronics adjustments at the valve in order to change the way it operates. For example, the modification of flow characteristics previously had to be done by changing mechanical cams and springs. Using digital communication it can now be done remotely through software, thereby significantly reducing cost and time, particularly during commissioning.

Other valve parameters, such as the opening and closing time of the valve, can also be set.

To align mechanical feedback systems based on leavers, cams, and potentiometers in older positioners was tedious and time consuming, especially in the field. Using digital communication, a setup command can simply be sent from a handheld terminal or software in the control system to auto-calibrate the positioner. This significantly shortens the setup time, as stroking need not be done manually, achieving reduced installation cost. Once the setup is initiated the valve automatically finds its own fully opened and fully closed positions.

A key advantage of Foundation Fieldbus is that control is also done in the field devices, typically in the valve. The decentralized architecture where control can be done in the field instruments is called Field Control System (FCS). Because control is done in the valve the number of loops in centralized controllers is reduced, and the number of

FIG. 6.11e
Discrete valve coupler incorporates solenoid and position feedback. (Courtesy of TopWorx.)

centralized controls can subsequently also be reduced. The savings achieved from the reduction in very expensive controllers are tremendous.

To enjoy maximum flexibility, the positioner can have dynamically instantiable function blocks. This allows the user to freely select from more than a dozen blocks in a library containing scores of blocks including PID, splitter, and arithmetic.

Valve Cycling and Stiction An inadequately functioning control valve will still degrade the overall loop performance and may cause the valve to wear out prematurely. Conversely, a healthy positioner ensures better control and less maintenance. The greatest savings potential for applying networked digital valves is in long-term operational costs such as maintenance and improved performance.

The push for higher valve performance has been seen particularly in the pulp and paper industries, because of their need for increasing product uniformity. Digital valve solutions have higher performance than their analog counterparts, resulting in lower process variability, which in turn translates into higher quality, less rejects, higher yield and productivity, and lower raw material and fuel consumption. Information from the digital valve may be used to fine-tune the loop for optimum production output and tighter product uniformity. The higher valve performance ultimately results in savings because of the resulting boost in production and yield.

Because aligning and calibrating analog positioners is tedious, this is often not done properly or not checked regularly. Automatic setup invoked through the fieldbus network makes sure it is done accurately, every time, so that positioning is precise. It is possible to remotely calibrate the positioner when conditions allow, e.g., while the process is down for any reason. As the valve seat and other parts wear and tear during operation, one can calibrate the valve to make sure that it operates and shuts off properly.

It is commonly seen that control valve cycling can cause the control loops to cycle and that this can destabilize production. Studies show that as many as two out of three control loops are oscillating due to the dead band of control valves. The operational statistics such as total valve travel and number of cycles that now are provided by the positioners are extremely helpful in detecting oscillation problems and enable one to eliminate them (Figure 6.11h).

FIG. 6.11f
Fieldbus electrical actuator. (Courtesy of Rotork.)

One can eliminate the sources of variability one by one. A hunting valve is easy to catch because one can see a great number of reversals counted in the positioner or coupler and communicated over the fieldbus. If this occurs, one knows that either the positioner or loop is poorly tuned. The servo gains and other parameters like the time required for opening and closing the valve can be set from a handheld or by a software tool to optimize positioner response to different valve sizes.

Please, press buttons to exec.:

Time_To_Close: .000 s
Time_To_Open: .000 s
Servo Kp: .000
ServoTr: .000 min/r

FIG. 6.11g
The display of a handheld tool for configuring a digital valve positioner. (Courtesy of SMAR; the screenshot is of SMAR HPC301.)

For good control loop performance a control valve must be able to respond to small steps in the control signal by matching it accurately with its stem within 1% or less. If not, a problem called “backlash” or “stiction” will occur. It is typically caused by packing friction (Section 6.5), which in turn may be a result of wear and tear.

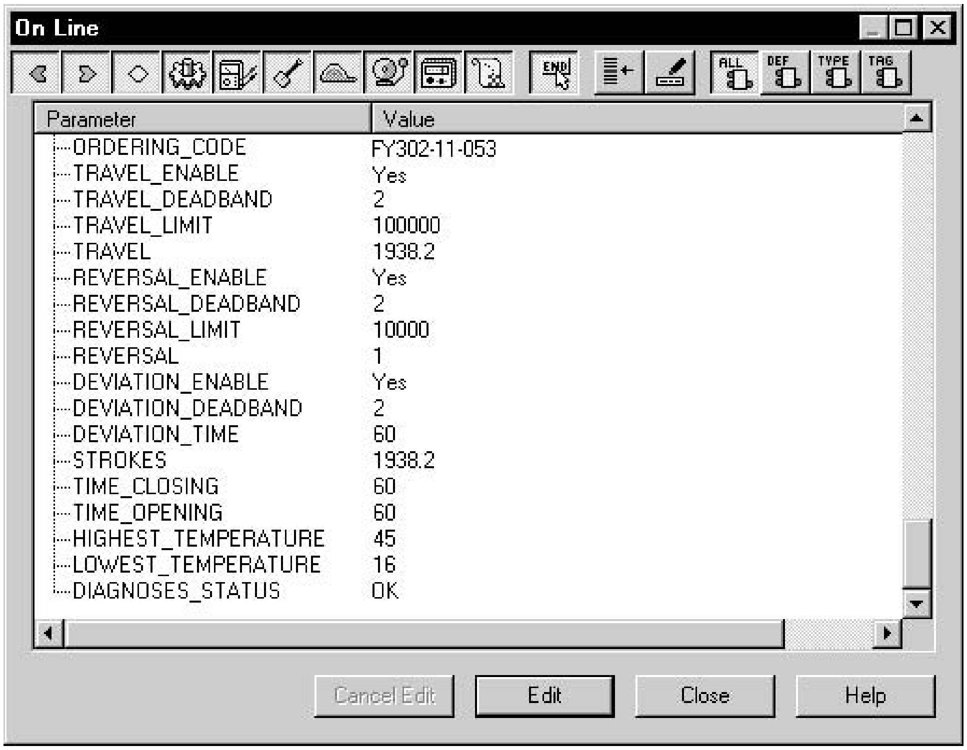
In addition to being able to predict wear based on operational statistics, software is now available to analyze the valve’s performance thanks to the improved communications capabilities to obtain data for such plots. The operational statistics may be used to more accurately predict the packing wear, as to strike the right balance of downtime due to replacement, and greater performance as a result of more frequent packing change.

Position Feedback Fieldbus positioners provide an actual position feedback value over the communication network that is fast enough to be used as part of advanced control strategies and also for true bumpless transfer and true reset windup protection based on actual valve excursion end points. Because the valve position is available over the communication network, the actual position can be put into use at a later stage, without having to modify the valve positioner, running wires, or adding any AI points to the control system, as was the case in the days of analog feedback.

Foundation Fieldbus thus makes actual position feedback possible for every control loop. This makes it feasible to provide true bumpless transfer back to the automatic control mode after the valve has been switched to manual control and was operated locally. Similarly, true integral windup protection can be provided when the valve has reached its endpoints or is unable to move. In the past this was done for some loops in conventional systems, but was not realistic for all. In fieldbus systems, this feature has no extra cost and therefore it is viable to optimize every loop and obtain the corresponding savings.

Lastly, systems and devices using Foundation Fieldbus or PROFIBUS-PA eliminate the need for D/A and A/D conversions in the transmission path, as no analog signals are used. This results in higher accuracy. Because the actual valve position is transmitted back to the workstation using digital communication, inaccuracies caused by conversion to analog are eliminated. Additional points of calibration are also eliminated.

Proactive or Predictive Maintenance For further details on valve diagnostics and predictive maintenance, refer to Section 6.8. Self-diagnostics done in the digital valves as well as using other information make it is possible to move to a proactive condition-based maintenance scheme that can lower maintenance expenses as compared to reactive or preventive maintenance schemes.



Parameter	Value
ORDERING_CODE	FY302-11-053
TRAVEL_ENABLE	Yes
TRAVEL_DEADBAND	2
TRAVEL_LIMIT	100000
TRAVEL	1938.2
REVERSAL_ENABLE	Yes
REVERSAL_DEADBAND	2
REVERSAL_LIMIT	10000
REVERSAL	1
DEVIATION_ENABLE	Yes
DEVIATION_DEADBAND	2
DEVIATION_TIME	60
STROKES	1938.2
TIME_CLOSING	60
TIME_OPENING	60
HIGHEST_TEMPERATURE	45
LOWEST_TEMPERATURE	16
DIAGNOSES_STATUS	OK

FIG. 6.11h
Operational statistics, such as listed here, is easily reviewed and can trigger alarms notifying the operators that the valve is due for service. (Screenshot is courtesy of SMAR Syscon.)

Asset management software can continuously poll digital valves over the network to find out their health immediately. This means that failed valves can be found faster and problems pinpointed with greater accuracy, making repair quicker and easier. Fewer resources spent means savings. For example, if the supply air to the positioner is lost, the operator is immediately notified.

The operational statistics such as total valve travel and number of cycles that now are provided by the positioners are leading indicators of wear and tear. The operational statistics are used for proactive maintenance to determine the optimum time for valve overhaul. Based on valve manufacturer recommendation for stem packing change, alarm limits for these statistics can be set, notifying the technicians that maintenance is due. Spares can be ordered in advance, several valves can be fixed all in one shutdown, and maintenance can be done before the valve fails. Fewer unscheduled shutdowns mean higher productivity. Major maintenance savings are possible.

In the past, many valves were removed, brought to the workshop, and torn down only to find nothing was wrong. Remote diagnostics reduce such unnecessary work by allowing the technicians to check the general health of the valve from the host before even going into the field. Valve instrumentation can be managed without having to leave the workstation. The cost that can be saved by not having to bring one valve into the workshop is substantial. Users can instead target their maintenance resources to the valves that really need attention.

It primarily is the asset management software that allows technicians to find out what is really wrong, and based on the valve condition it is possible to schedule maintenance only when really necessary, prioritizing the most immediate needs. Rather than running until essential equipment fails, maintenance can be carried out when needed, and only when needed. Maintenance efforts can be focused on problem valves, not the healthy ones.

Spare Parts and Flexibility The interoperability of HART, Foundation Fieldbus, and PROFIBUS-PA means that plants have many alternative third parties from which to obtain their replacement positioners and couplers. This ensures that spare prices will be market based, and not inflated, as was the case when proprietary protocols were used.

To achieve maximum flexibility and enable the same positioner or coupler to be used with any pneumatic control valve or actuator in a plant, more and more functions are performed in software as opposed to hardware, making it easier to adjust and adapt to the application. Users can standardize on one positioner or discrete valve coupler for all different kinds of control valves, keeping spares and training to a minimum.

Safety and Pollution

Apart from the direct economical benefits, digital technology can also enhance safety and protect the environment. An important part about the actual position interlocks in Foundation Fieldbus is that they are handled by the PID and AO blocks, without the user having to configure separate logic for interlocks and reset windup protection on actual position, as was done in old DCSs.

This is different from the old days, when some valves were connected to the DCS using 4–20 mA analog signals from the controller to the positioners. The DCS had no way of knowing the actual position of the valve. If the valve failed, nobody knew about it because there was no feedback or diagnostics. If somebody operated the valve by hand, nobody knew. This was very dangerous, because the operator was under the impression that automatic control was in operation, when it was not. Under these conditions, the controller can change its output, but the valve does not change, so the DCS still attempted to control even though there was no actuation.

Also, if the output of the automatic controller was 50% (12 mA), while some technician in the field has been operating the valve by hand and opened it to 75%, when the loop was switched back to the automatic controller, the valve has to jump 25% (from 75 to 50%). To avoid this, some critical loops had a 4–20 mA feedback of actual valve position from the positioner to the DCS AI module. In the DCS, then, one had to program logic to detect the deviation and in this case put the PID in manual and make sure the PID output follows the actual position feedback from the valve. That is very complex logic, but for Foundation Fieldbus valves it is all automatically taken care of.

Failure Detection Diagnostics from networked positioners and couplers detect failure in positioner, actuator, and valve including loss of supply air. Foundation Fieldbus communication will inform the operator of such events and will stop control. This allows for corrective action to be taken sooner, which might protect from dangerous situations and production loss. Thereby, Foundation Fieldbus makes systems safer.

Another example is a thermocouple failure detected by a temperature transmitter propagated to the positioner, which in response shuts the loop down to a predetermined safe state without the need for any central control action.

Another important form of diagnostics is partial stroke testing (Section 6.10). It verifies that the valve is not stuck, increasing the probability of a successful movement if called upon. This reduces the problem of valves stuck in one position.

Digital communication in conjunction with asset management software can ensure that the installed base of digital valves is maintained well and is experiencing fewer surprise failures. Thus, asset management enables plants to run uninterrupted for longer periods of time, subsequently increasing productivity, keeping costs down. Diagnostics provide an early warning for abnormal conditions and may be used as an indication of problems to come, allowing technicians to solve problems before they adversely affect the process. Users can switch to a proactive maintenance program, thereby minimizing plant downtime.

On-line Plant Asset Management

An inadequately functioning control valve will upset the overall loop performance or may cause the valve to wear out prematurely. Conversely, a healthy positioner ensures better control and less maintenance. A fieldbus control valve positioner takes a total life cycle view of the valve and is designed to enable a longer life for the valve package.

Positioners now come with built-in pressure sensors continuously monitoring the pressure at both the air supply input and the actuator chambers. This enhances simple text-based diagnostics such as loss of supply air with sophisticated chart- and graph-based analysis, such as valve signature. Thus, the valve is provided with additional on-board sensors, which determine its condition, the ambient conditions, and such external factors as the loss of supply air, and gives this information to the microprocessor.

The diagnostics is communicated to the on-line plant asset management tool using HART, Foundation Fieldbus, or PROFIBUS-PA communication. Asset management software takes the vast amount of diagnostics data in the positioner and turns it into information, which is useful to maintenance technicians. Together the asset management software and the positioners make it is easier to determine valve health and estimate their remaining life spans (Figure 6.11i).

Valve Signature The Valve Signature plot traces the actuator pressure required to put the valve in a desired position (Figure 6.11j). As was shown in Section 6.8, comparing changes in this behavior over time helps identify problem areas. Other charts include Hysteresis (a.k.a. “Positioner Signature” or “Dynamic Error Band”), step response (Section 6.9), and drive signal. It is a good idea to capture a “base line” signature for the valve package as it is new, and compare future signatures against this benchmark (Figures 6.8b and 6.10j).

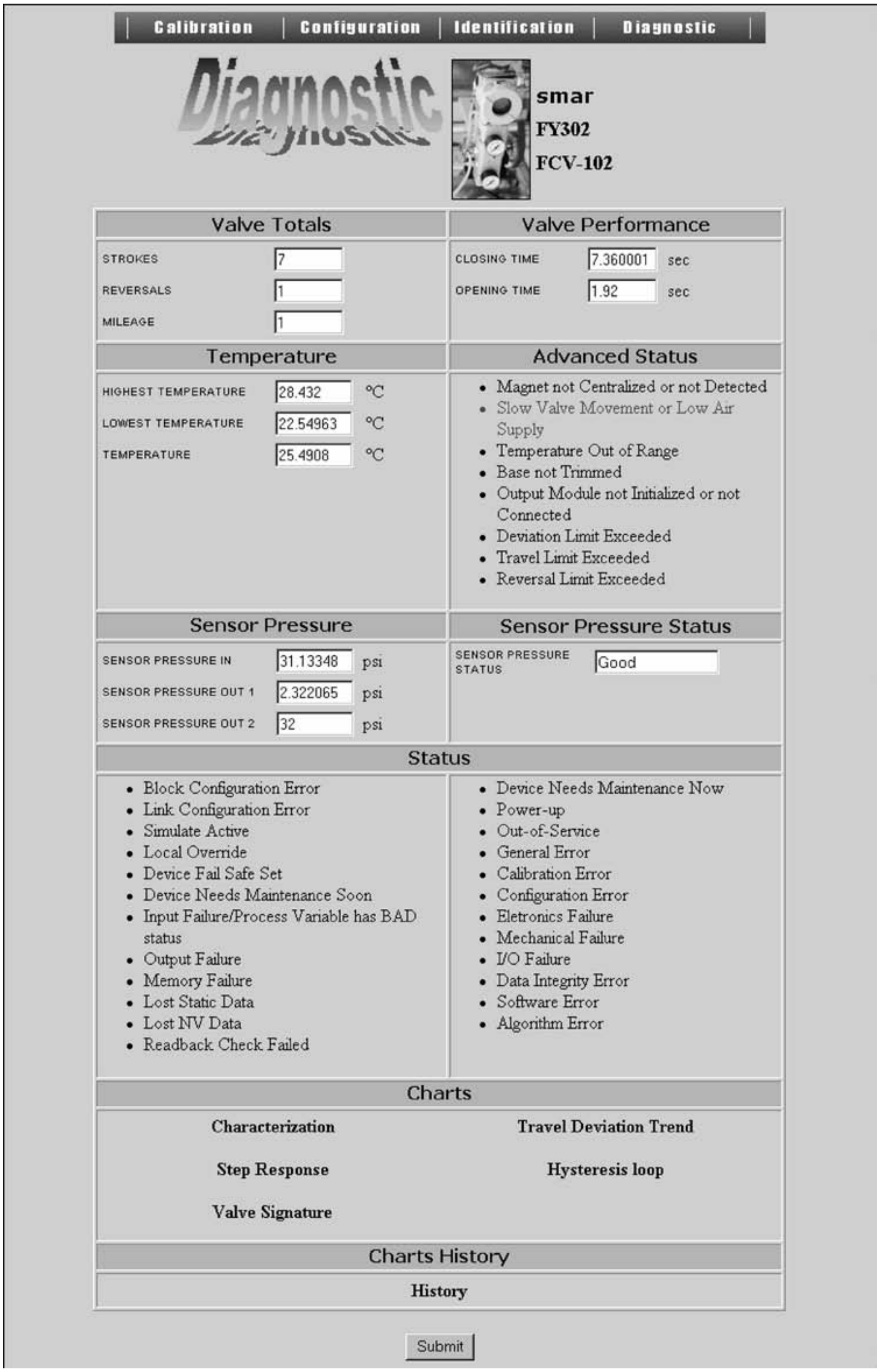
By using the above described techniques, maintenance can be done before the valves fail.

DIGITAL VALVE INSTRUMENTATION

The first generation of digital valve instrumentation was an electropneumatic device enhanced with a CPU and HART communication. Few other things changed. These devices still relied on the old lever and potentiometer for the actual valve position sensing for the feedback. They had the same old pneumatic gauges. High-current-consumption solenoids were used for flapper actuation. There was not much diagnostics to speak of. However, in the second generation the technology went beyond the mere incorporation of a CPU.

Second Generation

The microprocessor is at the center of every intelligent valve (Figure 6.11k). The firmware in digital valve instrumentation



smar
FY302
FCV-102

Valve Totals

STROKES	7
REVERSALS	1
MILEAGE	1

Valve Performance

CLOSING TIME	7.360001	sec
OPENING TIME	1.92	sec

Temperature

HIGHEST TEMPERATURE	28.432	°C
LOWEST TEMPERATURE	22.54963	°C
TEMPERATURE	25.4908	°C

Advanced Status

- Magnet not Centralized or not Detected
- Slow Valve Movement or Low Air Supply
- Temperature Out of Range
- Base not Trimmed
- Output Module not Initialized or not Connected
- Deviation Limit Exceeded
- Travel Limit Exceeded
- Reversal Limit Exceeded

Sensor Pressure

SENSOR PRESSURE IN	31.13348	psi
SENSOR PRESSURE OUT 1	2.322065	psi
SENSOR PRESSURE OUT 2	32	psi

Sensor Pressure Status

SENSOR PRESSURE STATUS: Good

Status

- Block Configuration Error
- Link Configuration Error
- Simulate Active
- Local Override
- Device Fail Safe Set
- Device Needs Maintenance Soon
- Input Failure/Process Variable has BAD status
- Output Failure
- Memory Failure
- Lost Static Data
- Lost NV Data
- Readback Check Failed
- Device Needs Maintenance Now
- Power-up
- Out-of-Service
- General Error
- Calibration Error
- Configuration Error
- Eletronics Failure
- Mechanical Failure
- I/O Failure
- Data Integrity Error
- Software Error
- Algorithm Error

Charts

Characterization
Travel Deviation Trend
Step Response
Hysteresis loop
Valve Signature

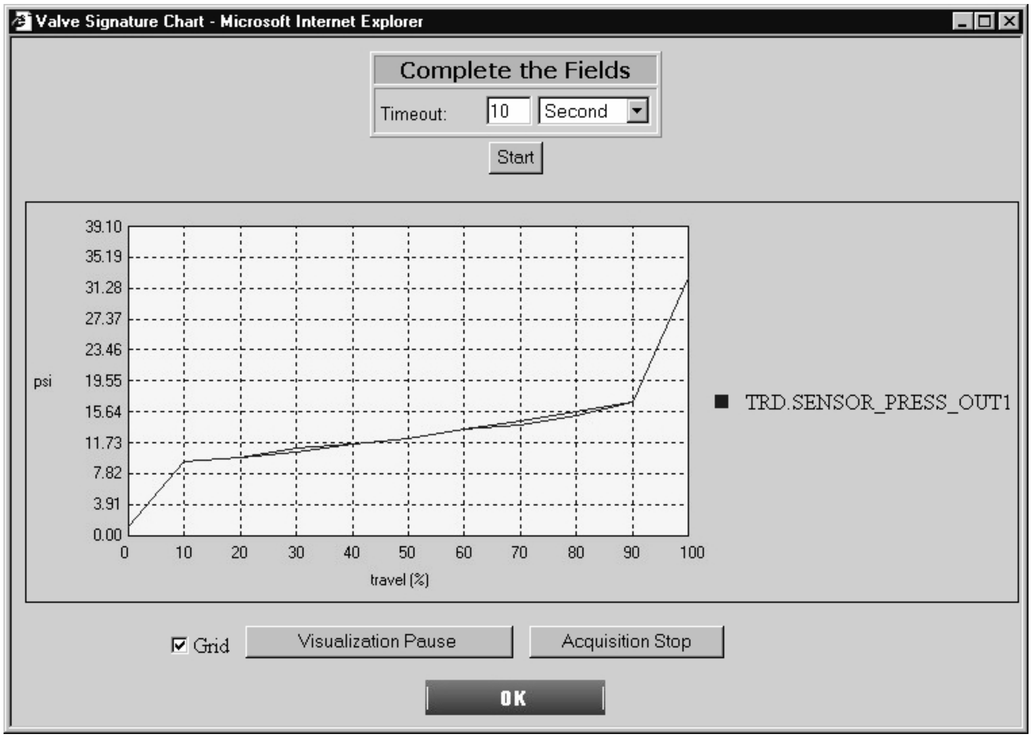
Charts History

History

FIG. 6.11i
Fieldbus positioner diagnostics on asset management software. (Screenshot is courtesy of SMAR AssetView.)

completely controls the valve. For example, instead of cam and spring mechanics on the feedback to apply flow characterization, these instruments do the characterization in software. Error due to mechanical imprecision is avoided, further improving performance. Likewise, split-range operation, digital display, and local operation are all provided using software.

Position sensing is a key aspect of valve control and is an important means to acquire the data for sophisticated diagnostics and to tap the full potential of digital instrumentation.



Timeout: 10 Second

☑ Grid

FIG. 6.11j
A valve signature. (Screenshot is courtesy of SMAR AssetView.)

See Section 7. 10 in Volume 1 of this handbook ("Linear and Angular Position Detection") for details on position sensing. Noncontact position sensing technologies such as Hall-effect eliminates linkages that can stick and gall; they are also less sensitive to vibration, thus improving hysteresis and reliability. Accurate position sensing allows the valve to be able to respond to very small demands in position change as required for tight control.

Power consumption is a key issue for digital valve instrumentation, and it is their only disadvantage as compared to analog instrumentation. For example, a traditional electropneumatic positioner has a voltage drop of only a few volts,

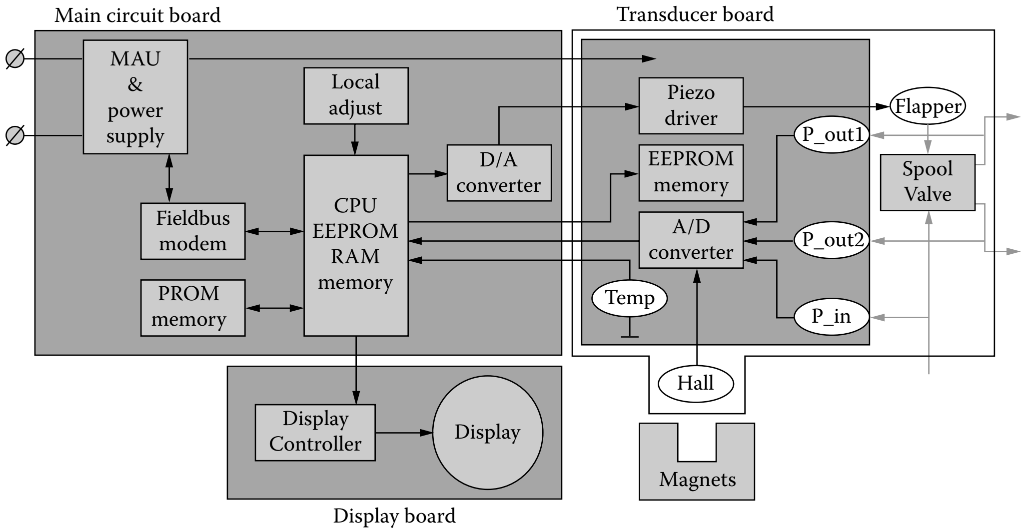


FIG. 6.11k
Block diagram of a Fieldbus pneumatic control valve positioner. (Courtesy of SMAR.)

permitting the positioner to be driven from a regular controller output over very long wires. It is even possible to put two analog positioners in a split-range scheme in series without concern.

However, smart positioners using HART typically have a voltage drop of 11 V or more and therefore Ohm's Law is putting a restriction on cable length, as controller open loop voltage may not be sufficient to drive the signal through the positioners plus cable resistance. Split-range applications generally require signal amplifiers. Current consumption is a concern in fieldbus installations where several devices are connected on the same network, causing voltage drops and restricting the number of devices on a safety barrier. Valve instrumentation that has I/P sections based on piezo technology typically have lower power consumption, making this less of a consideration.

Additional auxiliary sensors can be included in digital valve instrumentation in order to provide inputs for more advanced diagnostics. For example, they might include internal ambient temperature sensing as well as built-in pressure sensors to continuously monitor the pressure at the air supply input and at the two pneumatic outputs. The on-board sensors are dedicated to providing continuous information to the microprocessor, so that it can determine the condition of the valve assembly and ambient conditions.

CONCLUSIONS

The market for digital valve positioners in general, and for fieldbus in particular, looks very promising and is perhaps the single most important event that has occurred in the control valve industry during the past several decades.

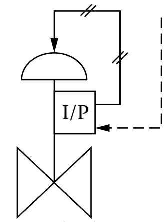
Standards and flexible software have made it possible for users to choose third-party positioners when the positioners being offered by the control valve manufacturers cannot meet their performance requirement. With the right positioners it is very easy to upgrade an existing control valve to make it smart or fieldbus-compatible. Even when an old DCS system is converted into a fieldbus system, all the user needs to do is to remove the old 4–20 mA positioner and place a Fieldbus-compatible positioner in its place.

References

1. Berge, J., *Fieldbuses for Process Control Engineering, Operation and Maintenance,* Research Triangle Park, NC: The Instrumentation, Systems and Automation Society (ISA), 2002.
2. Merritt, R., "Turned-On Valves: They're Using Embedded Intelligence and Digital Communication to Perform Data Acquisition, Diagnostics, and Control," *Control Magazine,* February 9, 2001.

6.12 Intelligent Valves, Positioners, Accessories

H. D. BAUMANN (2005)



Control valve
with
electropneumatic
positioner

Flow sheet symbol

Materials of Construction:	Housing and cover: Low copper aluminum (stainless steel optional)
Supply Pressure:	0.3 bar (5 PSIG) min., 10 bar (145 PSIG) max.
Ambient Temperature:	–40 to 85°C (–40 to 257°F).
Inaccuracy:	Min. dead band: 0.2% of span. Linearity: ±0.75% of span
Positioner Input Signals:	Analog: 4–20 mA, Digital: HART protocol or Fieldbus min. input current: 3.5 mA, load: 300 Ohms
Costs:	*$750.00 to $1,250.00 depending on functionality*
Partial List of Suppliers:	ABB Automation Products (Germany) (www.abb.com/industrialit) ARCA-Regler (Germany) (www.arca-valve.com) Dresser Valve Division (Masoneilan) (www.masoneilan.com) Emerson Process Management (Fisher) (www.Fisher.com) Flowserve (Valtec) (www.invensysflowcontrol.com) Foxboro/Invensys (www.foxboro.com) Metso Automation (Neles) (www.metsoautomation.com) Samson A.G. (Germany) (www.samson.de) Siemens A.G. (Germany) (www.sea.siemens.com/ia) Yamataki (Japan) (www.paxton@paxtoncorp.com) Young Tech Co. (Korea) (www.ytc.co.kr) Yokogawa (Japan) (www.yokogawa.co.jp)

INTRODUCTION

The reader is advised that various aspects of valve and positioner intelligence have also been discussed in the previous four sections, where control valve diagnostics (Section 6.8), dynamics (Section 6.9), safety (Section 6.10), and fieldbus interaction (Section 6.11) have been covered.

Intelligent positioners are primarily used to control the position of the control valve stem in relation to the process control signal. They are also known as “smart” or “digital” positioners. In case of most designs, the “digital” designation is somewhat of a misnomer, because in most cases, it is only the signal processing portion of the positioner that is digital, while most of the input signals are still analog (4–20 mA) and most output signals are either electrical or pneumatic.

Digital positioners arrived on the scene in the mid 1980s and thereafter rapidly replaced the former electronic, i.e., analog, positioners. Similarly to their analog predecessors, these intelligent devices also use air as their source of actuating power, and because of their enhanced electronics, they are about one third more expensive than their predecessors.

Advantages of Intelligent Positioners

Most intelligent positioners can be adapted to both sliding and rotary stem valves. On existing control valves, if built in conformance with IEC Standard 60534-4-1, these intelligent positioners can replace pneumatic or analog electronic positioners of the same supplier. Conformance to a new

FIG. 6.12a
Typical digital positioner with explosionproof housing meeting FM and CSA approval requirements. (Courtesy of the Dresser Valve Group, Inc. Masoneilan.)

German VDI/VDE Standard 3847 would even guarantee interchangeability between positioners made by different manufacturers.

The control signals into the smart positioners can be analog (4–20 mA) or digital (via bus systems), among which Foundation Fieldbus® predominates in the United States, while Prifibus® is more commonly used in Europe.

Similarly to electronic analog positioners, the digital units are also available in "intrinsic safe" designs or with explosionproof housings (see Figure 6.12a) and are protected from EMC interference. They typically meet U.S./CSA as well as CENELEC (European) and CESI (Japanese) standards.

In Figure 6.12a, one should note the presence of the explosionproof pushbuttons, which, in conjunction with the display window, enable the operator to perform on-site calibration and configuration. This is a feature that may not be available on all positioner models. In the illustrated design, calibration is done over the signal wires via HART protocol.

Listed below are some of the advantages of digital positioners and other field devices relative to their analog counterparts:[1]

Increased accuracy, 0.1–1% vs. 0.3–2% for analog
Improved stability, about 0.1% compared to 0.175%
Wider rangeability (without sacrificing accuracy) up to 50:1 compared to 10:1
Capable of performing multifunctions
Diagnostic and self-testing capability
Ability for miniaturization
Ability to be calibrated and adjusted without physical access to the instrument
Utilize information that could not be measured or accessed by analog means

Typical Performance Specifications

Analog input signal: 4–20 mA DC (4–12 or 12–20 mA split-range optional)
Signal voltage for HART communication: 11 V
Minimum current input: 3.5 mA
Load: 300 Ohm
Max. voltage: 30 V
Power source: extra-low voltage (SELV)
Reverse polarity protection
Actions: Single-direct or single-reverse (double acting, optional)
Hazardous area classifications: as required
Temperature limits: −40–85°C (−4–185°F)
Typical pneumatic supply pressures: 0.3 bar (5 PSIG) to 8 bar (120 PSIG)
Steady-state air consumption: 400 sL/hr (14 scfh) at 1.4 bar (20 PSIG)
Max. output capacity: 10,600 sL/hr (375 scfh) at 1.4 bar (20 PSIG)
Min. dead band: 0.2% of output span
Independent linearity: ±0.75% of output span.

Generating the Pneumatic Output

There are basically two ways to control the pneumatic output signal to the valve actuator. The first is to use a conventional I/P transducer coupled with an amplifying relay to increase the air flow to the valve actuator. The second is to generate the pneumatic output signal by pulsing either two piezoelectric valves, or two miniature solenoid valves (one for each stroke direction), and then amplify this output by a relay.

The advantage of the pulsed system is that there is less air consumption, at least when the valve is not moving. Other advantages include that the valve travel time (time constant) can be changed quite easily by varying the pulse rate of the miniature valves. This could improve the stability of the valve and the control loop's performance. Another advantage is a more compact positioner size.

A disadvantage is that these devices can have a larger dead band than their I/P-controlled cousins. This is necessary to keep the miniature valve from continuously pulsing if the signal is noisy. Unintended continuous pulsing can also happen, if there is a small air leak in the actuator or in the joints of the connecting air tubing. This could lead to an early failure of the "on/off" micro valves, which have a lifetime of about 10^8–10^9 pulses. In a worst-case scenario, this lifetime limit could be reached in a little over a year, in case of sustained high frequency instability!

Smart positioners offer a number of functional advantages, which are easily obtainable by digital processing and which could not be obtained from analog electronic positioners. Some of these features will be discussed in the following paragraphs.

Valve Performance Monitoring

As was discussed in Section 6.8, the ability to monitor the performance of the control valve is a widely advertised feature of intelligent positioners. Valve performance is monitored by checking some of the basic valve calibration parameters, which include the "zero" position and the travel span of the valve. Additional tests can monitor the air pressure in the actuator against travel. This "signature" is compared against data obtained when the valve was newly installed.

A major deviation from the "desired" characteristic could indicate potential problems such as the valve stuffing box being too tight, the valve stem being corroded, or the actuator spring being damaged. When such conditions evolve, they can cause an increase in the dead band and the dead time of the valve, thereby creating a potential for instability and cycling of the control loop.

This test data can be called up via HART protocol and can be a vital part of the plant's asset management system (also discussed in Section 6.11). Some believe such data could also be transmitted to a computer via a fieldbus digital signal while the valve is actively controlling the process. Others argue that in order to measure the valve's performance, one has to provide an artificial offset to the control signal. Therefore, some plant operators are understandably reluctant to utilize this feature, because they fear that this offset to the control signal would create too much disturbance in the control loop.

As a result, one typically has to wait until the valve is out of service (process is shut down) in order to perform functionality tests.

It also has to be realized that continuous monitoring of valve performances can create an information overload for the engineer or technician who may be responsible for up to 600 control loops! It therefore may make more sense to monitor the quality of the control signal (against preset limit values) at the controller. Any significant deviation could then be used to trigger a check of the control valve and its positioner.

Otherwise, such "call-up" of the valve's performance data, perhaps with a handheld HART, has to wait until it is time for the regular scheduled maintenance.

Future Trends and Tasks Future uses of the intelligent positioner as "data transmitter" will no longer be restricted to the monitoring of the condition of the valve actuator but may include

1. Monitoring the leakage of the valve's packing box or stem bellows using the "binary" output of suitably placed pressure sensors, or by "sniffing" the ambient air around the valve stem to detect toxic substances by means of a miniaturized chemical analyzer.
2. Checking for fluid leakage between plug and seat by feeding the signal of a sound pressure transducer the sound frequency within the valve housing against a known frequency profile to the positioner.
3. Another method of indirectly discovering excessive seat leakage is by comparing the controller output signal at "low flow" conditions with the same signal, which was measured for the same low flow rate when the valve was new. Excessive leakage would result in such a signal change, which will show less travel (less opening) to compensate for the increase in leakage.

An important feature of all positioners is the time they take to either evacuate all the air from the actuator (if the valve is "air-to-open"), or to deliver the maximum air pressure to the actuator (if the valve is "air-to-close") whenever the control signal calls for the valve to be closed. When this occurs, in order to avoid seat leakage, the positioner output signal must change rapidly to apply the maximum actuator force onto the valve plug.

Some digital positioners offer memory-embedded data such as serial number, date purchased, vendor, tag number, and valve size. These are all useful for asset management purposes and especially handy when the original valve serial plate is corroded and has become unreadable.

Future intelligent positioners may be able to receive their control signal via radio transmission. The difficulties with such a schema include reliability, path loss, RF interference, multipath radio echoes, and the required transmission power. Most of all, these positioner will still require sufficient electrical power to operate their pneumatic and electronic components. Consequently, the valve positioners will never be truly "wireless."

CONTROLLING THE PROCESS

Embedding the control algorithm (usually PID) inside the digital heart of the positioner transforms the control valve package into a dominant part of a "distributed control system." Such a smart control valve requires only an input from a process sensor/transmitter and a supervisory (adaptive, feedforward, and so on)or set point signal from a central computer.

Such imbedded algorithms initially tend to involve PID control only but may later include more advanced features to serve self-tuning or fuzzy logic functions.

Changing the Valve's Characteristics

Another common feature of digital positioners is their ability to modify the relationship between the controller's output signal and the pneumatic output signal to the valve actuator. The effect of this relationship is to alter the valve's inherent flow characteristic. Choices include "quick opening," "linear," "equal percentage" (Figure 6.7b), and a variety of "user-defined" characteristics that must be custom programmed.

Such a modification of the valve's inherent characteristic is similar to placing a mechanical cam in the travel feedback linkage of a pneumatic positioner. For example, it can convert

FIG. 6.12b
Typical intrinsically safe digital positioner that can provide single- or double-acting pneumatic output (note pneumatic pressure gauges). (Courtesy EMERSON Process Management–Fisher.)

the static flow characteristic of a butterfly valve, which has a nearly linear characteristic (C_v vs. travel is linear), into an equalpercentage one (C_v varies exponentially with controller signal).

The trouble is that the travel vs. flow coefficient (C_v) of a butterfly valve is still linear. The result is that the signal resolution at low flow becomes very poor, because with an equal percentage output signal at low flows, the valve travel is very small per given signal change. Therefore, in such cases the valve's dead band becomes high compared to its signal resolution.

Valve Gain and Time Constant The purpose of changing the flow characteristic is to make the "gain" (rate of flow change vs. signal change) of the control valve such that the loop gain will be more constant, in order to aid in maintaining the stability of the loop control (Figure 6.7a). This can be accomplished, at least in theory, by the above-described modification of the control signal.

In practice, this approach works only in applications where the time constants of the valve and of the process differ by at least a factor of five. The reason is that the "static" signal modification performed by the positioner results in vastly different stem travel speed requirements at low and at high flow rates. This in turn results in a continuous variation of the valve's time constant,^{3} which makes controller tuning very difficult.

It is hoped that future positioner designs will provide algorithms that will change the rate of output signal change in proportion to the rate of signal modification. This would mean that the travel speed of the actuator is changed so that the change in flow through the valve becomes proportional to the change in the controller's output signal. This way one would have not only a constant control valve gain but also a constant control valve time constant!

Another problem caused by the changing of the flow characteristic has to do with the dead band of the valve. For example, if one changes the inherent "linear" valve characteristic to "equal percentage," the valve travel for say the first 10% of signal range (0–10%) will be very small (typically 0–3%). This means that the apparent dead band of the valve stem will increase three times, and this can cause oscillation in the control loop.

OPERATION OF SMART POSITIONERS

From the outside there appears to be little difference between smart digital and analog electronic positioners (Figure 6.12b), other than a digital monitoring window on the outside of some smart positioners.

The real difference lies inside. As can be seen in Figure 6.12c, all communication as well as power supply travel through a single pair of wires from the controller or computer to the printed wiring board containing all the functional electronics. This signal may be in the form of a 4–20 mA analog, or in a digital form according to the selected fieldbus protocol.

The controlling electronic portion is usually encapsulated and field-replaceable. This printed wiring board receives feedback from a motion transmitter, which is sensing the position of the valve stem.

Any deviation from the desired valve stem position will be detected, and a correcting signal is sent (in this case) to an I/P transducer that converts the electronic signal into a

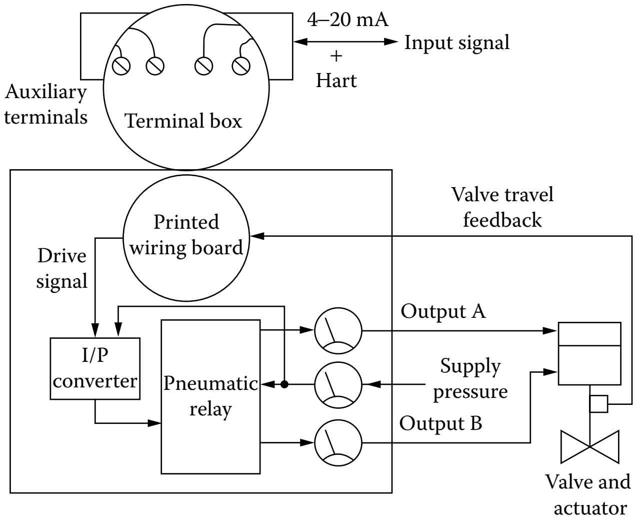


FIG. 6.12c
Schematic view of a digital valve positioner that is sending an electronic analog output signal into an I/P transducer, which in turn is generating a corresponding pneumatic output signal.

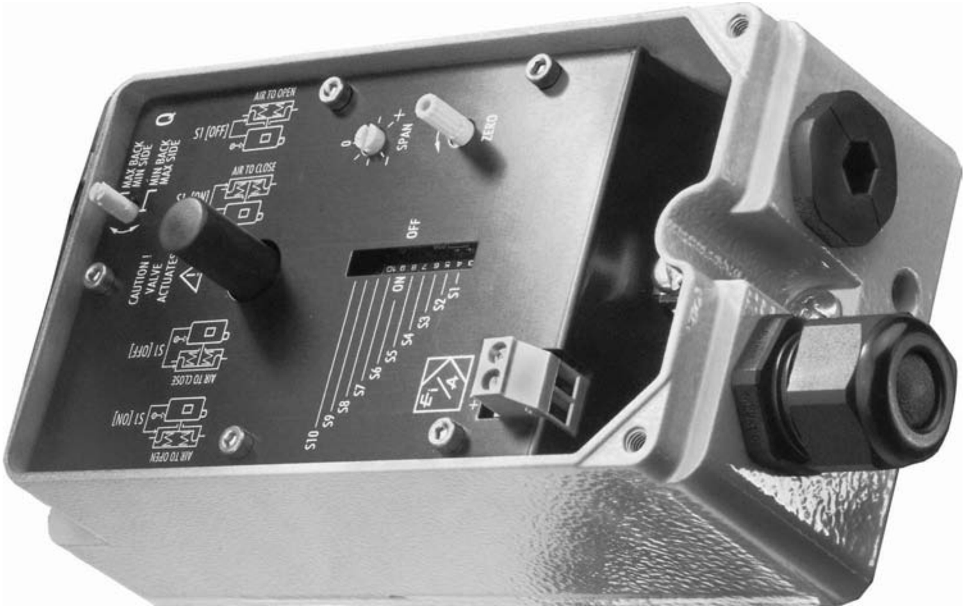


FIG. 6.12d
Digital valve postioner with cover removed. It shows the adjustment knobs for "zero," for "span" (travel), and for the type of action as "direct: air to open" or "reverse air to close." Finally, there is a setting for the low voltage internal limit switch. (Courtesy of SAMSON A.G.)

corresponding pneumatic one. This air signal is then amplified by a relay and fed to the valve's actuator.

Maintenance and Calibration

An additional pressure sensor (measuring the output of the air relay) can give additional feedback to the circuit board in order to aid in maintaining the stability of the loop. This pressure sensor also monitors the valve's performance by comparing the relationship between the air signal and travel with a stored relationship that was taken when the valve was new. As is illustrated in Figure 6.8c and discussed in Section 6.8, a major difference between the two relationships signals the need for maintenance.

After the positioner is mounted on the valve, its initial calibration can be done by a remote device such as a computer via a HART protocol, or alternatively by using internal adjustment knobs as shown in Figure 6.12d.

ACCESSORIES

The most common built-in accessories available as options in intelligent positioners are micro-switches, which typically operate on low-voltage levels (24 V). These switches provide the capability of a "safety shutdown" of the valve, which is initiated independently from the control signal. Other accessory options include analog or digital valve travel position transmitters.

In addition one also requires software, which is needed for the intelligent positioner to perform the desired diagnostic and remote calibration functions. Listed below are some of the capabilities that such a software package should support:

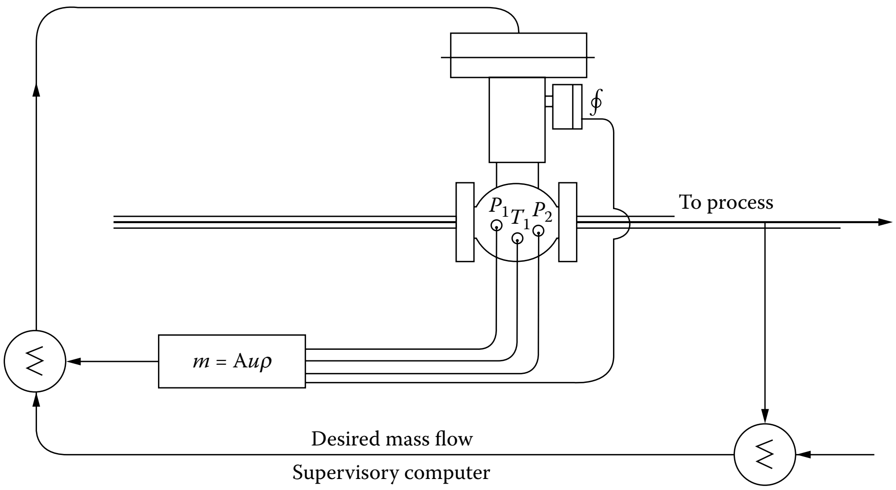
Reading of valve diagnostic data
Supporting the graphical diagnostic trends window displays
Valve configuration, and valve configuration database
Monitoring device variables
Configuration of valve characteristic
General process and valve database
Security levels for users
Diagnostic database

FLOW CONTROL BY SMART VALVE

It is possible to integrate a whole flow control loop into a control valve that is a self-contained part of a distributed control system. Such a device would include the valve, the flow sensor, the valve positioner. and the controller, all incorporated into a single entity.[4]

As shown schematically in Figure 6.12e, it includes a temperature sensor and a flow detector, which consists of two piezoelectric pressure sensors embedded within the valve housing and located up- and downstream of the valve orifice.

The mass flow through the valve is measured by detecting the differential pressure across the valve orifice and by applying a density correction to that differential on the basis of the detected temperature. This product is then multiplied by the known area of the vena contracta between orifice and valve plug at the prevailing inner valve travel position, and thereby



$$A = N_3C_v\phi(\text{ ft}^2);\ U = N_2\sqrt{P_1 - P_2}\ (\text{ ft/s});\ \rho = N_1\gamma P_1P_2/T_1\ (\text{lbs/ft}^3)$$

A = Area of valve opening
C_v = Valve capacity coefficient
m = Mass flow, lbs/s
N_1, N_2, N_3 = Numerical constants
P_1 = Absolute inlet pressure, lbs/ft^2
P_2 = Absolute outlet pressure, lbs/ft^2

T_1 = Inlet temperature, °R
U = Orifical velocity, ft/s
γ = Specific weight, lbs/ft^3 @ 70°F and 14.7 psia
ϕ = Fractional valve travel
ρ = Density of flowing gas (lb/ft^3)

FIG. 6.12e
Schematic view of a complete flow control loop built around an intelligent control valve.[4]

the on-board computer calculates the mass flow passing through the valve. Flow control is provided by a PID algorithm that compares the measured flow with a desired set point and adjusts the valve opening as needed to meet that set point.

Limitations

In order to make this type of flow measurement reasonably acceptable, one must locate the pressure sensors in a long and narrow channel in the body cavity to eliminate the effects of flow turbulence disturbances. While not an accurate flow sensor, it can detect a "deviation" from a desired flow rate and correct the flowing quantity.

Such a device can be used in temperature cascade loops, where this intelligent valve may control the flow rate of steam, while a temperature controlling cascade master adjusts the set point of this intelligent valve-based flow-controlling slave.

Such an intelligent valve provides ease of installation, space savings, and the elimination of extra wiring. While the price is higher than that of a conventional control valve, it is less than the combined cost of the components of a mass flow control loop.

References

1. Schneider, H.-J., "Digital Feldgeraete: Vorteile, Probleme und Anforderungen aus Anwendersicht," *ATP, Automatisierungstechnische Praxis, Germany,* April 1995), pp. 50–54.
2. Harris, J. "Assessment of Control Loop Performance," *Canadian Journal of Chemical Engineering,* Vol. 67, pp. 856–861, 1989.
3. Baumann, H. D., "Positioners Set the Course for Modern Control Valve Designs," *I&CA,* May 1997, pp. 41–45.
4. Baumann, H. D., *Control Valve Primer,* 3rd edition, Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1998, pp. 72–76.

Bibliography

Ali, R., "Smart Positioners to the Rescue," *InTech,* June 2002, pp. 72–73.

Ali, R. and Jero, L., "Smart Positioners In Safety Instrumented Systems," *Petroleum Technology Quarterly,* Winter 2002/2003.

Baumann, H. D., "Interkama 1999: The Status and Future of Control Valves: A Global Review," *ATP Automatesierungstechnische Praxis, Germany,* Vol. 42, No. 3, pp. 20–25, 2000.

DeGaspari, J., "A Tale of Two Buses," *Mechanical Engineering,* September 2001.

Emerson Process Management, www.emersonbprocess.com/Fisher/news/newproduct/newliter/pdfacilities.html.

Kiesbauer, J., "Detektion Der Inneren Leckage Bei Stellgeraeten," *ATP, Automatesierungstechnische Praxis, Germany,* Vol. 42, Heft 11, 2000, pp. 50–53.

Kiesbauer, J. and Hoffmann, H., "Anlagennahes Asset Management Bei Stellgeraeten-Eine Standortbestimmung," *ATP, Automatisierungstechnische Praxis, Germany,* Vol. 7, July 2002, pp. 30–37.

Miller, L.,"Valve Diagnostic Past, Present, and Future," *Fluid Handling Systems,* November 2002.

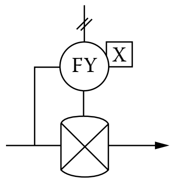
Render, N., "Digitally Mastered," *Control & Instrumentation,* November 2000.

Shinskey, F. G., "How Good Are Our Controllers in Absolute Performance and Robustness?" *Measurement and Control,* Vol. 23, No. 5, pp. 114–121, 1990.

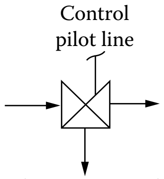
6.13 Miscellaneous Valve and Trim Designs

C. S. BEARD (1970, 1985) **B. G. LIPTÁK** (1995)

J. P. WILSON (2005)



Expansible tube or diaphragm valves



Fluid interaction valves

Flow sheet symbols

Valve and Trim Designs:	A. Anticavitation Valves B. Dirty Service Valves C. Low-Noise Valves D. High-Capacity Valves E. Cryogenic Valves F. High-Temperature Valves G. Steam Conditioning Valves H. Tank Mounted Valves I. Expansible Tube or Diaphragm Valve J. Fluid Interaction Valves
Available Range of Sizes:	A. 1 to 36 in. (25 to 915 mm) B. 1 to 36 in. (25 to 915 mm) C. 1 to 48 in. (25 to 1220 mm) D. 12 to 72 in. (300 to 1830 mm) E. 1 to 36 in. (25 to 915 mm) F. 1 to 20 in. (25 to 508 mm) G. 6 to 48 in. (152 to 1220 mm) H. 4 to 24 in. (100 to 600 mm) I. 1 to 12 in. (25 to 300 mm) J. $^1/_2$ to 4 in. (12.5 to 100 mm)
Design Pressure Limits:	A. Up to 8000 PSIG (550 bar) B. Up to 8000 PSIG (550 bar) C. Up to 4000 PSIG (275 bar) D. Up to 3000 PSIG (205 bar) E. Up to 1500 PSIG (100 bar) F. Up to 5000 PSIG (345 bar) G. Up to 5000 PSIG (345 bar) H. Up to 1500 PSIG (100 bar) I. Up to 1500 PSIG (100 bar) J. 100 PSIG (7 bar) or greater (no theoretical limit)
Maximum Operating Temperatures:	A. Up to 1100°F (595°C) B. Up to 1100°F (595°C) C. Up to 1100°F (595°C) D. Up to 1100°F (595°C) E. –325°F (–162°C) to 300°F (150°C) F. Up to 1800°F (980°C) G. Up to 1100°F (595°C)

H. Up to 1100°F (595°C)
I. Up to 150°F (66°C)
J. Can handle molten metals

Partial List of Suppliers:
ABB Process Automation (www.api-usa.com)
Emerson Process Management–Daniel Division (www.daniel.com)
Emerson Process Management–Fisher Valve Division (www.fisher.com)
Flowserve Corporation (www.flowserve.com)
Masoneilan Division of Dresser Flow Control (www.masoneilan.com)
Siemens (www. siemens.com)

INTRODUCTION

This section consists of two distinct parts. In the first part, a number of control valve designs are described, which are different in that they are neither linear nor rotary in their operation. These miscellaneous valve designs utilize the energy content of the flowing fluid for their operation. They depend on fluid interaction or the static pressure of the process fluid and use flexible elements to throttle gas pressure regulators and other control valves. These valves are used in specialized services, such as in sensitive level control, in gas pressure or flow regulation, and in toxic services.

In the second part of this section some of the special valve trim designs used in the traditional globe control valves are discussed in terms of their suitability for specialized applications. These applications include applications where noise; cavitation; flashing; high and low temperatures or pressures; or viscous, dirty, or slurry flows are involved.

Such trim designs and applications have already been discussed in Sections 6.1, 6.7, 6.14, and 6.15.

MISCELLANEOUS VALVE DESIGNS

Dynamically Balanced Plug Valves

This family of control valves has been developed for pressure and flow control applications where no external power is available to operate the valve, and therefore the static pressure of the process fluid is utilized to achieve throttling. Figure 6.13a illustrates an installation for upstream (back) pressure control.

Here, the upstream pressure is sent to the control pilot through port D. If the controlled upstream pressure drops, this lowers the pressure in the pilot chamber and the pilot spring moves the pilot poppet valve to the right. This opens port B and thereby upstream pressure is applied inside the plug chamber. This equalizes the pressure acting on the plug and allows the spring to move the plug to the left, to close the valve.

These valves provide nearly linear characteristics (Figure 6.13b) and high flow capacities ($C_v = 30d^2$). They are available in sizes up to 12 in. (300 mm), and special units

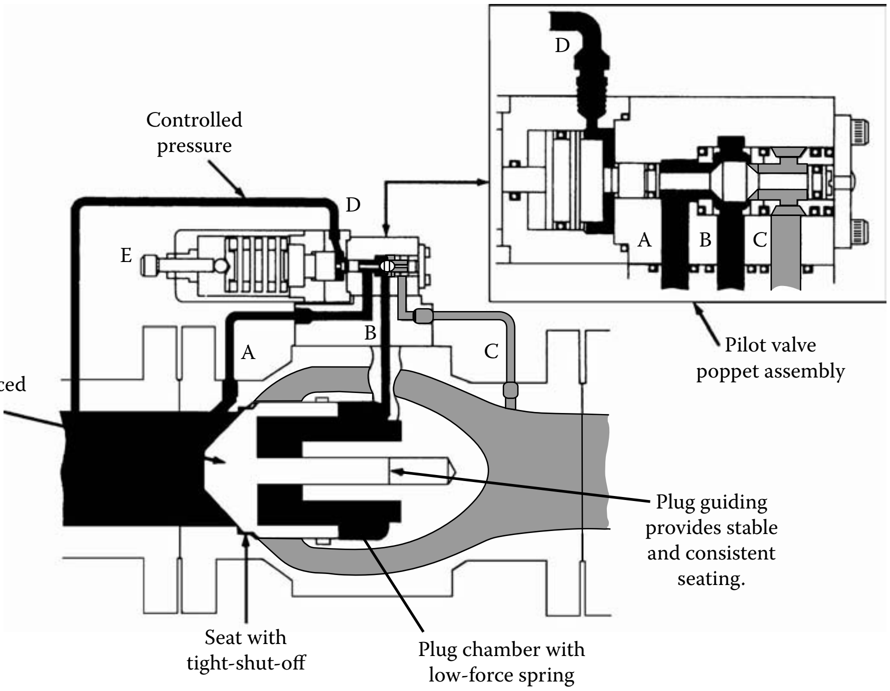


FIG. 6.13a
Dynamically balanced valve plug. (Courtesy of Daniel Industries.)

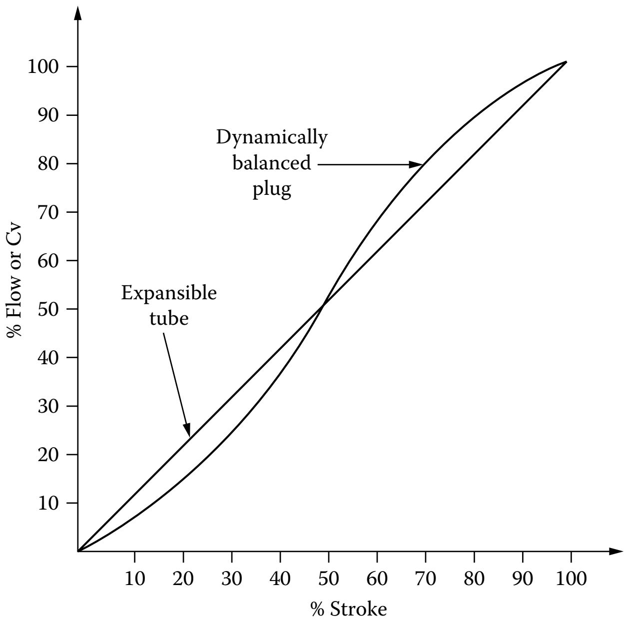


FIG. 6.13b
The characteristics of expansible tube and dynamically balanced plug valves are nearly linear.

have been made in up to 24 in. (600 mm) sizes. The valves can be provided with up to ANSI Class 1500 rating and with pressure control settings up to 3000 PSIG (205 bar).

The advantages of this design include its erosion and corrosion resistance, due to the hard facing of the plug; its good pressure recovery characteristics, due to the large and smooth annular flow area through the valve; its fast speed of response; and its bubble-tight seating.

Positioned Plug In-Line Valves

The positioned plug in-line valve, excluding control units, resembles a pipe spool. It is only necessary to inject pressure into its ports for positioning the valve plug. This simple design (Figure 6.13c) requires only three pressure seals. The plug is carried on a cylinder that also includes the piston. Pressure in one port causes closing, while the opposite port is used for opening. The valve has only one moving part and is available in sizes from 2–8 in. (50–200 mm) for use to 350 PSIG at 400°F (2.4 MPa at 204°C). Control quality is dependent upon the pilot valves and auxiliary units employed.

Another in-line valve available in small sizes (Figure 6.13c) carries the valve plug on a bridge in the operating cylinder, with the seat as part of a split body.

A spring-loaded version of this design uses the beveled end of the moving cylinder to seat on a replaceable soft seat, retained in a dam, held in position by struts from the inside wall of the valve body. The spring loading may cause fail-close or fail-open actions, as illustrated in Figure 6.13e. The unit can be powered with line fluid or by an external pressure source. A double-bleed feature can be incorporated to eliminate the possibility of actuation and line fluids combining if one of the dynamic seals should fail. All seats and seals are replaceable by separation at the body flange.

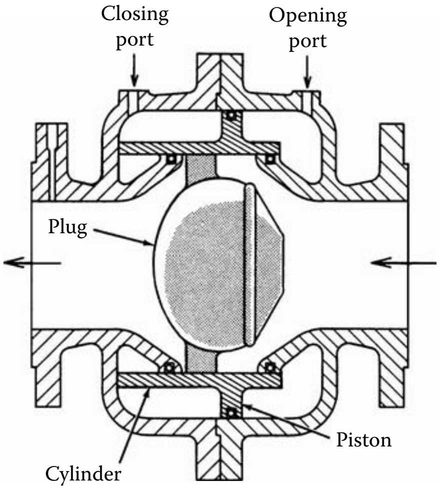


FIG. 6.13c
Positioned plug valve in its fully open position. (Courtesy of Eisenwerk Heinrich Schilling.)

The unit is particularly adaptable to fluids that are toxic or difficult to contain, such as nitrogen tetroxide, hydrogen, and others used in the aerospace industry. The unit is furnished in sizes from $1^1/_2$–18 in. (37.5–450 mm), with ratings to 2500 PSIG (17.3 MPa). All types of end connections are available, and control is dependent upon the auxiliary control components selected for the application. An explosionproof limit switch can be furnished for position indication.

Diaphragm-Operated Cylinder In-Line Valves An in-line valve using a low convolution diaphragm for positive sealing and long travel (Figure 6.13f) is designed particularly for gas regulation. The low level of vibration, turbulence, and noise of this in-line design makes it suitable for high-pressure gas service.

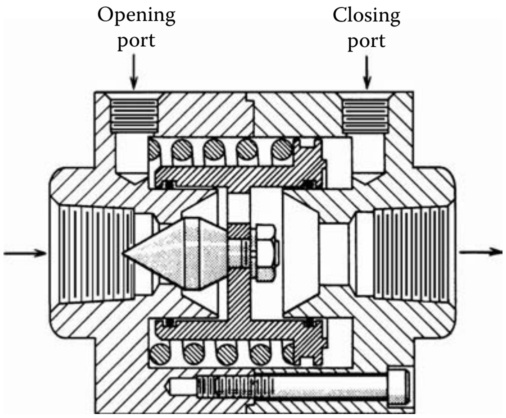
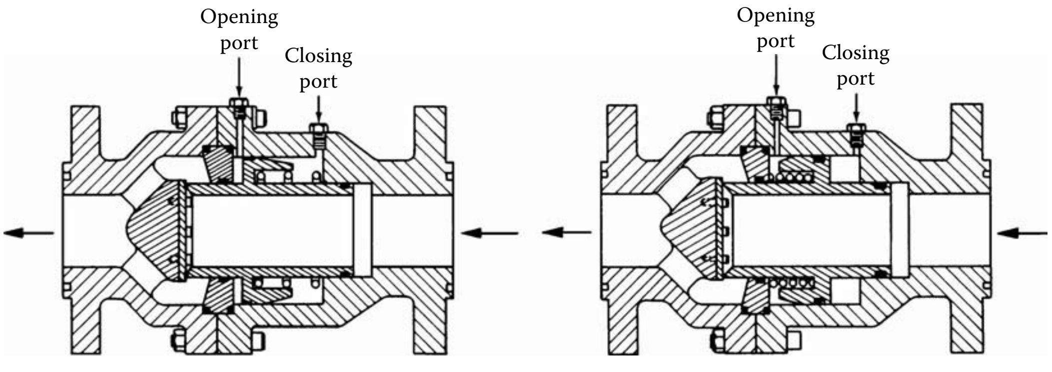


FIG. 6.13d
Positioned plug valve in closed position. (Courtesy of Control Air Inc.)



Fail closed design

Fail open design

FIG. 6.13e
Spring-loaded positioned plug valve.

Inlet pressures to 1400 PSIG (9.7 MPa) and outlet pressures to 600 PSIG (4 MPa) are possible in the 2 in. (50 mm) size. It is a high-capacity valve, as expressed by $C_v = 23\ d^2$. As a gas regulator the unit is supplied with a two-stage pilot to accept full line pressure. This pilot resists freeze-up and serves as a differential limiting valve. All portions of the pilot and line valve will withstand a full body rating of 600 PSIG (4 MPa).

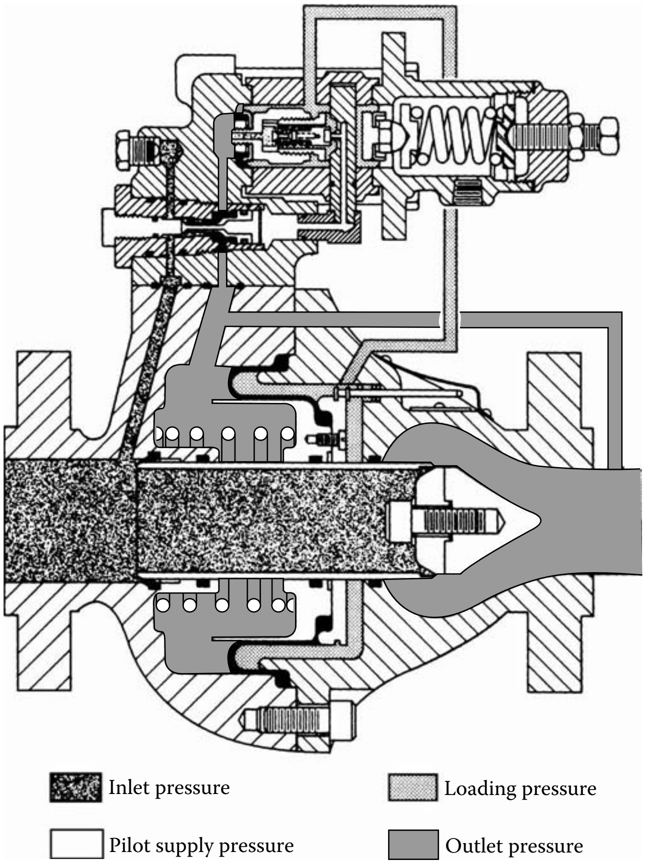


FIG. 6.13f
Diaphragm-operated cylinder-type in-line valve used in high-pressure gas regulation services.

Expansible Valve Designs

The expansible element in these valve designs can be a rubber cylinder, an expansible tube, or an expansible diaphragm. The common feature of all of these designs is that they utilize the process pressure to provide tight shut-off of gas flows.

Expansible Element In-Line Valves Streamlined flow of gas occurs in a valve in which a solid rubber cylinder is expanded or contracted to change the area of an annular space (Figure 6.13g). A stationary inlet nose and discharge bullet allow hydraulic pressure to force a slave cylinder against the rubber cylinder to vary its expansion. Control is from a diaphragm actuator, with the diaphragm plate carrying a piston. The piston acts as a pump to supply hydraulic pressure to the slave cylinder.

The rubber cylinder offers the seating ability of a soft seat valve. It has the capability of closing over foreign matter, and

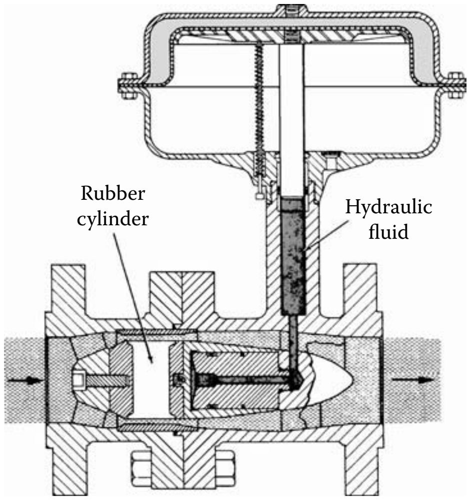


FIG. 6.13g
A rubber cylinder provides the soft seat in this low-noise, expandable element, pressure regulator in-line valve.

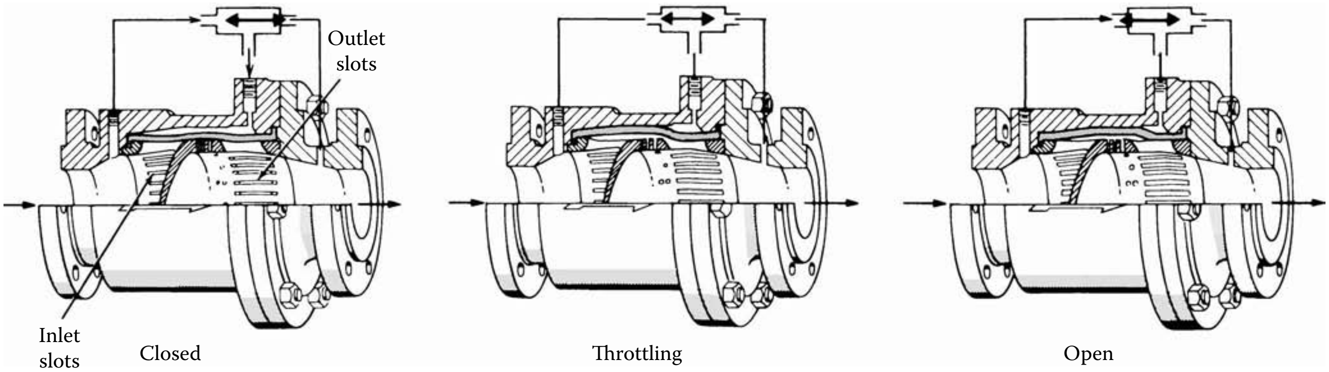


FIG. 6.13h
An expansible tube valve utilizes the process pressure for its operation, while being controlled by a three-way pilot that determines if the valve is to be closed (left), throttling (center), or open (right).

the design allows for the use of a restricted throat for reduced capacity. With this design, pressure drops as high as 1200 psid (8.3 MPa) have been handled with a low noise level. The valve may be utilized as a pressure reducer or for back-pressure control, depending upon the system requirements.

Available sizes are 1–6 in. (25–150 mm). The 1 in. (25 mm) valve can have screwed connections, while all sizes can be flanged. The body is steel with flange ratings to 600 PSIG (4 MPa).

A valve positioner can be used, if the stem position is calibrated as a function of the annular space reduction. If such calibration is provided, a controller output can throttle the valve to obtain accurate flow control.

Expansible Tube Valves Control of flow is obtained by use of an expansible tube that is slipped over a cylindrical metal core containing a series of longitudinal slots at each end and a separating barrier in between. The characteristics of such a valve were shown in Figure 6.13b.

A cylindrical, in-line jacket surrounds the tube so that the process pressure can be introduced between the jacket and the sleeve to cause the sleeve to envelope the slots. This valve will open if the space between the jacket and the sleeve is connected to the downstream (lower) pressure and will open if that space in connected to the upstream (higher) pressure.

With pressure connected to the downstream line (Figure 6.13h, right), the line pressure in the valve body will cause the valve to open fully. Control of the pressure on the sleeve creates a throttled flow condition by first uncovering the inlet slots and then progressively opening the outlet slots (Figure 6.13h, center). A continuous dynamic balance between fluid pressure on each side of the sleeve makes it possible to obtain wide rangeability between a no-flow and a full flow (fully open) condition.

The basic operation of the valve can be accomplished with a three-way pilot valve positioned from a remote location. A variety of automatic pilots give versatility to the basic valve. For reduced pressure control, a pilot is used to modulate the jacket pressure in response to the sensed pressure in the downstream pipeline.

As downstream pressure falls below the set point, the double-acting pilot positions itself to reduce the jacket pressure. This allows the valve to open to a throttling position. Therefore, downstream pressure increases to the set pressure, with attendant change in flow rate to maintain the set pressure. The static sensing line is separate from the pilot discharge line, in order to eliminate the pressure drop effect in the sensing line.

Another form simulates a conventional regulator, in that system gas is bled into the jacket annular space through a fixed orifice and bled off through the pilot regulator. In this form, the static sensing line and pilot output are common. Double-acting pilot systems use seven control ranges from 2 to 1200 PSIG (0.02–8.3 MPa) with corresponding inlet pressures up to 1500 PSIG (10.4 MPa). The fixed orifice design is available for control from 2 oz–600 PSIG (0.86 kPa–4 MPa).

Pilot Design Variations Back-pressure control and pressure relief are obtained in the same manner as pressure reduction is, except they sense the upstream pressure. By using a separate sensing and bleed port, a build-up from cracking to fully open can be varied from 3 to 14% of the set pressure. Return to normal operation causes the valve to create absolute shut-off. Emergency shut-off service may use an external pressure source piloted to obtain immediate shut-off upon abnormal conditions.

A diaphragm-operated, three-way slide valve may also control jacket pressure by proportioning the inlet and outlet pressures. In this design, the controller output signal is sent to the diaphragm actuator of the three-way slide valve, which causes the sleeve valve to open proportionally, in a manner similar to that of a conventional diaphragm control valve. In this manner, the valve becomes a throttling control valve (Figure 6.13i).

Differential pressure or flow control is accomplished by using a pilot valve in which the diaphragm is positioned by both upstream and downstream pressures. The differential

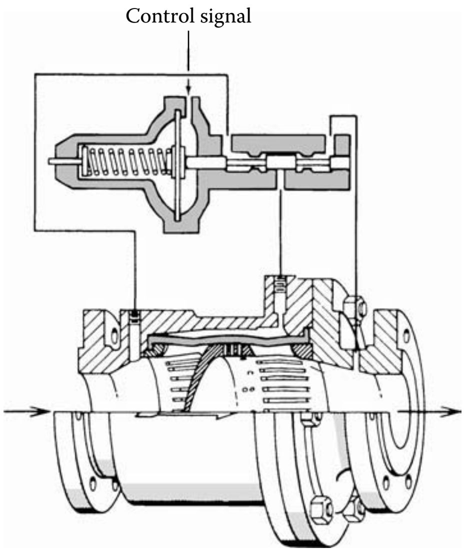


FIG. 6.13i
The expansible tube valve becomes a throttling control valve if it is provided with a three-way slide-valve-type control pilot.

static lines may be taken at the inlet and outlet of the line valve or, for flow control, across an orifice plate in the line.

Pressure Boosting A practically unlimited variety of pressure and flow regulators can be configured with the range of pilots that are available. One important application is in pressure boosting in gas distribution systems. By these boosters, the line pressure losses due to increased consumption can be counteracted automatically by increasing the set pressure of the distribution control valve.

In such a design, there can be as many as three pilots controlling the gas distribution control valve. In Figure 6.13j, the normal pressure is controlled by pilot #1. An increase in flow is sensed by the flow element (orifice, or flow tube), which opens a high-differential pilot to cut in a boost-pressure-control regulator (pilot #2), which has a higher set point than pilot #1. Return to normal flow cuts out the boost pressure control regulator and reinstates the normal control pilot.

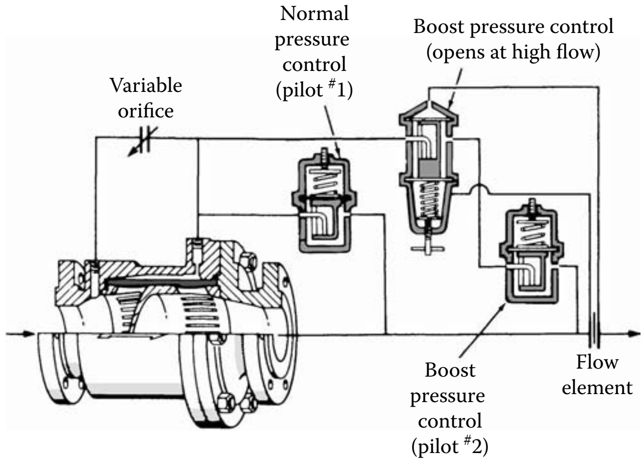


FIG. 6.13j
Expansible tube-type gas distribution control valve can automatically boost the distribution pressure, when the demand for gas increases.

The expansible tube valve is made in sizes from 1–12 in. (25–300 mm) with pressure rating from 200 PSIG (1.4 MPa) in iron to 1500 PSIG (10.3 MPa) in steel construction. It is made flanged or flangeless for insertion between line flanges. The flangeless body is cradled in the studs between the line flanges. Removal of the body is made easier by expanding the flanges about $^1/_8$ in. (3 mm) using nuts on the studs inside the flanges.

Tight shut-off or throttling requires a differential between the line pressure or external source used for closure and the downstream pressure sensed on the inside surface of the sleeve that is exposed to this pressure. This differential requirement for a special low-pressure 2 in. (50 mm) valve is 3.6 psid (24.8 kPa) and is 1.6 psid (11 kPa) for a 4 in. (100 mm) valve. The low-pressure series requires from 21 psid (145 kPa) for the 1 in. (25 mm) size to 4.6 psid (32 kPa) for the 10 and 12 in. sizes (250 and 300 mm). High-pressure models require from 58 psid (400 kPa) for the 1 in. (25 mm) size to 11 psid (76 kPa) for the 10 and 12 in. sizes (250 and 300 mm).

The body design allows tight shut-off, even with comparatively large particles in the flow stream. Freezing of the pilot by hydrates is not common because the intermittent and small bleed occurs only to open the valve and as such is not conducive to freezing.

The pilot may be heated or housed, or even located in a protected area close to the warm line. With its only moving part a flexible sleeve, this type of valve has no vibration to contribute to noise. The flow pattern also helps make this valve from 5 to 30 db more silent than most regulators. Flow capacities are comparable to those of single-seated as well as many double-seated regulators.

Expansible Diaphragm Design In an expansible diaphragm-type regulator, an expansible element (Figure 6.13k) is stretched down over a dome-shaped grid causing shut-off of the valve when pressure above this resilient member overcomes the line pressure under the element. Line pressure is evenly directed over the expansible area by a series of pressure channels.

Selection of the correct action on a pilot that supplies line or external pressure to the exterior of the expansible element causes the valve to control as a back-pressure or as a reducing control valve. For back-pressure control, or pressure relief, the static line is taken upstream of the valve. Increase in line pressure will increase the bleed from the annular space between the expansible element and the metal housing. Reducing regulation is accomplished by restricting the bleed upon increase in downstream pressure and increasing it upon decrease in downstream pressure.

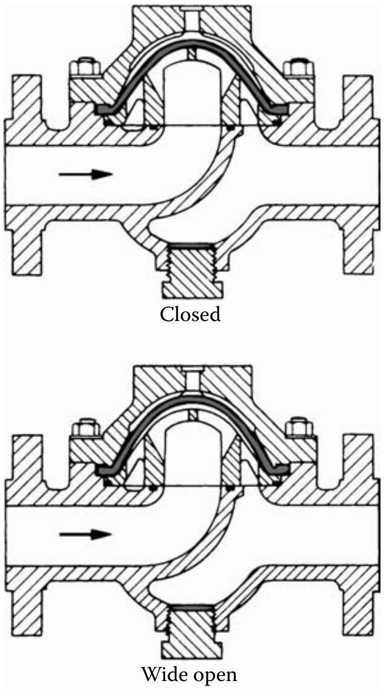


FIG. 6.13k
The expansible diaphragm valve can be used as either a pressure regulator or as a back-pressure relief valve.

The valve is available in iron or steel with ratings to 600 PSIG (4 MPa). Models are available from −10–150°F (−23–66°C), using a molded, Buna-N diaphragm. Relief valve pressures are from 30–300 PSIG (0.21–2.1 MPa), while reducing service varies from 5–150 PSIG (0.03–1 MPa). Capacity factors vary from $C_v = 11\ d^2$ in smaller sizes to $C_v = 14\ d^2$ in larger valves.

Fluid Interaction Valves

The Coanda effect, the basis of fluidics, is used in diverting valves from $^1/_2$–4 in. sizes (12.5–100 mm). The Coanda effect means the attachment of a fluid stream to a nearby side-wall of a flow passage. This effect can be used in a so-called “flip-flop” valve for diverting a stream from one discharge port to another.

Figure 6.13l shows the flow through the right-hand port due to both control ports being closed. Opening the right-hand port, to allow air or liquid to enter, will shift the flow. The industrial valve has rectangular diverting tubes, but the end connections may be circular. Control is maintained by opening or closing the control port or by injecting low-pressure air or liquid through a solenoid or other pilot valve.

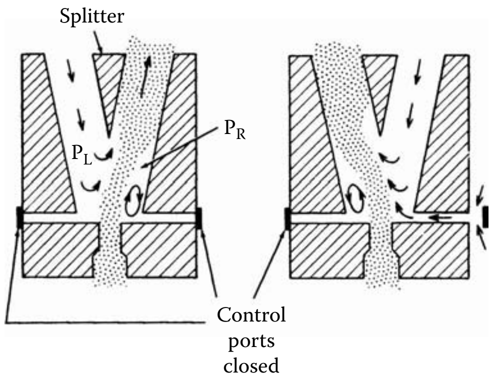


FIG. 6.13l
The Coanda effect is used in the “flip-flop” diversion of flow in the fluid interaction valves.

In this valve, with stream flowing in one diversion tube, the flow at the inlet contains some potential (pressure) energy and some kinetic (flowing) energy. Much of the potential energy is converted to kinetic energy at the nozzle. Up to 70% recovery of potential energy occurs in a diffuser section. Fifty percent recovery is guaranteed for commercial valves, and somewhat less for gases above critical flow and for viscous fluids ($\Delta P \sim P_1/2$).

Installation must allow the recovered pressure to create the desired flow against friction effects of piping or fittings. An uninhibited flow will create some aspirating effects in the open outlet; restriction causes blocking, while excessive restriction will cause a leak or diversion to the open port.

Fast Level Control Industrial valves, with their ability to divert in less than 100 ms, fill a wide variety of uses. The primary one is for level control in which the effluent not required for filling may be returned to storage. It is necessary only to use a dip tube set at the control point, as shown in Figure 6.13m. Lack of moving parts or of detrimental effects due to fast diversion action allows the system to provide close control.

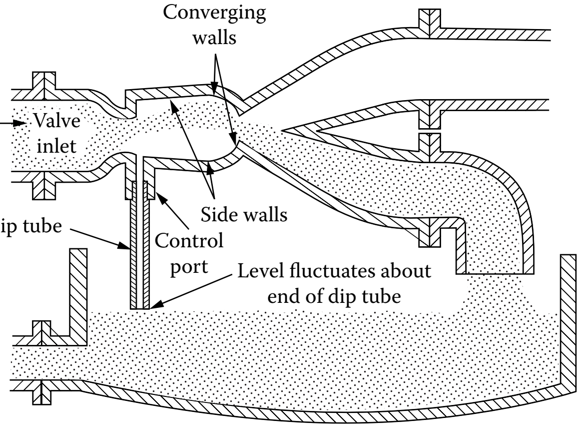


FIG. 6.13m
The response of level control with fluid interaction diverting valves can be very fast.

Numerous uses of diversion valves exist, such as tank filling, which is accomplished by using an external signal. Diversion of a process stream upon contamination, sensed by a pH or other analyzer, is important in paper mills and in chemical plants.

The ability to divert rapidly makes this valve applicable to oscillating flows. The valve may be used for space conditioning, if the total bypassing of a heating or cooling medium is adequate for space temperature control.

Fluidic valves can also be used for the diversion of engine exhaust gases from tailpipe propulsion nozzles to wing-mounted lift fans with ambient control flow. Other unique applications include a four-ported fluidic valve, which has been used for direction control of a missile. It seems that all the potential uses of this valve design have not been exploited yet.

SPECIAL VALVE APPLICATION

In the following paragraphs, the capabilities of one control valve manufacturer (Fisher Controls) are discussed in handling such demanding services as cavitation and sludge. This is somewhat redundant, because Section 6.1 has already covered all the special control valve applications. These included the products of all valve manufacturers and covered noise, cavitation, flashing, high and low temperature or pressure services, and viscous, dirty, or slurry flows.

Yet, the editor believes that this "manufacturer's perspective" that follows is still a valuable addition to the overall picture. The reader is reminded that Sections 6.14 and 6.15 provide in-depth information on control valve noise and on valve sizing aspects of the various special applications. For these reasons, the reader is advised to refer to the above-mentioned sections for an in-depth treatment of the subject matters that are discussed below.

Cavitation and Flashing

As it was already discussed in Section 6.1, a fluid will flash when the downstream pressure is below the vapor pressure of the flowing process fluid. The vapor bubbles that form when the pressure falls below the vapor pressure continue to grow as long as the pressure keeps dropping, and eventually the liquid changes or flashes to a vapor.

In connection with Figure 6.1w, cavitation has been also explained as it was also shown how vapor bubbles can be formed at the vena contracta and how these bubbles can implode and release powerful microjets that will damage any metallic surface, as the pressure rises downstream. In addition, the options available to the process control engineer to eliminate cavitation were also shown in connection with Figure 6.1x.

There are a number of ways to protect against cavitation. As was shown in Figure 6.1aa, one of the most common is to expose the fluid to a series of restrictions as opposed to a single restriction. Each subsequent restriction dissipates a certain amount of the available energy and reduces the inlet pressure to the next stage.

As was shown in Figure 6.1x, a well-designed pressure-staging device prevents cavitation by taking a large pressure differential and by maintaining the vena contracta pressure above the vapor pressure of the liquid, which prevents the liquid from cavitating.

The expanding flow area concept of damage control is closely related to the pressure drop staging approach (Figure 6.13n). Figure 6.13n shows a pressure vs. distance curve for flow through a series of fixed restrictions where the area of each succeeding restriction is larger than the previous. Notice that the first restriction takes the bulk of the pressure drop, and the pressure drop through successive sections decreases.

In the last restriction, where cavitation is most likely to occur, the pressure drop is only a small percentage of the total drop, and the pressure recovery is substantially lower. The expanding flow area concept requires fewer pressure drop stages to provide the same cavitation protection as does the concept that utilizes nearly identical areas for staging.

The most common approach to cavitation protection employs a drilled-hole cage that incorporates both pressure staging and expanding flow area concepts. Each drilled hole has a significant impact on the overall pressure recovery of the valve.

Drilled Trim and Multistage Designs Figure 6.13o shows the cross-section of three types of drilled holes that could be used in an anticavitation cage. The thin plate design is a very inefficient flow device, but it does provide a high F_L^2 value and therefore a low pressure recovery. The thick plate design is not only more efficient, but it also provides a high pressure recovery as denoted by its low F_L^2 value.

The Cavitrol® trim hole, designed by Fisher Controls, is a balance between the thick plate and the thin plate hole designs. It provides relatively high flow efficiency while still maintaining a high F_L^2 value, which results in a low pressure recovery. This design represents the optimal choice between capacity and cavitation control.

Figure 6.13p shows the cross-section of a three-stage, anticavitation trim. This particular design prevents the formation of damaging cavitation at pressure drops up to 3000 psid (207 bar) by utilizing a unique expanding flow area design, meaning that each stage has successively larger flow area.

When a series of drilled holes are used to control cavitation, it is also easy to characterize the trim. In the trim illustrated in Figure 6.13p, as the valve plug travels through the cage, the cage design changes. It begins as a pressure-staging device and transitions to a straight-through, low-restriction flow design. Consequently, the cavitation control ability of

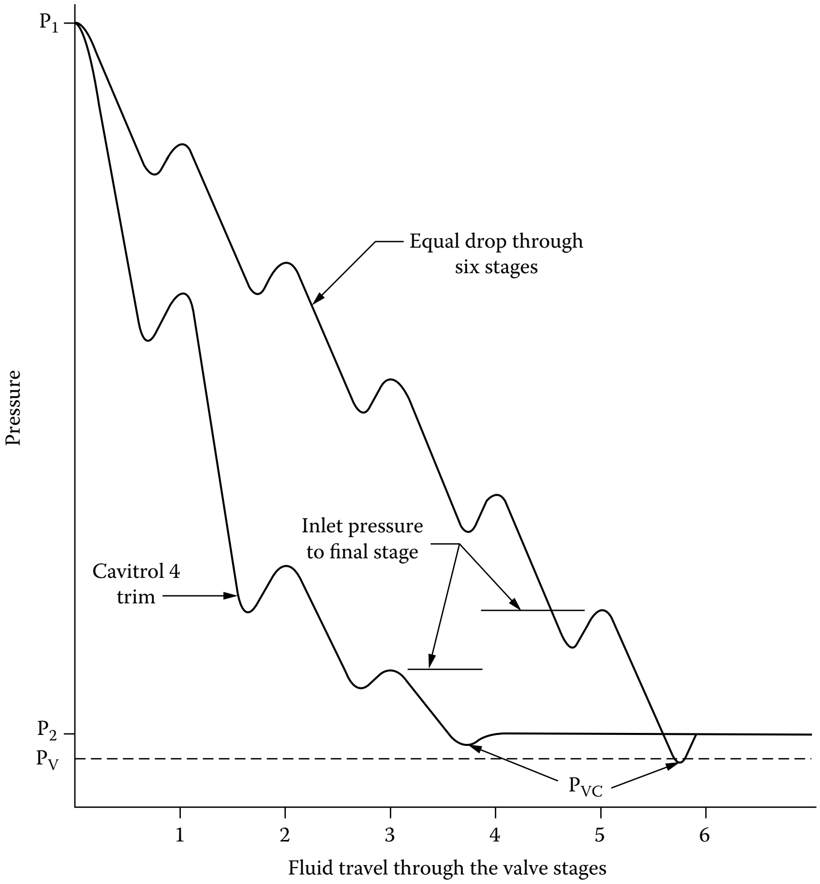


FIG. 6.13n
Comparison of staged pressure drops to prevent cavitation. (Courtesy of Fisher Controls.)

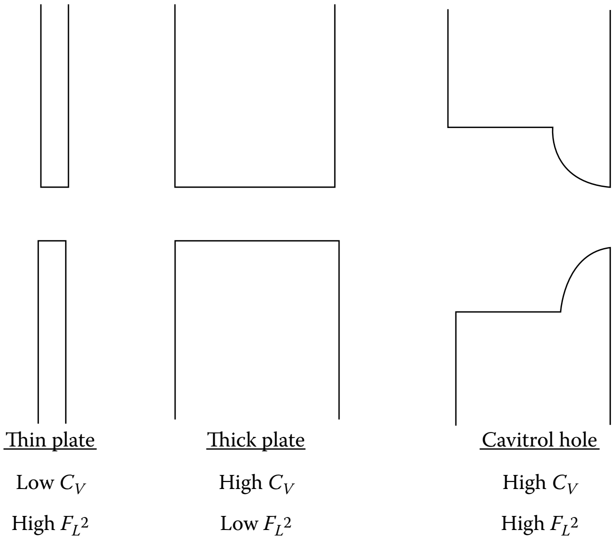


FIG. 6.13o
Comparing various drilled-hole-type anticavitation trim designs. (Courtesy of Fisher Controls.)

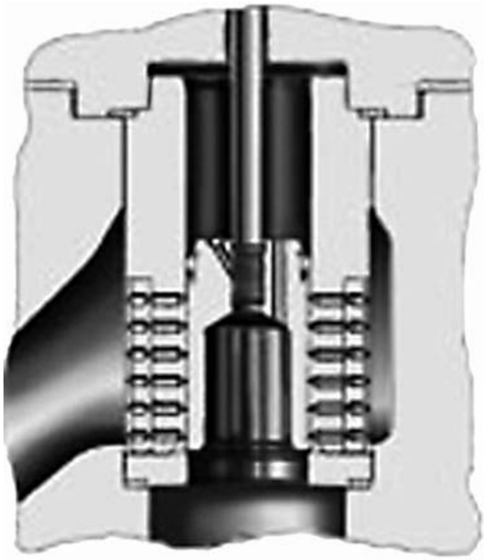


FIG. 6.13p
The cross-section of a three-stage anticavitation trim. (Courtesy of Fisher Controls.)

this trim design is the greatest at low travels and decreases as the valve plug travel increases.

In Figure 6.1y, a large number of anticavitation valve designs are shown. One disadvantage of most anticavitation trims is their potential for plugging. The flowing media often contains small particulate matter (e.g., sand) that can plug the passages, restricting or totally stopping flow through the valve. If this potential exists, the particles must be removed from the flow stream (e.g., by filtration), or an alternative approach to cavitation prevention should be taken.

In situations where it is impossible to remove the particulates via filtration or separation, valve designs that can pass the particulate and still resist cavitation and erosion should be considered. The next paragraph describes some of the control valves that can be considered for sludge or slurry services.

Dirty Process Services

In Section 6.1, Figure 6.1ee shows a number of control valve designs that can be considered for sludge and slurry services. These applications involving entrained particulates are some of the most challenging and can be found in all industries where entrained particulates are present in the process stream, which can cause extensive erosion to the valve trim and valve body.

Often these applications also involve high pressure drops that create the added potential for cavitation, flashing, and excessive noise and vibration. It is important to address all of these damage mechanisms in combination.

The most common valve selection for process streams containing entrained particulates is a rugged valve design utilizing erosion-resistant body materials and internal valve body liners. This brute force approach often does not incorporate any method of staging the pressure drop and therefore can cause valve and pipe vibration, if cavitating conditions exist.

Sweep-Angle and Rotary Ball Valves Figure 6.13q is an example of a valve that relies on "brute force" in being able to operate under dirty service conditions. This valve was initially designed for liquid asphalt service, but has also been successfully applied in separator let-down and slurry let-down applications.

This valve utilizes a venturi-type throat design that directs the fluid into the center of the valve and downstream piping. This prevents impingement of the fluid on the valve body and downstream piping. The expanded valve outlet reduces the velocity of the exiting fluid, thereby reducing any associated erosion effects. This is especially important in applications that experience flashing or out-gassing, where there may be some areas of high velocity.

The cylinder-guided valve plug ensures excellent controllability by providing plug stability, which results in better, more controlled flow of the fluid through the valve and reduced wear of the trim components.

FIG. 6.13q
Type 461 Sweep-Flo Valve by Fisher Controls.

For high-temperature applications, up to 1100°F (593°C), the valve body can be specified with an extension bonnet to reduce conduction of heat to the valve packing and the actuator. Connections on the side of the bonnet are available to flush particulate out of the extension bonnet cavity and back into the flowstream as the valve opens.

Another brute force selection for applications with lower pressure drops is to install a rotary ball valve. The plug in the rotary valve should be facing towards the downstream side of the valve body, so that the pressure drop will occur downstream of the valve body. However, with this approach it is important to provide either a durable downstream pipe liner or a sacrificial pipe section.

As rugged as these sweep-angle and ball valve designs are, they still can be exposed to conditions that cause excessive noise and vibration. This is because the process fluid is not staged through the valve trim in a manner that prevents the onset of damaging cavitation and the resultant noise and vibration.

Dirty Service Trim Another valve design that is available for dirty service applications (but not for heavy sludge, slurry services, or particle sizes exceeding 0.5 in., because of plugging) incorporates an axial flow path through a series of restrictions. These restrictions divide the pressure drop into stages, reducing the potential for cavitation and subsequent noise generation. Figure 6.13r shows one axial flow multistage trim design.

The dirty service trim (DST) trim eliminates damaging cavitation and resultant noise and vibration by staging the pressure across a properly determined number of stages. The

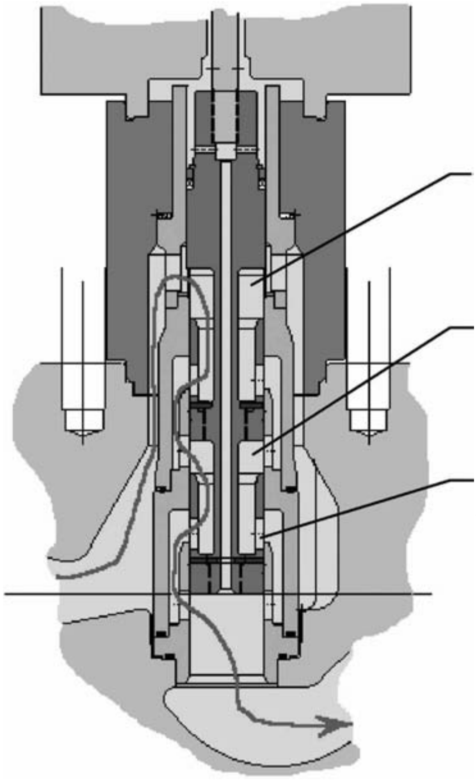


FIG. 6.13r
Dirty service trim (DST) design of Fisher Controls.

number of stages selected is dependent upon the pressure drop of the application and is designed in the same manner as was shown in Figure 6.13n.

The higher the pressure drop, the more stages become necessary. This trim design can pass particulates up to $^3/_4$ in. in diameter without plugging. The large open flow paths and expanded area staging design also compensate for volumetric expansion in flashing fluids, thus reducing velocities in the trim and the downstream piping.

Also included in this DST design is a protected seat, because the shut-off function of the trim is separate from its throttling areas. This separation is accomplished by not allowing any significant pressure drop to be taken until the fluid is downstream of the seating surface. This type of design also ensures that all clearance flow is subjected to a staged pressure drop. Unlike in linear cage-style, anticavitation trim sets, here no segment of the process flow can drop directly from the upstream (P1) to the downstream (P2) pressure after the valve.

High Noise

Section 6.14 covers all aspects of control valve noise, and Figures 6.14o to 6.14s show some low noise control valve designs from a variety of suppliers.

In many cases when the process fluid remains contained by the valve and piping, the noise generated becomes airborne only by its transmission through the valves and the adjacent piping. The sound field in the flow stream forces these solid boundaries to vibrate. These vibrations cause disturbances in the ambient atmosphere that are propagated as sound waves.

FIG. 6.13s
WhisperFlo noise abatement trim by Fisher Controls.

There are several ways to eliminate excessive valve noise (see Section 6.14), but this section will focus mainly on source treatment methods, which are obtained through specially designed control valve trim. There are many types of control valve trims that were designed to reduce valve noise, but there are only two ways to accomplish this reduction.

One method is referred to as frequency shifting, while the other is velocity control. The former approach, which has been used for many years, incorporates small, properly sized and spaced passages that raise the frequency of the fluid exiting the trim. The higher frequency reduces acoustic energy in the audible range by relying on the transmission loss of the piping. The latter approach (velocity control) can be accomplished in many ways, and the highest level of noise attenuation is usually accomplished by incorporating both methods.

Figure 6.13s shows the cross-section of a multistage noise reduction trim that utilizes a combination of noise reduction strategies and reduces valve noise by up to 40 dB.

Some of the features of this design include

- Unique passage shape reduces the conversion of total stream power generated by the valve into noise power.
- Multistage pressure reduction divides the stream power between stages and further reduces the acoustic conversion efficiency.
- Frequency spectrum shifting reduces acoustic energy in the audible range by capitalizing on the transmission loss of the piping system.
- Exit jet independence avoids noise regeneration due to jet coalescence.
- Velocity management is accomplished with expanding areas to accommodate the expanding gas.
- Complementary body designs avoid flow impingement on the body wall and secondary noise sources.

As explained in detail in Section 6.14, the amount of noise that will be generated by a control valve can be predicted quickly and reasonably by use of industry-standard methods. In order to obtain accurate noise predictions, it is important to utilize the standards described in Section 6.14.

High-Capacity Valves

Globe-style valves larger than 12-in., ball valves over 24-in., and high-performance butterfly valves larger than 48 in. are considered to be special valves. As valve sizes increase, shut-off static pressure loads also increase. Consequently, shaft strength, bearing loads, unbalance forces, and available actuator thrusts all become more significant with increasing valve size.

Noise levels must also be carefully considered in all large-flow installations because sound pressure levels increase in direct proportion to the magnitude of flow. To keep valve noise within tolerable limits, large cast or fabricated valve body designs have been developed. These bodies, which are normally cage-style constructions, use very long plug travel, a great number of small flow openings through the wall of the cage, and an expanded outlet line connection to minimize noise generation and to reduce fluid velocity.

With the increase in valve plug travels, the selection of the actuator used becomes more important. Typically, long-stroke, double-acting pneumatic piston actuators are selected. With these types of actuators, the accessories required to move the actuator also become more complex.

Cryogenic Valves

Cryogenic applications are those with temperatures that fall below –150°F (–110°C). Globe-style valves in cryogenic services are used in both cold box applications and non-cold box applications.

Cold boxes are commonly found in the air separation industry. Valves used in these applications feature bodies with welded extension necks and standard length bonnets to allow in-place trim maintenance from outside the cold box. For non-cold box applications, a cryogenic valve with an extension bonnet is used.

For control valves in cryogenic service, the correct selection of plastic and elastomeric components are important, because they often cease to function properly at temperatures below 0°F (–18°C). In these low-temperature ranges, components such as packing and plug seals require special consideration. For plug seals, a standard soft seal will become very hard and less pliable, thus not providing the shut-off that is required. Special elastomers have been applied in these temperatures, but require special loading to achieve a tight seal.

Packing is a concern in cryogenic applications because of the frost that may form on valves. Moisture from the atmosphere condenses on surfaces where the temperature is below freezing and will freeze into a layer of frost. As this frost and ice forms on the bonnet and stem areas of control valves and as the stem is stroked by the actuator, this layer of frost is drawn through the packing, causing tears and a loss of seal.

The solution is to use an extension bonnet (Figures 6.1t, 6.1u, 6.13t, 6.19p, and 6.19q) that allows the packing box area of the control valve to be warmed by ambient temperatures, thus preventing frost from forming on the stem and packing box areas. The length of the extension bonnet depends upon the operating process temperature and on the insulation requirements.

FIG. 6.13t
The construction of a cryogenic valve with extended bonnet. (Courtesy of Fisher Controls.)

When testing cryogenic valves for shut-off or in hydrostatic tests, the use of water-based tests should be avoided. If water tests are conducted, it is possible that moisture can be trapped inside the body or extension bonnet, which could ultimately form ice in the valve after it is cooled down. For these types of applications, in order to prevent freezing, the proper test medium is usually helium.

High-Temperature Valves

Valves that operate at temperatures above 450°F (232°C) experience many of the same limitations as do cryogenic valves. At elevated temperatures, the standard materials of control valve construction might be inadequate. For instance, plastics, elastomers, and standard gaskets generally prove unsuitable and must be replaced by more durable materials. Metal-to-metal seating materials are always used. Semimetallic or laminated flexible graphite packing materials are commonly employed, and spiral-wound stainless steel and flexible graphite gaskets are necessary.

It is important to select the trim materials and valve designs that will not experience sticking and gasket failure due to thermal expansion. If dissimilar trim materials are used, it is possible that one part will react to the high temperature faster than another, causing the components to gall. It is also important to allow the trim to grow axially in the valve body. Hanging the cage element from the top of the valve body does this. Some designs also incorporate internal springs or load rings to allow for the thermal expansion of the trim.

Similar to cryogenic applications, extension bonnets are used to protect the packing box parts from extremely high temperatures. For the selection of metallic and packing materials (Figure 6.1o) and the use of jacketed valve designs (Figures 6.1p to 6.1r), refer to Section 6.1.

Steam Conditioning Valves

Steam conditioning applications are examples of services where valves are exposed to high temperatures. These valves serve the function of simultaneously reducing the steam pressure and temperature to the level required for a given application. Frequently, these applications deal with high inlet pressures and temperatures and require significant reduction of both. They are, therefore, best manufactured as a forged or a fabricated body that can better withstand steam loads at elevated pressures and temperatures.

Forged materials permit higher design stresses, improve grain structure, and offer an inherent material integrity that is superior to cast valve bodies. The forged construction also allows the manufacturer to provide up to ANSI Class 4500, as well as intermediate and special class ratings, with greater ease vs. cast bodies.

Due to frequent extreme changes in steam properties as a result of the temperature and pressure reduction, the forged and fabricated valve body design allows for the addition of an expanded outlet to control outlet steam velocity at the lower outlet pressure. Similarly, with reduced outlet pressure, the forged and fabricated design allows the manufacturer to provide different pressure class ratings for the inlet and outlet connections to more closely match the adjacent piping.

Other advantages of combining the pressure reduction and desuperheating functions in the same valve versus two separate devices include:

- Improved spraywater mixing due to the optimum utilization of the turbulent expansion zone downstream of the pressure reduction element
- Improved rangeability
- Ease of installation and servicing of only one device

The manifold steam conditioning valve design (Figure 6.13u) is the most common form of steam conditioning valve available. This valve design offers all of the benefits of a combined valve, but features the ability to provide multipoint water injection utilizing an externally mounted manifold around the valve outlet. With this manifold, large quantities of water can be injected with a homogeneous distribution throughout the steam outlet flow.

Positioning of the valve plug within the control cage controls the steam pressure and flow. A signal from the pressure control loop to the valve actuator moves the valve plug within the control cage to increase or decrease the amount of free flow area. As the plug is lifted from the seat, steam passes into the center of the cage and out through the cage element.

The outlet section is outfitted with a water supply manifold. The manifold provides cooling water to a number of individual spray nozzles installed in the outlet section, which provide a fine spray mist that is injected radially into the highly turbulent stream of flowing steam.

FIG. 6.13u
The design of a steam conditioning valve, which is provided with an integral cooling manifold. (Courtesy of Fisher Controls.)

The combination of having a large surface area contacting the water and the high turbulence makes the mixing efficient and the vaporization rapid. For this control system an external water control valve is required, which is throttled by a downstream temperature controller to provide the required fine-tuning of the temperature control.

These types of valves are most commonly used in power (utility, cogeneration, and industrial) plant applications, which include the start-up of steam turbines, and the bypassing, dumping, venting, and exporting of steam.

Tank-Mounted Valves

Tank-mounted control valves are commonly found in the chemical industry. These are reverse acting (push down to open) control valves designed for handling corrosive fluids. The throttling element of the valve is positioned at the valve outlet and is placed inside the receiving tank so that flow exits directly into the vessel. In flashing applications or those with entrained gas, the erosion potential is greatly reduced.

Bibliography

Ball, K. E., “Final Elements: Final Frontier,” *InTech,* November 1986.

Carey, J. A., “Control Valve Update,” *Instruments and Control Systems,* January 1981.

"Control Valves—Globe, Plug, Pinch, Needle, Gate," *Measurements and Control,* February 1993.

"Control Valves, Regulators," *Measurements and Control,* June 1994.

Control Valve Handbook, 4th edition, Fisher Controls, Inc., 2003.

Control Valve Sourcebook, 3rd edition, Fisher Controls, Inc., 2003.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems,* January 1980.

O'Connor, J., "The Turbine Control Valve," *Instrumentation Technology,* December 1973.

Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

The Sourcebook of North American valve, Actuator and Control Manufacturers, 6th edition, 2004, Valve Manufacturers Association of America.

Wilson, J., "The Impact of Valve Outlet Velocity on Control Valve Noise and Piping Systems," *Valve World,* October 2001.

Wilson, J., "Tight Shutoff in Boiler Feedwater Control Valves," *Valve World,* April 2002.

6.14 Valves: Noise Calculation, Prediction, and Reduction

H. D. BAUMANN (1970) **J. B. ARANT** (1985) **B. G. LIPTÁK** (1995) **F. M. CAIN** (2005)

Valve Noise Types:	Mechanical vibration (usually below 100 dBA); hydrodynamic caused by liquid turbulence, cavitation, or flashing (usually below 110 dBA); aerodynamic (can reach 150 dBA)
Sizes:	1 to 24 in. (25 to 600 mm) in standard bodies; sizes above 24 in. in special castings or weldment fabrications
Design Pressure:	Up to ANSI Class 2500 (PN 420) standard; above Class 2500 in special designs
Materials of Construction:	Any machinable wrought or cast metal for body and trim approved for use in valves or pressure vessels
Special Features:	Balanced plugs, special seal designs, hard facings, piloted inner valve, characterized flow, dual (high/low) operating conditions, multistage trims
Cost:	Highly variable depending upon type of design, size, metallurgy, special features. Range may be from 2 to 10 times equivalent standard valve
Partial List of Low-Noise Valve and Diffuser/Silencer Suppliers:	ABB Control Valves (www.abb.com/controlvalves) Control Components Inc. (www.ccivalve.com) Dresser Flow Solutions (www.masoneilan.com) Emerson Process Management (www.emersonprocess.com/home/products) Flo-Dyne Limited (www.flo-dyne.net) Flowserve Corporation (www.flowserve.com/valves) GE-Nuovo Pignone (www.gepower.com/prod_serv/products/valves/en/std_control.htm) Industrial & Marine Silencers Ltd. (www.silencers.co.uk) Koso Hammel Dahl (www.rexa.com/hammel_dahl/hammeldahl_index.htm) McGuffy Systems, Inc. (www.mcguffy.com/products/manufacture/silencers/silenc.html) Metso Automation (www.neles.com) Samson AG (www.samson.de) SPX Valves and Controls (www.dezurik.com) Tyco Flow Control (www.tycovalves.com) Weir Valves & Controls (www.weirvalve. com) Welland & Tuxhorn (www.welland-tuxhorn.de)

INTRODUCTION

This section begins with an overview of general noise principles, followed by a description of the types of noise produced by fluid flow through control valves. The discussion of control valve noise mitigation includes both the treatment of the noise source (modifying the valve) and the treatment of the noise path (providing downstream insulation or silencers). Other options include protection of the receiver (by personal protective equipment such as earplugs or earmuffs) or the removal of the receiver (by placing a barrier or distance between the noise source and personnel). The section ends with a discussion about recent improvements in predicting and calculating probable noise levels.

Because most valve noise calculation standards avoid excessive detail, only the SI system of units will be used in this section. Users of U.S. Customary units should refer to Appendix A.1 and A.2 for the proper conversion factors, including gravitational units conversions (i.e., g_c) when necessary.

SOUND AND NOISE

A weed has been defined as an unwanted plant or flower. As an environmental analogy, noise may be considered as an unpleasant or unwanted sound. Sound, in the context of this discussion, is defined as pressure fluctuations generated in the air or other medium, which are capable of stimulating the physiological hearing response of the human ear and brain. For ease of understanding, we will hereafter refer to *sound* and *noise* as equivalent terms.

Most common sounds are a complex mixture of many frequencies at varying magnitudes. Pure tones have discrete frequencies. It is customary to model sound as pressure waves with sinusoidal characteristics such as frequency (f), magnitude (p), wavelength (λ), and speed (c). Of course, sound waves possess other more complex characteristics that are beyond the scope of this topic.

Frequency is expressed in cycles per second (cps) or Hertz (Hz), where cps and Hz are equivalent units. The magnitude of sound pressure is measured in units of pressure (Pascal in the SI system). The range of sound pressures that humans can discern from the threshold of hearing to the threshold of pain spans over 12 orders of magnitude! Therefore, it is more convenient to use a logarithmic comparison of an actual sound pressure to a standard pressure reference at the threshold of hearing and to define this comparison as a *sound pressure level*, L_p, expressed by Equation 6.14(1) in decibels (dB).

$$L_p = 10\log_{10}\left(\frac{p}{p_o}\right)^2 = 20\log_{10}\left(\frac{p}{p_o}\right) \qquad \mathbf{6.14(1)}$$

where p is the actual sound pressure, and the reference sound pressure, p_o, is defined as 2×10^{-5} Pascal (2×10^{-4} microbar or 29×10^{-8} psi.). Because the decibel is a logarithmic function, for every 10 dB increase, there is a tenfold increase in sound intensity. Thus, a 100 dB sound is 10 times as intense as 90 dB and 100 times as intense as 80 dB. However, the human ear perceives each 10 dB increase as an approximate doubling of loudness.

The sound pressure fluctuations must be generated by some energy source that transfers power into the air or other wave-conducting medium. (Sound waves cannot travel in a vacuum.) The total acoustic power created by the noise source is defined as sound power, W_a, usually expressed in watts (W).

The calculation of sound power will be used in this section to predict sound pressure levels in valve applications. It is worth remembering that, while sound is produced by a power source, it is sound pressure that the ear perceives. Sound power can also be presented as a sound power level, L_w, in decibels by logarithmic comparison with the standard reference level, W_o, of 10^{-12} W.

$$L_w = 10\log_{10}\left(\frac{W_a}{W_o}\right) \qquad \mathbf{6.14(2)}$$

Wavelength, λ, is the distance required for one complete pressure cycle.

Speed of Sound

The speed of sound, c, in any medium is a function of its mass density and elastic properties.

For a solid:

$$c = \sqrt{E/\rho} \qquad \mathbf{6.14(3)}$$

where E is the elastic modulus and ρ is the mass density.

For carbon steel pipe at 100°C, E = 198 GPa, ρ = 7.86 g/cm^3, and c = 5020 m/s. For CrMo steel alloy pipe at 100°C, E = 207 GPa, ρ = 7.84 g/cm3, and c = 5140 m/s. Austenitic stainless steel pipe (UNS S30400) at 100°C with E = 190 GPa and ρ = 8.03 g/cm^3 has c = 4860 m/s. So, for purposes of estimating the speed of sound in steel pipe, using 5000 m/s usually produces satisfactory results.[1]

For a liquid:

$$c = \sqrt{E_s/\rho} \qquad \mathbf{6.14(4)}$$

where E_s is the isentropic bulk modulus. It can be shown that at 20°C, speed of sound in fresh water is 1481 m/s, in seawater 1521 m/s, and in machine oil (sp. gr. = 0.90) 1297 m/s.

For a gas or vapor:

$$c = \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma RT}{M}} \qquad \mathbf{6.14(5)}$$

where γ is the ratio of specific heats, R is the universal gas constant (8 × 314 J/kmol × K), T is absolute temperature (Kelvin), and M is molecular mass of the fluid. Using this relationship, we find that the speed of sound in air at 0°C (273°K) is 331 m/s.

Wavelength, frequency, and speed of sound are related as shown in Equation 6.14(6).

$$c = \lambda f \qquad \mathbf{6.14(6)}$$

THE HUMAN EAR

The human ear is an intricate acoustic instrument that is described here in only general terms. The anatomy of the ear is divided into three major regions, each with unique functions: the outer ear, middle ear, and inner ear. The outer ear consists of the pinna, ear canal, and outer layer of the eardrum. It channels sound waves to the eardrum, where sound pressure waves are converted into mechanical energy by vibrating the eardrum.

---

[1] The most recent valve noise calculation standard and the field of acoustics in general use SI units. Users of U.S. Customary Units are cautioned to use the proper gravitational units conversions (i.e., g_c) when necessary. To avoid excessive detail, only the SI system of units will be used in this section.

The middle ear is an air-filled cavity containing the ossicles (bones) that connect to the oval window to the inner ear. The middle ear cavity is also connected to the Eustachian tube, which equalizes static pressure across the eardrum. The middle ear mechanism acts as an impedance-matching transformer. It is matching the impedance of the air in the ear canal to the impedance of the liquid of the inner ear.

The inner ear vestibule leads to the semicircular canals (providing sense of balance) and the "snail-shaped" cochlea, where the final energy transformation occurs. In the cochlea, mechanical energy is conducted through a traveling wave pattern on the basilar membrane, causing a shearing of the cilia of the outer and inner hair cells of the Organ of Corti. The design and stiffness gradient of the basilar membrane allow more efficient response to higher frequencies at the basal end, and progressively lower frequencies are detected along the membrane toward its apex.

The Organ of Corti is the sense organ that changes vibration energy into neural energy. This conversion takes place as shearing stress on hair cells induces a depolarization that generates neural impulses. The neural impulses are conducted by the auditory nerve to the brain, where they are processed and interpreted as sound. Damage to the hair cells connecting the basilar membrane and Organ of Corti usually produces permanent loss of hearing. Damage or deterioration can occur from sudden, loud noise (explosions), excessive exposure to moderately loud noise (industrial environments, loud music), physical injury (head trauma), advancing age, infections, or disease.

Loudness Perception

A healthy, young adult human is able to perceive sound over a wide range of frequencies from approximately 20 to 18,000 Hz, which is generally accepted as the audible range. The human ear, however, does not give equal weight (loudness perception) to the same sound pressure level across the frequency spectrum.

Studies of apparent loudness by many human subjects over the frequency spectrum when compared to a pure tone of 1000 Hz frequency has resulted in mapping the ear response. The loudness level in phons represents the pressure level in dB of a 1 kHz tone that a typical hearer feels is as loud as the sound in question. Figure 6.14a shows the loudness level map as function of frequency.

We can see from Figure 6.14a that a sound at 1000 Hz and 50 dB sounds equally loud as 67 dB at 100 Hz or 62 dB at 10 kHz. The resulting correction numbers, which are approximating the response of the human ear, are called "A" weighting. The corresponding decibel level is indicated as dBA, as shown in Figure 6.14b.

There are other weighting schemes for various purposes, but A weighting is used in governmental regulations on noise pollution. Hence, for the discussion of valve noise levels and environmental noise reduction, we will use the dBA scale. Noise levels of some common environmental sounds are compared in Table 6.14c.

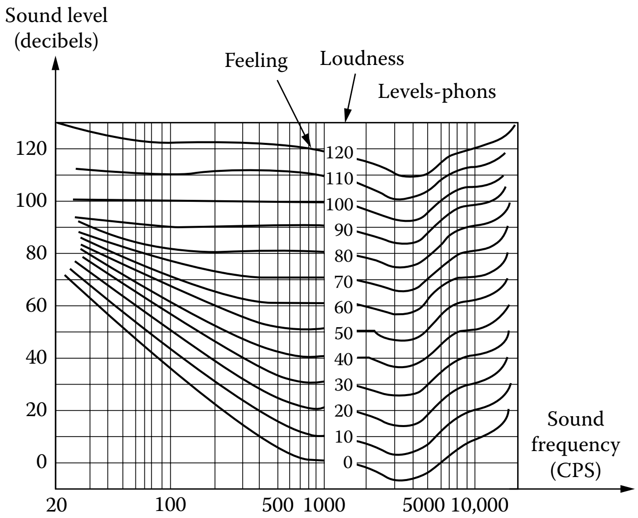


FIG. 6.14a
Apparent loudness contours for human hearing.

Limiting Valve Noise

There are several important reasons to limit the noise levels emitted by valves and piping. One of them is to prevent the harmful effects of environmental noise pollution, which includes hearing loss in people. As was noted earlier, we can tolerate much louder sounds at low and at very high frequencies than we can in the middle of the spectrum. This is represented in the A-weighting curve of Figure 6.14b.

Note that in the 500–7000 Hz range, the human ear is most responsive, and this is the area where high noise level exposure can do the most damage. For this reason, the U.S. government enacted the Occupational Safety and Health Act of 1970 (amended in 1998), establishing the Occupational Safety and Health Administration (OSHA). OSHA regulations limit a

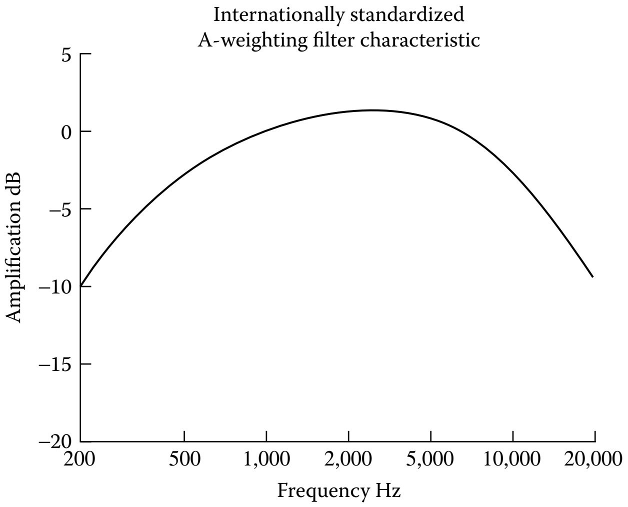


FIG. 6.14b
The A-weighting filter characteristic approximates the human ear's response to different sound frequencies.

TABLE 6.14c
Approximate Sound Pressures Levels of Typical Sounds

Source of Sound	*L_p (dBA)*
Near jet engine; artillery fire	140
50 hp victory siren at 30 m; threshold of pain	130
Rock-and-roll band; threshold of feeling	120
Jet flying overhead at 300 m	110
Air chisel; high pressure gas leak	100
Motorcycle at 15 m; subway train at 6 m; symphony	90
Inside sports car (100 km/h)	80
Loud conversation; noisy business office	70
Normal conversation; light traffic at 30 m	60
Private business office; normal conversation	50
Quiet conversation	40
Quiet home at night; still forest; soft whisper	30
Empty theater; rustling leaves	20
Inside a soundproof room; quiet breathing	10

weighted 90 dBA maximum level exposure to 8 hours per day. Table 6.14d below shows general exposure time limits established by OSHA.

Figure 6.14e shows a typical frequency octave band noise level contour that will meet this limit. Note that if the predominant noise frequency exposure is in the critical middle frequency range of 1000–5000 Hz, the allowable weighted noise level over 8 hours would be considerably less than 90 dBA.

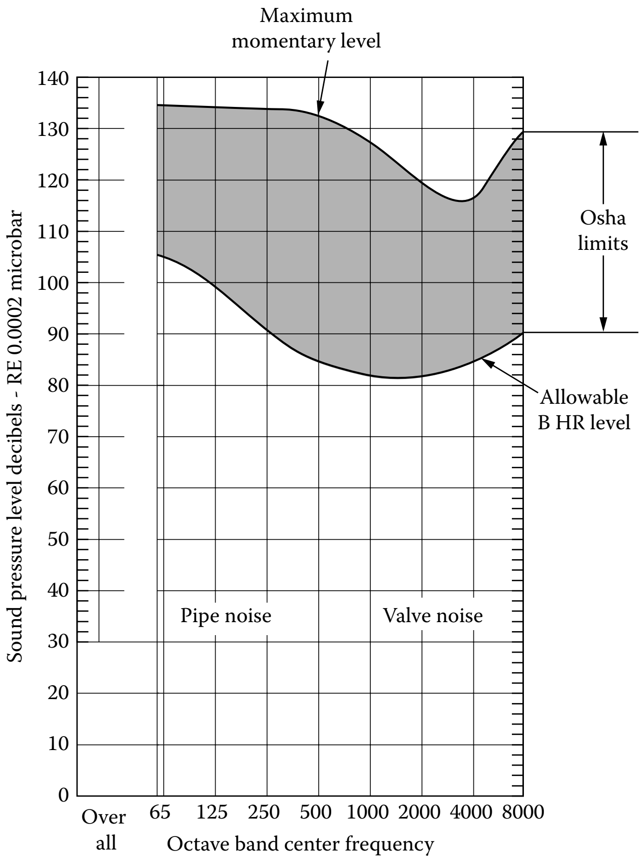


Fig. 6.14e
Frequency octave band noise level contours, which will result in the weighted average exposure to meet OSHA limits.

VALVE NOISE

While there are many noise sources in industrial and process plants, some of the main contributors can be control valves operating under conditions of high pressure drop. These are one of the few and sometimes the only sources of over 100 dBA sound levels found in process plants. To gain some perspective of how loud 100 dBA actually is, refer to Table 6.14c for a comparison of common environmental sounds.

However, even if people are removed from areas with high noise levels, other hazards are still created by excessive noise. High intensity noise can produce vibrations in structures, which become magnified when the natural (resonant) frequencies within the structure are close to the dominant frequency of the noise.

TABLE 6.14d
OSHA Exposure Time Limits for Various Noise Levels

Hours per Day	*dBA*
8	90
4	95
2	100
1	105
$^1/_2$	110
$^1/_4$	115 (max.)

Even without resonance, studies by Fagerlund have shown that the sound power that is produced downstream of valves can cause fatigue failures in valves and piping systems. This can occur when noise levels outside the pipe (1 m downstream of a valve and 1 m away from the pipe wall) exceed 110–115 dBA depending on pipe size.

Table 6.14f provides a summary of the likely causes of noise in valves and of their frequencies.

The five major sources of noise generated by control valves are as follows:

- Mechanical vibration
- Control element instability
- Resonant vibration
- Hydrodynamic noise
- Aerodynamic noise

Mechanical Vibration Mechanical vibration of valve internal parts is caused by unsteady flow and turbulence within the valve. It is usually unpredictable and is really a design

TABLE 6.14f
Sound Frequencies and Sources in Valves

Frequency (Hz)	*Octave Band Number*	*Sound Description*	*Typical Noise Source in Valves*
20–75	1	Rumble	Vertical plug oscillation
75–150	2		Cavitation*
150–300	3	Rattle	
300–600	4	Howl	Horizontal plug vibration
600–1200	5		
1200–2400	6	Hiss	Flowing gas
2400–4800	7	Whistle	
4800–7000	8	Squeal	Natural frequency vibration
20,000 and up		Ultrasonic	

* Cavitation frequencies vary widely from about 100 Hz to 15 kHz depending on valve and trim design.

problem for the manufacturer. Noise levels are typically low, usually well under 90 dBA, and in the 50 and 1500 Hz frequency range.

The problem is often not the noise, but the progressively worsening vibration as guides and parts wear. The solution can be in improving the valve design by adding heavy-duty stems and guides. Improvements in design may also include small changes in the flow path geometry of the trim, which can also eliminate some vibration problems.

Control Element Instability

Control element instability is usually due to mass flow turbulence impingement on the valve plug. The relationship between velocity and static pressure forces acting across the plug or disc face and the actuator force balance varies over time. Without sufficient stiffness in the actuator, valve, and mechanical connections, fluid buffeting forces may produce vertical stem oscillations in linear valves and torsional shaft oscillations in rotary valves, resulting in low-level rattle noise usually under 100 Hz. This instability is detrimental to control.

Correction requires changing the damping characteristics of the valve and actuator combination. This is done by providing a stiffer valve actuator or eliminating mechanical backlash. If the actuator is a spring-and-diaphragm type, then one can increase the nominal spring rate from 20–100 kPa (3–15 PSIG) to 40–200 kPa (6–30 PSIG).

For single-acting piston actuators, the cushion air loading can be increased. If these changes do not solve the problem, then either actuator can be replaced with double-acting air piston actuators, which are generally stiffer and allow use of higher air pressures. In extreme cases, a hydraulic snubber, an all-hydraulic actuator, or electromechanical actuator may be required.

Resonant Vibration

Resonant noise is characterized by a discrete tone and possibly a few harmonic multiples. Resonance can involve merely an acoustic interaction within the valve and piping geometry, with certain frequencies of the flow turbulence. Resonant frequencies from 200 Hz to 10 kHz can be excited acoustically by the flow in the same way that tones are produced in musical instruments.

Localized metal fretting or wear is likely on internal valve parts. In some cases, turbulence and acoustic resonance excites mechanical or structural natural frequencies, producing severe vibration capable of causing damage to piping, equipment, and supporting structures. Resonant noise levels exceed calculated predictions based on current prediction standards and may be in the 90–125 dBA range.

Work by Glenn refers to this discrete resonance as *screech*, and his research shows that screech is possible at pressure drops lower than required to produce sonic flow in gases or cavitation in liquids. It is possible for conditions to exist that produce screech-type resonance in virtually any valve type and brand. Glenn identifies several possible causes of screech in valves:

Higher than expected velocities in the valve, due to uncertainties in the pressure recovery characteristics as a function of valve opening. (Refer to the discussion of the pressure recovery factor F_L in Section 6.15.)
Excitation of harmonic pipe modes of vibration
Flow instabilities, due to
- Vortex shedding
- Tollmien-Schlichting waves in the laminar-to-turbulent flow transition
- Bi-stable flow separation
- Unstable shock waves
- Unstable vapor-liquid interface

Two approaches to solving these problems include 1) modifying the design of the flow path to change the characteristics of the turbulence, or 2) changing the stiffness and resonant frequencies of the valve trim.

Valve trims can often be modified by a change in stem diameter, change in the plug mass, or method of guiding. Flow paths can be modified sometimes by reversal of flow direction through the valve or by minor design changes to seats, plugs, or cages. These changes shift the natural frequency of the plug and stem out of the excitation range of the flow turbulence, or vice versa.

Valve manufacturers and users should collaborate to implement effective solutions in valves and piping when these problems arise. For example, one investigation identified the source of serious screech noise as a gap between valve and pipe flanges created by an oversized inside diameter of the gasket. Filling the gap with a properly sized gasket eliminated the problem.

Hydrodynamic Noise

Hydrodynamic noise is generally less troublesome and less severe than aerodynamic noise and usually only becomes excessive when accompanied by cavitation or flashing (discussed in Sections 6.1 and 6.15). Severe cavitation can produce noise in the range of 90–100 dBA or higher.

Problems with cavitation or flashing are usually avoided by use of a suitable trim or valve type with low-pressure recovery characteristics (high F_L). Noise caused by cavitation is the result of imploding vapor bubbles in the liquid stream. This noise can vary from a low-frequency rumble or rattling to a high-frequency squeal. This latter condition is due to acoustic or pipe resonance with cavitating fluid.

In most cases, the problem is not so much with noise as it is the destruction of the valve trim and piping from erosion and pitting by the imploding vapor bubbles. Reducing or eliminating the cavitation and its damage also eliminates the noise. Single-stage multiorifice valves (see Figure 6.1z) and multistage valves (see Figures 6.1y and 6.1aa) are typical solutions to cavitation erosion and noise.

The sizing of any liquid service control valve should include an evaluation of the cavitation potential, with emphasis on eliminating or mitigating the cavitation. Section 6.15 outlines methods for predicting the onset of cavitation. The standards VDMA 24422 and IEC 60534-8-4 include methods for calculating hydrodynamic noise, but these methods have been shown by Kiesbauer and Baumann to predict lower than actual noise in many cases. At the time of this writing, work is under way in the International Electrotechnical Commission (IEC) to improve the accuracy of hydrodynamic noise prediction.

Flashing is rarely a significant source of valve noise, although it can cause valve trim erosion damage in some cases. Flashing produces increasing valve exit velocity and downstream piping velocity as a result of the higher specific volume of the two-phase flow. In cases where sonic flow and shock cells develop in downstream piping, excessive noise can result. Expanded outlet valves and larger downstream piping will be required under conditions where a large percentage of the liquid undergoes flashing. At this time there is not a standard method for predicting noise from flashing.

Aerodynamic Noise

In control valve design, aerodynamic noise can be a major problem. It is a category of valve noise capable of generating noise levels of 120 dBA or greater. Noise produced by fluid turbulence in liquids is almost negligible as compared to the noise generated by the turbulence and shock cells due to the high velocity of gases and vapors passing through the valve orifice.

The mechanisms of noise generation in valves and transmission through pipe walls are highly complex and are still not completely predictable. As a result of the many variables influencing noise generation and the need for simplifying assumptions in calculations, predicting the noise levels from valves or atmospheric exhaust vents is an inexact science. However, universities, manufacturers, and interested technical societies have made much progress, which has resulted in better noise prediction methods based on scientific fundamentals, which will be discussed in the section on Noise Calculations.

Aerodynamic noise generation, in general, is a function of mass flow rate and the pressure ratio (p_1/p_2) across the valve. The point at which sonic speed is reached in the valve vena contracta is a function of the valve design and its pressure recovery coefficient, F_L, combined with the ratio of upstream to downstream absolute pressure (p_1/p_2). For example, valves with F_L values of 0.5 and 0.95 require pressure ratios of 1.15 and 1.80, respectively, to generate sonic flow in the valve.

When sonic velocity is reached at the vena contracta, the valves are said to be *choked*, because their capacity does not increase if the pressure ratio is increased while the upstream pressure is kept constant. Generally, choked valves are the sources of the highest noise levels, but subsonic flows can also generate high noise levels. Valves that are not choked operate in a subsonic flow regime. For a given mass flow, they are less noisy than choked valves, but the noise level will increase as the pressure ratio approaches the sonic level.

Velocity of the flow in downstream pipe can also generate significant noise starting at pipe velocities of about Mach 0.4 to Mach 1.0 (sonic). Noisy gas or vapor control valves can have acoustically induced and turbulence-induced vibration damage, trim wear, and control instabilities. High-intensity noise can produce vibration-related stresses at very high cycle rates (1,000–10,000 cps). Hence, noise-induced damage can drastically reduce valve service life, and in some cases, it can cause valve or piping failures in a matter of minutes or hours.

CONTROLLING NOISE

The transmission of a noise requires a source of sound, a medium through which the sound is transmitted, and a receiver. Each of these can be changed to reduce the noise level. In cases when the noise is from vibrating control valve components, the vibrations must be eliminated or they might result in valve failure. In cases when the source of noise is the hiss of a gas-reducing station, the acoustical treatment of the noise is sufficient.

Depending upon the magnitude of the aerodynamic noise and assuming that massive valve damage is not a factor, valve noise treatment can be accomplished either by path treatment or source treatment. Valve damage can only be reduced or eliminated by source treatment, which minimizes or eliminates the damage mechanism.

There is no absolute rule that will enable one to choose between path or source treatment. However, in general, if the noise is under 100 dBA, then either a path or source treatment is a possible solution. Noise above 100 dBA almost always requires source treatment to successfully solve the noise problem.

The proper choice of noise treatment method is not always easy to select, but with the help of improved noise predictions through frequency spectrum evaluation and with expertise based on experience it can be obtained. Conservative solutions are preferred, because reworking or retrofitting in case of poor design is often very expensive.

Path Treatment

Path treatment, as its name implies, does not focus on changing the noise source. The intent of path treatment is to attenuate the noise transmission from the source to the receiver (ear). There are several common path treatments: the use of heavy wall pipe; installation of diffusers, mufflers, or silencers; and application of acoustical insulation.

Path treatment is not always a more economical solution than source treatment, and economics must be evaluated for individual applications. For existing installations, path treatment may be used, not because it is the best solution, but because it may be the only feasible one.

Pipe Wall Thickness Heavy wall pipe reduces noise by increasing the transmission loss through the pipe wall. The amount of attenuation depends on the stiffness and mass of the pipe. The mechanisms are complex and beyond the scope of this text. However, as a simple rule for rough estimation, each doubling of pipe wall thickness results in approximately 6 dBA more attenuation depending upon pipe size (attenuation increases with pipe size).

Refer to the works of Fagerlund and Chow (1981) and Reethof and Ward (1986) for important foundation in calculating transmission losses through pipe walls. Sample calculations will be introduced in the Noise Prediction section.

Insulation and Absorption Another method of increasing transmission loss at the pipe wall is the use of acoustic insulation. Even thermal insulation can add 3–5 dBA attenuation. Proper selection and application of 1–2 in. (25–50 mm) of a good acoustic insulation can reduce the noise level by roughly 10 dBA.

Certain types of insulation are more effective at specific frequency bands, so this information is important for proper selection. Because sound travels down the pipeline with very little attenuation over long distances, increasing the pipe wall thickness or applying acoustical insulation can be a very expensive solution. This approach is most useful when downstream piping runs are short.

The higher the frequency of vibration, the more effective are the commercially available sound absorption materials. Figure 6.14g gives an example of acoustical treatment for the outside of a pipe.

It is often beneficial to cover the inside walls of the building with sound-absorbing materials to prevent the reflection and radiation of the sound waves from process equipment. If a valve is installed close to a single reflective surface (e.g., a hard floor or wall), the apparent noise increases by 3 dBA; with two reflective surfaces, noise increases by 6 dBA; and for three nearby reflective surfaces, noise increases by 9 dBA. A valve installed in a small room with all reflective surfaces can elevate noise levels by 30–40 dBA. When using walls as sound barriers, it is important to seal all openings.

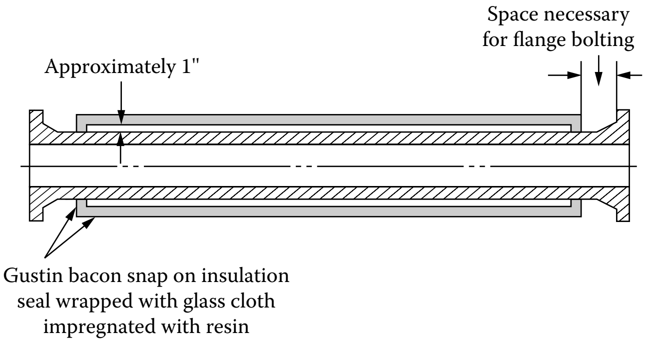


FIG. 6.14g
Acoustical treatment of pipe walls.

Isolation Locating a potentially noisy valve installation at a substantial distance from normal working areas may be effective and economical. If the valve can be located on top of a structure or pipe bridge, the distance attenuation can minimize the noise treatment required at the valve source and downstream pipe.

For example, instead of a control valve with a noise specification of 85 dBA, it may be possible to relax this to 90 or 95 dBA, which can considerably reduce the cost of the valve or noise treatment system. It is important to note that very little noise actually radiates from the valve itself, due to generally heavy wall thickness and rigidity of most valve bodies; downstream piping radiates the great majority of noise produced in the valve to the surroundings.

As a general rule, each doubling of a person's distance from the piping downstream of a valve will reduce the sound level by 3 dBA in a nonreflective environment. For example, the sound level that a person hears at 8 m from a valve will be 9 dBA quieter than the sound level at 1 m, and at 16 m it will be 12 dBA quieter than at 1 m.

Diffusers, Mufflers Diffusers located downstream of the valves (Figure 6.14h) can be helpful in both original installations and retrofit situations. These devices can aid in reducing exit flow turbulence or shock. Another important function of the multiple-hole design of diffusers depends on the fact that sound frequency increases as the size of the flow passage decreases. Using many small holes forces the dominant frequency of the turbulence into a higher range to which human hearing is less sensitive.

Diffusers can be designed to serve as pressure drop devices to reduce the pressure drop across the control valve

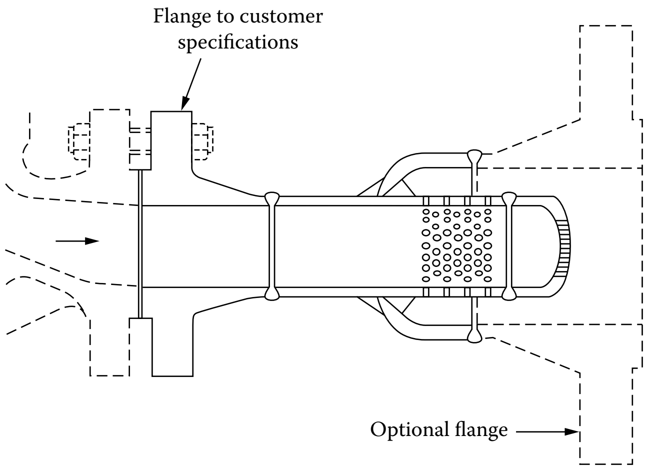


FIG. 6.14h
Acoustical diffusers are used to reduce the exit turbulence downstream of the valve. (Courtesy of Emerson Process Management.)

and thus reduce its noise generation. The valve and diffuser system works best in a situation when the flow rate is substantially constant or at least does not vary over a wide range. As a restrictor, the diffuser's effectiveness in generating backpressure on the valve decreases substantially as the flow rate drops. However, this shift of pressure drop back to the control valve does not necessarily increase noise, because when this occurs, the lower mass flow produces less noise.

Figure 6.14i illustrates a silencer design that can be installed downstream of a gas-regulating valve. Due to the resulting acoustical attenuation, it can reduce the sound pressure by a factor of five (e.g., from 96 to 82 dBA).

Mufflers or silencers (Figure 6.14j) can be used for in-line path treatment or for atmospheric vents. These are usually expensive devices, with the cost escalating dramatically with size.

Dissipative or dissipative/reactive silencers are most commonly used, but a comprehensive discussion of these devices is beyond the scope of this text. However, there are some rough guidelines for application. Inlet velocity must be subsonic, and the silencer cannot be sized to serve as a pressure reducer. An inlet diffuser (as shown in Figure 6.14j) can be helpful, because it breaks up turbulence or shock cell oscillations that often occur in downstream sound fields and reduce the effectiveness of the unit. The outer shell should have a thick enough wall to prevent resonance. Materials of construction are selected to meet process conditions and to retain absorptive materials.

Source Treatment

Source treatments reduce noise by limiting sound power generated at the source. In most cases, treatment consists of a special control valve and trim design, sometimes combined with a special diffuser or back-pressure element. While they may differ in concept, design, and manufacturing technologies, these special systems are designed for one or more of the following objectives:

- Reduce pressure drop in stages
- Limit fluid velocity to subsonic levels
- Reduce or eliminate the formation of high turbulence and shock cells
- Shift as much sound power as possible into higher frequency bands that have greater transmission losses in the pipe wall and have reduced response by human hearing

Depending upon the particular design, noise can be reduced with relatively inexpensive, simple elements by 7–10 dBA, whereas the more sophisticated valve designs or multielement systems can accomplish as much as 30–40 dBA attenuation from an untreated configuration.

Specification of source treatment valves or systems is not a simple matter. There are a number of design considerations, including the following:

- Application: in-line or vent
- Noise reduction required or maximum SPL allowed (dBA)

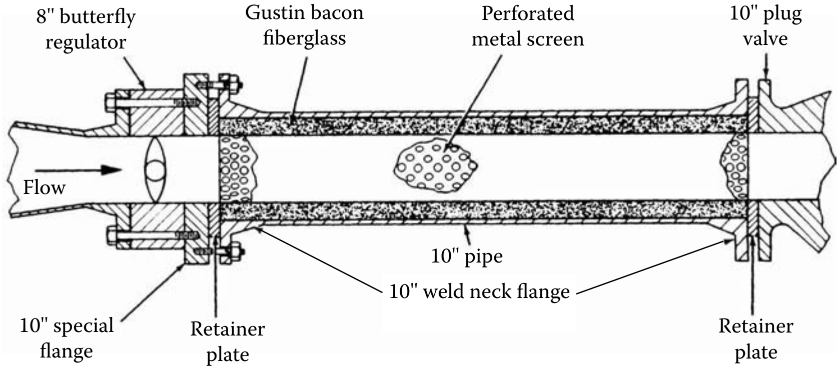


FIG. 6.14i
Silencer for gas regulating stations. (From King, C. F., "Control Valve Noise," Emerson Process Management.)

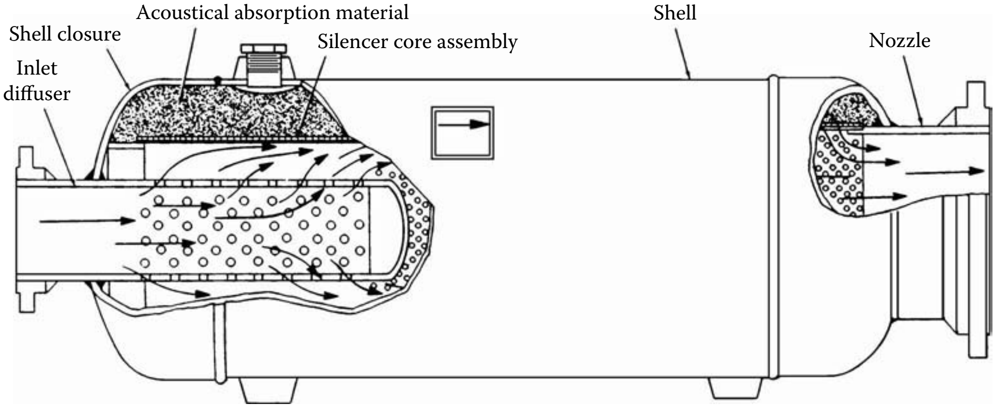


FIG. 6.14j
Silencers are installed in the flow path to dissipate the sound energy by absorbing it in an acoustical pack. They are designed to cause less than 1 PSI pressure drop. In-line silencers are often the most economical means of noise control in applications where the mass flow rate is high and the pressure drop is low. These units are normally installed immediately downstream of control valves, but in some cases they may also be required upstream of the valves. (Courtesy of Emerson Process Management.)

- Valve absolute pressure ratio: p_1/p_2 or $\Delta p/p_1$
- Pressure drop, Δp
- Fluid properties
- Temperature operating level and range
- Mass flow rate and turndown
- Metallurgy and mechanical design considerations
- Other potential velocity-induced problems
- Valve shut-off requirements
- Valve service life
- Valve location and orientation, piping arrangement, valve support, and maintenance access
- Actuation and control requirements
- Economics, including purchase, installation, and maintenance costs

The importance of each factor is a matter of judgment and experience, understanding all aspects of the application and plant operation. Thus it is up to the user to carefully weigh and evaluate all vendor proposals. If a vendor is using one of the standardized noise prediction methods, the user should verify that accurate input values were specified and used for the calculations. Vendors that offer proprietary noise prediction calculations should also offer empirical justification to support their noise prediction.

While initial cost is one factor, this is not the only consideration; it may be that the more expensive equipment is the most economical solution in the long run. Plant downtime and retrofit costs for deficient valve noise solutions are usually very expensive. With these caveats, there is much good equipment available, and vendor expertise and experience can be very valuable to the user with limited experience in controlling valve noise.

Basically, source treatment control valve designs fall into three categories: multipath, multistage, and combination of multipath/multistage. These designs are listed in order of sophistication and capability for noise reduction under severe operating conditions. As might be expected, cost also increases. The cost of these special designs tends to range from 2–20 times the cost of a standard valve with the same flow capacity.

Multipath Valves The multipath valve (Figure 6.14k) provides multiple orifices in parallel. A cylindrical plug to vary the flow rate through the valve uncovers these orifices. Although the path shape may vary by manufacturer, the principle consists of splitting the single-path flow into a large number of small paths. (See also the designs in Figures 6.1y and 6.1z.)

The noise radiated outside the pipe from the combined flow through multiple small paths is much less than that from the same flow rate through a single path restriction. Typical attenuation levels are 7–10 dBA but may reach 12–15 dBA in some applications. Variations of the multipath design are used for both hydrodynamic and aerodynamic valve noise with damage potential of low to moderate severity. Typically, compressible fluid pressure ratios (p_1/p_2) of 1.5–5 with valve exit velocities below Mach 0.33 are good candidates for this design.

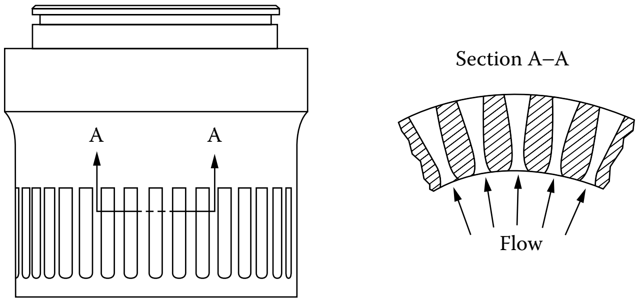


FIG. 6.14k
A multipath valve design, which can provide moderate noise reduction. (Courtesy of Emerson Process Management.)

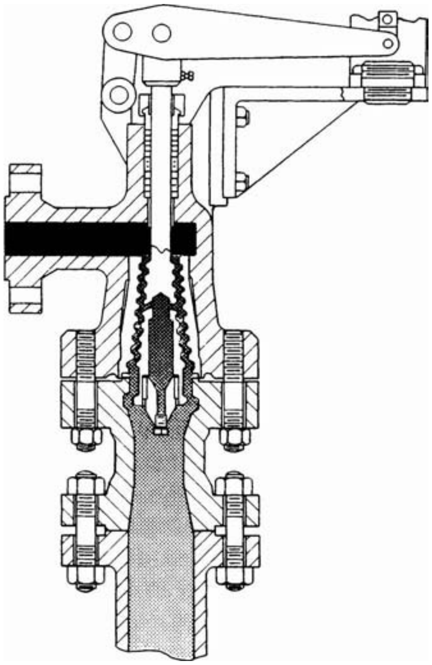


FIG. 6.14l
Multistage step trim valve for use on compressible fluids. Outlet is expanded to compensate for volume change. (Courtesy of Dresser Flow Solutions.)

Multistage Valves Pure multistage valves force the flow through a single path of two or more restrictions in series. (Figure 6.1y gives some examples.) An example of this design is shown in Figure 6.14l. The multiple orifices in series divide the total valve system pressure drop over several stages (typically three to nine). Thus, the reduced pressure drop per stage results in greater friction loss, reduced local velocities, and reduced noise.

The shape of the trim element allows an increasing effective flow area between the inlet and the outlet to compensate for the change in gas density and increase in specific volume. Thus, the outlet flange size is often larger than the inlet to limit the exit velocity to a level that will not regenerate excessive noise. Typically, these valves can provide noise attenuation up to 25 dBA, depending on pressure ratio and exit Mach number.

Resistor Elements In addition to diffusers, other special designs of multiple orifice restrictors are available (Figure 6.14m). These devices are built in a wafer design for installing between flanges and can be used in single- or in multistage configurations.

Such resistor elements can be installed in series, as shown in Figure 6.14n. These resistor plates are designed to work in series with the control valve to share the total pressure drop in a way that reduces pressure ratios on each element, thereby reducing the potential to generate noise.

The design of some of these devices forces the fluid through multiple changes in direction, acting like friction elements with noise attenuation capability of several dBA. These multiple orifice restrictors are very useful in valve noise control work, but like diffusers, they lose effectiveness with flow turndown.

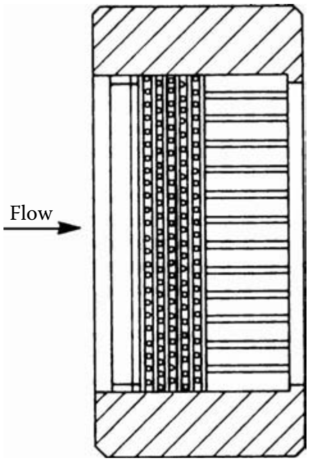


FIG. 6.14m
Resistor element used for valve back-pressure and noise reduction. (Courtesy of Dresser Flow Solutions.)

Combination Valves Combination multipath and multistage valves are usually required for the more severe and high noise producing services. These are the workhorse designs for the really tough applications, especially those that can cause extensive valve trim damage due to erosion or cavitation or noise levels in excess of 100 dBA. Various manufacturers have taken different approaches to the design of this type of valve.

Many are based upon multiple orifices in series and parallel with the flow controlled by a close-fitting cylindrical plug inside the cage for throttling (Figure 6.14o).

A variation of this design also incorporates a secondary diffuser element built into the valve (Figure 6.14p).

One manufacturer has designed a valve for moderate pressure drops using a standard valve trim assembly that eliminates the close-fitting cylindrical plug and uses a noise attenuator element ranging from one to seven stages (Figures 6.14q and 6.14r).

Another design utilizes the pressure loss producing effects of a fluid passing through a series of sharp turns

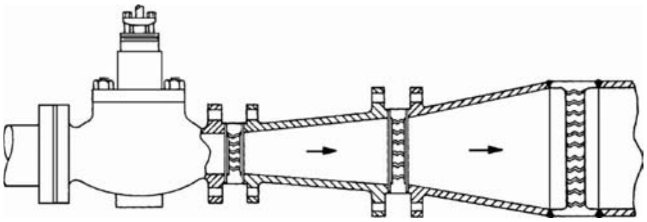


FIG. 6.14n
Noise can be reduced by resistor elements that are installed in series. (Courtesy of Dresser Flow Solutions.)

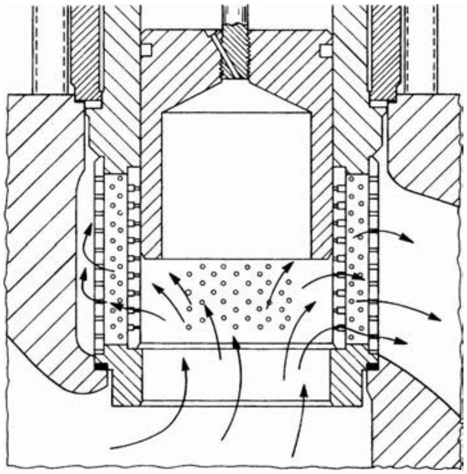


FIG. 6.14o
Multipath and multistage valve with shaped first stage holes. (Courtesy of Emerson Process Management.)

machined into a set of stacked disks (Figure 6.14s). Other similar designs are illustrated in Figure 6.1y and 6.1aa.

Depending upon the manufacturer and design, multistage, multiorifice valves will typically have 2–7 stages, although some might use 20 or more. However, the number of stages required by any specific design for a given application depends on the design principles employed and their effectiveness in the application. In other words, having more stages does not necessarily make a valve quieter than another design with fewer stages. Inexperienced users are advised to require some validation of manufacturers' claims of noise reduction for these critical service valves in addition to their noise calculations.

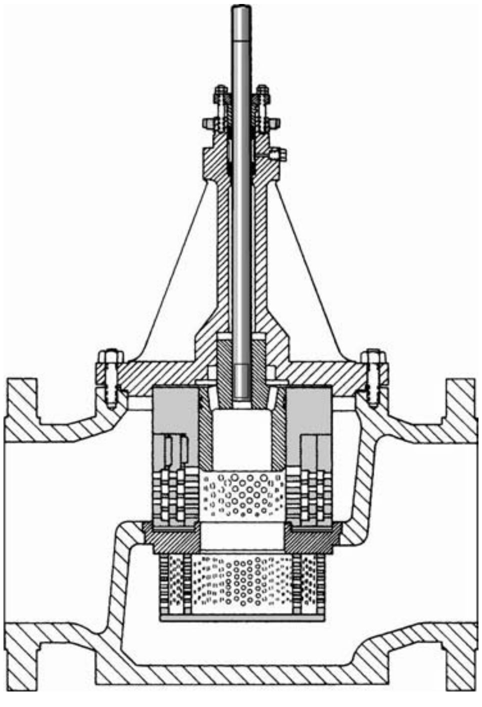


FIG. 6.14p
Multipath and multistage valve with integral secondary diffuser element. (Courtesy of ABB Control Valves.)

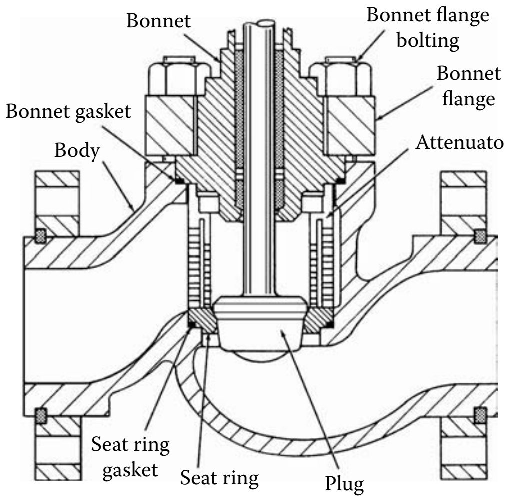


FIG. 6.14q
Two-stage noise attenuator in a valve, which was designed for use with a standard inner valve trim assembly. (Courtesy of Flowserve Corporation.)

AERODYNAMIC NOISE PREDICTION

Valve noise prediction is an inexact science because of the complex nature of noise generation by the control valve and the transmission of this noise through the pipe wall. So it is not surprising that a number of different prediction methods are used by manufacturers and others. What is surprising is that the various methods can give answers for the same application that differ by up to 20 dBA.

The subject of valve noise prediction is still subject to continuing research and evaluation. So what should we do?

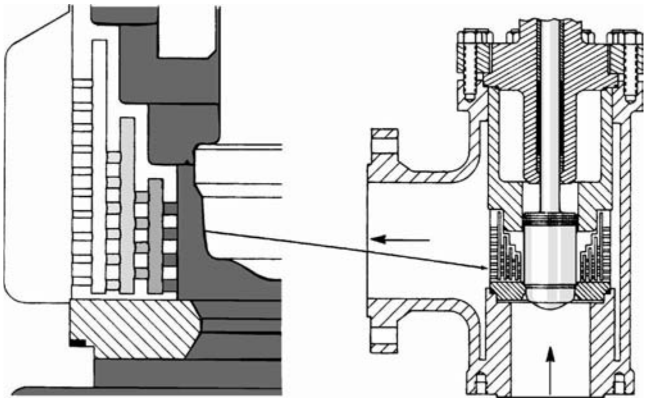


FIG. 6.14r
Multistage noise attenuator (and detail of that attenuator) in a valve, which was designed for use with a standard trim assembly with pressure balance. (Courtesy of Flowserve Corporation.)

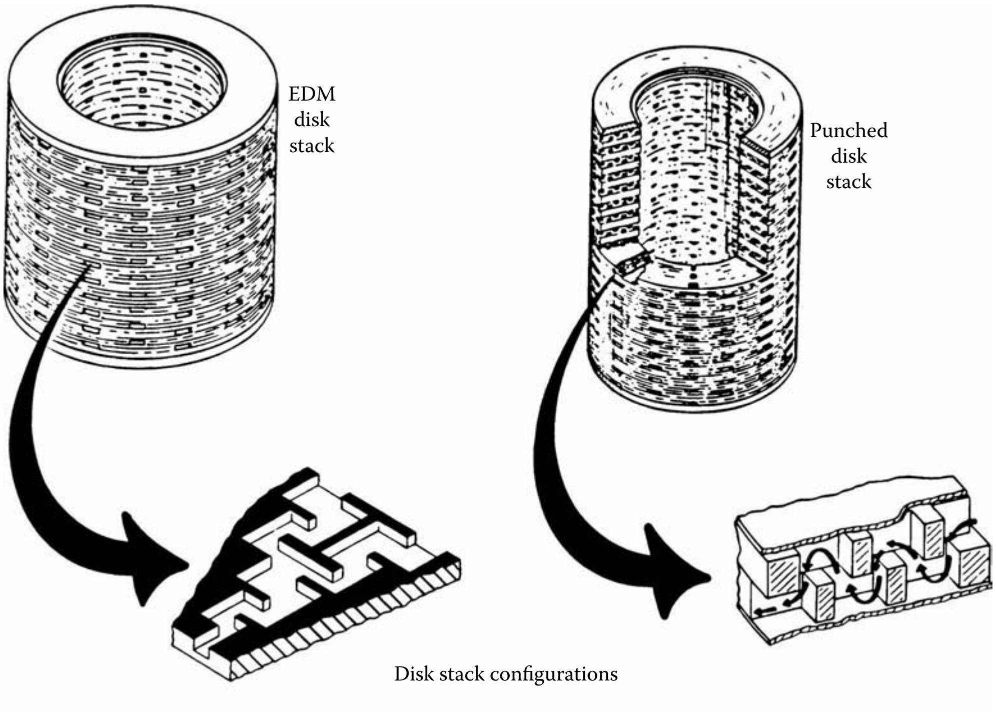


FIG. 6.14s
Special noise element design using labyrinth passages incorporated on plates. (Courtesy of Control Components Inc.)

Each manufacturer claims to be able to predict the valve noise level and provide a valve design solution. It finally falls upon the user to obtain the best possible process data, carefully evaluate all proposals, ask questions and resolve marked differences, and finally use good engineering judgment and experience in selecting the vendor for each application.

It is wise to err on the conservative side when making a final selection, because the cost of mistakes and of the required retrofit may far outweigh valve cost differentials. Fortunately, the noise prediction standards of the various standards organizations have helped to make comparisons of noise predictions somewhat easier for users and manufacturers alike.

Standards

The past quarter-century has seen continuous improvements in the standardized methods and in the industrial acceptance of noise prediction standards for valves. In 1979 the Verband Deutcher Maschinen- und Anlagenbau e.V. (VDMA) published the first standardized method of calculating the sound level for valves as Standard VDMA 24422, which addressed both hydrodynamic and aerodynamic noise. VDMA revised 24422 in 1989 to include calculations of the frequency domain.

The weakness of the VDMA method was that key valve noise parameters had to be determined experimentally; when testing was not practical, prediction accuracy was unsatisfactory. Meanwhile, other organizations were developing prediction methods based on free jet turbulence theories.

The Instrumentation, Systems, and Automation Society (ISA) published standard ISA-75.17 in 1989, and the International Electrotechnical Commission published the first edition of IEC 534-8-3 in 1995 and the second edition IEC 60534-8-3 in 2000. The basic methods in these standards are essentially the same. They both are based on the published works by Lighthill, Powell, Fowcs and Hawkins, Reethof and Ward, Shea, Fagerlund, Baumann, and the contributions of many others. These organizations update their respective standards when new information is validated.

The author has selected the more recent IEC Standard 60534-8-3 (2000) to demonstrate the basic calculation process. This standard and the field of acoustics in general use SI units. For conversion factors to U.S. Customary units, the reader should refer to Appendices A.1 and A.2 and is cautioned to use the proper gravitational unit conversions (i.e., g_c). To avoid excessive detail, only the SI system of units will be shown in this section.

Calculations

Acknowledgment: The author thanks the International Electrotechnical Commission for permission to reproduce information from its International Standard IEC 60534-8-3. All such extracts are copyright of IEC, Geneva, Switzerland. All rights reserved. Further information on the IEC is available from www.iec.ch. IEC has no responsibility for the placement and context in which the extracts and contents are reproduced by the author; nor is IEC in any way responsible for the other content or accuracy therein.

IEC 60534-8-3 The intent of this section is to familiarize the reader with the nomenclature and illustrate the basic procedure of the IEC 60534-8-3 aerodynamic noise prediction method. However, the standard covers additional special cases and details that are too extensive for coverage here. Like all theoretical methods, it is based on assumptions and limitations that must be applied with appropriate engineering skill and judgment to practical applications. Some of the stated assumptions and limitations of IEC 60534-8-3 include

- The valve is installed with steel or alloy steel piping upstream and downstream, possibly with pipe expanders, and that "the downstream piping is straight for a length of at least 2 m from the point where the noise measurement is made."
- The method assumes that the fluid properties can be modeled on the perfect gas laws.
- The method can be used for most valve types. However, it is not applicable for full-bore ball valves where the product F_pC (for the operating condition) exceeds 50% of the valve's rated flow coefficient.
- The method shown for standard single-stage trims in this section applies only for valve outlet and downstream pipe velocities with Mach numbers up to 0.3. Refer to IEC 60534-8-3 for multistage trims and higher velocities.
- Transmission loss of noise through the pipe wall is based on a simplified method due to the wide tolerances in pipe wall thickness for commercial steel pipe.
- The calculated sound pressure level assumes a location 1 m downstream from the valve or expander and 1 m from the outside of the pipe wall in an acoustic free field.
- The prediction does not guarantee actual results in the field. Validation tests were conducted with low pressure air and steam in laboratory tests, and the majority of test results were within 5 dBA of the predicted sound pressure level.

Nomenclature The following list of symbols, definitions, and units is a partial set of those used in IEC 60534-8-3.

Symbol	*Description*	*Unit*
A	Area of a single flow passage	m^2
A_n	Total flow area of last stage of multistage trim with n stages at given travel	m^2
C	Flow coefficient (K_v and C_v) (see IEC 60534-2-1)	Various (see IEC 60534-1)
c_{vc}	Speed of sound in the vena contracta at subsonic flow conditions	m/s
c_{vcc}	Speed of sound in the vena contacta at critical flow conditions	m/s
c_2	Speed of sound at downstream conditions	m/s
D	Valve outlet diameter	m
d	Diameter of a circular flow passage	m
d_H	Hydraulic diameter of a single flow passage	m
d_i	Smaller of valve outlet or expander inlet internal diameters	m
D_i	Internal downstream pipe diameter	m
D_j	Jet diameter at the vena contracta	m
d_o	Diameter of a circular orifice, the area of which equals the sum of areas of all flow passages at a given travel	m
F_d	Valve style modifier = d_H/d_o	Dimensionless
F_L	Liquid pressure recovery factor of a valve without attached fittings (see note 4)	Dimensionless
F_{LP}	Combined liquid pressure recovery factor and piping geometry factor of a control valve with attached fittings (see note 4)	Dimensionless
F_P	Piping geometry factor	Dimensionless
f_g	External coincidence frequency	Hz
f_o	Internal coincidence pipe frequency	Hz
f_p	Generated peak frequency	Hz
f_{pR}	Generated peak frequency in valve outlet or reduced diameter of expander	Hz
f_r	Ring frequency	Hz
l	Length of a radial flow passage	m
l_w	Wetted perimeter of a single flow passage	m
L_g	Correction for Mach number	dB (ref p_o)
L_{pAe}	A-weighted sound pressure level external of pipe	dBA (ref p_o)
$L_{pAe,1m}$	A-weighted sound pressure level 1 m from pipe wall	dBA (ref p_o)
L_{pi}	Internal sound pressure level at pipe wall	dB (ref p_o)
L_{wi}	Total internal sound power level	dB (ref W_o)
M	Molecular mass of flowing fluid	kg/kmol
M_j	Freely expanded jet Mach number in regimes II to IV	Dimensionless
M_{j5}	Freely expanded jet Mach number in regime V	Dimensionless
M_o	Mach number at valve outlet	Dimensionless
M_{vc}	Mach number at the vena contracta	Dimensionless
M_2	Mach number in downstream pipe	Dimensionless
$\dot{m}$	Mass flow rate	kg/s
$\dot{m}_s$	Mass flow rate at sonic velocity	kg/s
N	Numerical constants (see Table 6.14v)	Various
N_o	Number of independent and identical flow passages in valve trim	Dimensionless
p_a	Actual atmospheric pressure outside pipe	Pa (see note 3)
p_o	Reference sound pressure = 2×10^{-5} (see note 5)	Pa

p_s	Standard atmospheric pressure (see note 1)	Pa
p_{vc}	Absolute vena contracta pressure at subsonic flow conditions	Pa
p_{vcc}	Absolute vena contracta pressure at critical flow conditions	Pa
p_1	Valve inlet absolute pressure	Pa
p_2	Valve outlet absolute pressure	Pa
p_{2B}	Valve outlet absolute pressure at break point	Pa
p_{2C}	Valve outlet absolute pressure at critical flow conditions	Pa
p_{2CE}	Valve outlet absolute pressure where region of constant acoustical efficiency begins	Pa
R	Universal gas constant = 8314	J/kmol × K
r_w	Acoustic power ratio (see Table 6.14w)	Dimensionless
T_{vc}	Vena contracta absolute temperature at subsonic flow conditions	K
T_{vcc}	Vena contracta absolute temperature at critical flow conditions	K
T_1	Inlet absolute temperature	K
T_2	Outlet absolute temperature	K
TL	Transmission loss	dB
t_p	Pipe wall thickness	m
U_p	Gas velocity in downstream pipe	m/s
U_{vc}	Vena contracta velocity at subsonic flow conditions	m/s
W_a	Sound power	W
W_m	Stream power of mass flow	W
W_{ms}	Stream power of mass flow rate at sonic velocity	W
W_o	Reference sound power = 10^{-12} (see note 5)	W
α	Recovery correction factor	Dimensionless
β	Contraction coefficient for valve outlet or expander inlet	Dimensionless
γ	Specific heat ratio	Dimensionless
η	Acoustical efficiency factor (see note 2)	Dimensionless
ρ_1	Density of fluid at p_1 and T_1	kg/m^3
ρ_2	Density of fluid at p_2 and T_2	kg/m^3
Φ	Relative flow coefficient	Dimensionless

Note 1. Standard atmospheric pressure is 101.325 kPa or 1.01325 bar.

Note 2. Subscripts 1, 2, 3, 4, and 5 denote regimes I, II, III, IV, and V, respectively.

Note 3. 1 bar = 10^2 kPa = 10^5 Pa.

Note 4. For the purpose of calculating the vena contracta pressure, and therefore velocity, in this standard, pressure recovery for gases is assumed to be identical to that of liquids.

Note 5. Sound power and sound pressure are customarily expressed using the logarithmic scale known as the decibel scale. This scale relates the quantity logarithmically to some standard reference. This reference is 2×10^{-5} Pa for sound pressure and 10^{-12} W for sound power.

Method Outline Numerous equations are involved in the calculation procedure, but a brief outline of the five general steps will ensure continuity in the procedures.

1. Gather the necessary input data
    - Valve sizing data and dimensions of trim and body ports
    - Configuration and dimensions of adjacent piping
    - Service conditions and fluid properties
2. Calculate key pressures and pressure ratios, and determine the noise regime
3. Calculate the effective jet diameter
4. Calculate jet conditions, acoustic efficiency, sound power, peak frequency for the noise, and internal sound pressure level
5. Calculate pipe natural frequencies, pipe transmission loss, and external sound pressure level

Noise Regimes There are five key noise regimes identified in IEC 60534-8-3, which have different mechanisms governing the generation and transmission of sound. In order to determine which regime applies to a given set of conditions, several important pressures and pressure ratios must be calculated and compared with the actual downstream pressure, p_2.

The vena contracta is the region of flow constriction with maximum velocity and minimum pressure, p_{vc}, given in Equation 6.14(7).

$$p_{vc} = p_1 - \frac{p_1 - p_2}{F_L^2} \qquad \textbf{6.14(7)}^2$$

If the valve has attached fittings, F_L is replaced by the value of F_{LP}/F_P. At critical flow conditions, the vena contracta pressure is p_{vcc}.

$$p_{vcc} = p_1\left(\frac{2}{\gamma+1}\right)^{\gamma/(\gamma-1)} \qquad \textbf{6.14(8)}$$

The downstream pressure at the critical pressure drop where sonic flow begins at the vena contracta is P_{2C}.

$$p_{2C} = p_1 - F_L^2(p_1 - p_{vcc}) \qquad \textbf{6.14(9)}$$

At the break point pressure, P_{2B}, shock cell turbulent interaction begins to dominate noise generation. For downstream pressures greater than p_{2B}, turbulent shear flow generates most of the sound power.

$$p_{2B} = \frac{p_1}{\alpha}\left(\frac{1}{\gamma}\right)^{\gamma/(\gamma-1)} \qquad \textbf{6.14(10)}$$

where α is a correction factor defined as

---

[2] Equations 6.14(7)–(46) and 6.14(48)–(53) are used here by permission of IEC. Copyright © 2000, IEC, Geneva, Switzerland.www.iec.ch.

$$\alpha \equiv \frac{\left(\dfrac{p_1}{p_{2C}}\right)}{\left(\dfrac{p_1}{p_{\text{vcc}}}\right)} = \frac{p_{\text{vcc}}}{p_{2C}} \qquad \textbf{6.14(11)}$$

Regime V begins when downstream pressure drops to $P_{2\text{CE}}$ and is where acoustical efficiency becomes constant. Further reductions in downstream pressure will not increase the sound pressure level.

$$p_{2\text{CE}} = \frac{p_1}{22\alpha} \qquad \textbf{6.14(12)}$$

Table 6.14t summarizes the boundaries and characteristics of Regimes I through V.

Calculate Jet Diameter Determining the jet diameter requires information about the valve trim dimensions in order to calculate the valve-style modifier F_d, which is the ratio of hydraulic diameter, d_H, of a single flow passage to the circle diameter, d_o, corresponding to the total flow area.

$$F_d = \frac{d_H}{d_o} \qquad \textbf{6.14(13)}$$

$$d_H = \frac{4A}{l_w} \qquad \textbf{6.14(14)}$$

$$d_o = \sqrt{\frac{4N_oA}{\pi}} \qquad \textbf{6.14(15)}$$

TABLE 6.14t
Characteristics of IEC Noise Regimes

Regime	*Downstream Pressure*	*Description*
I	$p_1 > p_2 \geq p_{2C}$	Subsonic flow; isentropic recompression; turbulent shear noise
II	$p_{2C} > p_2 \geq p_{\text{vcc}}$	Sonic flow at vena contracta; isentropic recompression; turbulent shear noise
III	$p_{\text{vcc}} > p_2 \geq p_{2B}$	Supersonic flow past the vena contracta; no recompression; noise from shock turbulence and shear turbulence
IV	$p_{2B} > p_2 \geq p_{2\text{CE}}$	Sonic flow at vena contracta; supersonic Mach cone terminates in Mach disk at outlet; shock interaction dominates noise
V	$p_{2\text{CE}} > p_2$	Supersonic Mach cone reaches maximum Mach number; acoustical efficiency and noise level are constant

Lacking specific dimensions, approximate values of F_d are given in Table 6.14u.

The jet diameter, D_j, is calculated from Equation 6.14(16).

$$D_j = N_{14}F_d\sqrt{CF_L} \qquad \textbf{6.14(16)}$$

where units conversion factor N_{14} is found in Table 6.14v and depends on whether the required flow coefficient, C, is given as C_v or K_v. (Refer to Section 6.15 in this chapter for information about flow coefficients.)

TABLE 6.14u
Typical Values of Valve-Style Modifier F_d (Full-Size Trim)

	Relative Flow Coefficient Φ						
Valve Type	*Flow Direction*	*0.10*	*0.20*	*0.40*	*0.60*	*0.80*	*1.00*
Globe, parabolic plug	To open	0.10	0.15	0.25	0.31	0.39	0.46
	To close	0.20	0.30	0.50	0.60	0.80	1.00
Globe, 3 V-port plug	Either*	0.29	0.40	0.42	0.43	0.45	0.48
Globe, 4 V-port plug	Either*	0.25	0.35	0.36	0.37	0.39	0.41
Globe, 6 V-port plug	Either*	0.17	0.23	0.24	0.26	0.28	0.30
Globe, 60 equal diameter hole drilled cage	Either*	0.40	0.29	0.20	0.17	0.14	0.13
Globe, 120 equal diameter hole drilled cage	Either*	0.29	0.20	0.14	0.12	0.10	0.09
Butterfly, swing-through (centered shaft), to 70°	Either	0.26	0.34	0.42	0.50	0.53	0.57
Butterfly, fluted vane to 70°	Either	0.08	0.10	0.15	0.20	0.24	0.30
Butterfly, 60° flat disk	Either						0.50
Eccentric rotary plug	Either	0.12	0.18	0.22	0.30	0.36	0.42
Segmented ball 90°	Either	0.60	0.65	0.70	0.75	0.78	0.98

NOTE: These values are typical only; actual values are stated by the manufacturer.

* Limited $p_1 - p_2$ may apply in flow-to-close direction.

Copyright © 2000, IEC, Geneva, Switzerland. www.iec.ch.

TABLE 6.14v
Numerical Constants N

	Flow Coefficient	
Constant	K_v	C_v
N_{14}	4.9×10^{-3}	4.6×10^{-3}
N_{16}	4.23×10^{4}	4.89×10^{4}

Note: Unlisted numerical constants are not used in this standard

Regime I Calculations Calculate the following subsonic parameters for the vena contracta.

Gas velocity:

$$U_{\mathrm{vc}} = \sqrt{2\left(\frac{\gamma}{\gamma-1}\right)\left[1-\left(\frac{p_{\mathrm{vc}}}{p_1}\right)^{(\gamma-1)/\gamma}\right]\frac{p_1}{\rho_1}} \qquad \textbf{6.14(17)}$$

Stream power:

$$W_m = \frac{\dot{m}(U_{\mathrm{vc}})^2}{2} \qquad \textbf{6.14(18)}$$

Absolute temperature:

$$T_{\mathrm{vc}} = T_1\left(\frac{p_{\mathrm{vc}}}{p_1}\right)^{(\gamma-1)/\gamma} \qquad \textbf{6.14(19)}$$

Speed of sound:

$$c_{\mathrm{vc}} = \sqrt{\frac{\gamma R T_{\mathrm{vc}}}{M}} \qquad \textbf{6.14(20)}$$

Mach number:

$$M_{\mathrm{vc}} = \frac{U_{\mathrm{vc}}}{c_{\mathrm{vc}}} \qquad \textbf{6.14(21)}$$

With this information, calculate the acoustical efficiency factor, η_1, sound power, W_a, and peak frequency, f_p.

$$\eta_1 = (1\times 10^{-4})M_{\mathrm{vc}}^{3.6} \qquad \textbf{6.14(22)}$$

$$W_a = \eta_1 r_w W_m F_L^2 \qquad \textbf{6.14(23)}$$

where r_w is the acoustic power ratio taken from Table 6.14w.

TABLE 6.14w
Acoustic Power Ratio r_w

Valve or Fitting	r_w
Globe, parabolic plug	0.25
Globe, 3 V-port plug	0.25
Globe, 4 V-port plug	0.25
Globe, 6 V-port plug	0.25
Globe, 60 equal diameter hole drilled cage	0.25
Globe, 120 equal diameter hole drilled cage	0.25
Butterfly, swing-through (centered shaft), to 70°	0.5
Butterfly, fluted vane, to 70°	0.5
Butterfly, 60° flat disk	0.5
Eccentric rotary plug	0.25
Segmented ball 90°	0.25
Expanders	1

Peak frequencies in Regimes I and II are based on Strouhal's equation with the Strouhal number = 0.2.

$$f_p = \frac{0.2U_{\mathrm{vc}}}{D_j} \qquad \textbf{6.14(24)}$$

Common Calculations for Regimes II–V For sonic conditions in the vena contracta, calculate the following parameters.

Vena contracta temperature:

$$T_{\mathrm{vcc}} = \frac{2T_1}{\gamma+1} \qquad \textbf{6.14(25)}$$

Velocity of sound:

$$c_{\mathrm{vcc}} = \sqrt{\frac{\gamma R T_{\mathrm{vcc}}}{M}} \qquad \textbf{6.14(26)}$$

Stream power:

$$W_{\mathrm{ms}} = \frac{\dot{m}c_{\mathrm{vcc}}^2}{2} \qquad \textbf{6.14(27)}$$

Mach number in the freely expanding jet:

$$M_j = \sqrt{\frac{2}{\gamma-1}\left[\left(\frac{p_1}{\alpha p_2}\right)^{(\gamma-1)/\gamma}-1\right]} \qquad \textbf{6.14(28)}$$

Next, the acoustical efficiency factors, sound power, and peak frequency are calculated for the regime in question.

Regime II Acoustical efficiency factor:

$$\eta_2 = (1\times10^{-4})M_j^{6.6F_L^2} \qquad \textbf{6.14(29)}$$

Sound power:

$$W_a = \eta_2 r_w W_{\text{ms}}\left(\frac{p_1 - p_2}{p_1 - p_{\text{vcc}}}\right) \qquad \textbf{6.14(30)}$$

Peak frequency:

$$f_p = \frac{0.2M_j\, c_{\text{vcc}}}{D_j} \qquad \textbf{6.14(31)}$$

Regime III Acoustical efficiency factor:

$$\eta_3 = (1\times10^{-4})M_j^{6.6F_L^2} \qquad \textbf{6.14(32)}$$

Sound power:

$$W_a = \eta_2 r_w W_{\text{ms}} \qquad \textbf{6.14(33)}$$

Peak frequency is calculated from Equation 6.14(31).

Regime IV Acoustical efficiency factor:

$$\eta_4 = (1\times10^{-4})\left(\frac{M_j^2}{2}\right)\left(\sqrt{2}\right)^{6.6F_L^2} \qquad \textbf{6.14(34)}$$

Sound power:

$$W_a = \eta_4 r_w W_{\text{ms}} \qquad \textbf{6.14(35)}$$

Peak frequency:

$$f_p = \frac{0.35c_{\text{vcc}}}{1.25D_j\sqrt{M_j^2 - 1}} \qquad \textbf{6.14(36)}$$

Regime V Jet Mach number reaches it maximum:

$$M_{j5} = \sqrt{\frac{2}{\gamma - 1}[(22)^{(\gamma-1)/\gamma} - 1]} \qquad \textbf{6.14(37)}$$

The acoustical efficiency factor becomes constant:

$$\eta_5 = (1\times10^{-4})\left(\frac{M_{j5}^2}{2}\right)\left(\sqrt{2}\right)^{6.6F_L^2} \qquad \textbf{6.14(38)}$$

Sound power generated in this regime that radiates into downstream pipe is

$$W_a = \eta_5 r_w W_{\text{ms}} \qquad \textbf{6.14(39)}$$

Peak frequency:

$$f_p = \frac{0.35c_{\text{vcc}}}{1.25D_j\sqrt{M_{j5}^2 - 1}} \qquad \textbf{6.14(40)}$$

Noise Calculations The following calculations are used for all regimes.

Downstream mass density:

$$\rho_2 = \rho_1\left(\frac{p_2}{p_1}\right) \qquad \textbf{6.14(41)}$$

Downstream sonic velocity:

$$c_2 = \sqrt{\frac{\gamma R T_2}{M}} \qquad \textbf{6.14(42)}$$

where T_2 may be found from thermodynamic isenthalpic relationships. If fluid properties are not known, reasonable results can be obtained by assuming T_2 is approximately equal to T_1.

Mach number at valve outlet:

$$M_o = \frac{4\dot{m}}{\pi D^2 \rho_2 c_2} \qquad \textbf{6.14(43)}$$

If the outlet Mach number M_o is above 0.3, accuracy of this method diminishes. IEC 60534-8-3 clause 7 provides further procedures for high Mach number applications, which is outside the discussion of this basic process.

The internal sound pressure level, L_{pi}, referenced to 2×10^{-5} Pa is calculated in dB from the following:

$$L_{pi} = 10\log_{10}\left[\frac{(3.2\times10^9)W_a\rho_2 c_2}{D_i^2}\right] \qquad \textbf{6.14(44)}$$

Transmission through the Pipe The pipe wall must be made to vibrate in order for noise inside the pipe to radiate into the air outside the pipe. The mode of pipe vibration, for the purpose of this prediction method, is determined from the peak frequency of the noise source and the natural frequencies of the pipe.

The assumption is made that the shape of the sound frequency spectrum is an arc or "haystack"-shaped curve that reaches a pronounced maximum level at peak frequency, f_p. Although this is true for most valves, some configurations can possess "flatter" broadband spectra that could radiate more noise than the simplified model predicts.

Pipe natural frequencies are functions of the pipe diameter, wall thickness, and density. The transmission loss model used by the IEC standard is based on the work of Fagerlund and Chow. The important characteristic frequencies are explained in detail by Singleton and are summarized below.

Ring frequency, f_r, has a wavelength exactly equal to the circumference of the pipe, which produces a resonant stress wave around the circumference.

$$f_r = \frac{5000}{\pi D_i} \qquad \textbf{6.14(45)}$$

External coincidence frequency, f_g, corresponds to the external acoustic wave speed that matches the speed of a flexural wave in pipe wall. Assuming the speed of sound in steel is 5000 m/s and 343 m/s in air,

$$f_g = \frac{\sqrt{3}(343)^2}{\pi t_p(5000)} \qquad \textbf{6.14(46)}$$

First internal coincidence frequency, f_o, is the lowest natural frequency of the pipe wall and produces a longitudinal flexural wave that spirals along the length of the pipe.

$$f_o = \frac{1250}{\pi D_i}\left(\frac{c_2}{c_o}\right) = \frac{f_r}{4}\left(\frac{c_2}{343}\right) \qquad \textbf{6.14(47)}$$

Cutoff frequency, f_c, though not part of the IEC standard, is significant because at the cutoff frequency and below, the wavelengths are too long to reflect off the internal pipe wall, making them incapable of vibrating the pipe.

$$f_c = 0.586\frac{c_2}{D_i} \qquad \textbf{6.14(48)}$$

The relationship of the peak frequency in the flow stream to the pipe natural frequencies is used to calculate the frequency factors used in the transmission loss calculation. Table 6.14x, taken from IEC 60534-8-3, shows how frequency factors G_x and G_y are determined.

TABLE 6.14x
Frequency Factors G_x and G_y

$f_p < f_o$	$f_p \geq f_o$
$G_x = \left(\frac{f_o}{f_r}\right)^{2/3}\left(\frac{f_p}{f_o}\right)^4$	$G_x = \left(\frac{f_p}{f_r}\right)^{2/3}$ for $f_p < f_r$
	$G_x = 1$ for $f_p \geq f_r$
$G_y = \left(\frac{f_o}{f_g}\right)$ for $f_o < f_g$	$G_y = \left(\frac{f_p}{f_g}\right)$ for $f_p < f_g$
$G_y = 1$ for $f_o \geq f_g$	$G_y = 1$ for $f_p \geq f_g$

The transmission loss across the pipe wall is calculated from Equation 6.14(49).

$$TL = 10\log_{10}\left[(7.6\times10^{-7})\left(\frac{c_2}{t_p f_p}\right)^2 \frac{G_x}{\left(\frac{\rho_2 c_2}{415 G_y}+1\right)}\left(\frac{p_a}{p_s}\right)\right] \qquad \textbf{6.14(49)}$$

Next, calculate the downstream pipe velocity correction factor, L_g.

$$L_g = 16\log_{10}\left(\frac{1}{1-M_2}\right) \qquad \textbf{6.14(50)}$$

where M_2 should not exceed 0.3 and is calculated by

$$M_2 = \frac{4\dot{m}}{\pi D_i^2 \rho_2 c_2} \qquad \textbf{6.14(51)}$$

The A-weighted sound pressure level radiated from the outside surface of the pipe is given a 5 dB correction to account for all frequency peaks and is calculated below.

$$L_{\mathrm{pAe}} = 5 + L_{pi} + TL + L_g \qquad \textbf{6.14(52)}$$

Finally, a distance adjustment is made to calculate the sound pressure level in dBA at 1 m from the pipe wall.

$$L_{\mathrm{pAe,1m}} = L_{\mathrm{pAe}} - 10\log_{10}\left[\frac{D_i + 2t_p + 2}{D_i + 2t_p}\right] \qquad \textbf{6.14(53)}$$

Noise Calculation Example

These calculations are typically carried out with computer software and presented as part of the sizing calculations done by valve manufacturers. For a thorough understanding, a simple calculation example is tabulated below.

Example 1. Steam Valve

Inputs

Valve: NPS 4-in. (DN 100)	Fluid: Steam
Trim: Parabolic plug	$M = 18.02$ kg/kmol
$C = C_v = 152$ required (from sizing calculations)	$T_1 = 260$ °C $= 533$ K
$F_L = 0.90$	$p_1 = 2.5$ MPa
$\Phi = 0.60$	$p_2 = 1.7$ MPa
$F_d = 0.31$	$\dot{m} = 9.10$ kg/s
Maximum allowable noise level: 90 dBA	$\gamma = 1.32$
	From steam tables:
Pipe: NPS 4-in. (DN 100) Schedule 40 carbon steel	$\rho_1 = 11.17$ kg/m^3
$D_i = 0.102$ m	$T_2 = 247$ °C $= 520$ K
$t_p = 0.00602$ m	$\rho_2 = 7.58$ kg/m^3

Preliminary Calculations

Variable	*Equation*	*Results*
p_{vc}	6.14(7)	$2.5 - (2.5 - 1.7)/(0.9)^2 = 1.512$ MPa
p_{vcc}	6.14(8)	$2.5[2/(1.32 + 1)]^{1.32/0.32} = 1.355$ MPa
P_{2C}	6.14(9)	$2.5 - (0.9)^2\ (2.5 - 1.355) = 1.573$ MPa
α	6.14(11)	$1.355/1.573 = 0.861$
P_{2B}	6.14(10)	$(2.5/0.861)\ (1/1.32)^{1.32/0.32} = 0.924$ MPa
$P_{2\mathrm{CE}}$	6.14(12)	$2.5/[(22)\ (0.861)] = 0.132$ MPa
Regime?	Table 6.14t	$p_2 \geq p_{2C}$: $1.7 \geq 1.573$ $\therefore$ Regime I
D_j	6.14(16)	$(4.6 \times 10^{-3})(0.31)(152 \times 0.9)^{1/2} = 0.0167$ m
		Regime I Calculations
U_{vc}	6.14(17)	$\{2(1.32/0.32)[1(1.512/2.5)^{0.32/1.32}](2.5 \times 10^6/11.17)\}^{1/2} = 460$ m/s
W_m	6.14(18)	$(9.1)(460)^2/2 = 9.63 \times 10^5$ W
T_{vc}	6.14(19)	$(533)(1.512/2.5)^{0.32/1.32} = 472$ K
c_{vc}	6.14(20)	$[(1.32)(8314)(472)/(18.02)]^{1/2} = 536$ m/s
M_{vc}	6.14(21)	$460/536 = 0.858$
		Determine Internal Noise
η_1	6.14(22)	$(1 \times 10^{-4})(0.858)^{3.6} = 5.76 \times 10^{-5}$
r_w	Table 6.14w	0.25
W_a	6.14(23)	$(5.76 \times 10^{-5})(0.25)(9.63 \times 10^5) = 13.9$ W
f_p	6.14(24)	$(0.2)(460)/(0.0167) = 5{,}509$ Hz
ρ_2	6.14(41) or steam tables	7.58 kg/m^3
c_2	6.14(42)	$[(1.32)(8314)(520)/(18.02)]^{1/2} = 563$ m/s
M_o	6.14(43)	$(4)(9.1)/[\pi(0.1016)^2(7.58)(563)] = 0.263$
L_{pi}	6.14(44)	$10\log[(3.2 \times 10^9)(13.9)(7.58)(563)/(0.102)^2] = 162.6$ dB
		Determine Radiated Noise
f_r	6.14(45)	$(5000)/[\pi(0.102)] = 15.6$ kHz
f_g	6.14(46)	$(3)^{1/2}(343)^2/[\pi(0.00602)(5000)] = 2155$ Hz
f_o	6.14(47)	$(15600/4)(563/343) = 6401$
G_x	Table 6.14x	$f_p < f_o$: $(6401/15{,}600)^{2/3}(5509/6401)^4 = 0.303$
G_y	Table 6.14x	$f_o \geq f_g$: 1.0

TL	6.14(49)	$10\log\left[(7.6\times10^{-7})\left(\frac{563}{(0.00602)(5509)}\right)^{2}\left(\frac{0.303}{\frac{(7.58)(563)}{(415)(1)}+1}\right)(1)\right]=-52.3\text{ dB}$
M_2	6.14(51)	$4(9.1)/[\pi(0.102)^2(7.58)(563)] = 0.261$
L_g	6.14(50)	$16\log[1/(1-0.261)] = 2.10$ dB
L_{pAe}	6.14(52)	$5 + 162.6 - 52.3 + 2.1 = 117.4$ dBA
$L_{\text{pAe,1m}}$	6.14(53)	$117.4 - 10\log\{[0.102 + 2(0.00602) + 2]/[0.102 + 2(0.00602)]\} = 105$ dBA

Conclusion: Noise level exceeds desired maximum; consider noise reduction trim; consult manufacturer.

Applying Distance Corrections

Placing extra distance between noisy equipment and people is sometimes a viable alternative to expensive noise reduction treatment, if there are no other detrimental effects of the noise at the source.

If a noise source can be treated as a point in the acoustic far field, the sound radiates in a spherical pattern. Atmospheric vents can be treated this way. The reduced sound pressure level at some distance, *r*, from the center of a point source can be determined from a measured or calculated sound pressure level taken at a reference distance of r_o from the center (typically r_o = 1 m + pipe OD/2) using Equation 6.14(54) below.

$$L_{\text{pAe,r}} = L_{\text{pAe,1m}} - 20\log_{10}\left(\frac{r}{r_o}\right) \qquad \textbf{6.14(54)}$$

Because noise produced by valves radiates to the environment largely through the pipe for great distances downstream of a valve, this type of noise source is generally treated as a line source. Line sources radiate noise in a cylindrical pattern. The reduced sound-pressure level at distance, *r*, from a line source is

$$L_{\text{pAe,r}} = L_{\text{pAe,1m}} - 10\log_{10}\left(\frac{r}{r_o}\right) \qquad \textbf{6.14(55)}$$

Example 2. Valve Noise at a Distance Problem: Using the same valve and conditions from Example 1, what would be the sound pressure level for a worker 30 m away from the downstream pipe (centerline)?

Solution: From Example 1, $L_{\text{pAe,1m}}$ = 105 dBA, and pipe OD = 0.114 m. Use Equation 6.14(55).

$$L_{\text{pAe,30m}} = 105 - 10\log_{10}\left(\frac{30m}{1m + 0.114m/2}\right)$$

$$= 105 - 15 = 90\text{ dBA}$$

HYDRODYNAMIC NOISE PREDICTION

Noise prediction for liquid flow through valves should consider three major flow regimes. 1) Turbulent flow, which, without cavitation, rarely produces noise levels high enough to create dangerous structural vibration or noise pollution. 2) Cavitating flow, which produces noise from vapor cavity formation and collapse as well as from turbulence, and it frequently causes excessive vibration and noise in addition to erosion of valve and piping materials. 3) Flashing of liquid into vapor across a valve sometimes causes high levels of noise and vibration, if vapor velocities in the downstream piping approach sonic velocities. Piping systems should be so sized as to avoid vapor or two-phase velocities, which are high enough to cause noise and erosion.

Hydrodynamic noise prediction is currently in a state of development. Noise prediction Standards VDMA 24422 (1989) and IEC 60534-8-4 (1994) have been shown by Kiesbauer and Baumann to predict lower than actual noise in many cases. A more accurate method of hydrodynamic noise prediction has been proposed (Kiesbauer, J. and Baumann, H. D., "Recent Developments in the Prediction of Hydrodynamic Noise of Control Valves," *Valve World*, February 2004), which is being considered by the IEC as a revision to Standard 60534-8-4 at the time of this writing. This method includes calculations for turbulent flow and cavitating flow regimes. There are no standards or generally accepted methods at this time for predicting noise under flashing conditions. For calculation of noise, the reader is advised to study the Kiesbauer-Baumann method or later revisions of IEC Standard 60534-8-4.

Hydrodynamic noise predictions use a differential pressure ratio, x_F, to identify noise regimes.

$$x_F = \frac{p_1 - p_2}{p_1 - p_v} \qquad \textbf{6.14(56)}$$

The incipient cavitation index, x_{Fz}, corresponds to the differential pressure ratio at which cavitation in a valve begins and should be determined from cavitation tests, although some of the methods include ways of estimating x_{Fz}. This is the index that separates the turbulent flow regime from the cavitating flow regime. (If $x_F \geq 1.0$, the liquid is flashing.) Each of the methods discussed above follows a general process similar to that for aerodynamic noise prediction:

1. Gather the necessary input data
   - Valve sizing data and dimensions of trim and body ports
   - Configuration and dimensions of adjacent piping
   - Service conditions and fluid properties
2. Calculate key pressures and pressure ratios, and determine the noise regime.
3. Calculate the effective jet diameter and stream power.
4. Calculate acoustic efficiency and internal sound-pressure level.
5. Calculate pipe natural frequencies, pipe transmission loss, and external sound-pressure level.

Bibliography

Arant, J. B., "Special Control Valves Reduce Noise and Vibration," *Chemical Engineering,* March 6, 1972.

Arant, J. B., "How to Cope with Control Valve Noise," *Instrument Technology,* March 1973.

Arant, J. B., "Control Valve Noise," in *Control Valves—Practical Guides for Measurement and Control,* G. Borden and P. G. Friedmann (eds.), Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1998.

Ball, K. E., "Final Elements: Final Frontier," *InTech,* November 1986.

Barnes, R. W., "Understanding Automatic Control Valve Noise," *ISA Transactions*, Vol. 22, No. 1, Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1983.

Baumann, H. D., "On the Prediction of Aerodynamically Created Sound Pressure Level of Control Valves," ASME Paper 70 WA/FE-28, December 1970.

Baumann, H. D., "How to Estimate Aerodynamic Valve Throttling Noise: A Fresh Look," ISA Paper #82-902, 1982.

Baumann, H. D., "A Method for Predicting Aerodynamic Valve Noise Based on Modified Free Jet Noise Theories," ASME Paper 87-WA/NCA-7, presented at ASME Winter Annual Meeting, Boston, MA, December 1987.

Baumann, H. D., "A Firm New Handle on Valve Noise," *Chemical Engineering,* December 1996.

Baumann, H. D., "Predicting Control Valve Noise at High Exit Velocities," *InTech,* February 1997.

Baumann, H. D. and Hoffmann, H., "Method for the estimation of frequency-dependent sound pressures at the pipe exterior of throttling valves," *Noise Control Engineering Journal,* Vol. 47, No. 2, March–April 1999.

Baumann, H. D. and Page, G. W. Jr., "A Method to Predict Sound Levels from Hydrodynamic Sources Associated with Flow Through Throttling Valves," *Noise Control Engineering Journal,* Vol. 43, No. 5, September–October 1995.

Boyle, S. J., "Acoustical Design Considerations for Low-Noise Control Valves," *Proceedings of Noise-Con 83,* Institute of Noise Control Engineering, March 1983.

Brockbank, G. and Glenn, A., "Compressible Flow Noise," *Valve Magazine*, Fall 1992.

CEI/IEC 60534-8-4 (1994-05), "Prediction of Noise Generated by Hydrodynamic Flow," Geneva, Switzerland: International Electrotechnical Commission, 1994.

Chow, G. C. and Reethof, G., "A Study of Valve Noise Generation Processes for Compressible Fluids," ASME Paper 80 WA/NC-15, 1980.

CEI/IEC 60534-8-3 (2000-07), "Control Valve Aerodynamic Noise Prediction Method," 2nd edition, Geneva, Switzerland: International Electrotechnical Commission, 2000.

Fagerlund, A. C, "Use of Pipe Wall Vibrations to Measure Valve Noise," ISA Paper #74-829.

Fagerlund, A. C., "Recommended Maximum Valve Noise Levels, *InTech,* November 1986.

Fagerlund, A. C. and Chow, D. C., "Sound Transmission through a Cylindrical Pipe," *ASME Journal of Engineering for Industry,* Vol. 103, No. 4, November 1981.

Glenn, A. H., *Eliminating Screech-Type Noise in Control Valves*, doctoral dissertation presented to the Department of Mechanical Engineering, Brigham Young University, Provo, UT, July 1987.

Hynes, K. M., "The Development of a Low Noise Constant Area Throttling Device," ISA #839-70, October 1970.

ISA-75.17-1989, "Control Valve Aerodynamic Noise Prediction," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1989.

Kiesbauer, J. and Baumann, H. D., "Neues bei der Schallberechnung von flüssigketsdurchströmten Stellgeräten," *Industriearmaturen,* September 2002.

Kiesbauer, J. and Baumann, H. D., "Recent Developments in the Prediction of Hydrodynamic Noise of Control Valves," *Valve World*, February 2004.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V., *Fundamentals of Acoustics,* 3rd edition, New York: John Wiley & Sons, Inc., 1982.

Lighthill, M. J., "On Sound Generated Aerodynamically II, Turbulence as a Source of Sound," *Proceedings of The Royal Society of London,* Vol. 222, Series A, 1954.

Lodge, J. M. and Turner, J. T., "Aerodynamic Noise Generation in Control Valves—An Experimental Study," Paper F1, International Conference on Developments in Valves and Actuators for Fluid Control, sponsored by BHRA and BVAMA, Oxford, England, 10–13 September 1985.

Ng, K. W., "Control Valve Aerodynamic Noise Generation and Prediction," *Proceedings of the National Noise and Vibration Control Conference,* Chicago, IL: INCE, 1980.

Ng, K. W., "Control Valve Noise," *ISA Transactions 33,* 1994, pp. 275–286.

Page, G. W. Jr., "Predict Control Valve Noise–A Graphical Method for Liquid-Flow Systems Eases the Task," *Chemical Engineering*, July 1997.

Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

Reed, C. M., "Optimizing Valve Jet Size and Spacing Reduces Valve Noise," *Control Engineering*, September 1976.

Reethof, G., "Some Recent Developments in Valve Noise Prediction Methods," *ASME Transactions of Winter Annual Meeting,* November 1982.

Reethof, G. and Ward, W. C., "A Theoretically Based Valve Noise Prediction Method for Compressible Fluids," *Transactions of the ASME—Journal of Vibration, Acoustics, Stress, and Reliability in Design*, Vol. 108, pp. 329–338, July 1986.

Saitta, G. N., "A Realistic Look at Reducing Control Valve Noise Problems," *Oil and Gas Journal,* June 18, 1979.

Sawley, R. J. and White, P. H., "The Influence of Pressure Recovery on the Development of Valve Noise Description," Instrumentation, Systems, and Automation Society Paper #74-834, October 1974.

Seebold, J. G., "Reduce Noise in Process Piping," *Hydrocarbon Processing,* October 1982.

Shea, A. K., "A Comparative Study of Sound Level Prediction Methods for Control Valves," *Proceedings of Noise-Con 83,* Institute of Noise Control Engineering, March 1983.

Singleton, E. W., "The Impact of IEC 534-8-3 on Control Valve Aerodynamic Noise Prediction," *Measurement & Control,* Vol. 32, pp. 37–44, March 1999.

VDMA 24422 – 1989, *Richtlinien für die Geräuschberechnung Regel-und Absperrarmaturen,* Verband Deutscher Maschinen- und Anlagenbaue. V., January 1989.

6.15 Sizing

B. G. LIPTÁK (1970) **C. G. LANGFORD** (1985) **F. M. CAIN** (1995) **H. D. BAUMANN** (2005)

Reviewed by F. M. Cain and B. G. Lipták

INTRODUCTION

The control valve is an important component of the total process control loop, and proper valve sizing is essential for this final control element to fulfill its role. There are two important prerequisites for the control valve to be able to fulfill its role: First is to have the correct process data. This includes the accurate knowledge of the maximum and the minimum flow conditions, the available pressure drops across the valve at the various flow conditions, the maximum inlet pressure and inlet temperature, the viscosity of the fluid, and whether or not the fluid is two-phased. Correct selection of the valve's pressure drop, for example, depends on the head loss in the piping system or the characteristic of a pump, to name just two system variables (Figure 6.1b). As the reader will notice, just the selection of the process conditions involves considerable uncertainties.

The second criterion is the selection of the correct valve sizing equation for the prevailing operating conditions. Therefore, before starting the process of sizing, the process control engineer has to determine which of the following sizing basis are applicable. For liquid flow sizing, they include 1) turbulent non-choked, 2) turbulent, choked, 3) saturated, 4) laminar (viscous), 5) non-Newtonian, and 6) two-phase. For gas flows, the sizing conditions can be: 1) turbulent non-choked, 2) choked, and 3) laminar (small flow valves).

The Standard IEC 60534-2-1 and the Instrumentation, Systems, and Automation Society equivalent Standard S 75.01 have improved the accuracy of valve sizing over the years, lowering the sizing errors to less than 5% except for two-phase flow and for non-Newtonian flow conditions.

ABOUT THIS SECTION

For the definition of all the terms used in this section, refer to the "Nomenclature" listing provided at the end of this section. For capacity testing procedures see Section 6.6. The subjects of control valve characteristics and rangeability are discussed in Section 6.7, and the subject of control valve noise and its calculation is covered in Section 6.14. The engineering units used in this section are the U.S. Customary units. Use the numerical constants listed in Tables 6.15l and 6.15p for conversion to metric units.

This section begins with a definition of the valve coefficients (C_v and C_d) and with a brief description of fluid behavior during flashing, cavitation, and under turbulent or laminar flow conditions. After this general discussion the various liquid sizing factors (F_L, F_F, F_P, F_{LP}, and F_R) are defined, and then the liquid sizing equations are provided, as well as some examples. Some of the examples deal with viscous (laminar flow), sizing applications and pipe expander losses, and generalized approaches to liquid sizing. This is followed with a more detailed discussion of cavitation and flashing.

The second half of this section begins with a discussion of valve sizing for gas and vapor services. After the sizing equations are described, an example outlines a generalized approach to gas/vapor sizing. The section is concluded with a description of sizing for two-phase flow applications.

STANDARDS

The study of control valve sizing is based on the science of fluid dynamics. The information in this chapter is an overview of current knowledge, which is continually growing. This know-how has been developed, in large measure, from the following two standards of the Instrumentation, Systems, and Automation Society (ISA): S75.01-2000, "Control Valve Sizing Equations," and S75.02-1996, "Control Valve Capacity Test Procedure." The author is grateful to ISA for its permission to copy and adopt portions of these standards for this section.

The international equivalent of these U.S. standards are IEC 60534-2-1, "Industrial Process Control Valves, Part 2-1: Flow Capacity Sizing Equations for Fluid Flow under Installed Conditions," and IEC 60534-2-3, "Industrial Process Control Valves, Part 2-3, Flow Capacity Test Procedure." The equations in the international editions are essentially equivalent to those in the U.S. standards with the exception of using metric units.

The user is cautioned that the calculations are not perfectly accurate for all conditions. Accuracy is best for water, air, or steam applications using conventional valve designs, which are installed in straight piping. For non-Newtonian fluids or high-viscosity or low Reynolds number flows, these calculations are less accurate. In all cases, reliable calculations require reliable information about the characteristics of the valve and the process fluid, including the fluid's flow rates and pressures at normal and unusual operating conditions.

The English units are used in the various sample calculations and also in the "Nomenclature" listing at the end of the section. While this was necessary because conversion constants are embedded in the working equations, other units can be substituted with appropriate conversions found in Tables 6.15i and 6.15n. The notations used in this section are those used in the ISA standards.

General Principles

The earliest control valves were manual globe valves. Sizing was easy: a 4 in. (100 mm) valve "belonged" in a 4 in. (100 mm) line. Later it became obvious that at higher pressure drops such a valve had too much capacity for good control. Thus evolved the next rule-of-thumb practice, which was to make the control valve one size smaller than line size.

The limitations of such rule-of-thumb methods of valve sizing led to the development of a sizing coefficient based on Bernoulli's theorem for the conservation of energy, and the continuity equation for the conservation of mass. These equations can be combined to express the ideal flow rate through a pipe restriction such as a control valve:

$$Q = A_2 \sqrt{\frac{2g(\Delta H)}{1-(A_2/A_1)^2}} \qquad \textbf{6.15(1)}$$

Calculating the actual flow rate requires the introduction of a discharge coefficient, *C*, which accounts for the real losses in the specific valve. The head loss (ΔH) may also be expressed as pressure drop divided by weight density (γ). Thus,

$$Q = CA_2 \sqrt{\frac{2g(\Delta p/\gamma)}{1-(A_2/A_1)^2}} \qquad \textbf{6.15(2)}$$

By combining the area terms and discharge coefficient (*C*) with the other constants and expressing density in terms of specific gravity, G_f, the flow equation can be written as a function of a capacity parameter, C_v, which is a generally accepted valve capacity parameter called the valve coefficient.

$$Q = C_v \sqrt{\frac{\Delta p}{G_f}} \qquad \textbf{6.15(3)}$$

It should be noted that Equation 6.15(3) is only valid for turbulent and nonchoked (low pressure drop) liquid flows in sizing control valves.

It is important to note that C_v is not a dimensionless coefficient, and in the SI system K_v is used. The conversion between engineering units is covered in later paragraphs of this section. Table 6.15i provides the conversion constants for liquid and Table 6.15n for gas and vapor sizing equations. The C_v valve coefficient, which is based on English units, is widely used in the United States, while K_v, based on the SI system, is the preferred unit in Europe.

The Flow Coefficient

Prior to about 1946 control valves were sized using the flow area of the valve's orifice in square inches together with some, usually proprietary equations in order to estimate the amount of liquid or gas passing through the valve. Thereafter, a new coefficient was introduced called *Cv*, which started standardization in this field. This coefficient denotes the number of (U.S.) gallons of water that a valve would pass when the pressure drop across it is 1 lb/in.^{2}

Under turbulent flow conditions the water velocity is related to the square root of the pressure drop (inlet pressure minus outlet pressure). Therefore, one can multiply this quantity with the C_v number of the valve to obtain the expected flow of cold water in gallons per minute. Later in this section it will be shown how this simple definition of the C_v can be expanded to liquids having specific gravity other than cold water and with the use of conversion factors, for gases as well.

Correction Factors Yet, the turbulent C_v alone turned out to be inaccurate. Using it, occasionally there appeared to be severe sizing errors (up to 50%) that nobody could explain. This problem was resolved in 1963 by the introduction of a Critical Flow Factor.[1] This factor was later renamed the Liquid Pressure Recovery Factor and designated as F_L.

What this factor describes is the amount of pressure recovery, which is occurring downstream of the throttling orifice. A low F_L factor, which is usually the case with streamlined valves, will promote vaporization in liquid streams or cause the development of sonic velocity in gas streams. These effects tend to "choke' the flow and consequently, the valve will pass less than would be predicted if the F_L factor was not considered.

In 1983 two additional factors were introduced. One is called *Xt*, the Pressure Differential Factor of a Control Valve at Choked Flow.[2] The other is called the Expansion Factor *Y*, which enables the calculation of the correct gas density at the valve's orifice.

Yet another, albeit less serious, sizing error source has been noticed in applications where a valve, smaller than line size, was installed between reducers. The error was caused by the pressure drops across the reducers, which lowered the pressure drop available for the valve. In 1968, this problem was overcome by the introduction of a Piping Geometry Factor,[3] F_p.

One more factor was introduced for viscous fluid services, which is called the Reynolds Number Factor,[4] F_R. In 1993, this factor was redefined and its accuracy was improved.[5]

Finally, the Liquid Pressure Ratio Factor, F_F (based on the original work of Allen[6] and others), was introduced for boiling or flashing liquid applications.

Table 6.15a tabulates the various sizing equations for the overall orientation of the reader. The equations listed in this table will be derived in the following paragraphs. In this section, the equations are numbered consecutively, so the

TABLE 6.15a
Orientation Table: Summary of Valve Sizing Equations (for U.S. Customary Units: gpm, SCFH, psi, °F, lbm/hr, lb/ft^3, etc.)

Selection Basis		Fluid State	
		Liquid	Gas or Vapor
Nonchoked, turbulent flow in liquids: $\Delta p < F_L^2\ (P_1 - F_F P_v)$	Volumetric flow in gpm or SCFH	Eq. 6.15(35) $C_v = \frac{q}{F_P}\sqrt{\frac{G_f}{\Delta p}}$	Eq. 6.15(78) $C_v = \frac{Q}{1360 F_P P_l Y}\sqrt{\frac{G_g T_l Z}{x}}$ $Y = 1 - \frac{x}{3 F_k x_T};\ x = \frac{\Delta p}{P_l};\ F_k = \frac{k}{1.4}$
Nonchoked vapors and gases: $\Delta p / P_1 < x_T$	Mass flow in lbm/h	Eq. 6.15(36) $C_v = \frac{w}{63.3 F_P \sqrt{(\Delta p)\gamma_1}}$	Eq. 6.15(77) $C_v = \frac{w}{63.3\ F_P Y \sqrt{x P_l \gamma_l}}$
Choked flow due to cavitation or flashing liquids in liquids, or choked flow in gases	Volumetric flow in gpm or SCFM	Eq 6.15(31) $C_v = \frac{q_{\max}}{F_{LP}}\sqrt{\frac{G_f}{P_1 - F_F P_v}}$ $F_F = 0.96 - 0.28(P_v/P_c)^{1/2}$	Eq. 6.15(80) $C_v = \frac{Q_{\max}}{7320\ F_P P_l Y}\sqrt{\frac{M T_1 Z}{F_K x_T}}$
Choked flow due to sonic velocity in gases or vapors, or choked flow in liquids	Mass flow in lbm/h	Eq. 6.15(32) $C_v = \frac{w_{\max}}{63.3\ F_{LP}\ \sqrt{(P_l - F_F P_v)\gamma_l}}$	Eq. 6.15(79) $C_v = \frac{w_{\max}}{19.3\ F_P P_l Y}\sqrt{\frac{T_1 Z}{F_K x_T M}}$
Piping effect for above equations	Not choked	Eq. 6.15(22) $F_P = \left[1 + \frac{(\sum K) C_d^2}{890}\right]^{-1/2}$ ΣK See Eq. 6.15(21)	
	Choked	Eq. 6.15(23) $F_{LP} = \left[\frac{1}{F_L^2} + \frac{K_i C_d^2}{890}\right]^{-1/2}$ K_i See Eq. 6.15(20)	
Nonturbulent (viscous) flow	Volumetric flow in gpm	Eq. 6.15(63) $C_v = \frac{q}{F_R}\sqrt{\frac{G_f}{\Delta p}}$ where F_R is from Figure 6.15m and is a function of Re_v from Eq. 6.15(64 or 69)	Laminar, i.e., nonturbulent, conditions generally do not occur in gases or vapors except for small flow valves.
	Mass flow in lbm/h	Eq. 6.15(39) $C_v = \frac{w}{63.3\ F_R \sqrt{(\Delta p)\gamma_l}}$	

reader can easily find the derivation of any of the equations listed in the table.

Units In that part of the world where the SI system is used, the valve coefficient is called K_v and it is defined as the flow of cold water in cubic meter per hour units, when the a pressure drop across the valve is one bar. The conversion between C_v and K_v is

$$K_v = 1.17 C_v \qquad \textbf{6.15(4)}$$

The flow coefficient can be converted into an area that is the area of the vena contracta, or the net flow area inside a single stage valve orifice, in accordance with Equation 6.15(5):

$$A_v = C_v \times F_L \times 0.026 \text{ in square inches} \qquad \textbf{6.15(5)}$$

More refined equations are applicable for multistage valves but are not yet standardized.[7] Here the reader may refer to their respective vendors for assistance.

There has been substantial progress in increasing the accuracy of valve sizing, yet additional research is needed in the areas of mixed phase and non-Newtonian flows.

LIQUID SIZING

Fluid flowing through a control valve obeys the basic laws of the conservation of mass and energy. As the fluid stream approaches the valve restriction, fluid velocity increases in order to pass the same amount of flow through the restricted area. The restriction inside the valve is the result of moving a closure member (e.g., plug, disk, ball) closer to the valve seat.

Energy to accelerate the fluid comes from a corresponding decrease in static pressure, as illustrated in Figure 6.15b. Maximum velocity and minimum static pressure occur immediately downstream from the minimum area of valve restriction (the throttling point), the most constricted area of the flow stream. This minimum pressure point downstream from the throttling point is known as the vena contracta.

Figure 6.15b shows that as the velocity decreases after the vena contracta, some of the kinetic energy is converted back into static pressure. This conversion is called the pressure recovery of the valve, and its amount is equal to $P_2 - P_{vc}$. Valves with large pressure recovery relative to $P_1 - P_2$ are called high recovery valves. Such valves include most rotary and gate valves. The net pressure loss ($P_1 - P_2$) between the valve inlet and outlet is due to frictional effects (turbulence) and represents the permanent pressure loss.

Relative Valve Capacity Coefficient (C_d)

The valve capacity coefficient, C_v, increases as valve size increases, but valve geometry is also a major factor in the magnitude of the pressure loss for a given flow rate. It is therefore useful to define the relative valve capacity (C_d) in order to compare the effects of geometry of different valve designs. C_d is defined as

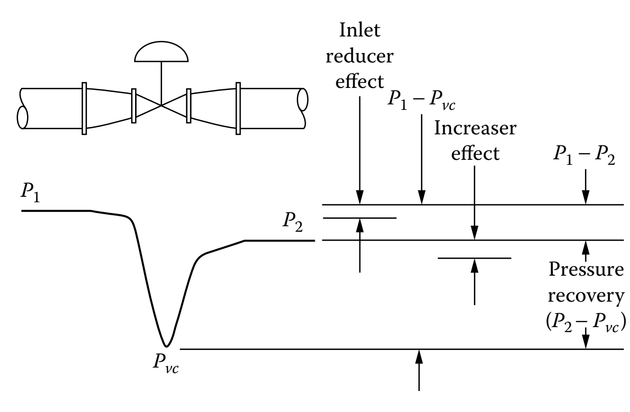


FIG. 6.15b
Pressure profile through a valve.

$$C_d = C_v / d^2 \qquad \textbf{6.15(6)}$$

Representative values of C_d for various valve styles (at full capacity) may be found in Table 6.15c.

C_d can also be useful for evaluating the relative magnitude of pressure recovery. Generally, the valve types with larger C_d values are also more likely to have higher pressure recovery ($P_2 - P_{vc}$); that is, they have a lower F_L factor and therefore tend to "choke" or have liquids evaporate at lower pressure drops ($P_1 - P_2$). The significance of this will become clearer in the discussions of sizing factors for choked flow.

Caution: The published C_v data from manufacturers is based on standard testing methods that include the friction losses associated with the test manifold pipe between the pressure taps (Section 6.6). When the C_d is under 20, the friction loss from eight diameters of this pipe adjacent to the valve is relatively insignificant. However, as the C_d increases above 20, the manifold pressure drop rises exponentially relative to the losses from the valve alone. Therefore, for valves with $C_d > 20$, one must also consider the pressure drop in the valve test manifold piping to avoid significant errors.

In the majority of valve applications the flow is turbulent, and the velocity profile across the restrictive (throttling) area of the valve is relatively uniform. Turbulent flow occurs when the Reynolds number is high (typically above 10,000). This occurs when the velocity is high and the viscosity is low [$\text{Re} = (\gamma U d/\mu)$]. In the turbulent range, the incompressible fluid flow is proportional to the square root of the pressure drop across the valve, as was shown in Equation 6.15(3). This proportionality is shown by the straight line on the left side of Figure 6.15d.

Factors F_L, F_F, F_P, and F_{LP}

Pressure Recovery Factor (F_L) F_L is the pressure recovery factor, which indicates the size of the pressure recovery relative to the valve pressure drop (Figure 6.15b). F_L is defined as

$$F_L = [(P_1 - P_2)/(P_1 - P_{vc})]^{1/2} \qquad \textbf{6.15(7)}$$

Note: F_L can also be used to calculate the pressure drop that causes sonic velocity in the valve orifice. As it has been discussed in Section 6.14, this is important for noise calculations, where

$$\Delta p_{\text{sonc}} = 0.5 \times F_L^2 \times P_1 \qquad \textbf{6.15(8)}$$

The values of F_L for various valve designs are listed in Table 6.15c.

TABLE 6.15c

Representative Values of Relative Valve Capacity Coefficients (C_d) and of Other Sizing Factors for a Variety of Valve Designs. The Cd Values Listed are for Valves with Full Area Trims, When the Valve is Fully Open

Valve Type	*Trim Type*	*Flow Direction**	X_T	F_L	F_d	F_s	$C_d = C_v/d^2$**	K_c
GLOBE								
Single-port	Ported plug, 4 port	Either	0.70	0.90	0.48	1.0	9.5	0.65
	Contoured plug	Open	0.72	0.90	0.46	1.1	11	0.65
		Close	0.55	0.80	1.0	1.1	11	0.58
	Characterized cage, 4 port	Open	0.75	0.90	0.45	1.1	14	0.65
		Close	0.75	0.85	0.41	1.1	16	0.60
	Wing-guided, 3 wings	Either	0.75	0.90	0.58	1.1	11	0.60
Double-port	Ported plug	Either	0.75	0.90	0.28	0.84	12.5	0.80
	Contoured plug	Either	0.70	0.85	0.32	0.85	13	0.70
	Wing-guided	Either	0.75	0.90	0.41	0.84	14	0.80
Rotary	Eccentric spherical plug	Open	0.60	0.85	0.42	1.1	12	0.60
		Close	0.40	0.68	0.42	1.2	13.5	0.35
ANGLE	Contoured plug	Open	0.72	0.90	0.46	1.1	17	0.65
		Close	0.65	0.80	1.0	1.1	20	0.55
	Characterized cage, 4 port	Open	0.65	0.85	0.45	1.1	12	0.60
		Close	0.60	0.80	1.0	1.1	12	0.55
	Venturi	Close	0.20	0.50	1.0	1.3	22	0.21
BALL	Segmented (throttling)	Open	0.30	0.80	0.98	1.2	25	0.25
	Standard port (diameter ≈ 0.8d)	Either	0.42	0.74	0.99	1.3	30	0.20
BUTTERFLY	60°, no offset seat	Either	0.42	0.70	0.5	0.95	17.5	0.39
	90°, offset seat	Either	0.35	0.60	0.45	0.98	29	0.32
	90°, no offset seat	Either	0.08	0.53	0.45	1.2	40	0.12

*Flow direction tends to open or close the valve, i.e., push the closure member away from or toward the seat.

**In this table, *d* may be taken as the nominal valve size, in inches.

Liquid Critical Pressure Ratio Factor (F_F) The complex geometry of most valves makes experimental measurement of the pressure at the vena contracta (P_{vc}) impossible. The ISA sizing equations[9] use the liquid critical pressure ratio factor, F_F, for approximating P_{vc} used for saturated liquid flow conditions.

$$P_{\text{vc}} = F_F P_v \tag{6.15(9)}$$

where F_F can be approximated by

$$F_F \cong 0.96 - 0.28\,(P_v/P_c)^{1/2} \tag{6.15(10)}$$

A graph of this relationship is plotted in Figure 6.15e. Equation 6.15(10) is based on the assumption that the fluid is always in thermodynamic equilibrium. However, a liquid does not remain in thermodynamic equilibrium as it flashes across the valve restriction; therefore, the actual flow rate can be greater than that predicted by Equation 6.15(10). Estimating P_{vc} by using F_F is not fully agreed upon by experts but it is accepted by the valve sizing standard (ISA S75.01) and is in common use in the control valve industry.

The pressure drop required to cause choked flow in liquids is given by the following equation:

$$\Delta P_{\text{choked}} = F_L^2\ (P_1 - F_F P_v) \tag{6.15(11)}$$

Piping Geometry Factor (F_P and F_{LP}) By convention, valve tests and calculations have included a portion of piping adjacent to the valve. P_1 is measured upstream of the pipe reducer,

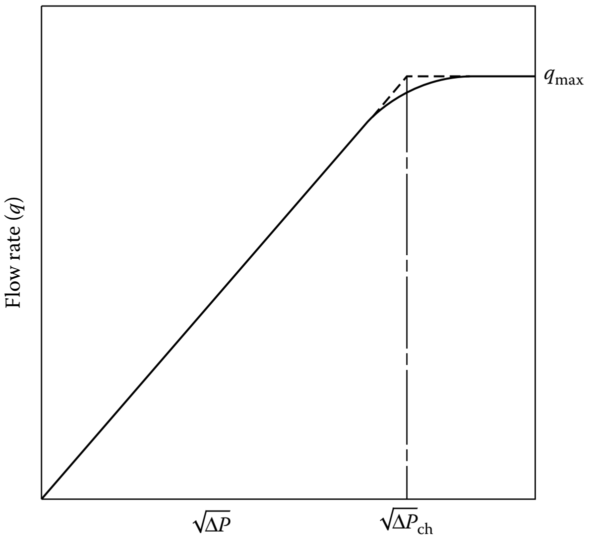


FIG. 6.15d
Typical flow rate vs. $\sqrt{\Delta p}$ *curve for liquid at constant upstream pressure and vapor pressure. The straight line segment on the left shows the behavior of turbulent liquid flow. As the pressure drop rises the pressure at the vena contracta drops and when it reached the vapor pressure of the liquid, vaporization starts. From this point on (right side of figure) the relationship is no longer linear, until at* q_{max} *the choked flow condition is reached.*

if used, and P_2 is downstream of the increaser. Some manufacturers supply tables for C_v for valves with standard pipe reducers. Where this data is based on actual test data, it should be used in preference to the approximations shown below. Elbows and block valves installed near the valve will upset the flow profile and cause some reduction in capacity. No procedure is available to predict these losses.

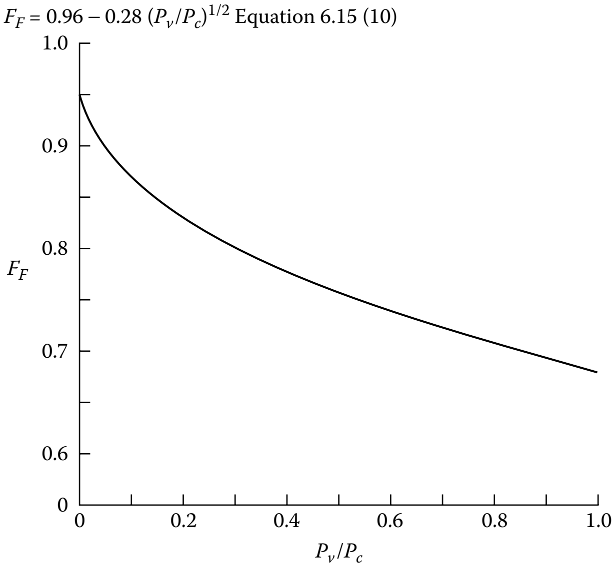


FIG. 6.15e
Liquid critical pressure ratio factor F_F*, plotted as a function of the* P_v/P_c *ratio.*

The standard reducer and increaser fittings result in an abrupt change in size. As a result of the change in cross-section, first there is an irreversible energy loss due to turbulence (friction loss), and second there is the conversion between pressure energy and velocity energy. If the control valve is smaller than the line size, the pressure at the valve inlet is reduced by the friction loss and is also reduced because the smaller flow area requires acceleration to a higher velocity.

If the downstream piping is the same size as the inlet piping, then the velocity-energy/pressure-energy exchange at the outlet is reversed and it cancels the effect at the inlet. Friction losses are always additives. For a reducer at the valve inlet, the ISA standard determines the inlet fitting friction loss coefficient:

$$K_1 = 0.5(1-(d/D_1)^2)^2 \qquad \textbf{6.15(12)}$$

For an increaser at the valve outlet the ISA standard defines the outlet fitting friction loss coefficient:

$$K_2 = (1-(d/D_2)^2)^2 \qquad \textbf{6.15(13)}$$

Although not explicitly stated by ISA S75.01, an analysis of the references shows that for the unusual case where an increase might exist at the valve inlet, K_1 is calculated as

$$K_1 = (1-(D_1/d)^2)^2 \qquad \textbf{6.15(14)}$$

and for a reducer at the outlet, K_2 is found as

$$K_2 = 0.5(1-(D_2/d)^2)^2 \qquad \textbf{6.15(15)}$$

For the velocity-pressure exchange (Bernoulli effect) for the usual case:

$$\text{reducer at inlet } K_{B1} = 1-(d/D_1)^4 \qquad \textbf{6.15(16)}$$

$$\text{increaser at outlet } K_{B2} = (d/D_2)^4 - 1 \qquad \textbf{6.15(17)}$$

It can also be shown from the same basic fluid mechanics for the unusual case:

$$\text{increaser at inlet } K_{B1} = (D_1/d)^4 - 1 \qquad \textbf{6.15(18)}$$

$$\text{reducer at outlet } K_{B2} = 1-(D_2/d)^4 \qquad \textbf{6.15(19)}$$

Certain combinations of these coefficients are used in calculations:

$$K_i = K_1 + K_{B1} \text{ (inlet combination)} \qquad \textbf{6.15(20)}$$

$$K = K_i + K_2 + K_{B2} \qquad \textbf{6.15(21)}$$

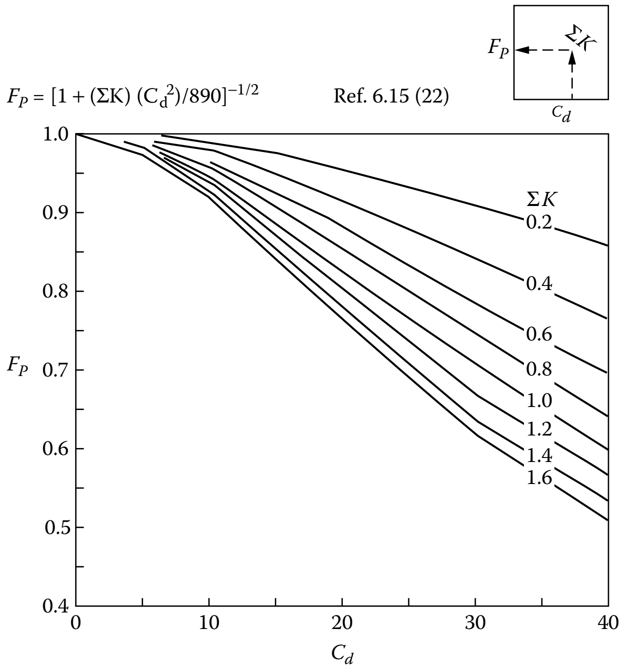


FIG. 6.15f
Piping geometry factor as a function of C_d and K.

A piping geometry factor is defined as:

$$F_P = [1 + (\sum K)(C_d^2)/890]^{-1/2} \qquad \mathbf{6.15(22)}$$

Valve flow capacity is directly proportional to this F_P factor (see Figure 6.15f).

Note that capacity is reduced for high values of ΣK combined with high C_d. Because the inlet fittings also affect the valve inlet pressure, a combination of F_L and F_P is used to predict choking or cavitation (see Figure 6.15g):

$$F_{LP} = [1/F_L^2 + (K_i)(C_d^2)/890]^{-1/2} \qquad \mathbf{6.15(23)}$$

Again note the large loss in capacity caused by reducers on valves with high C_d. This equation provides only an estimate. The actual point where choking starts is preferred to be determined by test.

$$(P_1 - P_{vc})(F_{LP}/F_P)^2 = \Delta p \text{ critical} \qquad \mathbf{6.15(24)}$$

For all practical purposes, the flow at rated valve capacity is always turbulent (otherwise the outlet velocity would be too high). This makes correcting for F_p much simpler and avoids most iterations. Figure 6.15h provides the piping configuration factor (F_p) for two pipe-to-valve diameter ratios. The curve on the left is for a $D/d = 2$, and on the right for a $D/d = 1.5$.

Installing a valve between reducers and expanders can lead to the distortion of the valve's inherent flow characteristic. This in turn can alter the gain of the valve (the percentage change in flow resulting from 1% change in control signal).

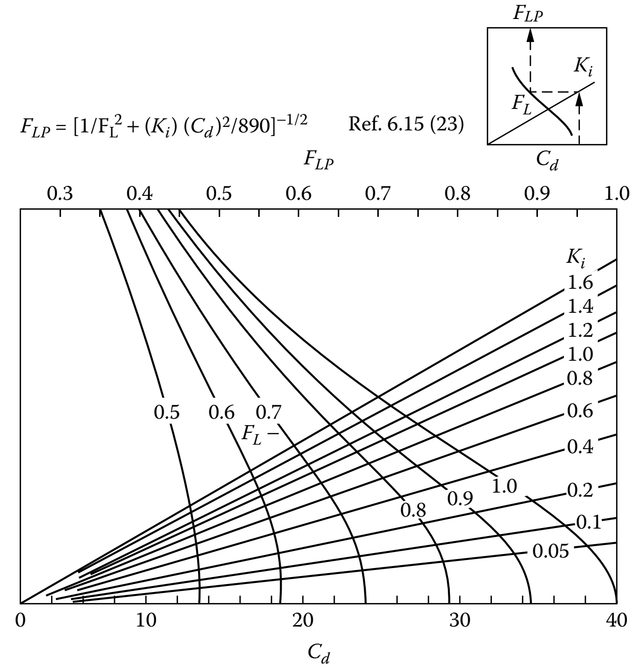


FIG. 6.15g
Combined factor (F_{LP}) for liquids as a function of C_d, K_i, and F_L.

F_p Correction Shortcut The correction to the required C_v for reducer and expander losses can also be obtained by the following steps: 1) Calculate the C_v for nonchoked liquid or gas flows. 2) Estimate the valve size d (inches) by determining the required C_d (dividing the calculated C_v by d^2). 3) Make sure that this required C_d is at least 40% smaller than the typical values listed in Table 6.15c. 4) Divide the pipe diameter by the chosen valve diameter to obtain the D/d ratio. 5) Select the applicable curve on Figure 6.15h. 6) Read F_p from the applicable graph using the calculated C_d (required C_v/valve size squared). 7) Divide the required system C_v by F_p. This then gives the required and corrected valve C_v. 8)

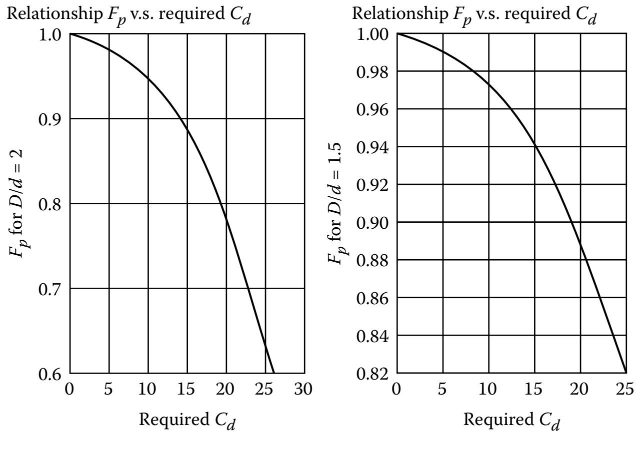


FIG. 6.15h
Correction factor for pipe reducers and expanders. The curve on the left is for a pipe to valve diameter ratio (D/d) of 2 and on the right, the curve is for D/d = 1.5.

Check the manufacturer's literature to make sure that the C_v of the chosen valve size is sufficient.

EXAMPLE

Assume that the required valve C_v was calculated from the flow data as 70 and the pipe size is 4 in. If a line size valve is assumed, the corresponding $C_d = 70/4^2 = 4.4$. This then could result in the selection of a 3– or 4-in. globe valve.

However, one might consider a less costly 2-in. ball valve. In that case the required $C_d = 70/2^2 = 17.5$. Consulting Figure 6.15h for a $D/d = 2$ (curve on the left), the factor F_p for a C_d of 17.5 is 0.84. Therefore, the required operating $C_v = 70/0.84 = 83.3$. For such a requirement, one needs to select a valve with a catalog C_v of 100 or more. Based on Table 6.15c, a 2-in. segmented ball valve can provide that, so the choice is acceptable.

Units Used in Valve Sizing

USA Units The valve sizing equations described here are based on Instrumentation, Systems, and Automation Society Standard S75.01, "Flow Equations for Sizing Control Valves."[9] For turbulent, nonvaporizing flow conditions, the flow rate in gallons per minute is calculated in accordance with Equation 6.15(25):

$$q = F_P C_v \sqrt{\frac{P_1 - P_2}{G_f}} \qquad \textbf{6.15(25)}$$

The valve coefficient may be calculated as

$$C_v = \frac{q}{F_P}\sqrt{\frac{G_f}{P_1 - P_2}} \qquad \textbf{6.15(26)}$$

where

P_1 and P_2 are in units of lb/in.^{2} absolute.

If flow rate, w, is given in lbm/hr and specific weight, γ_1, is in lbm/ft^3, then

$$w = 63.3\, F_P C_v \sqrt{(P_1 - P_2)\gamma_1} \qquad \textbf{6.15(27)}$$

or if solved for C_v:

$$C_v = \frac{w}{63.3\, F_P \sqrt{(P_1 - P_2)\gamma_1}} \qquad \textbf{6.15(28)}$$

Under choked flow conditions, the maximum flow rate is no longer dependent on the pressure drop but rather on the pressure drop at the onset of choking (Δp_{ch} in Figure 6.15d). See Equation 6.15(9). Therefore, the maximum volumetric flow rate, q, under choked conditions, in gallons/minute is

$$q_{\max} = F_L C_v \sqrt{\frac{P_1 - F_F P_v}{G_f}} \qquad \textbf{6.15(29)}$$

The maximum mass flow rate, w, in lbm/hr is

$$w_{\max} = 63.3\, F_L C_v \sqrt{(P_1 - F_F P_v)\gamma_1} \qquad \textbf{6.15(30)}$$

When reducers, increasers, or other fittings are installed between the pipe and the valve, a combined liquid pressure recovery factor, F_{LP}, is used. See the discussion on liquid sizing factors in connection with Equation 6.15(23).The volumetric flow rate becomes

$$\left.\begin{aligned} q_{\max} &= F_{LP} C_v \sqrt{\frac{P_1 - F_F P_v}{G_f}} \\ \text{or if solved for } C_v\text{:}\quad C_v &= \frac{q_{\max}}{F_{LP}}\sqrt{\frac{G_f}{P_1 - F_F P_v}} \end{aligned}\right\} \qquad \textbf{6.15(31)}$$

The mass flow rate is calculated by

$$\left.\begin{aligned} w_{\max} &= 63.3\, F_L C_v \sqrt{(P_1 - F_F P_v)\gamma_1} \\ \text{or if solved for } C_v\text{:}\quad C_v &= \frac{w_{\max}}{63.3\, F_{LP} \sqrt{(P_1 - F_F P_v)\gamma_1}} \end{aligned}\right\} \qquad \textbf{6.15(32)}$$

The pressure drop at the onset of choking (Δp_{ch}) must always be calculated to check if choking will occur. It is convenient to write the above Equations 6.15(25–32) using the allowable sizing pressure drop, Δp_a, which is defined as the smaller value of the actual pressure drop, $P_1 - P_2$, and the pressure drop at the onset choking, Δp_{ch}.

The Reynolds number factor, F_R, may also be included in the equation. Refer to the discussion on liquid sizing factors for nonturbulent in the later paragraphs discussing laminar and viscous flow. Thus, the following set of simplified equations can be written for volumetric flow rate (gpm):

$$q = F_P F_R C_v \sqrt{\frac{\Delta p_a}{G_f}} \qquad \textbf{6.15(33)}$$

and if solved for C_v:

$$C_v = \frac{q}{F_R F_P}\sqrt{\frac{G_f}{\Delta p_a}} \qquad \textbf{6.15(34)}$$

Other Engineering Units The equations presented previously were derived using U.S. Customary Units. In these equations C_v is not a dimensionless coefficient. Therefore, it is helpful to develop units conversion constants that will permit the use of these equations with other systems of units. ISA has chosen to write the equations using numerical constants (N_1, N_2, etc.) that take on unique values for specific

TABLE 6.15i
*Numerical Constants for the Conversion of Liquid Flow Equations**

Constant		Units Used in Equations					
	N	*w*	*Q*	*P*, ΔP	*d*, *D*	γ_l	ν *(nu)*
N_1	0.0865	—	m^3/h	KPa	—	—	—
	0.865	—	m^3/h	Bar	—	—	—
	1.00	—	gpm	psia	—	—	—
N_2	0.00214	—	—	—	mm	—	—
	890	—	—	—	in.	—	—
N_4	76,000	—	m^3/h	—	mm	—	Centistokes**
	17,300	—	gpm	—	in.	—	Centistokes**
N_6	2.73	kg/h	—	KPa	—	kg/m^3	—
	27.3	kg/h	—	Bar	—	kg/m^3	—
	63.3	lb/h	—	psia	—	lb/ft^3	—

*Reprinted by permission. Copyright © 1985, Instrumentation, Systems, and Automation Society. From ANSI/ISA-S75.01, "Flow Equations for Sizing Control Valves."

**To convert m^2/s to centistokes, multiply m^2/s by 10^6. To convert centipoise to centistokes, divide centipoise by G_f.

combinations of units. Equations 6.15(35–41) and Table 6.15i are from ISA S75.01.*

$$\left.\begin{aligned} q &= N_1 F_P C_v \sqrt{\frac{P_1 - P_2}{G_f}} \\ \text{or} \\ C_v &= \frac{q}{N_1 F_p}\sqrt{\frac{G_f}{P_1 - P_2}} \end{aligned}\right\} \qquad \textbf{6.15(35)}$$

$$\left.\begin{aligned} w &= N_6 F_P C_v \sqrt{(P_1 - P_2)\gamma_1} \\ \text{or} \\ C_v &= \frac{w}{N_6 F_P \sqrt{(P_1 - P_2)\gamma_1}} \end{aligned}\right\} \qquad \textbf{6.15(36)}$$

where:

$$F_P = \left[\frac{(\Sigma K)C_v^2}{N_2 d^4} + 1\right]^{-1/2} \qquad \textbf{6.15(37)}$$

The corresponding nonturbulent equations* are

$$\left.\begin{aligned} q &= N_1 F_R C_v \sqrt{\frac{P_1 - P_2}{G_f}} \\ \text{or} \\ C_v &= \frac{q}{N_1 F_R}\sqrt{\frac{G_f}{P_1 - P_2}} \end{aligned}\right\} \qquad \textbf{6.15(38)}$$

$$\left.\begin{aligned} w &= N_6 F_R C_v \sqrt{(P_1 - P_2)\gamma_1} \\ \text{or} \\ C_v &= \frac{w}{N_6 F_R \sqrt{(P_1 - P_2)\gamma_1}} \end{aligned}\right\} \qquad \textbf{6.15(39)}$$

The valve Reynolds number is defined* as

$$\text{Re}_v = \frac{N_4 F_d q}{\nu F_L^{1/2} C_v^{1/2}}\left(\frac{F_L^2 C_v^2}{N_2 d^4} + 1\right)^{1/4} \qquad \textbf{6.15(40)}$$

The equations* for determining maximum flow rate of a liquid under choked conditions for valves in straight pipes of the same size as the valve are as follows:

$$\left.\begin{aligned} q_{\max} &= N_1 F_L C_v \sqrt{\frac{P_1 - F_F P_v}{G_f}} \\ \text{or} \\ C_v &= \frac{q_{\max}}{N_1 F_L}\sqrt{\frac{G_f}{P_1 - F_F P_v}} \end{aligned}\right\} \qquad \textbf{6.15(41)}$$

Metric Capacity Coefficients, K_v and A_v Valve capacity coefficients, which have been derived in a similar manner as was the C_v but which have their basis in the metric or SI system of units, can be easily converted to their equivalent C_v values. If units for volumetric flow rate (q) are in cubic meters per hour (m^3/h) and the pressure drop (Δp), is given in bars (1 bar = 100 kPa), then the capacity coefficient K_v is

$$K_v = q(G_f/\Delta p)^{1/2} \qquad \textbf{6.15(42)}$$

---

* Reprinted by permission. Copyright © 1985, Instrumentation, Systems, and Automation Society. From ANSI/ISA-S75.01, "Flow Equations for Sizing Control Valves."

Simple units conversions lead to the following relationship between K_v and C_v:

$$K_v = 0.865C_v \qquad \textbf{6.15(43)}$$

The derivation of the capacity coefficient, A_v, yields an equivalent area in square meters (m^2) that is related to C_v by the conversion

$$A_v = 24 \times 10^{-6} C_v \qquad \textbf{6.15(44)}$$

It should be noted that valve capacity or sizing coefficients are defined in terms of a flow test with pressure taps located in straight pipe of the same nominal size as the valve and at specific distances upstream and downstream of the valve. The test configuration for determining C_v is well defined in ISA Standard S75.02[10] and is discussed in Section 6.6. The above conversions are valid only if K_v or A_v are determined using the same test manifold configuration and method as those specified in ISA-S75.02.

Sizing Example for Liquids

The above working equations can be used in a logical sequence to determine the required valve capacity for a given application. The following example illustrates the basic steps. However, there are other considerations when selecting a valve, such as noise, cavitation, corrosion, and actuator sizing, that are discussed in other sections of this chapter.

Example 1 Determine the required valve capacity coefficient (C_v) and valve size for the following application:

Liquid: Water
Critical pressure (P_c): 3206.2 psia
Temperature: 250°F
Upstream pressure (P_1): 314.7 psia
Downstream pressure (P_2): 204.7 psia
Flow direction (relative to valve plug): Flow-under-to-open
Line size/class: 4 in., ANSI Class 600
Flow rate (q): 500 gpm
Liquid vapor pressure (P_v): 30 psia
Kinematic viscosity (v): 0.014 centistokes
Valve characteristics: Equal percentage

Step 1: Calculate actual Δp.

$$\Delta p = P_1 - P_2 = 314.7 - 204.7 = 110 \text{ psi}$$

Step 2: Calculate choked pressure drop Δp_{ch} and determine allowable sizing pressure drop Δp_a. From Equation 6.15(10):

$$\begin{aligned} F_F &= 0.96 - 0.28\,(P_v/P_c)^{1/2} \\ &= 0.96 - 0.28(30/3206.2)^{1/2} = 0.93 \end{aligned}$$

In order to determine the Δp_{ch}, a preliminary valve-type selection must be made in order to establish the liquid pressure recovery coefficient F_L to be used in the calculation. Assume the use of a single-ported globe valve, which has an F_L equal to 0.9 (Table 6.15c). Solving for Δp_{ch} from Equation 6.15(11):

$$\Delta p_{\text{ch}} = F_L^2[P_1 - F_F P_v] = 0.9^2[314.7 - (0.93)(30)] = 232.3 \text{ psi}$$

The allowable sizing drop Δp_a is the smaller of Δp and Δp_{ch}. Therefore, we have determined that $\Delta p_a = 110$ psi, and the flow is not choked.

Step 3: Determine the specific gravity, G_f. The Instrumentation, Systems, and Automation Society sizing equations are based on water at 60°F with a density of 62.37 lb/ft^3. From water properties data we find the density of water at 250°F is 58.8 lb/ft^3.

$$G_f = 58.80/62.37 = 0.94$$

Step 4: Calculate the approximate required C_v assuming F_P and F_R equal 1.0. From Equation 6.15(34):

$$C_v = \frac{q}{F_R F_P}\sqrt{\frac{G_f}{\Delta p_a}} = 500(0.94/110)^{1/2} = 46.2$$

Step 5: Select the approximate body size based on the approximate C_v from Step 4. From manufacturers' catalogs it can be determined that the smallest globe valve body size that will accommodate C_v 46.2 is a 2 in. size with a maximum valve capacity coefficient of C_v51.

Step 6: Determine the piping geometry factor F_P. If manufacturer's test data are not available for the selected valve type, F_p may be estimated using Equations 6.15(12–22) or Figure 6.15f.

$$d = 2.0,\, D_1 = D_2 = 4.0, \quad \text{and} \quad d/D_1 = 0.50$$

Therefore, in Figure 6.15h, the left side ($D/d = 2$) curve should be used. For a reducer on the inlet and an increaser on the valve outlet the following values are calculated:

$$K_1 = 0.5[1-(0.5)^2]^2 = 0.2813$$

$$K_2 = [1-(0.5)^2]^2 = 0.5625$$

$$K_{B1} = 1-(0.5)^4 = 0.9375$$

$$K_{B2} = (0.5)^4 - 1 = -0.9375$$

$$\sum K = 0.2813 + 0.5625 + 0.9375 - 0.9375 = 0.8438$$

$$C_d = 46.2/(2)^2 = 11.55$$

$$\begin{aligned} F_P &= [1 + (\textstyle\sum K)(C_d)^2/890]^{-1/2} \\ &= [1 + (0.8438)(11.55)^2/890]^{-1/2} = 0.94 \end{aligned}$$

Step 7: Calculate the Reynolds number factor, F_R. For most process applications, this step is not necessary, because in most cases turbulent flow (i.e., $\mathrm{Re}_v > 10{,}000$) conditions exist. The calculation is shown here for illustration purposes. From Equation 6.15(64):

$$\mathrm{Re}_v = \frac{N_4 F_d q}{v F_L^{1/2} C_v^{1/2}}\left(\frac{(F_L C_d)^2}{N_2}+1\right)^{1/4}$$

$$\mathrm{Re}_v = \frac{(17300)(1)(500)}{(0.014)\sqrt{(0.9)(46.2)}}\left\{\frac{[(0.9)(11.55)]^2}{890}+1\right\}1/4$$

$$\mathrm{Re}_v = 98.6\times 10^6$$

Referring to Figure 6.15m, $F_R = 1$.

Step 8: Calculate the final required C_v using Equation 6.15(34) and the values for F_p and F_R that were calculated above:

$$C_v = \frac{q}{F_R F_P}\sqrt{\frac{G_f}{\Delta p_a}} = (500)/[(0.94)(1)](0.94/110)^{1/2} = 49.1$$

Note: It was noted in Step 5 that the maximum C_v for a 2-in. globe valve was found to be $C_{v(\max)} = 51$. The required $C_{v(\mathrm{req})} = 49.1$ is too close to the maximum, because it exceeds 85% of it. Therefore, a 3-in. globe valve should be selected.

Step 9: Calculate the fluid velocity at the 3-in. valve outlet using Equation 6.15(45) with appropriate units conversions.

$$U = q/A_o \qquad \textbf{6.15(45)}$$

where

U = velocity
q = volumetric flow rate
A_o = area of the body outlet port

If q is given in gallons per minute (gpm) and A_o is in square inches, then U can be calculated in feet per second:

$$U_{(\mathrm{fps})} = 0.321\, q_{(\mathrm{gpm})}/A_{o(\mathrm{sq.in.})} \qquad \textbf{6.15(46)}$$

For this example, the outlet velocity is

$U = 0.321(500\text{ gpm}/7.07\text{ sq. in.}) = 22.7\text{ fps}$

Evaluation: The valve sizing and outlet velocity of 22.7 fps is of consequence only under flashing (liquid/vapor) conditions, or where the fluid contains erosive particles. In such cases, the valve manufacturer should also be asked to double-check the sizing for the particular service.

The Cavitation Phenomenon

Cavitation occurs when static pressure anywhere in the valve drops to or below the vapor pressure of the process liquid. Vaporization begins around microscopic gaseous nuclei. The cavitation process includes the vapor cavity formation and sudden condensation (collapse) of the vapor bubble driven by pressure changes. The basic process of cavitation is related to the conservation of energy and Bernoulli's theorem, which describes the pressure profile of a liquid flowing through a restriction or orifice (Figure 6.15b).

In order to accelerate the fluid through the restriction, some of the pressure head is converted into velocity head. This transfer of static energy is needed to maintain the same mass flow through the reduced passage. The fluid accelerates to its maximum velocity, which corresponds to the point of minimum pressure (vena contracta). The fluid velocity gradually slows down as it expands back to the full pipe area. The static pressure also recovers somewhat, but part of it is lost due to turbulence and friction.

If the static pressure at any point drops below the liquid vapor pressure (P_v) corresponding to the process temperature, then vapor bubbles will form. If enough energy is imparted to the growing vapor bubble to overcome surface tension effects, the bubble will reach a critical diameter and expand rapidly. As the static pressure recovers to a point greater than the vapor pressure, the vapor will condense, causing the bubbles to collapse violently back into their liquid phase. The growth and collapse of the bubbles produce high-energy shock waves in the fluid.

The collapse stage of the process (the bubble implosion) produces the more severe shock waves. Shock waves and liquid microjets radiate for short distances from imploding cavities and erode nearby surfaces. Cavitation can cause erosion, noise, and vibration in piping systems and therefore must be avoided. Extensive cavitation also causes choked flow conditions in the valve.

Predicting and Mitigating Cavitation Sizing a valve in liquid service for choked flow allows one to determine its maximum flow capacity, but this is of limited value, because most liquid-service valves should not be operated under choked conditions. Special trim designs with multiple stages or multiple flow paths like those in Figure 6.1y are typically used to prevent severe cavitation and are better able to operate at or near choked conditions without damage.

Metal erosion from cavitation damage has a very distinctive appearance, like that of cinder block or sandblasting. No known material will withstand continuous, severe cavitation without damage and eventual failure. The length of time it will take is a function of the fluid, metal type, and severity of the cavitation. Without special trim geometry, mild or intermittent cavitation can be mitigated in standard valves by the use of extremely hard trim materials or overlays.

Sometimes it is feasible to increase the downstream backpressure, or limit the pressure drop by installing control valves in series to reduce the pressure drop in each valve. Another mitigating effect in some processes is a result of the fluid thermodynamic properties and operating conditions. As liquid operating temperature approaches its critical temperature, the growth and collapse rates of cavities slow down, as heat transfer effects become increasingly significant relative to the dominant inertial effects. This can greatly reduce the energy of cavity collapse and surface damage to the valve parts.

Some cryogenic and hydrocarbon applications are thought to behave this way, which may partly explain why cavitation damage in these cases is minimal or absent even when cavitation is present in the valve. Further information about cavitation and predicting its effects on valve performance can be found in the ISA Recommended Practice RP75.23.01, "Considerations for Evaluating Control Valve Cavitation."

Various methods are in use to establish the pressure drop at which cavitation starts. Two of the common techniques are described here. The first is based on a cavitation index, K_c, determined from flow capacity test data similar to that shown in Figure 6.15d. The second is based on a cavitation index, σ, and vibration test data.

Flow Curve Cavitation Index (K_c) Figure 6.15j illustrates the effect that cavitation has on the linear relationship between flow rate and the square root of pressure drop. The inflection point, the point at which the linear slope of the C_v curve marks the Δp, at which measurable amounts of vaporization exist is measured by a flow test.[10]

The calculation of the index K_c is based on the assumption that a valve may operate cavitation-free at any pressure drop that is less than that associated with K_c, which is defined as[15]

$$K_c = \frac{\Delta P_i}{P_1 - P_v} \qquad \textbf{6.15(47)}$$

where Δp_i is the pressure drop at the inflection in the tested C_v slope. K_c can then be compared to the index describing the actual service conditions of an application in question, K_{sc}:

$$K_{sc} = \frac{\Delta P}{P_1 - P_v} \qquad \textbf{6.15(48)}$$

If K_{sc} for the service is less than the K_c for the valve, the valve is assumed to be adequate for the service. Estimated K_c values for various valve types are shown in Table 6.15c. These values of K_c are only representative and should always be verified with the valve manufacturer. However, care must be exercised with this method, because it has not consistently indicated damaging cavitation for some valve designs.[2,15,18]

Vibration Curve Cavitation Index (σ) Noise and vibration measurements have identified four recurring regimes of cavitation in many valves.[19–22] Figure 6.15k illustrates the relationship between acceleration measurements taken on the downstream pipe and a cavitation index called *sigma* (σ).

There are several forms of sigma, all of which represent a ratio between the forces resisting cavitation and the forces promoting cavitation. The following forms of sigma have been used for valves.

$$\sigma_1 = \frac{P_1 - P_v}{P_1 - P_2} \qquad \textbf{6.15(49)}$$

$$\sigma_2 = \frac{P_2 - P_v}{P_1 - P_2} \qquad \textbf{6.15(50)}$$

Upstream and downstream pressures should be corrected for piping and fitting losses to obtain net values of P_1 and P_2 at the valve for calculating σ when $C_v/d^2 > 20$. The σ indexes

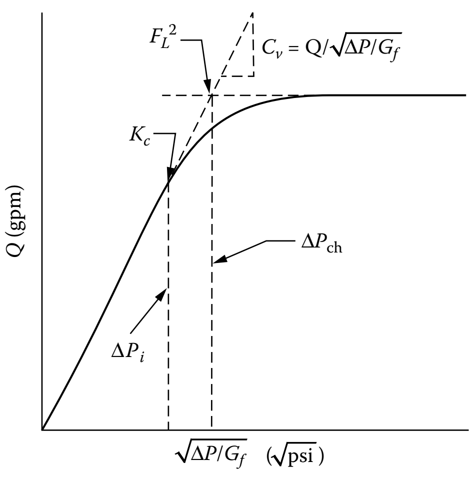


FIG. 6.15j
The flow rate curve showing the effect of choking and the methods how F_L and K_c are determined. (Copyright © 1991, Instrumentation, Systems, and Automation Society. From ISA Paper No. 91-0462, "Solving the Problem of Cavitation in Control Valves.")

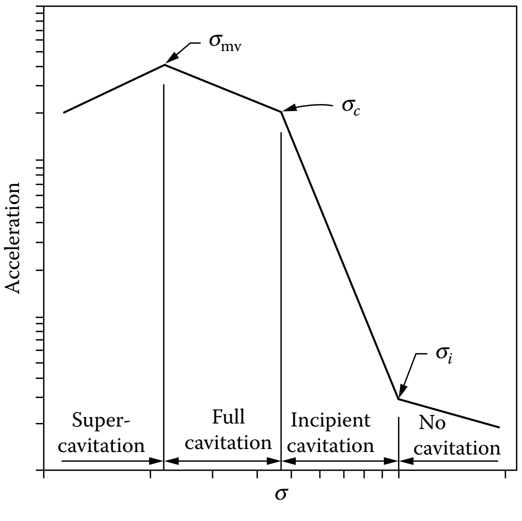


FIG. 6.15k
Classical cavitation level plot (log-log). (Copyright © 1991, Instrumentation, Systems, and Automation Society. From ISA Paper No. 91-0462, "Solving the Problem of Cavitation in Control Valves.")

can be related to the *K*-type indexes by using the relationship:

$$\sigma_1 = \sigma_2 + 1 = \frac{1}{K} \qquad \textbf{6.15(51)}$$

Although K_c can be converted to a σ value for subjective comparison, the meaning of the value does not change. There is no proven correlation between specific points on the flow curve Figure 6.15j and the inflection points on the vibration curve in Figure 6.15k. Various cavitation regimes can be identified for many (but not all) valves with the type of plot shown in Figure 6.15k.[22] The ISA Recommended Practice RP75.23.01 uses σ defined by σ_1 in equation 6.15(49), so where σ is used in this section without a subscript, σ_1 is implied.

The various regimes of cavitation identified in Figure 6.15k are determined in a test defined in ISA RP75.23.01, which begins at a maximum attainable pressure drop in the test valve. The pressure drop is then reduced in small steps while maintaining a constant upstream pressure and temperature; vibration, pressures, and flow rates are measured at each step until the noise, vibration, and ΔP indicate that a noncavitating flow regime has been attained. The results are plotted showing vibration as a function of σ.

The *noncavitating regime* in Figure 6.15k is distinguished by turbulent flow noise and mild vibration. The *incipient cavitation regime* begins at a level called incipient cavitation and progresses to another level called *constant*[22] (or *critical*[18]) *cavitation.* Incipient cavitation begins with intermittent cavity formation and collapse. The resulting noise may be scarcely audible above the background noise. Constant cavitation produces a steady process of cavity growth and collapse, so that the sound is no longer intermittent but is constant.

Cavitation in the incipient regime is not generally harmful to valve hardware,[18] and limiting valve operation below a level of constant cavitation is a conservative application approach—although sometimes overly conservative. In the *full cavitation regime*, vibration or sound intensity increases to the maximum vibration level. When damage testing can be done, the onset of cavitation damage, called *incipient damage*, is often found somewhere in this regime.[22]

In many cases, the value of $1/K_c$ plotted on this graph will also appear in this regime. Thus, K_c is sometimes referred to as "incipient choking." Noise and vibration in the *super-cavitation regime*[18] are less than at maximum vibration as a result of the way that large volumes of vapor in the flow affect the fluid density, compressibility, and vibration transmission properties. Super-cavitation may exhibit a vapor pocket extending from the valve downstream into the pipe, where it will collapse back into liquid as the static pressure recovers. This differs from true flashing, in which the downstream pressure remains less than the vapor pressure of the flowing fluid. Damage to downstream piping and fittings is common in this regime.

The cavitation index σ describing the service conditions is useful if it can be compared with a valve cavitation index that represents the resistance of the particular valve to potential cavitation. The coefficient σ_{mr} ("manufacturer's recommended σ") is defined in ISA RP75.23.01 to designate the limit for σ above which the valve may be operated.

If the operating pressures for the valve are very close to the pressures used to test the valve, a direct comparison could be made, such that $\sigma > \sigma_{\mathrm{mr}}$ indicates safe operating conditions. However, research has shown that valve σ coefficients for incipient, constant, and incipient-damage cavitation do not remain constant for all pressures or in different sizes of geometrically similar valves. Scaling equations have been developed that take these effects into consideration, and they have been adopted in the ISA recommended practice.

Pressure Scale Effect (PSE) The scaling factors PSE and size scale effect (SSE) are used to adjust the recommended valve coefficient, σ_{mr}, when the reference size and reference pressure are known for which σ_{mr} was determined. Equation 6.15(52) defines a scaled coefficient for the valve, σ_v.

$$\sigma_v = [\sigma_{\mathrm{mr}}(\mathrm{SSE}) - 1](\mathrm{PSE}) + 1 \qquad \textbf{6.15(52)}$$

σ_v is then compared to the service σ calculated by Equation 6.15(49). The service $\sigma \geq \sigma_v$ means that the valve will operate at a level of cavitation less severe than the level corresponding to the manufacturer's σ_{mr}.

The pressure scale effect, PSE, is

$$\mathrm{PSE} = \left[\frac{P_1 - P_v}{(P_1 - P_v)_R}\right]^a \qquad \textbf{6.15(53)}$$

The subscript *R* refers to reference pressures used in testing the valve. Typically, valves are tested at about $(P_1 - P_v)_R = 100$ psi; this value may be used as an estimate unless otherwise stated by the valve manufacturer. The exponent, *a*, is a constant determined for a specific valve. Table 6.15l shows representative values of *a* for different valve types and different levels of cavitation.[18,22] When the exponent *a* is 0, PSE reduces to a value of 1, indicating that there is no pressure scale effect for those conditions.

Size Scale Effect The size scale effect, SSE, can be estimated from Equation 6.15(54). The exponent *b* was derived from limited testing.[18,20] It should be noted that testing under choked conditions showed no size scale effect; i.e., SSE for choked flow has a value of 1.

$$\mathrm{SSE} = \left(\frac{d}{d_R}\right)^b \qquad \textbf{6.15(54)}$$

$$b = 0.068\left(\frac{C_v}{d^2}\right)^{1/4} \qquad \textbf{6.15(55)}$$

where

d is the inside diameter of the valve port in inches

d_R refers to the reference (tested) valve port diameter in inches

TABLE 6.15I
*Exponent "a" for Calculating Pressure Scale Effect (PSE)**

Valve Type	Cavitation Level	Exponent a
Quarter-turn valves (i.e., ball, butterfly)	Incipient	0.22–0.30
	Constant	0.22–0.30
	Incipient damage	0.10–0.18
	Choking	0
Segmented ball and eccentric plug	Incipient	0.30–0.40
	Constant	0.30–0.40
	Incipient damage	N/A
	Choking	0
Single-stage globe	Incipient	0.12–0.16
	Constant	0.12–0.16
	Incipient damage	0.10–0.14
	Choking	0
Multistage globe	Incipient	0.00–0.10
	Constant	0.00–0.10
	Incipient damage	N/A
	Choking	0
Orifice	Incipient	0
	Constant	0
	Incipient damage	0.20
	Choking	0

*Reprinted by permission. Copyright © 1991, Instrumentation, Systems, and Automation Society. From ISA Paper No. 91-0462, "Solving the Problem of Cavitation in Control Valves." Single-stage globe data modified per Tullis, Reference 18.

Size scale equations are valid only when the valves are the same style and have approximately the same relative capacity; i.e., C_v/d^2 must be about the same for both valves. There are additional considerations if the valve is larger or smaller than the pipe size with attached fittings (reducers, increasing). RP75.23.01 recommends additional calculations when $C_v/d^2 > 20$. The following example demonstrates how this method can be applied.

Example 2. Rotary Valve Service data:

Fluid: Water
$T = 74°\text{F}$
Line size = 10 in. Sch. 40
$P_1 = 82$ psia
$Q = 3500$ gpm
$G_f = 0.998$
$P_v = 0.41$ psia
$P_2 = 70$ psia

Results of valve sizing: $C_v = 1010$; 8 in., ANSI Class 150 throttling rotary disk valve; body outlet velocity = 22 fps; approximately 75% open with $C_v/d^2 = 1010/(8)^2 = 15.8$, so effects of fittings may be ignored.

Calculate σ using Equation 6.15(49).

$$\sigma = \left(\frac{82 - 0.41}{82 - 70}\right) = 6.8$$

Compare with manufacturer's cavitation data for the butterfly valve, which are given as:

$\sigma_{\text{mr}} = 5.1$
$(P_1 - P_v)_R = 100$ psi, $a = 0.12$
$d_R = 6$ in.

Calculate PSE, b, SSE from Equations 6.15(53–55).

$\text{PSE} = [(82 - 0.41)/100]^{0.12} = 0.976$
$b = 0.068(15.8)^{1/4} = 0.136$
$\text{SSE} = (8/6)^{0.136} = 1.040$

Calculate σ_v from Equation 6.15(52).

$$\sigma_v = [5.1(1.040) - 1](0.976) + 1 = 5.2$$

Evaluation: Since σ is 6.8 and is greater than σ_v of 5.2, the valve can operate without significant cavitation.

Rule of Thumb for Incipient Cavitation One definition of the state of incipient cavitation is when the pressure drop increases to the point where it is first audible that some cavitation bubbles burst. This condition is usually not a cause for concern, although the pressure ratio factor X_{Fzi} is important in noise calculations.[16]

$$X_{\text{Fzi}} = 0.9(P_{1\,\text{ref}}/P_{1\text{actual}})^{0.125}/[1 + 3F_d(C_v/1.17F_L)^{0.5}]^{0.5} \quad \textbf{6.15(56)}$$

In Equation 6.15(38), if $P_{1\text{actual}}$ is in psia, the value of $P_{1\,\text{ref}}$ is 87 psia. If $P_{1\text{actual}}$ is in bars, $P_{1\text{actual}}$ is 6 bars absolute. Using this pressure ration factor X_{Fzi}, the pressure drop causing incipient cavitation can be calculated using Equation 6.15(39):

$$\Delta P_{\text{inc}} = X_{\text{Fzi}}\,(P_1 - P_v) \quad \textbf{6.15(57)}$$

As a rule of thumb that can be used as a guide only in an attempt to avoid damage or noise, one should make sure that the valve's pressure drop does not exceed the following limit:

$$\Delta P_{\text{limit}} = 0.5(X_{\text{Fzi}} + F_L^{0.5})(P_1 - P_v) \quad \textbf{6.15(58)}$$

For more accurate cavitation calculations one should use Equations 6.15(47–55), while for accurate noise calculations, should consult Section 6.14.

Example 3 Assume that a 2 in. globe valve was selected for a particular flow condition where the inlet pressure was 195 psia and the outlet pressure 87 psia. Let us also assume that using Equation 6.15(25), the required C_v was calculated to be 38, and that the supplier's catalog for the particular valve showed an F_L of 0.87 and an F_d of 0.4 at the given valve travel. Finally, assume that the vapor pressure of the flowing fluid is 8.5 psia.

Using Equation 6.15(56), we can calculate the pressure differential ratio for incipient cavitation as

$$X_{\text{Fzi}} = 0.9\ (87/195)^{0.125}/[1 + 3 \times 0.4(38/1.17 \times 0.87)^{0.5}]^{0.5}$$
$$= 0.282$$

Using this X_{Fzi} value, the limiting pressure drop to avoid cavitation damage or noise, using Equation 6.15(58), is

$$\Delta P_{\text{limit}} = 0.5(0.282 + 0.87^2)(195 - 8.5) = 96.8 \text{ psi.}$$

Because the actual pressure drop is only 80 psi, we should not be concerned.

Flashing

Flashing begins in the same manner as cavitation. However, in case of flashing, the pressure downstream of the cavity growth region remains at or below vapor pressure of the process fluid. This causes the vapor to stay in the vapor state as it leaves the valve and enters the downstream piping. The specific volume increases as liquid changes to vapor, which in turn causes an increase in the fluid velocity. If enough vapor is formed, the resulting high valve outlet, or pipe, the resulting velocities can erode metals.

In many cases, flashing is a normal part of the process; it cannot be avoided, and special system and valve designs are required to accommodate it. The heat of vaporization comes from the process liquid, causing its temperature to decrease. The relative masses of liquid and vapor will thereby approach thermodynamic equilibrium. The amount of flashing can be calculated from an energy balance. Even small amounts of flashing (e.g., where the vapor equals 1–3% by weight) can significantly affect a valve's capacity, sizing, and selection; therefore, flashing should be stated in the valve specification data sheets.

Large amounts of flashing (e.g., 10–15% by weight) require special valve designs, such as oversized outlets, and a larger downstream pipe. Note, choking can occur in the downstream pipe due to the large volume. This in turn will increase the back-pressure, P_2, which in turn then causes the valve to undergo cavitation again.

When the valve outlet pressure, P_2, is less than or equal to the vapor pressure of the process liquid, some of the liquid "flashes" into vapor and stays in the vapor phase as it enters the downstream piping. The phase change from liquid to vapor may cause high velocities and erosion of metals at the outlet. Even small amounts of flashing can significantly affect valve sizing and selection. Large amounts of flashing (e.g., where the vapor equals 10–15% by weight) require special valve designs and materials.

In order to select the right valve, it is necessary to know the fraction of the liquid that will flash to vapor and the flowing velocity of the resulting vaporized mixture. If the vapor content and therefore the velocity is high, the resulting back-pressure at the valve outlet can cause "choked flow" in the downstream pipe. This can cause cavitation in the valve and can cause severe erosion of the valve trim.

Calculating the Flash Fraction (X) The method for calculating the vapor fraction is essentially the same for all liquids. Water is the most common liquid that is likely to flash, and data for water are readily available in the steam tables. Data for other fluids may be found in other references.[11]

The vapor fraction by weight can be calculated from Equation 6.15(59).

$$X = (h_{f1} - h_{f2})/h_{fg2} \qquad \textbf{6.15(59)}$$

where h_{f1}, h_{f2}, and h_{fg2} are the enthalpy of saturated liquid upstream, the enthalpy of saturated liquid downstream, and the enthalpy of evaporation at downstream pressure, respectively.

Calculating Velocity Velocity calculations of the liquid-vapor mixture downstream of the valve assume that the mixture is in thermodynamic equilibrium. In most cases, this assumption is a good enough approximation when compared to the accuracy of other valve sizing factors. The velocity, U, for the mixture is

$$U = (w/A)[(1 - X)v_{f2} + Xv_{g2}] \qquad \textbf{6.15(60)}$$

where all units are consistent and v_{f2} and v_{g2} are specific volumes of the saturated liquid and saturated vapor, respectively, at downstream conditions. If mass flow rate, w, is given in lb/hr, area A in in.^{2}, and specific volumes in ft^3/lb, Equation 6.15(60) for the calculation of velocity, U, in fps becomes

$$U = (0.040\ w/A)[(1 - X)v_{f2} + Xv_{g2}] \qquad \textbf{6.15(61)}$$

Similarly, if flow rate is given as q in gpm, then U in fps can be calculated by

$$U = (20.0\ qG_f/A)[(1 - X)v_{f2} + Xv_{g2}] \qquad \textbf{6.15(62)}$$

It has been the experience that velocities exceeding approximately 300–500 ft/s have caused erosion in downstream pipe fittings in flashing water applications. Therefore, flashing water applications often require angle body valves and piping made out of chromium-molybdenum-steel alloys to resist erosion from the flashing water that is moving at high velocities.

Sizing Example 4 A flashing application must always be sized on the basis of choked flow conditions. Piping downstream of a flashing valve should generally be larger than the valve size because it must accommodate the large increase in volume. Therefore, the corresponding combined pressure recovery and piping geometry factor, F_{LP}, must also be applied to the sizing equations. See Equation 6.15(23) and Figure 6.15g.

Using the same configuration and conditions as in Example 1, except that P_2 in this case is 20 psia, determine the valve size and required C_v. (The other conditions in Example 1 were: process fluid was water at 250°F, $P_v = 30$ psia, $G_f = 0.94$, $P_1 = 314.7$ psia, and $q = 500$ gpm.)

Step 1: Calculate actual Δp.

$$\Delta p = 314.7 - 20 = 294.7 \text{ psi}$$

Step 2: Calculate ΔP_{ch} and determine ΔP_a. From Example 1, $F_F = 0.93$. Assuming that an angle valve with a characterized cage and flow over the seat (flow to close) is the preliminary selection for this flashing application, from Table 6.15c one can obtain the values $F_L = 0.80$, $F_d = 1.0$, and $C_v/d^2 = 12$. From Equation 6.15(11)

$$\Delta p_{\text{ch}} = F_L^2 (P_1 - F_F P_v) = 0.8^2 [314.7 - (0.93)(30)] = 183.6 \text{ psi}$$

$$\Delta p_{\text{ch}} < \Delta p, \text{ therefore } \Delta p_a = \Delta p_{\text{ch}} = 183.6 \text{ psi}$$

Step 3: Determine specific gravity of liquid at upstream conditions. This was determined in Example 1; $G_f = 0.94$.

Step 4: Calculate the approximate required C_v assuming F_P and $F_R = 1$. From Equation 6.15(34)

$$C_v = \frac{q}{F_R F_P} \sqrt{\frac{G_f}{\Delta p_a}} = 500(0.94/183.6)^{1/2} = 35.8$$

Step 5: Select the approximate body size. Refer to manufacturers' catalogs or, by using estimates from Table 6.15c, $C_v/d^2 = 12$, in which case the minimum body size is

$$d \approx (35.8/12)^{1/2} \approx 1.73$$

Thus, a 2-in. body size is selected.

Step 6: Determine F_{LP}; $d = 2.0$ in.; $D_1 = D_2 = 4.0$ in.; $d/D_1 = 0.5$.

From Equations 6.15(12–22) and Example 1: $K_1 = 0.2813$; $K_{B1} = 0.9375$; $K_i = 0.2813 + 0.9375 = 1.2188$; $C_d = 35.8/(2)^2 = 8.95$.

Using Equation 6.15(23):

$$F_{LP} = 1/F_L^2 + K_i \, Cd^2/890)^{-1/2}$$

$$= [0.8^{-2} + (1.2188)(89.5)^2/890]^{-1/2} = 0.77$$

Step 7. Calculate the final required C_v from Equation 6.15(30).

$$C_v = (q_{\max}/F_{LP}) \sqrt{G_f/(P_1 - F_F P_v)}$$

$$= 500/0.77 \sqrt{0.94/(314.7 - (0.93)(30))} = 37.2$$

Therefore, the 2-in. valve size is adequate and the capacity required for the application is within the estimated limit.

Example 5 Given the flashing conditions described in Example 4, determine the mass fraction of vapor at the valve outlet.

Step 1: Determine the enthalpies of saturated liquid at upstream conditions and of saturated vapor and liquid at downstream conditions using the steam tables. From saturated steam temperature tables at 250°F for upstream conditions: $h_{f1} = 218.59$ BTU/lbm.

From saturated steam pressure tables at 20 psia for downstream conditions: $h_{f2} = 196.27$ BTU/lbm; $h_{g2} = 1156.3$ BTU/lbm; $h_{hfg2} = 960.1$ BTU/lbm.

Step 2: Using the assumption that the process is adiabatic and therefore the enthalpies immediately upstream and downstream of the valve are equal, Equation 6.15(59) can be used to calculate the vapor fraction, X, or steam "quality" downstream:

$$X = (h_{f1} - h_{f2})/h_{fg2} = (218.59 - 196.27)/960.1 = 0.023$$

Therefore, the amount of fluid mass flashing to steam is 2.3%.

Example 6 Given the flashing case in Examples 4 and 5, estimate the velocity of the two-phase mixture at the valve outlet.

Step1: From the steam tables, find the specific volumes of saturated liquid and vapor at downstream conditions.

$$v_{f12} = 0.016834 \text{ ft}^3\text{/lbm}; \; v_{g2} = 20.087 \text{ ft}^3\text{/lbm}$$

Step 2: Determine vapor fraction or steam quality at downstream conditions. For this example, the vapor fraction, X, was found in Example 5. $X = 0.023$.

Step 3: Find specific gravity (or density) of the mixture and the cross-sectional area at the valve outlet. From Example 4, $G_f = 0.94$. Flow area is the outlet area of a 2-in. ANSI Class 300 valve:

$$A = \pi (2.00)^2/4 = 3.14 \text{ in.}^2$$

Step 4: Calculate velocity using one of Equations 6.15(60–62). Selecting Equation 6.15(62), because the flow rate is in units of gpm and area in square inches,

$$U = [20 \, qG_f/A](1 - X) v_{f2} + X v_{g2}]$$

$$= [(20)(500)(0.94)/3.14][(1 - 0.023)(0.016834)$$

$$+ (0.023)(20.087)] = 1432 \text{ ft/s}$$

Evaluation: This is not a realistic velocity since the actual velocity cannot exceed the speed of sound (about 1300 ft/s). As a result of such high velocity, the pressure at the valve outlet will rise, which in turn will reduce the vapor content until an equilibrium is established at sonic velocity. The result will be choked flow in the pipe.

At that point, the back-pressure in the pipe can cause cavitation in the valve and in the pipe. Wet steam at 1300 ft/s can be erosive. An angle valve with hardened trim exhausting it into either a long, straight run of pipe or directly into a receiver vessel should be used in order to avoid erosion of downstream pipe and fittings.

Another alternative would be to use an expanded outlet with a venturi outlet on the valve to reduce the velocity before it exits into the pipe. A venturi outlet with the same exit area as the 4-in. downstream pipe would reduce velocity to approximately 360 ft/s. These velocities are estimates only because the assumption of thermodynamic equilibrium was used to determine fluid properties. Note: The longer the downstream pipe and the more elbows there are in your piping system, the more the valve's back-pressure will increase, and with it cavitation will also rise.

Laminar or Viscous Flow

The proportionality between flow rate and the square root of pressure drop holds true only for fully turbulent Newtonian fluid behavior. Non-Newtonian fluids include most polymers and many other fluids. If no experimental sizing data exist, one might specify a line size valve and purchase a number of reduced C_v trims for it, so that the final choice of trim would be determined by trial and error.

Laminar or transitional flow regimes may result from very low flows, small valves or passages, low pressure differentials, and high viscosity. The valve Reynolds number, Re_v, is calculated to determine the effect of laminar or transitional flow behavior on the required valve C_v.

Reynolds Number Factor (F_R) In the majority of control valve applications the flow velocities inside the valve are high, which in turn causes the Reynolds number to be high, and therefore turbulent conditions exist. Under turbulent conditions pressure drop relates to the square of flow.

On the other hand, when the viscosity of the process fluid is high or when the size of the valve and the flow velocity through it are low, the Reynolds number will also be low and laminar flow conditions will exist. Under laminar conditions pressure drop linearly relates to flow. Therefore, the same increase in pressure drop results in a larger increase in flow than it would under turbulent conditions.

Here the flow coefficient under laminar or transitional flow becomes

$$C_v = q/F_R\ (\Delta P/G_f)^{0.5} \qquad \textbf{6.15(63)}$$

The valve equivalent Reynolds number, Re_v, is defined as

$$\text{Re}_v = \frac{N_4 F_d q}{\nu F_L^{1/2} C_v^{1/2}} \left(\frac{(F_L C_d)^2}{N_2} + 1 \right)^{1/4} \qquad \textbf{6.15(64)}$$

(See Table 6.15i for the values of constants N_2 and N_4 for various units.)

$$F_R = 0.7 \times 77\ (\text{Re}_v)^{0.5}/F_L\ (C_v/d^2) \qquad \textbf{6.15(65)}$$

When the valve operates in the laminar regime and the diameter (d) is in inches,

$$F_R = 1 + (0.06\ (C_v/d^2)^{0.5} F_L^{0.5}) \log\ (\text{Re}_v/10{,}000) \qquad \textbf{6.15(66)}$$

Note: When operating in the transitional regime, use the smaller of the two F_R values calculated by using Equations 6.15(65) and 6.15(66).

If the diameter (d) is given in millimeters, the F_R numbers are determined by using Equations 6.15(67) and 6.15(68). For the laminar regime use Equation 6.15(67):

$$F_R = 0.0076(\text{Re}_v)^{0.5}/F_L\ (C_v/d^2) \qquad \textbf{6.15(67)}$$

For the transitional regime use Equation 6.15(68):

$$F_R = 1 + (1.53(C_v/d^2)^{0.5}\ F_L^{0.5}) \log\ (\text{Re}_v/10{,}000) \qquad \textbf{6.15(68)}$$

The curves provided in Figure 6.15m relate the valve Reynolds number (Re_v) to the Reynolds number factor (F_R).

Example 7 Assuming that in a particular application $C_v = 10$, $d = 25$ mm, $F_L = 0.9$, $\text{Re}_v = 500$, the F_R in the transitional

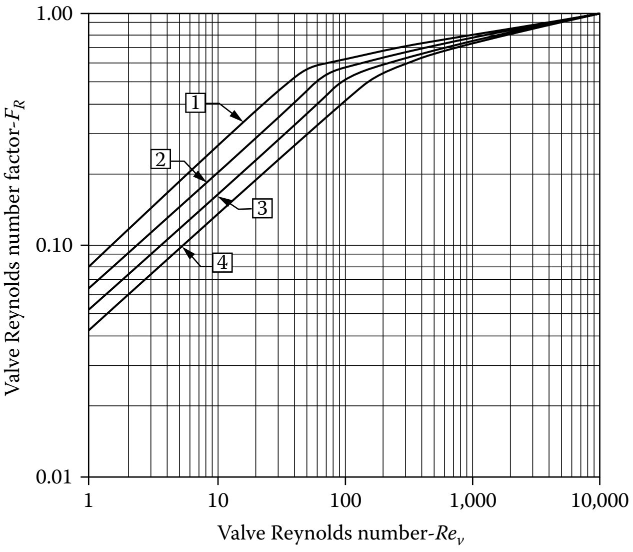


FIG. 6.15m

The valve Reynolds number (Re_v) can be corrected as a function of the Reynolds number factor (F_R) and of the type of valve used (if the diameter d is in inches).[8]

Curve No. 1 applicable for globe style valves $C_d \leq 10$.

Curve No. 2 applicable for globe valves and eccentric rotary plug valves, $C_d = 10$–15.

Curve No. 3 applicable for butterfly valves, $C_d = 15$–25.

Curve No. 4, applicable for ball valves $C_d \geq 25$.

regime is calculated as follows:

$$1 + (1.53(10/625)^{0.5}\ 0.9^{0.5})\ \log\ (500/10{,}000)$$
$$= 1 + 0.193 \times -1.3 = 0.748$$

In Figure 6.15m the straight diagonal lines extending downward from right to left, starting at an F_R value of approximately 0.5, are in the laminar region and operate under the conditions where laminar flow exists. At a valve Reynolds number above 10,000, all three curves in Figure 6.15m reach an F_R value of 1.0. At this number and at all higher values of Re_v, fully turbulent flow conditions exist. Between the laminar region (straight diagonal lines in Figure 6.15m), and the turbulent region (over $\mathrm{Re}_v = 10{,}000$, where $F_R = 1.0$), the flow regime is transitional (i.e., neither laminar nor turbulent).

Equation 6.15(69) can be used for determining the valve Reynolds number Re_v for globe valves and small flow valves. In these cases the Bernoulli correction for reducers (the bracketed term in Equation 6.15[64]) is not required.

$$\mathrm{Re}_v = N_4 F_d\ q/F_L^{0.5} C_v^{0.5} \qquad \mathbf{6.15(69)}$$

The valve-style modifier in Equation 6.15(69) is designated F_d. Baumann[17] defined F_d as the ratio of the equivalent hydraulic diameter(s) of a valve flow orifice to that of a circular one. For representative values please consult Table 6.15c.

Valve Sizing for Viscous (Laminar) Services The following treatment is applicable to valves with or without attached fittings. The F_p factor (Table 6.15c) is assumed to be 1.0, because of the absence of turbulence at the valve outlet. The first step in the valve sizing procedure is to calculate a pseudo valve flow coefficient C_{vt}, assuming turbulent flow and using Equation 6.15(70):

$$C_{vt} = \frac{q}{N_1 \sqrt{\frac{P_1 - P_2}{G_f}}} \qquad \mathbf{6.15(70)}$$

One may calculate the valve's Reynolds number by using either Equation 6.15(64) or 6.15(69) and then determine F_R from Figure 6.15m. The required C_v for the laminar or transitional flow conditions is then found by $C_{vL} = C_{vt}/F_R$.

The problem with this method is that it requires iteration, because one first has to guess the laminar valve C_v in order to calculate the Reynolds number, and later one has to repeat this procedure if the actual C_{vL} is too far off from the estimate.

In the sizing approach below *a simplified method* is presented for two typical valve classes. This approach, while ignoring transitional flow (a minor error), saves the time associated with the above-mentioned iteration.

Globe Valve (Example 8) For the calculations below, the values of $F_L = 0.9$ and $F_d = 0.46$ are assumed. In this simplified approach one would first calculate a pseudo Re_{vi} by modifying Equation 6.15(69) by also using Equation 6.15(70) for C_{vt}, in order to arrive at Equation 6.15(71) below:

$$\mathrm{Re}_{vi} = N4q/0.95 \nu C_{vt}^{0.5} \qquad \mathbf{6.15(71)}$$

Next, the required C_v is calculated under laminar flow conditions:

$$C_{vL} = C_{vt}/0.022\ (\mathrm{Re}_{vi})^{0.666} \qquad \mathbf{6.15(72)}$$

Assuming that $q = 4.13$ gpm, $\nu = 31{,}729$ cSt, $\Delta P = 8.41$ psi, and $G_f = 0.977$, one can, using Equation 6.15(70), calculate the C_{vt} of 1.407. Using that value in Equation 6.15(71), the Reynolds number comes out as 2.

Based on Equation 6.15(72), the C_v under laminar flow conditions is calculated as:

$$C_{vL} = 1.407/0.022\ (2)^{0.666} = 40.3$$

This C_v requirement can be met with a 2-in. globe-style control valve.

Ball or Butterfly Valves (Example 9) For the calculations below, the values of $F_L = 0.6$ and $F_d = 0.5$ are assumed. Here the required C_v under laminar flow conditions is

$$C_{vL} = C_{vt}/0.04\ (\mathrm{Re}_{vi})^{0.666} \qquad \mathbf{6.15(73)}$$

Using the same data as in Example 8 above, but using Equation 6.15(73), we get

$$C_{vL} = 1.407/0.04(2)^{0.666} = 22.$$

For this application, one could select a 1.5-in. ball valve instead of the 2-in. globe valve and thus obtain some cost saving. Note: If Re_v is greater than 10,000, the flow may be taken as turbulent and it can therefore be assumed that $F_R = 1.0$. In this case $C_{vL} = C_{vt}$.

For butterfly valves refer to Section 6.17.

Small Flow Valves (Example 10) The sizing of small flow valves is more complicated, because their F_d factor varies widely depending on the type of their trim. Laminar flow is likely to occur in small valves on liquid service, but at C_v values under 0.1, it is also possible to have laminar or transitional flow with gases.

The F_d values drastically change with valve type. For V-notch-type plugs $F_d = 0.7$, for flat seated trim $F_d = 0.3$, and for tapered needle trim $F_d = 0.09 \times (C_{vL}/d_o)^{0.5}$, where d_o = orifice diameter in inches.

A simplified sizing equation for *liquid* service in small valves is given below:

$$C_{vL} = 0.192(N_{11} G_f \nu q/\Delta P F_d)^{0.666} \qquad \mathbf{6.15(74)}$$

Assuming the following process conditions and the use of a flat seated trim: $q = 0.0035$ gpm, $G_f = 0.87$, $\Delta P = 50$ psi, $\nu = 16$ cSt, $F_d = 0.3$ (flat seat). Having this information, one would first calculate the turbulent C_v using $C_{vT} = q\ (G_f/\Delta P)^{0.5}$, which results in $C_{vT} = 0.00046$.

Next C_{vL} is calculated from 6.15(74) as $C_{vL} = 0.192\ (1 \times 0.87 \times 16 \times 0.0035/50 \times 0.3)^{0.666} = 0.004$.

Note that this C_{vL} value is substantially higher than the turbulent C_v of 0.00046, hence it indicates that laminar flow will occur and that a valve trim with a rated C_v that is above 0.004 has to be selected.

For small flow valves having *tapered needle trims* the sizing is even simpler. Here,

$$C_{vL} = 0.973(N_{12}\ qG_f \nu d_o/\Delta P)^{0.5} \qquad \textbf{6.15(75)}$$

Using the previous example but choosing a needle trim with an orifice diameter of $d_o = 0.055$ in., C_{vL} is calculated to be 0.0071 (a substantial difference from the 0.004 with the flat seat trim).

Sizing for Laminar Gas Flow In case of gases one should use *absolute viscosity* μ, because, in contrast with liquids, the kinematic viscosity varies with the absolute pressure of the gas. Sizing for the most common trim, the tapered needle, can be done as follows:

One should first calculate the turbulent C_v from the appropriate equation from the orientation table, Table 6.15a. Next, the laminar valve coefficient for gas service ($C_{vL\text{gas}}$) is to be determined by using the following equation, where d_o, valve orifice diameter in inches, has to be obtained from the manufacturer:

$$C_{vL\text{gas}} = Cv_T/0.0127(N_g q G_s/\mu d_o)^{0.5} \qquad \textbf{6.15(76)}$$

where

μ = absolute viscosity, centipoises

N_{11} = a constant for liquids = 1 if q = gpm, P = psi for misc. trim 30.7 if q = m^3/h, p = kPa

N_{12} = a constant for liquids = 1 if q = gpm, P = psi, and for tapered trim 1204 if q = m^3/h, P = kPa, and d_o = *m*, d_o = inches

N_g = a constant for gas = 1 if q = gpm, and d_o = inches for tapered trim 0.9 if q = m^3/h, and d_o = *m*

Example 11 Assuming the following process conditions: q = 1.6 scfh and the gas has a specific gravity, G_s, of 1.34. The absolute viscosity μ is 0.0215 cP, P_1 = 190 psia, P_2 = 170 psia. Based on these conditions, the calculated C_{vt} is found to be 0.0005.

If one learns from the manufacturer that the orifice diameter d_o = 0.197 in., one can proceed to Equation 6.15(76) and calculate the required laminar C_{vL} as:

$$\begin{aligned} C_{vL} &= 0.0005/0.0127(1 \times 1.6 \times 1.34/0.0215 \times 0.197)^{0.5} \\ &= 0.0017 \end{aligned}$$

This is 3.5 times the turbulent C_v requirement! Note: One should always calculate C_{vt} and C_{vL} for small valves, especially when C_v is below 0.01. After having made both calculations, one should select the larger of the two values.

GAS AND VAPOR SIZING*

The flow rate of a compressible fluid varies as a function of the ratio of the pressure differential to the absolute inlet pressure ($\Delta p/P_1$), designated by the symbol "x." At values of x near 0, the equations in this section can be traced to the basic Bernoulli equation for Newtonian incompressible fluids. However, increasing values of x result in expansion and compressibility effects that require the use of appropriate correction factors.

Equations for Turbulent Flow

The flow rate of a gas or vapor through a valve or the required C_v may be calculated by using any of the following equivalent forms of the equations, which differ only in the units used. If the flow is in mass units and the density is in specific weight units,

$$\left.\begin{aligned} & w = N_6 F_P C_v Y \sqrt{x P_1 \gamma_1} \\ \text{or} \quad & \\ & C_v = \frac{w}{N_6 F_P Y \sqrt{x P_1 \gamma_1}} \end{aligned}\right\} \qquad \textbf{6.15(77)}$$

For volumetric flow and specific gravity units:

$$\left.\begin{aligned} & Q = N_7 F_P C_v Y \sqrt{\frac{x}{G_g T_1 Z}} \\ \text{or} \quad & \\ & C_v = \frac{Q}{N_7 F_P P_1 Y} \sqrt{\frac{G_g T_1 Z}{x}} \end{aligned}\right\} \qquad \textbf{6.15(78)}$$

For mass flow and molecular weight units:

$$\left.\begin{aligned} & Q = N_8 F_P C_v P_1 Y \sqrt{\frac{xM}{T_1 Z}} \\ \text{or} \quad & \\ & C_v = \frac{w}{N_8 F_P P_1 Y} \sqrt{\frac{T_1 Z}{xM}} \end{aligned}\right\} \qquad \textbf{6.15(79)}$$

For volumetric flow and molecular weight units:

$$\left.\begin{aligned} & Q = N_9 F_P C_v P_1 Y \sqrt{\frac{x}{M T_1 Z}} \\ \text{or} \quad & \\ & C_v = \frac{q}{N_9 F_P P_1 Y} \sqrt{\frac{M T_1 Z}{x}} \end{aligned}\right\} \qquad \textbf{6.15(80)}$$

Note that the numerical value of x used in these equations must not exceed the choking limit ($F_k x_{TP}$), regardless of the actual value of x.

---

* Reprinted by permission. Copyright © 1985, Instrumentation, Systems, and Automation Society. From ANSI/ISA-S75.01, "Flow Equations for Sizing Control Valves."

TABLE 6.15n

*Numerical Constants to be Used in Gas and Vapor Flow Equations, which Reflect the Units Being Used**

Constant		Units Used in Equations					
	N	W	Q**	P, ΔP	γ_l	T_l	d, D
N_5	0.00241	—	—	—	—	—	mm
	1000	—	—	—	—	—	in.
N_6	2.73	kg/h	—	kPa	kg/m^3	—	—
	27.3	kg/h	—	Bar	kg/m^3	—	—
	63.3	lb/h	—	psia	lb/ft^3	—	—
N_7	4.17	—	m^3/h	kPa	—	K	—
	417	—	m^3/h	Bar	—	K	—
	1360	—	scfh	psia	—	°R	—
N_8	0.948	kg/h	—	kPa	—	K	—
	94.8	kg/h	—	Bar	—	K	—
	19.3	lb/h	—	psia	—	°R	—
N_9	22.5	—	m^3/h	kPa	—	K	—
	2250	—	m^3/h	Bar	—	K	—
	7320	—	scfh	psia	—	°R	—

*Reprinted by permission. Copyright © 1985, Instrumentation, Systems, and Automation Society. From ANSI/ISA-S75.01, "Flow Equations for Sizing Control Valves."

**Q is in cubic feet per hour measured at 14.73 psia and 60°F, or cubic meters per hour measured at 101.3 kPa and 15.6°C.

Constants for Engineering Units

The numerical constants, N, are chosen to suit the engineering units used in the equations. Values for N are listed in Table 6.15n.

Expansion Factor (Y)*

The expansion factor Y accounts for the change in density of a compressible fluid as it passes from the valve inlet to the vena contracta and for the change in area of the vena contracta as the pressure drop is varied (Contraction coefficient). Theoretically, Y is affected by all of the following:

1. Ratio of port area to body inlet area
2. Internal geometry of the valve
3. Pressure drop ratio, $x = \Delta p/P_1$
4. Reynolds number
5. Ratio of specific heats, k

The influence of items 1, 2, and 3 are defined by the factor x_T. Test data indicate that Y may be taken as a linear function of x, as shown in the following equation (and in Figure 6.15n for a valve without attached reducers or fittings):

$$Y = 1 - \frac{x}{3F_k x_T} \quad (\text{limits}: 1.0 \geq Y \geq 0.67) \qquad \textbf{6.15(81)}$$

For a valve with attached fittings, x_{TP} should be substituted for x_T. For the overwhelming majority of applications, Reynolds number effects may be disregarded when the process fluid is compressible except for small flow valves; i.e., if C_v is less than 0.01. The effect of the ratio of specific heats is discussed later.

Choked Flow*

If all inlet conditions are held constant and the differential pressure ratio (x) is increased by lowering the downstream pressure (P_2), the mass flow rate will increase till it reaches a maximum. Flow conditions where the value of x exceeds this limit are described as *choked flow*.

Choking occurs when the jet stream across the cross-sectional area of the vena contracta attains sonic velocity and when the vena contracta ceases to expand. Sonic velocity already occurs when P_1/P_2 exceeds about $0.5 \times F_L^2 \times P_1$, while choked flow occurs at P_1/P_2 ratios, which are higher than the ones causing sonic velocity. Therefore, x_T is always higher than the sonic pressure ratio.

The value of x at the inception of choked flow conditions (x_T) varies with the type of valve (Table 6.15c). It also varies with the piping geometry and with the thermodynamic properties of the flowing fluid. The factors involved are x_T, x_{TP}, and F_k.

Choking affects the use of Equations 6.15(77–81) in the following manner: The value of x used in the equations must not exceed $F_k x_T$ or $F_k x_{TP}$, regardless of the actual value of x. The expansion factor Y at choked flow ($x \geq F_k x_{TP}$) is at its minimum value of $^2/_3$.

---

* Reprinted by permission. Copyright© 1985, Instrumentation, Systems, and Automation Society. From ANSI/ISA-S75.01, "Flow Equations for Sizing Control Valves."

Critical Pressure Drop Ratio Factor (x_T) For maximum accuracy, the pressure drop ratio factor, x_T, must be established by using the test procedures specified in Reference 10. Representative x_T values for valves are tabulated in Table 6.15c. These representative values are not to be taken as actual. Actual values must be obtained from the given valve manufacturer.

When a valve is installed with reducers or other fittings, the critical pressure drop ratio factor of the assembly (x_{TP}) is different from that of the valve alone (x_T). For maximum accuracy, x_{TP} must be determined by test.[10] When estimated values are permissible, the following equation (or Figure 6.15p) may be used to determine x_{TP}:

$$x_{TP} = \frac{x_T}{F_P^2}\left(\frac{x_T K_i C_v^2}{N_5 d^4} + 1\right)^{-1} \qquad \textbf{6.15(82)}$$

In this equation, x_T is the pressure drop ratio factor for a given valve installed without reducers or other fittings, and K_i is the sum of the inlet velocity head coefficients ($K_1 + K_{B1}$) of the reducer or other fittings attached to the valve inlet.

Ratio of Specific Heats Factor (F_k) The ratio of specific heats of a compressible fluid affects the flow rate through a valve. The factor F_k accounts for this effect. F_k has a value of 1.0 for air at moderate temperatures and pressures, where its specific heat ratio is about 1.40. Both theoretical and experimental evidence indicates that for valve sizing purposes, F_k may be taken as having a linear relationship to k. Therefore,

$$F_k = \frac{k}{1.40} \qquad \textbf{6.15(83)}$$

Compressibility Factor (Z) Equations 6.15(78–80) do not contain a term for the actual specific weight of the fluid at upstream conditions. Instead, this term is inferred from the inlet pressure and temperature based on the laws of ideal gases. Under some conditions, real gas behavior can deviate markedly from the ideal. In these cases, the compressibility factor Z shall be introduced to compensate for the discrepancy.

Z is a function of both reduced pressure and reduced temperature. In this section, reduced pressure P_r is defined as the ratio of the actual inlet absolute pressure to the absolute thermodynamic critical pressure for the fluid in question:

$$P_r = P_1 / P_c \qquad \textbf{6.15(84)}$$

The reduced temperature is defined similarly. That is,

$$T_r = T_1 / T_c \qquad \textbf{6.15(85)}$$

Absolute thermodynamic critical pressures and temperatures for most fluids and curves from which the compressibility factor, Z may be determined can be found in many reference handbooks of physical data.[11]

Velocity of Compressible Fluids

The velocity of sound in a compressible fluid is an important parameter. Pressure changes travel through an ideal gas at the velocity of sound. Most valve and piping systems are restricted to fluid velocities less than the speed of sound, due to the nature of the geometric discontinuities (supersonic wind tunnels being a notable exception). In valves, the throttling element and seat produce the vena contracta.

As flow rate and fluid velocity increase, the effects of compressibility cause the relationship of mass flow $(x)^1/_2$ to depart from a straight line. Once sonic velocity is attained in the vena contracta and the pressure ratio of x_T is reached, choking occurs. Reducing the downstream pressure will not increase the flow rate or the velocity at the vena contracta.

However, if local downstream geometry permits (e.g., similar to diverging nozzles), it is possible for velocity to increase downstream of the vena contracta to supersonic levels. Shock waves from supersonic velocities can produce undesirable and even dangerous vibration in process systems; use of supersonic velocities in system or valve applications should be avoided. The ratio of the actual fluid velocity, U, to the sonic velocity, c, is the Mach number, M_N. For valves with standard trim M_N should not exceed 0.3 at the outlet of the valve body. In order to avoid excessive noise. (See Section 6.14 for details.)

$$M_N = U / c \qquad \textbf{6.15(86)}$$

The sonic velocity can be determined from the first law of thermodynamics and the ideal gas law. Gas constants and ratios of specific heat are required in the calculation.

$$c = \sqrt{\frac{g_c k R T_1}{M}} \qquad \textbf{6.15(87)}$$

Flow rate for gases is usually given in terms of mass flow rate, w, or volumetric flow rate, Q. The velocity is

$$U = wv / A = Q / A \qquad \textbf{6.15(88)}$$

From these equations and from appropriate units conversions, the Mach number can be derived as follows:

For gases (using molecular weight, M):

$$M_N = \frac{Q}{5574\,A\sqrt{\frac{kT_a}{M}}} \qquad \textbf{6.15(89)}$$

For gases (using specific gravity, G_g):

$$M_N = \frac{Q}{1035\,A\sqrt{\frac{kT_a}{G_g}}} \qquad \textbf{6.15(90)}$$

For air:

$$M_N = \frac{Q}{1225\,A\sqrt{T_a}} \qquad \textbf{6.15(91)}$$

For steam:

$$M_N = \frac{wv}{1515\,A\sqrt{T_a}} \qquad \mathbf{6.15(92)}$$

where

Q = actual volumetric flow, ft^3/hr (not scfh)
A = flow area of the valve outlet, in.^{2}
T_a = absolute temperature, °R (°F + 460)
w = mass flow rate, lbm/hr
v = specific volume, ft^3/lbm
G_g = specific gravity of the fluid at standard conditions relative to air at standard conditions
M = molecular weight
k = ratio of specific heats

The calculated exit velocity or Mach number should be compared to the limits set by the process or the valve manufacturer for the specific valve and application. Mach numbers exceeding 0.3 can cause significant noise problems.

Sizing for Compressible Fluids (Example 12)

The above working equations can be used in a logical sequence to determine the required valve capacity for a given application. The following example illustrates the basic steps. However, there are other considerations when selecting a valve, such as noise, erosion, corrosion, and actuator sizing, that are discussed in other sections of this chapter and in the references.

For the example below, assume that the process conditions are the following:

Fluid: Steam
Temperature: 450°F
Upstream pressure (P_1): 150 psia
Downstream pressure (P_2): 65 psia
Specific volume downstream (v): 7.98 ft^3/lb
Flow rate (w): 15,000 lb/hr
Flow characteristic: Equal percentage
Line size: 3 in., Class 600

Step 1: Select the appropriate equation based on the available process information and its units. As the information is given in mass flow units, Equation 6.15(79) can be used and N_8 can be selected from Table 6.15n as 19.3. Thus, the working equation becomes

$$C_v = \frac{w}{19.3\,F_P P_1 Y}\sqrt{\frac{T_1 Z}{xM}} \qquad \mathbf{6.15(93)}$$

In order to be able to use this equation, one must obtain other physical data for critical pressure and temperature, for molecular weight, and for the ratio of specific heats. Table 6.15q lists such data for a variety of gases and vapors.

For steam:

Critical pressure (P_c): 3208.2 psia
Critical temperature (T_c): 1165.5°R
Molecular weight (M): 18.02
Ratio of specific heats (k): 1.33

Step 2: Check for choked flow. Determine the critical pressure drop ratio, x_T, from manufacturer's data or from the estimates in Table 6.16c. In order to do that, a preliminary valve-type selection is required. Assuming the use of a single-seated globe valve with contoured plug for this type of application and referring to Table 6.15c, $x_T = 0.72$. The ratio of specific heats factor, F_k, must be calculated from Equation 6.15(83):

$$F_k = k/1.40 = 1.33/1.40 = 0.95$$

Calculate the actual pressure drop ratio, $x = \Delta p/P_1$: $x = (150 - 65)/150 = 0.57$.

Choked flow occurs when x equals or exceeds the value of $F_k x_T$ or $F_k x_{TP}$ (when the valve is installed with reducers or other fittings). The flow is not choked when x is less than $F_k x_T$ or $F_k x_{TP}$,

$$F_k x_T = (0.95)(0.72) = 0.684$$

Note: If the value of x had been greater than $F_k x_T$, then the value of $F_k x_T$ would replace the value of x in the calculation of Y and C_v below.

Step 3: Calculate the expansion factor, Y, from Equation 6.15(81) or obtain it from Figure 6.15o.

$$Y = 1 - x/(3F_k x_T) = 1 - 57/[(3)(0.684)] = 0.72$$

Step 4: Determine the reduced pressure, P_r, and the reduced temperature, T_r using Equations 6.15(84–85). The compressibility factor, Z, is found from generalized compressibility charts available in reference books such as Perry and Chilton's *Chemical Engineer's Handbook.*[11]

$$P_r = P_1/P_c = 150/3208.2 = 0.04$$

$$T_r = T_1/T_c = 910/1165.5 = 0.78$$

$$Z = 1.0$$

Step 5: Calculate a preliminary C_v using the equation identified in Step 1 and assuming a line size valve with $F_P = 1$.

$$C_v = \frac{15{,}000}{19.3(1)(150)(0.72)}\sqrt{\frac{(910)(1)}{(0.57)(18.02)}} = 67.7$$

Step 6: Select the approximate body size based on the calculated preliminary C_v. Manufacturers' tables are the best source for C_v values, but estimates can be made based on C_d($C_d = C_v/d^2$) values from Table 6.15c. Observing that

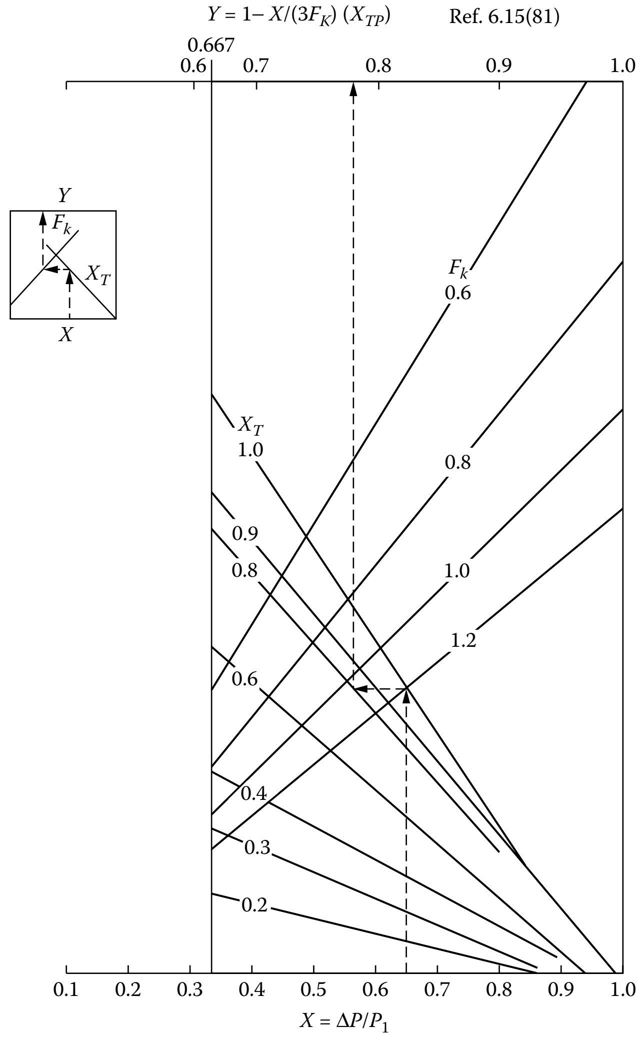


FIG. 6.15o
Graph for determining the expansion factor Y.

a single-ported, plug-type globe valve has a C_d = 11, d = (67.7/11) = 2.48. Thus, a 3-in. valve is selected for further evaluation.

Step 7: Determine the effect of piping geometry. If the valve had been smaller than line size and installed with reducers, F_p and x_{TP} would need to be known or estimated. Where appropriate, F_p can be estimated using Equations 6.15(12–22) or Figure 6.15f. The factor, x_{TP}, can be estimated from Equation 6.15(82) or Figure 6.15p. Y and C_v values may require recalculation using the value of $F_k x_{TP}$, if this product is less than x, to replace x in calculating Y and C_v. In this case, the line size and the valve size are 3 in. NPS, so this step is not required.

Step 8: Calculate the final required C_v using the actual or estimated F_P. In this example, $F_P = 1$, so the required C_v remains 67.7. The body size must be verified if the C_v changes at this step.

Note: The Reynolds number factor, F_R, is generally not considered when sizing for gases and vapors. However, in case of very small valves, with $C_v < 0.01$, F_R should be applied (when it becomes significant) in the same manner as for liquids.

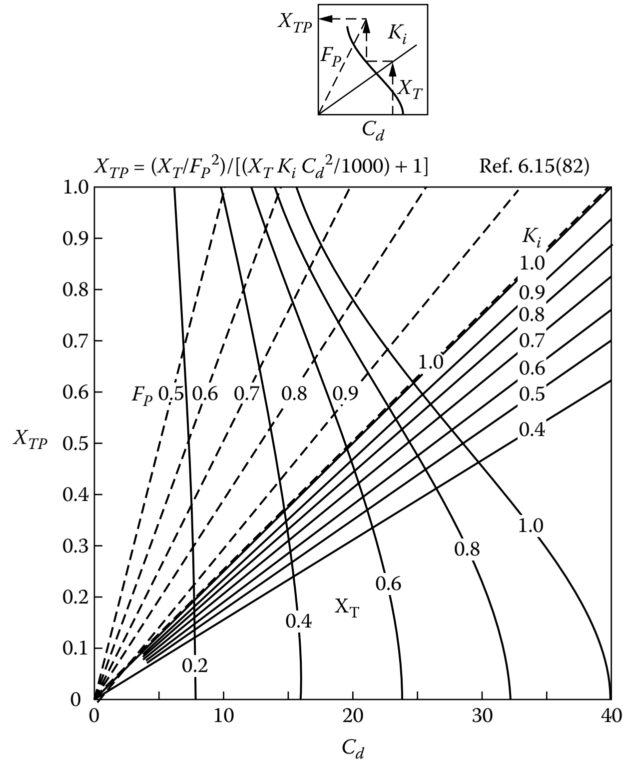


FIG. 6.15p
Graph for determining the critical pressure drop ratio factor for the combined assembly of valve and fittings (x_{TP}).

Step 9: Calculate valve exit velocity to evaluate the likelihood of velocity-related problems in the valve or in the downstream piping. Excessive velocity can result in noise, structural vibration, erosion, equipment damage, and personal injuries. Using Equation 6.15(92), with a flow area A = 7.07 in.^{2}

$$M_N = (15000)(7.98)/[(1515)(910)^{1/2}(7.07)] = 0.37$$

This is a high Mach number for the application; its acceptability should be verified with the manufacturer and with the process system engineer. Valve noise should be predicted and appropriate noise reduction measures taken to safeguard against hazards of excessive noise, as discussed in Section 6.14.

TWO-PHASE FLOW

Valve sizing for two-phase flow is approximate at best, because the effects of thermal, mass, and energy transfer between phases of the mixture are not easily quantified. The methods commonly used by the control valve industry[12,13,14] will be explained here. More elaborate experimental and numerical methods may produce more accurate results for specific cases.

Flow through a control valve is considered isenthalpic (adiabatic) overall and isentropic between the valve inlet and

TABLE 6.15q
Physical Data for a Number of Common Gases and Vapors

Gas	Critical Pressure, P_c (psia)	Critical Temperature, T_c (°R)	Molecular Weight, M	Ratio of Specific Heats, k
Air	492.4	227.1	28.97	1.40
Ammonia	1636.1	729.8	17.0	1.31
Argon	707.0	271.4	39.9	1.67
Butane	551.0	765.2	58.1	1.09
Carbon dioxide	1070.2	547.5	44.0	1.29
Carbon monoxide	507.1	239.2	28.01	1.40
Chlorine	1120.0	751.0	70.91	1.34
Ethane	708.5	549.7	30.0	1.18
Ethylene	730.5	508.3	28.05	1.21
Helium	33.2	9.45	4.00	1.67
Hydrogen	188.1	59.9	2.02	1.40
Methane	673.0	343.9	16.04	1.32
Natural gas	673.0	343.9	16.04	1.32
Nitrogen	492.4	227.1	28.02	1.40
Oxygen	736.0	278.6	32.0	1.40
Propane	615.9	665.9	44.1	1.12
Steam	3208.2	1165.5	18.02	1.33

the vena contracta. It is also assumed for this analysis that the velocities of the liquid and gas or vapor are essentially equal at the throttling point or at the vena contracta. This discussion considers two cases of mixed-phase flow: (1) a liquid and a gas of different chemical species, and (2) a liquid and vapor of the same chemical species.

Liquid-Gas Mixtures

The procedure for sizing valves for liquid-gas mixtures applies to a homogeneous mixture of a nonvaporizing liquid and a noncondensible gas. As the mixture passes through the valve, the specific volume of the liquid remains essentially constant, but the volume of the gas expands, thereby increasing the effective specific volume, v_e, of the mixture.

$$v_e = f_g v_g / Y^2 + f_f v_f \qquad \textbf{6.15(94)}$$

where f_g and f_f are the weight or mass fractions of gas and liquid, respectively, and Y is the expansion factor from Equation 6.15(81). Equation 6.15(94) expresses v_e as the specific volume of an equivalent incompressible fluid, allowing use of v_e in equations for incompressible fluids.

The required valve capacity, $F_P C_v$, assuming $F_R = 1$, can be obtained from rearranging Equation 6.15(36):

$$F_P C_v = \frac{w}{N_6 \sqrt{(P_1 - P_2)\gamma_1}} \qquad \textbf{6.15(95)}$$

By substituting $1/v_e$ for γ_1 and generalizing the equation for both choked and non-choked flow using Δp_a as the smaller of $P_1 - P_2$ and Δp_{ch} (see Equation 6.15[11]), the valve capacity is

$$F_p C_v = \frac{w}{N_6}\sqrt{\frac{v_e}{\Delta p_a}} \qquad \textbf{6.15(96)}$$

These equations have provided the required C_v values within an error of about ±10% based on limited testing on air and water mixtures in small globe valves in horizontal pipe.[15]

Example 13 The goal in this example is to find the required valve capacity ($F_p C_v$) for the conditions listed below:

Air flow rate: 600 lb/hr
Water flow rate: 26,000 lb/hr
Upstream pressure, P_1: 150 psia
Pressure drop, Δp: 50 psi
Temperature: 90°F (550°R)
Line size: 3 in. schedule 40

Step 1: Make a preliminary selection of valve type and determine the critical pressure drop ratio factor (x_T) for the valve. Assume a single-seated globe valve with a contoured plug with flow under the plug (to open). Using Table 6.15b, obtain an estimate of $x_T = 0.72$. (Manufacturers' catalogs may also be used.)

Step 2: Determine the relative mass fractions of gas and liquid, f_g and f_f. The total mass flow rate is $w = 600 + 26{,}000 = 26{,}600$ lb/hr.

$$f_g = 600/26{,}600 = 0.0226$$

$$f_f = 26{,}000/26{,}600 = 0.9774$$

Step 3: Calculate the pressure drop ratio, (x), the ratio of specific heat factor (F_K), and the expansion factor $(Y) \cdot x = \Delta p/P_1 = 50/150 = 0.333$. Because $x < x_T$, and the gas flow is not choked, $\Delta p_a = \Delta p = 50$ psi.

From Equation 6.15(83), using $k = 1.40$ for air, $F_k = 1.40/1.40 = 1.0$. Using Equation 6.15(81):

$$Y = 1 - \frac{x}{3F_k x_T} = 1 - 0.333/[(3)(1)(0.72)] = 0.846$$

Step 4: Determine the effective specific volume of the mixture at upstream conditions using Equation 6.15(94). The specific volume of the air can be calculated from the gas law equation of state:

$$v_g = \frac{RT_1}{MP_1} \qquad \textbf{6.15(97)}$$

$$v_g = \frac{\left(1545\,\frac{\text{ft}\cdot\text{lbf}}{\text{lb}\cdot\text{mol}\cdot{}^\circ\text{R}}\right)(550^\circ\text{R})}{\left(29\,\frac{\text{lbm}}{\text{lb}\cdot\text{mol}}\right)\left(150\,\frac{\text{lbf}}{\text{in.}^2}\right)\left(144\,\frac{\text{in.}^2}{\text{ft}^2}\right)} = 1.357\,\frac{\text{ft}^3}{\text{lbm}}$$

From saturated steam temperature tables at 90°F, the saturated liquid specific volume is $v_f = 0.016099$ ft^3/lbm. The mixture effective specific volume from Equation 6.15(94) is $v_e = (0.0226)(1.357)/(0.846)^2 + (0.9774)(0.016099) = 0.0586$ ft^3/lbm.

Step 5: Calculate valve capacity from Equation 6.15(96).

$$F_P C_v = \frac{w}{63.3}\sqrt{\frac{v_e}{\Delta p_a}} = \frac{26{,}600}{63.3}\sqrt{\frac{0.0586}{50}} = 14.4$$

Liquid-Vapor Mixtures

Liquid-vapor mixtures are a more difficult, if not impossible, problem owing to the difficulty in determining the inlet vapor fraction and because of the nonequilibrium conditions prevailing as the fluid passes through the valve.

Tests have shown that the flow rate of low-quality (vapor fraction $X < 0.03$) steam-water mixtures can be as much as ten times the amount predicted using calculations based on thermodynamic equilibrium, and about twice the flow predicted assuming no change of state between valve inlet and vena contracta.[15]

Calculations produced better predictions for very high-quality (vapor fraction $X > 0.9$) steam-water mixtures. For high vapor fraction mixtures, the same procedure is used as in Equations 6.15(94–96). For low vapor mixtures, while recognizing the possibility of large errors, use the effective specific volume at inlet conditions, v_e, and Equation 6.15(98).

$$C_v = \frac{w}{N_6 F_{LP}}\sqrt{\frac{v_e}{P_1(1 - F_F)}} \qquad \textbf{6.15(98)}$$

F_{LP} is the combined piping and pressure recovery factor from Equation 6.15(23).

Example 14 Let us assume that the following conditions are given for a liquid-vapor mixture in near-equilibrium conditions. Find the required valve capacity, $F_{LP}C_v$.

Fluid: Ammonia
Vapor flow rate: 200 lb/hr
Liquid flow rate: 9000 lb/hr
Upstream pressure, P_1: 110 psia
Pressure drop, Δp: 5 psi
Temperature: 60°F
Line size: 2 in. schedule 40

Step 1: Make a preliminary selection of valve type, and determine the valve x_T. Assume a throttling (segmented) ball valve. Using Table 6.15c for an estimate, $x_T = 0.25$.

Step 2: Determine the relative mass fraction of gas and liquid. The total mass flow rate is $w = 200 + 9000 = 9200$ lb/hr.

$$f_g = 200/900 = 0.0217$$

$$f_f = 9000/9200 = 0.9783$$

Step 3: Calculate pressure drop ratio (x), ratio of specific heat factor (F_K), and expansion factor $(Y) \cdot x = \Delta p/P_1 = 5/110 = 0.0455$. Since $x < x_T$, the vapor flow is not choked. From Equation 6.15(83) and Table 4.17q, using $k = 1.40$ for air and 1.31 for ammonia, $F_k = 1.31/1.40 = 0.936$. Using Equation 6.15(81):

$$Y = 1 - \frac{x}{3F_k x_T} = 1 - 0.0455/[(3)(0.936)(0.25)] = 0.935$$

Step 4: Calculate F_F from Equation 6.15(10). The critical pressure of ammonia from Table 6.15q is $P_c = 1636.1$ psia. The vapor pressure from ammonia tables[11] at 60°F is $P_v = 107.6$ psia.

$$\begin{aligned} F_F &= 0.96 - 0.28(P_v/P_c)^{1/2} \\ &= 0.96 - 0.28(107.6/1636.1)^{1/2} = 0.89 \end{aligned}$$

Step 5: Determine the effective upstream specific volume at upstream conditions using Equation 6.15(94). The specific volume of the ammonia vapor taken from thermodynamic tables[11] is only approximate, because the vapor is not in true thermodynamic equilibrium. However, using the values for saturated ammonia provides sufficient accuracy due to the relative uncertainty of the capacity estimate. From saturated ammonia tables,[11] $v_g = 2.751$ ft^3/lb and $v_f = 0.02597$ ft^3/lb. Therefore, one can calculate v_e from Equation 6.15(94).

$$v_e = (0.0217)(2.751)/(0.935)^2 + (0.9783)(0.02597) = 0.0937 \text{ ft}^3/\text{lb}$$

Step 6: Calculate valve capacity from Equation 6.15(98).

$$F_{LP}C_v = \frac{w}{63.3}\sqrt{\frac{v_e}{P_1(1-F_F)}} = \frac{9200}{63.3}\sqrt{\frac{0.0937}{110(1-0.89)}} = 12.8$$

CONCLUSIONS

Due to the uncertainties in correctly determining the process conditions, there is a tendency to be overcautious; i.e., to overestimate pressure losses in pipes and to increase the normal flow rates by safety factors. This tends to oversize the control valves. This in turn affects the installed gain of the valve and greatly reduces the turn-down, or rangeability of the valve.

It is recommended that the pump heads be selected on the basis that the control valve will add no more than 10% to the pressure drop in the piping system at maximum flow. This will not only reduce the size of the pump, but will also save on the required operating energy. If this approach is taken to determine the required pump head, the maximum differential pressure across the control valve will be the pump head at maximum flow minus the sum of the pressure drop across the piping system and any altitude changes, plus 10% of the pipe friction losses.

There are instances when the flow conditions are not accurately known. This might be the case if an old valve is to be replaced in a piping system and its serial number or source is not known. In such cases, the "best guess" approach would be to select the same type of valve having the same pressure rating as the surrounding pipe and having a flow coefficient (C_v) one size smaller than that listed for a full size orifice of a line size valve.

Another important rule is to make sure that the maximum required flow capacity for the valve does not exceed 85% of the valve's maximum rated capacity. This is because there is a typical tolerance of ±10% between what is advertised as the maximum C_v number and what actually is the C_v of the particular valve.

Similar rules apply to the minimum required flow coefficient, which is obtained by considering a process condition when the flow is the minimum and the valve pressure drop is the maximum. Once this $C_{v(\min)}$ is determined, one has to make sure that the valve travel at the required minimum flow coefficient is not less than 10% of the rated travel (or twice the expected dead band of the valve, whatever is higher) away from touching the valve seat. This safety margin is necessary to make sure that the plug does not touch the seat during any excursion caused by a disturbance in the control loop.

If the minimum flow capacity of the larger valve is insufficient to provide the required rangeability for the loop, use two valves in parallel by split-ranging the control signal among them, as shown in Figure 6.1i.

NOMENCLATURE

A	Area
A_1	Cross-sectional area of pipe, in.^{2}
A_2	Area of valve opening, in.^{2}
a	Pressure scale exponent
A_v	Metric capacity coefficient, m^2
b	Size scale exponent
C_d	Relative valve capacity or valve discharge coefficients, $C_d = C_v/d^2$ (Table 6.15c)
C_v	Valve capacity coefficient. Its value is 1.0 when the valve passes 1 gpm of cold water, having a specific gravity of 1.0, at a pressure drop of 1 psid.
C_{vt}	Pseudo valve coefficient
d	Valve body port diameter, inches
d/D	Valve to pipe diameter ratio
D_1	Upstream pipe inside diameter, inches
D_2	Downstream pipe inside diameter, inches
F_d	Valve-style modifier
F_F	Liquid critical pressure ratio factor
F_f	Mass fraction of liquid in a liquid-gas mixture
F_g	Mass fraction of gas in a liquid-gas mixture
F_k	Ratio of specific heats factor; k/1.4
F_L	Liquid pressure recovery factor of a valve without attached fittings
F_{LP}	Combined liquid pressure recovery factor and piping geometry factor of a valve with attached fittings
F_P	Piping geometry factor of a valve with attached fittings
F_R	Reynolds number factor
F_s	Laminar (streamline) flow factor
g	Gravitational acceleration, 32.174 ft/s^2
g_c	Units conversion constant, 32.174 (lbm-ft)(lbf-s^2) or 1.0 (kg-m)/(N-s^2)
G_f	Specific gravity relative to water at 60°F
G_g	Specific gravity relative to air 1 atm. and 60°F
H	Enthalpy
K	Ratio of specific heats for gases; 1.4 for air
K_{B1}	Inlet fitting Bernoulli coefficient (static pressure to velocity exchange)
K_{B2}	Outlet fitting Bernoulli coefficient

K_c	Valve cavitation coefficient
K_I	Inlet fitting combined loss coefficient
K_{sc}	Cavitation index of service conditions
K_v	Metric valve sizing coefficient, $(m^3/h)/(bar)^{1/2}$
K_1	Inlet fitting friction loss coefficient
K_2	Outlet fitting friction loss coefficient
m	Mass, lbm
M	Molecular weight
M_N	Mach number
N_1 to N_9	Numerical constants
P	Pressure, psia
P_1	Valve inlet pressure, psia
P_2	Valve outlet pressure, psia
P_c	Critical pressure, psia
PSE	Pressure scale effect for cavitation
P_{vc}	Pressure at the vena contracta, psia
P_v or P_{vp}	Vapor pressure at flowing conditions, psia
P_r	Reduced pressure (used with compressibility curves); defined as P_1/P_c
q	Volumetric flow rate, gpm
Q	Volumetric flow rate, ft^3/hr
R	Universal gas constant, 1545 (ft-lbf)/(lb-mol-°R) or 1.987 cal/(gm-mol-°K)
Re	Pipeline Reynolds number
Re_v	Valve Reynolds number
SSE	Size scale effect for cavitation
T	Temperature, °F or °C
T_a	Absolute temperature, °R
T_c	Critical temperature, °R or °K
T_r	Reduced temperature; defined as T_1/T_c
T_1	Absolute temperature at valve inlet, °R or °K
U	Velocity, ft/s
v	Specific volume, ft^3/lbm
V	Volume, ft^3
v_e	Effective specific volume of a liquid-gas mixture, ft^3/lbm
v_f	Specific volume of saturated liquid, ft^3/lbm
v_g	Specific volume of gas or saturated vapor, ft^3/lbm
W	Mass flow, rate, lbm/hr
X	Vapor portion of a liquid-vapor mixture, related to steam quality
X_Fzi	Differential pressure ratio for incipient cavitation at actual inlet pressure.
X_T	Critical pressure drop ratio factor (inception of choked flows)
X_{TP}	Critical pressure drop ratio factor for a valve and fitting assembly
Y	Expansion factor for gases
Z	Compressibility factor

ΔH	Head loss, ft.
Δp	Pressure difference (drop), psi
Δp_a	Allowable sizing pressure drop (the smaller of $P_1 - P_2$ and Δp_{ch}), psi
Δp_{ch}	Pressure drop at onset of choked flow defined in terms of F_L, psi
Δp_I	Pressure drop at flow curve inflection point where C_v begins to decrease due to cavitation
γ (gamma)	Weight density, lbm/ft^3
γ_1	Specific weight at upstream conditions, lbm/ft^3
μ (mu)	Absolute viscosity, centipoise = 10^{-3} N·sec/m^2
ν (nu)	Kinematic viscosity, centistokes = 10^{-6} m^2/sec
σ (sigma)	Cavitation index at flowing conditions
σ_c	Valve coefficient for constant cavitation
σ_I	Valve coefficient for incipient cavitation
σ_{id}	Valve coefficient for incipient damage cavitation
σ_{mv}	Valve coefficient at maximum vibration cavitation
σ_R	Recommended operating cavitation coefficient for a valve at a reference pressure
σ_v	Valve cavitation coefficient adjusted for pressure and size scale effects

References

1. Baumann, H. D., "The Introduction of a Critical Flow Factor for Valve Sizing," *ISA Transactions*, Vol. 2, 1963.
2. Driskell, L. R., *Control Valve Selection and Sizing*, Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1983.
3. Baumann, H. D., "Effect of Pipe Reducers on Control Valve Capacity," *Instruments and Control Systems*, December 1968.
4. McCutcheon, E. B., "A Reynolds Number for Control Valves," *Symposium on Flow, Its Measurement and Control in Science and Industry*, Vol. 1, Part 3, 1974.
5. Baumann, H. D., "A Unifying Method for Sizing Throttling Valves under Laminar or Transitional Flow Conditions," *Journal of Fluids Engineering*, Vol. 115, No. 1., pp. 166–168, March 1993.
6. Allen, W. F., Jr., "Flow of a Flashing Mixture of Water and Steam through Pipes and Valves," *Journal of Basic Engineering*, April 1951.
7. Singleton, E. W., "Multistage Valve Sizing Made Easy," *InTech*, August 1997, pp. 47–51.
8. George, J. A., "Evolution and Status of Nonturbulent Flow Sizing for Control Valves," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 2002.
9. ANSI/ISA S75.01, "Flow Equations for Sizing Control Valves," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, Revised 1995.
10. ANSI/ISA S75.02-1988, "Control Valve Capacity Test Procedure," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1996.
11. Perry, R. H. and Chilton, C. H., *Chemical Engineer's Handbook,* 5th edition, New York: McGraw-Hill Book Company.
12. *Masoneilan Handbook for Control Valve Sizing,* Masoneilan-Dresser, 2000.
13. *Catalog 10, Technical Information,* Fisher Controls Company.
14. Monsen, J. F., "Spreadsheet Sizes Control Valves for Liquid/Gas Mixtures," *InTech,* December 1990.
15. Hutchison, J. W. (ed.), *ISA Handbook of Control Valves,* 2nd edition, Research Triangle Park, NC: *Instrumentation, Systems, and Automation Society*, 1976.

16. Kiesbauer, J. and Baumann, H. D., “Recent developments in the prediction of hydrodynamic noise of control valves,” *Valve-World,* February 2004.
17. Baumann, H. D., “Determination of Peak Internal Sound Frequency Generated by Throttling Valves for the Calculation of Pipe Transmission Losses,” *Noise Control Engineering Journal,* Vol. 22, No. 1, pp. 75–82, March–April, 1991.
18. Cain, F. M. and Barnes, R. W., “Testing for Cavitation in Low Pressure Recovery Control Valves,” *ISA*, Volume 25, No. 2, 1986.
19. Tullis, J. P., *Hydraulics of Pipelines: Pumps, Valves, Cavitation, Transients,* New York: John Wiley, 1989.
20. Rahmeyer, W., “Cavitation Testing of Control Valves,” Instrumentation, Systems, and Automation Society, Paper No. C.I. 83-R931, presented at International Conference in Houston, TX, October 1983.
21. Riveland, M. L., “The Industrial Detection and Evaluation of Control Valve Cavitation,” Instrumentation, Systems, and Automation Society, Paper No. C.I. 82-909, presented at International Conference in Philadelphia, PA, October 1982.
22. Cain, F. M., “Solving the Problem of Cavitation in Control Valves,” *Advances in Instrumentation and Control—Proceedings of the ISA/91 International Conference,* Vol. 46, Part 2, 1991, Paper #91-0462, pp. 1133–49.

Bibliography

Baumann, H. D., *Control Valve Primer*, 3rd edition, Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1998.

Cain, F. M., “Solving the Problem of Cavitation in Control Valves,” *Advances in Instrumentation and Control—Proceedings of the ISA/91 International Conference,* Vol. 46, Part 2, 1991, Paper #91-0462, pp. 1133–49.

Catalog 10, Technical Information, Fisher Controls Company.

ANSI/ISA S75.02-1988, “Control Valve Capacity Test Procedure,” Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1996.

Driskell, L., *Control-Valve Selection and Sizing,* Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, Revised 1995.

ANSI/ISA S75.01, “Flow Equations for Sizing Control Valves,” Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1985.

Handbook on Control Valve Sizing, Baumann Inc., Portsmouth, NH.

Hutchison, J. W. (ed.), *ISA Handbook of Control Valves,* 2nd edition, Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1976.

ISA S20 Specification Forms for Process Measurement and Control Instrument, Primary Elements and Control Valves, Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1981.

IEC Standard 60534-2-1, “Flow Capacity: Sizing Equations for Fluid Flow under Installed Cconditions,” Geneva, Switzerland: International Electronical Commission, September 1998.

Masoneilan Handbook for Control Valve Sizing, 7th edition, Masoneilan-Dresser, 1987.

Monsen, J. F., “Spreadsheet Sizes Control Valves for Liquid/Gas Mixtures,” *InTech,* December 1990.

Morse, P. M. *Thermal Physics,* W.A. Benjamin, Inc., 1964.

Perry, R. H. and Chilton, C. H., *Chemical Engineer’s Handbook,* 5th edition, New York: McGraw-Hill Book Company, 1973.

Rahmeyer, W., “Cavitation Testing of Control Valves,” Instrumentation, Systems, and Automation Society, Paper No. C.I. 83-R931, presented at International Conference in Houston, TX, October 1983.

Rahmeyer, W. and Driskell, L., “Control Valve Flow Coefficients,” *Journal of Transportation Engineering,* Vol. 111, No. 4, ASCE Pipeline Division Paper No. 19882, 1985.

Riveland, M. L., “The Industrial Detection and Evaluation of Control Valve Cavitation,” Instrumentation, Systems, and Automation Society, Paper No. C.I. 82-909, presented at International Conference in Philadelphia, PA, October 1982.

Shinskey, F. G., “When You Have the Wrong Valve Characteristic,” *Instruments and Control Systems,* October 1971.

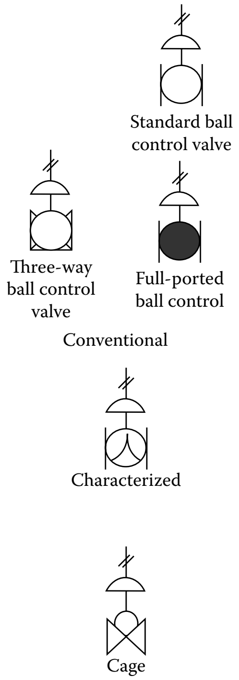
Technical paper No. 410, Crane Company, 2001.

Tullis, J. P., *Hydraulics of Pipelines: Pumps, Valves, Cavitation, Transients,* New York: John Wiley, 1989.

Valtek Control Valve Sizing and Selection, Flowserve Inc. (Valtek Inc.), 1992.

6.16 Valve Types: Ball Valves

C. E. GAYLER F. D. MARTON (1970, 1985) **B. G. LIPTÁK** (1995)
H. L. MILLER (2005)



Flow sheet symbols

Types of Ball Valves:	A. Conventional B. Characterized C. Cage
Size and Design Pressure:	A. ½ to 42 in. (12 to 1180 mm) in ANSI Class 150; to 12 in. (300 mm) in ANSI Class 2500 B. Segmented ball—1 to 24 in. (25 to 600 mm) in ANSI Class 150; to 16 in. (400 mm) in ANSI Class 300; to 12 in. (300 mm) in ANSI Class 600 C. ¼ to 14 in. (6 to 350 mm) C. Up to ANSI Class 2500
Design Temperature:	A. Varies with size and material, typically from −250 to 600°F (−155 to 315°C), with special designs available from −300 to 1800°F (−185 to 1020°C) B. From −50 to 300°F (−45 to 150°C); special units available from cryogenic to 1000°F (540°C) C. From −425°F to 1800°F (−255°C to 980°C)
Rangeability:	Refer to Section 6.7; generally claimed to be about 50:1
Characteristics:	See Figure 6.16a
Capacity:	A and B. Standard ball: $C_v = 30\ d^2$ to $C_v = 45\ d^2$; segmented ball: $C_v = 24\ d^2$ to $C_v = 30\ d^2$; full bore ball: $C_v = 35\ d^2$ to $C_v = 100\ d^2$; see Table 4.3c for details C. $C_v = 20\ d^2$ (noncritical flow)
Materials of Construction:	A. *Body:* Cast or bar stock brass or bronze, carbon steel, stainless steel, ductile iron, aluminum, Monel, titanium, Hastelloy C, plastics, glass; also hafnium-free zirconium (for nuclear applications) and ceramic for abrasives. *Ball:* Forged naval bronze, carbon steel (also plated), stainless steel, plastics, glass, ceramics, Alloy 20, Monel, Hastelloy C, aluminum, titanium. *Seats:* Teflon, Kel-F (both tetrafluoroethylene), Delrin, buna-N, neoprene, Perbunan, Hypalon, natural rubber, graphite.

	B. Body, ball, seal ring, and shaft are available in 316 stainless steel. Chrome and tungsten carbide plating available for ball and carbon steel for valve bodies. All-ceramic valves are also available. C. Stainless steel, other materials available.
Leakage:	A. ANSI V (see Section 6.1). B. Metal seats, ANSI IV to VI; composition seats, ANSI V or better.
Costs:	A. Low-cost ½ to 1 in. PVC valves with on/off actuators can be obtained for $300 or with positioned throttling actuator for about $1200. For process control-quality valve costs, see Figure 6.16b below. Cost data are based upon ANSI Class 150 flanged bodies with single-acting, spring-return actuator for on/off service. Other operators can be furnished and positioners added for throttling service. Ball and stem materials are 316 SST for both carbon and stainless steel bodies with TFE seat seals. Other alloys are available at higher cost. B and C. See Figure 6.16b for on/off; modulating with metal seats for high temperature will be three to four times higher.
Special Features:	A. Full-ported, three-way, split body, two-directional. B. Depending on "contour edge" of ball the flow characteristics vary slightly between suppliers. Slurry design provides for continuous purging of low-activity zone of valve to prevent build-up of solids, dewatering, or entrapment. C. Good resistance to cavitation and vibration.
Partial List of Suppliers:	A Plus Development Co., Ltd. (A) (www.acndc.com) Actuation Valve & Control Ltd. (A) (www.actuation.co.uk) Armstrong International Inc.(A) (www.armstrong-int.com) Assured Automation Inc. (A) (www.assuredautomation.com) Bardiani Valvole SpA (A) (www.bardiani.com) China Zhejiang Chaoda Valve Co. Ltd (www.chinavalve.com) Circor Int., Inc. (Circle Seal controls) (A) (www.circleseal/com) Cole-Parmer Instrument Co. (A) (www.coleparmer.com) Combraco Industries Inc. (www.combraco.com) Control Components Inc. (www.ccivalve.com) Cooke Vacuum Products Inc. (A) (www.cookevacuum.com) Cooper Cameron Valves (A) (www.ccvalve.com) Crane Valve Group (Stockham, Xomox) (A) (www.craneco.com) Cyclonic Valve Co. Inc. (A) (www.cyclonic.com) Dresser (Masoneilan) (www.dresser.com) Derex Company (C) (www.derex.com) Emerson Process Management (Fisher) (www.Emersonprocess.com) Eurovalve s.r.l.(A) (www.eurovalve.it) Flowdyne Controls, Inc. (A) (www.flodynecontrols.com) Flowserve Corp (Anchor/Darling, Valtek, Worcester, McCanna,) (A, B) (www.flowserve. com) Fujikin of America (A) (www.fujikin.com) Hartmann KG (A) (www.hartmann-valve.com) Hoke Inc. (A) (www.hoke.com) Honeywell (A) (www.honeywell.com) ITT Industries, Engineered Valves (A) (www.engineeredvalves.com) Kitz Corporation of America (A) (www.kitz.com) Marpac Inc. (A) (www.mccannainc.com) Metso Automation (Jamesbury) (A,B) (www.metsoautomation.com) Milwaukee Valve Co. (A) (www.milwaukeevalve.com) Mogas Industries Inc. (A, B) (www.mogas.com) Nibco Inc. (A) (www.nibco.com) Nihon Koso (A,B) (www.koso.co.jp) Nordstrom Valves Inc. (A) (www.nordstromvalves.com) Oliver Valves Inc. (A) (www.olivervalves.com) Parker Hannifin Corp. (A) (www.parker.com)

PBM Inc.(A) (www.pbmvalve.com)
Plast-O-Matic Valves, Inc. (A) (www.plastomatic.com)
Power and Pumps, Inc. (A) (www.powerandpumps.com)
Spirax Sarco Ltd. (A) (www.spiraxsarco.com)
SPX Valves (Daniel, Dezurik) (A,B) (www.spxvalves.com)
Tyco Flow Control (A) (MCF) (www.tycovalves-na.com)
Valvtechnologies Inc. (A) (www.valv.com)
Velan Valve Corp. (A) (www.velan.com)
Voss/Europower Inc. (A) (www.vossinc.com)
Watts Regulator Co. (A) (www.wattsregulator.com)
Wier Valves and Controls (A) (www.weirvalve.com)
Zurn Industries (A) (www.zurn.com)

INTRODUCTION

The rotary ball, butterfly, and plug valves, which in the past were considered only as on/off shut-off valves, are extensively used today as control valves. Relative to the traditional globe valve, their advantages include their lower cost and weight and higher flow capacity (two to three times that of the globe valve, as listed in Table 6.16c). Other features, such as tight shutoff, fire-safe designs, and low stem leakage, make it easier to meet governmental regulatory requirements from OSHA and EPA in the United States and the Pressure Equipment Directive (PED) in the EEC. Some ball valve designs, such as the characterized ball valve, also provide a near-equal-percentage characteristic. For a comparison of ball and globe valves in terms of their features and performance, see Reference 1.

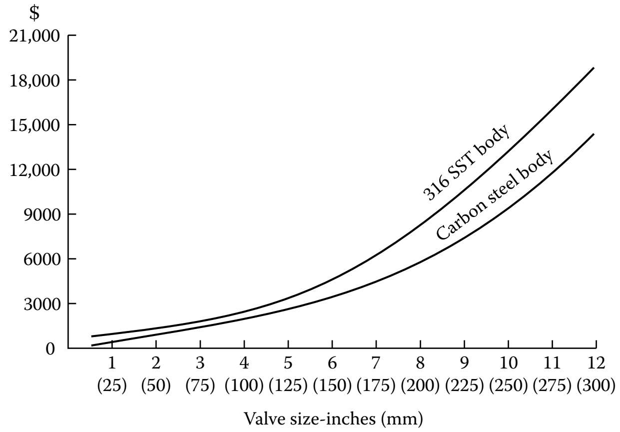


FIG. 6.16b
The costs of conventional ball valves with on/off actuators are provided above. For characterized ball valves add 10%, and for cage-type ball valves add 20% to these costs.

Throttling Ball Valves

When used for throttling service, some of their disadvantages are a direct consequence of the above-listed advantages. Their high capacity results in either using oversized valves or installing small valves in large pipes. This means a substantial waste of pumping energy caused by the reducer pressure drops. Also, the characteristic of a high-pressure recovery results in low vena contracta pressures. This in turn increases the probability of cavitation and noise.

Reduction of these problems has been achieved by adding perforated parallel plates in the ball valve openings (see Figure 6.16o) or by adding other tortuous path features (see Figure 6.16p) to the initially fully open flow path. These design improvement have added measurably to the range of trouble-free applications.

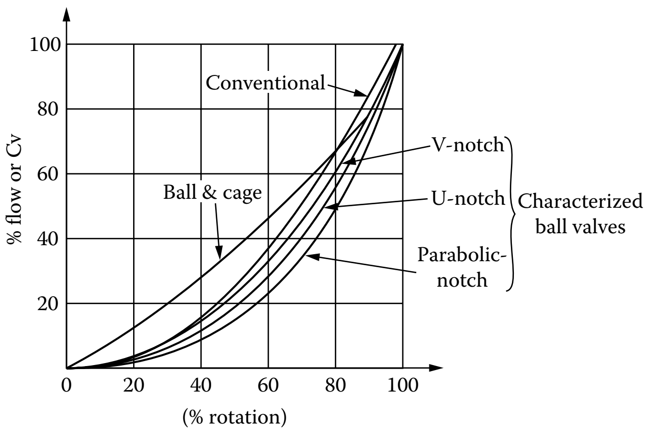


FIG. 6.16a
The characterized ball valve with a parabolic notch is near-equal-percentage, while the ball-and-cage valve characteristics are closer to linear, when used on water service. On gas service at critical velocities, the characterized ball valve's performance lines move closer to linear.

In a ball valve, critical flow occurs when the pressure drop through the valve rises to about 15% of the inlet pressure to the valve. In operating rotary valves, the linear movement of cylinder- or spring/diaphragm-type actuators must be converted by linkages, which introduces hysteresis and dead play. In addition, a nonlinear relationship exists between actuator movement and the resulting rotation. These considerations make the use of positioners essential, which on fast processes can lower the quality of control.

TABLE 6.16c
*Valve Sizing Coefficients (C_v) of Full Ported and Reduced Ported Ball Valves**

	Full Bore Valves, Ball Opening (%)									
Valve Size	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
½	0.52	1.17	2.0	2.8	3.9	5.7	9.1	13	19.8	26
¾	1.00	2.25	4.0	5.5	7.5	11.0	17.5	25	38	50
1	1.88	4.23	7.5	10.3	14.1	20.7	32.9	47	71.5	94
1½	5.2	11.7	20.8	28.6	39	57.2	91	130	197.6	260
2	9.6	21.6	38.4	52.8	72	105.6	168	240	364.8	480
2½	15.0	33.8	60	82.5	112	165	262.5	375	570	750
3	26	58.5	104	143	195	286	455	650	988	1300
4	46	103.5	184	253	345	506	805	1,150	1,748	2300
6	108	243	432	594	810	1,188	1,890	2,700	4,100	5400
8	200	450	800	1,100	1,500	2,200	3,500	5,000	7,600	10,000
10	320	720	1,280	1,760	2,400	3,520	5,600	8,000	12,160	16,000
12	480	1,080	1,920	2,640	3,600	5,280	8,400	12,000	18,240	24,000
14	628	1,413	2,512	3,454	4,710	6,908	10,990	15,700	23,860	31,400
16	860	1,935	3,440	4,730	6,450	9,460	15,050	21,500	32,680	43,000
18	1,140	2,565	4,560	6,270	8,550	12,540	19,950	28,500	43,320	57,000
20	1,460	3,285	5,840	8,030	10,950	16,060	25,550	36,500	55,480	73,000

	Reduced Bore Valve, Ball Opening (%)									
Valve Size	*10*	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
1½	1.9	4.4	7.8	10.8	14.7	21.6	34.3	49	74.5	98
2	3.6	8.1	14.4	19.8	27	39.6	63	90	136.8	180
2½	5.6	12.6	22.4	30.8	42	61.6	98	140	212.8	280
3	8.4	18.9	33.6	46.2	63	92.4	147	210	319.2	420
4	15.4	34.7	61.6	84.7	115.5	169.4	269.5	385	585	770
6	36	81	144	198	270	396	630	900	1,368	1800
8	68	153	272	374	510	748	1,190	1,700	2,584	3400
10	118	265.5	472	649	885	1,298	2,065	2,950	4,484	5900
12	160	360	640	880	1,200	1,760	2,800	4,000	6,080	8000
14	240	540	960	1,320	1,800	2,640	4,200	6,000	9,120	12,000
16	280	630	1,120	1,540	2,100	3,080	4,900	7,000	10,640	14,000
18	360	810	1,440	1,980	2,700	3,960	6,300	9,000	13,680	18,000
20	440	990	1,760	2,420	3,300	4,840	7,700	11,000	16,720	22,000

*Courtesy of KTM Products, Inc.

The torque characteristics of these valves are also highly nonlinear (Figure 6.4v), and because of the high “break-torque” requirement, the actuator is usually oversized for the operation in the throttling range. Some pneumatic piston actuators have a torque characteristic that has a high torque delivery at the closed position, therefore allowing a better matching of the torque needs (see Figure 6.4w).

The ball valve contains a spherical plug that controls the flow of fluid through the valve body. The three basic types of ball valves that are manufactured are (1) the conventional or quarter-turn pierced ball type, (2) the characterized type, and (3) the cage type.

CONVENTIONAL BALL VALVES

The quarter-turn (90°) required to fully uncover or fully cover an opening in the valve body can be imparted to the ball either

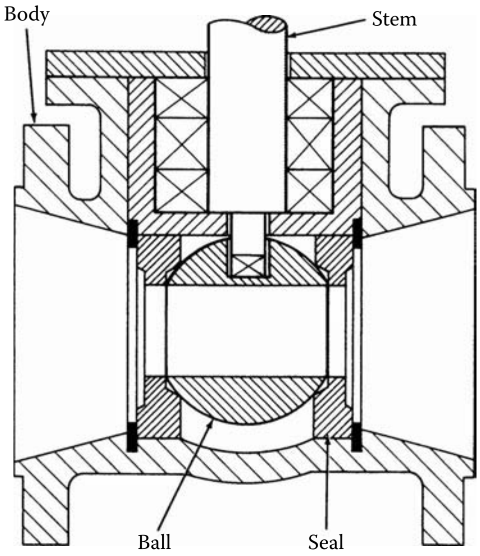


FIG. 6.16d
Top-entry pierced ball valve.

manually by turning a handle, or mechanically by an automatic valve actuator. Actuators used for ball valves may be the same as those used to control other valve types. As was discussed in Sections 6.3 and 6.4, they can be pneumatic, electric (including electronic or digital), hydraulic, or a combination.

The latter types include electropneumatic, electrohydraulic, electromechanical, and pneumohydraulic actuations. Most of the ball valves on the market are available with built-in, or integral, valve actuators. The valves are designed so that they can be used with or without an actuator, and they can be fitted with other manufacturers' actuators.

The spherical plug lends itself not only to precise control of the flow through the valve body but also to tight shut-off. Thus the ball valve may assume the double role of control and block valve (Figure 6.16d). Special materials used for valve seats help achieve these functions.

Ball valves are available with the features listed in the front of this section. Their tight shut-off characteristics correspond to ANSI Class IV and VI. The valve body can be configured as two-way, three-way, or split-body. Figure 6.16e illustrates some of the available multiport configurations.

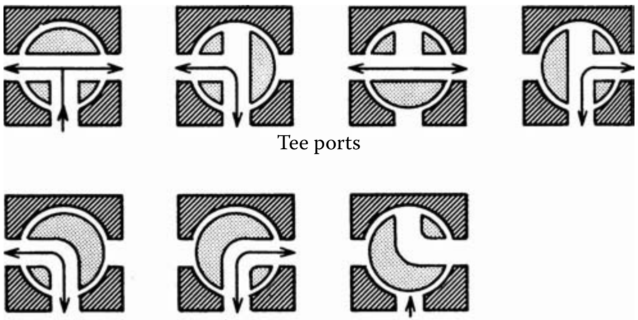


FIG. 6.16e
Porting arrangements of various multiport ball valve designs.

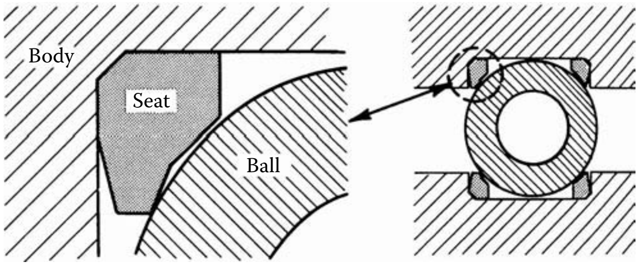


FIG. 6.16f
The design of ball valve seats.

The Valve Trim

The ball in a ball valve is cradled by seats on the inlet and the outlet side. The seats are usually made of plastic and are identical on both sides, especially in double-acting valves. Tetrafluoroethylene materials are preferred for seat materials, because of their good resilience and low-friction properties (Figure 6.16f).

In some valve designs the plastic seats are backed up by metallic seats in order to ensure tightness in the event that the soft seat gets damaged by high temperature, such as in a fire (Figure 6.16g). Such precautions are imperative on shipboard, for nuclear installations, and in cryogenic applications.

The fire-safe design of ball valves can be certified to API-607, which specifies the types of secondary seats that are acceptable to control the leaking of flammable fluids, when the primary seat (usually PTFE) sublimes during fire. The secondary metal seats of the fire-safe designs are also useful on erosive or abrasive services and on saturated steam service.

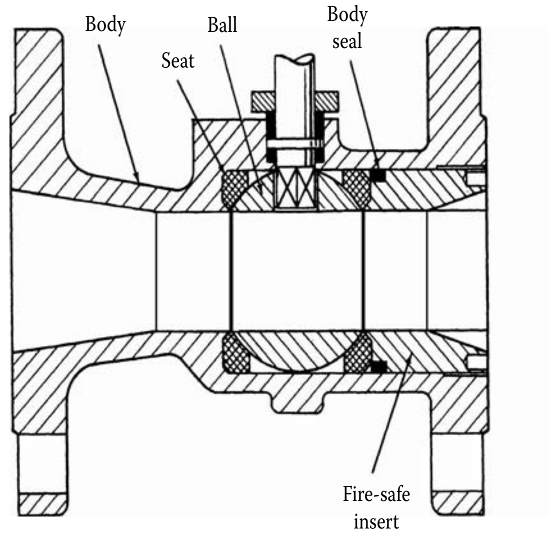
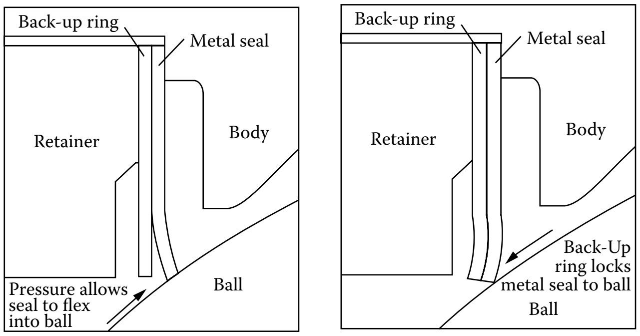
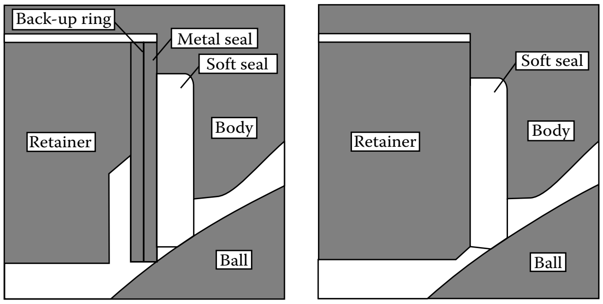


FIG. 6.16g
Fire-safe ball valve design.



Bi-directional flex-loc seal



Soft seals

FIG. 6.16h
Ball valves can be sealed by using either flexible metal seals or soft seals. (Courtesy of Valtek Inc.)

Some seat designs are such that the seat is always under compression, which allows the use of such nonresilient seat materials as carbon graphite.

Seats and Their Maintenance Other seat designs utilize the flexing of metal seals or soft PTFE seats, usually backed up by stainless steel or Inconel metal seals (Figure 6.16h). Where fluids of high temperature are handled, graphite seats are recommended. They hold tight up to 1000°F (540°C).

There have been significant advancements in the manufacturing processes that allow the use of primary metal seats in direct contact with the ball (see Figure 6.16i). Such designs can deliver better than ANSI Class V performance in terms of their seat leakage.[1]

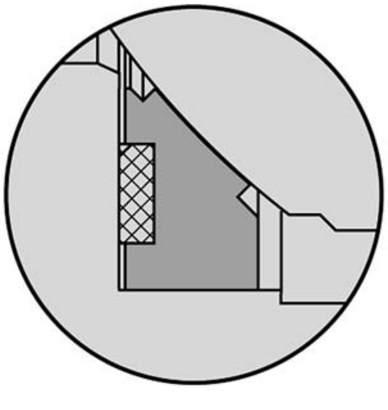


FIG. 6.16i
Contoured metal ball seal. (Courtesy of Metso Automation.)

The ball valves are designed so that lubrication is unnecessary and the torque required to turn the ball is negligible. Both upstream and downstream seats of the pierced ball can sometimes be freely rotated in order to reduce wear. In some designs the seats are forcibly rotated a fraction of a turn with each quarter turn of the ball. Thus seat wear, which is concentrated at the points where the flow begins or ends on opening or closing of the valve, is distributed over the periphery of the seat.

To facilitate cleaning or replacing worn seats in some designs the whole seating assembly is made in the form of a tapered cartridge. If the valve has top-entry design, the cartridge can be removed without disturbing the valve arrangement. O-rings usually close off stem and seats, and thrust washers made of tetrafluoroethylene compensate for axial stem thrust due to line pressure and reduce stem friction to a minimum.

Some seats are preloaded by springs or are made tapered for wear compensation and leak-tight closure.

Ball and stem are often machined from one piece. Other designs use square ends on the stem to engage in square recesses of the ball. In this case, the ball is made floating in fixed seats, while other designs provide a fixed location of ball and stem through the application of top and bottom guiding and ball bearings.

Balls are subject to wear by friction. Where long life and dead-tight closures are of paramount importance, the design is recommended that provides for lifting the ball off its seat before it is turned. This measure also prevents freezing or galling. Lift-off is achieved by mechanical means such as an eccentric cam. Valves of this design facilitate the handling of slurries and abrasive fluids, and they can be used for high pressures.

The proper materials for body and trim depend on the application. For handling chemicals or corrosive fluids, all wetted parts will possibly require stainless steel, plastics, or glass (borosilicate glass is preferred for impact strength).

Ball valves are made with the same connections as used in all other valve types. Where screwed connections must be used, valves with ends that take the place of unions are preferred.

Flow Characteristics

The flow characteristics of a ball valve approximate those of an equal-percentage plug (Figure 6.16a). These characteristic curves compare favorably with those of other rotary-stem valves. The flow path through a ball valve includes two orifice restriction locations (Figure 6.16j). Balls characterized either by a notch or by a noncircular bore give somewhat better characteristics (Figure 6.16a). Critical flow in ball valves is encountered at $\Delta P = 0.15P_1$, far below the usual figure of 50% of absolute inlet pressure.

Sizing of ball valves proceeds along the lines described in Section 6.15, with the possible exception that due to the

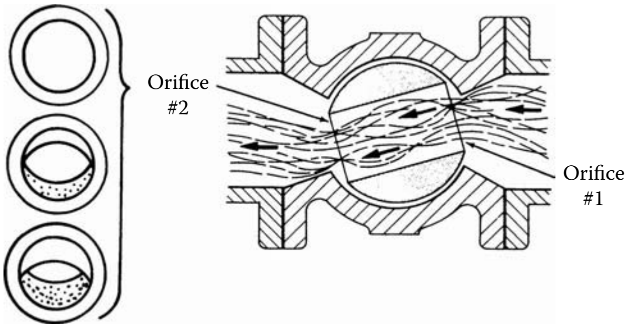


FIG. 6.16j
Two restricted orifice locations are formed when a ball valve is used for throttling the flow.

essentially straight-through flow feature, a ball valve can be chosen whose size is equal to the nominal pipe size, which usually is an advantage on slurry service. This can only be done if an oversized valve can be tolerated, which also implies the acceptance of reduced sensitivity and increased cost. The low-pressure loss and high-pressure recovery of a ball valve must also be considered in the calculations.

The application of ball valves for control requires some caution. Where a lot of noise is encountered some have found it necessary to bury the valve. Such may be the case with natural gas flowing at high speed. Valves handling liquids could cavitate and erode the pressure boundary as well as produce substantial noise.

On the other hand, the use of ball valves to control liquid oxygen in the experimental X-15 air-space vehicle as early as 1961, and later in the Atlas rocket, attests to the precise controllability with ball valves. A special valve with a dual-ball design for mixing liquid oxygen with liquid ammonia in precise proportions has also been used in the Atlas rocket system.

CHARACTERIZED BALL VALVES

The characterized ball valves can be V-notched, U-notched, parabolic, and anticavitation-antinoise designs. The notched trim valves and the partial-ball trim valves were introduced partially in an effort to solve the problem of valve clogging and dewatering in paper stock applications. Since then these valves have come into more widespread use as a result of increased valve rangeability and the shearing action at the sharp edges of the valve as it closes.

In essentially all characterized ball valves, the “ball” has been modified so that only a portion of it is used (Figure 6.16k). The edge of the partial ball can be contoured or shaped to obtain the desired valve characteristics. The V-notching of the ball in Figure 6.16k serves this purpose as well as the purpose of shearing the process stream.

This shape or contour of the valve’s leading edge is the main difference between the various manufacturers’ products. The ball is usually closed as it is rotated from top to bottom, although this action can be reversed.

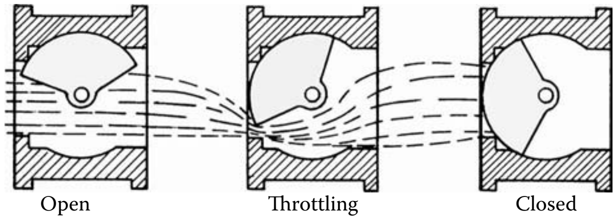


FIG. 6.16k
The open, throttling, and closed positions of the characterized ball valve.

Construction

Mechanically, the characterized ball valves are very similar to their ancestor, the conventional full ball designs. However, because of the asymmetrical design, the characterized ball valve has some design problems that are not significant with the conventional ball valves. A typical characterized ball valve is shown in Figure 6.16l, in its end and side views. The main parts of a characterized ball valve are described below.

The controlling edge of the ball can be notched or contoured to produce the desired flow characteristics. Characterized ball valves are available as U-notched, as V-notched, and as a parabolic curved designs. Mechanically, this part can create problems by bending under pressure and thus introducing movement into the shaft seals. Also, the stub shafts can be distorted by the bending of the partial ball under operating loads.

Early bodies were not designed for high-pressure services or for installations other than insertion between flanges. Today they are available with up to 12 in. (300 mm) flanges with up to ANSI Class 600 ratings.

The seal ring and seal-retaining ring are usually held in place by companion flanges. Damage due to overtightening of flange bolts sometimes occurs. Figure 6.16m illustrates a special sealing arrangement useful in slurry applications, due to the purging effect created by the flow into the otherwise low-activity zone, through the indent in the ball plug.

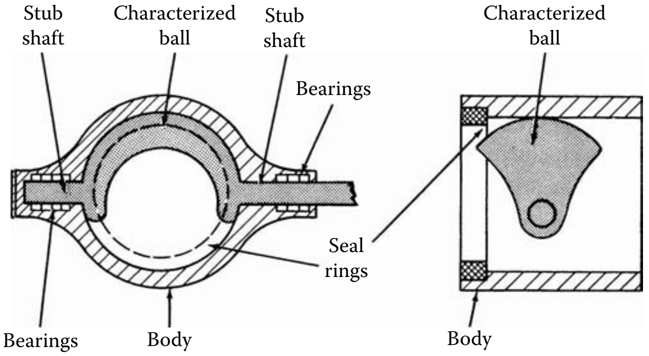


FIG. 6.16l
The component parts of a characterized ball valve.

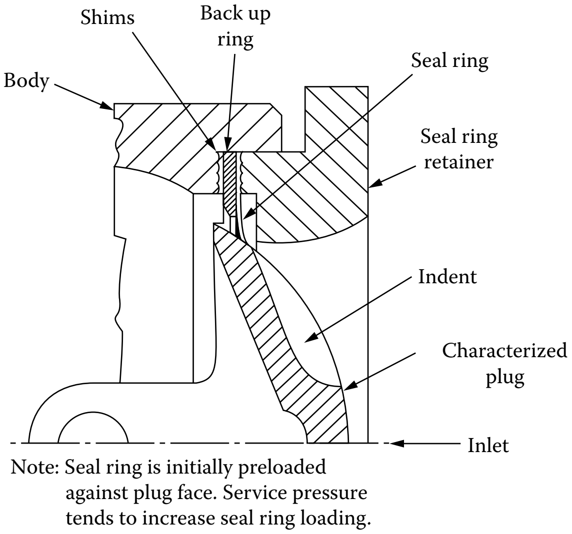


FIG. 6.16m
Special seal ring arrangement, where service pressure increases seal ring loading.

Characteristics

The flow characteristics are dependent upon the shape of the edge of the partial ball and on the installed flow direction. The shape of the V-notch at the edge of the valve varies from concave for small openings to convex for large openings. Figure 6.16n illustrates this characteristic together with the corresponding shapes for the parabolic ball valves.

The flow characteristics for parabolic, U-notched, and V-notched valves are given in Figure 6.16a. These curves are based on water flow and are also applicable to compressible fluid flow at less than critical (choking) velocities. If the characteristics were evaluated using compressible fluids at critical velocities, these curves would be flatter, closer to linear.

The characteristics of the conventional ball valves are also modified by the anticavitation and antinoise designs. One such design approach is illustrated in Figure 6.16o. Here the attenuator is placed inside the ball, so that when the valve is throttling, the fluid has to pass the attenuators, creating a number of pressure drop stages. The size, location, and distribution of perforations on the attenuator plates can be modified to obtain changes in the valve characteristics. These valves can also handle impurities in the process fluid.

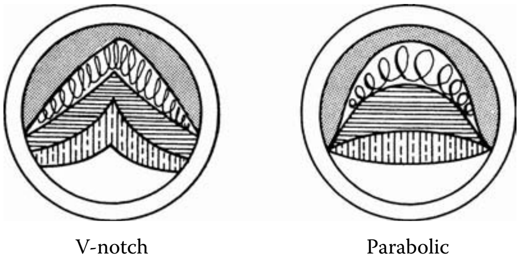


FIG. 6.16n
Shapes of throttling areas of the V-notched and U-notched characterized ball valves.

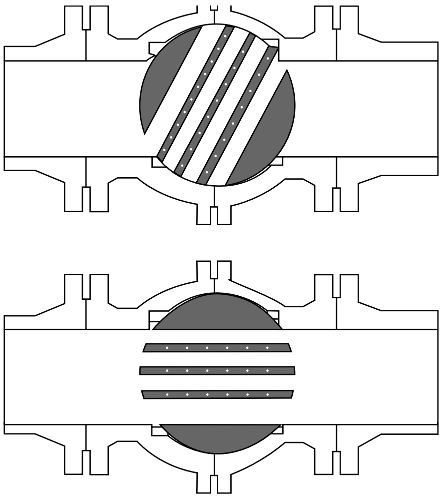


FIG. 6.16o
Internal attenuation plates reduce noise/cavitation and can also modify the throttling characteristics of ball valves.[1]

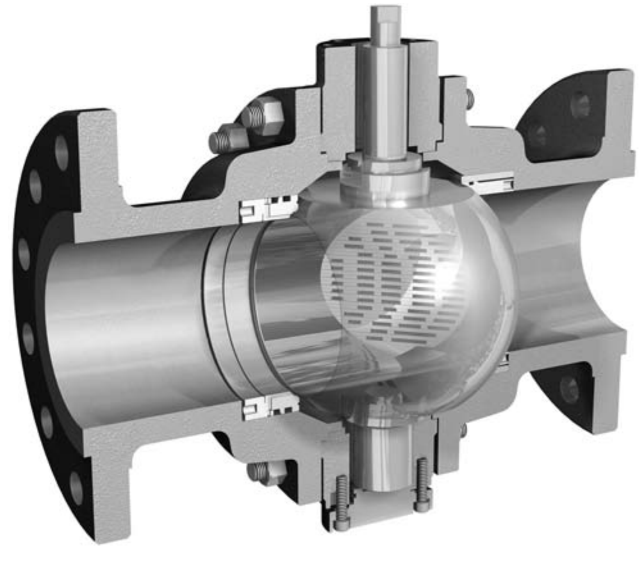
Another approach to the anticavitation, antinoise designs is to add a multistep tortuous trim feature in the ball so that, because of the small openings, the fluid velocity is maintained below damaging levels. Such a design having 16 discrete stages of pressure drop is shown in Figure 6.16p. This design provides a rangeability of up to 300:1 and operating noise levels of less than 75 dBA. The flow passages are designed to continually expand so that any solids that enter the trim will pass through and solids that are blocked will be swept through the valve at larger openings.

The anticavitation and antinoise designs are particularly useful in eliminating the need for low-flow bypass systems, which have been used in the start-up and shut-down of many gas pressure regulation applications.

BALL AND CAGE VALVES

Positioning of a ball by a cage, in relation to a seat ring and discharge port, is also used for control (Figure 6.16q). This valve design consists of a venturi-ported body, two seat rings, a ball that causes closure, a cage that positions the ball, and a stem that positions the cage. Seat rings are installed in both inlet and discharge, but only the discharge ring is active. The body can be reversed for utilization of the spare ring.

The cage rolls the ball out of the seat as it is lifted by the stem, positions it firmly during throttling, and lifts it out of the flow stream for full opening (Figure 6.16r). The cage is contoured for unobstructed flow in the open position. Cage design includes four inclined control surfaces. The two surfaces next to the downstream seat lift the ball out of the seat and roll it over the top edge of the seat ring as the valve is opened.



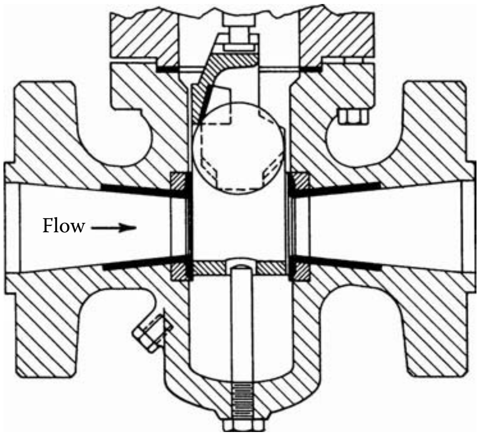


FIG. 6.16q
The ball is positioned by a cage in this valve design.

FIG. 6.16p
Multistage low-flow control valve design with anticavitation and antinoise capability. (Courtesy of Control Components Inc.)

As the valve opens farther, the ball rolls down the first two inclined surfaces to the center of the cage to rest on all four inclined surfaces. The Bernoulli effect of the flowing stream holds the ball cradled in this position throughout the rest of the stroke. A nonrotating slip stem is guided by a bushing at the bottom and by a gland at the top of the bonnet. A machined bevel near the base of the stem acts as a travel limit and allows for back-seating.

Sizes and Other Features

Ball and cage valves are furnished in sizes from 1/4–14 in. (6–350 mm), with ratings from 150–2500 PSIG (1–17 MPa). Reported flow coefficients (C_v) are consistently high. The flow characteristic reflects the increasing enlargement of the crescent between the surface of the ball and the discharge port (Figure 6.16a). With a flow characteristic starting at zero flow, the rangeability is very high, over 50:1, depending only upon the ability of the actuator to position the cage.

Tight shut-off occurs over a long operating life due to the continual rotation of the ball at each operation, which offers a new seating surface each time it is closed. Closure is positive due to the wedging of the cage in addition to line pressure. Although tightly closed, the stem force for opening is approximately 25% of a single-seated globe valve due to the manner in which the inclined surfaces of the cage roll the ball away from the seat.

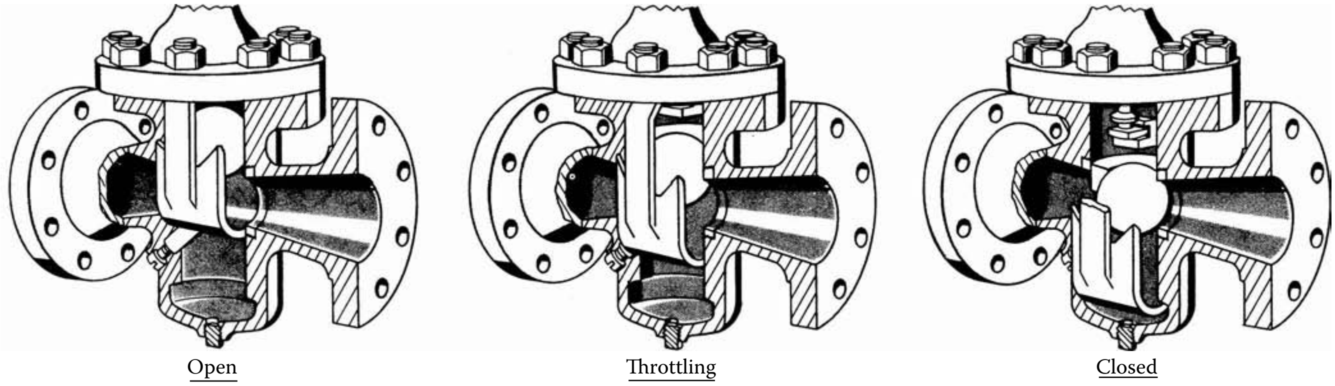


FIG. 6.16r
The open, throttling, and closed positions of a cage-positioned ball valve.

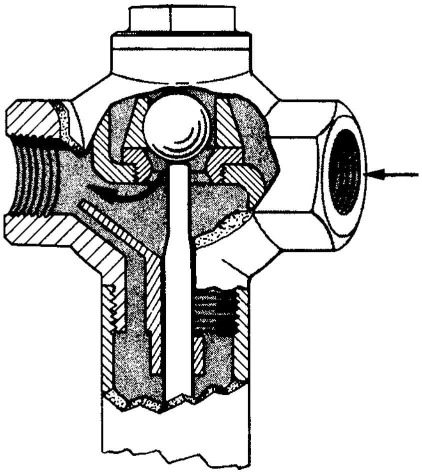


FIG. 6.16s
Valve design in which the ball is unseated by the valve stem. (Courtesy of Powers Process Control.)

Opening and closing force factors have been determined for all sizes of the valve. The low opening force requirement (9520 lb for a 4 in. valve at 2000 PSIG, or 42,400 N for a 100 mm valve at 13.8 MPa) is beneficial in being able to use a relatively small actuator. The design is conducive to minimizing cavitation effects because flow tends to follow the curve of the ball, thus reducing turbulence. Cavitation tends to occur in the venturi passage, not at the seat, allowing use of hardened or replaceable throats. The expanding venturi discharge assists in handling flashing liquids.

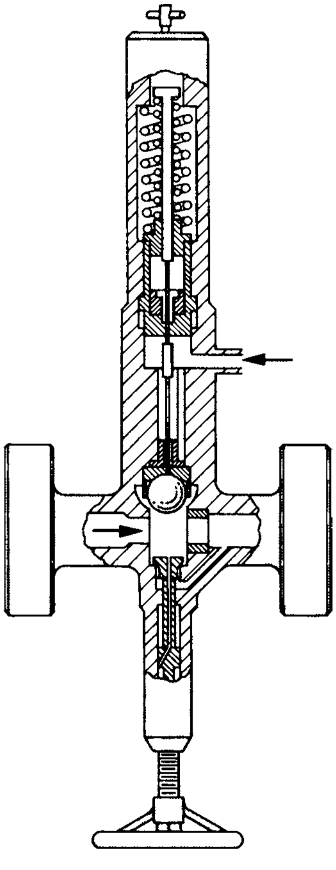


FIG. 6.16t
Ball-and-cage valve design, which is used to provide emergency closure.

The bonnet design lends itself to the adaptation of a variety of linear or rotary actuators. Because the ball must be moved completely out of the flow stream, stem travels are at least as much as the diameter of the valve throat.

Ball Unseated by Stem

The ball cage has also been used for regulators. The ball is cradled in the cage with the valve installed in the vertical position (Figure 6.16s). The stem of the regulator, coming from below the ball, forces the ball away from the seat. Flow is around the ball through the annular space, similar to the flow in a single-seated valve.

Ball Gripped by Cage

A variation of the ball-and-cage design is used for emergency closure (Figure 6.16t). Separate springs and ejection pistons allow high and low limit settings. Pressure above the high setting pushes the piston down to eject the ball from the holder into the seat. Low pressure allows the low-pressure spring to push the piston down. The ball is held firmly on the seat by the differential pressure. An internal bypass is opened to equalize the system pressures. Rotation of the bypass handwheel moves the ball back into the holder, the reset rod is retracted, and the bypass valve closed for normal operation.

References

1. Bruckent, F. W., “Using Ball Valves in Control Applications,” *Hydrocarbon Processing,* August 2002.
2. Husu, M., “Trimming Control Valve Noise/Cavitation,” *InTech,* December 1990.

Bibliography

Baumann, H. D., “Trends in Control Valves and Actuators,” *Instruments and Control Systems,* November 1975.

Carey, J. A., “Control Valve Update,” *Instruments and Control Systems,* January 1981.

Cox, J., “All Thanks to Metallic Sealing Systems: Ball Valves Withstand the Toughest Operating Conditions,” Valve World 98 Conference, KCI Publishing BV, The Netherlands, 1998.

Dobrowolski, M., “Guide to Selecting Rotary Control Valves,” *Instrumentation Technology,* December 1981.

Fernbaugh, A., “Control Valves: A Decade of Change,” *Instruments and Control Systems,* January 1980.

Hammitt, D., "Rotary Valves for Throttling," *Instruments and Control Systems,* July 1977.

Holton, A. D., "Control Valve Update," *Instruments and Control Systems,* January 1982.

Lucitti, A. "Choosing Valves for Isolation: The Case for Ball Valves," *Valve Magazine,* Winter 2004.

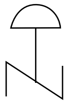
Monsen, J. F., "Valve Wars — Rising Stem vs. Rotary," *Plant Services,* January 1999.

Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 Instrumentation, Systems, and Automation Society Conference, Houston, October 1992.

Ytzen, G. R., "Ball Valves for Throttling Control," *19th ISA Conference,* Preprint No. 11.6-2-64.

6.17 Valve Types: Butterfly Valves

C. E. GAYLER (1970) **B. G. LIPTÁK** (1985) **J. B. ARANT** (1995, 2005)



Flow sheet symbol

Types of Designs:	A. General-purpose, aligned shaft B. High-performance, offset (eccentric) shaft
Sizes:	A. 2 to 48 in. (51 mm to 1.22 m) are typical, but units have been made in sizes from 0.75 to 200 in. (19 mm to 5 m) B. 4 to 16 in. (0.1 to 0.4 m) are common, but units are available from 2 to 80 in. (50 mm to 2 m)
Design Pressures:	A. Most are available through ANSI Class 300 ratings and for up to 200 psid (1.4 MPa) pressure drop. Special units have been designed for up to 6000 PSIG design pressure B. For installation purposes most are available through ANSI 600 ratings and for up to 720 psid (5 MPa) pressure drop
Design Temperature:	A. –450 to 1000°F (–268 to 538°C). Special refractory lined units have been made for up to 2200°F (1204°C) B. –320 to 450°F (–196 to 232°C) for Teflon-seated valves; 1200°F (649°C) for metal-seated ones. Special units are available up to 1700°F (927°C)
Body/Disc Materials:	A. Iron, ductile iron, carbon or alloy steels, stainless steel (302–316), aluminum bronze, Alloy 20, Monel, Hastelloy C, titanium, chrome plating, nickel plating, Kynar, Nordel, Viton, EPDM, Buna-N, neoprene elastomer lining, TFE encapsulation B. Steel, 316 stainless steel, alloy steel, Durimet 20, aluminum bronze, Alloy 20, Monel, Hastelloy C, titanium, tungsten titanium carbide (TTC) coating
Seal Materials:	A and B. TFE, Kel-F, EPT, polyethylene, PTFE with titanium, Inconel, or 316 stainless steel or other metals
Characteristics:	See Figure 6.17a
Rangeability:	Generally claimed as 50:1
Leakage:	A. Unlined, 2 to 5%; lined, ANSI V (see Table 6.1gg) B. Metal seat, ANSI IV; soft (toggle) seat, ANSI VI
Capacity:	A. With 60° rotation, $C_v = (17 \text{ to } 20)d^2$. Typical for throttling with 75° rotation, $C_v = (25 \text{ to } 30)d^2$; with 90° rotation, $C_v = (35 \text{ to } 40)d^2$ B. $C_v = (20 \text{ to } 25)d^2$; see Figure 6.17c
Special Features:	Reduced torque disc designs, fire-tested seals, reduced noise disc, special disc seal designs
Cost:	See Figure 6.17b
Partial List of Suppliers:	ABB Kent Inc. (www.abb.com) AMRI, Inc (www.amrivalves.com) Bray Controls (www.bray.com) Cashco Inc. (www.cashco.com) Circle Seal Controls (www.circle-seal.com) DeZurik/SPX Valves & Controls (www.spxvalves.com) Fisher Controls International Inc. (www.fisher.com) Flowserve, Flow Control Div.—Valtek (www.flowserve.com) FMC Blending & Transfer (www.fmcblending.com)

Foxboro-Invensys (www.foxboro.com)
George Fischer Inc. (www.us.piping.georgefischer. com)
Halliburton Energy Services (www.halflow.com)
Honeywell Industry Solutions (www.iac.honeywell.com)
ITT Industries, Engineered Valves (www.engvalves.com)
Keystone International (www.ikeystone.com)
Love Controls Corp. (www.love-controls.com)
Metso Automation (www.jamesbury.com)
MKS Instruments Inc. (www.mksinst.com)
Nibco Inc. (www.nibco.com)
North American Mfg. Co. (www.namfg.com)
Tyco Flow Control (www.tycovalves-na.com)
Ultraflo Corp.
Xomox (www.xomox. com)

INTRODUCTION

The orientation table (Table 6.1a) compares the main features of butterfly valves to other control valve designs. The rotary valves such as butterfly, ball, and plug valves were once considered to be only on/off valves. In recent decades the rotary valves in general and the butterfly design in particular have been used more and more as throttling control valves.

Relative to the traditional globe control valve, the butterfly valves have the advantages of lower cost and weight, two to three times the flow capacity of globe valves (as shown in Figure 6.17c), fire-safe designs, and low stem leakage, which enables them to more easily meet the Federal Clean Act. Their leakage is high, unless special soft-seated configurations or high-performance designs are used. Some of them can also be provided with near-equal-percentage characteristics and tight shut-off.

When used for throttling service, some of their disadvantages are a direct consequence of the above advantages. Their high-capacity design results in either using oversized valves or having small valves mounted in large pipes. If small valves are used, this means substantial waste of pumping energy caused by the reducer pressure drops. Also, their high-pressure recovery nature results in low vena contract pressures, which in turn increase the probability of cavitation and noise.

Reduction of these problems has been attempted by adding flutes to the butterfly disc, but the problems have not been fully resolved. In operating rotary valves, the linear movement of cylinder or spring/diaphragm actuators must be converted by linkages, which introduces hysteresis and dead

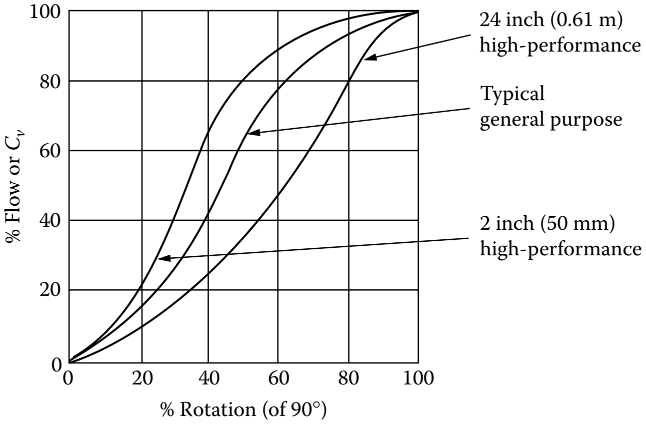


FIG. 6.17a
The flow characteristics of butterfly valves are affected by the location of the shaft (aligned or eccentric) and by the relative size of the shaft compared to the valve. The characteristics of high-performance designs are also slightly affected if the shaft is moved from the upstream to the downstream side of the disc. For throttling purposes the rotation of the valve is usually limited to move between the 0° and 60° positions.

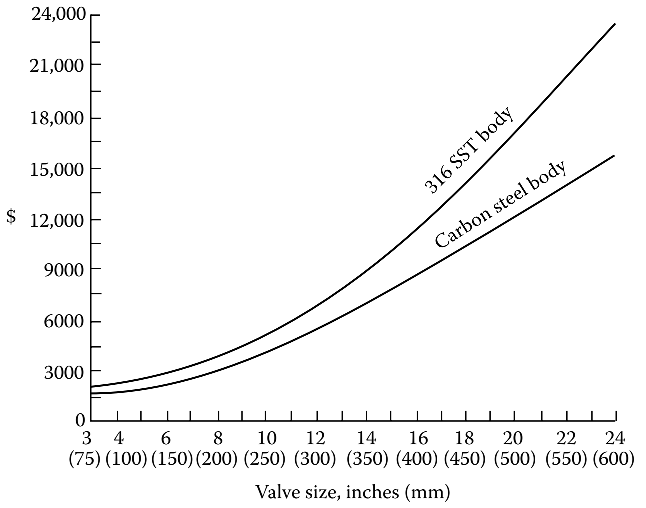
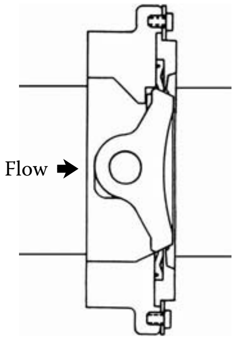
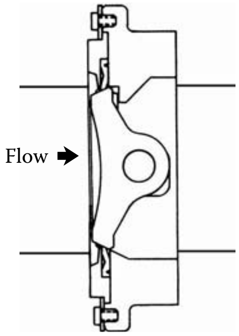


FIG. 6.17b
Approximate costs of high-performance butterfly valves provided with throttling actuators and positioners. Cost data are based upon standard high-performance, eccentric disc, soft-seat valves with double-acting piston operator and positioner. Carbon steel bodies up through 8 in. (200 mm) size have 316 SST disc and 17-4PH shaft. Above 8 in. size, discs are chrome-plated carbon steel. Stainless steel bodies have 316 SST disc with 17-4PH shaft in all sizes. Other operators can be furnished. Other alloys are available at higher cost.



	C_v at degrees open								
Size	90	80	70	60	50	40	30	20	10
2.0	59	59	58	56	50	40	28	14	2
3.0	220	209	198	176	139	90	55	26	11
4.0	420	400	376	338	265	172	104	52	20
6.0	910	800	660	490	350	235	155	90	35
8.0	1720	1620	1290	998	740	482	310	172	69
10.0	2780	2610	2080	1610	1200	778	500	278	111
12.0	4000	3820	3100	2420	1860	1240	750	410	170
14.0	6640	6240	5050	3980	2920	1990	1200	664	266
16.0	8400	7640	6130	4700	3700	2520	1510	840	336
18.0	10350	9730	7870	6210	4550	3100	1860	1040	414
20.0	13670	12850	10390	8200	6020	4100	2460	1370	547
24.0	20200	19000	15400	12100	8890	6060	3649	2020	808



	C_v at degrees open								
Size	90	80	70	60	50	40	30	20	10
2.0	55	55	54	52	48	40	27	12	3
3.0	185	174	157	135	109	88	67	41	15
4.0	330	310	282	240	194	158	118	74	25
6.0	840	760	620	450	320	225	160	100	45
8.0	1620	1470	1180	891	664	470	308	194	81
10.0	2640	2400	1930	1450	1080	766	502	317	132
12.0	3860	3560	2820	2200	1700	1140	750	430	200
14.0	6380	5810	4660	3570	2680	1850	1210	766	319
16.0	8070	7340	5890	4520	3390	2340	1530	968	404
18.0	9950	9060	7260	5570	4180	2890	1890	1190	498
20.0	13300	12100	9690	7430	5570	3850	2520	1590	664
24.0	19600	17900	14300	11000	8240	5690	3730	2350	981

FIG. 6.17c

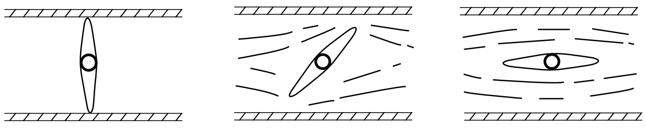
Listed in the tables are the valve capacity coefficients of high-performance butterfly valves at various degrees of opening. The C_v is different if the shaft of the disc is upstream (top) or downstream of the disc (bottom). (Courtesy of Valtek, Inc.)

play. In addition, a nonlinear relationship exists between actuator movement and the resulting rotation. These considerations make the use of positioners essential, which on fast processes can lower the quality of control.

The torque characteristics of these valves are also highly nonlinear (Figure 6.4v), and because of the high "break-torque" requirement, the actuator is oversized when operated in the throttling range. The characteristics of butterfly valves (Figure 6.17a) are somewhere between linear and quick-opening.

CONVENTIONAL BUTTERFLY VALVES

The butterfly valve is one of the oldest types of valves still in use. The dictionary defines the butterfly valve as a "damper or throttle valve in a pipe consisting of a disc turning on a diametral axis" (Figure 6.17d). Butterfly valves are not only used in industry, but variations are found in consumer products such as furnace dampers, automobile carburetors, and shower heads.



Closed
(Damper
perpendicular
to flow)

Throttling

Open
(Damper
parallel
to flow)

FIG. 6.17d

The vane positions of butterfly valve when closed, throttling, or open.

Wide use of butterfly valves dates back only to the 1920s, when improved designs resulted in their acceptance for public waterworks applications. The valve was particularly applicable to the low-pressure on/off service usually encountered in waterworks applications. Today's modern butterfly valve designs are suitable for a wider variety of fluid applications, including those with higher pressure drops, tight shut-off, and corrosive characteristics.

The straight-through design has a high capacity and has advantages when erosion is a consideration. Especially since the development of the eccentric shaft and disc designs known as the high-performance butterfly valve (HPBV), it is one of the fastest-growing segments of the valve industry.

Operation

Butterfly valve operation is basically simple, because it involves only rotating the vane, disc, louver, or flapper by means of the shaft to which it is fastened. This may be done manually by a lever handle on smaller valves or by a handwheel and rotary gear box on larger sizes. Automatic operation may be accomplished by pneumatic, hydraulic, or electrical motor drives attached to the shaft by various methods.

Unfortunately, some of these methods of attachment do not provide good valve control, and the design of the operator-to-shaft connections must be closely examined. For the connection, some manufacturers just use the same square shaft end and clamp with set-screw that is used on manual valves. This type of connection is very prone to "wear play," which is not acceptable for automatic control, where dynamically responsive valve operation is needed.

The proper way to make a suitable connection is by using a valve shaft with a spline end and a corresponding mating connection at the operator shaft. Linkages should be the "self aligning" ball end type for an overall "best" connection.

As the disc moves through a 90° rotation, the valve moves from fully closed to fully open (Figure 6.17d). The area open to flow increases as the disc rotates from closed to open, and this variation is used for throttling. The characteristic curve, which is a plot of the free area vs. percentage vane rotation, is shown in Figure 6.17a for a general-purpose butterfly valve and for some high performance designs.

However, modern valve operator positioners are available either with cams or signal conversion units that can be programmed to give almost any valve characteristic that is desired. For a detailed discussion of intelligent positioners, refer to Section 6.12.

Construction

Mechanically, butterfly valves vary widely in their construction features. However, common to all are the valve body, the disc and shaft, shaft support bushings or bearings, shaft packing, and a means of attaching an operator to the shaft. Butterfly valves also fall into two basic categories, swing-through and shut-off designs.

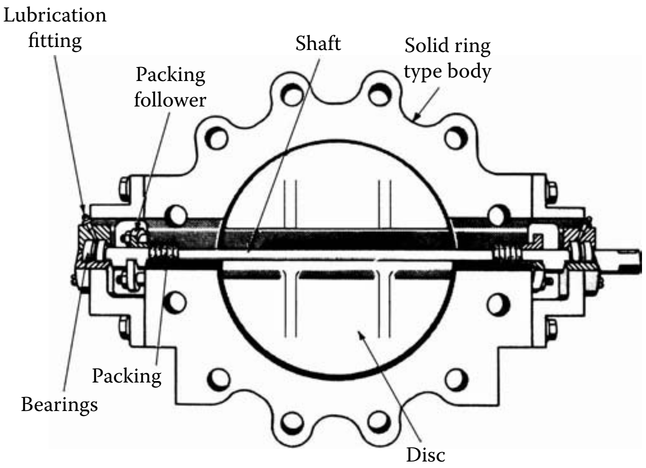


FIG. 6.17e
The design of a swing-through butterfly valve. (Courtesy of Emerson Process Management, Fisher Controls Company.)

Most swing-through designs (Figure 6.17e) have a symmetrical disc and shaft design with a certain clearance required between disc and body. The body is usually the solid ring type, which is mounted between pipe flanges. It can be either the wafer type or the single flange lug pattern, where the flange bolting also goes through the valve body.

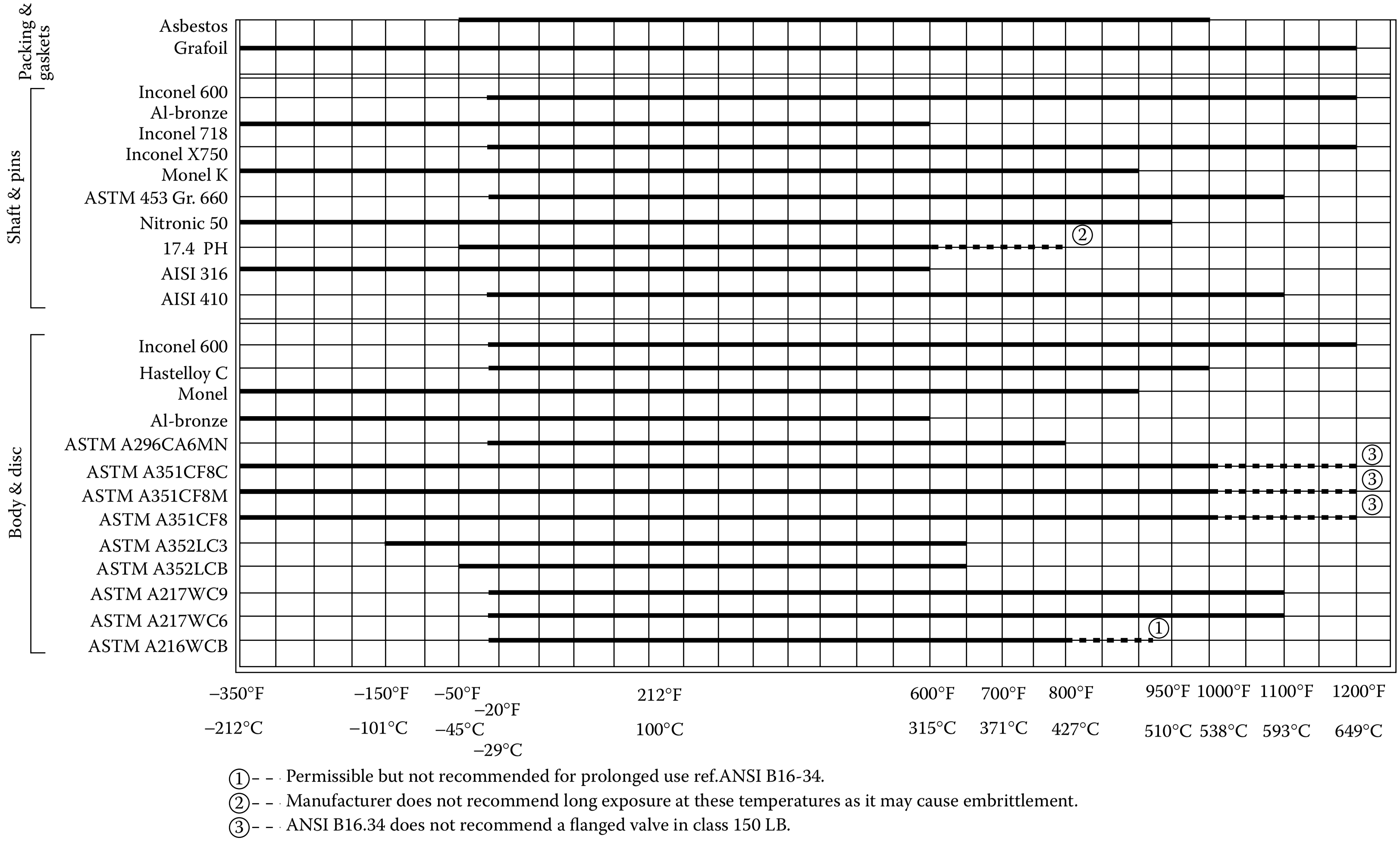
Discs are cast in one piece. The thickness of the disc and hub along with the diameter of the shaft is a function of the maximum pressure drop and torque required. Careful alignment of the body, bushings, shaft, and disc eliminates binding. Hard facing materials can be applied to the disc edge and body bore where erosive fluids such as steam are involved. Refractory-type linings are also available for the body. The operating temperature ranges of the various materials used in the construction of butterfly valves are given in Figure 6.17f.

The swing-through butterfly valve designs are available with a variety of disc shapes that serve to reduce the required torque and to increase throttling angle range (Figure 6.17g).

Swing-through butterfly valves are normally limited to a maximum throttling of angle about 70° open for the standard patterns and 60° open for the heavy patterns, due to their larger diameter shafts (Figure 6.17h). This is because the disc profile projection tends to "disappear" into the shaft area as the valve opens.

HIGH-PERFORMANCE BUTTERFLY VALVES

The most significant design advance in butterfly valves was the development of the high-performance butterfly valve. This design concept combined the tight shut-off of the lined valves, the reduced operating torque and excellent throttling capabilities of the swing-through disc shapes, and the ability to operate with relatively high pressure drops.



①- - Permissible but not recommended for prolonged use ref.ANSI B16-34.
②- - Manufacturer does not recommend long exposure at these temperatures as it may cause embrittlement.
③- - ANSI B16.34 does not recommend a flanged valve in class 150 LB.

FIG. 6.17f
Operating temperature ranges of various materials used in the construction of high-performance butterfly valves.

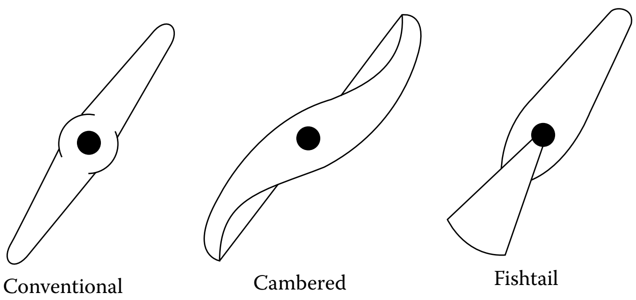


FIG. 6.17g
Cambered and fishtail disc shapes are used to reduce the required torque and to increase the throttling angle range.

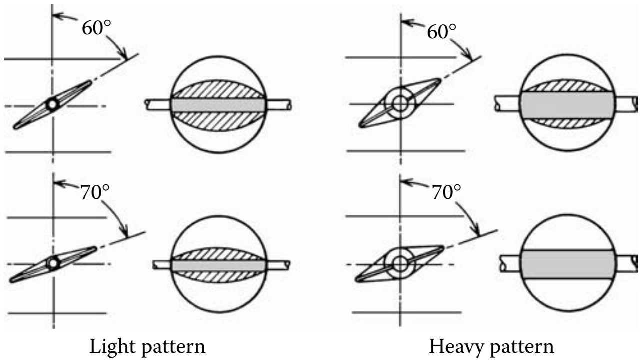


FIG. 6.17h
Effect of design pattern and shaft diameter on the flow area of butterfly valves.

The compact size, reduced weight, and lower cost have made the HPBV a formidable competitor to other control valve designs in sizes 3 in. (75 mm) and larger. There are many designs available, which usually all have the characteristics of 1) a separable seat ring contained in the body and 2) an eccentric cammed disc (Figure 6.17i). This camming action enables the disc to back out of and into the seat before and after the disc rotation when throttling. This is accomplished by having the shaft offset from both the centerlines of the disc and the valve body.

Tight Shut-off Designs

Special shut-off seals, such as piston rings for high temperatures to 1500°F (816°C) and T-ring seals for tight shut-off, are available (Figure 6.17j). These seals are not as popular as they used to be before the development of the HPBV valves and other designs with improved high-temperature seals.

Butterfly valves designed for tight shut-off fall into two categories. One is the valve that is provided with an elastomer or plastic liner. In this configuration, the disc is also encapsulated in some cases (Figure 6.17k).

The other tight shut-off design is the HPBV with the cammed disc and a separate seal ring clamped into the body (Figure 6.17l).

In addition, there are some special designs with laminated seal rings located on the disc edge that wedge into a conical seat in the valve body (Figure 6.17m). These laminated seal designs are especially suitable for high pressure and temperature shut-off.

Some time back, the lined butterfly valves were the only butterfly valves designed for tight shut-off. For their lining,

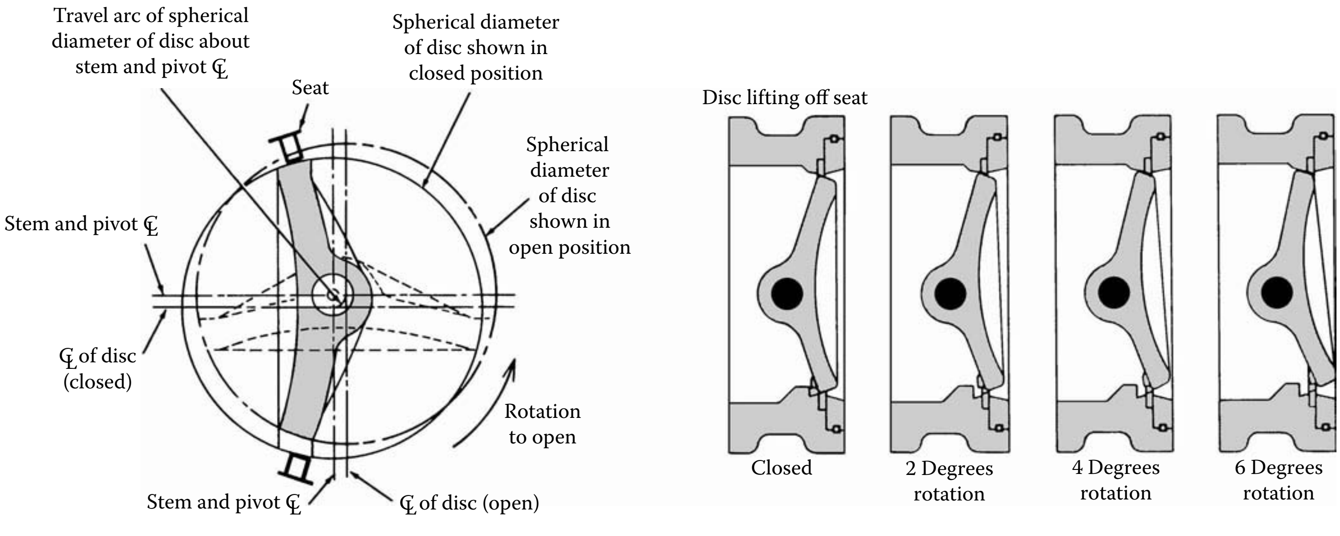


FIG. 6.17i
The high-performance butterfly valve with eccentric shaft and cam action disc operation is shown on the left. On the right it is illustrated how the disc is lifting off the seat as the valve begins to open. (Courtesy of Flowserve-Valtek, Inc.)

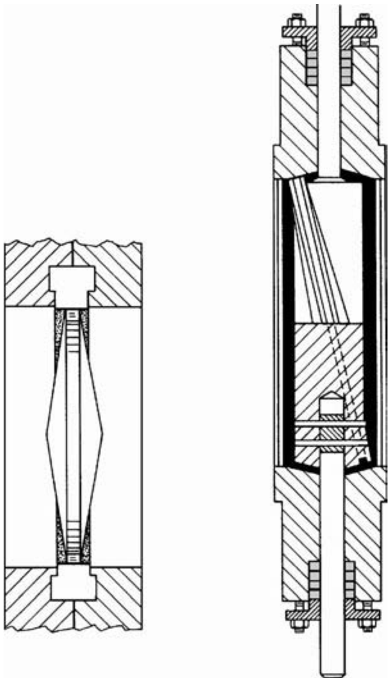


FIG. 6.17j
Special butterfly seal designs for tight shut-off. (Courtesy of Fisher Controls Company.)

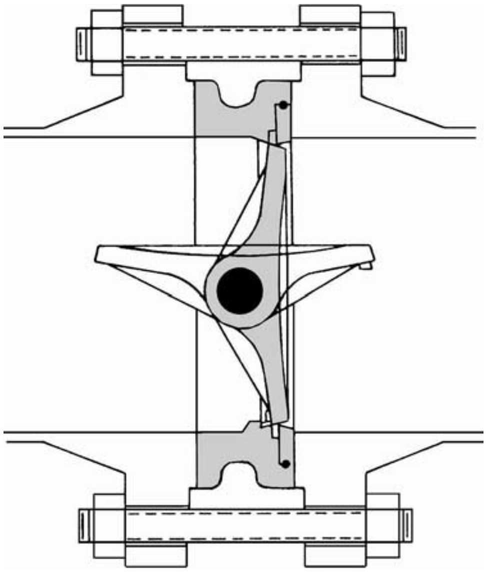


FIG. 6.17l
High-performance butterfly valve design provided with cammed disc. (Courtesy of Flowserve-Valtek, Inc.)

various elastomer materials were used, with a more rigid backup ring, which completely lined the bore of the valve and the valve gasket face area. Some liners can be bonded to the body and others are removable. Plastic liners, such as Teflon, are suitable for corrosive applications. In some cases the disc is also encapsulated in the elastomer or Teflon.

Sealing in these valves is usually accomplished by a wedging action of the disc edge into the elastomeric or plastic seat. The discs may be symmetrical on the shaft (similar to swing-through), offset from the shaft, or canted on the shaft. The objective of the latter two designs is to give a 360° seal contact on the disc edge.

Caution must be exercised in the selection of elastomers because they may be subject to attack by the process fluid. This attack may result in softening, swelling, cracking, or other effects. Plastic liners are not immune to these problems if the process fluid seeps past liner shaft seals and attacks the backing material or the body metal. By their very nature, these materials are also temperature limiting and can seldom exceed 350°F (177°C).

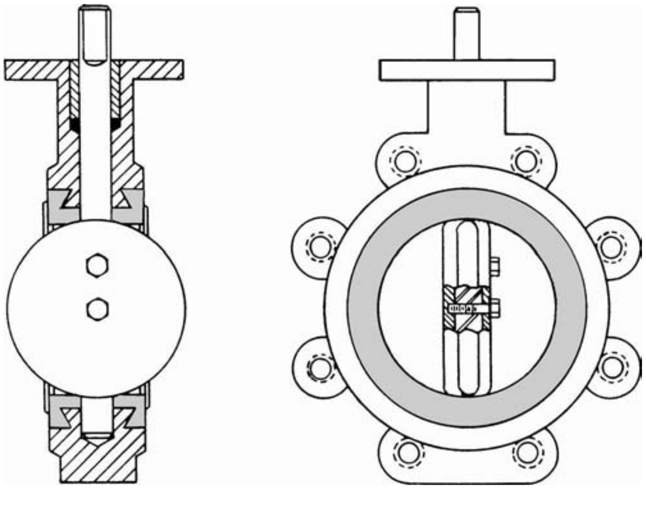


FIG. 6.17k
Lined butterfly valve design. (Courtesy of Keystone International, Inc.)

Leakage Ratings

The seat retainer rings are fastened to the body by various means. The most common method is by countersunk screws or bolts. Usually these fasteners are within the gasket area, and it is advisable to evaluate this interference in the gasket area for the intended service. Normally, this does not pose a problem, but it may require special gaskets or a change in gasket type to effect a proper seal. Other methods, such as a snap ring, friction fit, or retaining pins, do not intrude into the gasket area and may offer a better choice in some applications.

Seats are commonly made of plastic materials such as Teflon or various elastomers. Each manufacturer has a specific idea of how this seat seal should be designed and configured. These designs range from very simple to very complex shapes, and in some cases may incorporate elastomers behind the Teflon to effect a pressure-energized seal.

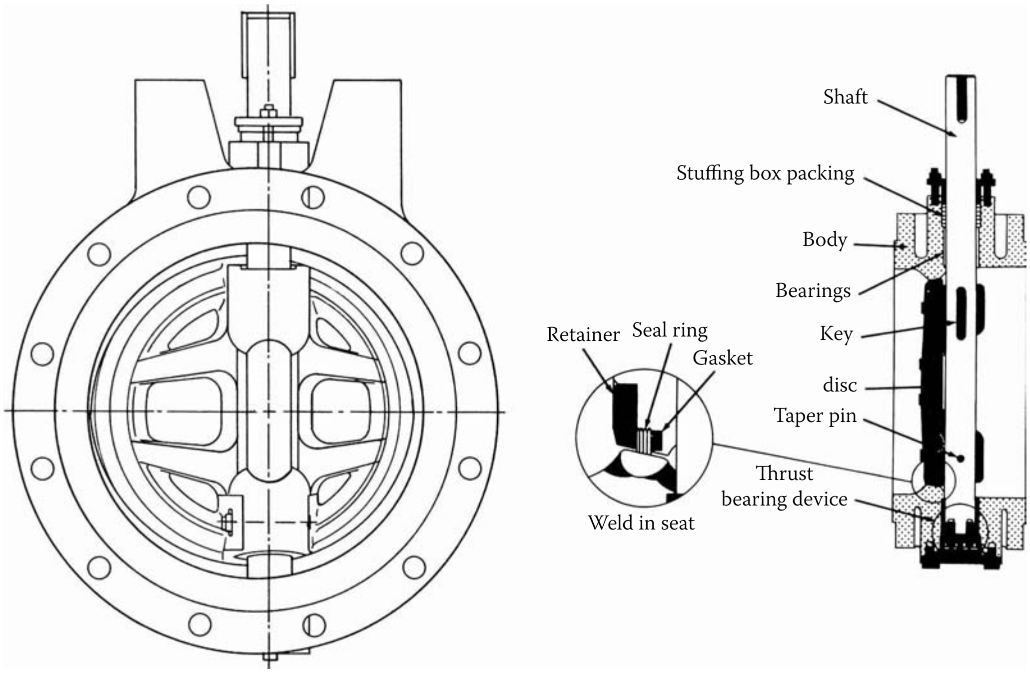


FIG. 6.17m
Eccentric disc butterfly with laminated disc seal ring. (Courtesy of Tyco-Vanessa, Inc.)

Basically, these various seat designs are all classified as ANSI Class VI, which is commonly construed to be "bubble-tight." Where temperatures are too high for the soft seat materials, metal seats are available that can also provide excellent shut-off equivalent to ANSI Class IV. Some metal seat designs can even approach ANSI Class V. Table 6.17n gives the seat leakage options for a particular HPBV design.

Fire-Safe Designs

A special design variation known as a "fire-safe" seat is also available from many manufacturers. This combines the soft seat discussed above with a backup metal seat. If the fire destroys the soft seat, the metal seat will serve to minimize leakage (Figure 6.17o). These designs also incorporate special fire-resistant gaskets and stem packing to minimize external leakage during and after a fire.

TABLE 6.17n
*Seat Leakage Choices in a High-Performance Butterfly Valve**

Type of Seat	*Leakage*
Metal seat	ANSI Class IV
Jam-lever toggle soft seat	ANSI Class VI
Flow ring	2 % of rated C_v
Dual seat	ANSI Class IV

*Courtesy of Flowserve-Valtek, Inc.

TORQUE CHARACTERISTICS

Operating torque requirements of butterfly valves require more careful consideration than do any other types of control valves (Figure 6.4v). The disc acts much like an airfoil or the wing of an aircraft. However, the special disc shapes and the HPBV designs already discussed have much lower torque requirements due to the shape effects, much like "spoilers" on an airfoil.

The conventional symmetrical disc behavior as an airfoil results in different pressure distributions around the face of the disc, producing a torque that tends to close the valve. Only at 0 and 90° are the pressures equal on both sides of the disc. Between 0 and 90° the thrust load of the disc wing turned toward the upstream side is larger than that on the downstream side. This is called the unbalanced, or hydraulic, torque, and its magnitude is a function of the pressure drop and disc diameter.

The valve operator must have enough power to overcome the unbalanced torque and, in addition, overcome the friction of the bearings and packing on the valve shaft. The total is known as combined torque. The combined torque required to open the valve is larger than that required to close the valve because the unbalanced torque helps to close the valve. This difference is known as torque hysteresis.

In fact, when the symmetrical disc closes, the torque may become negative somewhere around 30–35° from the closed position, and the valve will tend to close itself. The torque characteristics also indicate that at about a 75–80° opening, the torques for both opening and closing maximize. Above this rotation angle, the torque falls quickly to zero. Thus, the

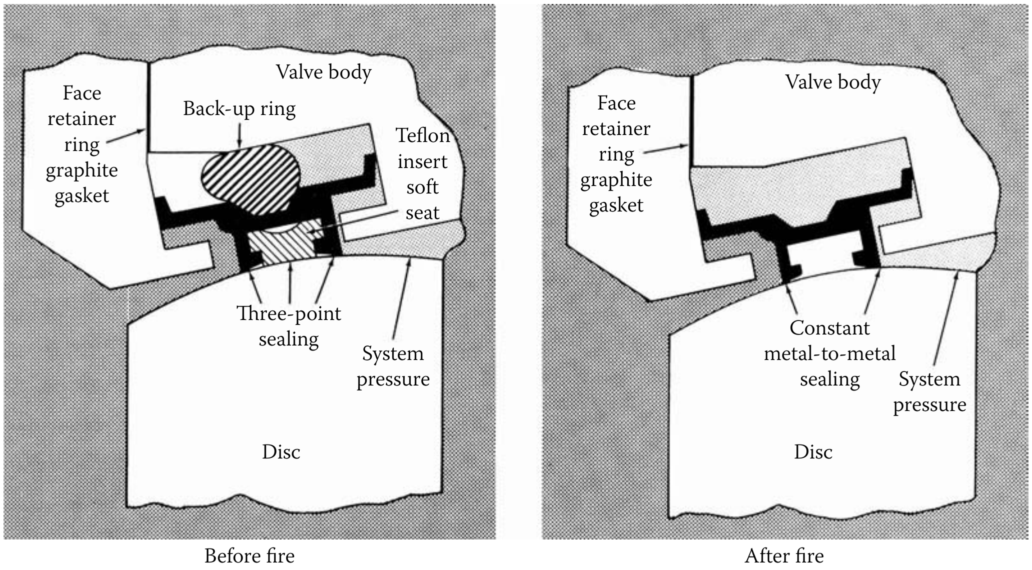
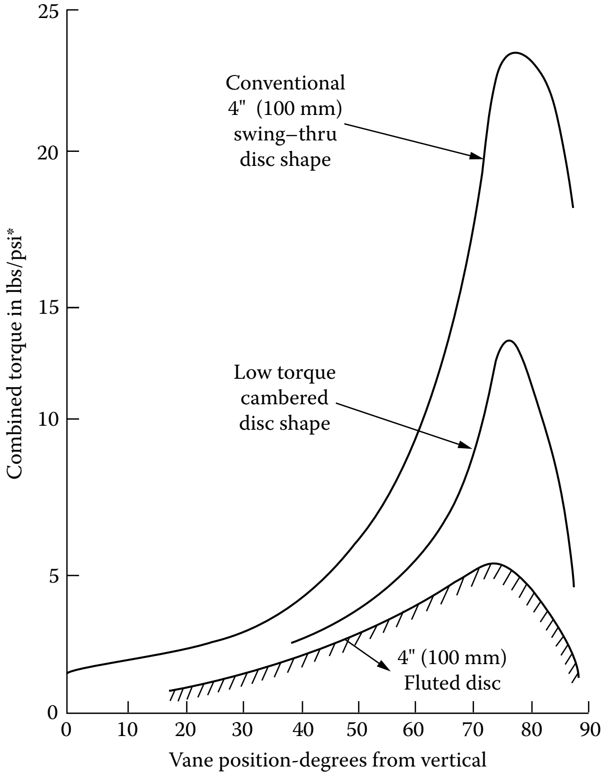


FIG. 6.17o
One type of "fire-safe" disc seal seat design.



*See section A.I for SI units

FIG. 6.17p
Combined torque requirements (sum of shaft friction and unbalanced torque) of a number of butterfly disc shapes at various degrees of rotation. (Courtesy of H. D. Baumann Co.)

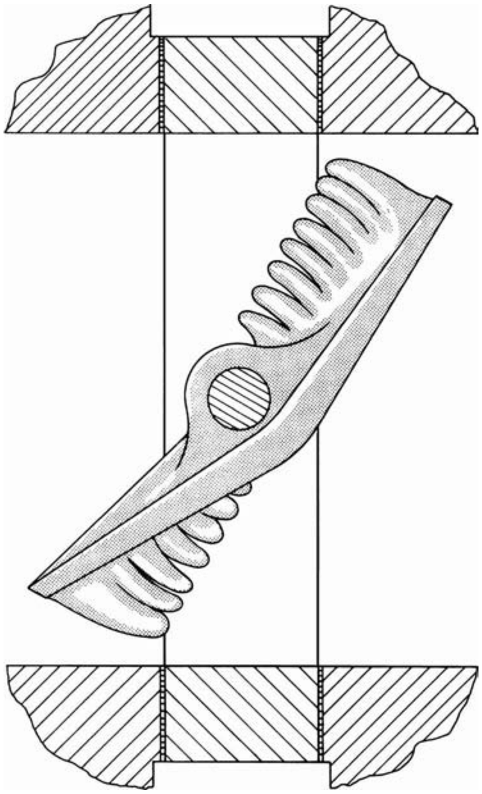


FIG. 6.17q
Substantial noise reduction can be obtained by the fluted butterfly disc design.

torque characteristics are highly nonlinear. They pose a considerable burden on the valve automatic operator, because it must cope with sharp increases and decreases in torque as well as positive and negative forces.

Typical torque curves for conventional symmetrical, special shape, and fluted discs are shown in Figure 6.17p. It should be noted that while HPBV disc designs can be considered a reduced torque disc, the amount of torque and its characteristic behavior curve is a function of whether the valve is installed with the shaft upstream or downstream. Flow tends to open the valve with the shaft downstream and tends to close the valve with the shaft upstream. However, with the shaft downstream, dynamic torques with the disc open are much lower. On gas service, the shaft can be in either location, but on liquid service the shaft should be downstream only because of the effects of liquid inertial forces.

Lined butterfly valves are not only subject to the above torque considerations but also encounter the additional torque required to seat and unseat the disc into the liner. The manufacturers usually rate the torque requirements of the various butterfly valve designs, because they are subject to so many variables. The manufacturer's recommendations should be followed for selecting the appropriate operator size required to operate the particular valve. It is best to be conservative when sizing butterfly valve operators, because operating and seating torques can often be greater than predicted.

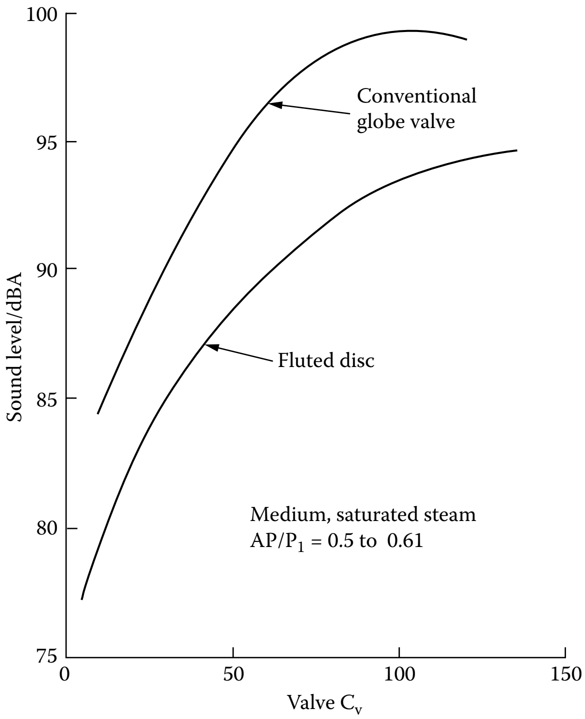


FIG. 6.17r
Noise reduction produced by a fluted disc in comparison to the noise level produced by a conventional globe valve.

NOISE SUPPRESSION

Butterfly valves will generate noise, as will any other valve when throttled at high flow rate and pressure drop, as is discussed in full detail in Section 6.14. The noise characteristics of some special disc designs, such as the cambered and the fishtail designs shown in Figure 6.17g, are improved in comparison to the conventional swing-through disc shape designs. However, with higher mass flows and pressure drops the generated noise can still be substantial.

A newer design development for the butterfly is the fluted disc shown in Figure 6.17q.

For compressible fluid applications, the fluted disc design is capable of delivering noise reductions of up to 10 dBs on the A-weighted scale (dBA), as shown in Figure 6.17r. In addition, the airfoil "spoiler" effect of the flutes enables this disc to have the lowest operating torque requirements of any disc design. It can be provided for valve sizes 2 in. (25 mm) through 16 in. (400 mm).

The fluted disc design is more expensive than others but is a very useful alternate where needed. The cost of manufacturing the noise-reduction spoilers has been somewhat reduced by relocating the flutes, as shown in Figure 6.17s.

The addition of swing-through vanes to the butterfly disc not only reduce noise emissions and cavitation, but it also increases valve rangeability and improves valve characteristics (Figure 6.17t). The valve characteristics can be changed by varying the area, spacing, and number of the flow orifices. It is possible to provide such spacing of the flow orifices that the two semispherical contours will generate near-equal-percentage characteristics. The addition of a diffuser pack (shown in Figure 6.17t) brings the characteristics closer to linear.

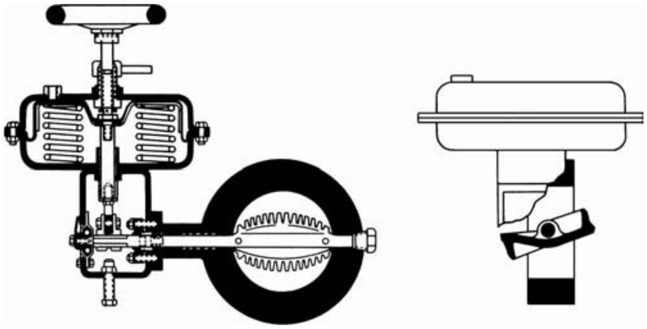


FIG. 6.17s
Flutes can be added to the butterfly disc to act as noise-reduction spoilers. (Courtesy of Flowserve-Valtek, Inc.)

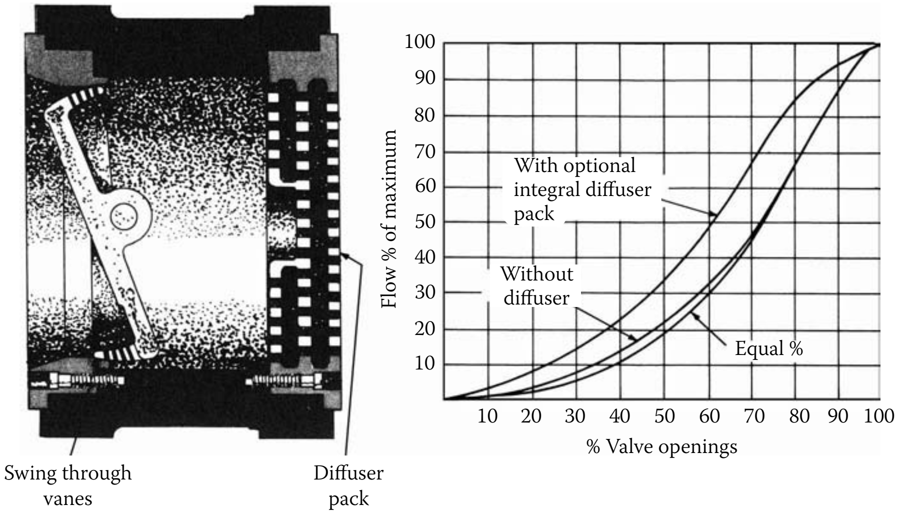


FIG. 6.17t

The addition of perforated swing-through vanes not only reduces noise and cavitation but also improves valve characteristics. (Courtesy of ABB Kent.)

Bibliography

Ball, K. E., "Final Elements: Final Frontier," *InTech,* November 1986.

Baumann, H. D., "The Case for Butterfly Valves in Throttling Applications," *Instruments and Control Systems,* May 1979.

Baumann, H. D., "Trend in Control Valves and Actuators," *Instruments and Control Systems,* November 1975.

Boger, H. W., "Low-Torque Butterfly Valve Design," *Instrumentation Technology,* September 1970.

Cain, F. M., "Solving the Problem of Cavitation in Control Valves," Paper #91-0462, 1991 ISA Conference, Anaheim, CA, October 1992.

Carey, J. A., "Control Valve Update," *Instruments and Control Systems,* January 1981.

Daneher, J. R., "Sizing Butterfly Valves," *Water & Wastes Engineering,* July and September 1970.

Dobrowolski, M., "Guide to Selecting Rotary Control Valves," *Instrumentation Technology,* December 1981.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems,* January 1980.

Hammitt, D., "Rotary Valves for Throttling," *Instruments and Control Systems,* July 1977.

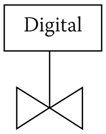
Page, G. W., "Predict Control Valve Noise," *Chem. Eng.,* New York 107(9): 23–26 August 2000.

Passage, D., "Butterfly Control Valves," *Instruments and Control Systems,* March 1964.

Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 ISA Conference, Houston, TX, October 1992.

Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

6.18 Valve Types: Digital Valves



Flow sheet symbol

D. R. A. JONES (1985) **B. G. LIPTÁK** (1995, 2005)

Size: $^3/_4$ to 10 in. (19 to 250 mm) in-line and angle pattern

Design Temperature Range: Cryogenic to 1250°F (677°C)

Design Pressure Limits: Up to 10,000 PSIG (690 bars)

Capacity: $C_v = 13\ d^2$

Applications:
Where high speed is required, as in surge control
Where accurate flow control is needed, such as in provers or in the blending of expensive ingredients
Where great rangeability is needed, such as in pH control
Where flow is to be both controlled and accurately measured, such as in natural gas regulator stations at high-pressure drops or in power plants
Where tight shutoff is needed

Rangeability:

No. of "bits"	8	10	12	14	16
Resolution	255:1	1023:1	4095:1	16,383:1	65,535:1

Speed: 25 to 100 ms

Characteristics: Unlimited

Leakage: ANSI V

Materials of Construction:
Body—Aluminum, carbon steel, stainless steel, titanium
Seals—Buna, rubber, Viton, TFE, Kel-F, Derlin, Hypalon, and graphite

Cost: Typical prices for ANSI Class 600 carbon steel body, 12-bit valves are \$5000 for $^3/_4$ in. (20 mm), \$5000 to \$10,000 for 1, $1^1/_2$, and 2 in. (25, 32, 40, and 50 mm), \$12,500 for 3 in. (75 mm), \$15,000 for 4 in. (100 mm), \$20,000 for 6 in. (150 mm), and \$30,000 for 8 in. (200 mm). For 316 SST bodies, add \$1000 for $^3/_4$ in., \$2000 for 1, $1^1/_2$, and 2 in., \$3000 for 3 in., \$5000 for 4 in., \$7000 for 6 in., and \$14,000 for 8 in.

Partial List of Suppliers:
ABB Kent, Introl Valves Div. (www.abb.com)
Emco-Digital Valve (www.emcoflow. com/digitalvalve.htm)
Emerson-Daniel (www.emersonprocess.com/daniel)
Herion Inc. (www.herionusa.com)
Hoke Inc. (www.hoke.com)
Instrutech Inc. (www. instrutechinc.com)

INTRODUCTION

These days, when one refers to digital control valves, people think of intelligent control valves that are provided with fieldbus interaction capability (Figure 6.18a). Fieldbus interaction is not the topic of this section; it is covered in Section 6.11. The digital control valves discussed in this section (while they can also be operated by digital networks) are multiported valves, with the number of ports ranging from 8 to 16.

A digital valve contains a group of valve elements assembled into a common manifold. The elements have a binary relationship to each other; i.e., starting with the smallest, each

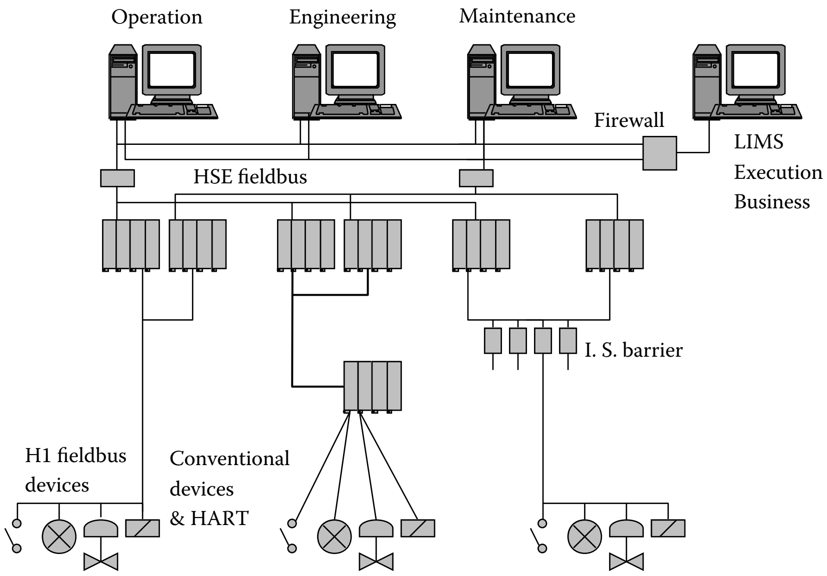


FIG. 6.18a
Networked system architecture.

increasing size element is twice as large as its next smallest neighbor (Figure 6.18b). Each element is controlled by an individual electric or electronic signal. Thus, an 8-bit digital valve requires 8 parallel, on/off electric (or electronic) signals, a 12-bit digital valve will require 12 parallel signals, and a 16-bit digital valve will require 16 parallel signals.

The main advantages of digital control valves are their speed, high precision, and practically unlimited rangeability. Their main disadvantages are their high cost and their suitability for only clean services, because the smaller ports plug very easily.

HISTORY

In the early digital control valves, the ports were distributed circumferentially, as shown in Figure 6.18c. These designs were difficult to manufacture. In the later designs the circumferential distribution of ports was replaced by top-entry, vertical ports.

Balanced Piston Digital Control

Balanced piston type designs (Figure 6.18d) are also called digital control valves, because their piston-operated main inner valve is positioned by the opening or closing of two solenoid pilots. Their advantages include their positive shutoff and fail-safe nature, as they close on loss of electrical failure. These valves provide precise flow rate and batch controls, including preprogrammed low flow start-up, high-rate delivery, and low-flow shutdown.

Top-Entry Design

An 8-bit digital valve throttles the flow by controlling the openings of eight flow elements in the valve body (Figure 8.18e) and provides a flow rangeability of 255:1. This body, an in-line, top-entry design, includes an air reservoir to ensure the fail-safe operation of the cylinder actuators. The design includes adequate manifold area to ensure consistent performance of the individual elements. The manifolds are large enough to minimize the possibility of cavitation and resulting erosion.

Each element in the array is on/off. Flow throttling is accomplished by opening enough ports to provide the exact flow area required by the controller output signal. There is a 1:1 relationship between the binary weighted signal and the binary weighted flow area. Figure 6.18b illustrates schematically the size relationship between binary elements in a digital valve.

Applications The main applications of digital valves are ones where speed, rangeability, and precision are critical. Such applications include the accurate blending and batching of both gases and liquids. The high-speed operation of compressor surge controls and all flow control applications, where the process fluid is clean and the rangeability of the

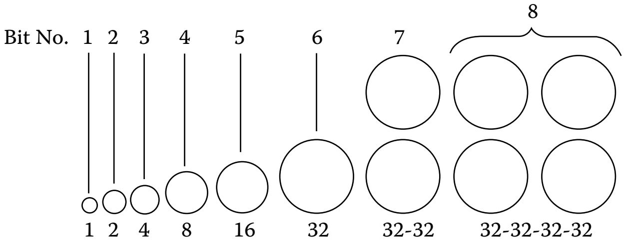


FIG. 6.18b
An 8-bit digital valve is provided with a balanced distribution of ports and guarantees a resolution of 255:1, which corresponds to a flow control accuracy within 0.39% of the total valve C_v. (Courtesy of Emco-Digital Valve.)

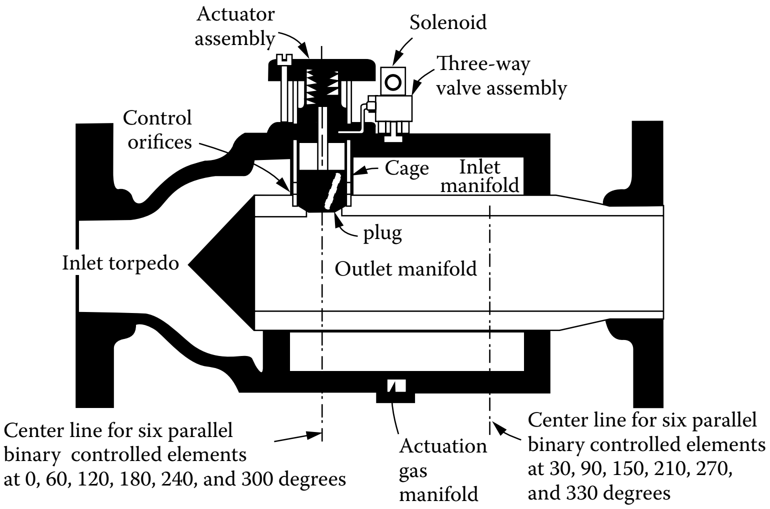


FIG. 6.18c
Early digital control valve design with circumferentially distributed ports.[1]

flow exceeds the capabilities of conventional valves. Such applications include but are not limited to environmental chamber controls, aircraft pressure cycling, and the operation of both liquid and gas provers.

Resolution and Throttling Because a binary array of 12 bits will provide resolution of 1 in more than 4000 (the area of the smallest flow element in a 12-bit digital valve is less than 0.025% of the total flow area), it is essential that the digital valve be leak-tight (Figure 6.18f). Leakage rates that are acceptable in standard globe valves would make the above-mentioned resolution impossible to achieve. It is for these reasons that digital valves are commonly manufactured to extremely high leak-tight standards.

In order to smoothly change the valve opening, say between 49% and 50%, it is usual to use two 25% elements instead of one 50% element for the largest bit. Thus, an 8-bit digital valve would have nine elements—three 25%, one 12.5%, one 6.25%, one 3.125%, one 1.56%, one 0.78%, and one 0.39% element, as was shown in Figure 6.18b.

FIG. 6.18d
Balanced piston operated digital control valve. (Courtesy of Emerson Process Management–Daniel.)

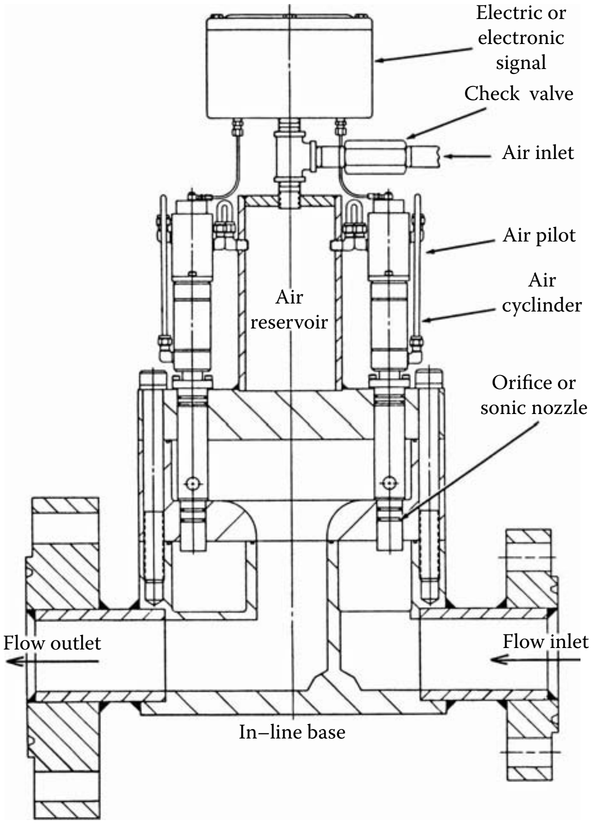


FIG. 6.18e
An 8-bit, 3 inch, carbon steel digital control valve designed for high pressure service (3000 ANSI).

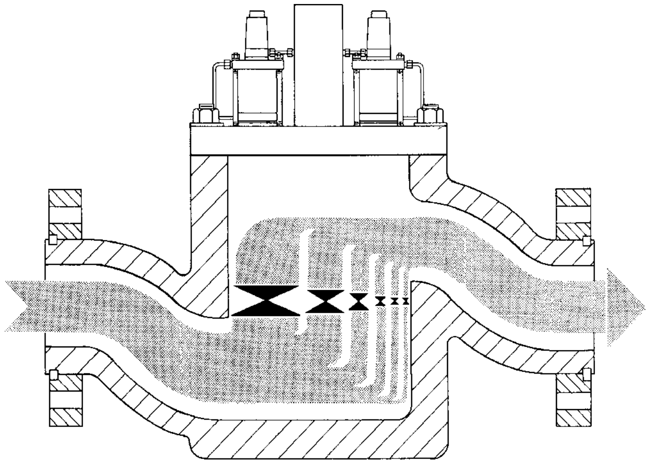


FIG. 6.18f

In a digital valve, the area of each valve port is half that of the previous port.

In higher-resolution digital valves, multiple elements are common; i.e., a 12-bit digital valve could have 16 elements, and a 14-bit digital valve could have 18 elements. The largest element then would be 12.5%. Each such valve would have seven 12.5% elements.

The 6 in. (150 mm) cast steel body digital valve (Figure 6.18g) uses 12 elements to provide 8-bit performance. The binary series of control element capacities is 1, 2, 4, 8, 16, 32-32, 32–32, 32–32. In this way no individual element handles more than 12.5% of the total flow. The largest bit (50%) is handled by four elements spread out around the body.

This arrangement improves redundancy in operation and uniformity in individual element sizing. Noise generation is reduced by breaking the flow up into many flow paths. The 8-bit computer word remains the same but operates 12 control elements.

This method of breaking the flow into many streams has a number of advantages. Each element actuator (electric, hydraulic, or pneumatic) is small and very quick. The multiple element arrangement provides redundancy, usually permitting operations to continue even if one or more elements are disabled. Flow calibration requirements are affected, because the calibration unit need be only $^1/_8$ as large to qualify 100% capacity.

Leakage, Ratings, and Speed Flow difference between individual elements is determined by the control orifice or the nozzle size. Each element consists of a plunger and a seat. The plunger is operated by a solenoid or solenoid-piloted

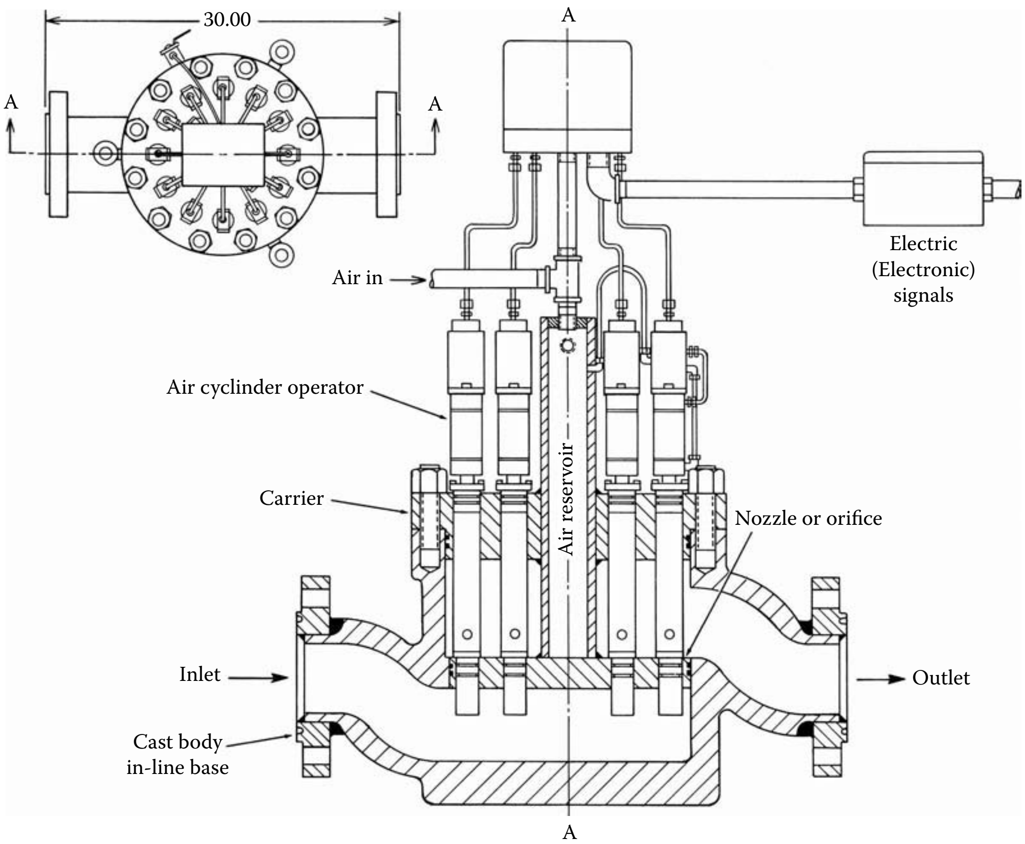


FIG. 6.18g

A 6 inch, 8-bit, explosion-proof cast steel (900 ANSI) digital valve with 12 elements. (Courtesy of Emco-Digital Valve.)

cylinder (pneumatic or hydraulic). Elements are usually dynamically balanced for low actuating force and are designed to assure tight shut-off.

Because of the necessary leak-tight construction, digital valves of very small capacity coefficients (C_v) are practical. Standard digital valve body sizes range from $^3/_4$ in. NPT (19 mm) through 10 in. (254 mm) pipe size. Pressure ratings up to 10,000 PSIG (68, 788 kPa) have been supplied.

Digital valves have been used effectively in cryogenic service, and special high-temperature models are under development (to 1250°F, or 677°C). Fluid temperatures above 450–500°F (232–260°C) require special seals and seal design.

Digital valves provide exactly repeatable performance, because each digital command causes the opening of a precisely defined port area. Nominal transfer time from any one position to any other position is usually under 100 ms, and the transfer time is uniform from one position to another.

The binary relationship between elements provides a linear increase of port area with a linearly increasing digital signal command. Any desired valve characteristic can be obtained by the correct programming of the digital control command.

FLOW METERING

A valve-flowmeter is a version of the digital valve that is equipped with flow sensing nozzles or orifices in each element (flow port).

Gas Flow

When gas is flowing through a "sonic venturi" and the inlet pressure is high enough to induce sonic flow at the vena contracta (throat) of the venturi, that condition is called "choked flow." Under such conditions, the rate of gas flow will only be dependent on the inlet conditions (i.e., absolute temperature and absolute pressure) of the flowing gas. When a "sonic venturi" is installed in each of the ports of a digital valve (Figure 6.18h), the "choked flow" at each port is independent of the downstream pressure.

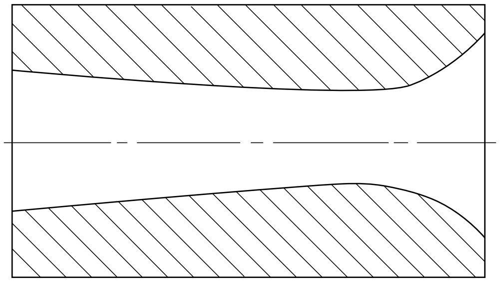


FIG. 6.18h
The sonic venturi element.

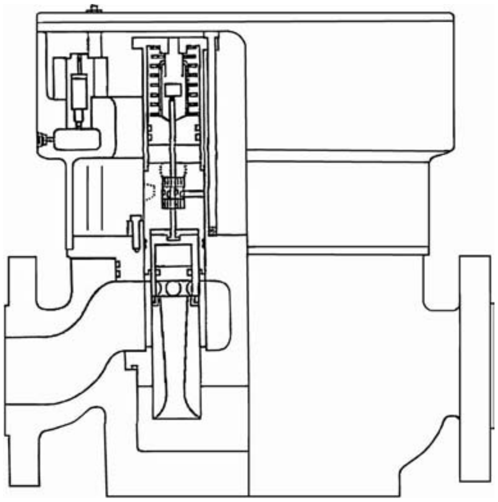


FIG. 6.18i
Extremely wide rangeability can be provided by using sonic venturi-type digital flowmeters.

As long as the downstream pressure is less than the critical pressure at the vena contracta, choked flow can usually be maintained with a pressure drop across the venturi of 15% of the inlet absolute pressure. The divergent portion of the sonic venturi recovers part of the velocity head. Thus, variations in the downstream pressure do not affect the flow rate.

The flow measurement is so accurate under these conditions that a digital valve-based flowmeter can be used as a transfer standard for calibrating other flow meters. An example of such installations is an 8 in. (200 mm) digital valve-flowmeter equipped with sonic venturis that has been installed in a power plant fuel line to act as a combination pressure regulator and fuel flowmeter (Figure 6.18i).

Liquid Flow

An orifice can be installed in each element, and the differential pressure across the digital valve can be used as a measure of the fluid flow through the open port area. Digital valves can be used as wide-range flowmeters by controlling the valve opening to maintain a constant pressure drop across the valve. Thus, a low-pressure drop-based measurement can provide a very wide flow range (up to 16,000:1 or more).

CONCLUSIONS

In summation, digital valves can provide high resolution, very fast response, exact repeatability, and very wide range. They can also be equipped with flow elements to provide both measurement and control of flow. Recommended applications are flow blending, compressor surge control, gas meter-regulators, transfer-standard flow provers, and precise liquid flow rate measurement and control. In fact they are appropriate any place where speed, accuracy, and high resolution are needed and the process fluid is clean.

Reference

1. Langill, A. W., "New Control Valve Accepts Digital Signals," *Control Engineering,* August 1969.

Bibliography

Alspach, W. J. and Maurer, G., "Consider Digital Valves," *Hydrocarbon Processing,* December 1991.

ANSI/ISA-75.25.01-2000, "Test Procedure for Control Valve Response Measurement from Step Inputs," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 2000.

Ball, K. E., "Final Elements: Final Frontier," *InTech,* November 1986.

Beeson, J. and Alspach, W. J., "How Sonic Nozzle Proving Works," *Gas Industries,* February 1990.

Britton, C. L., "Sonic Nozzles," Appalachian Gas Measurement Short Course, August 1978.

"Calibrating Control Valve Accuracy," *Mech Eng,* 122 (6): 33-33, June 2000.

Corey, J. A., "Control Valve Update," *Instrument and Control Systems,* January 1981.

Clark, H. L., "Turbine Meter Testing," AGA Transmission Measurement Conference, May 1979.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instrument and Control Systems,* January 1980.

Grumstrup, B., "Digital Valve Mounted Instrumentation—Opening a Window to the Process," ISA/94 Technical Conference, Anaheim, CA, 1994.

Harrold, D., "Calibrating Control Valves," *Control Engineering Europe,* Vol. 3, Issue 3, p. 43, June–July 2002.

Jones, D. R. A., "Digital Valves," AGA Transmission Measurement Conference, May 1977.

Langford, C. G., "A Method to Determine Control Valve Dynamic Requirements," Instrumentation, Systems, and Automation Society, 2002.

Miller, C. E. and Alspach, W. J., "Digital Valves and Digital Flow Technology," *Proceedings of the Instrumentation Symposium for the Process Industry,* Texas A&M University, 1991.

Miller, L., "Valve Diagnostic Past, Present, and Future," *Fluid Handling Systems,* November 2002.

Morris, W., "Digital Valve as a Transfer Standard," *Gas Magazine,* September 1980.

Page, G W., "Predict Control Valve Noise," *Chem Eng-New York* 107 (9): 23–26, August 2000 (Reprinted, July 1997).

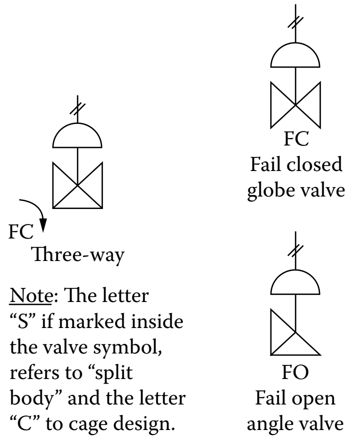
Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 ISA Conference, Houston, TX, October 1992.

Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

"'Smart' Actuator Incorporates all the Controls in One Package," Chemical Engineering Magazine, January 1, 2003.

6.19 Valve Types: Globe Valves

H. D. BAUMANN (1970) **J. B. ARANT** (1985)
B. G. LIPTÁK (1995) **F. M. CAIN** (2005)



Note: The letter "S" if marked inside the valve symbol, refers to "split body" and the letter "C" to cage design.

Flow sheet symbols

Types:	A. Single-ported with characterized plug B. Single-ported, cage-guided C. Single-ported, split body D. Double-ported, top-bottom-guided or skirt-guided plug E. Angle F. Y-type G. Three-way H. Eccentric plug, rotary globe
Sizes:	A. Typically NPS 1/2 to 16 in. (DN 15 to DN 400); available up to NPS 48 in. (DN 1200) B. Typically NPS 1/2 to 16 in. (DN 15 to DN 400); available up to NPS 48 in. (DN 1200) C. NPS 1/2 to 10 in. (DN 15 to 250) D. NPS 1/2 to 16 in. (DN 15 to DN 400) E. Typically NPS 1/2 to 16 in. (DN 15 to DN 400); available up to NPS 48 in. (DN 1200) F. NPS 1 to 16-in. (DN 25 to DN 400) G. Typically NPS 1/2 to 6 in. (DN 15 to DN 150); available up to NPS 24 in. (DN 600) H. NPS 1 to 12 in. (DN 25 to DN 300)
Design Pressure Ratings:	DN = diameter, nominal—mm assumed. Typically all ratings are available from ANSI Class 150 (PN 20) to Class 2500 (PN 420) with special designs up to Class 4500; types C and H are limited to ANSI Class 600 (PN 100) PN = Pressure, nominal—bar assumed.
Maximum Pressure Drop:	Up to maximum allowed by body pressure rating depending on limitations of actuator size and trim design and materials
Design Temperature:	Depends on material properties. Generally from –20 to 1200°F (–29 to 538°C). Cryogenic designs for temperatures down to –423°F (–253°C). Special valves have been designed for operation up to 1600°F (871°C).
Materials of Construction:	*Body and bonnet materials:* Most cast and forged grades of carbon steel, low-alloy steel, stainless steel, Alloy 20, duplex stainless steel, nickel and nickel alloys, bronze, titanium, and zirconium. See Table 6.19ww. Fluoropolymer lining also available for corrosion protection. *Trim materials:* Generally available in stainless steel, nickel, nickel alloys, bronze, titanium, and zirconium. Hard facing is available for erosive applications. See Table 6.19j. *Seal and soft seat materials:* FEP, PFA, PTFE, PCTFE, ETFE, EPT/EPDM, Fluoroelastomers, Nitrile, polyethylene, polyurethane, UHMWPE, compressed graphite, and soft metals.

Leakage:	(See Table 6.1gg for FCI leakage classes.) Metal seats in double-ported designs are Class II, while in single-seated designs they can meet Class IV or Class V. Soft seats in double-ported designs can meet ANSI Class IV or V, while in single-seated globe valves they can give Class VI performance.
Characteristics:	Refer to Section 6.7 for details; see Figure 6.19a.
Rangeability:	Based on Instrumentation, Systems, and Automation Society-75.11 or IEC 60534-2-4, it seldom exceeds 30:1. Special designs can achieve 50:1 or higher by increasing precision of control at small valve openings. See the discussion under Rangeability below and Section 6.7 for details.
Capacity:	C_v/d^2 = 10 to 15 with single-ported designs closer to the bottom of the range and with double-ported and eccentric disc designs closer to the top of the range (see Table 6.19c).
Cost:	See Figure 6.19b.
Partial List of Suppliers: *(Includes both manual and control valves)*	ABB Control Valves (www.abb.com/controlvalves) American Valve, Inc. (www.americanvalve.com) ARI-Armaturen Richter (www.ari-armaturen.com) Asahi-America (www.asahi-america.com) Cashco Inc. (www.cashco.com) Collins Instrument Co. (www.collinsinst.com) Control Components Inc. (www.ccivalve.com) Conval Inc. (www.conval.com) Crane Valves (www.cranevalve.com) Curtis Wright Flow Control (www.cwfc.com) Dresser Flow Solutions (www.masoneilan.com) Emerson Process Management (www.emersonprocess.com/home/products) Flowserve Corporation (www.flowserve.com/valves) GE-Nuovo Pignone (www.gepower.com/prod_serv/ index.htm) Invalco (www.fmcinvalco.com) Kitz Corp. (www.kitz.com) Koso Hammel Dahl (www.kosoamerica.com) Metso Automation (www.neles.com) Milwaukee Valve Co. (www.milwaukeevalve.com) Nibco Inc. (www.nibco.com) Powell Valves (www.powellvalves.com) Richards Industries Valve Group, Inc. (www.jordanvalve.com) Samson AG (www.samson.de) Severn Glocon Ltd. (www.severnglocon.com) Spirax Sarco, Inc. (www.spriaxsarco.com/us) SPX Valves and Controls (www.dezurik.com) Tyco Flow Control (www.tycovalves.com) Velan Valve Corp. (www.velan.com) Warren Controls Corporation (www.warrencontrols.com) Weir Valves & Controls (www.weirvalve.com) Welland & Tuxhorn (www.welland-tuxhorn.de) Yamatake Corp. (www.yamatake.com)

VALVE TRENDS

When this handbook was first published some 35 years ago, the overwhelming majority of throttling control valves were the globe types, characterized by linear plug movements and actuated by spring-and-diaphragm operators. At that time, the rotary valves were considered to be on/off shut-off devices. Globe valves are still widely used, but their dominance in throttling control applications has been diminished by the less expensive rotary (ball, butterfly, and plug) valves as a result of improvements in rotary valve and actuator designs.

Generally, globe valves use a linear-motion stem connected to a plug head that controls the flow area through a stationary seat ring. One exception to this is the rotary globe valve, which rotates an eccentric spherical plug into the seat ring; this type of rotary stem valve will be discussed later. Unless otherwise noted, the discussion of globe valve characteristics will apply to the linear-stem globe valve.

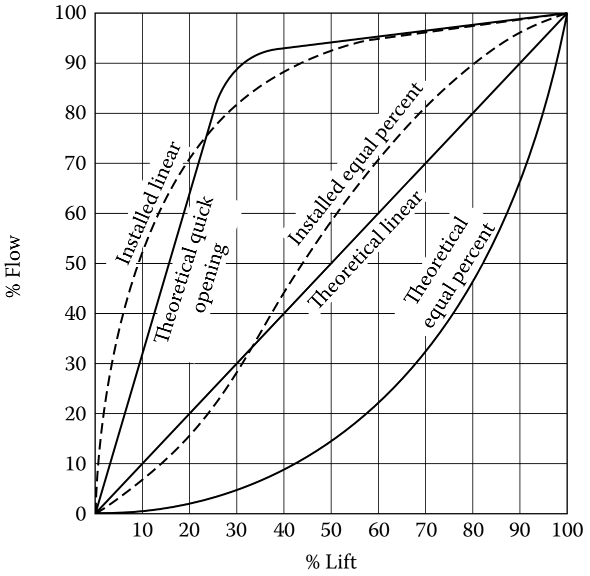


FIG. 6.19a
The theoretical valve characteristics shift as a function of installation. The dotted lines reflect such a shift in a mostly friction process where at 100% flow, 20% of the pressure drop was assigned to the control valve.

The main advantages of the traditional globe design include

1. The simplicity of the pneumatic actuator designs
2. The availability of a wide range of valve characteristics
3. The relatively low likelihood of cavitation and noise
4. The availability of a wide variety of specialized designs for corrosive, abrasive, and high-temperature or high- pressure applications
5. Relatively small amounts of dead band and hysteresis

The main reason for the increasing popularity of rotary valves is their lower manufacturing cost and higher capacity ($C_v/d^2 = 20–40$ for rotary vs. $C_v/d^2 = 10–15$ for globe). They generally weigh less, some designs can act as both control and shut-off valves, and they can be easier to seal at the stem to meet clean air requirements.

The limitations of globe valves, in addition to their higher cost per unit C_v, include their greater weight and dimensional envelope relative to their flow capacity. For the valve coefficients of globe valves, refer to Table 6.19c.

One major disadvantage of rotary valves is their higher tendency to cavitate and produce excessive amounts of noise (Section 6.14). They are also more likely, due to their smaller size per unit C_v, to have larger pipe reducers with the associated waste of pressure drop and distortion of characteristics. Their control quality can suffer from the linkages, which can introduce substantial hysteresis and dead band.

As a result of advances in distributed control system (DCS) technologies, both rotary valves and globe-style (linear) valves generally benefit from the use of positioners with either single-acting or double-acting pneumatic actuators (diaphragm or piston types). Previous problems with positioners in fast processes are largely relics of old-fashioned control systems, but much misinformation has perpetuated an aversion to positioner use on fast processes, even though DCS technology overcame these issues decades ago.

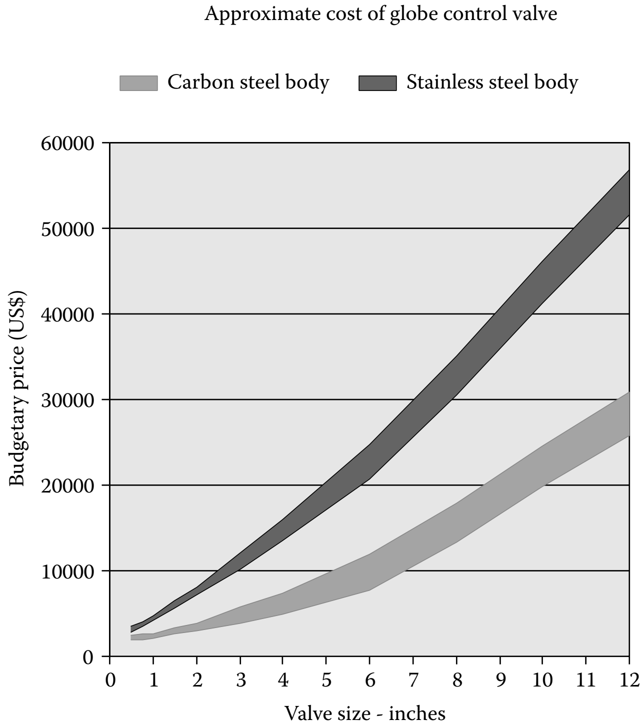


FIG. 6.19b
Approximate cost data are based upon typical globe control valves with Class 300 flanged bodies, double-acting piston actuator, and positioner with I/P transducer. As an example, Figure 6.19b would estimate the average cost of a 4 in. globe valve with a steel body and cage-type trim to be about $6000. Other actuators such as single-acting piston, spring-and-diaphragm, and electric motor drive are available. Bodies and trim are available in numerous metal alloys but usually at higher prices.

For more information about control theory in general and controller tuning in particular, consult Chapter 2, and about DCS systems refer to Chapter 4. For conventional and for intelligent positioners, refer to Sections 6.2 and 6.12 respectively, in this chapter.

TRIM DESIGNS

The valve trim consists of the internal parts contained within the body and wetted by the process fluid. The main components are the plug and stem and the seat ring(s). Some globe valve body designs also incorporate other parts such as cages or seat retainers, spacers, guide bushings, and special elements. The trim parts create the flow restriction or throttling action responsible for most of the pressure loss dissipated in the valve. The trim design also serves to determine the inherent flow characteristics of the valve. The various aspects of trim design, construction, and selection will be discussed.

TABLE 6.19c

Valve Coefficients (C_v) for Single-Ported, Equal-Percentage, Unbalanced Globe Valves with Flow under the Plug.[*]

Valve Size (in.)	Trim Size (in.)	Stroke (in.)	C_v At Percent Open										
			5	10	20	30	40	50	60	70	80	90	100
1	0.81	0.75	0.33	0.66	0.94	1.3	1.9	2.8	4.1	6.6	9.2	12	13
	0.72	0.75	0.23	0.46	0.69	1.0	1.5	2.3	3.5	5.6	8.0	11	12
	0.62	0.75	0.15	0.29	0.45	0.70	1.1	1.7	2.5	3.9	6.1	9.0	9.7
	0.50	0.75	0.10	0.19	0.29	0.47	0.77	1.2	1.8	2.7	4.5	5.9	6.5
	0.38	0.75	0.065	0.13	0.19	0.29	0.43	0.66	1.0	1.5	2.3	3.4	3.9
	0.31	0.75	0.040	0.079	0.12	0.18	0.26	0.47	0.71	1.1	1.6	2.4	2.8
	0.25A	0.75	0.030	0.059	0.084	0.14	0.19	0.30	0.47	0.72	1.2	1.6	1.8
	0.25B	0.75	0.0071	0.014	0.027	0.046	0.081	0.13	0.21	0.31	0.48	0.73	1.2
	0.12A	0.5	0.0070	0.014	0.025	0.040	0.062	0.086	0.13	0.20	0.33	0.50	0.51
1.5	1.25	1	0.64	1.3	1.9	2.9	4.4	6.8	11	16	24	28	30
	1.00	0.75	0.45	0.90	1.3	1.9	3.0	4.5	6.6	11	17	22	22
	0.81	0.75	0.16	0.33	0.59	0.93	1.6	2.6	4.5	6.1	9.4	14	16
	0.62	0.75	0.13	0.27	0.43	0.78	1.1	1.9	3.2	5.3	6.3	8.3	10
	0.38	0.75	0.044	0.088	0.14	0.23	0.36	0.60	0.88	1.3	1.9	3.2	3.7
2	1.62	1.5	1.1	2.2	3.1	4.6	7.0	10	16	30	41	45	47
	1.25	1	0.62	1.2	1.9	2.8	4.3	6.4	10	15	24	29	30
	1.00	0.75	0.44	0.89	1.3	1.9	3.0	4.7	6.8	12	18	22	23
	0.81	0.75	0.30	0.61	0.87	1.2	1.8	2.7	4.0	6.0	8.8	14	17
	0.62	0.75	0.14	0.28	0.44	0.74	1.1	1.8	2.7	4.5	6.7	9.0	10
	0.38	0.75	0.073	0.15	0.21	0.32	0.49	0.80	1.2	1.9	2.6	2.9	3.0
3	2.62	2	2.2	4.5	7.8	13	21	36	71	88	92	101	108
	2.00	1.5	1.6	3.3	5.2	8.7	14	25	43	63	71	77	82
	1.62	1.5	1.1	2.2	3.2	4.8	7.4	11	17	28	42	47	49
	1.25	1	0.61	1.2	1.7	2.5	4.0	6.7	11	18	24	30	33
4	3.50	2.5	5.3	11	15	24	36	57	114	156	168	183	195
	2.62	2	3.4	6.8	10	17	27	42	67	100	111	123	133
	2.25	2	2.1	4.2	6.3	10	16	26	41	67	81	90	98
	1.62	1.5	1.1	2.2	3.2	4.8	7.4	11	17	28	42	52	56
6	5.00	3	6.8	14	18	35	62	148	249	297	339	374	400
	3.50	2.5	5.4	11	16	26	41	65	114	152	176	202	224
	3.00	2	3.9	7.8	11	17	25	37	73	115	134	153	170
	2.62	2	2.8	5.6	8.9	15	23	37	68	93	106	119	130
8	6.25	4	14	28	46	76	115	184	334	497	589	643	691
	5.00	3	7.0	14	19	36	65	134	212	298	359	412	457
	3.50	2.5	5.5	11	16	26	41	65	114	161	196	220	245
	2.62	2	3.0	6.0	8.7	15	23	37	68	100	117	132	143
10	8.00	4	20	41	70	112	191	424	603	723	817	921	1013
	6.25	4	14	28	46	76	115	184	335	498	590	644	693
	5.00	3	7.0	14	19	36	65	134	212	298	378	434	482
12	9.50	4	29	58	99	158	269	534	766	960	1140	1290	1410
	7.38	4	22	44	65	97	142	267	480	637	756	858	935
	6.25	4	14	28	46	76	115	184	334	508	612	683	750

[*] For each valve size, the values given in the first line correspond to full-area trim; the values for reduced trims follow in descending order. Courtesy of Flowserve Corporation.

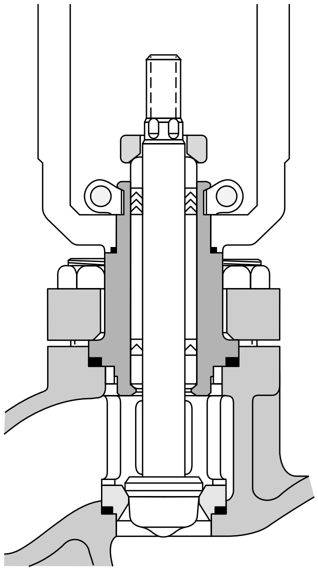


FIG. 6.19d
Valve with plain bonnet and separable bonnet flange design. (Courtesy of Flowserve Corporation.)

Typical components that make up a globe valve are shown in Figure 6.19d. Globe valve trims are available in many design variations, depending upon manufacturer and the intended application of a particular valve. A complete discussion of every trim design variation is impossible, but those with the most general application use will be covered, including special trims for severe services such as noise, cavitation, and erosion. See Sections 6.1 and 6.14 for discussions of low-noise control valves.

Trim Flow Characteristics

All control valves are pressure-reducing devices; in other words, they have to throttle the flowing fluid in order to achieve control. The most widely used form of throttling is with a single-stage orifice and plug assembly. Multiple-stage orifice elements are usually found in trim designs for combating noise, erosion, and cavitation (Figure 6.19g). In all cases, the valve trim is the heart of the valve and operates to give a specific relationship between flow capacity and valve plug lift. This relationship is known as the valve flow characteristic and is achieved by different cage orifice patterns (Figure 6.19e) or valve plug contours (Figure 6.19f).

The term “flow characteristic” usually refers to the “inherent” characteristic, which is a function of a number of valve design and manufacturing parameters. The inherent characteristic is determined by testing the valve flow vs. valve lift using a constant differential pressure across the valve throughout the test. These types of tests are standardized by ANSI/ISA-75.02-1996 and IEC 60534-3-2: 1997.

Therefore, the manufacturers’ trim characteristic curves or tables should not be confused with the installed flow characteristic in the actual process fluid flow loop. In actual service, the differential pressure across the valve varies throughout the valve lift and flow range as a function of the system characteristics. This variation is due to such factors as pump head changes with flow, piping friction losses, and the hydrostatic resistance of pipe fittings, block valves, flow measurement devices, heat exchanges, and other system elements.

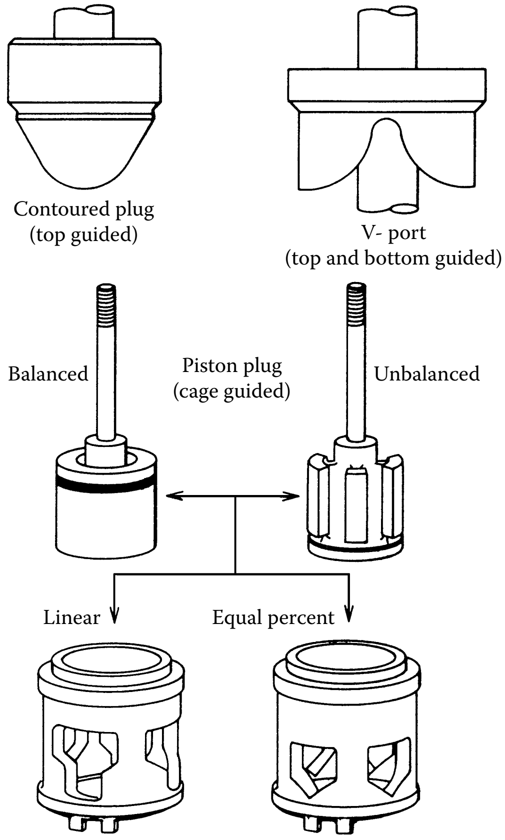


FIG 6.19e
Types of valve plug configurations for various valve designs. (Courtesy of ITT Conoflow.)

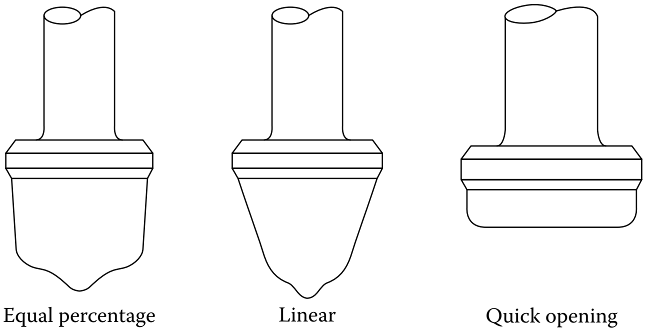


FIG. 6.19f
Valve plug shapes to produce the three common flow characteristics: equal percentage, linear, and quick opening.

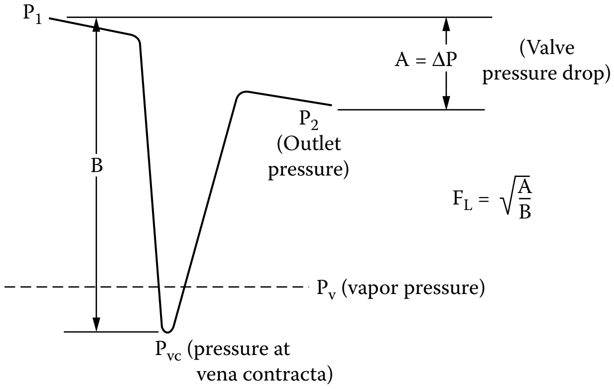


FIG. 6.19g
Pressure profile of a single-seat valve that is experiencing cavitation.

Control valve inherent characteristic data are expressed in graphs or tables, such as shown in Table 6.19c, where a flow coefficient ($C_v = K_v/1.17$) is expressed as function of the percentage of valve opening. It is important to understand the meaning of these flow coefficients. A detailed discussion of control valve sizing is provided in Section 6.15, which discusses control valve sizing. In its simplest form, the valve capacity coefficient, C_v (or K_v), for liquids can be expressed as

$$C_v(\text{or } K_v) = Q\sqrt{\frac{G_f}{\Delta p}} \qquad \mathbf{6.19(1)}$$

where Q is volumetric flow rate in gpm (or m^3/h for K_v), G_f is specific gravity relative to water, and Δp is the pressure differential (the lesser of the actual Δp or the choked Δp) across the valve in lb/in.^{2} (or bar for K_v). Important observations should be made from this simplified expression.

1. C_v and K_v are not dimensionless coefficients. They have units of (volume/time) × (area/force)$^{1/2}$ = (length)^4/[(time)(force)$^{1/2}$].
2. Valve manufacturers publish valve inherent characteristic in terms of C_v (or K_v) vs. lift (or percentage open).
3. Valve users need to know flow rate vs. lift for the installed characteristic. They must determine the system pressure differential allocated to the valve for the full range of flow rates in order to calculate the flow rate vs. lift.

Control valve manufacturers commonly furnish three types of inherent characteristic valve trims along with some minor variations (Figure 6.19a). These are idealized curves and do not accurately reflect the actual characteristic as determined by test. Examination of actual test data will show deviations in lift vs. flow of 10% or more, slope variations, and other distortions from the ideal curve.

This is due to a number of factors; in order of their significance, they include a) the trim type and design, b) valve body geometry effects, c) test variations and repeatability, and d) manufacturing variations. For practical purposes, these distortions, if kept within reasonable limits, do not materially affect the valve in actual service. Allowable limits on variations in flow characteristic are established in industry standards ANSI/ISA-75.11 and IEC 60534-2-4.

The typical inherent characteristic (i.e., C_v or K_v vs. lift) test data and pressure loss vs. flow rate data for the static elements of the process system can be used to approximate the valve characteristic behavior in the installed system. This can be used to select the best valve trim for the controlled process, which will keep the control loop gain constant or optimized for process control (see Figure 6.7a in Section 6.7).

A traditional rule of thumb is to use a linear trim if the control valve pressure drop is relatively constant (such as in pure pressure reducing). Where there is significant system and valve pressure drop variation as flow changes, the equal-percentage trim is recommended.

Rangeability

Rangeability can be expressed in different ways with different meanings. *Inherent rangeability* is the ratio of the largest controllable flow coefficient (C_v or K_v) to the smallest controllable flow coefficient within specific deviation allowances.[1]

Valve *flow rangeability*, referred to as *turndown*, is the ratio between the valve's maximum and minimum controllable flow rate at stated operating pressures. Generally, the minimum controllable flow is considered to be about twice the minimum clearance flow as the plug lifts off the seat. For single-seated contoured plug control valves, manufacturers often state inherent rangeability from 30:1 to 50:1 based on C_v vs. lift tests.

In practice, these numbers are merely benchmarks. When applied to processes with high system pressure losses, the installed flow rangeability of the control valve is more likely to be in the order of 7:1 to 15:1. This is usually enough because most processes do not operate much over 5:1 turndown. Some processes can require flow rangeability beyond the capability of one valve and will require parallel valves with split-ranging. For more details, see Figures 6.1i and 6.1j.

While linear and equal-percentage trims are designed to throttle essentially over the full valve travel, the quick-opening characteristic is designed to act more like a "bath stopper" plug. The flow characteristic develops approximately 80% of its C_v capacity in a nearly linear manner over the initial 20–30% of valve lift. The remaining capacity is added over the balance of the lift. This plug can be used for on/off service, for short-stroke valves such as self-contained pressure regulators, or in some process applications where a decreasing valve gain is required as the load increases.

---

[1] See ANSI/ISA-75.11-1985 and IEC 60534-2-4: 1989.

Standard Trim Configurations

The valve plug configurations used on modern control valves include the contoured plug, ported plug, or piston plug (see Figure 6.19e). The turned or contoured plug is probably the most common, followed by the piston and ported plugs. The contoured plug is simple to machine out of bar stock from stainless steels and special alloys and can be hard-faced easily for erosive services. The contoured plug is usually used with single-seat valves, but it is also available in double-seat designs.

The ported plugs usually have cast or forged stainless steel or special alloy heads mechanically attached to a similar stem material. They are more difficult to hard-face than designs with uninterrupted control surfaces. This plug is most common in double-seat valves, but it is also available in single-seat designs. Both the contoured and ported plugs are shaped to obtain the desired flow characteristic as they move in and out of the seat ring.

The piston plug is relatively easy to fabricate and is usually made from a hardenable type of stainless steel such as 410, 440, or 17-4PH™. Other materials such as nickel alloy 718, alloy K-500, and austenitic stainless steel with hard-facing are also available. Plug heads are usually a hardened material, or hard-facing must be done, because these plugs are normally guided in a cage assembly. The flow characteristic for this trim design is incorporated into the cage.

Special Trim Configurations

There are severe applications that may require special trim configurations. These applications usually involve noise, cavitation, erosion, or combinations of these problems. The special trim designs for all of these are often similar in concept although they will differ in design detail. Because noise is an especially difficult and complex problem to deal with, it is covered in depth in Section 6.14. The following discussion will touch upon the noise reduction trims, but will primarily concentrate on cavitation and erosion services.

Cavitation Liquid cavitation is a complex fluid dynamic reaction to pressure change, which is also discussed in Sections 6.1, 6.14, and 6.15. It has many aspects that have not been successfully explained, but the fundamentals are relatively simple. The basic process of cavitation is related to the conservation of energy and Bernoulli's theorem, which describe the pressure profile of a liquid flowing through a restriction or orifice (Figure 6.19g).

In order to accelerate the fluid through the restriction, some of the pressure head is converted into velocity head. This transfer of static energy is needed to push the same mass flow through the smaller passage. The fluid accelerates to its maximum velocity, which is also the point of minimum pressure (vena contracta). The fluid velocity gradually slows down as it again expands back to the full pipe area. The static pressure also recovers, but part of it is lost due to turbulence and friction.

If the static pressure at any point drops below the liquid vapor pressure (P_v) for that temperature, then vapor bubbles will form. As the static pressure recovers to a point greater than the vapor pressure, the vapor bubbles collapse back into their liquid phase. The cavitation process includes the vapor cavity formation and sudden condensation (collapse) driven by pressure changes. The growth and collapse of the bubbles produce high-energy shock waves in the fluid. The collapse stage of the process (the bubble implosion) produces more severe shock waves.

These implosions generate noise, fluid shock cells, and possible microjets that impinge upon the trim parts. This generates highly concentrated impact forces that cause surface fatigue and localized fractures that destroy the metal. This erosion process gives cavitation damage a very distinctive appearance, like that of cinder block or sandblasting.

No known material will withstand continuous, severe cavitation without damage and eventual failure. The length of time it will take is a function of the fluid, metal type, and severity of the cavitation. Without special trim geometry, some of the possible mitigating actions include the use of extremely hard trim materials or overlays, increasing the downstream back-pressure, or limiting the pressure drop by installing control valves in series to reduce the pressure drop in each valve.

Another mitigating effect in some processes is a result of the fluid thermodynamic properties and operating conditions. As liquid operating temperature approaches its critical temperature, heat transfer effects become increasingly significant relative to the dominant inertial effects, which causes the growth and collapse rates of cavities to slow down. This can greatly reduce impingement stresses on the valve parts. Some cryogenic and hydrocarbon applications are thought to behave this way, which may partly explain why cavitation damage in these cases is minimal or absent even when cavitation is present in the valve. Further information about cavitation and predicting its effects on valve performance can be found in Section 6.15 of this text and in the Instrumentation, Systems, and Automation Society (ISA) Recommended Practice RP75.23.01, "Considerations for Evaluating Control Valve Cavitation."

Some of the special trim designs for combating cavitation are the Drag®, Cavitrol®, Turbo-Cascade®, VRT®, Hush®, ChannelStream®, and various other staged or step-type plugs and orifices, which are shown in Figures 6.1y, 6.1z, and 6.1aa. The multihole Cavitrol-style trims are designed to break the flow into multiple fluid jets and force the jets to impinge upon themselves with extreme turbulence. This turbulence converts part of the upstream energy (static pressure) into heat energy. Bubbles from the many small fluid jets are small and tend to implode within the turbulent fluid core away from the internal surfaces, which greatly reduces trim damage. However, these trims are suitable for only moderate cavitation and moderate pressure drops.

The balance of the trims listed can be considered variations of staged trims. These trims reduce the total valve pressure drop in multiple steps such that the vapor pressure of the fluid is not reached in any stage. In some cases, the fluid is also forced to undergo multiple changes in direction to promote gradual head loss and energy conversion.

It should be pointed out that virtually all anticavitation trims are designed on the basis of data from tests with water. Therefore, actual performance with different fluids cannot be absolutely extrapolated. However, because water is one of the most destructive fluids under high pressure drop conditions, it is likely that a particular valve trim that is tested with water will be reasonably effective with most other fluids. In most applications, actual field experience is necessary to demonstrate the adequacy of a particular trim design.

Erosion Erosion of the valve trim can also be caused by high-velocity liquid impingement, abrasive particles, and erosive-corrosive combination action. Erosion damage is roughly proportional to some power (e.g., $1.4 \le n \le 6$) of velocity, which depends on the erosive environment and the boundary material. As in cavitation, one major key to the solution is to reduce the velocity through the trim. It is no surprise that the valve and trim designs discussed above are also useful on most erosion problems, although there may be some variations of design and materials of construction needed to better cope with a particular problem.

A properly chosen and specified control valve and trim type can be one of the most reliable pieces of equipment in process service. Indeed, many companies no longer use control valve bypasses, except in some unique situations. However, it is a requirement that there be close consultation with the control valve manufacturer in specification of special trims. High-velocity liquid impingement erosion is usually associated with high pressure drop coupled with undesirable valve geometry. High-velocity fluid jets developed through the seat area will often result in erratic flow patterns that allow the liquid to impinge directly on the valve trim and body. Such damage is often confined to specific areas in the valve. Liquid droplets in a vapor stream can also cause impingement erosion, but it is generally spread over a greater area. Impingement damage is characterized by relatively smooth grooves and pockets worn into the metal.

Abrasive erosion occurs when the fluid stream contains solid particles that are harder than the trim surface and are traveling at sufficient velocity. This erosion can be likened to a type of scouring action that wears away metal, similar to a file or grinder. Solutions to the problem involve the use of harder trim materials, streamlining the flow pattern, and reducing velocity. However, abrasive erosion can only be reduced in magnitude and not entirely eliminated.

Good valve and trim service life can be obtained in some cases, but in severe problems other alternatives should be considered. If the fluid and operating conditions are compatible with elastomers, it might be better to consider pinch valves (see Section 6.20). With high velocities or corrosive fluids, ceramic-lined valves or chokes can be used.

Erosion-Corrosion Metals in most ambient and process environments resist corrosion by means of a protective metal-oxide film. Rust is a form of protective film on iron and steel, even though it has some undesirable characteristics. If the protective film is damaged, worn, or dissolved, the base metal is exposed to further corrosive action and a new oxide film is formed at the expense of the base metal. Protective films can be damaged by particle abrasion, mechanical wear, cavitation, chemical attack, and fluid velocity or turbulence.

Common rust is a relatively weak film that is easily disturbed chemically or mechanically. Small additions of alloys, such as copper, to the steel can increase the stability of the iron oxide against atmospheric corrosion, such as in "weathering" steel. Substantial additions of chromium to steel create a "stainless" steel that forms a relatively strong protective chromium oxide film instead of iron oxide, which protects against further attack in a wide range of chemical environments and is referred to as the *passive* layer or film.

The flow of fluid through a piping system and especially through valves and fittings can have both electrochemical and mechanical effects on protective films. Velocity, turbulence, and impingement can increase the polarization rates of the oxidation and reduction reactions at the metal-electrolyte interface, which can weaken or dissolve metal-oxide films.

The mechanical effects of fluid velocity can more easily remove a weakened or thinned protective film, leaving the base material exposed to further corrosion. When these effects are combined in a way that accelerates the rate of corrosion from a static state, it is called *erosion-corrosion* or *flow-accelerated corrosion*. This is a complex phenomenon, and actual service experience may be needed to determine which alloys and trim configuration will give the best service.

Depending on the actual metal, chemical, and velocity characteristics of a specific application, erosion-corrosion may appear in different ways. In some cases, the eroded surface may appear like a sandy beach with wave-like ripples or ridges. When there is more direct impingement or severe turbulence, the surface might have deep gouges, undercuts, or gullies. Some cases exhibit a pattern of elongated pits. Erosion-corrosion has been observed in the following applications.

- Deoxygenated water (condensate and boiler feedwater) or wet steam in the temperature range of 212–480°F (100–250°C) in carbon steel valves and fittings of fossil-fueled and nuclear power plants
- Polluted or silty, salt or brackish waters with dissolved or entrained gases with low levels of sulfur compounds in copper, bronze, or brass systems of oilfields and wastewater applications
- Slurry flow or cavitation in stainless steel systems

Users should consult experienced corrosion specialists or metallurgists regarding alloy selection, system design, and process chemistry (or water treatment) to prevent or correct these situations.

Trim Materials

The most popular general service trim material is austenitic type 316 stainless steel, which is commonly used with and without hard-facing up to about 800°F (427°C) and in special cases up to 1200°F (649°C). Other harder materials are frequently required in special trims, higher temperatures, and trim parts that might gall because of close tolerance metal-to-metal sliding action (cage guiding and guide bushings). Among these materials are 17-4PH, 410, 416, and 440C stainless steels; hardenable Ni-Cr-Fe-Mo alloys (e.g., Inconel® 718); cobalt-chromium alloys (e.g., Stellite®); nickel-boron alloys (e.g., Colmonoy® and Deloro® hard-facing); tungsten carbide; and ceramics.

For very corrosive services, more noble or high alloy metals are used to advantage. Among these are Alloy 20 stainless steel, nickel, titanium, tantalum, zirconium, 70Ni-30Cu alloys (e.g., Monel® 400, Monel K-500), Ni-Cr-Fe-Mo alloys (e.g., Inconel), Ni-Cr-Fe alloys (e.g., Incoloy®), Ni-Mo alloys (e.g., Hastelloy®-B/B2), and Ni-Cr-Mo alloys (e.g., Hastelloy-C/C276, -C22). It is difficult to generalize on recommended materials or material combinations for valve trims because of the wide range of valve designs and process application requirements.

The specifying engineer should utilize not only his or her own knowledge, but also enlist the experience and expertise of the manufacturer and material specialists and metallurgists when needed. Fortunately, many applications are reasonably straightforward, and standard trim material combinations set forth by the manufacturer can be used.

Some general guidelines can be given for specifying trim. When specifying hard-faced plugs and seats, the plug can be supplied with hard-facing alloy on the seat surface only (Figure 6.19h). This may be sufficient if the valve is subject to high pressure drop primarily during shut-off. However, for continuous high pressure drop throttling, the full face or contour of the plug should be completely overlaid with hard-facing alloy, unless the base material is already hardened.

If the plug is stem-guided or post-guided, the lower guide area should also be hard-faced for high pressure drop throttling or if the fluid temperature is above 750°F (400°C). As with the plug, the seat ring can be hard-faced only on the seating surface or over the entire bore surfaces, depending on the severity of the pressure drop and temperature. A variety of cobalt-chromium and nickel-chromium-boron alloys are available for hard-facing. In some cases, coating techniques such as flame or plasma spraying will be used. The valve manufacturer's recommendations are valuable guidance.

Commonly used hardenable alloys for trim materials include

- 410 and 416 stainless steels hardened up to about 38 Rockwell C (HRC)
- 440C stainless steel with hardness up to 60 HRC
- 17-4PH stainless steel, a precipitation-hardened material combining good corrosion resistance with a range of tensile properties and hardness between 28 and 42 HRC depending on heat treatment

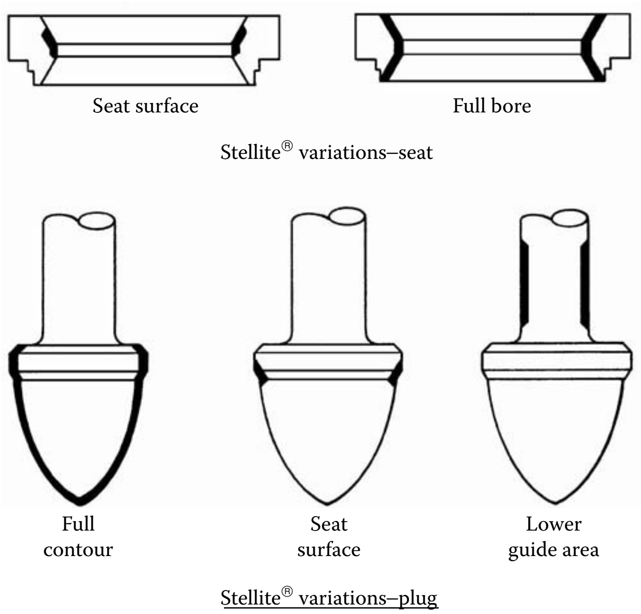


FIG. 6.19h
Hard-facing of plug and seat rings typically used in erosive services. (Courtesy of Flowserve Corporation.)

The 400-series (martensitic) stainless steels have limited corrosion resistance suitable for most water and steam service up to about 700°F (371°C). They are generally used for trim parts that can be made from bar stock or forgings, but some types are available as castings. Stainless steel 17-4PH is available in cast and wrought forms and is a usable general-purpose alloy up to 750°F (400°C). It is not as corrosion resistant as austenitic 316 stainless steel, but it is significantly better than most common 400-series types. Where 316 or other soft alloys are indicated as the proper alloy for a cage-guided trim, hard-facing will be required to minimize metal-to-metal galling problems.

Leakage

Control valves have varying degrees of shut-off capability, depending upon the valve and internal trim design, material, and manufacturing methods. Tight shut-off is not always a requirement, especially for throttling control valves. However, as reliability of control valves has increased, more applications are requiring tight shut-off for control valves in order to minimize the number of isolation valves and bypass loops. Another consideration should be the cost of lost or contaminated product and wasted energy resulting from leakage through valves. Users should consider all of these factors along with the cost and operating requirements of the valve before specifying the shut-off requirements.

Shut-off requirements are usually specified as an allowable volumetric leakage rate measured in a standardized test.

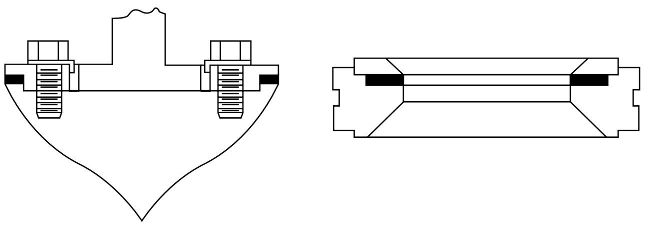


FIG. 6.19i
Typical soft-seat insert designs. (Courtesy of Dresser Flow Solutions.)

While it is permissible for a user to specify any value for the allowable leakage at a specified pressure drop, it is more common to specify one of the standard *leakage classes* defined by industry standards. The two most commonly applied standards, FCI 70-2 (formerly ASME B16.105) and IEC 60534-4, define several classes, with Classes II, III, IV, V, and VI being most commonly used for globe valves. Class V and Class VI represent the smallest allowable leakage depending on the pressure drop and test method specified. Refer to Section 6.1 for information about determining allowable leakage in each class.

Good single-seated valve trims can give Class IV or V shut-off in nonbalanced plug designs. Leakage Classes III or IV are typical for balanced plugs, but elastomer seals and special pilot-operated balanced trims have been used to achieve Class V shut-off. Figure 6.1gg gives the seat leakage class tabulation. Typically, Class VI shut-off is specified for soft seat inserts such as polytetrafluoroethylene (PTFE) or Teflon®, ETFE (Tefzel®), or other plastics.

Figure 6.19i shows two typical configurations where the soft insert can be located in either the plug or the seat ring. Class VI leakage is also achievable with special lapped-in or precision-fit metal plugs and seats.

Note that operating service temperatures, high or low, will limit the use of soft seat materials. The pressure drop across the valve is another limiting factor, although some "protected insert" designs (Figure 6.19j) will operate at very high pressure drops.

Seat leakage tests and classes are defined only for new valves, and it should not be assumed that the same level of tightness can be maintained in service. When a valve is placed in service, the seating surfaces can be worn or damaged by high velocity, pipe scale, process solids, corrosion, or vibration, which cause leakage to increase over time. Wear-resistant materials and careful installation and start-up practices can minimize wear and damage to seating surfaces.

Plug Stems

The valve stem connects the plug head to the actuator stem or coupling. It has several requirements as part of a carefully balanced mechanical system consisting of the plug head and stem, seat ring, cage, bonnet, stem bushings, stem packing, and actuator. It must be strong enough to transfer the load from the actuator to plug and bear the seat loading forces without cyclic fatigue. It must be stiff enough under maximum actuator loading to prevent buckling or significant deflections that promote packing wear and leakage. Yet it cannot be so big as to generate excessive packing friction or unnecessary static pressure forces that make actuation and control difficult.

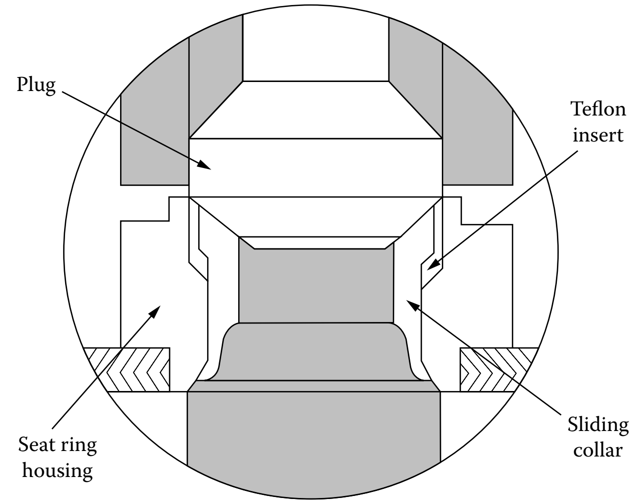


FIG. 6.19j
Protected soft seat insert for high pressure drop and tight shut-off. (Courtesy of Dresser Flow Solutions.)

The proper function of the valve also requires precise alignment of the plug and stem with the seat, bonnet, and actuator in order to prevent excessive friction and binding and to ensure minimal leakage at the seat or through the packing. Materials, design, and surface finish of the stem and bushings must prevent galling and minimize friction and packing wear. Material selection of the stem includes consideration of the process environment inside the valve, the range of environmental conditions outside the valve, and a mixture of those environments in the packing interface. In some unusual corrosive services, moisture or chemicals in the atmosphere can react with process chemicals to create a more corrosive mixture in the stem-packing interface. This situation may require a stem material with more corrosion resistance than the plug head that is exposed only to the process fluid.

Valve stems are normally the same material as the plug head, but they can be different based on the design criteria discussed above. Depending upon the type of valve design, the valve stem may be integral (one piece) with the plug, or it can be threaded into the plug and then pinned to prevent unscrewing. Because the manufacturer does a number of things to insure proper alignment and a solid, vibration-resistant threaded connection, good maintenance practice may dictate replacement as a unit rather than separate pieces.

Stem guides in one form or another are an integral part of the trim assembly. Metal stem guide or bushing material is selected to minimize metal-to-metal wear and galling

TABLE 6.19k
Common Trim Material Characteristics

Trim Material	*Hardness Rockwell C*	*Impact Strength*	*Corrosion Resistance*	*Recommended Max. Temperature*	*Erosion Resistance*
316 stainless steel (without hard-facing)	80–90 Rockwell B	Excellent	Excellent	600°F 316°C	Fair
316 stainless steel (with hard-facing)	Varies with hardfacing	Excellent	Excellent	1200°F 649°C	Excellent
410/416 stainless steel	35–42	Good	Fair	700°F 371°C	Good
420 stainless steel	50	Good	Fair	800°F 427°C	Excellent
17-4PH H1025	38	Good	Good to excellent	800°F 427°C	Good
440C stainless steel	55–60	Fair	Fair	800°F 427°C	Excellent
Ni-Cu Alloy K-500	32	Good	Good to excellent	600°F 316°C	Fair to good
Tungsten carbide	~72	Fair to good	Good (with nickel alloy binders)	1200°F 649°C	Excellent
Co-Cr-W Alloy 6 (hard-facing, cast, or wrought)	42	Fair	Good	1500°F 816°C	Good
Ni-Cr-B No. 5 (hard-facing)	45–50	Fair	Good	1200°F 649°C	Good

Courtesy of Flowserve Corporation.

against the stem. For lower temperatures and light duty, nonmetallic guide bushing materials, such as reinforced PTFE fluoroplastic or compressed graphite, are common choices. Metal guides may be of such materials as 17-4PH, 440-C, Stellite, or hard-chrome plated or nitrided stainless steel, bronze, and aluminum-bronze. For the main characteristics of the common trim materials refer to Table 6.19k.

BONNET DESIGNS

The valve bonnet is the top closure assembly for the globe valve, as well as for several other valve body design types. In addition to closing the valve body, the bonnet also provides the means for mounting the actuator assembly to the valve body and sealing the valve stem against process fluid leakage. The various bonnet designs will be discussed along with the subject of stem sealing utilizing packing materials, lubricants, and special seal designs.

In addition to considerations of pressure containment, manufacturers design bonnets to provide features that predominate in their particular philosophy of valve design. The depth of the stuffing box and surface finish, provision of guides, method of operator attachment, packing design flexibility, and packing follower design are all details that vary from manufacturer to manufacturer or even among various body designs offered by one manufacturer.

Some low-pressure valves, especially in sizes below 2 in. (DN 50), can be provided with a threaded bonnet. This reduces the weight and is more economical in first cost than the flanged and bolted design. Depending on the materials of construction, it can be difficult to remove a threaded bonnet after extended service in high temperature or corrosive service, making this type more costly to maintain in these types of applications. Economic justification for using a valve with a threaded bonnet is sometimes based on a life-cost model of “discard and replace,” whereas larger valves, bolted bonnet valves, and control valves typically require a “maintain and repair” cost analysis. In special cases the bonnet-to-body joint can be seal welded to prevent leakage of highly reactive, lethal, or radioactive fluids.

Bolted Bonnets

The most common valve bonnet design is the bolted bonnet, like the one shown in Figure 6.19l. It is usually fastened to the valve body by high-strength stud bolts and heavy nuts. Removal of the bonnet gives complete access to the valve trim for maintenance purposes. Because the bonnet is a pressure-retaining part of the overall fluid containment system,

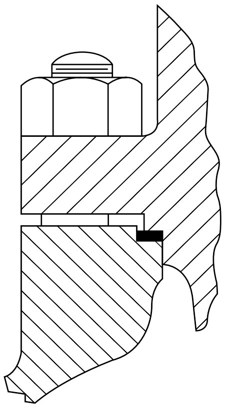


FIG. 6.19l
Bolted bonnet joint showing retained gasket design.

the design and materials are determined in accordance with an applicable pressure vessel standard. For example, the ASME Pressure Vessel Codes, ASME B16.34, API 6D, AD Merkblätter, European norms, ISO standards, and numerous individual nations' pressure vessel standards give material requirements and design criteria for flanges, wall thickness, and flange bolting.

Bonnet Gaskets The seal between the valve bonnet and body can take several forms depending upon the valve design and application range, including containment pressure and fluid temperature. Fluid corrosion is also a factor that must be considered. The most common seal is a contained gasket, either a flat or spiral-wound composite design (Figure 6.19l).

Other designs that may be found are the API ring joint (oval or octagonal cross-section), lens type, delta gasket, and Bridgeman gasket (Figure 6.19m). These latter designs are metal gaskets. The flat gasket is usually either graphite-based or a reinforced PTFE. The spiral-wound gasket usually consists of graphite ribbon or PTFE wound between thin metal strips. The metal windings of a spiral-wound gasket are usually stainless steel, but they are available in a variety of corrosion resistant and high-temperature alloys.

Asbestos fillers were commonly used in gaskets until the late 1970s, largely due to the lack of effective substitutes for many high-temperature or corrosive sealing applications. Since then, sealing materials and technologies have improved to the point where there should be no further requirement for using this potentially hazardous material in gaskets. Proper handling and disposal are required if very old valves are encountered in order to reduce the risks of asbestos exposure.

Pressure Seal Bonnets

Another typical bonnet seal is called the *pressure seal,* and it is used in high-pressure valves Class 900 (PN 150) and higher to minimize size and weight of the bonnet-to-body connection. The pressure-sealed bonnet does not rely on heavy bolting and a thick flange to retain pressure, like the bolted bonnet. Instead, a wedge-shaped graphite-composite gasket is installed on top of a bonnet sealing lip and is compressed between the bonnet and a set of antiextrusion or spacer rings and heavy retaining ring segments that are indexed into the body. During assembly, the bonnet is pulled up tightly against the seal, spacer ring, and gasket retainer segments by the pull-up bolting and bonnet retainer to create a radial seal between the body neck and bonnet, as shown in Figure 6.19n.

When the valve is pressurized in service, the upward force from pressure under the bonnet further energizes the radial sealing. The design and materials of pressure-sealed

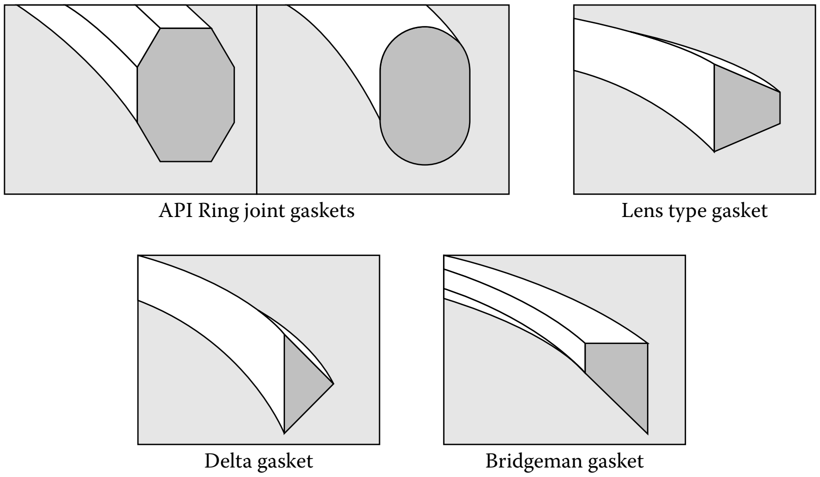


FIG. 6.19m
Special bonnet gaskets for high pressure and temperature.

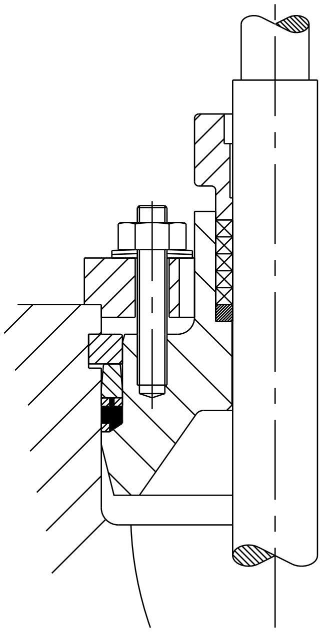


FIG. 6.19n
Pressure seal bonnet configuration showing composite wedge gasket. (Courtesy of Flowserve Corporation.)

gasket joints are especially important. If the gasket material bonds so tightly in service to the inside bore of the body that it becomes difficult to remove for maintenance, it may be time-consuming or impossible to free the bonnet without damage to the body. Fortunately, recent design and material improvements have minimized these types of maintenance problems. Pressure-sealed bonnet joints have advantages in high-pressure, high-temperature applications, because they permit more uniform body wall thickness at the body neck opening, which minimizes thermal stresses in the body during temperature transients. The valves are frequently lighter in weight than comparable bolted bonnet valves, and they are relatively easy to assemble.

Bonnet Classification

Bonnets fall into basically three classifications. These are standard, extended for hot or cold service, and special designs such as cryogenic extensions and bellows seals. These classifications along with the stem seal systems, guides, and bushings will be discussed in more detail later in this section.

Standard Bonnet The standard or plain bonnet (Figure 6.19d) is the normal bonnet design furnished on most valves. It covers the range of pressures and temperatures compatible with standard seal gaskets and stem packing materials. Generally, this includes valves designed for ANSI Pressure Classes 150 (PN 20) through 2500 (PN 420) and temperatures from −20 to 600°F (−30 to 315°C). Above 450°F (230°C) graphite-based packing or extended bonnets are recommended. The reason for setting a temperature limit for standard-length bonnets is to limit the temperatures to which the packing and actuator are exposed. Over 90% of all control valve applications can be handled by the plain bonnet design. It may or may not incorporate stem guides or bushings and may have very broad or very limited packing configurations available, depending upon the specific manufacturer and valve design.

Extended Bonnet The extended bonnet (Figure 6.19o) is usually required when the fluid temperature is outside the plain limitations of the standard bonnet temperature. Even when normal process temperatures are within the plain bonnet limits, it may be necessary to use the extended bonnet to protect the packing and actuator against temperature excursions during occasional process upsets.

Originally, the extended bonnet used different designs for hot and cold service. The hot service extended bonnet was provided with “cooling fins,” while the cold service extended bonnet was a plain casting without fins. Over many years it was demonstrated that fins on the bonnet added only marginal capability to heat dissipation in the packing area. It made a costly and complex casting and has been largely abandoned in favor of the plain extension.

In most modern control valve designs, the bonnet extension is similar for hot or cold service, except where deep cryogenic temperatures under −150°F (−100°C) are encountered. Some manufacturers offer two standard bonnet extension

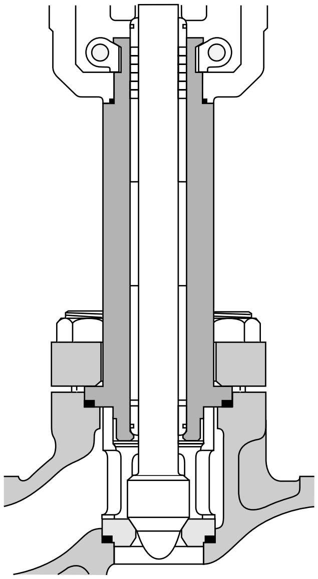


FIG. 6.19o
Standard high-temperature extension bonnet. (Courtesy of Flowserve Corporation.)

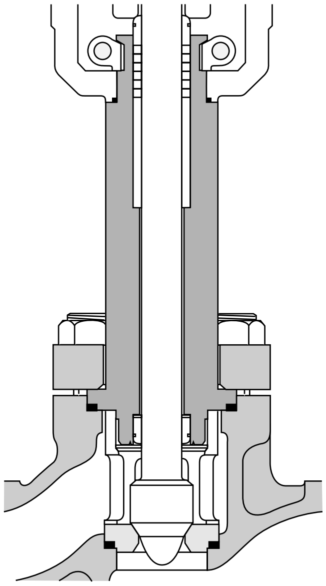


FIG. 6.19p
Standard cold extension bonnet. (Courtesy of Flowserve Corporation.)

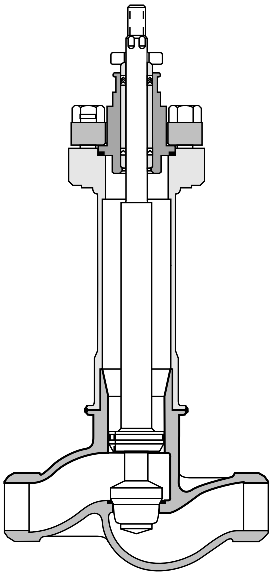


FIG. 6.19q
Typical cryogenic cold-box valve. (Courtesy of Flowserve Corporation.)

lengths (other than cryogenic) depending upon the operating temperature. In general, the standard extension bonnet is suitable from −20 to 800°F (−30 to 425°C) in carbon steel construction and from −150 to 1500°F (−100 to 815°C) in austenitic stainless steel construction.

Cryogenic Bonnet The cryogenic bonnet or cold-box bonnet (Figure 6.19p) is a special design adaptation of the extended bonnet. Depending upon a manufacturer's valve design, this style bonnet may be required with operating temperatures ranging from −150 to −300°F (−100 to −185°C) down to −425°F (−255°C) using a bolted bonnet. The main purpose for extending the bonnet for cold service is to keep the packing and upper plug stem warm enough to prevent icing or frost around the packing area and stem. Ice crystals on the stem can be pulled into the packing, damage the seal, and create a leak path.

The bonnet length is selected for the application based on valve body size, piping requirements, and operating temperature needs; it will generally range from 12 in. (300 mm) to 36 in. (900 mm). The standard cryogenic bonnets are distinctly different from cold-box designs or extended-neck designs (Figure 6.19q). The standard design is usually similar to the standard extension bonnet, except much longer. It can be cast or fabricated from multiple pieces by welding.

For extreme temperatures near −454°F (−270°C), the body neck is often extended to remove the bonnet gasket away from the extremely low temperature and to make the bonnet joint accessible outside of the cold-box. The extended neck of the cold-box body is usually fabricated from thin-walled stainless steel tubing (to reduce cool-down weight and heat leakage into the cryogenic process). The extension is welded directly to the body casting or forging (Figure 6.19q). At the top of the extension is the flange for connecting with the bonnet. In general, these applications are limited to ANSI Pressure Class 600 or below.

Special designs are also available for high-pressure service up to ANSI Class 4500 and above. Because the extreme cold requires good impact resistance, materials of construction are limited to the austenitic stainless steels (Types 304L or 316L) and bronze. These valves incorporate a stem seal system at the plug end of the bonnet to keep the cryogenic liquid out of the bonnet and packing area. This seal may be vented or nonvented, but it must allow a pocket of vapor to exist below the bonnet as insulation against severe convective heat loss from the warmer bonnet area if exposed to cryogenic liquid.

The seal design must also allow any build-up of gas pressure in the warmer bonnet area to relieve back into the valve body. Vented designs are typically used if the valve is installed with the neck and stem oriented vertically up to within 30° of horizontal to prevent liquid from getting to the packing area. Nonvented designs incorporate a unidirectional seal and are used when the valve neck is installed at or near horizontal. In some cases, where additional insulation is needed to reduce outside heat flow, the bonnet can be fitted with a vacuum jacket.

Bellows Seal Bonnet When no stem leakage can be permitted, many globe valve manufacturers provide extended bonnet designs that incorporate a bellows seal around the stem (Figure 6.19r). Bellows seals are justified in applications involving toxic or radioactive fluids, where leakage to the

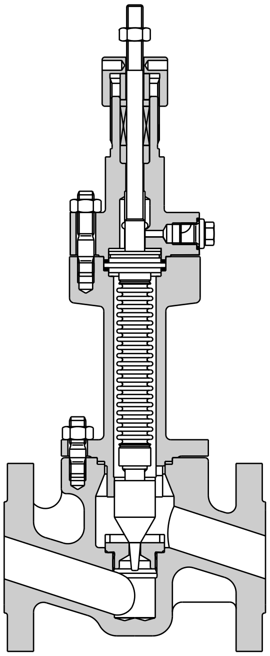


FIG. 6.19r
Typical bellows seal bonnet with formed bellows. (Courtesy of Flowserve Corporation.)

outside would pose personnel safety hazards. The bellows is usually made of stainless steel or other corrosion-resistant nickel alloys such as Hastelloy C-276 and Inconel 625. They can be hydraulically or mechanically formed (as shown in Figure 6.19r), or they can be made by stacking many individual leaf segments that are welded together at their outer edges and are known as welded or nested bellows.

The bellows is attached to one end of the stem by welding; the other end is welding to a clamped-in fitting with an antirotation device. The antirotation device prevents the bellows from being twisted during assembly and disassembly or by vibration of the plug in service.

Bellows seals are usually leak-tested from atmospheric pressure to vacuum using a mass spectrometer to detect helium leakage rates below 1×10^{-6} cc/sec. The service life of a metal bellows depends on the design, material, manufacturing processes, service pressure and temperature, corrosion effects, and stem cycle history. Failures of metal bellows usually occur by cyclic fatigue.

Bellows manufacturers that have tightly controlled production processes have cycle-tested their designs and are able to make reasonable predictions of expected cycle life. Bellows-sealed bonnets are backed up with a standard stem packing set and a leakage monitoring port between the bellows and the packing in order to prevent catastrophic release of hazardous fluid in the event of a bellows leak.

Metal bellows seals have pressure and temperature limitations. Ratings of about 150 PSIG (1030 kPa) at 100°F (40°C) or 90 PSIG (620 kPa) at 600°F (315°C) are typical. The average full stroke cycle life can vary from 50,000 cycles for size 1 in. (DN 25) and smaller valves to 8000 cycles for sizes 3–6 in. (DN 75–DN 150). In some cases, cycle life can be improved by reducing operating pressures or by using special short stroke valve plugs.

Operating pressures can be increased to as much as 2900 PSIG (20,000 kPa) and temperatures up to 1100°F (590°C) by multiple or heavy-wall bellows and selection of a high-strength metal alloy. However, this reduces cycle life considerably. As a result of these various factors, metal bellows bonnet seals are selected for relatively few applications. Improved environmental packing systems are specified for a majority of hazardous fluids, except for dangerous applications where the risks justify the additional costs of purchase and maintenance for metal bellows seals.

Bonnet Packing

In order to seal the valve stem against leakage of process fluid to the atmosphere, the upper part of the bonnet contains a section called the stuffing or packing box. This assembly consists of a gland flange, packing follower, packing spacer or lantern ring, lower packing retainer, and a number of packing rings. The surface finish of the stem should be very fine, on the order of 8 Ra (micro-inch roughness average), and the internal finish of the bonnet stuffing box should be 16 Ra or better. Various valve packing materials are available, but for control valves use three general groups of materials: fluoroplastics (PTFE, FEP, PCTFE), carbon graphite (compressed rings and braided yarn), and synthetic polymer fibers (PBI® and aramid).

Control valve packing must be compatible with the process fluid, seal the stem and bonnet, produce minimum starting and sliding friction, and give long service life in modulating service. The most popular material that meets all of these conditions over the broadest range of fluid applications is PTFE, used as V-rings or braided filament. As a result, the majority of control valve manufacturers provide a variety of packing configurations with this material.

Fluoroplastic Packing Polytetrafluoroethylene (or Teflon) is the most common packing material in the family of fluoroplastics. It is normally formed or machined as chevron or V-rings from virgin (not reprocessed) material. For special needs, a shape variation known as "cup-and-cone" is available. Note that in V-ring and cup-and-cone configurations, the top and bottom rings are adapter rings with one flat surface. Braided packing is also formed from PTFE filament. For higher temperature or pressure applications, PTFE can be reinforced or filled with up to about 25% by volume of glass fibers, silica, carbon, or graphite and other fillers to add strength and stiffness and to improve its resistance to cold-flow or creep. PTFE braided packing can be reinforced by using PBI or aramid fiber in the corner braids. As noted earlier, PTFE is limited to a maximum exposure temperature

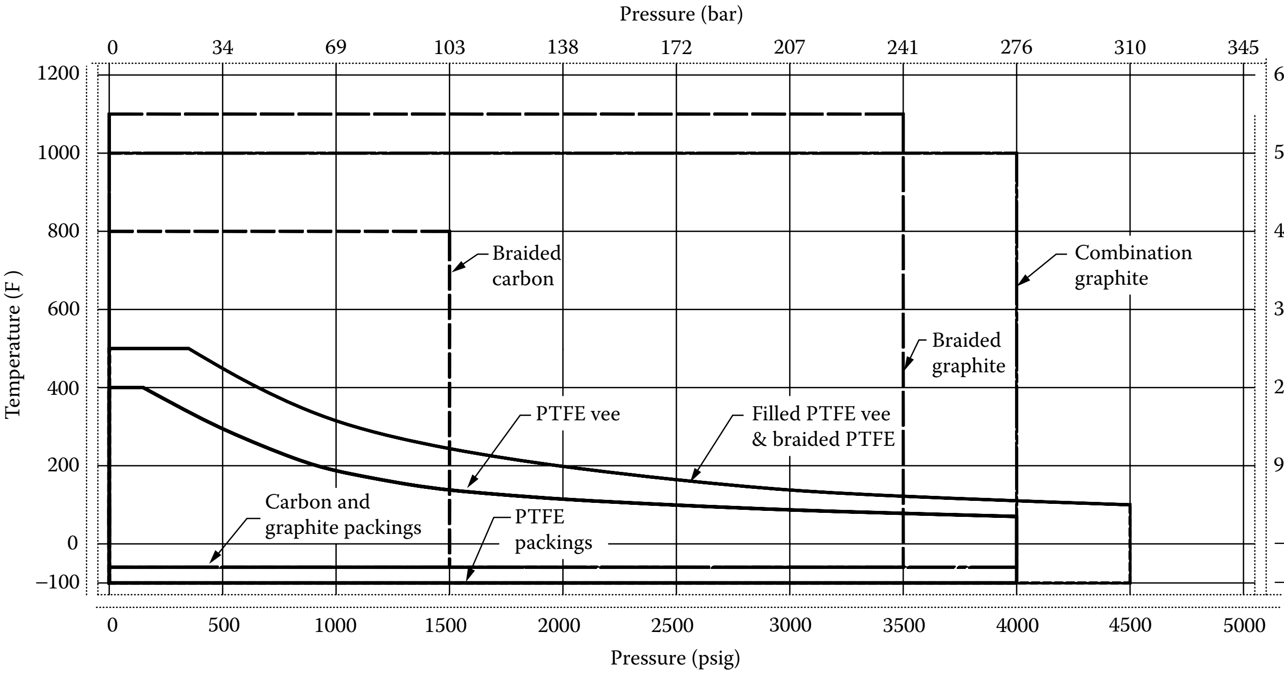


FIG. 6.19s
Recommended temperature-pressure limits for some common packing materials. (Courtesy of Flowserve Corporation)

of about 450°F (230°C) in a plain bonnet, depending on the service pressure.

When used in extended bonnets, some users have successfully used reinforced PTFE packing with process temperatures up to about 850°F (455°C), providing the packing box temperature remains below 450°F (230°C). Although PTFE packing has useable properties down to –320°F (–195°C), its practical lower limit for packing in a standard length bonnet ranges from –50°F (–46°C) to –100°F (–73°C), depending on bonnet design. See Figure 6.19s for pressure-temperature limits for PTFE and other packing in standard bonnets. In order to prevent ice damage to packing at lower temperatures, extended bonnets or cold-box-style bodies are used (Figures 6.19p and 6.19q).

Contrary to some claims, PTFE packing does not require *live-loading* springs to be effective as a stem seal. Normal packing follower loading and adjustment is all that is required, especially with the V-ring or chevron shape (Figure 6.19t). The cup-and-cone style requires a higher loading to effectively energize the seal. Live-loading springs are beneficial in special applications, such as in temperature cycling duty and for some types of environmental seals to reduce volatile organic chemicals (VOCs) and toxic emissions.

The packing rings and stuffing box dimensions should be very accurate for proper contact with the sealing lips of the V-rings. Dimensional tolerances are even more critical for the PTFE cup-and-cone rings because of the high degree of stiffness and packing loading required. PTFE packing has extremely low friction characteristics and does not require supplemental lubrication. However, excessive overtightening of the gland bolting can transform the V-ring structure into a solid compression packing capable of transferring high radial forces and excessive friction to the stem.

Asbestos Packing Asbestos is one of the oldest packing materials and was in wide use until the late 1970s, even though the health hazards of airborne fibers were known before then. Health risks were generally insignificant when handling new packing with binders to keep the fibers from fraying. However, in high-temperature service, the binders disappear, and removal of old packing from stuffing boxes

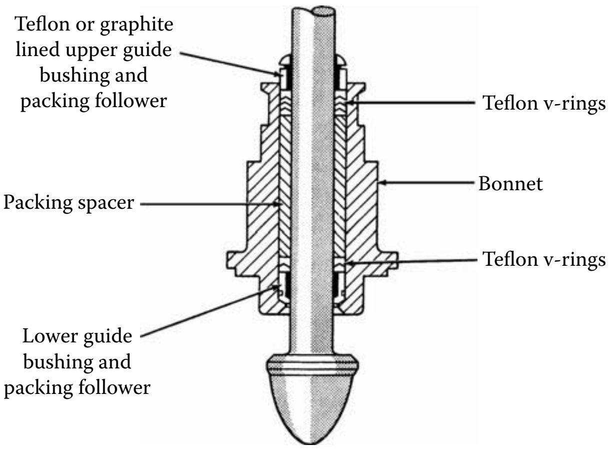


FIG. 6.19t
Typical Teflon V-ring arrangement.

had the potential of creating airborne fibers unless special dust mitigation procedures were employed. The lack of effective asbestos substitutes for many high-temperature or corrosive sealing applications made it difficult to abandon its use. Since then, sealing materials and packing technologies have virtually eliminated further need for using this potentially hazardous material. In order to reduce the risks of asbestos exposure, proper handling and disposal are required if very old valves are encountered.

Graphite Packing The improvement and optimization of graphite foil and yarn for use as valve packing came about largely out of the need to find substitutes for asbestos packing. Flexible graphite packing is available in formed rings of laminated graphite foil. Rings are formed in square, wedged, and cup-and-cone cross-sections.

A wide variety of braided constructions are available. Braided graphite can include dry lubricants to reduce friction. Braided graphite can be reinforced for high-pressure service using synthetic fibers like PBI or aramid. Reinforcing of the graphite braid with an Inconel wire core or fine wire-encapsulated yarn is common for high-temperature, high-pressure service. Several typical graphite packing configurations consist of a combination of formed laminated rings with braided rings above and below the laminated set to act as wiper rings.

Flexible graphite packing can be used in many fluids including reducing and mildly oxidizing acids (not strong oxidizers), caustics, hydrocarbons, solvents, water, steam, and gases. It has high-temperature capability up to 1200°F (649°C) in steam; and in inert, nonoxidizing media (i.e., oxygen-free environments like nitrogen or carbon dioxide) pure graphite it is useable up to 4500°F (2500°C). In atmosphere and oxidizing media, the high-temperature limit can vary between 650°F (343°C) and 850°F (454°C) depending on the specific materials; consult with packing manufacturer for specific applications. If an extended bonnet is used, the upper temperature can be extended to 1200°F (650°C) on both oxidizing and nonoxidizing service. The low temperature limit for pure graphite braided packing can be as low as –400°F (–240°C).

Packing and valve manufacturers have undertaken considerable developmental work in an attempt to overcome or mitigate past problems with graphite. Among these shortcomings are the following:

- Relatively high stem friction
- Difficulty in "energizing" the packing to give an effective stem seal
- Low cycle life without leakage
- Electrolytic pitting of stainless steel stems in conductive or high-temperature services
- Shortened packing life due to graphite plateout on the stem

There are several methods that can be used to extend the life and improve the performance of graphite packing. The packing or valve manufacturer should be consulted for proper methods for their specific sealing system. Some general things that have helped to improve graphite packing performance include the following:

1. Use combination packing assemblies consisting of laminated graphite and braided graphite fiber rings. The braided graphite rings help as antiextrusion or wiping rings and prevent graphite plateout on the stem.
2. Use sacrificial zinc washers where possible or hard chrome plating or stem materials resistant to pitting attack. Some packing include a zinc powder as anodic protection for the stem.
3. Carefully torque the packing flange nuts to the minimum torque recommended by the valve manufacturer. Overtorque will result in excessive valve stem friction and may even lock the stem.
4. Remove the packing while the valve is in storage or out of service for extended time periods.
5. Clean graphite plating from the stem before installing new packing.
6. Avoid trapping air between the rings during installation. Leave the ring level with the chamfer of the stuffing box cavity and install the next ring on top.
7. "Breaking in" new packing by tightening the gland in gradual steps while cycling the valve at least 10 times may help. A break-in lubricant applied to the packing rings and stem (e.g., nickel antiseize or silicone grease) is also recommended.

One of the oldest methods for reducing stem friction with graphite or any braided packing (such as asbestos) involved a lubricator fitting that was added to the bonnet. In this design, a compatible grease compound is injected into the *lantern ring* area by a packing lubricator assembly (Figure 6.19u). A loading

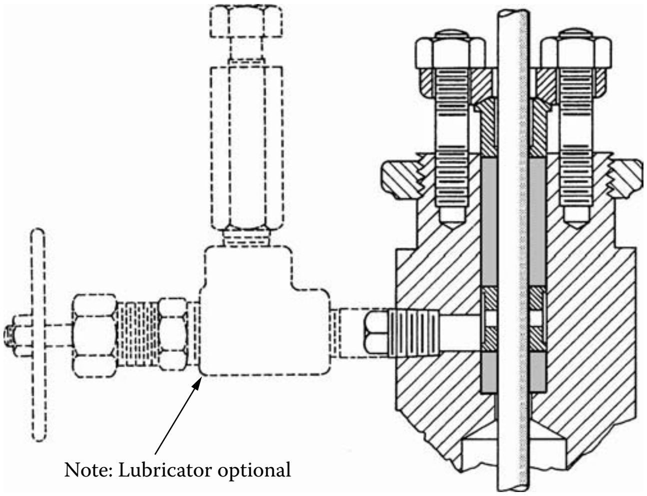


FIG. 6.19u
Stuffing box assembly with external lubricator.

bolt is turned to force the grease sticks into the packing. An isolating valve should be used for safety. External lubrication works reasonably well, although it is cumbersome and requires constant maintenance for checking and reloading of the lubricator. Also, it is sometimes difficult to find lubricants compatible with the process fluid.

Packing Arrangements There are a number of packing arrangement systems in use to suit various types of packing and fluid containment problems (Figures 6.19v). In general, the arrangements are equivalent for either Teflon V-rings or square rings (braided and laminated), although the number of rings will vary for each packing type and manufacturer's sealing system. In a standard packing arrangement, a lower set of rings serves to minimize ingress of solids from the process into the stuffing box and acts as a stem wiper. The upper set provides the primary seal and consists of several rings.

For harder-to-hold fluids or when a leak detection port, bonnet purge port, or lubricator is used, a packing arrangement known as the twin-seal is often used (Figures 6.19w). A variation of the twin seal arrangement is also recommended for vacuum service. In the twin seal, there are two full sets of packing installed. This system requires a bonnet with deeper stuffing box, because space is needed for spacers or lantern rings and a lower guide bushing as well as the packprimary stem seal. However, in vacuum service, the upper set is inverted with the V-opening toward the atmospheric side as the primary seal, because positive pressure is from atmosphere to the negative pressure inside of the valve.

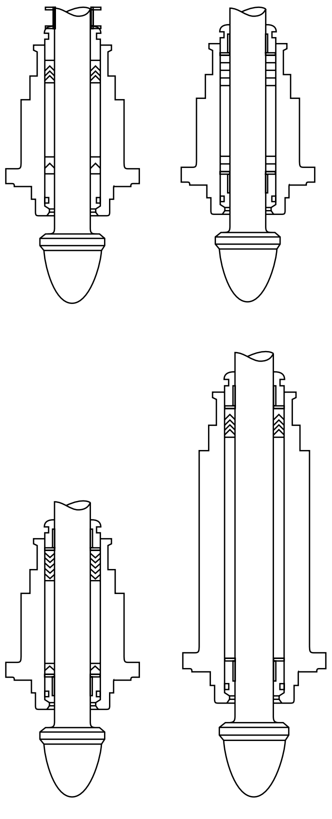


FIG. 6.19v
Typical standard packing arrangements. (Courtesy of Flowserve Corporation.)

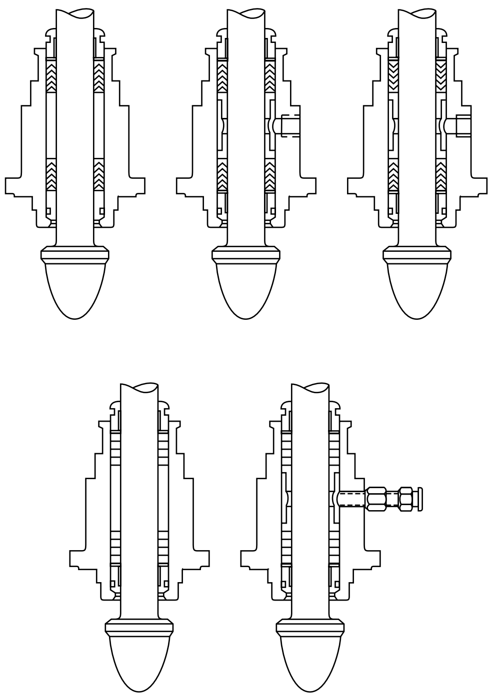


FIG. 6.19w
Typical twin- or double-packing arrangements. (Courtesy of Flowserve Corporation.)

For toxic or radioactive fluid services, the bonnet can be tapped at the lantern ring area, and this connection can be used in three different ways, depending upon specific design requirements. 1) The tap can be used for a leakage monitor connection using a pressure gauge or switch. 2) It can be used for sampling the process. 3) More commonly, the tap is used as a leak-off connection, whereby any process leakage is piped to a disposal header or nonpressurized waste container.

In cases where the fluid is a slurry or tends to solidify, the tap can be used as a purge connection to keep process media out of the bonnet. Here, an inert gas or liquid (depending upon the process) is introduced at a pressure well above the highest expected process pressure. This purge material provides an additional pressure seal, and any packing leakage will be in the direction of purge into the process or from the purge to atmosphere.

Environmental Packing The need for more aggressive protection of the environment against volatile organic chemicals and clean air regulations in Europe (e.g., TA-Luft standards), United States (e.g., EPA standards), and most other countries have motivated the development of special packing configurations and

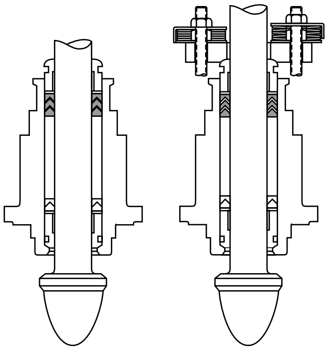


FIG. 6.19x
Typical configurations of environmental (fugitive emissions) packing. (Courtesy of Flowserve Corporation.)

materials. These regulations generally establish compliance below benchmark levels varying from 500–100 parts per million (ppm) of VOC. Figure 6.19x shows only two of many different environmental packing system designs.

Environmental packing systems often consist of a combination of ring designs and different materials. For low to moderate temperatures, a common packing set can have unfilled PTFE or chemically inert elastomer V-rings stacked between stiffer thermoplastic or reinforced PTFE chevrons for a tighter and more durable seal. For higher temperatures, various combinations of graphite braided rings with laminated graphite rings (square, wedged, or cup-and-cone) are used. Any of these configurations can include live-loading spring arrangements to help maintain constant packing loads and make up for consolidation or cold-flow of the packing over time. Antiextrusion rings are also used in some systems to promote longer packing life. Most valve manufacturers offer a variety of environmental packing styles to suit a wide range of applications.

Other Packing Materials Elastomeric O-rings are used for stem sealing in some globe valve designs, usually for less low-pressure, low-temperature utility services such as water. O-ring stem packing tends to be more common for rotary valves than for sliding stem valves. Proper selection of an elastomer that is compatible with the process fluid chemistry and temperature can also be a problem. In gas service or with liquids containing gases, many elastomers will absorb some gas. Depressurizing the valve will result in sudden expansion of this gas. If the gas cannot diffuse rapidly out of the elastomer, the resulting internal pressure can rupture or split the O-ring. This effect is called explosive decompression (ED). Most O-ring manufacturers now offer special grades of some elastomers that are resistant to ED failures.

The O-ring gland design requires a fully retained groove and probably backup rings to minimize extrusion or blowout. Installation of O-rings requires great care to avoid cutting, twisting, or other damage. Lubrication may be needed to overcome sliding friction, and gradual loss of this lubrication in service can cause excessive O-ring wear or sticking in dry service. A standard stem seal packing ring system is usually preferred in sliding stem valves.

Packing Temperature The relationship between the process and the packing temperature is a function not only of the type of bonnet used but also of the valve metallurgy and the valve and bonnet physical relationship. Heat is transmitted to the packing area by conduction through the metal, by convection via the process fluid, and by the relative heat radiation balance with the ambient environment. Stainless steel, for example, has a much lower heat conductivity coefficient than carbon steel and thus is about 20–30% less "efficient" in conducting heat into the packing area. This does not mean that the packing temperature rating can be increased, but it may serve to reduce some heat load and increase packing life. (See Figure 6.1o in Section 6.1 for further discussion and illustration of packing temperature determination.)

In some cases, the valve can be installed upside down so that the bonnet is below the valve body. In liquid service, this reduces convection, and heat is transferred to the bonnet mainly by conduction. In vapor service (with or without superheat), it may be possible in the stem-down orientation to condense sufficient vapor in the bonnet to create a condensate seal and lower the packing temperature to a suitable level. The potential corrosive effects of alternate wetting and drying must be considered with this approach. However, this approach to controlling packing temperature has serious disadvantages where maintenance is concerned, and the benefits generally do not justify the potential maintenance difficulties except in unusual cases.

BODY FORMS

The actual pressure containment and fluid conduit portion of a control valve is called the valve body assembly. This assembly consists of the body, a bonnet or top closure, sometimes a bottom flange closure, and the internal elements known as the trim. The body can have flanged, threaded, or welded end (butt weld or socket weld) connections for installation into the piping. The trim consists of such elements as the plug and stem, guide bushings, seat rings, cages or seat retainers, and stuffing box lantern ring.

The body configuration can be in-line, angle, offset in-line, Y-type, and three-way. Some of these will be discussed in more detail.

The shape and style of the valve body assembly is usually determined by the type of trim elements it contains, piping requirements, and the function of the valve in the process system. There are a large number of body designs on the market, including a number of special-purpose designs. The end result is a device that can be fitted with an actuator and used to modulate the flow of process fluid to regulate such things as pressure, flow, temperature, liquid level, or any other variable in a process system. Examples of some of the more widely used body configurations are shown in Figure 6.19y.

Double-Ported Valves

The double-ported (double-seated) balanced valve (Figure 6.19z) was one of the first globe valves developed during the early 20th century. It is still available today, but has been replaced in most applications by single-seated globe valves. Size for size, it is much larger and heavier than its single-seated counterpart. Shut-off is poor because it is not practical to have both plugs in tight contact with the seats at the same time, but the valve was intended for throttling control rather than for tight shut-off. Some special seat designs have been developed to help overcome this, but application is limited.

The double-ported valve is considered semibalanced; i.e., the hydrostatic forces acting on the upper plug partially cancel out the forces acting on the lower plug. The result is less actuator force requirement, and a smaller actuator can be used. However, there is always an unbalanced force due to the difference between the upper and lower plug diameters required for assembly. In addition, unbalance forces are generated by the effect of dynamic fluid forces acting on the respective throttling areas of each plug. Such forces can be quite high, particularly with the smooth contoured plugs; these can reach as much as 40% of the forces of an equivalent unbalanced single-seated valve plug. Double-ported valves have been built in sizes up to 24 in. (DN 600), although most manufacturers now limit them to 12 in. (DN 300) as a maximum.

Figure 6.19z shows that the valve can be converted from the push-down-to-close configuration shown to a push-down-to-open design. This is done by removing the bottom closure flange, bonnet, and stem, and by inverting the entire assembly and reinstalling the stem, flange, and bonnet. This, coupled with the use of direct-acting (air pushes the stem down) and reverse-acting (air pushes the stem up) actuators, gives full flexibility to provide the required valve failure mode.

Single-Seated Valves

Single-seated valves are the most widely used of the globe body patterns. There are good reasons for this. They are available in a wide variety of configurations, including special-purpose trims. They have good seating shut-off capability, are less subject to vibration due to reduced plug mass, and are generally easy to maintain. There are three general types of seat construction.

1) The *floating seat ring* sits in machined bore in the body with a gasket to seal the joint between the body and seat ring. It must be retained in the body by a cage or seat retainer to maintain gasket tightness and concentricity with the plug. 2) The *screwed-in seat ring* is threaded into matching body threads with a special tool; a gasket may or may not be required. A separate seat retainer is not required, but some designs use a cage for guiding the plug or characterizing the flow. 3) The seating surface can be machined directly into the body; this is called an *integral seat*. The floating and screwed-in seat rings can be replaced after they wear out, but the integral seat requires resurfacing or machining of the body to repair wear or damage.

Single-seated valve plugs are guided in one of four ways: post-guided (Figure 6.19aa), top-and-bottom-guided (Figure 6.19bb), stem-guided (Figure 6.19cc), and cage-guided (Figure 6.19dd). The most popular globe valves are stem-, cage-, or post-guided types, which require only one body opening for the bonnet and have one less closure gasket subject to leakage than the top-and-bottom-guided configuration.

The stem-guided and post-guided designs provide more streamlined flow and are less subject to fouling in dirty service. The stem-guided valve minimizes stagnant fluid cavities and may be a better selection than the post-guided and cage-guided valves when dealing with fluids containing solids, sticky or viscous fluids, or highly corrosive fluids. The top-and-bottom single-seated valve (Figure 6.19bb), like its double-seated counterpart (Figure 6.19z), has similar limitations, but some users still prefer this design where the plug is held by two, widely spaced bushings. The top-entry single-seat globe valve is most commonly used in sizes from 1 in. (DN 25) through 12 in. (DN 300) from most manufacturers.

Some top-entry designs are manufactured with bodies suitable for slip-on flanges (Figure 6.19y) rather than with integral cast flanges. This type of flange construction is discussed in more detail below, under Split-Body Valves and Valve Connections.

Cage Valves The cage valve is a variant of the single-seated valve and is the most popular design used in the process industries. The top-entry bonnet and trim design makes it extremely easy to change the trim or to do maintenance work. Cages are used with floating seat rings and with screwed-in seat rings. The design is very flexible in that it allows a variety of trim types to be installed in the body. This includes such variations as reduced trim, anticavitation, and low-noise trims (Sections 6.1 and 6.14).

The overall design is very rugged, and with proper specification of trim type and materials, cage valves provide relatively trouble-free service for extended time periods. These valves may eliminate the need for block and bypass valves in some cases, because their service life can be as good as or better than most other components in the process that require periodic maintenance.

There are two basic design configurations available for cage valves. One type uses the cage solely as a seat retainer to clamp a floating seat ring into the valve body (Figure 6.19cc). This design is usually stem-guided or post-guided, and the valve plug is characterized and does not guide or control flow through the cage.

The other type uses the cage to guide the plug head, and the cage openings are shaped to provide the desired flow characteristic as the valve plug exposes the ports (Figure 6.19dd), and it is called a *cage-guided valve*. In the cage-guided valve,

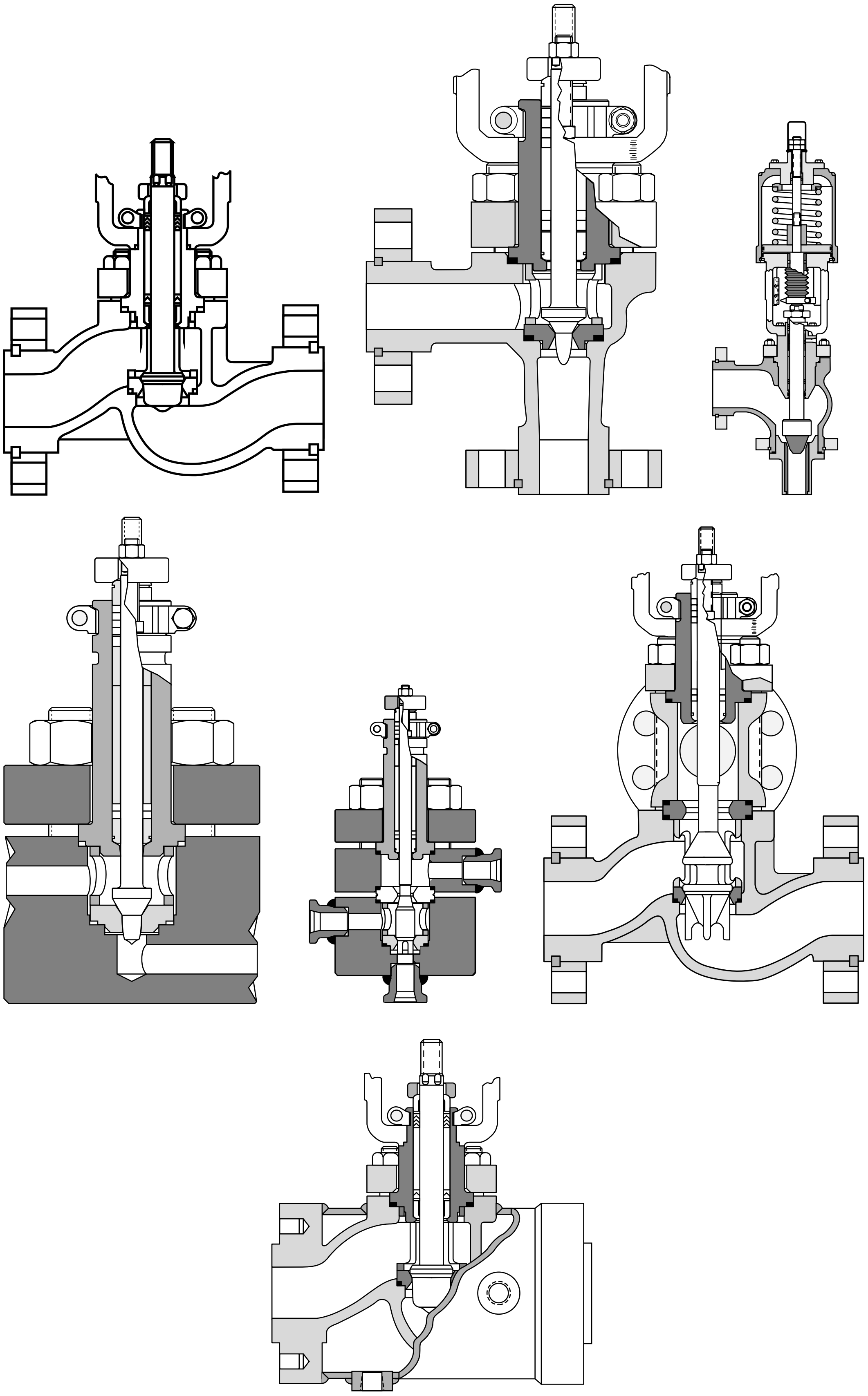


FIG. 6.19y
Typical globe body configurations. (Courtesy of Flowserve Corporation.)

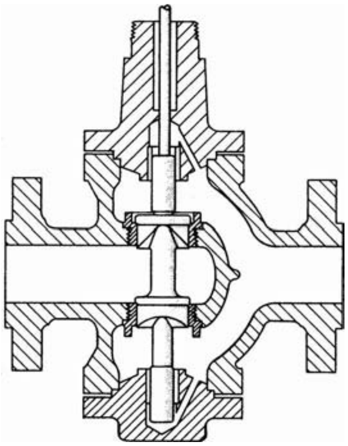


FIG. 6.19z
Top- and bottom-guided invertible double-seated globe valve.

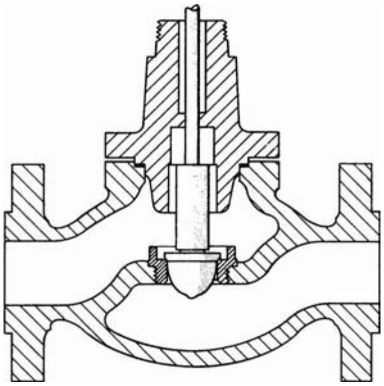


FIG. 6.19aa
Top-entry, post-guided plug, threaded single-seated globe valve.

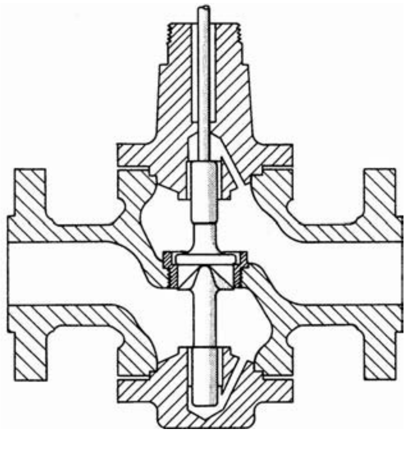


FIG. 6.19bb
Top- and bottom-guided invertible single-seated globe valve.

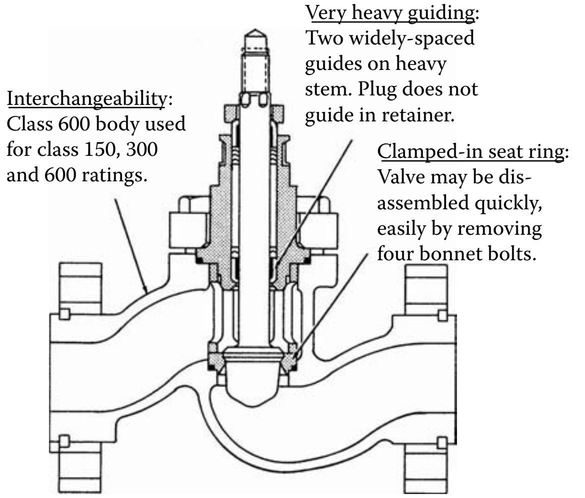


FIG. 6.19cc
Cage valve with clamped-in seat ring and characterized plug. (Courtesy of Flowserve Corporation.)

the cage may be used to clamp a floating seat ring, or it may be used with a screwed-in seat where the cage is part of the bonnet assembly. Both of the above cage designs are usually provided in smaller valve sizes with unbalanced plugs for best shut-off.

The stem-guided design (Figure 6.19cc) has advantages when handling fluids with solids, sticky fluids, fluids that coat or plate out onto the trim, and corrosive fluids. This is because the large plug-to-cage clearances in the stem-guided design are not sensitive to debris or build-up, and corrosion-resistant trim materials (316 stainless steel and nickel alloys) are more practical to use without special treatments to prevent metal-to-metal galling.

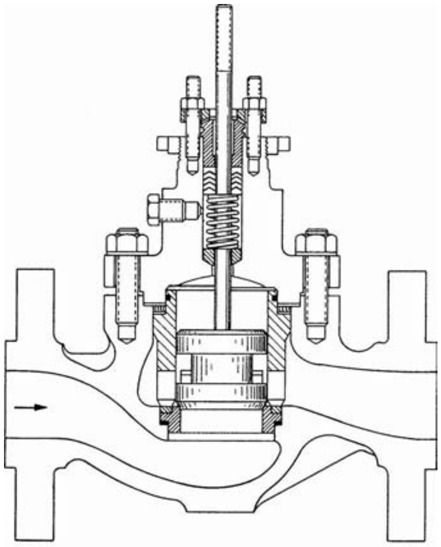


FIG. 6.19dd
Cage valve with unbalanced plug and characterized cage ports. (Courtesy of Emerson Flow Management.)

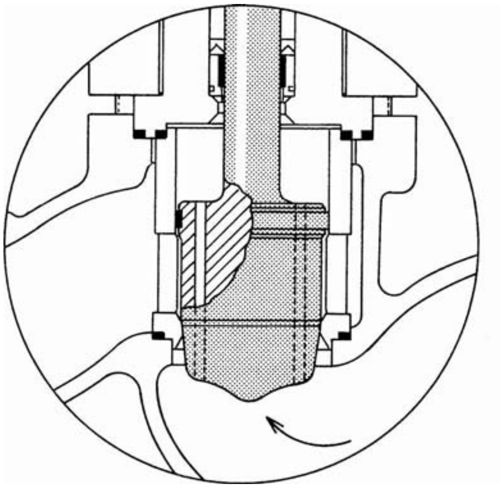


FIG. 6.19ee
Cage valve with balanced and characterized plug. (Courtesy of Flowserve Corporation.)

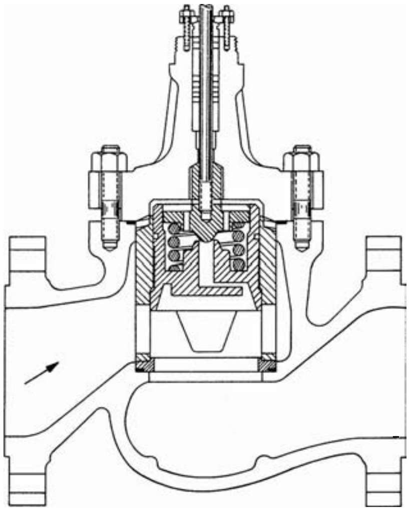


FIG. 6.19gg
Cage valve with pilot-balanced construction and characterized cage ports. (Courtesy of Dresser Flow Solutions.)

Tight clearances are required in the cage-guided version, which requires that metal surfaces must be hardened by heat treatment, hard-faced with Co-Cr or Ni-Cr-B alloys, plated with hard chrome or electroless nickel, or surface treated (e.g., nitriding) to eliminate metal-to-metal galling.

Pressure-Balanced Valves Pressure-balanced trims are also available in stem-guided (Figure 6.19ee) and cage-guided (Figure 6.19ff) versions.

Pressure-balanced designs help provide better control of high-pressure drops, reduce the magnitude of unbalanced plug forces, and help to reduce actuator size. The balanced plug designs can be provided with a variety of balance seal styles and materials to meet service conditions. These may be plastics such as PTFE, elastomeric O-rings, and metal piston rings. Generally, the use of balanced plugs will degrade the valve shut-off capability to some degree. For example, an unbalanced valve rated for Class V shut-off may drop to Class IV with balanced trim. This is due to leakage past the balance seal. (Refer to the discussions on leakage in this section and in Section 6.1.)

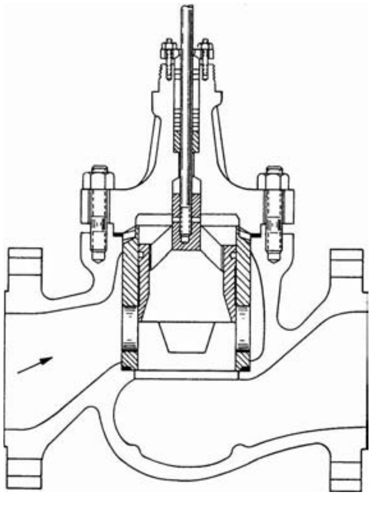


FIG. 6.19ff
Cage valve with balanced plug and cage port characterization. (Courtesy of Dresser Flow Solutions.)

One variation of the pressure-balanced plug that has improved shut-off capability is shown in Figure 6.19gg. This is called the *pilot-balanced* plug. This design is particularly helpful when dealing with high-temperature, high-pressure drop situations when tight shut-off is difficult to achieve without the use of elastomeric seals. When the valve is shut off, the actuator has compressed the pilot spring and closed the pilot.

There is provision for a pressure path from the upstream pressure side of the plug to the area above the pilot plug, allowing upstream pressure on the upper plug area to assist the actuator in achieving a tight shut-off. In the shut-off position, there is no pressure drop across the conventional balance seals. When the actuator begins to open the valve, the stem first lifts the small pilot plug off its seat. This allows the pressure above the plug to vent to the downstream side, and the valve operates as a conventional pressure-balanced valve. The spring shown in this particular design holds the pilot valve off the pilot seat during normal operation. When the valve begins to close for shut-off, the plug first seats off on the valve seat, and a small amount of additional stem travel compresses the spring and seats the pilot valve.

Special Designs Many customized, special-purpose designs are possible to meet unique service and operating requirements. The example in Figure 6.19hh shows a cage and seat that are installed from the bottom of the valve and supported by a bottom flange. This design was intended as a "flangeless"

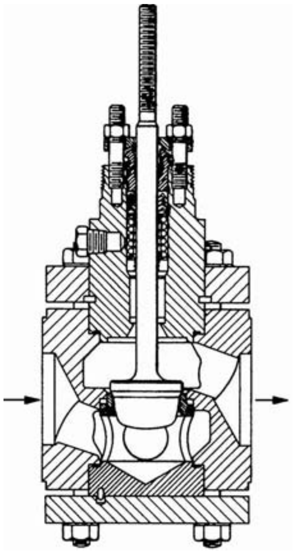


FIG. 6.19hh
Quick-change trim valve designed for bottom trim removal. (Courtesy of Emerson Process Management.)

valve for small sizes of forged or bar stock corrosion-resistant alloys such as Alloy 20, nickel, and Hastelloy C-276 for corrosive service. It allows for a bottom drain and quick inspection of the valve trim without removal of the actuator and bonnet. Similar bodies can be machined in custom configurations and may be more economical in small sizes for special alloys than other globe designs.

Split-Body Valves Another type of single-seated globe body is the split-body valve (Figure 6.19ii). This design was the original stem-guided chemical service valve intended for hard-to-handle services involving slurries, sticky fluids, and corrosive services. The seat ring is clamped between the body halves, and the body is easily disassembled for replacement of the seat and plug.

Another feature is the adaptability for building the body to use slip-on flanges. This results in cost savings when corrosion-resistant alloy castings or forgings are required for the wetted body. Because the flanges are not normally subject to wetting by the process fluid, they can be of carbon steel or lower alloys than the body. In small sizes, it is economical to cast or forge the body for a standard ANSI Class 600 rating and install Class 150, 300, or 600 slip-on flanges as needed. It is also possible to rotate the lower body 90 degrees to the line axis and eliminate a pipe elbow, although this is rarely done. The design of the split body makes it possible to be molded in structural plastics for low-pressure corrosive applications. Refer to the discussion on lined and thermoplastic valves below.

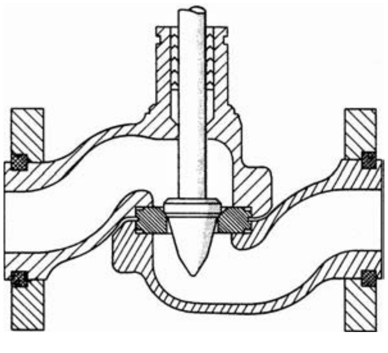


FIG. 6.19ii
Split-body valve with separable end flanges. (Courtesy of Dresser Flow Solutions.)

While still popular for some applications, split-body valves have limitations on installing special trim options. The bolted body joint may leak if exposed to large piping stresses from severe thermal cycling, and it should not be welded into the piping because maintenance requires body separation.

Split body valves are available with flanged ends up to size 10 in. (DN 250), but separable (slip-on) flanges are only available up to size 4 in. (DN 100), and threaded ends are available up to size 2 in. (DN 50). Availability is generally limited to small sizes in high-pressure classes. ANSI Class 900, 1500, and 2500 ratings (PN 150, PN 250, and PN 420) are in sizes 2 in. (DN 50) and smaller.

Angle Valves Angle valves (Figure 6.19y) are often used in a flow-to-close direction for high-pressure drop service. This is favorable to the valve body and trim but requires careful design of the downstream piping to avoid erosion problems in high-velocity liquid or two-phase flow. Depending on actual downstream velocities, these applications can require a larger pipe size than the valve and up to 20 diameters of straight pipe before the first elbow.

Angle valves are also used to accommodate special piping arrangements to aid drainage, on erosive services to minimize solids impingement problems, and on other special applications such as coking hydrocarbons. Figure 6.19jj shows a coking valve design with a streamlined, sweep-angle flow path and a replaceable venturi outlet sleeve. This body flow path is designed to reduce flow velocity in the body to minimize erosion.

Trim materials such as tungsten carbide and ceramics can be selected to resist erosion due to the higher trim velocities. The venturi-style outlet has high-pressure recovery and low vena contracta pressures, which makes the trim and downstream pipe highly susceptible to cavitation on liquid service, even with moderate pressure drops. In order to avoid damage to piping, these valves can be installed directly on vessel inlets or with a larger downstream pipe to mitigate the effects of cavitation or flashing.

Y-Type Valves The Y-type valve has application in several special areas. Among these applications are those where good drainage of the body passages or high flow capacity is required, such as in controlling molten metals or polymers, cryogenic fluids, and liquid slurries. Cast or forged Y-type valves are available up to size 16 in. (DN 400) and with pressure ratings up to Class 4500 (PN 760). The valve can be installed in

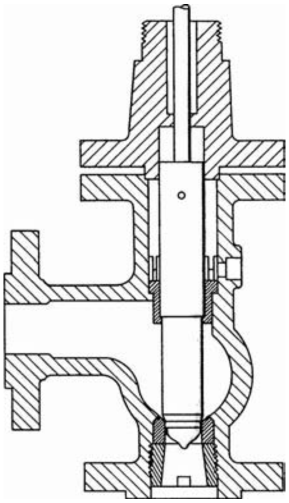


FIG. 6.19jj
Streamlined (sweep) angle valve with lined venturi outlet.

horizontal, vertical, or angled piping to suit the application requirements. Because of the simple body shape and bonnet design, they are easy to fit with thermal or vacuum jackets.

Figure 6.19kk shows a Y-type valve with a vacuum jacket that has been specially designed for low-pressure cryogenic liquid service. The vacuum jacket provides maximum thermal insulation for such applications as liquid hydrogen. This body is designed for minimum and uniform wall thickness, which enables rapid cool-down rates. The single-seated design allows good shut-off and can be provided with soft inserts for exceptionally tight shut-off. Note that the vacuum jacket is designed with a metal bellows to allow for mechanical tolerances or dimensional changes during temperature transients and has a provision for welding to a matching jacket around the adjacent piping.

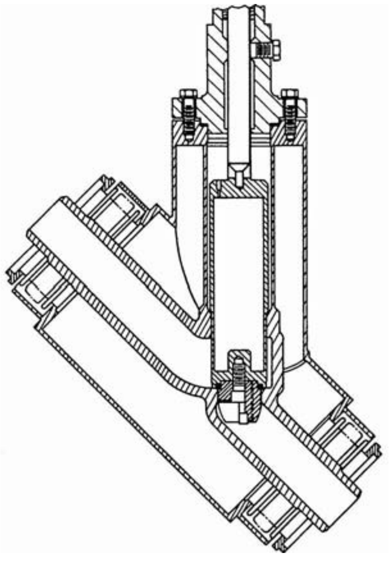


FIG. 6.19kk
Y-pattern valve fitted with vacuum jacket for cryogenic service.

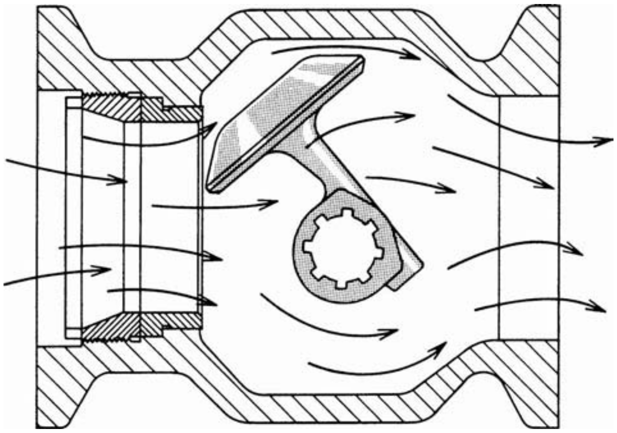


FIG. 6.19ll
Rotary globe with eccentric spherical segment plug. (Courtesy of Dresser Flow Solutions.)

Eccentric Plug Rotary Globe Type The rotary globe valve (Figure 6.19ll) is mentioned here, because its design and performance combine features of conventional rotary valves with conventional globe valves. Compared to most rotary valves, the eccentric spherical segment plug valve has the advantage of lower torque requirements. The seating surface of the plug has the form of a spherical segment. The plug design makes exaggerated use of the offset center to obtain contact at closure without rubbing on the seat, like ball valves. It is capable of substantial seat contact loading for tight shut-off approaching that of conventional globe valves.

The flow characteristic approaches linear (Figure 6.19a). Changes in control characteristic can be accomplished with a cam in the positioner or by modifying the controller output signal. Capacity (C_v) is between that of a double-ported globe and a high-performance, double-offset butterfly valve. High-flow capacity is achieved with only moderate pressure recovery in the body, so the critical flow coefficient (F_L) is higher than that of a butterfly valve throughout its throttling range and therefore less likely to cavitate.

The valve is made in sizes from 1–16 in. (DN 25–400) in pressure Classes 150, 300, and 600 (PN 20, PN 50, PN 100). Flanged and flangeless versions are available in most sizes. Allowable operating pressure drops depend on the manufacturers' designs, materials, and actuators. In most cases, the valves are rated for operating pressure differentials significantly less than the body's maximum pressure rating, so it is necessary to consult the manufacturer's data for specific limitations.

C_v rangeability data are stated in ranges from 100:1 up to 160:1. However, these ratios are achievable in service only if the actuator and control system can control in the range of 1–5% open, which is possible but often not practical. From a practical control standpoint, assuming a control range from 5–95% open, the C_v rangeability for eccentric rotary plugs is about 35:1 or lower, which is comparable with conventional globe valves.

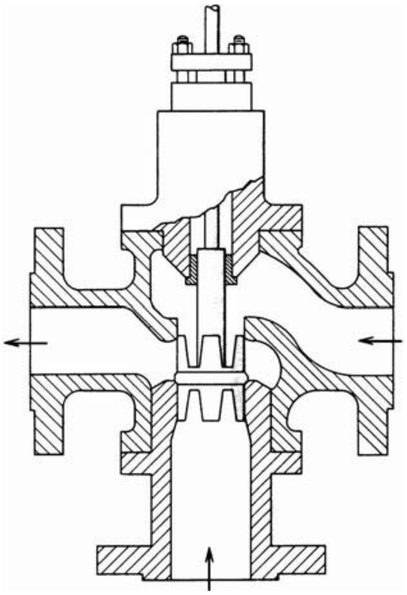


FIG. 6.19mm
Three-way valve for mixing service.

Three-Way Valves

Three-way valves are a specialized double-seated globe valve configuration. There are two basic types. One is for mixing service, i.e., the combination of two fluid streams passing to a common outlet port (Figure 6.19mm).

The other is for diverting service, i.e., taking a common stream and splitting it into two outlet ports (Figure 6.19nn). The three-way design shown in Figure 6.19y uses an adapter to convert a conventional single-seated globe into mixing or diverting configurations.

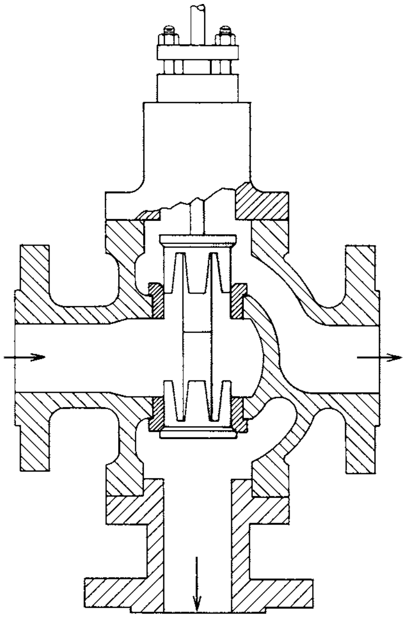


FIG. 6.19nn
Three-way valve for diverting service.

A typical mixing valve application would be the blending of two different fluids to produce a specific outlet end product. A diverting valve might be used for switching a stream from one vessel to another vessel or for temperature control on a heat exchanger. In the latter, the valve would direct one portion of the flow through the exchanger, and the balance of the flow would bypass the exchanger. The relative split would provide the heat balance needed for temperature control.

The forces acting on the double-seated, three-way plug do not balance in the same way that a double-seated, two-way plug is balanced. This is because different pressure levels exist in each of the three flow channels. Also, there are different dynamic forces acting on each plug head. Therefore, these valves are not normally used in high-pressure drop service. The valve plugs are usually seat-guided and post- or stem-guided to maintain stability under fluctuating dynamic forces. Due to the larger size and piping complications, some users prefer to use two opposite-acting, two-way globe valves operating from one controller to do the same job as a three-way valve.

Lined and Thermoplastic Valves

For extremely corrosive services, high alloy metals such as Alloy 20, nickel, nickel alloys, titanium, and zirconium are often required where austenitic stainless steels are inadequate. However, special alloy valves are very expensive and the service life may still be limited by corrosion.

An alternative to special alloys could be the lined globe valve (Figure 6.19oo). The lined valve uses a carbon steel or stainless steel shell for retaining pressure, but the inside surfaces of the body and bonnet are bonded with a chemically inert or corrosion-resistant fluoroplastic, such as PFA or PTFE. These valves are commonly for 1 in. (DN 25) through 2 in. (DN 50) sizes with capacity from $C_v = 0.33$ to 45.

For very low C_v flow capacities, the design shown in Figure 6.19pp is available in 1 in. size (25 mm) with trim selections covering a C_v range from 0.001 to 1.0. These valve designs are limited to operating temperatures below 300–400°F (149–204°C) and pressure drops less than 125–250 PSI (8.5–17 bar). Lined valves are also available in various rotary valve types and pinch valves (see Section 6.20).

Conventional globe patterns and split-body designs can be molded entirely from structural thermoplastics for corrosive applications with low to moderate pressures and temperatures. Bodies can be molded in polyvinylidene fluoride (PVDF), polyvinyl chloride (PVC), polypropylene (PP), and other plastics. Allowable pressures and temperatures vary with material and body design.

Typically, PVDF is limited to 300 PSIG (21 bar) at 73°F (23°C) and 150 PSIG (10.3 bar) at 225°F (106°C). PVC has a limit of 300 PSIG (21 bar) at 73°F (23°C) and 150 PSIG (10.3 bar) at 140°F (60°C). Polypropylene can operate up to 300 PSIG (21 bar) at 73°F (23°C) and 150 PSIG (10.3 bar) at 175°F (79.4°C). End connections are available in Class 150 flanged,

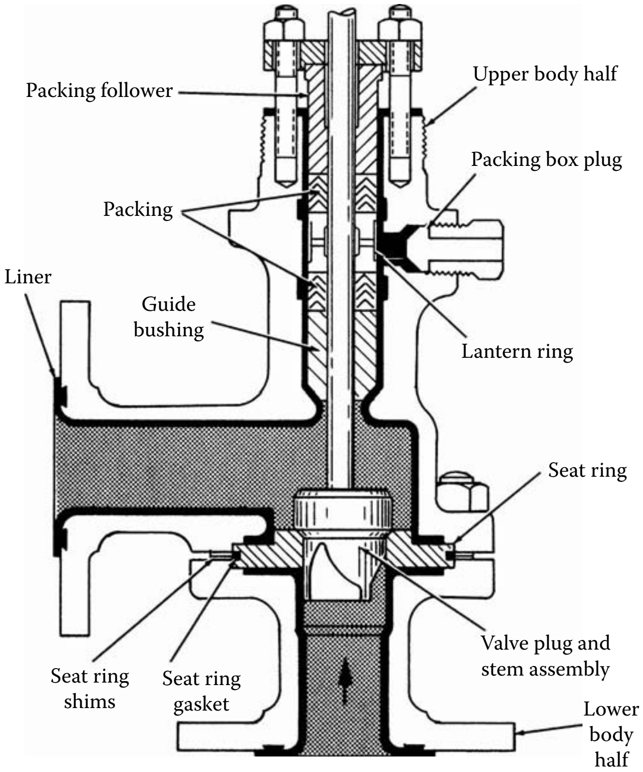


FIG. 6.19oo
Standard Teflon-lined valve with Teflon-encapsulated trim. (Courtesy of Emerson Process Management.)

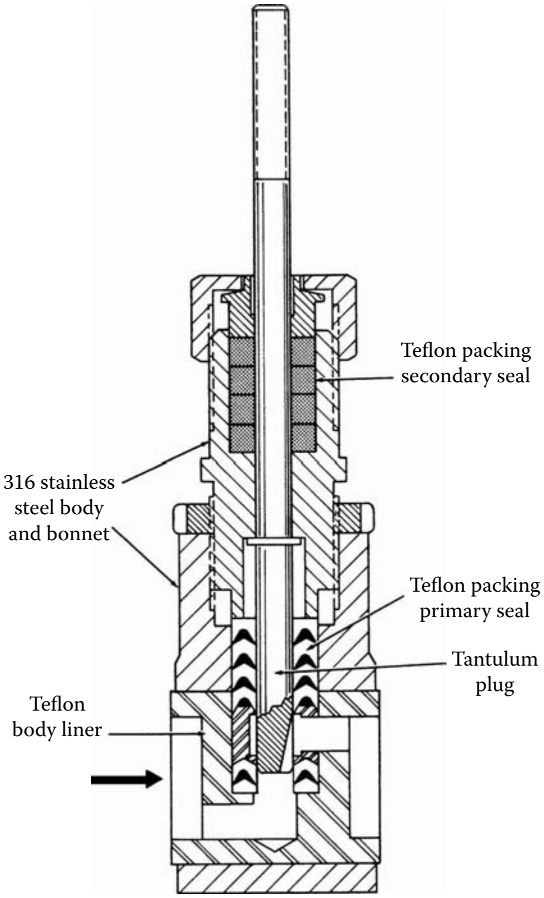


FIG. 6.19pp
Small-flow Teflon-lined valve with tantalum plug. (Courtesy of Emerson Process Management.)

flangeless (for clamping between pipe flanges with through-bolting), or threaded ends. Packing and gaskets are generally made of PTFE or perfluoroelastomer (e.g., Kalrez®) materials.

VALVE CONNECTIONS

Valve connections to adjacent piping can be categorized into four general types: 1) flanged, 2) welded, 3) threaded, and 4) clamped. Each style incorporates elements that contain pressure, bear piping loads, and seal the joint between the valve and piping.

As a general rule, valves smaller than 2 in. (DN 50) can use threaded connections, while sizes 2 in. (DN 50) and larger use flanged connections. In a few process systems where fugitive emissions or process leakage is not a problem (such as water systems), threaded connections up to size 4 in. (DN 100) have been used. Most applications require both ends of the valve to have identical connections, but on some vent and drain valves, the upstream port and the downstream port may require different sizes or types of connections. Typical end connections are discussed below.

Flanged Ends

Flanges are the most common valve connection to piping. Most countries and piping codes accept the flange design and ratings from the standard ASME B16.5, Pipe Flanges and Flanged Fittings commonly referred to as “ANSI” flanges (because the earlier standards were published by the American National Standards Insitutue). Some countries and applications use other flange standards, such as American Petroleum Institute (API) standard for oil production equipment and the International Standards Organization (ISO) ISO-7005 flanges (adopted from the German DIN standards).

These rating systems are different and are not interchangeable; they use different rating systems, bolting, gaskets, and flange dimensions. The ASME (or ANSI) flanges use a Pressure Class rating system (e.g., steel flanges are Class 150, Class 300, Class 400, Class 600, Class 900, Class 1500, and Class 2500). API flanges are designated by a nominal pressure system (e.g., 2,000 PSI, 3,000 PSI, 5,000 PSI, 10,000 PSI, 15,000 PSI, and 20,000 PSI). The ISO (and DIN) standards also use a nominal pressure or PN numbering with typical ratings PN 10, PN 16, PN 20, PN 40, PN 64, PN 100, PN 160, PN 250, PN 320, and PN 400, which indicate maximum nominal pressure in bar (1 bar = 14.504 PSI = 100 kPa).

Flanges may be flat face, raised face, ring-type joint (RTJ), tongue and groove, male and female, or other configuration to suit the application. Cast iron, ductile iron, and bronze are usually flat face; carbon steel and alloys are usually raised face; and above ANSI 600, the RTJ is fairly common.

Figure 6.19qq shows three common configurations of flanges: separable flange-raised face, integral flange-raised face, and integral flange-RTJ. Raised face and flat face flanges use gaskets from sheet stock, such as graphite or PTFE composites, or a spiral wound thin metal strip with graphite, PTFE, or other mineral filler between each metal winding. RTJ gaskets are oval or octagonal in cross-section made of any suitable metal softer than the flange.

Valves with separable (slip-on) flanges retain the flange on the valve body ends with two circular half-rings held in the body grooves. Separable flanges are used mainly on small sizes from NPS $^1/_2$ in. up to 4 in. With separable end flanges, the body can be designed for an ANSI Class 600 rating and then adapted as needed with ANSI Class 150, 300, or 600 flanges.

Separable flanges are not wetted by the process, which permits less expensive carbon steel or stainless steel flanges to be used on high alloy valves. Separable flanged valves are easy to install with the mating piping, because the flanges can be rotated to fit mismatched line flange hole patterns. However, care must be taken during installation to prevent the valve from rotating in the line until the flange bolts are properly tightened.

Flangeless bodies (clamped between pipe flanges) are sometimes used in the small bar-stock bodies. While they permit lower cost where expensive alloys are involved, they require care with bolting, gasket, and piping alignment. The tie-rod bolting should be high tensile strength material, and the valve must be carefully centered to permit proper gasket sealing and loading. The longer the bolt studs are, the more they are affected by longitudinal thermal expansion differences with the valve body and piping, which can lead to gasket leakage in thermal shock or severe temperature cycling.

Welded Ends

Welded ends are not common in the chemical process industries, but they are generally recommended in power generation and other applications where high piping stresses or thermal shock conditions exist. They are also used in hazardous services when no leakage is allowed from gasketed joints. Socket-weld ends (Figure 6.19rr) are easy to align and may be used in small sizes of 2 in. (25 mm) or under.

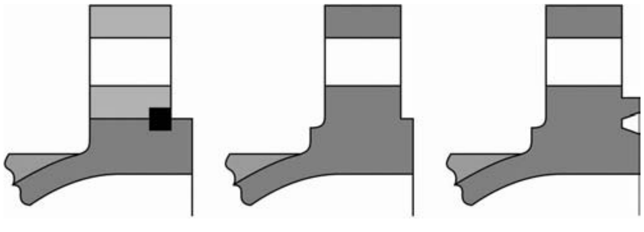


FIG. 6.19qq
Flange configurations: separable raised-face, integral raised-face, ring-type joint (RTJ) integral flange. (Courtesy of Flowserve Corporation.)

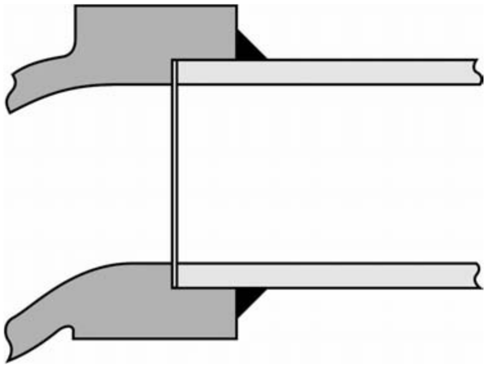


FIG. 6.19rr
Socket-weld end valve and pipe joint with fillet weld. (Courtesy of Flowserve Corporation.)

However, most valves are installed with butt-weld ends where maximum joint strength is achieved by full penetration of the weld (Figure 6.19ss). Butt-weld joints are commonly checked by radiographic inspection (X-ray) to confirm full penetration and lack of defects such as cracks, voids, and slag.

The valve body material should be selected so that it is suitable for welding to the adjoining piping. In many cases the valve and adjacent pipe require a post-weld, stress-relieving heat treatment. The mating ends of the body and pipe are usually machined at an angle to form a V-joint for ease of welding. The design of joint (end-preparation) depends on the thickness of the mating parts. ASME B16.25 is the most common standard for design of the butt-weld end preparation, but some users and engineering contractors choose their own dimensions for the end preparation to suit their particular fabrication practices.

Threaded Ends

Threaded connections with NPT (tapered pipe threads per ASME B1.20.1) threads are common on valve sizes 1 in. (DN 25) and smaller and are sometimes used on valves up to 2 in. (DN 50). Threaded connections are usually used in pressure ratings up to ANSI Class 600 (PN 100). The threaded connection normally consists of the valve body with female NPT threads; then the pipe with male NPT threads is screwed into the body end. Materials commonly used in

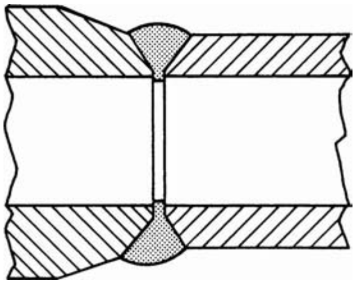


FIG. 6.19ss
Butt weld end valve and pipe joint showing full penetration weld.

threaded end valves are brass, bronze, and carbon steel. Stainless steel bodies usually do not have NPT threads, because they gall easily with stainless steel pipe. An antigalling thread sealant compound should be applied to the threads during installation to prevent leakage and seizing and to allow for future disassembly.

Threaded connections are not normally used with process valves because they can leak more easily, they are more difficult to make up in the piping, and they may corrode or gall together, making disassembly difficult. Threaded connections do not work well on corrosive fluids, because the corrosion attack is generally more severe in the threaded crevice, causing the threads to corrode together and making it difficult to disassemble. Accelerated corrosion in the threads also promotes leakage. Threaded joints also tend to loosen in services where the temperature fluctuates.

Threaded ends are not recommended for services with severe thermal shock or thermal cycling that occurs above 500°F (260°C) or below −50°F (−46°C). However, some small valves under size 2 in. (DN 50) are furnished with threaded end connections for utility services such as low-pressure steam, water, and gas. Where it is necessary, threaded end valves can be converted to flanges by welding a flange to a pipe nipple, threading it into the valve, and seal-welding the nipple to the body.

Special End Fittings

In very high pressures, usually above 5000 PSIG (345 MPa), some operating companies use proprietary fitting and flange designs. One widely used high-pressure connection design is the lens-ring-type fitting shown in Figure 6.19tt. This is a self-energizing type of seal; i.e., the lens-ring deforms to give a tighter seal with increase in-line pressure. (Further details are given in Section 6.1.)

In other high-pressure applications, special clamped fittings are very effective. One proprietary clamped fitting called the Grayloc® fitting (developed by the Gray Tool Company) greatly reduces the bulk of the valve end as compared to the large high-pressure, bolted flanges. The Grayloc fitting utilizes a lens-ring-type seal, similar to one shown in Figure 6.19tt, but the special hub design is fastened together with a bolted clamp-type fitting instead of flanges. See Figure 6.19uu.

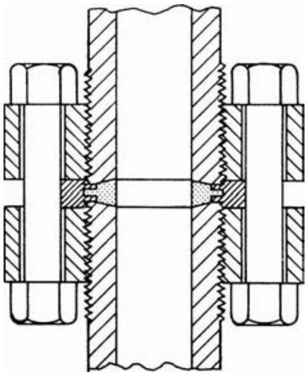


FIG. 6.19tt
High-pressure lens-ring-type joint.

These fittings are also available in 10,000 and 15,000 PSIG (69 and 104 MPa) designs. The ASME Boiler and Pressure Vessel code permits the use of clamped connections and includes design rules in Section VIII, Division 1, Appendix 24.

For standards on globe valve end connections, see Table 6.19vv.

MATERIALS OF CONSTRUCTION

A valve body assembly is a pressure containment vessel. As such, design and material selection must follow guidelines of a pressure vessel code that is recognized by the country or state in which it will be operated. Some of the

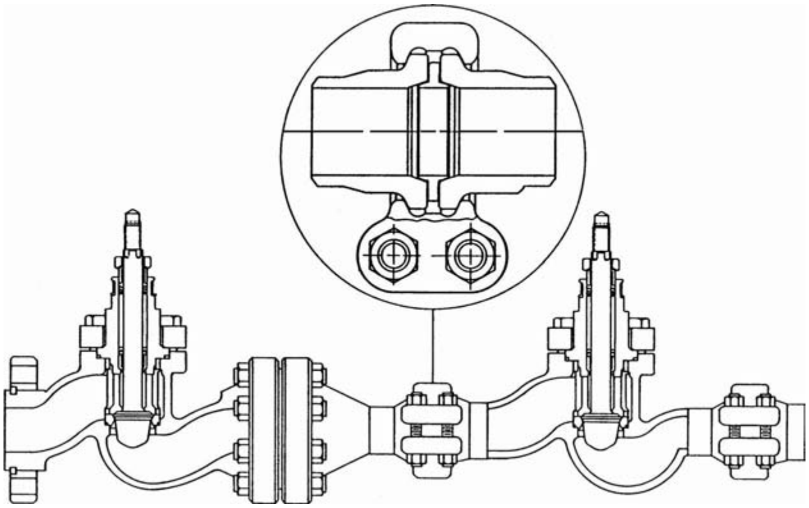


FIG. 6.19uu
Grayloc end connection detail and comparison with high-pressure flange connection. (Courtesy of Grayloc Products.)

TABLE 6.19vv
End Connection Standards used for Globe Valves

Type	*Standards Organization*	*Designation**	*Title*
Flanged	ASME International	ASME B16.5	Pipe Flanges and Flanged Fittings (NPS $^{1}/_{2}$ through NPS 24)
		ASME B16.47	Large Diameter Steel Flanges (NPS 26 through NPS 60)
		ASME B16.1	Cast Iron Pipe Flanges and Flanged Fittings
		ASME B16.42	Ductile Iron Pipe Flanges and Flanged Fittings, Class 150 and 300
		ASME B16.24	Cast Copper Alloy Pipe Flanges and Flanged Fittings, Class 150, 300, 400, 600, 900, 1500, and 2500.
	American Petroleum Institute (API)	API Spec 6A	Specification for Wellhead and Christmas Tree Equipment
	European Committee for Standardization (CEN)	EN 1092-1	Flanges and Their Joints: Circular Flanges for Pipes, Valves, Fittings, and Accessories, Part 1: Steel Flanges PN Designated
		EN 1092-2	Flanges and Their Joints: Circular Flanges for Pipes, Valves, Fittings, and Accessories, Part 2: Cast Iron Flanges PN Designated
		PREN 1759-1	Flanges and Their Joints: Circular Flanges for Pipes, Valves, Fittings, and Accessories, Class Designated – Part 1, Steel Flanges NPS $^{1}/_{2}$ to 24
	International Standards Organization (ISO)	ISO 7005-1	Metallic Flanges – Part 1: Steel Flanges
		ISO 7005-2	Metallic Flanges – Part 2: Cast Iron Flanges
		ISO 7005-3	Metallic Flanges – Part 3: Copper Alloy and Composite Flanges
		ISO 10423	Petroleum and Natural Gas Industries – Drilling and Production Equipment – Wellhead and Christmas Tree Equipment
	Manufacturers Standardization Society (MSS)	MSS SP-6	Standard Finishes for Contact Faces of Pipe Flanges and Connecting End Flanges of Valves and Fittings
		MSS SP-44	Steel Pipe Line Flanges
	ASME International	ASME B16.25	Butt Welding Ends
Welded	European Committee for Standardization (CEN)	ASME B16.11	Forged Fittings, Socket-Welding and Threaded
		EN 12627	Industrial Valves – Butt Welding Ends for Steal Valves
		EN 12760	Valves – Socket Welding Ends for Steel Valves
Threaded	ASME International	ASME B1.20.1	Pipe Threads, General Purpose (Inch)
		AMSE B1.20.3	Dryseal Pipe Threads (Inch)
		ASME B16.11	Forged Fittings, Socket-Welding and Threaded
Clamped	ASME International	ASME B&PVC Sec.VII-Div. 1-Appendix 24	Design Rules for Clamp Connections

* Always use the current edition of any standard, unless governmental or construction specifications require otherwise.

prominent standards include the ASME Unfired Pressure Vessel Code (Section VIII and Section III), ASME B31.1, "Power Piping," and ASME B31.3, "Process Piping." Most other countries have similar standards, and the European Community has moved to harmonize and unify some of these standards under a European Pressure Equipment Directive (PED) 97/23/EC. Most of these codes recognize ASME B16.34, "Valves—Flanged, Threaded, and Welding End," as an accepted standard for design, materials, and testing of valves.

The selection of proper materials for valves requires consideration of all factors related to the process fluid and operating conditions. This can include forces applied from actuators and piping, (wind and seismic loads, hoists and rigging, accidental loads), changes in process and ambient temperatures, and the erosive and corrosive potential of both the process fluid and the outside environment. The complex scope of material selection is outside the purpose of this text, but a few of the common globe valve materials are listed in Table 6.19ww with typical ASTM specifications and process applications.

TABLE 6.19ww
Common Valve Body Materials, Specifications, and Applications

Common Material Designation	Body Material Specification (Castings)	Body Material Specification (Forgings)	Max. ANSI Pressure Class Rating (ASME B16.34a-1998)	Temperature Limit °F (°C)*		Typical Service/Applications
				Min.	Max.	
304 SS	A351-CF8	A182-F304	Class 4500	−450 (−268)	1000 (538)	General corrosion, organic acids, nitric acid, hydroxides, hydrocarbons, high temperature; poor resistance to halogens (chlorides, fluorides, etc.), poor resistance to pitting and crevice attack
304L SS	A351-CF3	A182-F304L	Class 4500	−450 (−268)	850 (454)	Same as 304, low to moderate temperature
316 SS	A351-CF8M	A182-F316	Class 4500	−425 (−254)	1000 (538)	Same as 304, high temperature; fair resistance to pitting and crevice attack
316L SS	A351-CF3M	A182-F316L	Class 4500	−425 (−254)	850 (454)	Same as 316, low to moderate temperature
316H SS	A351-CF10M	A182-F316H	Class 4500	−325 (−198)	1500 (816)	Same as 316, high temperature, creep resistant
Alloy 20	A351-CN7M	A182-F20	Class 4500	−325 (−198)	600 (316)	Sulfuric acid, organic acids, hydroxides, nonhalogenated organic chemicals
Bronze	B61 (C92200)	n/a	B16.24, Class 300	−325 (−198)	400 (204)	Steam, fresh and chloride water, oxygen
Aluminum-Bronze	B148-C95200	n/a	B16.24, Class 2500	−425 (−254)	600 (316)	Brine; fresh, brackish, and salt water; chlorides
Carbon Steel	A216-WCB	A105	Class 4500	−20 (−28)	800 (427)	Neutral and alkaline waters, steam, dilute caustic, hydrocarbons
Carbon Steel	A216-WCC	A350-LF3	Class 4500	−20 (−28)	800 (427)	Neutral and alkaline waters, steam, dilute caustic, hydrocarbons
Cast Iron	A126-A	n/a	B16.1, 400 PSIG	−20 (−28)	406 (208)	Neutral and alkaline waters, low-pressure steam, dilute caustic
Chrome-Moly WC9	A217-WC9	A182-F22 Cl.3	Class 4500	−20 (−28)	1200 (649)	Mildly corrosive, high temperature, resists erosion by steam and flashing water
Chrome-Moly C12A	A217-C12A	A182-F91	Class 4500	−20 (−28)	1200 (649)	Mildly corrosive, high temperature, resists erosion by steam and flashing water
Duplex SS 22% Cr	A351/A995-CD3MN (J92205)	A182-F51	Class 4500	−20 (−28)	600 (316)	Corrosive, brine, salt water, polluted water, acid-chlorides, good against pitting and crevice attack
Duplex SS 25% Cr	A351-CD4MCu	A182-F61	Class 4500	−20 (−28)	600 (316)	Corrosive, brine, salt water, polluted water, acid-chlorides, better against pitting and crevice attack
Hastelloy B/B2	A494-N-7M	B335-N10001/ N10665	Class 4500	−325 (−198)	800 (427)	Superior in hydrochloric acid up to boiling, hydrofluoric acid, strong reducing chemicals
Hastelloy C/C-276	A494-CW-6M	B564-N10276	Class 4500	−325 (−198)	1250 (677)	Oxidizing and reducing acids, hypochlorite, chlorine, seawater, acid-chlorides, brines, excellent against pitting and crevice attack

(*continued*)

TABLE 6.19ww
(Continued)

Common Material Designation	*Body Material Specification (Castings)*	*Body Material Specification (Forgings)*	*Max. ANSI Pressure Class Rating (ASME B16.34a-1998)*	*Temperature Limit °F (°C)**		*Typical Service/Applications*
				Min.	*Max.*	
Inconel 600	A494-CY-40 Class 2	B564-N06600	Class 4500	−325 (−198)	1200 (649)	Caustics with chlorides or sulfides, mild oxidizers, excellent resistance to chloride SCC
Low Temp CS	A352-LCB	A350-LF3	Class 4500	LCB: −50 (−46)	650 (343)	Neutral and alkaline waters, steam, dilute caustic, hydrocarbons, low-temperature impact strength
Low Temp CS	A352-LCC	A350-LF3	Class 4500	LCC: −50 (−46)	650 (343)	Neutral and alkaline waters, steam, dilute caustic, hydrocarbons, low-temperature impact strength
Monel	A494-M35-1	B564-N04400	Class 4500	−325	900 (482)	Hydrofluoric acid, caustic, seawater, brine; not for oxidizing service
Nickel	A494-CZ-100	B160/B564-N02200	Class 4500	−325	600 (316)	Hot concentrated caustics, SCC resistant in chlorinated organics; not for oxidizing service
Titanium	B367-C-3	B381-F-3	Class 2500	−75	600 (316)	Better than Hastelloys for pitting and crevice attack, wet chlorine, dilute HCl, bleaches, brines; not for dry chlorine or fluorides
Zirconium 705	B752-705C/ Flowtherm HT	B550-R60705	Class 2500	−75	700 (371)	Dry chlorine, hot hydrochloric, -sulfuric, nitric, and acetic acids; not for fluorides or wet chlorine

*Temperature limits based on ASME B16.34 and ASME B31.3.
Courtesy of Flowserve Corporation

Trademarks

17-4PH is a trademark of Armco, Inc.
Cavitrol is a registered trademark of Emerson Process Management.
ChannelStream is a registered trademark of Flowserve Corporation.
Colmonoy is a registered trademark of the Wall Colmonoy Corporation.
Drag is a registered trademark of Control Components, Inc.
Grayloc is a registered trademark of Grayloc Products.
Kalrez is a registered trademark of DuPont Dow Elastomers.
VRT is a registered trademark of Dresser Industries, Inc.
Hastelloy and C-22 are registered trademarks of Haynes International, Inc.
Hush is a registered trademark of DeZurik/Copes Vulcan, a unit of SPX Corporation.
Inconel, Incoloy, and Monel are trademarks of the Special Metals Corporation group of companies.
PBI is a registered trademark of Celanese AG.
Stellite and Deloro are registered trademarks of Deloro Stellite, Inc.
Teflon and Tefzel are registered trademarks of the DuPont Company.
Turbo-Cascade is a registered trademark of Yarway Corporation, a unit of Tyco-Flow Control.

Reference

1. See ANSI/ISA-75.11-1985 and IEC 60534-2-4: 1989.

Bibliography

Adams, M., “Control Valve Dynamics,” *InTech,* July 1977.
ANSI/ISA-75.02-1996, “Control Valve Capacity Test Procedures,” Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1996.
ANSI/ISA-75.01.01-2002, “Flow Equations for Sizing Control Valves,” Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 2002.
Ball, K. E., “Final Elements: Final Frontier,” *InTech,* November 1986.
Baumann, H. D., “How to Assign Pressure Drop Across Control Valves for Liquid Pumping Services,” *Proceedings of the 29th Symposium on Instrumentation for the Process Industries,* Texas A&M Univ., 1974.

Borden, G. and Friedmann, P. G. (eds.), *Control Valves—Practical Guides for Measurement and Control,* Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1998.

Borden, G. and Zinck, L., "Control Valve Seat Leakage," *InTech,* November 1986.

Buckley, P. S., "A Control Engineer Looks at Control Valves," *Proceedings of the 1st ISA Final Control Elements Symposium,* Wilmington, DE, 1970.

Carey, J. A., "Control Valve Update," *Instruments and Control Systems,* January 1981.

Control Valve Handbook, 3rd edition, Marshalltown, IA: Fisher Controls International Inc., 2001.

Control Valve Seat Leakage, FCI 70-2-1998, Cleveland, OH: Fluid Controls Institute, Inc., 1998.

"Control Valves: Globe, Plug, Pinch, Needle, Gate," *Measurements and Control,* February 1993.

Davis, J.R. (ed.), *Corrosion–Metals Handbook, Ninth Edition, Volume 13,* Metals Park, OH: ASM International, 1987.

Cunningham, E. R., "Solutions to Valve Operating Problems," *Plant Engineering,* September 4, 1980.

Engel, H. O., *Control Valves for Process Automation,* Maintal, Germany: Honeywell Regelsysteme GmbH, 1998.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems,* January 1980.

Fontana, M. G. and Greene, N. D., *Corrosion Engineering,* 3rd ed., New York: McGraw-Hill, 1985.

George, J. A., "Seat Leakage Standard Revised," *InTech,* October 2003.

Hutchison, J. W. (ed.), *ISA Handbook of Control Valves,* 2nd ed., Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1976.

IEC 60534-2-4: 1989, "Inherent Flow Characteristic and Rangeability," Geneva, Switzerland: Research Triangle Park, NC, International Electrotechnical Commission, 1989.

Inherent Flow Characteristic and Rangeability in Control Valves, ANSI/ISA-75.11-1991, Research Triangle Park, NC, Instrumentation, Systems, and Automation Society, 1991.

ISA-RP75.23.01-1995, "Considerations for Evaluating Control Valve Cavitation," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1995.

Lyons, J. L., *Lyons' Valve Designers Handbook,* New York: Van Nostrand Reinhold, 1982.

Pyotsia, J., "A Mathematical Model of a Control Valve," 1992 ISA Conference, Houston, TX, October 1992.

Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

Shinskey, F. G., "Control Valves and Motors," Foxboro Publication No. 413-8.

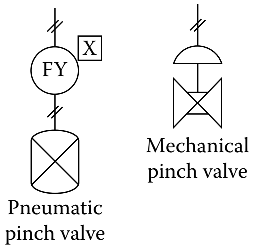
Whatever Happened To... A Guide to Industrial Valve and Actuator Companies - Past and Present, Washington, D.C.: Valve Manufacturers Association, 1997.

Wolter, D. G., "Control Valve Selection," *InTech,* October 1977.

6.20 Valve Types: Pinch Valves

C. S. BEARD (1975) **J. B. ARANT** (1985) **B. G. LIPTÁK** (1995)

J. B. ARANT, D. GARDELLIN, B. G. LIPTÁK (2005)



Flow sheet symbols

Applications:	Abrasives, minerals in suspensions, hydraulically transported solids, slurries, viscous, or food industry products
Sizes:	Generally 1 to 24 in. (25 mm to 0.61 m); special units from 0.1 to 72 in. (2.5 mm to 1.8 m)
Design Pressures:	Generally up to ANSI Class 150, with special units available with up to ANSI Class 300 ratings. See Figure 6.20a
Maximum Pressure Drop:	Generally about 10 to 200 psid (69 to 138 kPa)
Design Temperatures:	–20° to 350°F (–29° to 204°C)
Materials of Construction:	*Sleeve materials:* Buna-N, Chlorobutyl, EPDM, Hypalon, Neoprene, PGR, Polyurethane, Teflon, Viton *Bodies:* Aluminum, cast grey iron, cast ductile iron, cast steel, stainless steel, Polyamide blend plastic
Characteristics:	For the characteristics of reduced port and Clarkson pinch valves, refer to Figures 6.20v and 6.20w; for others, refer to Figure 6.20b
Costs:	See Figure 6.20c
Capacity:	$C_v = 60\ d^2$ for full-ported pinch valves and $C_v = 20\ d^2$ for reduced-ported pinch valves; see Tables 6.20e and 6.20f
Rangeability:	From 5:1 to 10:1
Leakage:	Generally ANSI Class IV or Class V (see Table 6.1gg in Section 6.1for definitions)
Partial List of Suppliers:	Clarkson Co. (www.tycovalves.com) Elasto-Valve Rubber Products Inc. (www.evrproducts.com) Ever-Flex (www.ever-flax.net) Larox (Finland–USA) (www.larox.fi) Onyx Valve Co. (www.onyxvalve.com) Red Valve Co. (www.redvalve.com) Richway Industries (www.richwayind.com)

INTRODUCTION

These valves are called either pinch or clamp valves, depending upon the configuration of the flexible tube and on the means used to compress the tube. The compression can be done by mechanical clamping mechanisms or by external pneumatic or hydraulic power within a metal jacket enclosure. Pinch valves have been improved due to introduction of plastic tubes, elastomers, and reinforcing fabrics.

Tubes can be fabricated from pure gum rubber or from a variety of rubber-like elastomers such as Buna-N, butyl, neoprene, Nordel, hypalon, Viton, silicone, polyurethane, polypropylene, white butyl, and odorless and tasteless white neoprene. The latter two materials are often used in the food and allied industries. Reinforcing fabrics may include some of the materials used in automobile tire fabrication, such as cotton duck, rayon, nylon, fiberglass, and Kevlar®, which is a new arimid polymer material that is as strong as steel at

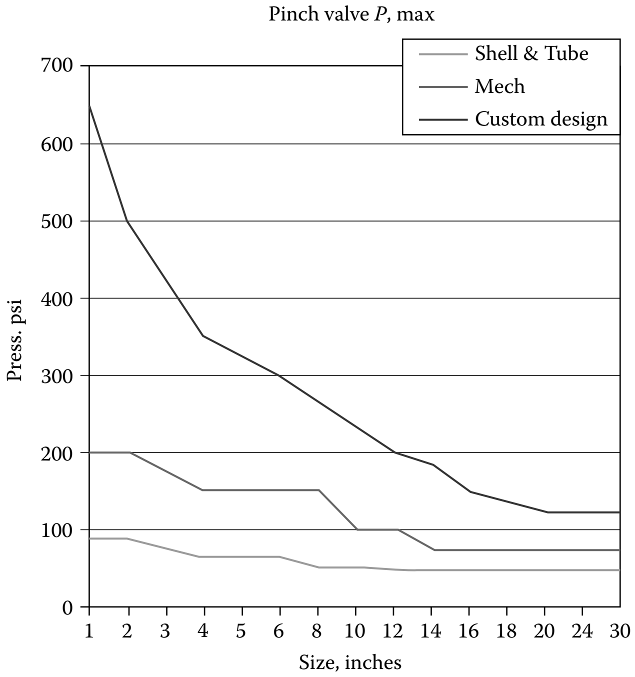


FIG. 6.20a
Pinch valve pressure rating drops as size increases. The bottom curve gives the rating of shell and tube, the center of mechanical and the top of custom designs.

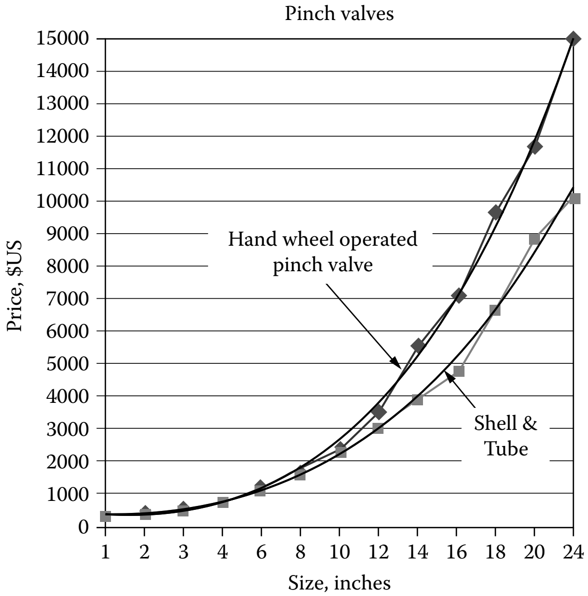


FIG. 6.20c
The shell and tube-type pinch valves are less expensive than cylinder or electric-operated valves and generally cost less than hand-operated valves. Cost data are based upon standard flanged cast iron housing, pure gum rubber sleeve, and on/off operation. Numerous other elastomers are available for the sleeve, a range of material for the body, and a variety of actuators and drives.

one sixth the weight. Teflon plastic tubes are uniquely capable of handling highly corrosive or sticky fluids.

As was shown in Figure 6.20b, pinching a standard full round sleeve from 100% to 50% of travel has little effect on the flow through the valve. The advantage of the prepinched tube design is that this "dead" part of the valve control is nearly eliminated. As can be seen from Table 6.20e and Table 6.20f, fluid capacity is sacrificed, because the shape of a full open

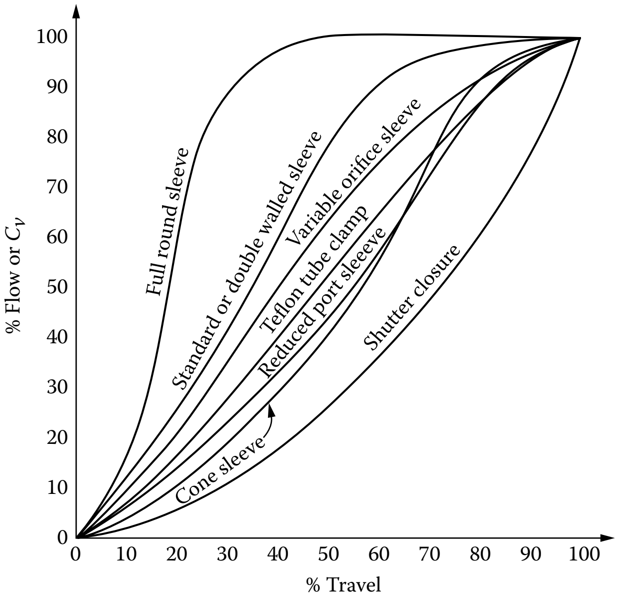


FIG. 6.20b
The characteristics of a full round sleeve is nearly quick opening. Double-walled or cone-shaped sleeve cross-sections bring the characteristics closer to linear.

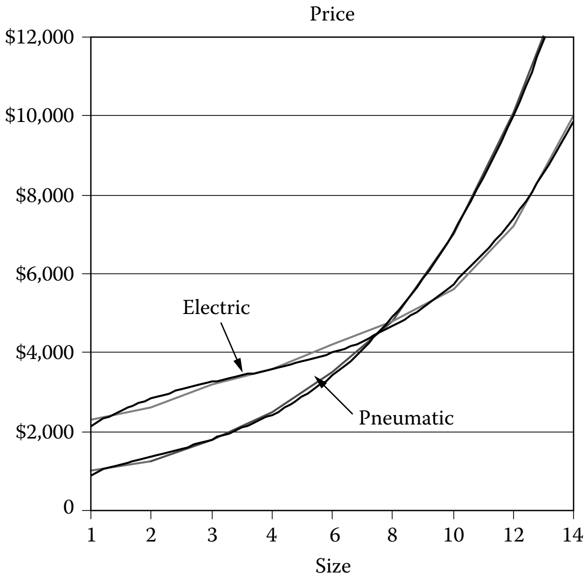


FIG. 6.20d
Price comparison pinch valves provided with electric or pneumatic actuators.

TABLE 6.20e

*Valve Coefficients (C_v) of Standard and Reduced-Port Pinch Valves during Throttling (% of Travel)**

Standard and Double-Wall Sleeve Flow Chart C_v

Valve Size, in.	*Valve Opening—% of Total Travel* 10	20	30	40	50	60	70	80	90	100
1/2	2.7	5.0	7.1	8.9	10.2	11.5	12.5	13.1	13.5	14
3/4	8.0	18	22	27	29	30.5	31	32	33	34
	9.0	18	28	41	50	61	64	65	66	67
1	19	40	62	91	112	137	143	145	147	148
1 1/2	34	70	109	159	196	240	252	255	257	260
2										
2 1/2	53	108	169	247	304	372	390	395	398	402
3	74	152	237	347	427	523	548	554	560	565
4	10	235	389	532	656	759	791	803	811	817
5	169	394	714	961	1,092	1,176	1,216	1,237	1,251	1,259
6	229	552	1,038	1,390	1,527	1,594	1,643	1,670	1,691	1,700
8	405	979	1,843	2,466	2,706	2,827	2,913	2,961	2,998	3,014
10	634	1,531	2,883	3,856	4,233	4,420	4,552	4,629	4,686	4,710
12	1,034	2,496	4,701	6,288	6,902	7,207	7,422	7,548	7,641	7,680
14	1,381	3,335	6,280	8,400	9,224	9,624	9,911	10,083	10,209	10,260
16	1,804	4,355	8,202	10,971	12,047	12,569	12,944	13,170	13,333	13,400
18	2,302	5,558	10,467	14,000	15,373	16,040	16,519	16,806	17,015	17,100
20	3,405	8,223	15,486	20,713	22,745	23,731	24,440	24,865	25,174	25,300
24	4,215	10,180	19,171	25,642	28,157	29,378	30,255	30,781	31,164	31,320

Reduced Port Sleeve Flow Chart C_v

Valve Size, in.		Port Size, in.	*Valve Opening—% of Total Travel* 10	20	30	40	50	60	70	80	90	100
1/2	×	1/4 3/8	0.09	1.3	2.1	2.8	3.0	3.4	3.7	3.8	3.9	4
1/2	×	3/8	0.43	2.5	3.6	4.7	5.8	6.3	7.0	7.5	7.8	8
3/4	×	1/2	0.2	1.9	3.0	4.0	5.0	5.6	6.0	6.4	6.5	6.75
3/4	×	1/2	0.31	2.2	3.2	4.3	6.2	8.1	10.3	12.4	13.6	14
1	×		0.30	1.0	2.0	4.0	6.0	8.0	10	12	13	14
1	×	3/4	6.0	11	16	20	25	28	30	31	32	33
1 1/2	×	3/4	2.0	5.0	10	14	18	21	28	30	31	32
1 1/2	×	1	5.0	10	17	28	34	43	58	60	62	64
2	×	1	4.0	7.0	12	20	25	30	35	40	43	44
2	×	1 1/2	10	19	35	53	68	79	96	107	118	119
2 1/2	×	1 1/2	7.0	14	25	36	48	59	70	80	85	87
2 1/2	×	2	15	33	57	81	105	132	161	181	189	196
3	×	1 1/2	5.0	11	19	27	36	45	52	58	62	64
3	×	2	11	25	42	60	79	102	135	166	175	181
4	×	2	8.0	18	29	42	57	73	82	95	99	102
4	×	3	33	65	110	163	220	269	314	358	383	391
5	×	3	21	43	72	110	149	180	224	238	252	266
5	×	4	53	113	181	270	364	459	540	594	621	630
6	×	4	45	92	152	216	308	386	470	523	550	551
6	×	5	55	142	295	415	544	664	753	922	1230	1253

(continued)

TABLE 6.20e
(Continued)

*Valve Coefficient (Cv) for Standard and Reduced-Port Pinch Valves**

Reduced Port Sleeve Flow Chart C_v

Valve Size		*Port Size*	*Valve Opening—% of Total Travel*									
			10	*20*	*30*	*40*	*50*	*60*	*70*	*80*	*90*	*100*
8	×	4	32	89	146	210	270	337	391	431	457	460
8	×	6	110	250	407	614	851	1,107	1,250	1,372	1,491	1,502
10	×	6	151	230	405	603	750	9,01	1,036	1,062	1,14	1,123
10	×	8	241	300	1,009	1,375	1,780	2,175	2,472	2,490	2,521	2,570
12	×	8	277	421	742	1,105	1,374	1,650	1,898	1,945	2,041	2,057
12	×	10	359	919	1,545	1,516	2,726	3,331	3,736	3,313	3,861	3,936
14	×	10	451	686	1,209	1,799	2,238	2,689	3,091	3,169	3,324	3,351
14	×	12	540	1,346	2,242	3,220	3,992	4,878	5,544	5,584	5,654	5,764
16	×	12	673	1,325	1,805	2,687	3,342	4,015	4,616	4,732	4,964	5,004
16	×	14	726	1,808	3,041	4,144	5,364	6,555	7,450	7,504	7,597	7,745
18	×	14	903	1,375	2,421	3,603	4,482	5,384	6,190	6,345	6,656	6,710
18	×	16	969	2,413	4,060	5,532	7,161	8,751	9,946	10,018	10,143	10,344
20	×	16	1,211	1,844	3,247	4,832	6,010	7,220	8,301	8,509	8,925	91,25
20	×	18	1,255	3,125	5,258	7,164	9,273	1,1332	12,880	12,973	13,183	13,394
24	×	18	1,438	2,189	3,854	5,736	7,134	8,570	9,853	10,100	10,594	10,821
24	×	20	2,007	3,056	5,380	8,007	9,959	11,963	13,758	14,099	14,789	15,120

*Courtesy of Red Valve Co.

prepinched tube has approximately 80% of the area of a full open standard round tube.

Pinch valves can be mechanically pinched or be of the shell and tube type. The mechanically pinched design can have a frame that is open or enclosed and can be provided with handwheel, pneumatic, or electric operators. If a pneumatic actuator is selected for a mechanically pinched valve, that design is available with either an open or a closed failure position. The shell and tube designs can be pinched pneumatically or hydraulically and can only fail open. A summary of pinch valve features is given in Figure 6.20g.

THE SLEEVE

A pinch valve consists of a short section of flexible hose combined with a mechanism for throttling or completely closing the hose. The hose is made of a composite matrix of elastomer and a fabric carcass. The fabrication and materials of construction are similar to automobile tire manufacturing, but inside out, in the sense that the tire has the wear surface on the outside, a pinch valve has the wear surface on the inside of the tube.

Much of the art of making pinch valves lies in the design and manufacture of the elastomer sleeve. By balancing the number of fabric layers, the type and orientation of the fabric, and the ratio of rubber to fabric, the sleeves can be made soft and pliable for low-pressure or tough and rigid for high-pressure applications. Table 6.20h lists the properties of a variety of elastomers used as pinch valve sleeves.

The tensile strength of the sleeve is of particular importance because it limits the ability to resist tearing and delamination of the rubber and fabric. As can be seen from Table 6.20h, natural gum rubber is superior to all the synthetic compounds for abrasion resistance, tensile strength, and elongation. It is also the least expensive. This makes it the material of choice for most pinch valve sleeves.

In addition to the elastomers listed in Table 6.20h, most manufacturers offer rubber compounds in FDA White Odorless and Tasteless versions for food and cosmetic applications. The white pigments in these compounds compromise temperature resistance and tensile strength. Pinch valves are available with conductive gum rubber sleeves that safely dissipate static electricity. These are routinely used in the manufacture of nitro-glycerin, gunpowder, and ordinance materials.

Some manufacturers offer proprietary extreme abrasion resistant compounds such as Linatex Red Crepe© and Onyx Blue©. Urethane compounds are available on special order; they offer good abrasion resistance but are plagued by delamination and splitting at the pinch points. Kalrez© has been debated for years as a pinch valve material but its extreme cost has made it prohibitive even for small pinch valves.

Reinforcing fabrics for pinch valves borrow heavily from the tire industry and include polyester, nylon, cotton, rayon,

TABLE 6.20f

The Valve Capacity Coefficients (C_vs) are Listed for Full and Reduced Ported Pinch Valves between the Valve Sizes from 0.5 to 30 in.

Valve	Port	C_v	Valve	Port	Prepinch Full	Full Round C_v	Valve	Port	Prepinch C_v	Full Round C_v
1/2	Full	12	4	Full	944	1133	14	Full	8,550	10,260
	0.37	9		3.0	460	552		12	5,764	6,917
	0.25	4		2.5	186	223		10	3,351	4,021
3/4	Full	32		2.0	120	144		8	1,837	2,204
	0.50	14		1.5	51	61		6	943	1,132
	0.37	8								
1	Full	42	5	Full	1364	1637	16	Full	11,200	13,440
	0.75	31		4.0	714	857		14	7,745	9,294
	0.62	28		3.0	287	344		12	5,004	6,005
	0.50	18		2.5	163	196		10	3,002	3,602
	0.37	6		2.0	94	113		8	1,711	2,053
1.5	Full	86	6	Full	1843	2212	18	Full	14,300	17,160
	1.00	35		5.0	1180	1416		16	10,344	12,413
	0.75	18		4.0	567	680		14	6,710	8,052
	0.62	12		3.0	280	336		12	4,500	5,400
	0.50	11		2.5	152	182		10	2,790	3,348
				2.0	91	109		8	1,632	1,958
2	Full	170	8	Full	3040	3648	20	Full	17,600	21,120
	1.50	79		6.0	1436	1723		18	13,400	16,080
	1.00	30		5.0	832	1000		16	9,125	10,950
	0.75	17		4.0	460	553		14	6,000	7,200
	0.50	8		3.0	233	280		12	4,166	4,999
				2.5	143	172		10	2,650	3,180
2.5	Full	346	10	Full	4520	5424	24	Full	26,100	31,320
	2.00	156		8.0	2534	3040		20	15,120	18,144
	1.50	70		6.0	1149	1379		18	10,830	12,996
	1.00	29		5.0	800	960		16	7,536	9,043
	0.75	18		4.0	420	504		14	5,167	6,200
	0.50	10		3.0	208	250		12	3,780	4,536
								10	2,475	2,970
3	Full	576	12	Full	6400	7,680	30	Full	42,500	51,000
	2.50	293		10	3936	4,723		24	21,250	25,500
	2.00	148		8.0	2057	2,468		20	12,000	14,400
	1.50	66		6.0	1003	1,203		18	8,900	10,680
	1.00	29		4.0	402	482		16	6,461	7,753

Kevlar®, and occasionally fiberglass. They can be built up in either a bias-ply or radial pattern, or a hybrid combination where different layers in the same sleeve have different orientation.

The rubber sleeve is the most critical component of the pinch valve and it determines the maximum time between service intervals. Sleeve quality varies considerably between manufacturers, and price tends to reflect quality.

There are two methods for making pinch valve sleeves. One is the vulcanization process, where they are held together with heat tape, or they can be compression molded.

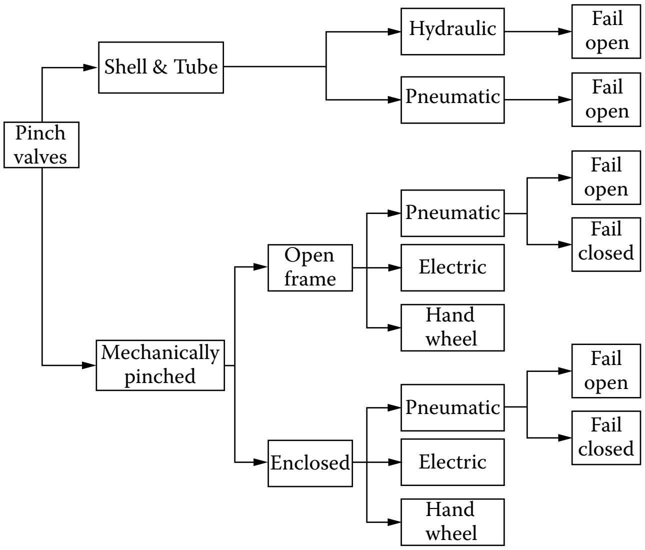


FIG. 6.20g
Pinch valves design variations and feature options.

An important feature of the sleeve design is the positive opening feature (POF), which is illustrated in Figure 6.20i. This feature is provided by a pair of rubber and fabric tabs, which are molded into the trunk of the pinch valve sleeve. These tabs are clamped to the pinch bars. They force the sleeve to open properly when there is no process pressure on the inside to open it, which is of value if the sleeve has been closed for an extended period of time.

PINCH VALVE TYPES

Pressure Limitations

As was shown in Figure 6.20a, shell and tube-type pinch valves have the lowest pressure rating. Their main limitation is the available pressure of the compressed air used to close the valve.

Standard mechanically pinched valves can operate against line pressure in the range of 75 to 200 PSIG, depending on size. The limitation is the strength of the fabric reinforcement in the rubber sleeve. Special-order pinch valves can be fabricated for high-pressure service. Manufacturers use exotic fabrics including Kevlar and the latest engineering software to create hybrid sleeves built up with a mix of bias ply and radial layers to increase working pressure.

Shell and Tube Design

The simplest pinch valve is the *shell and tube* type, which is also referred to as a *direct operated* or as a jacketed pinch valve. This design consists of a cast iron, aluminum, or plastic shell with a rubber sleeve inside. For a cost comparison between shell and tube-type and handwheel-operated pinch valves, refer to Figure 6.20c.

As shown in Figure 6.20j, injecting compressed air into the housing causes the rubber sleeve to collapse, thereby blocking flow of the process fluid through the valve. The valve housing is usually configured to coax the sleeve into a straight-line closure to operate with the lowest possible differential pressure.

Flanged bodies are available in 1 through 72 in. sizes. Body material for flanged valves is cast iron, ductile iron,

TABLE 6.20h
Comparative Properties of Elastomers Used to Make Pinch Valve Sleeves

ANSI/ASTM D1418-77	*NR/IR*	*CR*	*NBR*	*CHR*	*CSM*	*EPDM*	*FKM*	*AFMU*	*SI*
Trade Names	*Poly Isoprene Pure Gum*	*Neoprene*	*Nitrile Buna-N*	*Chloro Butyl*	*Hypalon*	*Nordel*	*Viton Fluorel*	*Teflon*	*Silicone*
Durometer-A (typical)	37–45	50	55	55	70	60	73	52D	40–80
Abrasion Resistance	Outstanding	Excellent	Good	Good	Good	Very good	Fair	Poor	Poor
Tensile Strength, psi	3000	1800	1700	—	1800	1800	1470	2500	1500
Elongation%	640	550	460	—	—	350	300	250	-
T range °F	–30 → +180	–20 → +220	–20 → +230	–40 → +250	–10 → +275	–50 → +325	–15 → +375	–120 → +450	–60 → +460
Acid Resistance Diluted	Fair to good	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Best	Excellent
Acid Resistance Concentrated	Fair to good	Good	Good	Good	Very good	Excellent	Excellent	Best	Fair
Oil Tolerance	Poor	Good	Excellent	Poor	Good	Poor to good	Excellent	Excellent	Fair

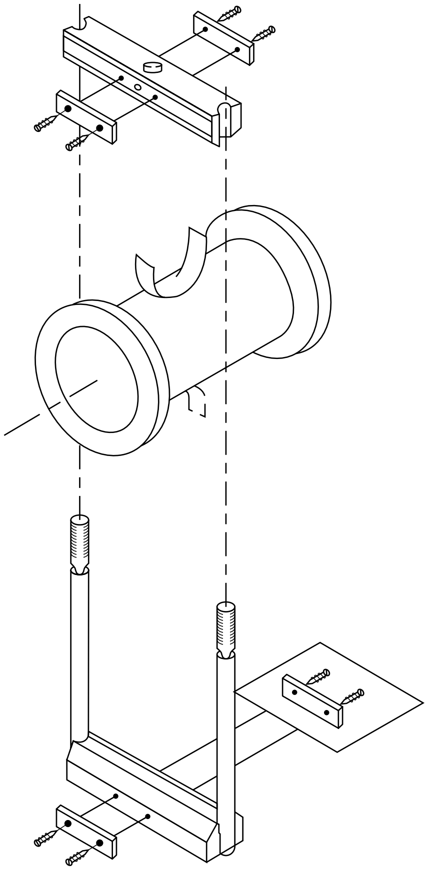


FIG. 6.20i
Pinch valve provided with the positive opening feature (POF). (Courtesy Larox Co.)

cast aluminum, and fabricated carbon steel. Small valves in the range of 1/8 through 3 in. are available with screw-on pipe connections with steel, stainless steel, and plastic bodies. Valves in the 1/2 through 4 in. size are available with slip-on connections, where the mating pipe is simply inserted into the valve body and is retained solely by friction with no threads or grooves required. These are generally made in cast aluminum.

Advantages

1. Simple operation, sleeve is the only moving part
2. Light weight
3. Compact
4. Inexpensive
5. Closes around suspended particles, rags, powder, pellets, and other debris without jamming or leaking

Caveats

1. Limited line pressure capability. This is because a higher pressure is needed on the outer surface of the sleeve to force it closed. Typically a 35 to 40 psi differential between internal and external pressure is required to close the sleeve. In a compressed air system operating at 100 psi, this limits the process pressure to 60 psi.
2. The valve must fail open (FO). There is no practical way to put a spring or other energy storage device inside the valve so it is not available in a true fail-closed version. Auxiliary devices can be added to the valve; for example, a solenoid valve can be configured for fail closed on loss of *electric* power. Likewise, designers add air reserve tanks and other peripheral devices to close the valve on compressed air failure. None of these devices are functional when the sleeve wears through; at that point there is no way to sustain air pressure in the valve housing and the valve opens.
3. Low-pressure and vacuum systems can inhibit full opening or cause valve to “flutter,” thereby compromising flow capacity. There are ways to force the valve open using vacuum jet pumps but they add complexity.
4. Disappointing performance in throttling applications. Because these valves have no valve stem, there is no place to attach a positioner feedback lever. The alternative is to use the valve in concert with an I/P transducer coupled to a pneumatic amplifier. This converts a

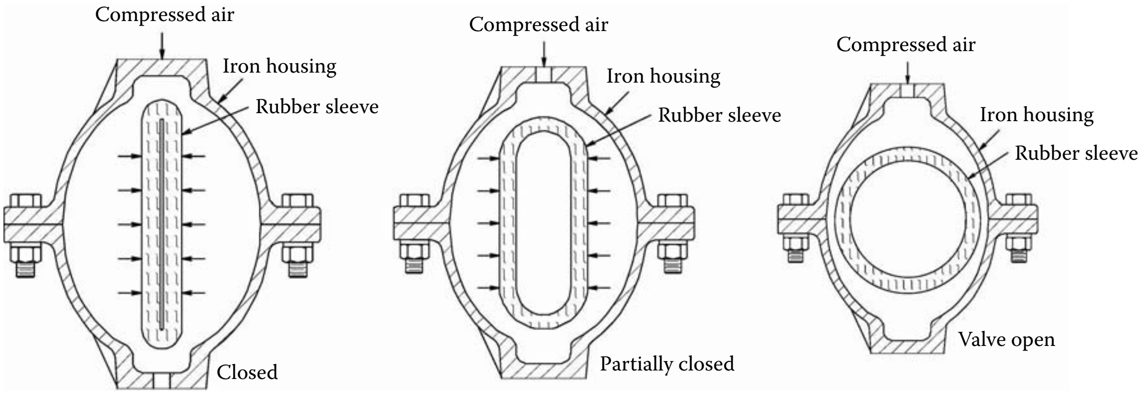


FIG. 6.20j
The throttling of shell and tube-type pinch valves.

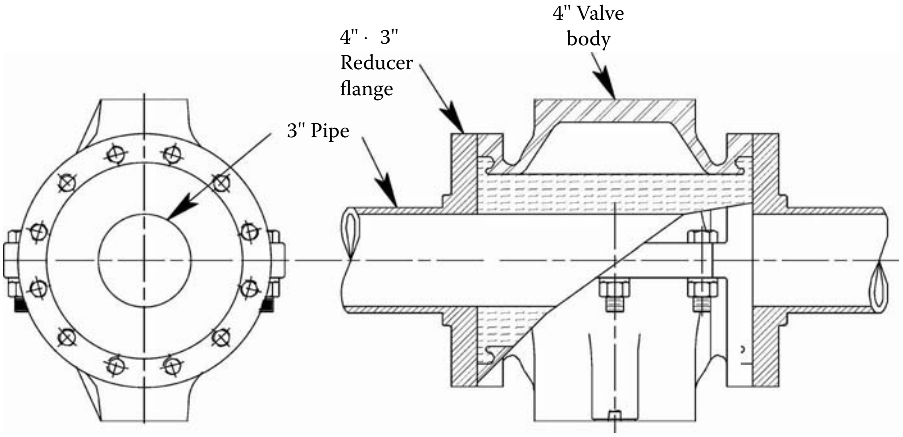


FIG. 6.20k
The design of a double walled, 3 in. size, shell and tube pinch valve.

4–20 mADC electronic signal to a pneumatic 0–100 PSIG signal that is introduced into the valve shell. This varies the opening of the rubber sleeve, but without any feedback in the valve. The only means of closing the loop is through the measuring element and controller. This makes for sluggish response, erratic stability, and shallow turndown. Modulating shell and tube valves are generally limited to slow, linear applications like level control on large vessels. Exception to the rule: For pH control, use a shell and tube pinch valve with a solenoid valve instead of an I/P transducer. Configure the controller for pulse width modulation (PWM) with a *s-l-o-w* time base around 10 sec. This introduces a response time lag but cranks the turndown up to 100:1 or better with zero sleeve abrasion. Want well all-around modulating control? Read on and see how well *mechanically* pinched valves work in throttling applications.

5. Limited sleeve life. In mechanically operated pinch valves, designers add extra fabric for ample reserve strength. Designing a shell and tube valve does not offer this luxury. Additional fabric adds stiffness and increases the differential pressure needed to close the valve. This drives the sleeve design to the minimum number of fabric layers. The result is that the fabric operates right up to its yield point. After a certain number of open-close cycles the fabric fails through fatigue and the sleeve delaminates. Sleeve life in a shell and tube pinch valve operating at its maximum rated pressure is usually around a mean of 75,000 cycles with a variation of ±25,000.

In comparison, mechanical pinch valves operate through practically an unlimited number of cycles because the fabric is operating at reduced stress. Mechanical pinch valves operating for millions of cycles over a period of 20 years are not uncommon.

One method to increase the sleeve life is to regulate the compressed air pressure to precisely 40 psi over process pressure. For example, a 4 in. valve might be rated for 65 psi max process pressure with 105 psi max compressed air pressure. However, if the actual process pressure is 25 PSIG, reducing the air pressure from 105 down to 65 PSIG greatly reduces stress in the reinforcing fabric. This increases sleeve life by about 50%. The additional cost of an air pressure regulator pays for itself in improved sleeve life.

Another method of improving sleeve life is to double the sleeve's wall thickness. As illustrated in Figure 6.20k, this can be done by increasing the valve size, underboring the hole through the sleeve, and doubling the thickness of the rubber sleeve. This requires the use of reducer flanges to make the connection between the pipe and valve.

The Clarkson-C Valve The Clarkson division of Tyco makes a unique variation of the shell and tube valve (Figures 6.20l)

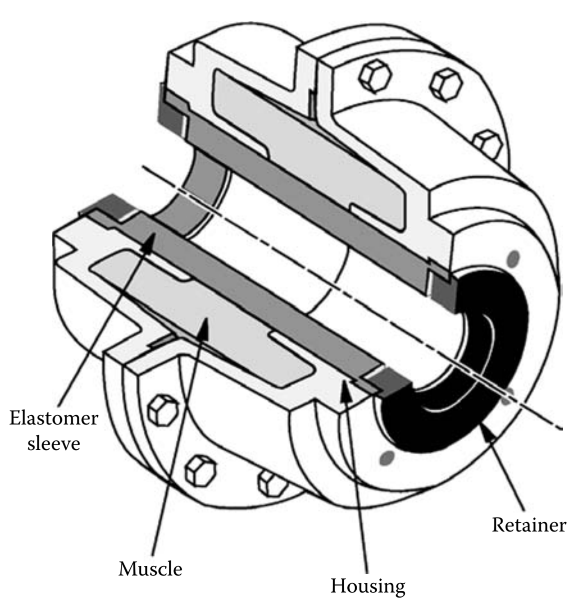


FIG. 6.20l
The Clarkson-C valve.

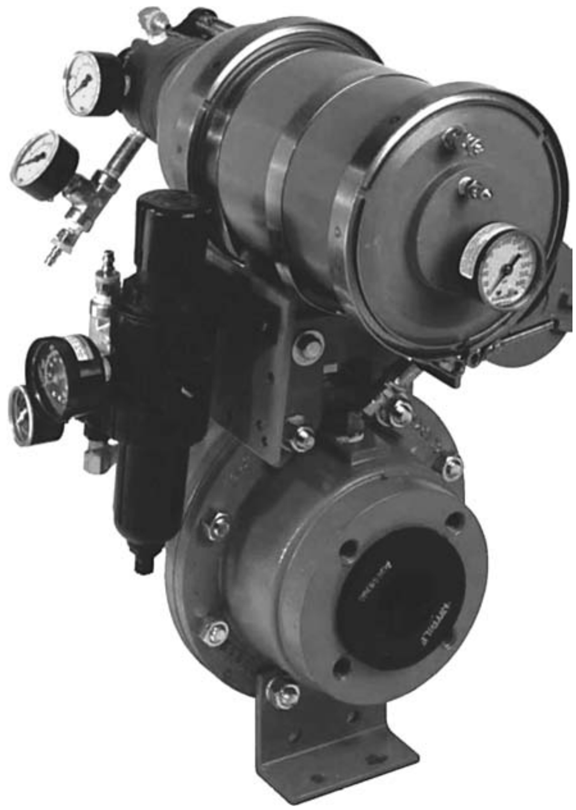


FIG. 6.20m
The Clarkson "C-Valve" showing the air-powered hydraulic pump, amplifier, tank, positioner, check valve, pressure regulator, and gauges.

and 6.20m. It is called the "Clarkson-C" or "shutter valve." This rather unusual valve has been around since the late 1960s. Instead of compressed air, this valve uses hydraulic fluid, which is pressurized to several hundred PSIG, in order to be able to collapse the rubber sleeve. The assembly includes an air-powered hydraulic pump, valves, positioner, air filter-regulator, and gauges piggybacked onto the valve.

The hydraulic fluid is separated from the sleeve by a thick-walled rubber cylinder called the "muscle." To throttle the slurry, the muscle is compressed inward by hydraulic pressure, reducing the orifice size of the sleeve in a 360° squeeze. As it constricts, a round concentric aperture is maintained throughout the usable throttling range. Compared to conventional shell and tube pinch valves, this valve can handle higher pressure and more abrasive fluids, and it offers more precise throttling.

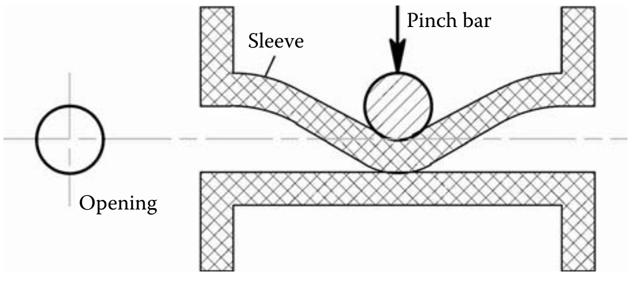


FIG. 6.20n
Full ported and fully round single pinch design.

Naturally, the use of hydraulics adds another level of complexity and, therefore, both maintenance and first-investment costs.

Mechanically Pinched Valves The alternative to the shell and tube-type pinch valve described above is the mechanically pinched valve. In this design, steel pinch bars clamp the rubber sleeve closed. A handwheel, lever, electric actuator, or pneumatic cylinder actuates the pinch bars. The mechanically pinched valve has ample energy available to effect closure so the rubber sleeve can incorporate more reinforcing fabric. This enables the valve to handle higher process pressure and extends potential sleeve life to millions of cycles.

A pneumatic actuator can be a double-acting cylinder, which, if you combine it with a two-coil solenoid valve, is inherently a fail-in-last position device. Pneumatic actuators are available for fail-open and fail-closed operation, using mechanical springs or air reservoirs to store the energy required to operate in the absence of compressed air pressure.

As a general rule, pneumatic actuators are the least-expensive for pinch valves up to 6 in. size, electric actuators have the advantage for valves over 10 in., and in between it is roughly a draw.

The majority of mechanically pinched valves use flanged connections, although there are some designs that use a slip-on connection with a hose clamp. 125/150 lb flanges are the most common, with 250/300 lb flanges also available.

A word of explanation on the difference between 125 and 150 lb flanges: The bolt circles are identical between 125 and 150 lb flanges. The difference is the housing material: If the valve housing is cast grey iron or aluminum, ANSI 125 lb specifications apply; if the housing is steel, ductile iron, or stainless steel, then ANSI 150 lb specifications apply. The same relationship holds for 250 and 300 lb flanges.

Single Pinch Design As illustrated in Figure 6.20n, the simplest mechanical pinch valve design consists of a rubber tube and a single steel bar that pinches the tube from above.

When the valve is open, the inside is a round circle with its diameter of opening equal to the diameter of the adjacent pipe. Therefore, a full open 2 in. valve can pass a 2-in.-diameter sphere. This design offers good abrasion resistance, allows the pipe to drain completely, and can be cleaned with a pipe pig. This design is also a simple and economical mechanism.

The main limitation of this design is valve size: It works for pinch valve sizes up to and including 3 in. Another problem is due to the fact that as the steel bar pinches the tube closed, the rubber and reinforcing fabric have to stretch. On small valves, this is not a problem. But on larger valves, the change in length exceeds the allowable elongation of the rubber and fabric matrix.

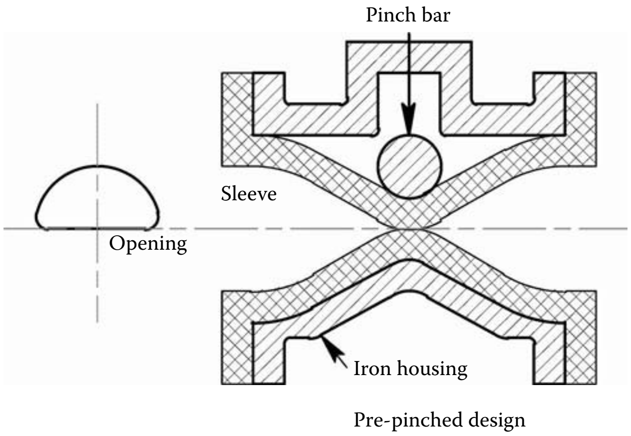


FIG. 6.20o
Prepinched design.

There are three possible solutions to this dilemma:

1. Make the valve longer. This reduces the elongation on the sleeve. This solution is controversial, because longer valves take up more space, and in today's competitive environment, plant space is a valuable commodity.
2. Prepinch the valve.
3. Use two moving pinch bars.

Prepinched Design The valve in Figure 6.20o is a prepinched design. The term "pre-pinched" refers to the iron housing. The valve itself is "full ported," because the sleeve has a full round shape before insertion into the housing. The iron housing has a weir cast into it that pinches the valve to centerline, resulting in a "D" shaped opening. This valve can pass a sphere equal to half nominal line size. If you want to pass a 4-in.diameter grapefruit through a 4 in. valve, you have to cut it in half and push the pieces though flat side down.

Advantages

- Simple operation.
- Reduced cost (compared to dual pinch valves).
- Shorter actuator stroke, reducing the cost of the associated actuator.
- Good throttling characteristic.

Disadvantages:

- Largest particle that the valve can pass is half nominal valve size.
- Reduces valve capacity by 20%.
- With valve full open, adjacent pipe will not drain below centerline.
- You cannot pig out the line.
- Accelerated sleeve wear in high-velocity, high-abrasion applications

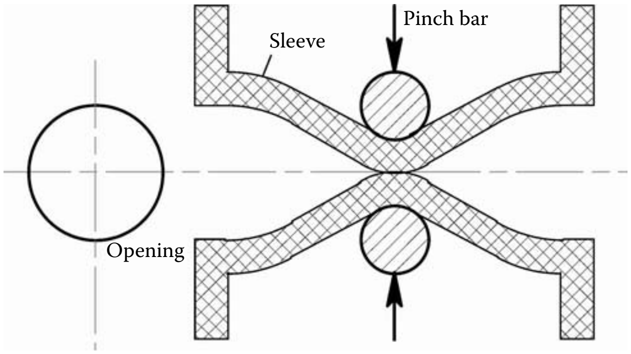


FIG. 6.20p
Fully round and full ported dual pinch mechanism.

Dual-Pinch Mechanism As shown in Figure 6.20p, an alternate design is the dual-pinch mechanism, where two separate pinch bars squeeze the rubber tube closed. Here, there are two moving pinch bars that meet at the centerline of the valve to pinch off the rubber sleeve. "Full round" refers to the mechanical parts of the valve; "full port" refers to the sleeve design.

Advantages

- Excellent abrasion resistance.
- Minimal friction.
- Valve can pass particles equal to line size.
- Adjacent pipe can drain completely.
- Valve can pass a cleaning pig.

Disadvantages

- Mechanism is more complex.
- More expensive than prepinched.
- Actuator moves when stroking valve.
- Requires longer stroke actuator compared to prepinched design.

Open and Enclosed Designs Most mechanical pinch valves have a housing that surrounds the rubber sleeve. As shown in Figure 6.20q, this housing connects the flanges, supports the handwheel or actuator, and protects the rubber sleeve from the environment. This housing also serves to contain the process fluid when the rubber sleeve wears out.

An alternative is the open frame design. This pinch valve consists of a rubber sleeve, a clamping mechanism, and little else. Its light weight makes it popular for use in nonmetallic piping systems. It is also popular in mining operations as a "throw-away" valve.

Advantages

- Easy visual inspection of the rubber sleeve
- Economy
- Lightweight

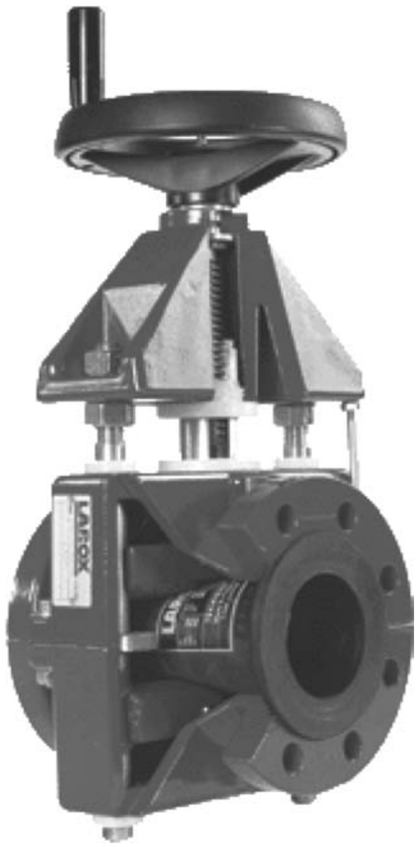


FIG. 6.20q
Enclosed pinch valve. Part of the housing has been cut away to show the operation of the guide rods. (Photo courtesy Larox Company.)

Caveats The rubber sleeve is a sacrificial part. When the sleeve wears out, whatever is being pumped just sprays all over the place. These valves are well suited for sand mining and stormwater applications, because if one inadvertently dumps a ton or two of sand on the ground, one just sends a crew to clean it up, but no serious harm is done.

On the other hand, always use a *fully enclosed* pinch valve when the process fluid is scalding hot, acidic, caustic, toxic, or flammable. Always consider the consequences of leakage, and never compromise safety.

Actuators Mechanical pinch valves can be actuated by:

- Handwheel
- Handwheel and gearbox
- Pneumatic actuators
- Electric actuators

Handwheel When a pinch valve is to be buried or installed in a pit, it can be equipped with a torque tube extension and a handwheel operator. From an esthetic point of view, the occasional handwheels protruding from the ground are acceptable, but the cost can be substantial when one has to dig up the valve for servicing.

Gearbox and Handwheel Pinch valves require extraordinary amounts of thrust to operate. The pinch valve requires brute force and lots of it to close the sleeve against the pressurized process fluid, which is a similar task as trying to crush a pressurized truck tire. For example, a 10 in. hand-operated pinch valve closing against, say, 75 PSIG process pressure requires about 5 tons of force at the seat. Using a normal acme-threaded stem, this works out to 150 ft lb_f of torque.

If one uses an unassisted handwheel, one will require a 6-foot diameter handwheel to reduce the rim pull down to 50 ft lb_f. Therefore, once the valve size exceeds 6 or 8 in. and depending on the process pressure, consider the use of a gear assist for manual valve operation.

This is particularly true in case of chain wheel operators. Cutting corners by eliminating the gearbox or using a small-diameter chain wheel can cause serious problems, because once one exceeds the requirement of 150 ft lb_f rim pull, the plant operator might not be able to provide it.

Pneumatic Actuators Pneumatic actuators work particularly well with pinch valves. They are simple, economical, and dependable. One should use pressure regulator controls so that the closing thrust can be set precisely. This is necessary, on the one hand, to provide sufficient thrust to close the valve drop tight, and on the other hand, to avoid overpinching to the point where one might damage the rubber sleeve. Also, as the sleeve wears from erosion, a pneumatic actuator has reserve travel to compensate for the lost material.

The most common pneumatic actuator used on pinch valves is the conventional double-acting air cylinder. When the air supply fails, a double-acting air cylinder fails in its last position.

Using an Air Cushion The pneumatic cylinder can be supplied with a spring or air cushion on either top or bottom for fail-closed or fail-open operation. Air cushions are cheaper than springs but are less reliable, because a minor leak can void the actuator's ability to execute the proper fail action.

The operation of the air cushion is shown in Figure 6.20r, where compressed air is supplied through tube a to pressure regulator b, which reduces the pressure to 45 PSIG. Air flows through check valve c and into cylinder head d and is trapped in the cylinder head by check valve c at a pressure of 45 PSIG. If the air supply fails and, therefore, the pressure below the piston drops, this drives the piston cylinder rod e down, closing the valve.

To retract the cylinder rod, one needs both the air supply to be on and the three-way solenoid (f) to be energized. When this is the case, 90 PSIG air is injected underneath the piston, which lifts it up, compressing the air in the cylinder head (d) to about 80 PSIG.

If the solenoid valve is de-energized or the compressed air supply fails, the 80 PSIG air that is trapped air above the piston will drive the piston down, thereby closing the valve. As the air cushion expands, its 80 PSIG pressure falls back

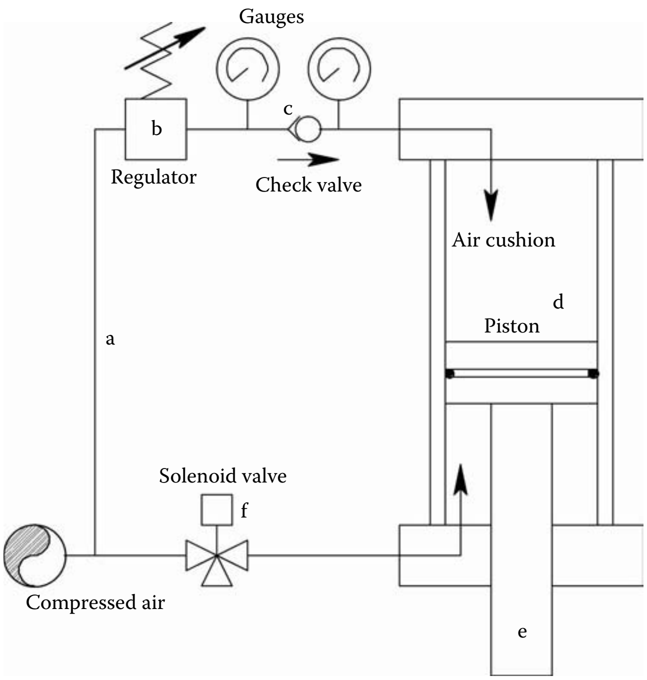


FIG. 6.20r
Double-acting cylinder can be configured to fail closed on air failure, when it is provided with an air cushion.

to 45 PSIG. Therefore, the cylinder has to be sized to operate at this reduced pressure.

Suspension Bag One pinch valve manufacturer has devised a particularly clever pneumatic actuator, where instead of a conventional air cylinder, a rubber pneumatic suspension bag is used as the actuator (Figure 6.20s). One advantage of this actuator on modulating applications is that it has a nonlinear thrust output. The reason for this is that as the suspension bag expands vertically, it "necks in" around its circumference. This eliminates breakaway friction and overtravel, enabling them to position a valve with 1/1,000th in. precision.

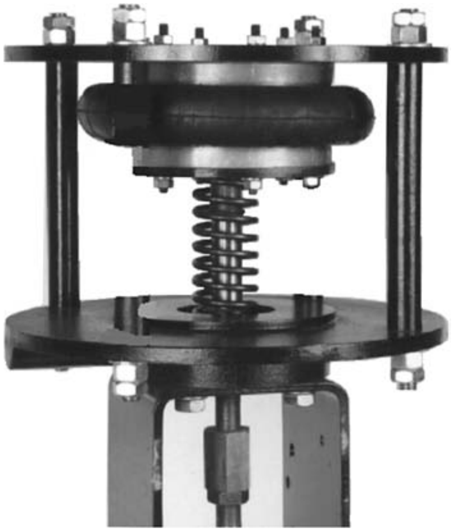


FIG. 6.20s
Suspension bag actuator with its cover removed to clearly show its components. (Courtesy Onyx Valve Company.)

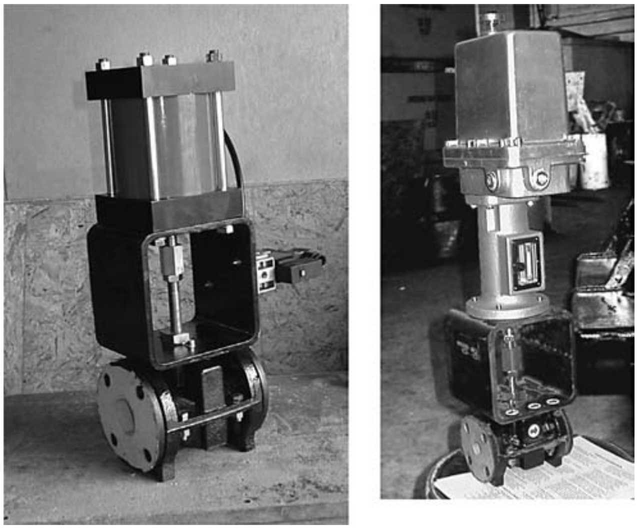


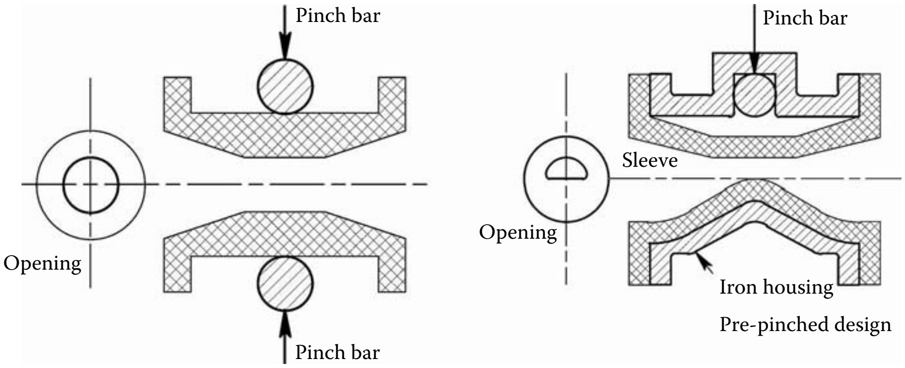
FIG. 6.20t
1 in. mechanically operated pinch valves, with pneumatic actuator on the left and electric on the right. As can be seen in Figure 6.20d, the cost of the electric valve in this size range is higher. For its cost, the user can purchase the pneumatic valve, plus a 2 hp compressor, and have $700 left over.

Pneumatic suspension bags are normally used in trucks and buses as shock absorbers. They are visible on tractor-trailers between the frame and axel assembly. Air bags have a decades-long track record of reliable operation in temperature extremes, rain, snow, sleet, dust, oil, vibration, and shock. They tolerate contamination of the compressed air system without adverse effects. They require no lubrication and operate virtually friction free.

Electric Electric actuators also work well with pinch valves (Figure 6.20t). They avoid problems associated with compressed air including, compressor maintenance, condensation freezing, and the need to run airlines throughout the plant. In smaller valves, pneumatic actuators are more economical. As it was shown in Figure 6.20d, at around 6 or 8 in., the balance of economy tips towards electric actuators.

Throttling Characteristics

The performance of the first throttling pinch valves was very poor, because they were grossly oversized. Until 1985, when David Gardellin ran the first capacity tests on pinch valves, nobody had a clue as to what their C_v and F_L values were. As it turns out, pinch valves offer higher C_v values per valve size than any other valve style (see Tables 6.20e and 6.20f). This is not unexpected, because they have a smooth streamlined internal configuration with no disk, seals, gate tracks, packing cavity, or other sources of friction or turbulence.



Full round reduced port

Pre-pinched reduced port

FIG. 6.20u
Pinch valves can be provided with reduced-port sleeves in two ways. The full round reduced port version is shown on the left and the prepinched reduced port version is shown on the right.

The reason for the poor performance of the full-ported modulating pinch valves is their excess capacity. The consequence of it being oversized is that the automatic controller tends to operate it nearly fully closed, down near the seat. This precipitates two problems:

One is accelerated sleeve erosion caused by the high velocity flow. Pumping abrasive slurry through a pinch valve running just off the seat is like firing a shotgun through a mail slot. The sleeve doesn't last very long.

The other problem is poor control quality. This was because if a 6 in. valve is used, but it needs to open only 1/4 in. to pass the required maximum flow, throttling over such narrow range is next to impossible, and no positioner can correct that.

Reduced Port and Clarkson Sleeve One way to improve the control performance of pinch valves is to use reduced port sleeves. As shown in Figure 6.20u, the reduced port can be provided either in a "full round" form or in a "prepinched" form.

On the left of Figure 6.20v, the inherent throttling characteristics of a full round reduced port valve is described. It is not exactly linear, but using a characterizing positioner with the X^2 function gets it close enough to provide a nearly linear and stable control in most situations.

On the right of Figure 6.20v, the inherent throttling characteristics of a prepinched reduced port pinch valve is shown. It is "prepinched" by the weir in the cast iron housing. Because the sleeve is prepinched, the opening is "D" shaped. It is "reduced port" because of the taper molded into the rubber sleeve.

The prepinched design offers considerable cost savings and a more linear inherent throttling characteristic (on the left of Figure 6.20v) than a full round valve.

Figure 6.20w illustrates the throttling characteristics of the Clarkson C-1 pinch valve, which is between linear and equal percentage.

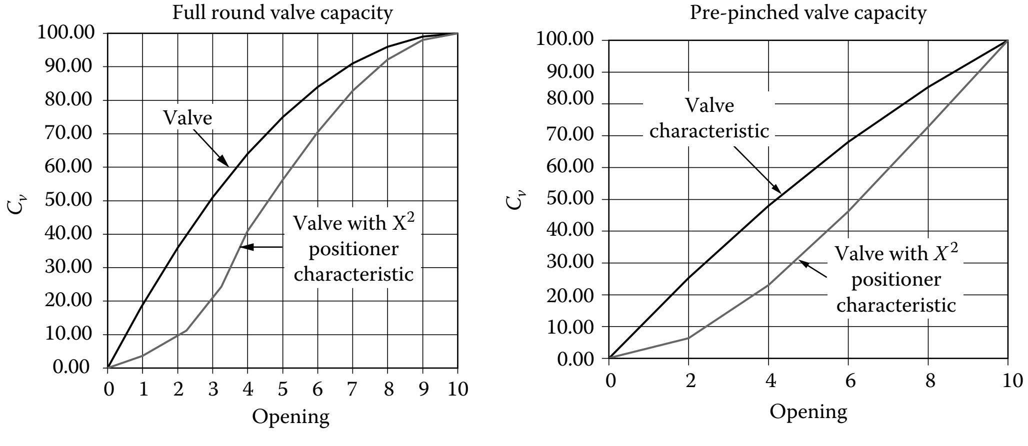


FIG. 6.20v
Characteristics of reduced port pinch valves: On the left the characteristics of the full round, on the right the prepinched design is shown.

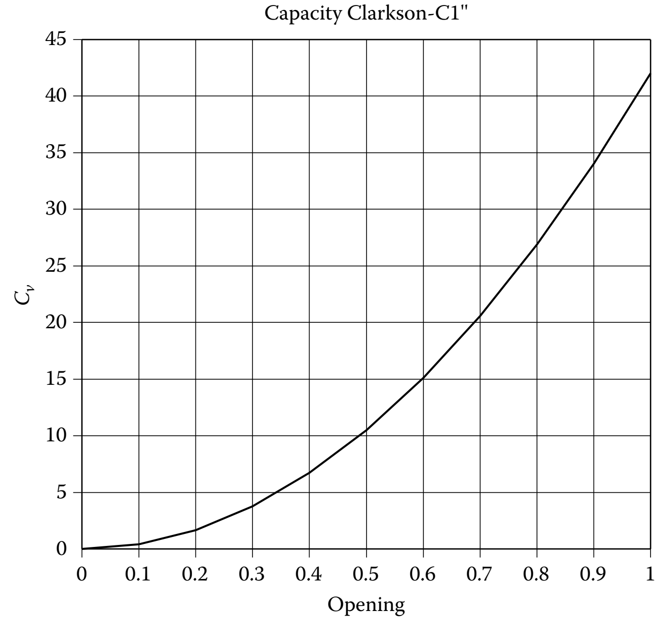


FIG. 6.20w
Characteristic curve of a 1 in. Clarkson "C" valve.

Example When compared to ball, butterfly, or plug valves, the reduced port and Clarkson pinch valves have higher valve coefficients (C_v and F_L) and better throttling characteristics. In many installations, they are also more cost effective.

As an example, let us assume that we have an application that requires a maximum C_v of 336, and the pipe size is 6 in. From both Tables 6.20e and 6.20f, we find that the C_v of a full ported 3 in. valve is the most suitable (C_v = 570). This application, therefore, works out to require a 6 in. × 3 in. pinch valve. There are three ways to design this system:

Option 1 Buy a 6 in. × 3 in. reduced port pinch valve that (according to Table 6.20f) has a C_v measured from 6 in. flange to 6 in. flange of 336. This option (shown on the left of Figure 6.20x) is the most expensive solution, but it also offers the most flexibility, because if the plant capacity increases, one can just change the rubber sleeve to a 6 in. × 4 in. to double the C_v to 680. If more capacity is required, a 6 in. × 5 in. sleeve will give a C_v = 1723 and a full 6-in. sleeve will provide a C_v = 2200, which is a sixfold increase in capacity without changing the valve or the fittings.

Option 2 The second option (shown on the right side of Figure 6.20x) is to purchase a 3 in. full ported valve and install it within 6 in. × 3 in. reducer fittings. This installation will also provide a C_v (measured from 6 in. flange to 6 in. flange) of 336. This second option is the least expensive, but it also offers the least flexibility in terms of accommodating future increases in plant capacity.

Option 3 One can compromise between the above options by installing a 4 in. × 3 in. pinch valve inside two 6 in. × 4 in. pipe reducers. This solution might be the best of both worlds. It offers better economy than buying a 6 in. reduced ported pinch valve, but it does have flexibility for increased capacity by changing out the 4 in. × 3 in. reduced port rubber sleeve to a 4 in. full ported one.

This change would increase the C_v from 336 to 680: double the original capacity without having to change the valve, actuator, or adjacent fittings. Such options do not exist with ball, butterfly, or plug valves.

APPLICATIONS

Pinch valves are used in a surprising range of industries and applications. Their primary application is in controlling the flow of abrasive slurry streams, but they are also used elsewhere. Some examples will be listed below.

Wastewater

Pinch valves can control sewerage and stormwater streams, which might contain oil, paper, rags, sticks, disposable razors, garbage, grit, and other contaminates.

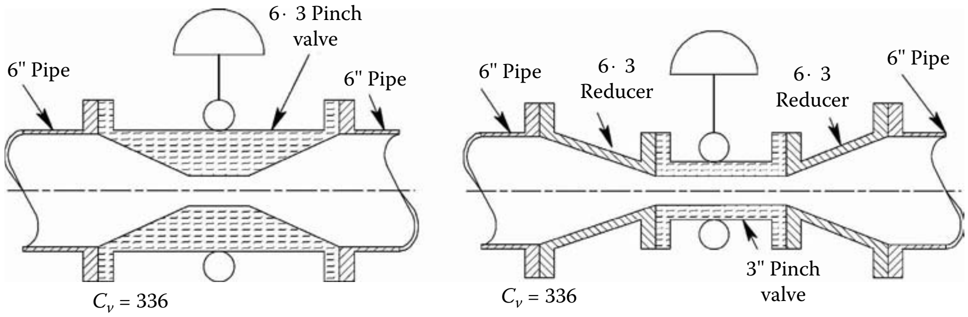


FIG. 6.20x
The same valve coefficient (C_v = 336) can be obtained by installing a 6 in. × 3 in. reduced ported pinch valve (left) or by using a 3 in. full ported pinch valve inside 6 in. × 3 in. reducers (right).

Flue-Gas Desulfurization

Pinch valves are suited for lime slurry service because they offer a straight-through design without any crevices or cavities for material to accumulate. Closing a pinch valve causes the rubber sleeve to stretch. As the rubber sleeve stretches, any scale that has built up on the surface will flake off. The fluid acceleration that occurs as the valve is closing helps to wash these flakes off the rubber surface, so the interior of the valve is self-cleaning. Unlike ball and plug valves, there is no dead space in the valve where a plug of slurry can solidify and block the line.

Mine Slurries

Pinch valves have numerous applications in mining applications, including:

Sand mining	Coal and water slurry	Diatomaceous earth
Copper tailings	Coal and oil fuel	Tar sands
Gold slime	Borax	Mercury
Taconite slurry	Phosphate slurry	Molybdenum
		Kaolin and other clays

Paper and Tile Manufacturing

Pinch valves can handle countless fluids in paper manufacturing, including bleaching and coating applications. As shown in Figure 6.20y, pinch valves can control the levels in head boxes in ceiling tile manufacturing applications, where the thickness of the finished tiles is within 1/16th in. of the specifications.

Toxic Gas Applications

Pinch valves are frequently used to handle carbon monoxide, methane gas from sewerage and landfill applications, and so on. Abrasion is not an issue; the pinch valve is preferred because there is NO PACKING BOX that could leak toxic or flammable gasses into the atmosphere. Consequently, fugitive emissions are not a problem with pinch valves.

Pigments, Paint, and Ink

Pinch valves routinely handle thick viscous fluids that would cement out in other valves after prolonged closure. Pinch valves never jam or clog.

Glue

Pinch valves are used to modulate the pressure in a closed-loop system that sprays the glue onto plywood layers prior to lamination. A pinch valve can provide both throttling control and reliable drop-tight closure even on wood glue.

FIG. 6.20y
A 10 in. modulating pinch valve used for throttling pulp stock and chopped glass fibers to make acoustic ceiling tiles. (Courtesy Onyx Valve Co.)

Food

Pinch valves are used to convey tomatoes, chili peppers, live shrimp, chicken feet (for export to China), mustard, chicken bones and entrails (pet food stock), pig and cattle hooves (for glue), and grains.

Powders and Grinding Compounds

Pinch valves are routinely used to control air-conveyed cement, dry lime, dust, detergent powder (they fill containers very precisely), and fertilizer. They are also used to transport grinding compounds including diatomaceous earth, aluminum oxide, and garnet slurry used to polish teeth, TV screens, auto bodies, and the kitchen sink.

Chemicals

Pinch valves can handle a variety of chemicals without corrosion and with zero leakage because there is no packing box in their design.

CAVITATION

The phenomenon of cavitation has been discussed under the subjects of control valve applications (Section 6.1), control valve noise (Section 6.14), and control valve sizing

(Section 6.15). For that reason, its discussion here will be brief, and the reader is referred to the noted sections for a detailed treatment.

The Phenomenon

The pressure recovery factor (F_L) relates to the ratio between the pressure drop across the valve and the pressure difference between the inlet pressure and the vena contracta pressure (P_{vc}).

$$F_L = [(P_1 - P_2)/(P_1 - P_{vc})]^{1/2} \qquad \mathbf{6.20(1)}$$

The higher the pressure recovery factor (F_L), the better the cavitation resistance of a particular valve design (see Figure 6.1v for a range of valve designs). The cavitation coefficient (K_c) is the ratio of the difference between the in- and outlet pressures and the difference between the inlet and the vapor pressure of the flowing fluid (P_v).

$$K_c = (P_1 - P_2)/(P_1 - P_v) \qquad \mathbf{6.20(2)}$$

If the pressure at the vena contracta falls below the vapor pressure of the liquid, gas pockets form in the flowing liquid. The process reverses itself when the liquid emerges from the restriction, because as the fluid decelerates its pressure recovers. Pressure does not recover to its original magnitude, but if it recovers beyond the vapor pressure, the cavities collapse and the vapors reliquefy. This is cavitation.

Figure 6.1w illustrates a pressure profile where cavitation takes place as vapor bubbles form at the vena contracta and then implode as the pressure recovers and exceeds the vapor pressure. The microjets generated by these implosions cause the damage to metallic surfaces in the area. The resulting micro shockwaves generate localized impact pressures over 200,000 PSIG, which no material can withstand. The location where the cavitation damage occurs is where the bubbles start to collapse, which is downstream of the vena contracta.

Figure 6.15d shows that the formation of these bubbles reduces the process flow until it reaches the "choked flow" condition, when an increase in inlet pressure to the valve will not increase the flow through it. The pressure drop, which will cause choking (Δp_{choked}), can be approximately calculated as

$$\Delta p_{\text{choked}} = F_L^2 (P_1 - 0.93P_v) \qquad \mathbf{6.20(3)}$$

The Pinch Valves

Pinch valves are one of the "high recovery" valve designs with a pressure recovery factor around $F_L = 0.68$. Reduced port valves have higher F_L coefficients, and the more severe reduction in port size, the higher it will be. Figure 6.20z shows the F_L coefficient of a reduced port pinch valve as a function of its percentage opening.

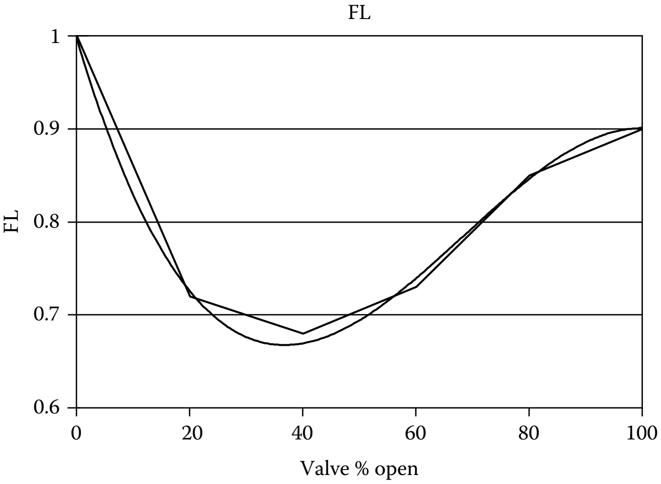


FIG. 6.20z
The recovery coefficient F_L as a function of the opening of a reduced ported pinch valve.

Some manufacturers suggest that one way to avoid cavitation is to size the control valve, not for the available total pressure drop ($P_1 - P_2$), but the pressure drop that would cause choking, or some fraction of that ($0.9\ \Delta p_{\text{choked}}$). This approach will redistribute the total system drop and eliminate cavitation in the valve (see C in Figure 6.2aa), but it does it by oversizing the valve, which will also deteriorate the quality of control.

Metal valves and rubber-lined valves respond differently to cavitation. Metal valves begin to deteriorate at a point between incipient cavitation and choked flow. Ironically, rubber fares better than metal under cavitation conditions. Rubber absorbs much of the shock of the imploding bubbles, so pinch valves can wade far into the chocked flow zone without any adverse effects.

Limiting or Eliminating the Damage

Figure 6.20aa illustrates five ways how cavitation can be eliminated by making changes in the installation or in the valve design.

One pinch valve manufacturer provides a calculated approximation based on the available total pressure drop ($P_1 - P_2$) and the pressure drop that would just begin to cause choking (Δp_{choked}). They argue that the difference between these two pressure drops is the energy source that drives cavitation (the formation and collapsing of bubbles), and therefore a cavitation damage factor (G) can be calculated by using this difference:

$$G = [(P_1 - P_2) - (\Delta p_{\text{choked}})](Q/17.44\text{d}) \qquad \mathbf{6.20(4)}$$

It is claimed that if G is calculated by using Q in gpm units of ambient water and d as the diameter of a rubber pinch valve, Table 6.20bb will give acceptable selection guidance for cavitation.

Using a Choke Fitting Cavitation does not start at the throat of the valve, but develops past the throat where pressure

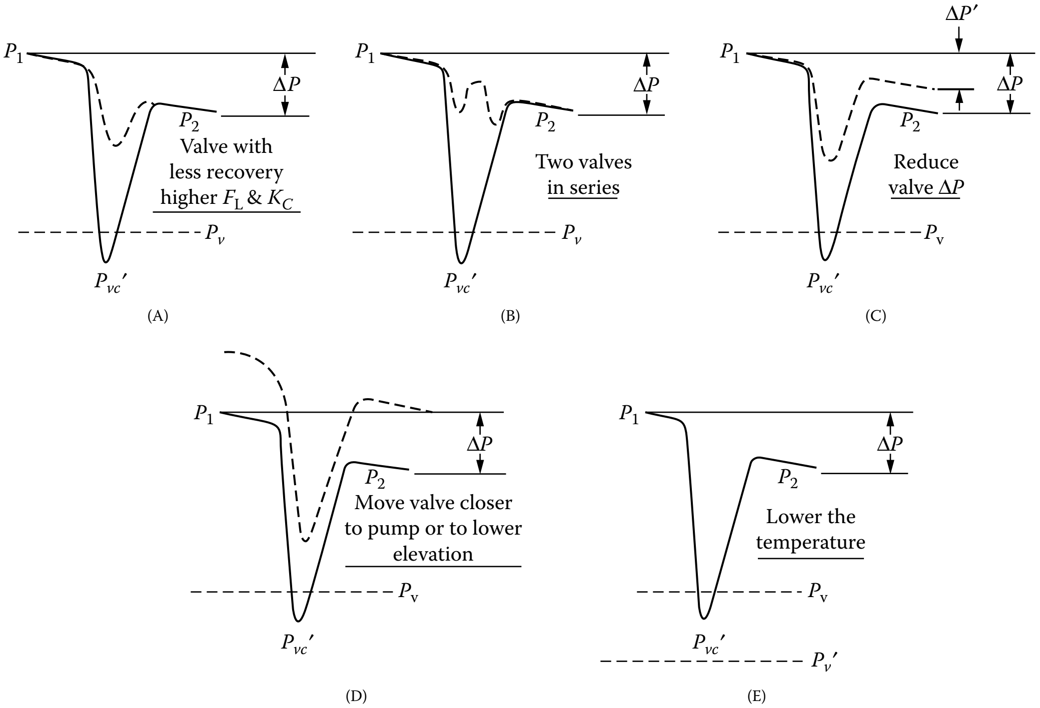


FIG. 6.20aa
Ways to eliminate cavitation: A) use valve with higher Kc and F_L, B) use two valves, C) reduce valve pressure drop, increase valve size, D) move valve to lower elevation or closer to pump, and E) reduce the temperature of the flowing process fluid.

begins to recover. As a result, cavitation damage occurs downstream of the valve throat and destroys the pinch valve's sleeve between the pinch point and the valve exit. Modulating pinch valves are usually supplied with a reduced port sleeve. The sleeve is molded with a venturi shape that tapers down to a small diameter in the center of the valve. The sleeve is normally symmetrical, tapered on both inlet and outlet so the valve can be installed for either flow direction. This design is vulnerable to cavitation damage.

Revising the sleeve design to an asymmetrical shape will enable the valve to withstand high levels of cavitation (Figure 6.20cc). This design is called a "trumpet mouth" or "cone sleeve." The thicker rubber at the valve outlet can absorb more cavitation damage. There are several ways to exploit this feature to tolerate even higher levels of cavitation with minimum sleeve damage.

One method is to install the valve at the end of the piping run, allowing the pressure to recover as the fluid emerges from the piping system. Here the cavities form after they have emerged from the choke fitting, where they cascade harmlessly into the open tank.

If it is not practical to install the valve at the end of the pipe run, some pinch valve manufacturers recommend the use of a trumpet mouth-type valve, followed by a short spool piece of pipe having a diameter that equals the valve's port size. The minimum spool piece length is typically 10 times the nominal diameter of the valve. The spool is connected to

TABLE 6.20bb
The Predicted Degree of Cavitation Damage Based on the Size of the "G Factor"

Cavitation Damage Factor (G)	*Predicted Degree of Cavitation Damage*
$0 \le G \le 100$	Cavitation damage will be undetectable.
$100 \le G \le 300$	Some damage will be observable after 1 year of operation.
$300 \le G \le 500$	Sleeve damage will be observable in 6 months.
$500 \le G \le 750$	Frequent maintenance and sleeve replacement is likely.
$750 \le G$	Consult factory.

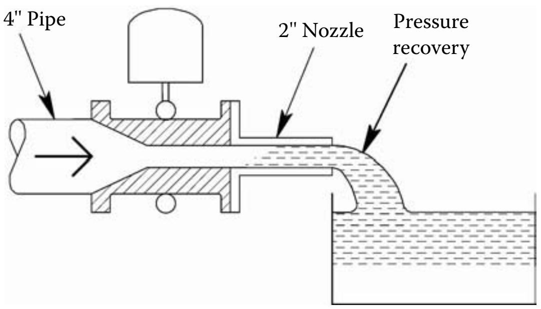


FIG. 6.20cc
If the valve can be located at the end of a pipeline, the addition of a choke fitting at the end of an asymmetric sleeve can move the zone of pressure recovery (cavitation) away from the valve.

the downstream full-size piping by a reducer flange, which creates a sharp, sudden expansion.

This design is not desirable because it does not eliminate the cavitation, but only moves its location downstream of the valve. Yet, some manufacturers argue that if the expansion downstream the spool is made with a rubber hose-type pipe joint, it can absorb a fair amount of wear and can be replaced once or twice a year at relatively low costs.

CONCLUSIONS

The phenomenon of cavitation is discussed in more detail in connection with control valve applications (Section 6.1), control valve noise (Section 6.14), and control valve sizing (Section 6.15). In general, one should note that the difference between the total available pressure drop ($P_1 - P_2$) and the pressure drop that would cause choking (Δp_{choked}) is the energy that drives cavitation (the formation and collapsing of bubbles). Therefore, instead of trying to find ways of harmlessly wasting this excess and unnecessary energy, one should concentrate on finding ways to NOT introduce it in the first place.

Bibliography

Carey, J. A., "Control Valve Update," *Instruments and Control Systems,* January 1981.

"Control Valves–Globe, Plug, Pinch, Needle, Gate," *Measurements and Controls*, February 1993.

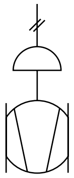
Dobrowolski, M., "Guide to Selecting Rotary Control Valves," *Instrumentation Technology,* December 1981.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems,* January 1980.

Gardellin, D., "Valve Cavitation," in *Encyclopedia of Chemical Processing and Design,* Vol. 61, Marcel Dekker, 1997.

6.21 Valve Types: Plug Valves

C. S. BEARD (1970) **R. D. BUCHANAN** (1985) **B. G. LIPÁTK** (1995, 2005)



Flow sheet symbol

Sizes:	1/2 to 36 in. (12.5 mm to 0.96 m)
Types:	V-ported, three-way, four-way, five-way, fire-sealed
Design Pressure:	Typically from ANSI Class 125 to ANSI Class 300 ratings and up to 720 PSIG (5 MPa) pressure, with special units available for ANSI Class 2500. The retractable seat design is suited for 10,000 PSIG (69 MPa) service.
Design Temperature:	Typically from −100 to 400°F (−73 to 204°C), with special units available from −250 to 600°F (−157 to 315°C)
Rangeability:	Refer to Section 6.7; generally 20:1
Characteristics:	See Figure 6.21a
Capacity:	$C_v = (25 \text{ to } 35)d^2$; see Tables 6.1a and 6.21b
Leakage:	Metal seats ANSI Class IV, composition seats ANSI Class V; see Table 6.1gg for definitions
Materials of Construction:	Iron, forged and alloy steel, chrome plating, 302 through 316 stainless steel, Alloy 20, Ni-resist, Monel, nickel, Hastelloy B and C, and zirconium, plus rubber or plastic, including PTFE linings
Costs:	Costs vary drastically with design and accessories. In general, the cost of conventional plug-type control valves is about half that of globe valves, while the cost of eccentric rotating plug valves is about the same as that of globe valves of the same size and materials. (For the costs of carbon steel and stainless steel globe valves, refer to Figure 6.19b.)
Partial List of Suppliers:	ABB Inc. (www.abb.com) Anchor/Darling Valve Co. (www.flowserve.com) Cashco Inc. (www.cashco.com) Circle Seal Controls Inc. (www.circle-seal.com) Dezurik/ SPX Valves & Controls (www.spxvalves.com) Emerson Process Management (www. emersonprocess.com) FMC Fluid Control Div. (www.fmcblending.com) Halliburton Services (www.halflow.com) Honeywell Industrial Controls (www.honeywell.com) Hydril Co. (www.hydril.com) Jordan Valve (www.jordancontrols.com) Mar-In Controls (www.mar-in-controls.com) Nordstrom Valves Inc. (www.nordstromaudco.com) Offshore Technology (www.offshore-technology.com/index.html) Spirax Sarco Inc. (www.spiraxsarco-usa.com) Xomox/Tufline (www.xomox.com)

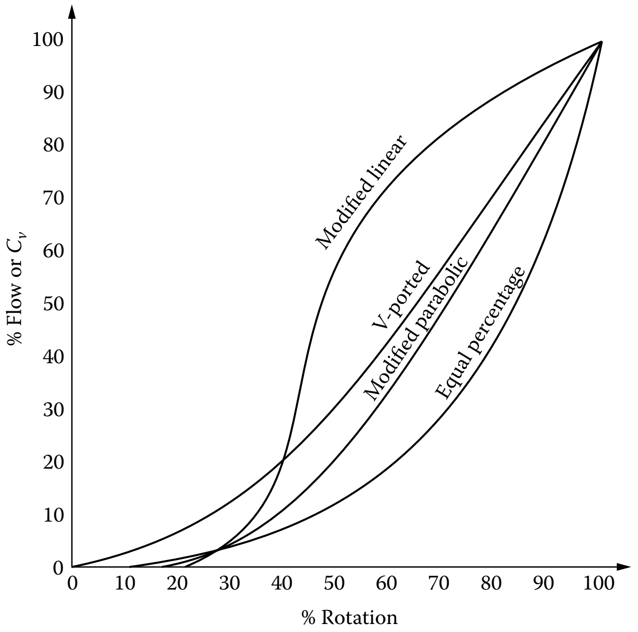


FIG. 6.21a
Plug valve characteristics are a function of the type of the particular V-port or of the shape of the throttling plate used.

GENERAL CHARACTERISICS

The rotary plug valves (similar to ball and butterfly valves) used to be considered only as on/off shutoff valves, but today they are also used as control valves. Table 6.1a shows how they compare in their characteristics and applications to some of the other control valve designs. Plug valves are well suited for corrosive, viscous, dirty, fibrous, or slurry services, while they are generally not recommended for applications where cavitation or flashing is expected.

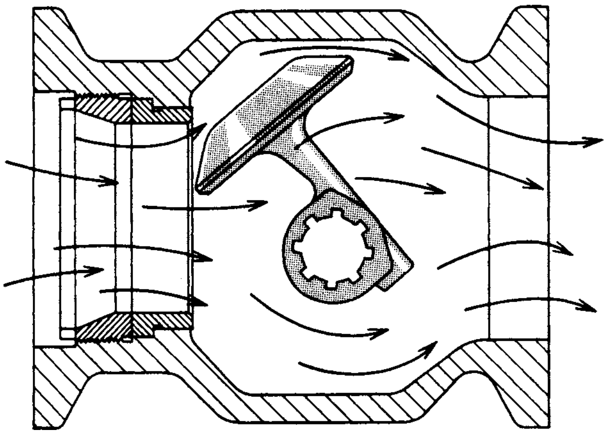


FIG. 6.21c
The self-cleaning nature of the eccentric rotating plug valve makes it a good option for slurry service applications.

Relative to the traditional globe valve, the advantages of conventional plug valves include their lower cost and weight, higher flow capacity, which can be two to three times that of the globe valve, if the plug is not characterized. In addition, they provide tight shutoff, fire-safe designs, and low stem leakage, which meets OSHA and EPA requirements.

The designs using characterized or eccentric rotating plugs provide good control performance and a self-cleaning flow pattern, which also reduces noise and cavitation. Actually, the performance of the rotating spherical segment-type valve is just as good as that of a globe valve, and for that reason some refer to it as a globe valve (Figure 6.21c).

In addition to their superior stem-sealing capability, plug valves are also suited for such corrosive applications as chlorine, phosgene, hydrofluoric acid, and hydrochloric acid. Plug valves are widely used on lethal and toxic services and can be made fire-safe by the use of Grafoil packing and can meet the external leakage requirement limits of API 607.

TABLE 6.21b
*Valve Coefficient (C_v) Values of Standard and Characterized Plug Valve Designs**

Valve Size-inch (mm)	*½ (12)*	*¾ (19)*	*1 (25)*	*1½ (38)*	*2 (51)*	*3 (76)*	*4 (102)*	*6 (153)*	*8 (224)*	*10 (254)*	*12 (305)*
Standard	9	9	25	55	144	254	433	900	1400	2100	3100
Modified parabolic plate				33	51	115	201	402	625	940	1385
Modified linear plate				12	20	46	81	162	252	380	560
Three-way	7	7	20	40	70	100	175	350	475	650	
Equal-percentage cage			20	44	97	178	307	641			
Modified parabolic cage			12	33	51	115	201	402			
Modified linear cage			5	12	20	46	81	162			
V-ported			35		150	320	560	1200			

*Courtesy of Xomox Corp./Tufline and DeZurik.

Plug Valve Features

The plug valve is a type of quarter-turn valve that is among the oldest designs known in engineering. Wooden plug valves were used in the water distribution systems of ancient Rome and probably predate the butterfly valve. Although no longer as popular as ball or butterfly valves, they lend themselves to special designs that work very well in specific control applications.

Conventional plug valves are usually lower cost and lighter weight than comparable gate or globe valves. Plug valves afford quick opening or closing with tight, leakproof closures under conditions ranging from vacuum to pressures as high as 10,000 PSIG (69 MPa). Some, including the various characterized and Y-ported or diamond design, can be used for throttling, while others, like the multiport, are used for diverting and bypass applications.

Plug valves are used on gas, liquid, and nonabrasive slurry services. Lubricated plug valves can also be used for abrasive slurries, and eccentric plugs are also used on applications involving sticky fluids. Plug valves are also used for applications requiring the contamination-free handling of foods and pharmaceuticals. In general, plug valves can handle applications with the following requirements:

- High flows at low pressure drops
- Low flow control
- Flow diversion
- High- or low-temperature applications
- Vibration-free operation
- Throttling control, only with eccentric and V-ported characterized designs.

The conventional plug valves are generally undesirable for the following types of applications:

- Flow modulation or continuous, exact flow throttling
- Maintenance-free operation (occasional lubrication is usually required and plugs may wear)

Throttling and Actuator Considerations

When used for throttling service some of the above-mentioned advantages of rotary valves, such as their high-capacity, can become disadvantages. Their high flow capacity can result in installations where small valves are mounted in large pipes. This results in a substantial waste of pumping energy, as the pump has to overcome the reducer pressure drops.

Also, the high-pressure recovery provided by most plug valve designs results in low vena contracta pressures, which in turn increase the probability of cavitation and noise. These problems, which are even more pronounced with pinch, butterfly, and ball valves, have been reduced by various means. In case of ball valves, perforated parallel plates have been inserted into the ball valve openings (Figure 6.16o) or flutes have been added to the butterfly disc (Figure 6.17s); however, the cavitation problems associated with high recovery valve designs have not been fully resolved.

In operating rotary valves, the linear movement of cylinder or spring-and-diaphragm actuators must be converted by linkages, which introduces hysteresis and dead play. In addition, a nonlinear relationship exists between actuator movement and the resulting rotation. These considerations make the use of positioners essential, which on fast processes can lower the quality of control. The torque characteristics of these valves are nonlinear (Figure 6.4v), and because of the high "break-torque" requirements, the actuators can be oversized relative to the torque requirements in the throttling range.

DESIGN VARIATIONS

The first plug valves consisted of a tapered or straight vertical cylinder containing a horizontal opening or flow-way inserted into the cavity of the valve body (Figure 6.21d). They have developed through time into numerous shapes and patterns, depending on the application, but almost all are adaptations of the cylindrical or tapered plug. Within that plug, however, the ports may be round, oval, rectangular, V-, or diamond-shaped, and can be the flow-through type two-way valves or multiport. These make up the special designs described in subsequent paragraphs.

Plug valve designs can be categorized as lubricated or nonlubricated. In the lubricated type, the thin film of lubricant serves not only to reduce friction between the plug and the body, but also to form an incompressible seal to prevent gas or liquid leakage. Because the seating surfaces are not exposed in the open position, gritty slurries may be handled. The lubricant hydraulically lifts the plug against the resilient packing to prevent sticking. A special lubricant must be injected periodically while the valve is either fully open or fully closed.

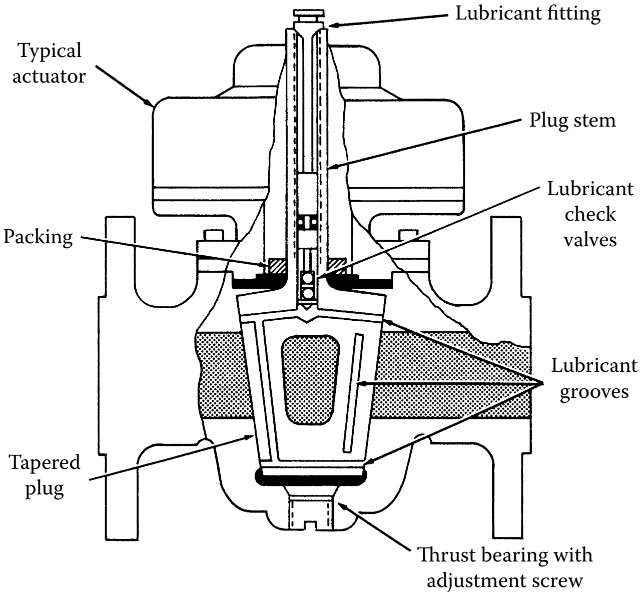


FIG. 6.21d
Conventional, lubricated plug valve with tapered plug.

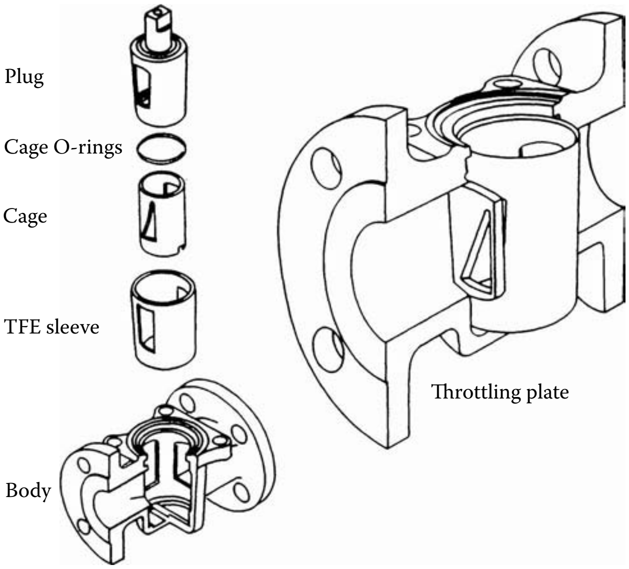


FIG. 6.21e
Throttling plates and characterizing cages can both be used to modify the inherent characteristics of plug valves. (Courtesy of Xomox/Tufline.)

The plugs of nonlubricated plug valves are treated with coatings such as Teflon or are specially heat-hardened and polished to prevent sticking. Often they are constructed so the tapered plug may be lifted mechanically from the seat for easier operation.

Characterized Plug Valves

Plug valves can be characterized by the use of characterizing cage or plate inserts (Figure 6.21e). The resulting characteristics are a function of the shape of the opening on the cage or plate.

Some examples of available plug valve characteristics are illustrated in Figure 6.21f. Rotation of the plug is inside a TFE sleeve, which is locked into the body in such a way that recessed areas minimized. Although a rangeability of 20:1 is claimed, this is made possible only if the valve can be fully open in order to provide full flow. The valve is available in 1/2–12 in. (12.5–300 mm) sizes and up to 600 PSIG ANSI (4.1 MPa) rating for use up to 400°F (204°C).

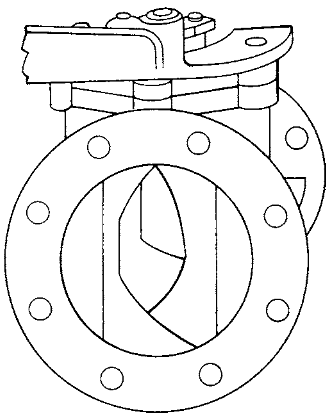


FIG. 6.21g
The design of a V-ported plug valve.

V-Ported Design

The V-ported plug valve (Figure 6.21g) is used for both on/off and throttling control of slurries and fluids containing solid concentrations in suspensions greater than 2%. These applications occur principally in the chemical and pulp and paper industries. A diamond-shaped opening is created by matching a V-shaped plug with a V-notched body.

Straight-through flow occurs on 90° rotation, when the plug is swung out of the flow stream. Shearing action and a pocketless body make the valve applicable for use on fibrous or viscous materials. The opening develops a modified linear flow characteristic with C_v capacities approximating 17d^2. Valves are flanged from 3–16 in. (75–400 mm) in bronze, corrosion-resistant bronze, or stainless steel. The body may be rubber-lined with a rubber-coated plug. A cylinder actuator and valve positioner are used for throttling control.

A variation is the true V-port opening (Figure 6.21h). It is obtained by a rotating segment that is closing against a

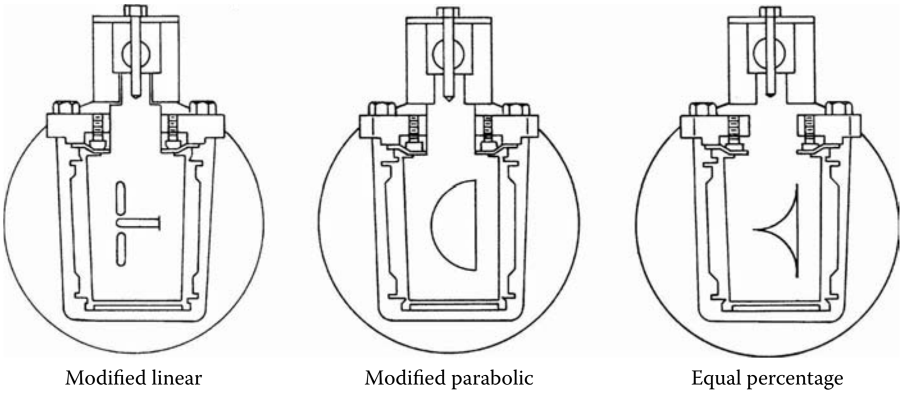
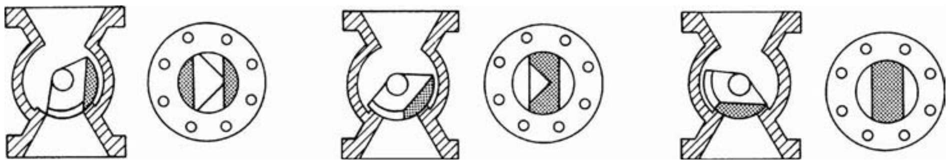


FIG. 6.21f
Plug valve characteristics can be modified to linear, parabolic, equal-percentage, and so on.



Open

There are no sharp corners or narrow openings to pack with stock. Large port area and clean interior design assure high flow capacity.

Throttling

Close plug-to-seat clearance remains constant. The V-orifice retains its shape throughout the cycle.

Closed

Flow continues until the V-orifice in the leading edge of the plug rotates past the seat, throttling is smooth down to the shutoff position.

FIG. 6.21h
Illustration of how throttling is provided by a V-ported plug valve.

straight edge. The valve can be smoothly throttled on thick stock flows without the stock packing or interfering. The valve is available in sizes from 4–20 in. (100–500 mm) and with C_v stated as more than $20d^2$. The valve is available in many body and trim materials for use in the chemical or pulp and paper industries. A cylinder-operated rack and pinion is used for on/off service with the addition of a valve positioner for throttling services.

Adjustable Cylinder Type

In another form of quarter-turn valve, flow is varied by rotating the core and by raising or lowering a “curtain” with an adjusting knob (Figure 6.21i). Proportional opening at any curtain position is made with the control handle, which may be attached to an actuator for automatic control. Various openings are obtained by these manipulations.

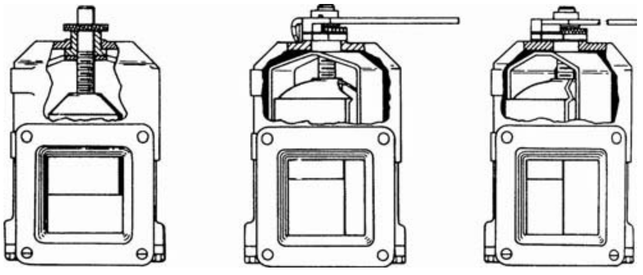
The valve is used for combustion control and for mixing applications, in which case valves are “stacked” on a common shaft or operated by linkages from the same actuator. To obtain linear flow with constant pressure drop, a port adjustment technique is used. After installation, the curtain is closed until the pressure drop across the valve is one sixth of the total pressure drop of the system using the control handle in wide open position. This provides a flow characteristic approximating linear without decreasing sensitivity by limiting valve stroke. The percentage of flow is equalized by manipulation of the linkage to the actuator.

Semispherical Plugs for Tight Closure

Various designs have been developed to obtain tight closure while eliminating the continuous friction of the seals during rotation, as with most ball valves. The valve design illustrated in Figure 6.21j uses an eccentric ball. In the closed position, the rectangular end of the stem protrudes into the ball and the closure face is wedged toward the seating surface. As the stem is rotated to open the valve, the closure surfaces separate and the pin moves into a vertical slot so that rotation occurs. A nonlubricated seal can be provided with a primary Teflon seal enclosed in a body seat retainer ring. The valve is adapted for automatic operation by connecting the stem to a diaphragm actuator.

Expanding Seat Plate Design

As shown in Figure 6.21k, metal-to-metal or resilient seats can be provided in two seating segments. These segments are



Curtain ½ open, core full open

Curtain ¾ open, core rotated ¼

Curtain ¾ open, core rotated ½

FIG. 6.21i
Illustrations of both the curtain and the core openings of an adjustable cylinder-type plug valve.

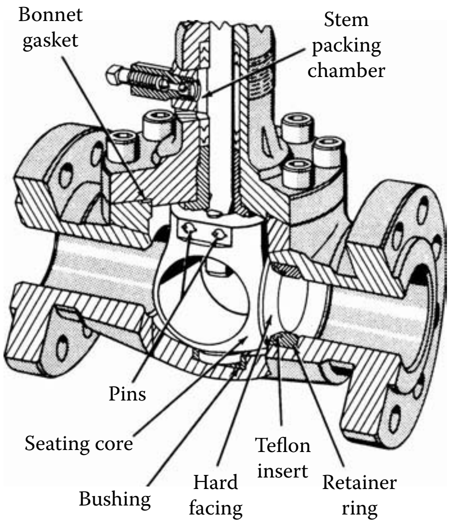


FIG. 6.21j
The semispherical plug in this plug valve design provides tight closure. (Courtesy of Offshore Technology, formerly Orbit Valve Co.)

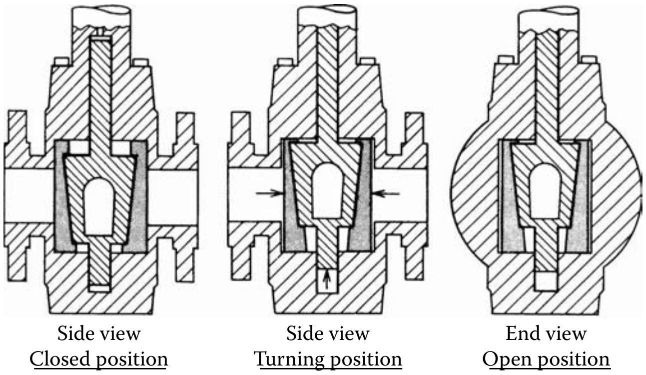


FIG. 6.21k
Plug valve with expanding seat plate design.

carried on a rail that is tapered so that downward stem movement forces the plates against the inlet and outlet ports. When the valve is being opened, the first few turns of the actuator cause the retraction of the plates and then plug rotation proceeds. These plates can be removed by merely removing the bottom plate of the valve V-body.

Retractable Seat Type

Positioning a movable seal after a spherical plug is in the closed position creates tight closure with sliding friction. In the retractable seat plug valve, a trunion-mounted partial sphere is operated by spur or worm gears (Figures 6.21l). The gear system rotates the plug until it is in a closed position, at which point additional rotation of the drive creates a camming action to compress the packing ring.

When the valve is being opened, the packing ring is released before the rotation of the plug occurs. Although also applicable to low-pressure service, this valve is particularly useful up to API 5000 PSIG (34.5 MPa) rating for 10,000 PSIG (69 MPa) service. It is used mostly in oil fields.

Overtravel Seating Design

By fabricating a cylindrical flow passage within a tapered cylindrical plug, a plug valve can be built at least up to the 60 in. size, without undue weight and attendant inertia to rotation. In the overtravel seating design shown in Figure 6.21m, the rotation of the inner valve is caused by a rod operated by a piston pushing down on a rotator. Continuous movement of the rod closes the plug, but leaves a small crescent, because the plug is slightly raised from seating. For tight closure, the rod contacts a seating adjustment to force the plug into a tapered seating surface.

In this design, the restriction of the flow area is rapid; 30% rod travel causes about 65% closure. Complete rotation of the plug occurs at about 65% stroke, with the additional stroke being utilized for seating. If rapid closure to about 20% opening does not create serious surge pressures, this valve can be used for emergency closure without much consideration of the piston speed. The rangeability of the valve exceeds 50:1 for proportional control. Free rotation and relatively low plug weight contributes to lower power requirements.

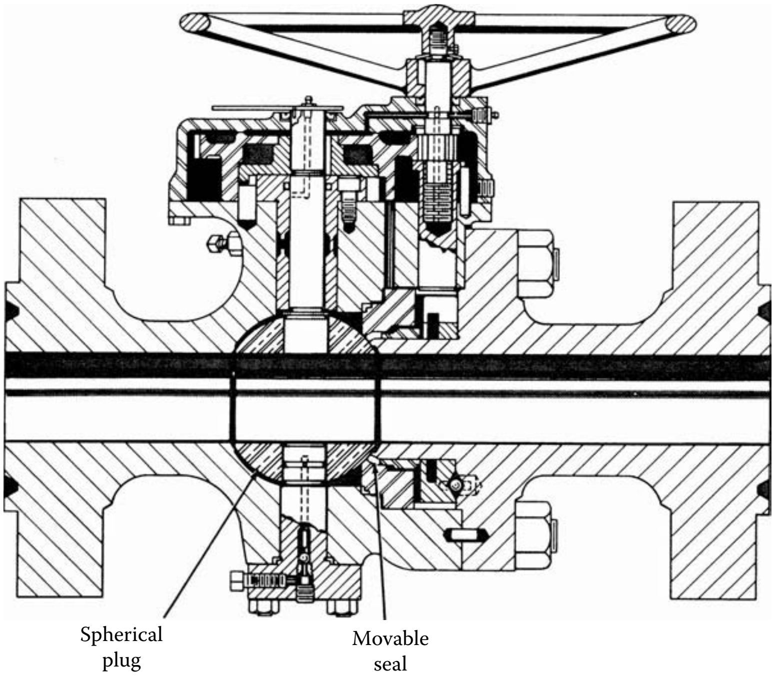


FIG. 6.21l
Spherical plug and retractable seat for tight shut-off. (Courtesy of Hydril Co.)

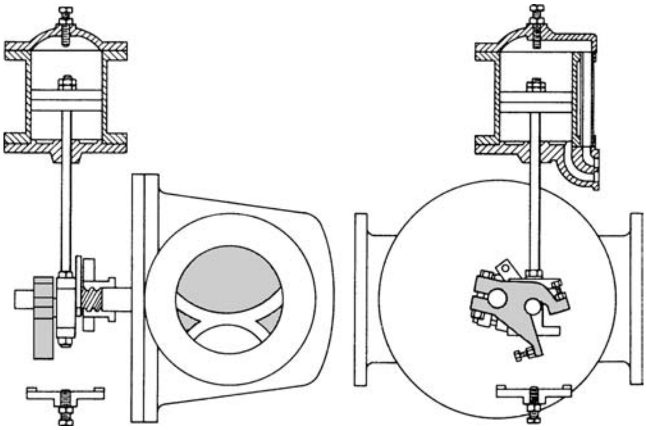
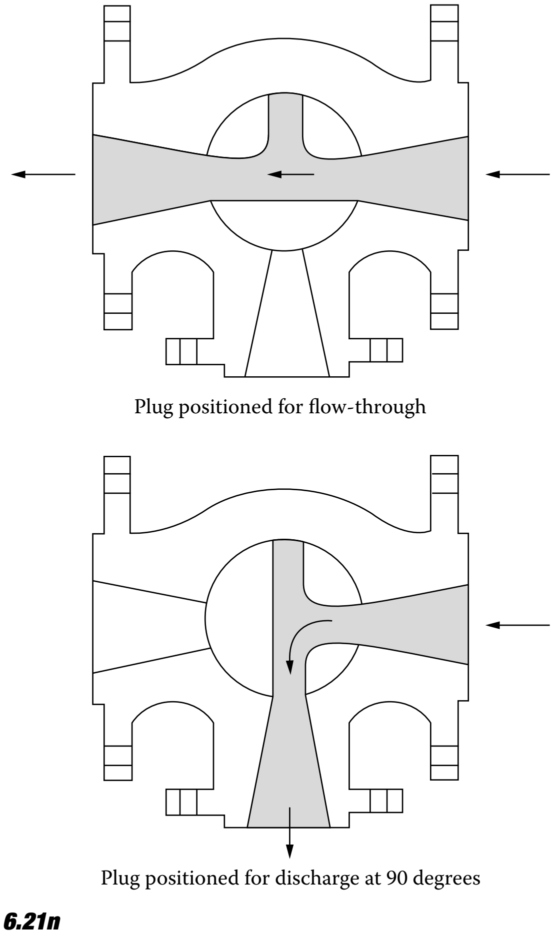


FIG. 6.21m
Two views of a plug valve of the overtravel seating design, shown in a throttling position.



Plug positioned for flow-through

Plug positioned for discharge at 90 degrees

FIG. 6.21n
The design of a three-way plug valve.

Multiport Design

As shown in Figure 6.21n, a three-way plug valve is obtained by providing the plug with an extra port at 90° from the inlet, so that flow can be directed in either of two destinations. A multitude of directions can be achieved by nesting combinations of the simple multiport valves or by using more complex designs. These include a multistoried arrangement with the plug extending upward to connect to a series of tiered outlets. In such multistoried configurations, the plug has a long, vertical passageway connecting the horizontal ports.

Another method of increasing the number of flow-directions is to design the plug with a diameter that is sufficiently larger than the ports so that intermediate ports can be placed at 45° or even 30 and 60°. In that case, the actuators can be programmed to serve a variety of process applications.

Bibliography

Ball, K. E., “Final Elements: Final Frontier,” *InTech,* November 1986.

Baumann, H. D., “Trends in Control Valves and Actuators,” *Instruments and Control Systems,* November 1975.

Bialkowski, B., Coughran, M., Beall, J., “Control Valve Performance Update — The Last 10 Years,” *Pulp Pap-Canada, (*102):21–22, November 2001.

Bishop, T., Chapeaux, M., Jaffer, L., et al., “Ease Control Valve Selection,” *Chem Eng Prog,* 98 (11): 52–56, November 2002.

Borden, G. and Friedmann, P. G. (Eds.), *Control Valves — Practical Guides for Measurement and Control,* Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1998.

Cain, F. M., “Solving the Problem of Cavitation in Control Valves,” Paper #91-0462, 1991 ISA, Anaheim, CA, October 1992.

Control Valve Handbook, 3rd edition, Marshalltown, IA: Fisher Controls International Inc., 2001.

Control Valve Seat Leakage, FCI 70-2-1998, Cleveland, OH: Fluid Controls Institute, Inc., 1998.

“Control Valves—Globe, Plug, Pinch, Needle, Gate,” *Measurements and Control,* February 1993.

Dobrowolski, M., “Guide to Selecting Rotary Control Valves,” *Instrumentation Technology,* December 1981.

Hammitt, D., “Rotary Valves for Throttling,” *Instruments and Control Systems,* July 1977.

Hutchison, J. W. (ed.), *ISA Handbook of Control Valves*, Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1993.

Kawamura, H., “Selecting Valves for the Hydrocarbon Processing Industry,” *Hydrocarbon Processing,* August 1980.

Lyons, J., *Lyons’ Valve Designer’s Handbook,* New York: Van Nostrand Reinhold, 1982.

Monsen, J. F., “Valve Wars—Rising Stem vs. Rotary,” *Plant Services,* January 1999.

Pyotsia, J., “A Mathematical Model of a Control Valve,” 1992 ISA Conference, Houston, TX, October 1992.

Rahmeyer, W., “The Critical Flow Limit and Pressure Recovery Factor for Flow Control,” *InTech,* November 1986.

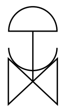
Shinskey, F. G., “Control Valves and Motors,” Foxboro Publication No. 413-8.

Zappe, R., *Valve Selection Handbook,* Houston, TX: Gulf Publishing, 1981.

Whatever Happened To... A Guide to Industrial Valve and Actuator Companies—Past and Present, Washington, D.C.: Valve Manufacturers Association, 1997.

6.22 Valve Types: Saunders Diaphragm Valves

C. E. GAYLOR (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

Design Types:	Weir, full-bore, straight-through, dual-range
Applications:	Slurries, corrosive fluids at low pressure drops
Sizes:	Standard units from $^1/_2$ to 12 in. (12 to 300 mm); special units up to 20 in. (500 mm)
Maximum Operating Pressure:	In sizes up to 4 in. (100 mm), 150 PSIG (10.3 bar); 6 in. (150 mm), 125 PSIG (8.6 bar); 8 in. (200 mm), 100 PSIG (6.9 bar); and 10 or 12 in. (250 or 300 mm), 65 PSIG (4.5 bar). See Figure 6.22a for limits.
Vacuum Limits:	Mechanical damage can occur when opening valve against process vacuum
Temperature Limits:	With most elastomer diaphragms from 10 to 15°F (–12 to 65°C); with PTFE diaphragm from –30 to 350°F (–34 to 175°C). See Figure 6.22a.
Materials of Construction:	*Body materials:* iron, ductile iron, steel, 302 to 316 stainless steel, Alloy 20, bronze, Monel, Hastelloy C, aluminum, titanium, graphite, plastic such as PTFE lining or solid plastics *Diaphragm materials:* Teflon, Buna-N, neoprene, hypalon
Characteristics:	See Figure 6.22b
Capacity:	$C_v = 20\ d^2$; see Table 6.22c
Rangeability:	About 10:1; see Section 6.7
Leakage:	ANSI Class IV or V; for definitions see Table 6.1gg
Costs:	Costs vary drastically with design and accessories. In general, the cost of conventional Saunders-type control valves is about half that of globe valves of the same size and materials. (For the costs of carbon steel and stainless steel globe valves, refer to Figure 6.19b.)
Partial List of Suppliers:	ABB Inc. (www.abb.com) Emerson Process Management (www.emersonprocess.com) Foxboro-Invensis (www.foxboro.com) ITT Industries, Engineered Valves (www.engvalves.com) McCanna/Marpac (www.mccannainc.com) Nibco Inc. (www.nibco.com) Teledyne Engineering (www.teledyne.com) Velan Valve Corp. (www.velan.com)

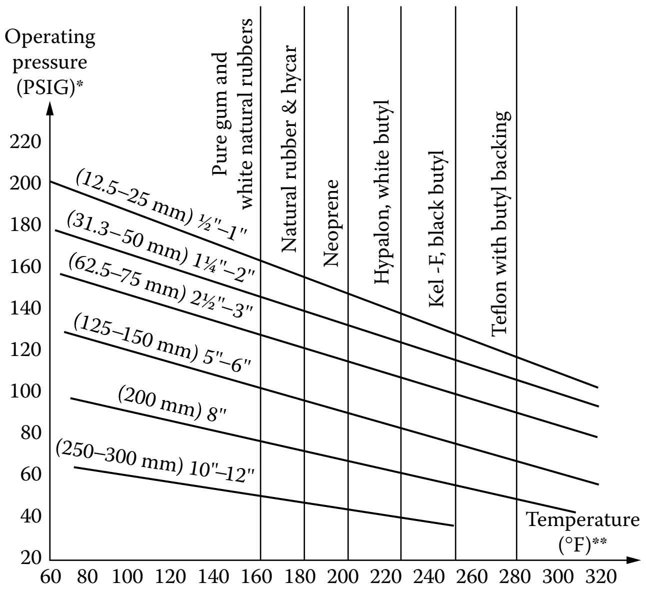
INTRODUCTION

The advantages of Saunders valves include their relatively low costs, tight shut-off, and suitability for corrosive, dirty, viscous, or slurry services (Table 6.1a). Their limitations include their poor control characteristics, although the use of the dual-range design improves it. They are also limited in their temperature (high and low) and pressure (high) ratings and are not suited for cavitating or flashing services.

The valve coefficients for a number of Saunders valve sizes with a variety of connection types and lining options are tabulated in Table 6.22c.

SAUNDERS VALVE CONSTRUCTION

The Saunders valve is also referred to as a diaphragm valve or less often as a weir valve. Conventional Saunders valves utilize both the diaphragm and the weir for controlling the



*1 PSIG = 6.9 kPa

$$**°C = \frac{°F - 32}{1.8}$$

FIG. 6.22a
Pressure and temperature limitations of the various diaphragm materials used in Saunders valves, as a function of valve size.

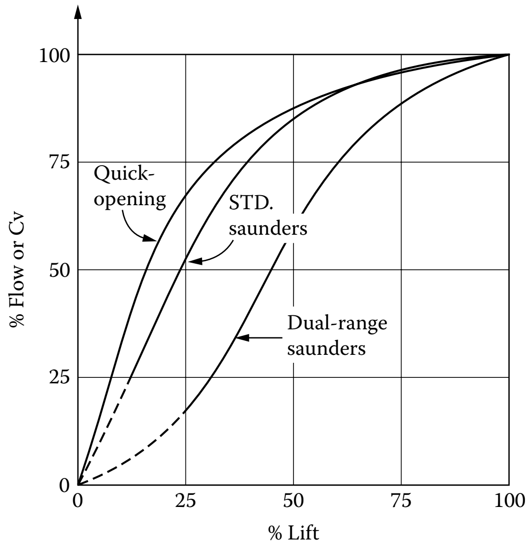


FIG. 6.22b
The characteristics of conventional Saunders valves are nearly quick opening, while the characteristics of dual-range Saunders valve designs are closer to linear.

flow of the process fluid (Figure 6.22d), while straight-through and dual-range ones do not necessarily use a weir.

The conventional Saunders valve is opened and closed by moving a flexible or elastic diaphragm toward or away from a weir. The elastic diaphragm is moved toward the weir by the pressure applied by a compressor element on the

TABLE 6.22c
*Valve Coefficients (C_v) of Conventional Saunders Valves**

Nominal Valve Size				*Flanged Ends*			
mm	*in.*	*Sanitary Ends*	*Threaded Ends*	*Unlined*	*Neoprene or Rubber Lined*	*Glass Lined*	*Polypropelene Kynar or Tefzel Lined*
15	$^1/_2$	3.5	2.6	3.5	2	3	—
20	$^3/_4$	19	8	21	7	22	—
25	1	21	16	21	7	22	11
30	$1^1/_2$	29	26	29	12	39	22
40	$1^1/_2$	50	44	50	26	54	39
50	2	66	66	66	54	83	69
65	$2^1/_2$	150	—	150	105	170	99
80	3	165	—	165	133	235	160
100	4	290	—	290	235	365	280
150	6	—	—	550	495	805	670
200	8	—	—	1050	1050	1625	1000
250	10	—	—	1725	1650	—	—
300	12	—	—	2250	2075	—	—

* Courtesy of The Foxboro Co.

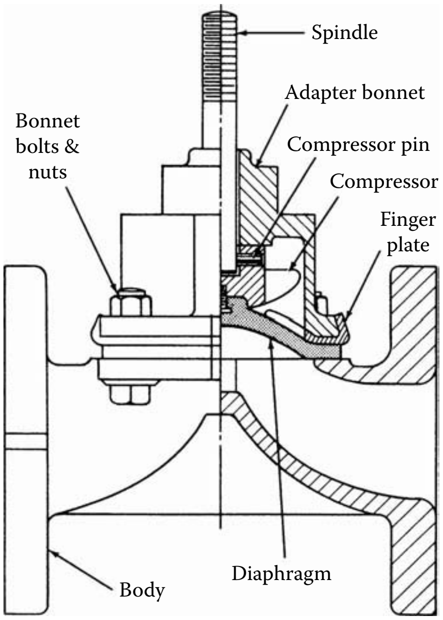


FIG. 6.22d
The main components of a weir-type Saunders valve.

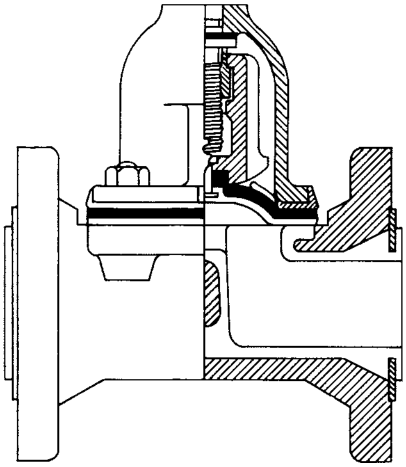


FIG. 6.22f
Teflon-lined Saunders valve.

diaphragm. The compressor is connected to the valve stem, which is moved by the actuator. The diaphragm, which at its center is attached to the compressor, is pulled away from the weir when the compressor is lifted.

For high-vacuum service it is often desirable to evacuate the bonnet of the Saunders valve in order to reduce the force that is pulling the diaphragm away from the compressor. This is especially desirable for large valves, where the vacuum might be sufficient to tear the diaphragm from the compressor.

The compressor is designed to clear the finger plate, or diaphragm support plate, and to contour the diaphragm so that it matches the weir (Figure 6.22d). The purpose of the finger plate is to support the diaphragm when the compressor has been withdrawn. The finger plate is utilized for valve sizes 1 in. (25 mm) and larger. For valves larger than 2 in. (50 mm), the finger plate is built as part of the bonnet.

A Saunders valve can be considered as a half pinch valve (Section 6.20). The pinch valve operates as if two diaphragms were moved toward or away from each other, whereas the Saunders valve has only one diaphragm and a fixed weir. Because of their design similarity, their flow characteristics (see Figures 6.20v and 6.20w for pinch valves) are also similar, as it was shown above in Figure 6.22b.

Figure 6.22e shows the three basic positions of a conventional Saunders valve.

Materials of Construction

The body of a conventional Saunders valve (Figure 6.22d), because of its simple and smooth interior, lends itself well to lining with plastics, glass, titanium, zirconium, tantalum, and other corrosion-resistant materials (Figure 6.22f). Valve bodies

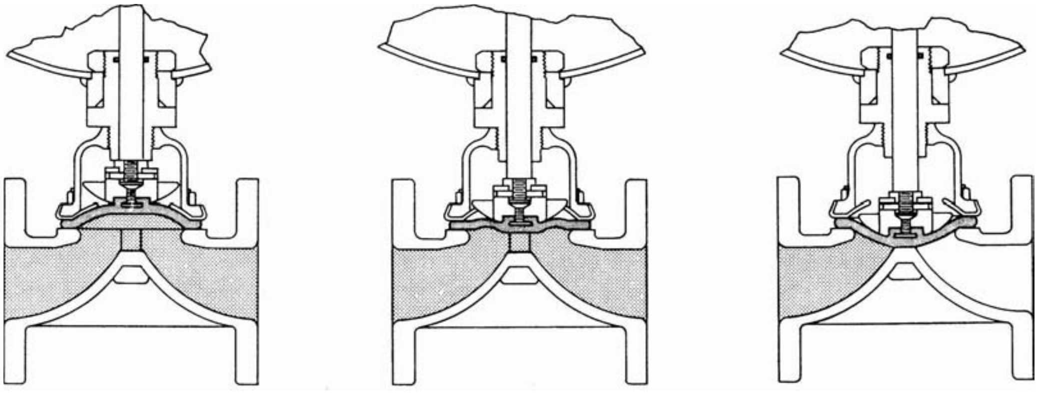

Streamline flow in open position
Flow control in throttling position
Leak-tightness in closed position

FIG. 6.22e
The open, throttling, and closed positions of a conventional Saunders valve.

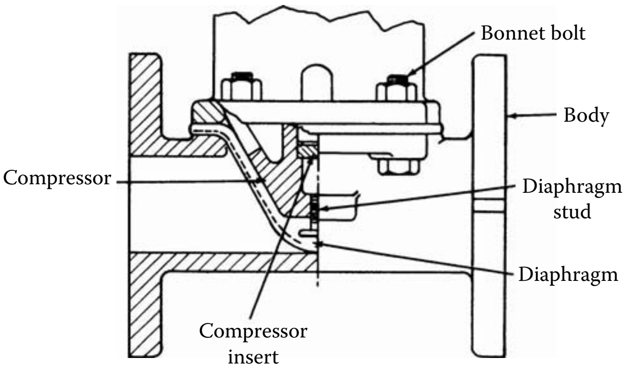


FIG. 6.22g
The design of a straight-through Saunders valve.

are available in iron, stainless and cast steels, alloys, and plastics. Iron bodies are lined with plastic, glass, special metals, and ceramics. As it was shown in Table 6.22c, lining lowers the valve capacity of smaller Saunders valves (under 2 in., or 50 mm) by about 25% below that of the unlined ones (Table 6.22c).

The diaphragm for the conventional Saunders valve is available in a wide range of materials. These include polyethylene, Tygon, white nail rubber, gum rubber, hycar, natural rubber, neoprene, hypalon, black butyl, KEL-F, and Teflon, with various backings, including silicone. Some of these diaphragms also contain reinforcement fibers.

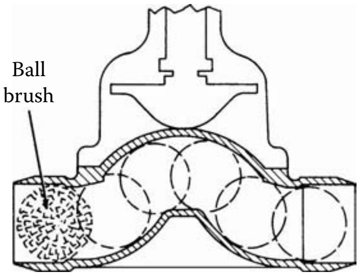


FIG. 6.22h
Full-bore Saunders valve.

Maintenance requirements of Saunders valves are mainly determined by the diaphragm life, which is a function of the diaphragm's resistance to the controlled process fluid (which may be corrosive or erosive) and also of the operating pressure and temperature (Figure 6.22a).

Straight-Through Design

The valve seat of the straight-through diaphragm valve is not the conventional weir. Here the compressor is contoured to meet the walls of the body itself (Figure 6.22g). The longer stem stroke of the straight-through valve necessitates a very flexible diaphragm. The increased flexure requirement tends to shorten the life of the diaphragm, but the valve's smooth, self-draining, straight-through flow pattern makes it applicable for hard-to-handle materials, such as slurry.

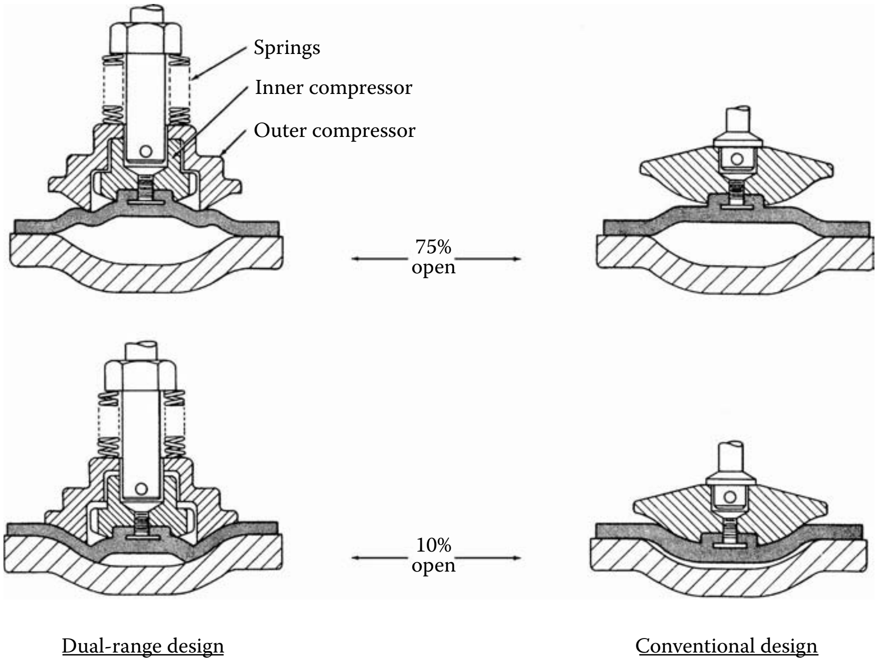


FIG. 6.22i
The shape of the openings of a 10 and 75% open Saunders valves are compared in the dual-range (left) and single-range (right) designs.

The flow characteristic of the straight-through design is more nearly linear than those of the conventional Saunders valves.

Full Bore Valve

The body of a full-bore Saunders valve is modified to provide a special shape to the weir. As a result, the opening of the internal flow path is fully rounded at all points, permitting ball brush cleaning (Figure 6.22h). This is an important feature in the food industry, where a smooth, easy-to-clean interior surface is required.

Dual-Range Design

The rangeability and flow characteristics of a conventional Saunders valve are rather poor, and so it is not suitable if high precision control is required. The flow characteristics of the dual-range design is an improvement, in comparison to the characteristics of the conventional Saunders.

The dual-range valve contains two compressors, which provide independent control over two areas of the diaphragm (Figure 6.22i). The first increments of stem travel raise only the inner compressor from the weir. This allows flow through a contoured opening in the center of the valve. This is superior to the operation of the single-range design, where the corresponding flow is the result of a slit across the entire weir.

This improvement in the shape of the value opening helps prevent clogging and the dewatering of stock and it also keeps abrasion at a minimum. In this dual-range design, while springs hold the outer compressor firmly seated, the inner compressor may be positioned independently to provide accurate control over small amounts of flow.

When the inner compressor is opened to its limit, the outer compressor begins to open. From this point on, both compressors move as a unit. When wide open, this valve provides the same flow capacity as its conventional counterpart.

Bibliography

Bialkowski, B., Coughran, M., and Beall, J., "Control Valve Performance Update—The Last 10 Years," *Pulp Pap-Canada,* (102):21-22, November 2001.

Bishop, T., Chapeaux, M., Jaffer, L., et.al., "Ease Control Valve Selection," *Chem Eng Prog,* 98 (11): 52–56, November 2002.

Borden, G. and Friedmann, P. G. (eds.), *Control Valves—Practical Guides for Measurement and Control,* Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1998.

Cain, F. M., "Solving the Problem of Cavitation in Control Valves," Paper #91-0462, 1991 ISA Conference, Anaheim, CA, October 1992.

Carey, J. A., "Control Valve Update," *Instruments and Control Systems,* January 1981.

Control Valve Handbook, 3rd edition, Marshalltown, IA: Fisher Controls International Inc., 2001.

Control Valve Seat Leakage, FCI 70-2-2003, Cleveland, OH: Fluid Controls Institute, Inc., 2003, (www.fluidcontrolsinstitute.org).

"Control Valves—Globe, Plug, Pinch, Needle, Gate," *Measurements and Control,* February 1994.

Cunningham, E. R., "Solutions to Valve Operating Problems," *Plant Engineering,* September 4, 1980.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems*, January 1980.

Monsen, J. F., "Valve Wars—Rising Stem vs. Rotary," *Plant Services,* January 1999.

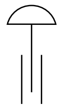
Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

Sanderson, R. C., "Elastomer Coatings: Hope for Cavitation Resistance," *InTech,* April 1983.

Whatever Happened To... A Guide to Industrial Valve and Actuator Companies—Past and Present, Washington, D.C.: Valve Manufacturers Association, 1997.

6.23 Valve Types: Sliding Gate Valves

C. S. BEARD (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

Types:	A. Knife gate B. V-insert C. Plate and disc (multi-orifice) D. Positioned disc
Sizes:	A. On/off; 2 to 120 in. (50 mm to 3 m) B. Throttling: $^1/_2$ to 24 in. (12 to 600 mm) C. Throttling: $^1/_2$ to 6 in. (12 to 150 mm) D. Throttling: 1 to 2 in. (25 and 50 mm)
Design Pressures:	A and B. Up to ANSI Class 150; higher with "wedge within wedge" design C. Up to ANSI Class 300 D. Up to 10,000 PSIG (69 MPa)
Design Temperatures:	A and B. Cryogenic to 500°F (260°C) C. −20 to 1125°F (−29 to 607°C)
Rangeability:	A. 10:1 B. 20:1 C. Up to 50:1 is claimed; see Section 6.7
Characteristics:	See Figure 6.23a
Capacity:	A. $C_v = 45\ d^2$ B. $C_v = 30\ d^2$ C. $C_v = (6 \text{ to } 10)\ d^2$ See Table 6.23b
Leakage:	A and B. ANSI Class I or II with metal seat; better with soft seat or lining C. ANSI Class IV; see Table 6.1gg for definitions
Materials of Construction:	A and B. Ductile iron, cast iron, carbon steel, 304, 316, 317 stainless steel, Alloy 20, Hastelloy B or C. Seating can be metal to metal, nylon, or RTFE C. Body: Ductile iron, bronze, carbon steel, stainless steel, aluminum, Monel Trim: Stainless steel is standard, which can be chrome-plated for hardness or Teflon-coated for corrosion resistance; Monel or Hastelloy trims are also available
Costs:	The cost of V-insert-type slide gate valves is similar to, but generally less than that of, single-seated globe valves, which are given in Figure 6.19b. The cost of plate and disc valves is given in Table 8.23j.
Partial List of Suppliers:	Anchor/Darling Valve Co. (www.flowserve.com) DeZurik/SPX Valves (www.spxvalves.com) ITT Industries, Engineered Valves (www.engvalves.com) Jordan Valve (www.joprdanvalve.com) Kurimoto Valves (www.kurimoto.co.jp) Red Valve Company Inc. (www.redvalve.com) Richards Industries (www.richardsind.com) Stockham Valves & Fittings (www.stockham.com) Zimmermann & Jansen Inc. (www.zjnc.com)

INTRODUCTION

The knife gate-type slide gate valves are relatively inexpensive, have high capacity, and are suited for slurry and dirty services. On the other hand they have poor control characteristics, do not provide tight shut-off, and are not suited for corrosive services. The V-insert type variation of this design has similar features, but as illustrated in Figure 6.23a, has better control characteristics.

The positioned sliding disc designs are ideal for high-pressure (up to 10,000 PSIG), cavitating, abrasive, or erosive services, but are relatively expensive and are not suited for sludge, slurry, viscous, or fibrous services. The multiport plate and disc type valves are similar, but provide superior control characteristics. These valves are available as pump governors or as unusually high rangeability (> 200:1) control valves in sizes from 0.5 to 6 in.

SLIDING GATE VALVE DESIGNS

Knife Gate Valves

Changing the process fluid's flow rate by sliding a plate past a stationary hole is one of the oldest and most basic approaches to throttling flows. The most common valve, the sliding gate valve, operates like this. Although occasionally used for automatic control, it is not considered to be a throttling control valve. It is a form of "guillotine"-type gate valve (Figure 6.23c) and is much used in the pulp and paper industry due to its shearing ability and nonplugging body design.

The "slab-type" sliding gate is provided with a round opening (Figure 6.23d) and therefore its characteristics are determined by the two converging circles. Its characteristics approximates equal-percentage behavior up to about 70% of its flow capacity, and above that it becomes nearly linear. The flow rate of 70% is reached by opening the valve to about 30% of its stroke.

On critical services, such as in catalytic cracking, reforming, isomerization, or coal gasification applications, the "double" gate valves are often considered (Figure 6.23e). In this

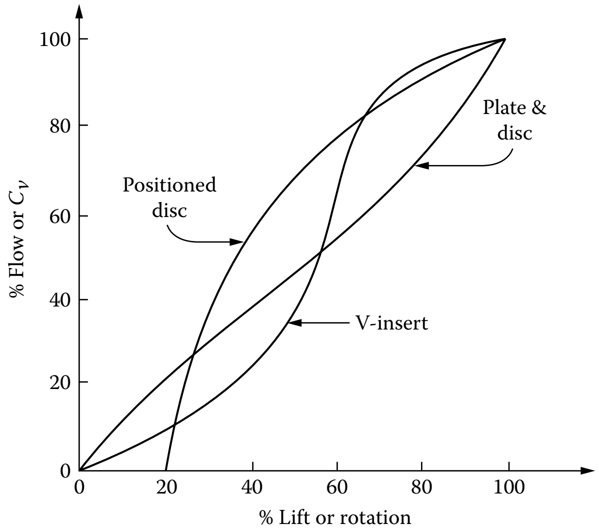


FIG. 6.23a
The characteristics of the various types of sliding gate-type valve designs.

TABLE 6.23b
Valve Capacity Coefficients of Open Gate, V-Insert, or Plate-and-Disc-Type Slide Gate Valves

	Valve Design		
Size in. (mm)	Plate and Disc	Open Knife Gate	V-Insert Type
0.5 (12.5)	0.0005 to 2.5		
1.0 (25.4)	0.0005 to 9		
1.5 (38)	9.0 to 17.5		
2 (51)	17.4 to 30.5	161	88
3 (75)		362	196
4 (100)		646	353
6 (150)		1450	794
8 (200)		2570	1410
10 (250)		3450	1935
12 (300)		5060	2850
14 (350)		6630	3725
16 (400)		8680	4870
18 (450)		11,350	6370
20 (500)		14,000	7850
24 (600)		20,550	11,550
30 (750)		31,900	17,900
36 (900)		45,700	26,200

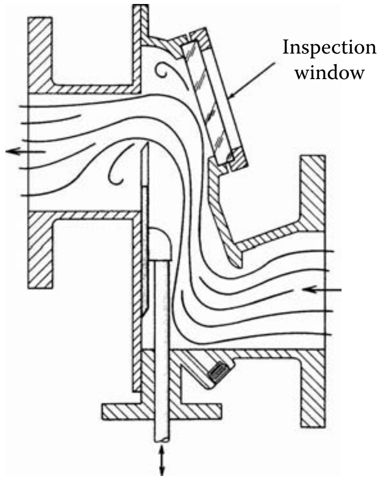


FIG. 6.23c
The guillotine-type sliding gate valve.

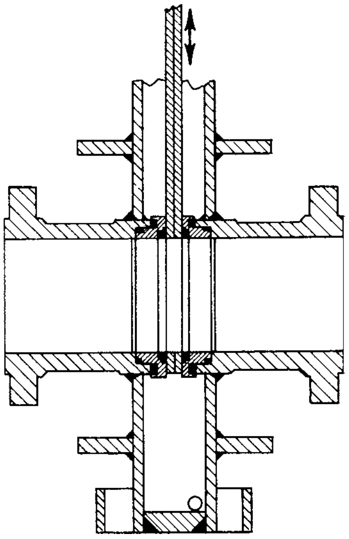


FIG. 6.23d
The design of a slab-type sliding gate valve.

design a "wedge" is provided within the wedge-shaped sliding gate. This inner wedge forces the two sliding gates on its two sides against the two seats, thereby guaranteeing tight closure.

V-Insert Type The addition of a V-shaped insert (Figure 6.23f) in the valve opening creates a parabolic flow characteristic. As shown in Figure 6.23a, this characteristic is somewhat similar to that of the V-ported globe valve. The performance of these valves is much dependent on the type of actuator and positioner used, because the quality of control is dependent upon the ability to provide very accurate positioning of the sliding gate.

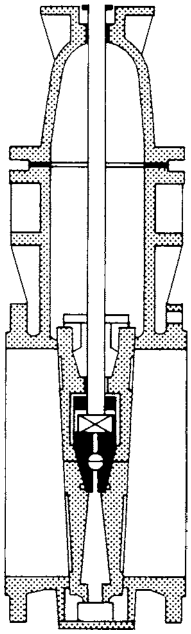


FIG. 6.23e
The design of the double gate valve, which is also called the "wedge in the wedge" design. (Courtesy of Zimmermann & Jansen Inc.)

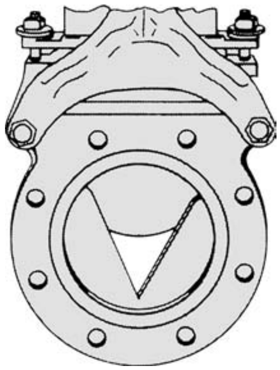


FIG. 6.23f
Sliding gate valve with V-insert.

Positioned-Disc Valves

Rotation of a movable disc with two holes, which if rotated can progressively cover two holes in the stationary disc, can successfully throttle flow (Figure 6.23g). This variable choke was designed to control flow from high-pressure oil wells.

The use of ceramic or tungsten carbide discs allows it to handle pressures up to 10,000 PSIG (69 MPa). Such valves are presently furnished in 1 and 2 in. (25 and 50 mm) sizes with areas from 0.05 in.^{2} with 0.25 in. hole (32 mm^2 with 6.3 mm hole) to 1.56 in.^{2} with two 1 in. holes (1006 mm^2 with two 25 mm holes).

An angle version of this valve design is used for proportioning control, with an actuator capable of controlling the discharge flow at quarter-turn movement. Both linear and rotary type actuators can be used. The valve opening (the relationship between the discs) remains in the last position if power fails.

A stepping actuator (Figure 6.23h) positions the inner valve disc in 1° increments as a function of a pneumatic controller input to a double-acting, spring-centered piston. Rotation occurs through a rack and pinion assembly. Limit

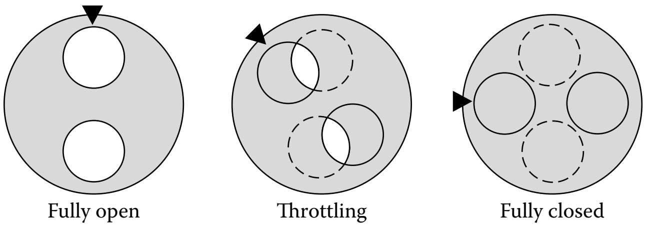


FIG. 6.23g
High-pressure process streams can be throttled by the positioned-disc-type slide gate valve.

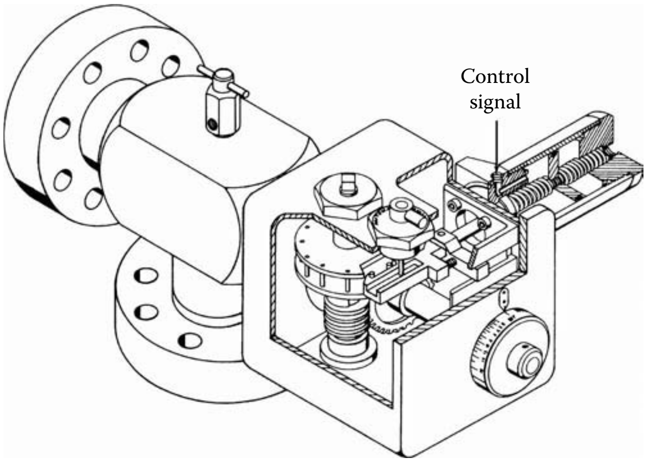


FIG. 6.23h
The actuator of a positioned-disc-type angle valve. (Courtesy of Willis Oil Tool Co.)

switches are provided, and a stepping switch can be used for position transmission and for position feedback in automatic control systems.

Plate and Disc Valves

A wide variety of flow characteristics are available using a stationary plate in the valve body and a disc that is moved by the valve stem. The plate (Figure 6.23i) is readily replaceable by removing a flanged portion of the body, which retains the plate with a pressure ring. Areas of the plate are undercut to reduce friction. A circumferential groove provides flexibility and allows the plate to remain flat in spite of differential pressures or expansion or contraction of the body. The stem contacts the disc by a pin through a slot in the plate.

The disc is held in contact with the plate by upstream pressure and by retaining guides. The contacting surfaces of the disc and plate are lapped to light band flatness. The chrome-plated surface of the stainless steel plate has a hardness comparable to 740 Brinell to resist galling and corrosion and obtain smooth movement of the disc. The material, with the registered name of Jordanite, is reported to have an extremely low coefficient of friction, is applicable to high-pressure drops, and has great resistance to heat and corrosion.

Flow occurs through mating slots in the disc and the plate. Positive shut-off occurs when the slots are separated (Figure 6.23i). Flow increases on an approximately linear relationship until the slots are lined up for maximum flow. Capacities are about $C_v = 6.5\ d^2$ through the 2 in. (50 mm) size and about $C_v = 12\ d^2$ through the 6 in. (150 mm) size.

Stem travels to obtain full flow are very short due to the slot relationship, and low-lift diaphragm actuators can be used for positioning. Forces needed for positioning are low, requiring only sufficient power to overcome friction between the plate and disc, which is right-angle motion and not opposed to the direction of flow.

Valve bodies are offered in sizes between $^1/_4$ and 6 in. (6.3 and 150 mm) and with ratings through 300 PSIG (2.1 MPa), depending on the material, with a selection of trims and packings. Many styles of actuators are used, including one with a thermal unit and cam actuation. This body design has been adapted for extensive use in self-contained pressure or temperature regulators.

This valve is also used to control the steam flow to steam-driven pumps, so as to maintain the pump discharge pressure constant. The sizes and costs of these pump governors are given Table 6.23j.

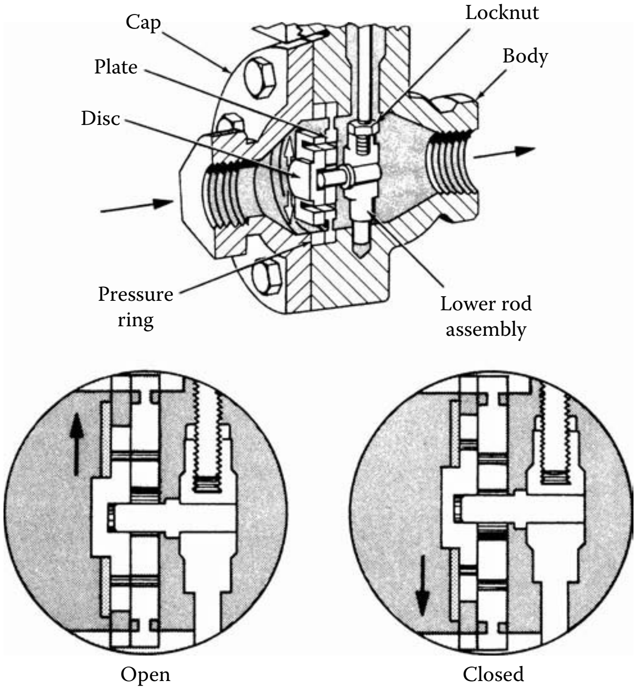


FIG. 6.23i
This throttling control valve provides wide rangeability by the use of a plate-and-disc inner valve. (Courtesy of Jordan Valve, a division of Richards Industries.)

TABLE 6.23j
*Multiple-Orifice, Plate- and Disc-Type Pump Governor Steam Valve Costs**

Valve Size (Inches)	*Carbon Steel*	*Stainless Steel*
0.5	$2,000	$2,600
1.0	$2,300	$3,300
1.5	$2,700	$3,600
2.0	$3,100	$4,000

* All valves are provided with 150# RF connections
(Courtesy of Jordan Valve.)

Bibliography

Arant, J. B., "Fire-Safe Valves," *InTech,* December 1981.

Ball, K. E., "Final Elements: Final Frontier," *InTech,* November 1986.

Bialkowski, B., Coughran, M., and Beall, J., "Control Valve Performance Update—The Last 10 Years," *Pulp Pap-Canada,* (102):21-22, November 2001.

Bishop, T., Chapeaux, M., Jaffer, L., et.al., "Ease Control Valve Selection," *Chem Eng Prog,* 98 (11): 52–56 November 2002.

Carey, J. A., "Control Valve Update," *Instruments and Control Systems,* January 1981.

Control Valve Handbook, 3rd edition, Marshalltown, IA: Fisher Controls International Inc., 2001.

Control Valve Seat Leakage, FCI 70-2-1998, Cleveland, OH: Fluid Controls Institute, Inc., 1998.

Borden, G. and Friedmann, P. G. (eds.), *Control Valves—Practical Guides for Measurement and Control,* Research Triangle Park, NC: ISA, 1998.

"Control Valves—Globe, Plug, Pinch, Needle, Gate," *Measurements and Control*, February 1994.

Fernbaugh, A., "Control Valves: A Decade of Change," *Instruments and Control Systems*, January 1980.

Monsen, J. F., "Valve Wars—Rising Stem vs. Rotary," *Plant Services,* January 1999.

Rahmeyer, W., "The Critical Flow Limit and Pressure Recovery Factor for Flow Control," *InTech,* November 1986.

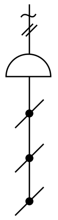
Whatever Happened To… A Guide to Industrial Valve and Actuator Companiess—Past and Present, Washington, D.C.: Valve Manufacturers Association, 1997.

Regulators and Final Control Elements

7

7.1 Dampers and Louvers

A. BRODGESELL (1970) **B. G. LIPTÁK** (1985, 1995, 2005)



Flow sheet symbol

Types of Designs:	A. Multiple blade or louver B. Rotating disc, including multiple disc C. Radial vanes D. Iris-type variable orifice (Note: Some butterfly valves and guillotine-type slide gate valves are similar in their functions to dampers. They are discussed in Chapter 6.)
Design Pressure:	Type A designs can usually handle up to 10 in. H_2O (2.5 kPa) shut-off differentials; type D units can handle up to 15 psid (103 kPa)
Materials of Construction:	Steel, galvanized steel, aluminum, and fiberglass; stainless steel is used in special cases
Sizes:	Type A dampers are available up to 6 ft × 8 ft (1.8 m × 2.4 m) for HVAC applications and in even larger sizes for boilers and other industrial applications
Flow Characteristics:	See Figure 7.1b
Leakage through Each ft^2 (0.092 m^2) of Damper Area:	At 3 in. H_2O (0.75 kPa) shut-off pressure differential: the leakage of standard type A units is 50 SCFM (250 l/s/m^2); special low-leakage type A, 5 SCFM (25 l/s/m^2); positive-steel type B, 0.5 SCFM (2.5 l/s/m^2)
Costs:	Type A costs range from \$100 to \$250 per square foot of area (\$1000 to \$2500 per square meter) as a function of size, features, and accessories
Partial List of Suppliers:	Air Clean Damper Co. (www.aircleandamper.com) Arrow United Industries (www.arrowunited.com) Babcock & Wilcox Co. (www.babcock.com) Bachmann Industries (www.bachmannusa.com) Belimo Air Controls (www.belimo.org) Damper Design Inc.(www.damperdesign.com) Flextech Industries (www.flextech-ind.com) FMC Corp. (www.fmc.com) Honeywell Sensing and Controls (www.honeywell.com/sensing) Johnson Controls (www.jci.com) Louvers & Dampers Inc. (www.louvers-dampers.com) Miracle Vent Inc. (www.miraclevent.com) Polymil Products Inc. (www.polymil.com) Ruskin Air & Sound Control (www.ruskin.com) Safe-Air/Dowco (www.safeair-dowco.com) Vent Products Co. (www.ventprod.com) Young Regulator Co. (www.youngregulator.com)

INTRODUCTION

Dampers and louvers are used to control the flow of gases and vapors. These streams usually flow in large ducts at relatively low static pressures. There are both "commercial" and "process control" quality dampers on the market. Commercial-quality units are used for the less-demanding applications, such as heating, ventilation, and air conditioning (HVAC), while the process control-quality units can handle higher pressures, higher temperatures, and corrosive vapors. The process control-quality units also can provide superior leakage and control characteristics.

Dampers are also used to control the flow of solids or to throttle the capacity of fans and compressors. There is no

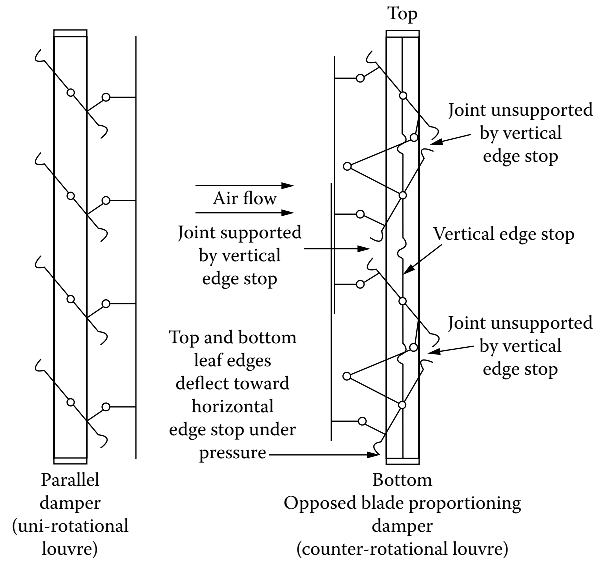


FIG. 7.1a

The designs of the parallel-blade and opposed-blade dampers, which are also referred to as uni-rotational and counter-rotational louver designs.

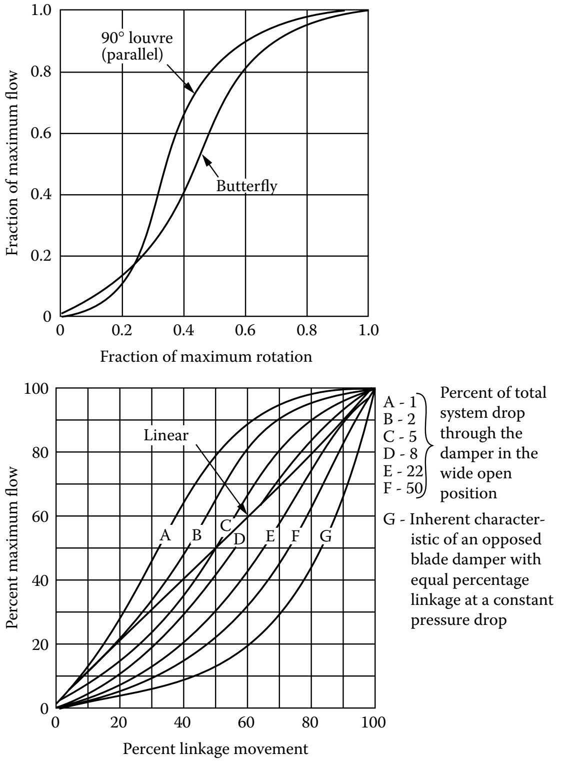


FIG. 7.1b

The flow characteristics of a parallel-blade damper are similar to those of a conventional butterfly valve (see top part of this figure). The flow characteristics of an opposed-blade damper approach equal-percentage characteristics when the total system pressure drop is through it and it shifts its characteristics toward quick opening, as the damper receives less and less of the total pressure differential.

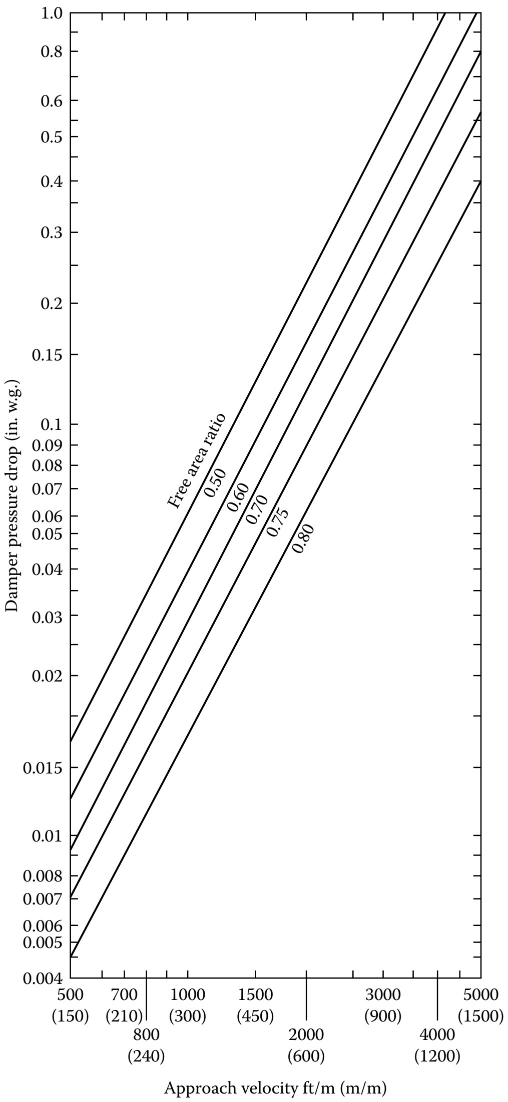


FIG. 7.1c

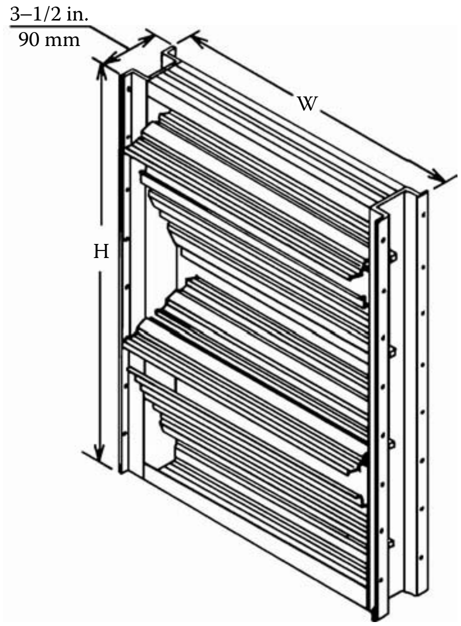
Pressure drop through wide-open dampers. The free-area ratio of an open damper is the total open area between the blades, divided by the nominal area.

clear distinction between butterfly valves and butterfly dampers or between slide-gate valves and guillotine dampers. The design features of these dampers are quite similar to their control valve counterparts, which are discussed in Chapter 6.

Dampers in general are large in size, and their operating and shut-off pressures are limited to lower values. The diameters of the largest dampers can exceed 20 ft (6 m).

DAMPER DESIGNS

Dampers and louvers can be grouped according to their shapes into parallel-blade, disc and multiple-disc, radial vane, and variable-orifice designs. Within each design category, there can be subdivisions according to leakage rates, materials of construction, actuator designs, or accessories provided.



Width (W) $\frac{\text{in.}}{\text{mm}}$

Height (H) $\frac{\text{in.}}{\text{mm}}$

Nlanimo		12 304	15 381	18 456	21 533	24 609	27 686	30 760	33 838	36 914	42 1067	48 1219	60 1524
	Actual	11–15/16 303	14–59/64 379	17–29/32 453	20–57/64 530	23–7/8 606	26–55/64 682	29–27/32 756	32–53/64 832	35–13/16 909	41–25/32 1061	47–3/4 1213	59–11/16 1516
6 152	5–31/32 151	.5 .046	.62 .058	.75 .070	.87 .081	1 .093	1.12 .104	1.25 .116	1.37 .127	1.5 .139	1.75 .163	2 .186	—
12 304	11–15/16 303	1 .093	1.25 .116	1.5 .139	1.75 .163	2 .186	2.25 .209	2.5 .232	2.75 .255	3 .279	3.5 .325	4 .372	5 .465
18 456	17–29/32 453	1.5 .139	1.87 .174	2.25 .209	2.63 .244	3 .279	3.37 .313	3.75 .348	4.12 .383	4.5 .418	5.25 .488	6 .557	7.5 .697
24 609	23–7/8 606	2 .186	2.5 .232	3 .249	3.5 .325	4 .372	4.5 .418	5 .465	5.5 .511	6 .557	7 .650	8 .743	10 .929
30 760	29–27/32 756	2.5 .232	3.12 .290	3.75 .348	4.37 .406	5 .465	5.62 .522	6.25 .581	6.87 .638	7.5 .697	8.75 .813	10 .929	12.5 1.16
36 914	35–13/16 907	3 .279	3.75 .348	4.5 .418	5.25 .488	6 .557	6.75 .627	7.5 .697	8.25 .766	9 .836	10.5 .975	12 1.11	15 1.39
42 1066	41–25/32 1061	3.5 .325	4.37 .406	5.25 .486	6.12 .569	7 .650	7.67 .731	8.75 .813	9.62 .894	10.5 .975	12.25 1.14	14 1.30	17.5 1.63
48 1219	47–3/4 1213	4 .372	5 .485	6 .557	7 .650	8 .743	9 .836	10 .929	11 1.02	12 1.11	14 1.30	16 1.49	20 1.86
60 1524	59–11/16 1516	—	—	—	—	10 .929	11.25 1.05	12.5 1.15	13.75 1.28	15 1.39	17.5 1.63	20 1.86	25 2.32
72 1829	71–5/8 1820	—	—	—	—	12 1.11	13.5 1.25	15 1.39	16.5 1.53	18 1.67	21 1.95	24 2.23	30 2.79
84 2133	83–9/16 2122	—	—	—	—	14 1.30	15.75 1.46	17.5 1.63	19.25 1.79	21 1.95	24.5 2.28	28 2.80	35 3.25
96 2438	95–1/2 2425	—	—	—	—	16 1.49	18 1.67	20 1.86	22 2.04	24 2.23	28 2.80	32 2.97	—

FIG. 7.1d

Standard commercial damper frame sizes and areas, ft^2/m^2. (Courtesy of Ruskin, previously Johnson Controls.)

Parallel-Blade Dampers and Louvers

Multiblade dampers consist of two or more rectangular vanes mounted on shafts that are one above the other, which are interconnected so they rotate together (Figure 7.1a). The vanes are operated by an external lever, which can be positioned manually, pneumatically, or electrically. In the uni-rotational louver (parallel damper) design, the vanes remain parallel at all rotational positions. In a counter-rotational louver (opposed blade), alternate vanes rotate in opposite directions. Both designs are illustrated in Figure 7.1a.

Flow guides are sometimes installed between adjacent vanes in order to improve the effectiveness of throttling. In the top of Figure 7.1b, the blade-angle vs. flow characteristics of a parallel-blade damper and a butterfly valve are shown. The sensitivity of this design is very high at mid-flow while the last 30° of rotation is relatively ineffective. The flow characteristics of butterfly valves are similar though somewhat superior to those of louvers.

The lower portion of Figure 7.1b shows an opposed-blade damper with equal percentage linkage. As less and less of the total system pressure drop is assigned to the damper, the characteristics of this damper shift toward quick opening.

Figure 7.1c gives the pressure drop across wide-open dampers. Ideally the wide-open pressure drop should be between 4 and 8% of the pressure difference across the closed damper. If the damper is sized so that closed pressure difference is between 12 and 25 times the pressure drop, when the damper is open, its apparent characteristic will be nearly linear (see curves C and D in Figure 7.1b).

Figure 7.1d provides some dimensional data for standard commercial damper frames, including their areas.

Low-Leakage Designs The parallel-blade damper cannot provide tight shut-off because of the long length of unsealed seating surfaces. The leakage characteristics of unsealed standard dampers are given in the lower portion of Figure 7.1e. In low-leakage damper designs, blade seals are installed along the seating surfaces of the blades, resulting in the reduced leakage characteristic shown in the upper portion of Figure 7.1e.

There are a number of variations in the blade-edge seal designs. Some of these designs are illustrated in Figure 7.1f.

Corrosion Resistant Designs For corrosive services, both the parallel- and the opposed-blade designs are available in sizes from 12 by 24 in. up to 60 by 120 in. These units are made of fiberglass-reinforced polymer with 304 stainless steel hardware. Some design variations are shown in Figure 7.1g.

Actuators and Accessories Damper actuators can be manual, electric, hydraulic, or pneumatic. Standard pneumatic actuators vary their effective diaphragm areas from 2 to 24 in.^{2} (13 to 155 cm^2), while their stroke lengths range from 2 to 6 in. (51 to 152 mm). The amount of force they produce ranges from about 10 to 300 lbf (4.5 to 136 kgf). The standard spring ranges for dampers include the spans of 3–7, 5–10, and 8–13 PSIG (0.2–0.48, 0.34–0.68, and 0.54–0.88 bars). Electronic actuators can be operated by 4 to 20 mA DC analog or by digital signals.

For more accurate throttling, the actuators can also be provided with positioners. If remote indication of damper status is desired, limit switches can be installed to detect the blade angle. These can be pneumatic sensors of nozzle backpressure or mechanically actuated position sensors. The damper position switch can be furnished with an adjustable mounting flange, which allows the unit to be mounted through a duct wall with the trip lever positioned so that it is actuated by the damper blade itself.

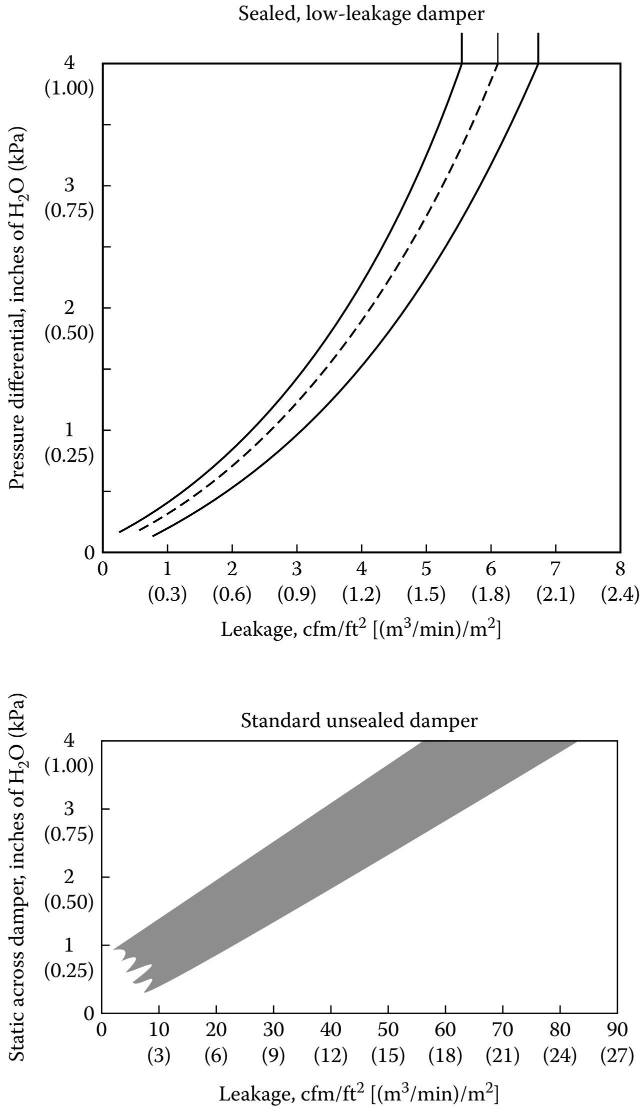


FIG. 7.1e
The leakage rates through sealed, low-leakage dampers are shown in the top, while the leakage rates of unsealed dampers are given in the bottom portion of the figure. (Courtesy of Honeywell and Ruskin, formerly Johnson Controls.)

Rotating Disc Dampers

The rotating disc damper designs are very similar to the butterfly valve designs, which were discussed in detail in Section 6.17. These dampers are usually installed in circular ducts and can be operated both manually or automatically. A corrosion-resistant version of this design is made of fiberglass-reinforced plastics materials and is illustrated in Figure 7.1h.

Multiple-Disc Dampers

A unique variation of the butterfly design is the multiple rotating disc damper. In this design several disc elements are distributed over an area (Figure 7.1i). One advantage of this configuration is improved flow control characteristics,

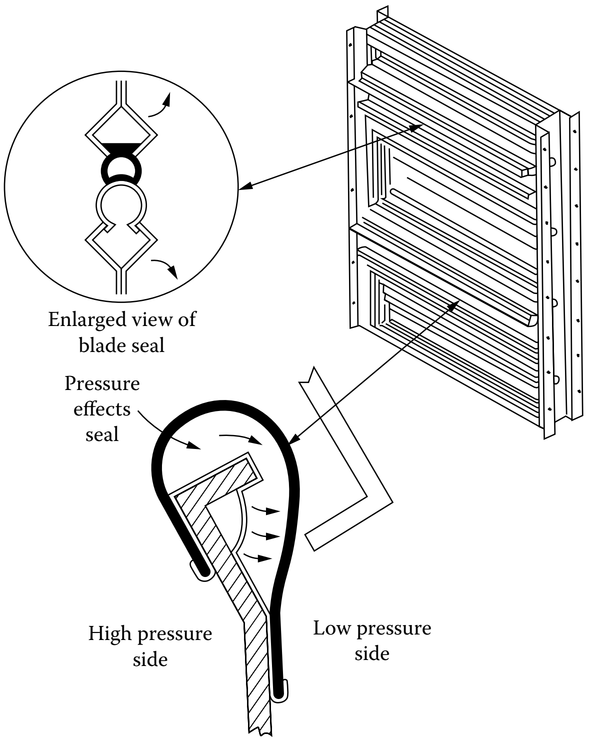


FIG. 7.1f
Low-leakage damper designs tend to increase the efficiency of HVAC systems. In this figure, two blade-edge seal designs are illustrated. (Courtesy of Honeywell and Ruskin, formerly Johnson Controls.)

because each disc can have its own unique spring range and failure position.

Another major advantage is the substantial reduction in leakage compared to the parallel-blade design. At a static pressure of 3 in. H_2O (0.75 kPa) the leakage can be estimated as 0.01% of full damper capacity, which corresponds to about 0.5 SCFM (2.5 $l/s/m^2$) leakage per ft^2 (0.092 m^2) of damper area.

Fan Suction Dampers

On blowers and fans, when throughputs must be controlled, radial vane dampers can be utilized. The damper consists of a number of radial vanes arranged to rotate about their radial axis (Figure 7.1j).

The radial vane dampers do not provide high-quality control, and their closed position leakage rates are also high. Their control applications include furnace draft control.

Usually a positioner is furnished with these units, which provides a linear relationship between the control signal and the blade pitch angle. In certain packages the positioner is factory set in the reverse acting mode, meaning that an increasing control air signal will reduce the air flow by decreasing the blade pitch. In such packages, one has to install a reversing relay between the positioner and the actuator, if direct action is desired.

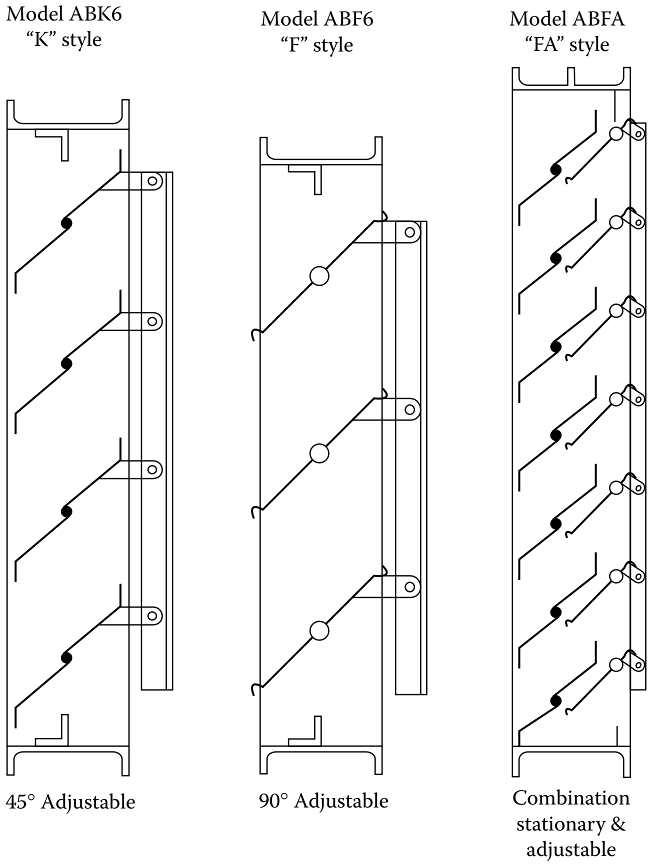


FIG. 7.1g
Corrosion-resistant fiberglass-reinforced polymer damper designs for both adjustable and stationary applications. (Courtesy of Polymil Products Inc.)

Variable-Orifice Dampers

Variable-orifice dampers use the same principle as the iris diaphragm of a camera. In order to achieve control action, the closure element moves within an annular ring in the damper body and produces a circular flow orifice of variable

FIG. 7.1h
Corrosion-resistant single-disc damper (Courtesy of Polymil Products Inc.)

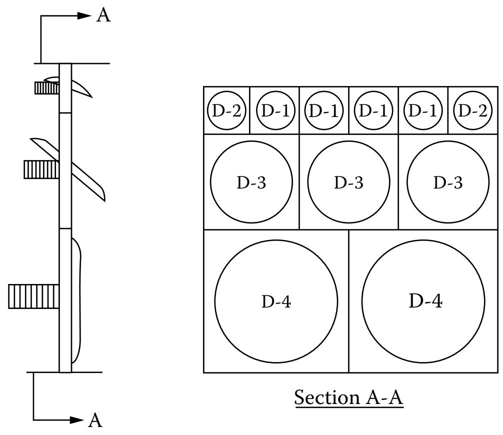


FIG. 7.1i
Multiple-disc dampers provide better sealing and control characteristics.

diameter (Figure 7.1k). The flow characteristics are similar to those of a linear valve.

However, tight shut-off is not possible, and leakage rates are comparable to or greater than those of a butterfly valve of equal size. Maximum pressure differential is limited to approximately 15 psid (104 kPa). Dual valve units can be provided with a common discharge port for applications involving the blending of two streams.

For solids service, the variable-orifice valve can be used for throttling, if the valve is installed in a vertical line. In horizontal lines, the shutter mechanism of the valve forms a dam, which makes the valve unsuitable for solids service. Standard sizes range from 4 to 12 in. (100 to 300 mm).

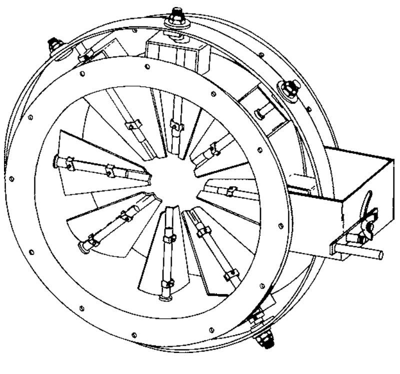


FIG. 7.1j
Radial vane dampers are used to throttle the flow on the suction side of air fans and blowers.

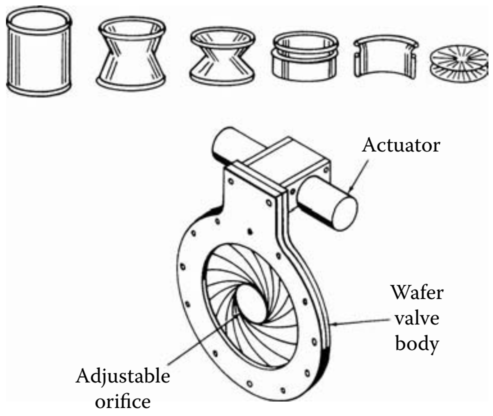


FIG. 7.1k
Variable-orifice or iris damper. The sleeve can be made of nylon or of other materials and is provided with a built-in retaining ring at each end. When the upper end is fixed and the lower end is rotated, this gradually reduces the orifice opening. At 180° of rotation the orifice is completely closed. If the sleeve is first turned back on itself, partly "inside out," the effect of rotation is exactly the same, but operates in a much more compact form, as a pleated duplex diaphragm.

CONCLUSIONS

Dampers are suitable for control of large flows at low pressures where high control accuracy is not a requirement. Typical applications of these units include air conditioning systems and furnace draft control. Variable-orifice or iris dampers are smaller than other dampers, offer better control quality, and can also be used to control vertical solid flows.

Bibliography

Ball, K. E., "Final Elements, Final Frontier," *InTech,* November 1986.

Brown, E. J., "Air Diffusing Equipment," in *ASHRAE Handbook, Equipment Volume,* Chapter 2, 1979.

"Catalog of U.S. & Canadian Valves & Actuators," Washington, D.C.: Valve Manufacturers Association of America, 2003.

Dickey, P. S., "A Study of Damper Characteristics," Reprint No. A-8, Bailey Meter Co.

Daryanani, S., et al., "Variable Air Volume Systems," *Air Conditioning, Heating and Ventilating,* March 1966.

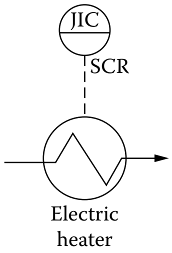
"Evaluation Finding for Oiles Multiple Viscous Shear Dampers," Cerf. Report: Hitec 99-04, American Society of Civil Engineers, 1999.

Lipták, B. G., "Reducing the Operating Costs of Buildings by Use of Computers," *ASHRAE Transactions,* Vol. 83, Part 1, 1977.

Lipták, B. G., "Optimization of Semiconductor Manufacturing Plants," *Instruments and Controls Systems,* October 1982.

7.2 Electric Energy Modulation

A. BRODGESELL (1970) **L. D. DINAPOLI** (1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

Types:	A. On/off switches, relays B. Saturable core reactors or variable voltage controllers C. Silicone-controlled rectifiers (SCRs) C1. Phase-angle-fired C2. Zero-angle-fired
Power Range:	From a few kilowatts up to a few megawatts
Cooling:	Up to 20 kW by convection cooling; up to 500 kW by fan cooling; and above 500 kW by water cooling
Inaccuracy or Linearity:	Better than 1% of full scale
Costs:	A. For the costs of on/off switches see the *Process Measurement* volume of this handbook; for the costs of relays, refer to Chapter 5 of this volume B. A saturable core reactor for the manual regulation of electric power to small motors, heaters, or lamps, up to 1200 W or 1 hp, costs about $200. Proportional control for a 5–10 kW load costs about $800 C. SCR power controller cost range for a 300 kW load is from $3000 to $5000
Partial List of Suppliers:	Allen-Bradley Co. (A) (www.ab.com) Barber-Colman Co. (C) (www.barber-colman. com) BEC Controls Corp. (C) (www.beccontrols.com) Cole-Parmer Instrument Co. (B) (www.coleparmer.com) Eagle Signal Controls (A) (www.danahercontrols.biz/ eagsignal.htm) Enercorp Instruments Ltd. (A) (www.enercorp.com) Eurotherm Corp. (C) (www.eurotherm.com) General Electric Co. (C) (www.ge.com) Honeywell Sensing and Controls (C) (www.honeywell.com/sensing) Love Controls Corp. (C) (www.lovecontrols.com) Lumenite Control Technology Inc. (A) (www.lumenite. com) Mag-Con Engineering Inc. (B) (www.magconeng.com) Pearson Electronics Inc. (B) (www.pearsonelectronics.com) Reliance Electric Controls (A, C) (www.reliance.com) Robicon (C) (www.robicon.com) Superior Electric Co. (B) (www.superiorelectric.com) Testmart (B) (www.testmart.com) Thermo Electric (C) (www.thermo-electric-direct. com) Transmagnetics (B) (www.kron.de) Traid Technologies Inc. (A) (www.triadtec. com) United Electric Controls (A) (www.ueonline.com) Westinghouse Electric Corp. (A) (www.westinghouse.com)

INTRODUCTION

Devices for modulating the flow of electrical heating power can be grouped into on/off and proportional control devices. Examples of on/off final control elements include solid-state relays (SSRs) or solid-state contactors (SSCs) and electromechanical power relays or contactors. Proportional power controllers are mostly silicon-controlled rectifiers (SCRs), with some saturable core reactors, power amplifiers, and a few ignitrons also still being used.

An SCR or thyristor is a semiconductor device that switches AC power on and off. On/off control always produces a cyclic

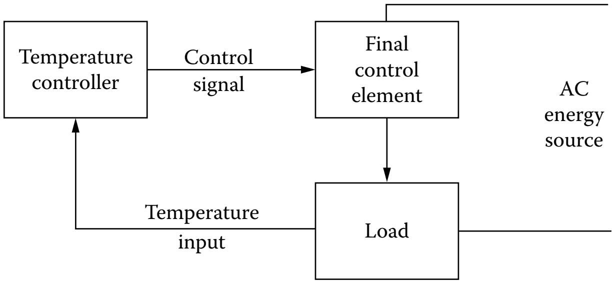


FIG. 7.2a
Temperature control loop, using electric energy modulation as the means of control.

response, while properly tuned proportional throttling control improves control quality by eliminating cycling. SCRs are almost silent in operation and can eliminate the switch-on surge.

Electric power throttling devices are often used as the final control elements in temperature control loops, as shown in Figure 7.2a. The load, to which the electric power is throttled by this final control element, can be an electric or ultraviolet heater, but can also be a variable-speed motor or any other electric load. A comparison of the characteristics of the various electric energy modulation devices is summarized in Table 7.2b.

ON/OFF POWER CONTROL

Electric circuits can be opened and closed as a function of the value of temperature or other controlled variables. Such on/off control devices have been discussed in several sections of the three volumes of this handbook. In the *Process Measurement* volume, Section 2.7 covered flow switches, Section 4.11 discussed temperature switches, and Section 5.13 dealt with pressure switches. Relays and contactors are covered in Chapter 5 of this volume.

The common feature of these on/off devices is that they apply electric power intermittently and therefore cause cycling in the controlled variable, which usually is temperature. This mode of control is acceptable only in such secondary processes as HVAC, where steady and close control is not required and cycling can be tolerated.

TABLE 7.2b
Selection Guide of Electric Energy Modulation Devices for Various Services

Service/Control Element	*SCR*	*Ignitron*	*Saturable Core Reactor*	*Contactor or Relay*
DC output required	✓	✓		
Operating voltage above 600 V		✓	✓	✓
Small size	✓			✓
Severe overload or transients		✓	✓	
Overall efficiency	✓			
Low cost per watt	✓			✓

THROTTLING POWER CONTROLLERS

A power controller contains solid-state switching devices, heat sinks, protection circuits, and the electronics to control the solid-state switching devices. The power controller accepts standard industrial control signals, which can either be analog, such as 0–5 mA, 1–5 mA, 4–20 mA, 0–5 V, 0–10 V, or digital and which controls the power or voltage delivered to an electrical load in proportion to the control signal.[1]

For temperature control applications both the saturable core reactors and the silicon-controlled rectifiers (SCRs) are used. Both will be described below.

Saturable Core Reactors

A saturable core reactor is a device consisting of an iron core, an AC or gate winding, and a control winding (Figure 7.2c). The impedance of the gate winding can be changed by means of the magnetic flux produced by the DC current in the control winding. A DC control signal of a few watts can therefore control hundreds of kilowatts of AC power. Because the control signal requires watts of power, a magnetic amplifier is always used to boost the control signal from the process controller up to the power level required by the control winding of the saturable core reactor.

Saturable core reactors are relatively inefficient (90%), large in size, and require water or air cooling, depending upon their power rating. Their main advantage is their immunity to modest transients and overloads. Well-designed SCRs can offer similar reliability at greater efficiency and reduced size.

The main advantage of saturable core reactors is that it is not possible to burn out the unit. Even if the controlled electric heater is short-circuited, it will not allow enough current to flow to burn out the reactor. The disadvantages of the saturable core reactors, in addition to their size and weight, is that they cannot adjust the line voltage from 0 to 100%.

Because the reactor relies on the impedance of its core to attain the shut-off power, a shut-off condition will still

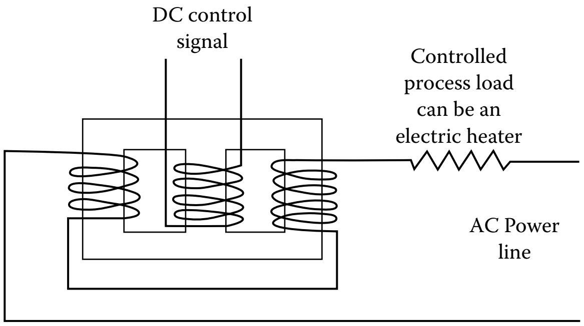


FIG. 7.2c
The basic components of a saturable core reactor controller.

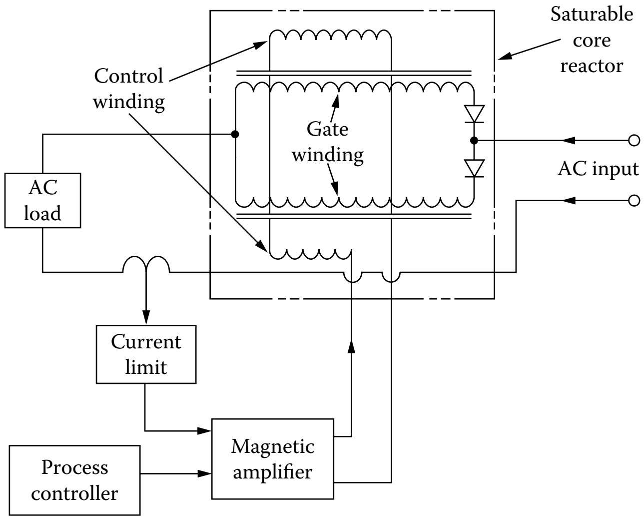


FIG. 7.2d
Current limiting applied to a saturable core reactor control loop.

yield about 3% of full power. Similarly, because the reactor depends on the saturation of its core to provide full power, this full power state usually yields only about 90% of full power. Therefore, when the control signal is 0, the load voltage is about 5%, and when the control signal is 100%, the load voltage is only about 90% of the line voltage.

As a consequence of this limitation, the use of saturable reactors requires a very good understanding of the nature of the load, because the saturable core reactor must be so selected that the load wattage and voltage will always be within 3 and 90% of its power control range. In order to reduce the severity of this limitation, reactors can be tapped for several ratings.

Saturable core reactors can be provided with current limiting by means of a current transformer and an appropriate input to the magnetic amplifier. A diagram of a typical saturable core reactor design is shown in Figure 7.2d.

SCR Power Controllers

The power switching element in most power controllers is a silicon-controlled rectifier—thus the general term "SCR power controller." The SCR is a four-layer solid-state silicon structure consisting of alternating layers of negatively and positively doped material. It is a three-terminal device consisting of an anode, a cathode, and a gate, encapsulated and bonded to a thermally conductive base (Figure 7.2e).

In order for the SCR to conduct, the anode must be positive with respect to the cathode and a trigger signal at the gate must initiate conduction. Once the gate has been pulsed by a DC voltage, conduction is self-sustained until the current flow through the SCR tries to reverse. When current attempts to reverse, the SCR "commutates" and returns to the blocking state and will remain "off" until triggered again, provided the anode is positive with respect to the cathode. In this manner an SCR can efficiently control loads from a few hundred watts to megawatts of electrical energy.

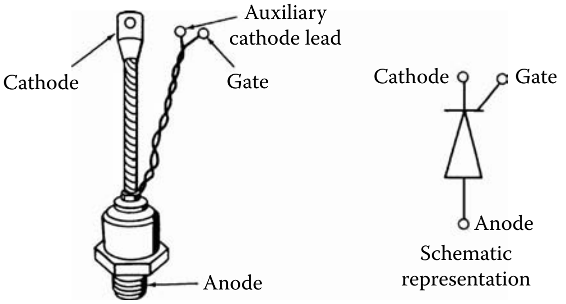


FIG. 7.2e
The components of a silicon-controlled rectifier.

For DC applications, two SCRs are combined with two diodes to provide a full wave rectified, modulated output. For AC loads, two SCRs are connected in an antiparallel fashion to provide modulated AC output.

Below a kilowatt, some AC power controllers use a triac as the power switching element. A triac is a four-layer device with three terminals that behave as a pair of SCRs connected in an antiparallel fashion. Power controller configurations for DC loads, as well as SCR and triac AC loads, are shown in Figure 7.2f.[2]

SCR-type power controllers operate either by turning on each AC cycle for part of the full cycle (phase-angel firing) or by sending only a percentage of the full cycles to the load (zero-voltage firing). Zero-voltage firing is like turning a switch on and off while the power is off. Phase-angle firing can be compared to opening and closing a switch while the power is on.

Relative to the saturable core reactors, the SCRs are superior in the following ways: They are smaller in size and weight; they have higher efficiencies because complete shut-off of electric heaters can be attained; and they do not require load matching, but are able to control any load that is below their rating. The disadvantages of SCRs are due to their being semiconductor devices, which necessitates that they be protected against transient surges, high inrush currents, load short circuits, over-temperatures, and other circuit component failures.

Phase-Angle-Fired SCR Controllers An SCR, once gated properly, will continue to conduct until the current through it goes to zero. Phase-angle-fired units modulate the power delivered by triggering the SCR between 180° and 0° in the appropriate half cycle of the AC supply. As shown in Figure 7.2g, the point at which the SCR is triggered is defined as the firing angle. The SCR will deliver power to the load until the end of the half cycle. Thus, for half power the SCR is gated on at 90°; for 75% power the firing angle is gated at 60°; and the full-power firing angle is 0°.

The same action occurs for the negative half cycle, thus resulting in the waveforms shown. Thus, by varying the point in each half cycle where the SCRs are turned on, the average power delivered to the load can be varied from zero to full power, in proportion to the control signal, with infinite resolution.

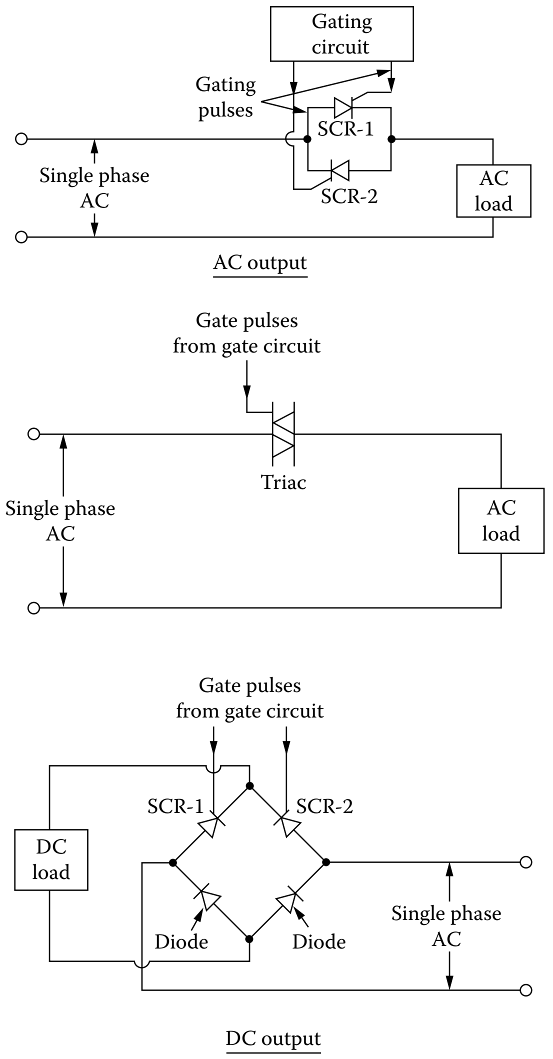


FIG. 7.2f
Power controller configurations: SCR and triac circuits for AC loads on the top and SCR control circuits for DC loads at the bottom.

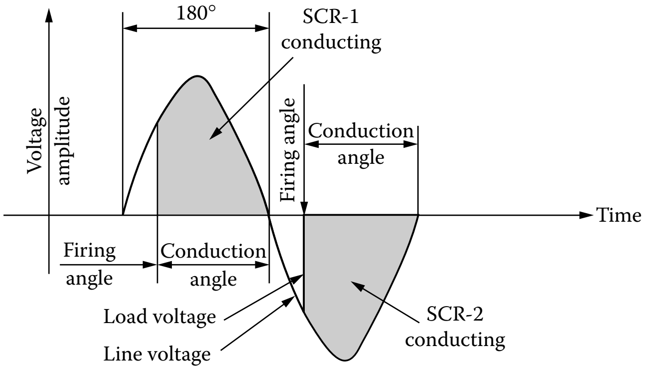


FIG. 7.2g
Definition of firing and conduction angles.

Radio Frequency Interference (RFI): Phase-angle firing can generate radio frequency interference and power line spikes, which can cause interaction between controllers if two or more of them operate on the same line. Because a phase-angle-fired SCR power controller switches from the off state to the on state very quickly (less than 1 ms) at substantial voltage levels, harmonics in the radio frequency area of the spectrum are created. This radio frequency interference can interfere with proper control loop performance. In severe cases or sensitive applications, filtering of the RFI by means of line chokes and bypass capacitors may be necessary.

Current Limit Feature: With the addition of a current transformer and some control circuitry, an adjustable current limit feature can be used with phase-angle-fired power controllers for loads that exhibit high inrush currents. These types of loads are heaters whose cold resistance is very low compared to the resistance at normal operating temperatures.

Current limiting works by always gating the SCRs on at a 180° firing angle, then moving the firing angle back toward that desired by the control signal. As the firing angle is backed away from 180°, the current is monitored and will be kept below the preset current limit before the firing angle is permitted to back away toward higher current output firing angles. This feature is also useful for transformer-coupled loads to ensure that the transformer is not driven into saturation and drawing excessive current.

Soft Start Feature: With phase-angle-fired power controller units, a soft start feature will start the firing angle at 180° and ramp up slowly (typically 0.5–1 sec) to the firing angle desired by the control signal. This feature can be incorporated with little or no additional control circuitry and is usually standard on most phase-angle-fired power controllers. It helps to prevent high inrush currents experienced with transformer-coupled loads, coupled loads, and nonlinear heating elements.

ZERO-VOLTAGE-FIRED SCR CONTROLLERS

In zero-voltage-fired power controllers the SCRs are gated on at the beginning of each half cycle. Thus, each SCR will conduct for a full cycle. Generally, power modulation is accomplished by varying the number of full half cycles on vs. the number of full cycles off over some time base (usually 0.5–1 sec).

If we assume a 1 sec time base, then at 50% power the load would receive 30 full cycles on and 30 full cycles off. At 20%, power would be on for 12 full cycles and off for 48 full cycles. Sometimes zero-voltage-fired units are called burst-fired units because in this type of unit the load receives a burst of energy followed by some off time.

Some units proportion energy without a fixed time base by varying both the on time and the off time. For this type of unit, 50% would have one full cycle on and one full cycle off, 20% would be one full cycle on and four full cycles off. Typical waveforms of both methods are shown in Figure 7.2h.

Because a zero-fired SCR controller gated the SCR on at the beginning of the cycle, current limiting cannot be done with

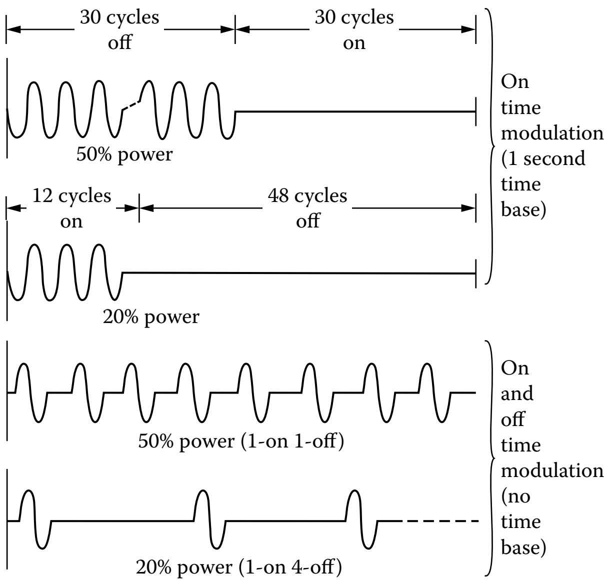


FIG. 7.2h
Output waveforms when zero-voltage firing is used.

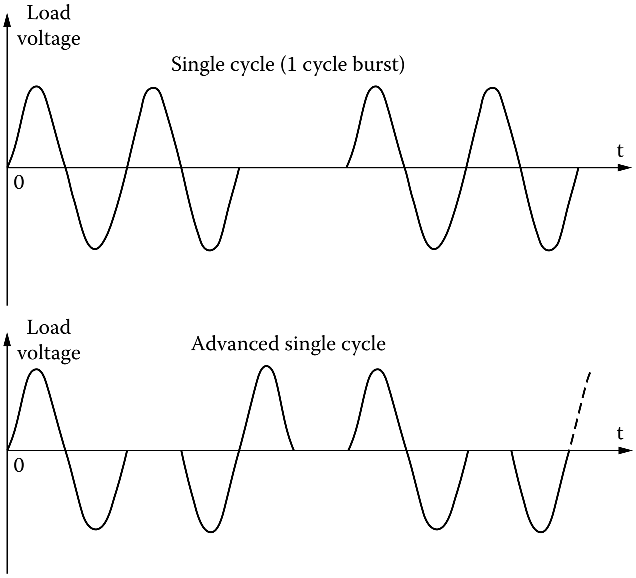


FIG. 7.2i
On top, the conventional single cycle firing method is illustrated, while on the bottom the advanced half cycle firing is shown. (Courtesy of Invensys Eurotherm Controls.)

this type of unit. For this reason zero-fired units are usually applied to resistive loads and not transformer-coupled loads.[3]

Because the SCR is gated at or near zero voltage, almost no RFI is generated by this type of SCR power controller. This is the chief advantage of this type of SCR unit over the phase-angle-fired type.

Controlling Infrared Tungsten Heaters

Infrared radiation occurs at wavelengths from 0.76–1000 μm. The *radiant* heat output of an IR heat source relates to the fourth power of its temperature, and the wavelength of IR radiation depends on the temperature of the heat source. The wavelength of the IR radiation also determines its ability to penetrate or to be absorbed. The shorter wavelengths emitted by tungsten lamps at 4000°F are more penetrating than those emitted by nickel chrome filaments operating at 1800°F.

Different processes benefit differently from the use of different wavelengths. For example, the heating of transparent materials such as water requires longer wavelengths. On the other hand, the drying of paint benefits from the greater penetration of shorter wavelength IR radiation.

Traditional IR Controls The traditional method of controlling IR heaters was by phase-angle-fired SCRs. They provided smooth, stable application of the power, and their soft start eliminated inrush problems.

This method also had the disadvantages of higher equipment cost and higher electricity consumption, because of the reduced power factor. In addition, phase-angle firing required larger transformers due to harmonics, and it also generated RF interference, which can harm the electronics equipment.

Advanced Half Cycle Firing and V^2 Power Control Figure 7.2i illustrates both the conventional single cycle and the advanced half cycle firing. In single cycle firing the on and off times are reduced to a single whole cycle of the AC supply. In contrast, the advanced method switches half cycles and, thereby, reduces the off time at 60 Hz to 8.33 ms. This not only reduces flicker, but provides a more stable heater temperature, as the tungsten filament has less time to cool.

One possible disadvantage of the half cycle firing method is the generation of DC current or voltage if the number of switched positive and negative half cycles is not equal. To eliminate this requires intelligent half cycle firing.

The relationship between tungsten filament temperature and radiated energy output is not linear, but the relationship between power output and the square of the voltage (V) is linear, as shown in Figure 7.2j. Therefore, for the best overall performance of IR heater controls, the intelligent half cycle firing should be combined with voltage-squared power control.

Common SCR Limitations

Regardless of the type of power controller chosen, some common limitations exist that require protective devices or circuitry. These are usually provided by the manufacturer and some of them are described below.

Transient Voltage Protection SCRs can be damaged by the line voltage transients that exceed the voltage capability of the device. Most power controllers offer metal oxide varistor (MOV)

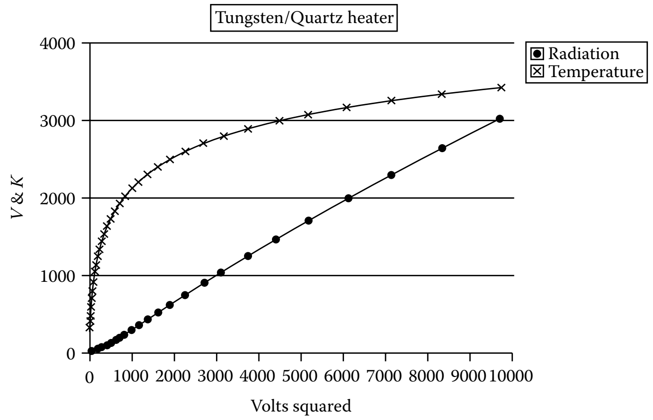


FIG. 7.2j
The radiated energy related linearly to the square of voltage. (Courtesy of Invensys Eurotherm Controls.)

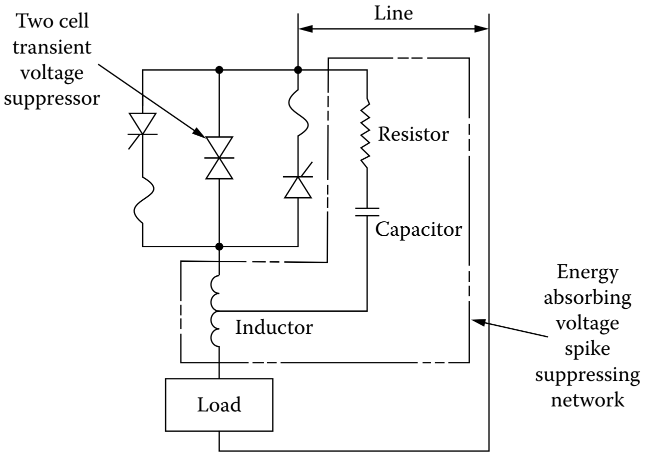


FIG. 7.2k
Energy-absorbing network using voltage spike suppressors.

protection circuitry across the SCR. The MOV will short-circuit the voltage spike around the SCR and prevent a false turn-on or permanent damage to the device.

Current Limit Fuses SCRs will fail due to load shorts, just as mechanical contactors will. SCRs have fault current ratings that are specified by the terms I^2T and I surge. These ratings allow proper selection of the special fuses that clear in less than a half cycle in order to protect SCRs. The I^2T and I surge value of the fuse should be less than that of the SCR it is to protect.

dv/dt Protection A large change in a short period of time can cause problems with SCR performance. High rates of voltage change (dv/dt) can trigger an SCR into an unscheduled turn-on by inducing internal gate trigger currents via interjunction capacitance. Transformer-coupled loads can present dv/dt problems due to a lagging current in the inductive load. In this case the SCR will not commutate (turn off) when the gate trigger is stopped.

Transient voltage spikes are also a form of dv/dt problems. To prevent dv/dt problems, gate-to-cathode capacitors are used to shut internally coupled gate currents from the gate, or R-C snubber circuits are used in parallel with the SCR (Figure 7.2k).

di/dt Protection A high anode current flow during the turn-on of the SCR can stress the junctions of the device. If a load requires near-rated current before the entire junction area of the SCR is turned on, then the current density in the partially conducting junction is too high. This can cause localized heating and eventually lead to failure of the device. Usually, gate drive circuitry ensures good junction turn-on with a large fast-rising trigger pulse.

If a load is purely resistive, then for phase-angle-fired units some small inductance should be put in a series with the load, to prevent di/dt problems around a 90° firing angle. Usually zero-fired power controllers do not have di/dt problems, because the voltage is near zero when the SCR is turned on.

Cooling and Heat Sinking The temperature of an SCR will increase during normal operation due to the power dissipated in it. This power loss stems from the fact that there is a voltage drop of approximately 1 V, which the device is conducting. The current that an SCR can safely carry is dependent on the heat sinking provided. Operating at excessive temperatures will cause failures.

Most power controllers up to 20 kW are convection cooled. Fan cooling is usually required for units between 30 and 500 kW. Above 500 kW water cooling is employed to maintain safe operating temperatures for the SCR. Suppliers can also supply thermostats mounted to the heat sinks to signal if the temperature of the SCR is getting too high.

IGNITRON TUBE

Ignitron tubes, in function, are similar to the SCRs, but quite different in construction. As shown in Figure 7.2l, the ignitron is a mercury arc rectifier consisting of a graphite anode, a cathode containing a pool of mercury, and an electrode called

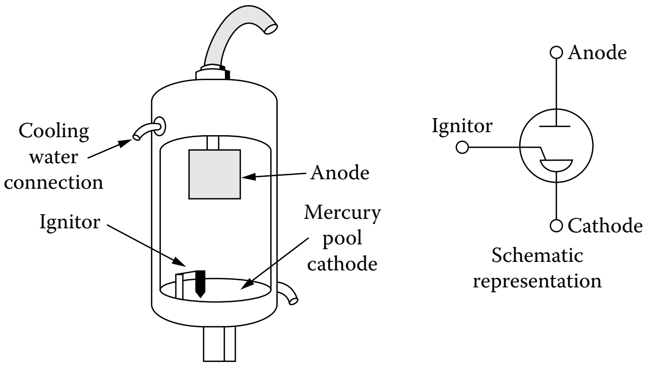


FIG. 7.2l
Ignitron tube.

the ignitor. Ignitron operation is achieved by applying a DC voltage pulse to the ignitor. The arc produces a large number of mercury ions that permit current to flow from anode to cathode, if the required potential exists. The ignitron can be subjected to 200% overloads for up to a minute without damage.

Because of the lower efficiency of the ignitron, the amount of cooling required here is much greater than for an SCR. Normally, forced air or water cooling must be employed with the ignitron. In general, the ignitron is inferior to SCR in terms of size, efficiency, life expectancy, and vibration resistance. It is for these reasons that ignitrons have largely been replaced by SCRs even on existing installations.

POWER AMPLIFIER

A power amplifier is a combination of active devices, transistors, operational amplifiers or tubes, and passive elements, resistors, capacitors, and inductors to yield power amplification with certain characteristics such as linearity, frequency response, and gain. Such amplifiers are not usually used as final control elements. Reproduction of an input waveform, while important in communications, is not necessary in process control, where only modulation of power is desired. For this reason the previously discussed devices offer more cost-effective and efficient solutions than does the power amplifier.

References

1. Hardy, M., "Selecting SCR Power Controllers," Nelson Publishing, 2002.
2. *Machine Design,* Electronics Reference Issue, Chapter 4, May 13, 1982.
3. *Thyristors Manual,* General Electric Co., 1982.

Bibliography

"Analog Controllers, Meter Relays," *Measurements and Control,* February 1993.

Anderson, P. M., Fouad, A. A., *Power System Control and Stability,* 2nd edition, Hoboken, NJ: Wiley-IEEE Press, 2002.

Graham, D. R., *SCR Manual,* 6th edition, General Electric, Englewood Cliffs, NJ: Prentice Hall, 1982.

Hall, J., "Motor Drives Keep Pace with Changing Technology," *Instruments and Control Systems,* September 1982.

Janki, C., "What's New in Motors and Motor Controls?" *Instruments and Control Systems,* November 1979.

Kaiser, J., *Electric Power: Motors Controls, Generators, Transformers,* Tinley Park, IL: Goodheart-Wilcox, 1997.

Kinligh, S. S., "ABCs of SCR Controllers," *Instrumentation Technology,* February 1976.

Miller, T. J. E., ed., *Reactive Power Control in Electric Systems,* Hoboken, NJ: Wiley Interscience, 1982.

McNair, W. L., *SCR and New Technology,* Tulsa, OK: Pennwell Publishers, 1991.

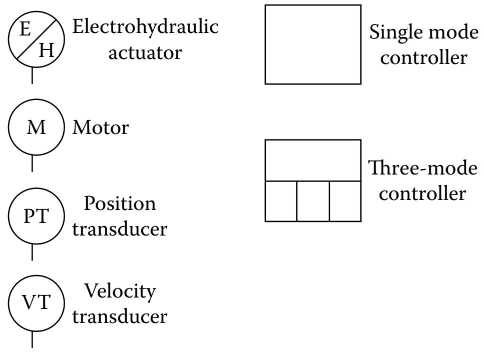
Pendarvis, M. T., "SCR Control of Saturable Resetors," *Instruments and Control Systems,* August 1970.

Saccomano, F., *Electric Power Systems: Analysis and Control,* Hoboken, NJ: Wiley-IEEE Press, 2003.

Valentine, R., *Motor Control Electronics Handbook,* New York: McGraw-Hill Professional, 1998.

7.3 Linear and Angular Positioning of Machinery

D. S. NYCE (2005)



Flow sheet symbols

Range:	*Linear position:* Linear motion is typically driven through the use of a leadscrew, ballscrew, or worm drive, which are available with ranges from less than an inch (< 25 mm) to over 20 ft (> 6 m). Linear sensors for position feedback in the lower range include LVDTs, magnetic, and optical encoders. Longer stroke linear feedback devices include encoders and magnetostrictive position transducers. Optical encoders are practical up to 6 ft (2 m). Magnetostrictive position transducers can be used up to 60 ft (20 m). *Angular position:* Gearboxes are normally used to transmit rotary motion, but direct drive is also used. The motion, often driven by a stepper or DC motor, can be 0–360°, or can be multiturn or continuous rotation. Typical angular feedback sensors have a range of 270–360° rotation, and when using a 360° sensor, the turns can be counted for multiturn applications.
Inaccuracy:	A complete control loop for one axis of motion comprises several components, typically including a motor, motor driver, controller, and one or more feedback elements. The position inaccuracy is the sum of errors from all the above-listed components. There are both static and dynamic sources of error. Generally speaking, inaccuracy can range from about 1.0%, in systems where speed or cost are the driving factors, down to less than 0.01% where precision is the main consideration.
Costs:	A combination of motor and controller can range from less than $200 for a stepper motor and open-loop printed circuit controller card, to well over $1000 for a linear motion stage with motor, driver, controller, and feedback element.
Partial List of Suppliers:	Advanced Motion Controls (a-m-c.com) AMCI (amci.com/motion.asp) Bosch Rexroth (boschrexroth.com) Cleveland Motion Controls (cmccontrols.com) Control Technology Corp. (ctc-control.com) Galil (galilmc.com) Heidenhain (heidenhain.com) Maxon Precision Motors (mpm.maxonmotor.com) Micromo Electronics (micromo.com) MTS Systems Corp. (Temposonics.com) National Instruments (ni.com) Newport (newport.com) Oregon Microsystems (omsmotion.com) Ormec Systems (ormec.com) Portescap Danaher (Danaher.com) Revolution Sensor Company (rev.bz) Simple Step (simplestep.com) Solutions Cubed (solutions-cubed.com) The Motion Store (themotionstore.com)

OPEN- AND CLOSED-LOOP POSITIONING

Movable components of industrial machinery can require positioning in relation to the three linear axes (*x*, *y*, and *z*), and in rotations about these same axes. Normally, for a particular movable component, the motion is desired in relation to only one or two axes, and the machine is designed to minimize unwanted motion in relation to the remaining axes.

Positioning systems can use either an open-loop or a closed-loop control system. In an open-loop system, such as can be built utilizing a stepper motor or piezoelectric actuator, the drive system can output a precise predetermined command for movement. In such a system, moving to a new position then is accomplished by consecutively executing a given number of predetermined steps. In open-loop positioning, there is no sensor that is continuously detecting the resulting position, and therefore no control is exercised during the move. The motion continues until the control input is discontinued, or until a limit switch is activated.

In closed-loop motion control, the position of the moving part is continuously detected and reported. Based on this feedback, the position and velocity of movement can both be controlled. The reported position can be continuously compared to the desired position, and the component can be moved in a direction that will reduce the error. This is called servo control. Figure 7.3a shows a block diagram of a basic closed-loop motion control loop, which continues to adjust the component's positions, from the beginning of the motion until it is completed as commanded.

The input (or desired) position is compared to the actual position, which is reported by the feedback device. The difference between these positions is the error, *E*. The error signal can be positive or negative, where the sign indicates the direction. The error signal is amplified (the amplification factor is called gain) and used to drive the motor.

The complete control system for positioning on an actual control axis includes several levels of control loops, which are acting in a synchronized manner, as shown in Figure 7.3b. The intermediate control loops are called embedded loops, and together they guarantee that the original command for a specified movement is correctly completed.

In this example, the position input (set point) is compared to the detected actual position by the feedback sensor. The detected difference is used to generate a velocity command. A differentiator determines the actual rate of change in the position of the controlled component. This derivative of the position feedback signal is used as the velocity feedback. The velocity error (the difference between the desired and the actual velocities) becomes the input to the motor driver. The motor driver is provided with a current feedback from the DC motor circuit to maintain continuous control of the motor.

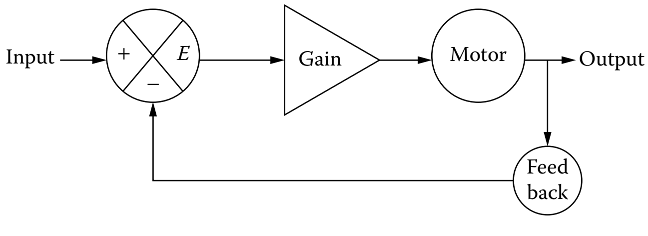


FIG. 7.3a
Basic block diagram of a closed-loop motion control system.

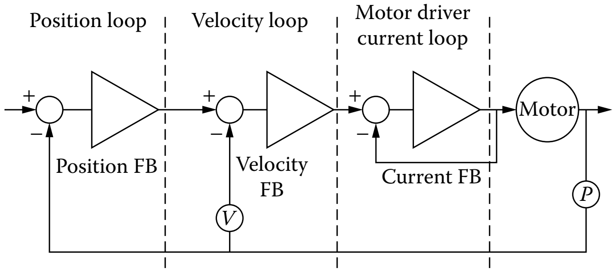


FIG. 7.3b
Motion system with embedded loops.

Any command for a movement that is executed under closed-loop control will result in a *motion profile*. This profile defines how the component is intended to be moved. Figure 7.3c illustrates a standard trapezoidal velocity profile.

In Figure 7.3c, the velocity is zero when the component to be moved is in its starting position, but is not moving yet. As time passes, the velocity of the component is ramped up at a constant acceleration until a limiting velocity is reached. This velocity remains constant until the desired end (stopping) position is approached. When it does, the velocity is reduced at a constant rate of deceleration until zero velocity is reached as the moved element arrives at the end position.

POSITIONING SYSTEM COMPONENTS

In closed-loop systems for positioning of industrial machinery, one of the components used is a motor or actuator. The motor driver receives its commands from a controller. When the controller receives a command to initiate a move, it compares the real-time readings of position or velocity to the

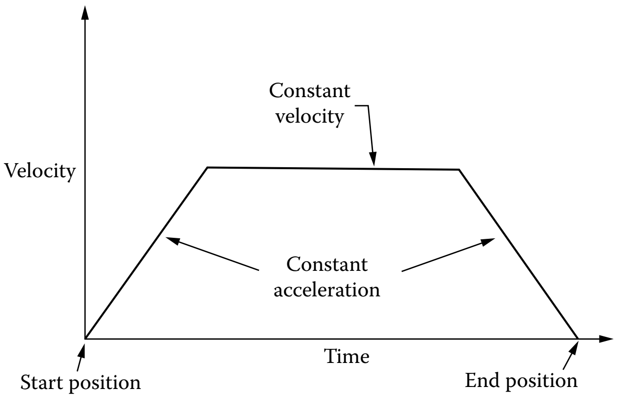


FIG. 7.3c
The standard velocity profile is trapezoidal.

desired values in the move command and sends the appropriate signal to the motor driver.

The motor driver adjusts the driving current (or other power source used) to operate the motor in the direction and at the rate that are commanded by the controller. The motor moves the machine axis, and one or more feedback devices report the actual position and speed back to the controller. The components in such control loops will be described in more detail in the following paragraphs.

Motors and Actuators

The power supply required for moving components can be electrical, hydraulic, or pneumatic. Of the three, electric drives are the most popular.

Hydraulic actuators are fast and powerful and can provide stiffness. These properties result from the incompressibility of the hydraulic fluid and from the energy stored as the momentum of the moving fluid, but the use of hydraulic actuators also necessitates the use of hydraulic pumps and plumbing. The disadvantages of hydraulic actuators is their greater space requirement, audible noise, and potential for leakage.

Pneumatic actuators are not as powerful or fast as hydraulic ones, but can still handle the less demanding applications that do not require much stiffness in the holding position. Compared to hydraulic designs, one advantage of the pneumatic actuators is that they do not leak oil.

Electric motors and actuators are the most widely used and are available in the largest variation of designs. The electric motor can either be synchronous types of stepper motors utilizing permanent magnet, variable reluctance and hybrid designs, or DC motors of the brushed or brushless types.

Electric Motors

Permanent Magnet Stepper Because opposite magnetic poles attract and like ones repel each other, stepper motors can use magnetic fields to move a rotor with respect to a stator by the exploitation of magnetic attraction and repulsion. An electrical pulse sent to a stepper motor causes it to rotate by a small amount, or step. The number of pulses received results in a corresponding amount of motion, and the frequency of the pulses determines the speed of that motion.

In a permanent magnet-type stepper, the permanent magnets are located on the rotor. The movement of the rotor magnets is caused by the magnetic fields of the stator coil windings. By energizing four stator fields in sequence, rotation at 90° intervals is achieved. Some permanent magnet steppers use 45° steps.

Externally mounted electronic circuits direct which stator windings will receive each pulse. This pulse direction or switching is called commutation. When at rest and receiving no pulses, the stepper has a holding torque that tends to maintain the last position, even when no power is being applied.

Variable Reluctance Stepper Variable reluctance steppers have many more poles of permeable material on the rotor. These are formed by teeth, each of which forms a permanent magnetic pole when energized by the nearby stator coil. The teeth align themselves with the respective poles of the stator windings. This provides a finer resolution per step, typically 15° or less.

Hybrid Synchronous Stepper Hybrid synchronous steppers have a high number of permanent magnetic teeth on the rotor, with an equally high number of teeth magnetized by the stator windings. The stator windings can be driven in either direction, forming either a North or South magnetic pole. This further improves resolution over either the permanent magnet or variable reluctance types of stepper motors.

Microstepping Even higher angular resolution is possible through a technique called microstepping. Instead of applying full power to one set of windings and no power to the others, microstepping involves the provision of percentages of power to more than one set of windings. This enables the determination of positions in-between the normal steps. When microstepping, more power is used due to the powering of more coils at any given time. Also, there is no holding torque in a microstepped position when power is lost (the holding torque will act to maintain the closest normal step position).

Brushed DC Motors The brushed DC motor has been around for a long time. It is simple and cheap, and its speed and torque are easily controlled. The magnetic field of the stator (stationary field) is produced by field windings or by permanent magnets.

Most DC motors of up to a few horsepower are made by using permanent magnets for the stator field. The armature field is the moving field associated with the rotating shaft. It is generated by windings on the rotor. Due to the two magnetic fields of the stator and rotor, the rotor moves toward alignment of the fields (North attracting South poles). A larger number of poles in the rotor and stator fields results in a more smoothly operating motor.

The electric current powering the rotor fields must be constantly switched as pole alignment is approached, in order to maintain rotation. The switching is called commutation. A set of stationary conductive brushes rubs onto contacts near one end of the rotating shaft to provide connection to the rotor coils.

Torque is easy to control because it is directly related to the armature current. Speed is easy to control because it is related to the voltage. One disadvantage of DC brushed motors is the wear of the brush material, requiring maintenance and generating dust. Another is the electrical noise generated during the commutation. In addition, it is difficult to maintain smooth operation when controlling at low speeds.

Brushless DC Motors Brushless DC motors have most of the same attributes as the brushed type, except that they do not have brush wear. They are more expensive and more complex because of the electronics of the brushless commutation. They operate in the same way as a brushed type, but use sensors to indicate the rotational position of the armature. The armature field is generated by permanent magnets, and the stator fields use coil windings. When the armature sensors indicate that the next set of stator field windings need to be energized, an electronic switching circuit makes the change.

Motor Drivers

Once the motor type and size are selected, a suitable motor driver is needed. The motor driver must accommodate the performance and power requirements of the specific motor to be used. This relationship can be complex.

The motor driver is an amplifier that converts the signal from the controller into the voltage and current signals needed to drive the particular motor type. It is usually best to select a compatible motor driver model that is supplied or recommended by the motor manufacturer. Sometimes the motor driver is mounted inside the motor housing, so that the motor and drive are purchased as a single unit.

Controllers

The positioning servo controllers can use the types of position and velocity control loops that were shown in Figures 7.3b and 7.3c or can use proportional, integral, derivative (PID) controls.

In a proportional-only controller (P), the controller output is equal to the error signal multiplied by the gain (the gain is usually signified as “A”), as was shown in Figure 7.3a. So, the farther away the positioned component is from the desired (commanded) position, the greater will be the amplitude of the motor drive output signal.

The gain can be set high in order to obtain a fast response, but this can result in more overshoot, which occurs if the motion continued past the commanded position and has to come back in order to reach it. If the gain is set low, then the positioning accuracy may suffer because the positioning is too slow or there is not enough power to overcome friction and reach the exact position as commanded.

A PID control solves these problems, because it does not only respond to the state at the moment, but its integral mode also looks at the past history or error, while its derivative mode predicts and corrects for potential future errors. The derivative mode responds to the rate of change of position so that the speed of changing the position can be increased faster than that which is possible with proportional gain control only. The integral function responds to the lower frequency requirements, including that of a steady state, to apply enough restoring force to reach and maintain the desired final position. In a PID control, the proportional, integral, and derivative modes are all incorporated into the gain function, A.

Sensors for Feedback

In a closed-loop control system, position feedback to the controller is required. For velocity feedback either a velocity sensor can be used, or the velocity can be obtained by taking the derivative of the position signal. Popular position feedback devices include potentiometers, optical and magnetic encoders, linear variable differential transformers (LVDTs), and magnetostrictive position transducers. Potentiometers and encoders can be linear or rotary.

A rotary type of LVDT is called a rotary variable differential transformer (RVDT). It is possible to make rotary magnetostrictive position transducers, but these are not common as yet. Magnetostrictive position transducers are normally used only for linear measurements. They are very cost effective for longer strokes, because the rod assembly is relatively inexpensive when compared to the electronics head.

Potentiometers Potentiometers detect the position by making physical contact with the moving component. A potentiometer utilizes a conductive element and a metallic contact that traverses over the range of positions that the measured component can travel. It is a contact device, meaning that the metal contact rubs against the resistive element. This also means that the contact and element are subject to wear during normal use.

There are three electrical connections to a potentiometer, whether it is a linear or a rotary measuring device. Two connections are made to the ends of the resistive element, and the third to the moving contact, called the wiper. The resistive element is mounted in a housing that includes an internal structure to guide the wiper carrier for even wiper contact along the element surface. An actuator rod is often attached to the wiper carrier, and this is used to provide a mechanical connection to the moving part that is to be measured (Figure 7.3d).

A DC voltage is normally applied across the resistive element, and the voltage of the wiper is the position measurement signal. Some potentiometers can have a calibrated accuracy on the order of 0.1%, but the wear of the rubbing parts can require frequent maintenance. Potentiometers are used in lower cost systems where long-term reliability is not as important as initial system cost. The other position sensors listed here are noncontact, so they have nearly unlimited lifetime.

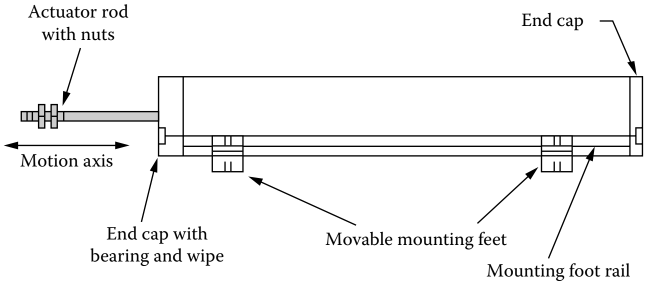


FIG. 7.3d
Components of a linear potentiometer position transducer.

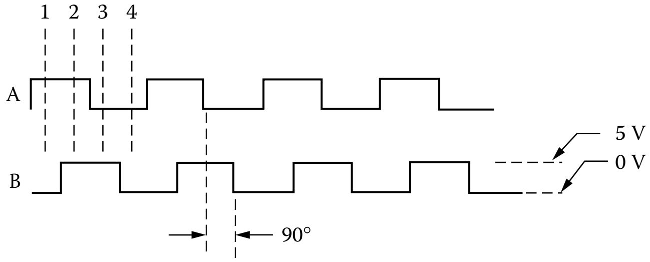


FIG. 7.3e
Incremental encoder pulse trains, showing the output trains A and B, which are separated by 90°.

Encoders Position can be detected by optical or magnetic encoders, which generate pulses in proportion to the amount of movement detected. Encoders can be incremental or absolute reading. An incremental encoder outputs a number of pulses corresponding to the amount of motion. There are usually two outputs, called A and B, that are separated in time by 90° in order to indicate the direction of travel. This is shown in Figure 7.3e.

The two outputs being separated by 90° are called quadrature outputs. A counter counts up all of the increment and decrement counts to arrive at the total count that represents the present position. There are four states of the combination of A and B levels per the 360° of one cycle. The phase relationship between outputs A and B during a state change indicates the direction of motion (increment or decrement of the count).

The drawback of an incremental encoder is that the count can be corrupted by power interruption or electrical noise. Then, in order to restart the system, it needs to be homed in order to re-establish the zero count position. A pictorial representation of an incremental rotary optical encoder is shown in Figure 7.3f.

In an optical encoder, a linear pattern or rotor disk is provided with adjacent transparent and opaque sections, allowing pulses of light to reach the detectors as the move progresses. These light pulses produce the electrical output pulses that indicate the linear or angular position change and direction.

An absolute encoder is more complex than an incremental one, in that it must have a number of light paths and light detectors equal to the number of bits of resolution, but it has the advantage of maintaining an accurate position even through power loss and other disturbances.

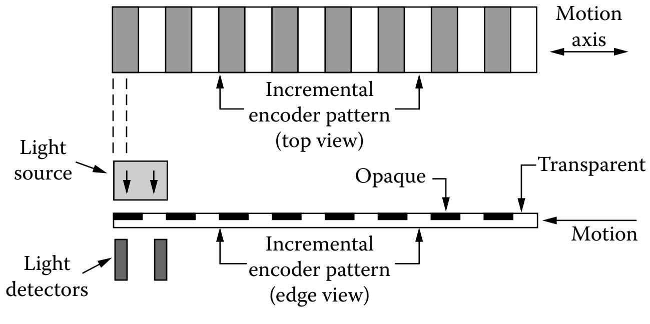


FIG. 7.3f
Incremental rotary optical encoder.

FIG. 7.3g
The outside appearance of a typical LVDT-type position sensor. (Courtesy of Revolution Sensor Company.)

A rotary encoder is often attached to the motor shaft. Linear encoders can be mounted so that they directly measure the output axis, eliminating errors due to gear backlash.

Although Figure 7.3f illustrates an encoder with optical sensors, encoders are also available with magnetic sensors. In a linear magnetic encoder, a tape is laid out along the axis of the motion and read by a reading head that moves with that axis. Alternatively, the tape and reading head can be incorporated into a slide assembly and purchased as a unit.

LVDT A linear variable differential transformer is suitable for measuring relatively short travel distances at high accuracy, in the range of microinches (or microns) to several inches (≈100 mm). As shown in Figure 7.3g, an LVDT is usually cylindrical in shape and is provided with a core that moves into the center bore of the cylinder.

An LVDT consists of three symmetrically spaced coils, which are carefully wound onto an insulated bobbin.[1] As shown in Figure 7.3h, the primary coil is in the center and it is surrounded by the two secondary coils.

The primary coil is driven by an AC waveform, usually a sine wave of 250 Hz to 10 kHz. The two secondaries are connected in series-bucking so that the output is also an AC waveform. The secondary difference output is demodulated and amplified to produce a DC output voltage. The output

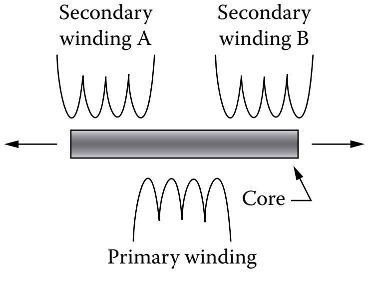


FIG. 7.3h
The primary and the two secondary windings of an LVDT.

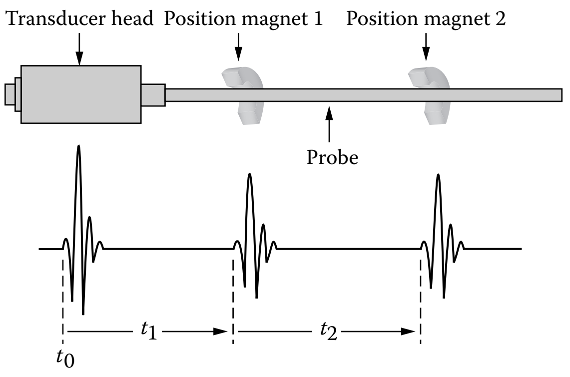


FIG. 7.3i
The components of a magnetostrictive position transducer.

voltage amplitude indicates the magnitude of the position, and its polarity indicates the direction from the center or null.

The AC waveform generator, demodulator, and amplifier comprise a signal conditioner. The signal conditioner may be purchased as a separate component when using an AC LVDT. Alternatively, if a DC LVDT can be used, it includes the signal conditioner within the same housing as the LVDT.

Magnetostrictive Position Transducer A magnetostrictive position transducer can have as high a resolution as 1 μ and a range from less than an inch (< 25 mm) to over 60 ft (> 20 m). Errors as low as 0.01% of range are typical. Figure 7.3i illustrates that it consists of a transducer head, a probe section, and one or more position magnets.

The transducer probe houses a waveguide made from magnetostrictive material and measures the location of a permanent magnet called the position magnet. A magnetostrictive material is one that changes shape or size when magnetized.

The operating sequence is such that at time t_0 in Figure 7.3i, the electronics module within the transducer head applies a current pulse to the waveguide, which forms a magnetic field around it. At the location of the position magnet, the interaction of the two fields imparts a mechanical torsion onto the waveguide due to the vector sum of the fields. The torsional force travels as a sonic wave along the length of the waveguide at a speed of approximately 9000 ft/s (3000 m/s). For each measurement cycle, one current pulse is applied, which instantaneously forms a magnetic field on the full length of the waveguide. The torsion pulse is generated at the location of the position magnet.

At the time of applying the current pulse, an electronic timer is also started. When the above-mentioned sonic wave is detected at the transducer head, the timer is stopped.[2] The elapsed time indicates the location of the position magnet (time t_1 or $t_1 + t_2$, in Figure 7.3i).

The update sequence can occur at frequencies of up to 4 kHz, depending on the length of the sensing probe. Magnetostrictive position transducers are accurate and are very cost effective, especially for longer stroke lengths of over 3 ft (> 1 m). They also have the capability to read the location of more than one position magnet while using only one sensing probe. This reduces system cost.

COMMUNICATION PROTOCOLS

Positioning systems can use a network communication configuration, where the components operate as nodes on a network. Alternatively, they can use a direct backplane interface, where the motion control system uses a PC card that plugs into a PC bus. Wireless communication networks are not yet fast enough to allow these sensors to be used in closed-loop position control.

Network communications protocols include ARCnet, CANbus, Device net (a version of CANbus), Ethernet, Profibus, IEEE 1394 (Firewire), IEEE 1451, Interbus-S, SERCOS, and Seriplex, among others.

PC bus-type protocols include the normal backplane ISA/EISA (PC-XT/AT) connection for a PC, Mac PCI (Nubus) for Macintosh computers, Multibus, PC 104, PCI bus, compact PCI (cPCI) bus, PCMCIA, VME bus, and VXI.

Each of the above communication protocols have fairly sophisticated definitions, and so the reader is directed to contact the product manufacturers to determine which are available in the specified product, and to learn about the advantages offered by each in the particular application. For a detailed description of network and bus protocols, the reader can also refer to the third volume of this handbook, *Process Software and Digital Networks.*

CONCLUSIONS

A motion control system comprises of a matched set of components including a motor, motor drive, feedback device, controller, and software. The best approach to implementing such a system is to first specify a motor and drive that are capable of meeting the performance requirements. Once they are specified, select the remaining parts of the system to match the requirements of the motor and drive combination with the help of their manufacturer, because the manufacturer can select the compatible components for optimum performance.

Often a custom system will be used. In addition to the performance of the control axis, networking capability, user interface, and environmental considerations should also be addressed.

References

1. Herceg, E., *Handbook of Measurement and Control,* New Jersey: Schaevitz Engineering, 1976, pp. 3–11.
2. Nyce, D. S., *Linear Position Sensors, Theory and Application,* New York: John Wiley & Sons, 2003, p. 140.

Bibliography

Anderson, N. A., *Instrumentation for Process Measurement and Control,* 3rd edition, Boca Raton, FL: CRC Press, 1998.

Astrom, K. J., *Automatic Tuning of PID Controllers,* Research Triangle Park, NC: Instrument Society of America, 1988.

Lipták, B. G., ed., *Instrument Engineers Handbook, Process Measurement and Analysis,* 4th edition, Boca Raton, FL: CRC Press, 2003.

Lloyd, T. C., *Electric Motors and their Applications,* New York: John Wiley & Sons, 1969.

Nyce, D. S., "Magnetostriction-Based Linear Position Sensors," *Sensors Magazine,* 11(4), 1994.

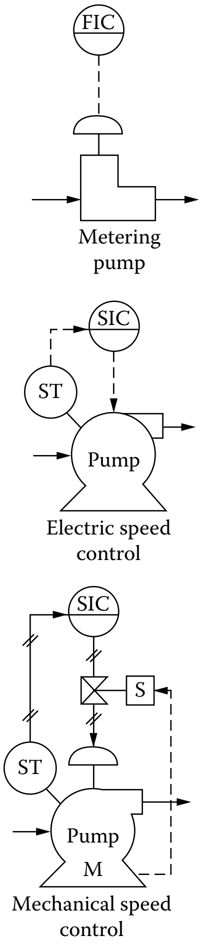
Nyce, D. S., "Position Sensors for Hydraulic Cylinders," *Hydraulics & Pneumatics,* November 2000.

Shinskey, F. G., *Simulating Process Control Loops Using Basic,* Research Triangle Park, NC: Instrument Society of America, 1990.

Webster, J. G., ed., *The Measurement, Instrumentation, and Sensors Handbook*, Boca Raton, FL: CRC Press, 1999.

7.4 Pumps as Control Elements

A. BRODGESELL (1970) **R. D. BUCHANAN** (1985)
B. G. LIPTÁK (1995) **I. H. GIBSON** (2005)



Flow sheet symbols

Types of Pumps:	A. Radial-flow pumps B. Axial- and mixed-flow pumps C. Peripheral or regenerative turbine pumps D. Pitot or jet pumps E. Progressing cavity pumps F. Flexible-rotor pumps G. Peristaltic pumps H. Gear pumps I. Plunger and piston pumps J. Diaphragm pumps K. Eductors L. Blow-egg and air lifts *Note:* Types A, B, and E are available in both conventional and submersible designs.
Rangeability:	*Variable-speed drives*: From 4:1 to 40:1 (electrical types), from 4:1 to 10:1 (mechanical designs), up to 40:1 (hydraulic types).
Metering pumps:	Can exceed 100:1 combining variable speed and variable stroke.
Efficiencies:	Pump efficiencies range from 85% for large-capacity centrifugals (types A and B) to below 50% for many of the smaller units. For types I and J, the efficiency ranges from 30% and up, depending on power and number of heads. For type J the efficiency is at most 30%, and for other types as low as 25%.
Materials of Construction:	For water using type A or B pumps: normally bronze impellers, bronze or steel bearings, stainless or carbon steel shafts, cast iron housing. For industrial process services, stainless steel and cast steel. For corrosive services, engineering plastic materials are common.

Costs: Varies with size and horsepower. For example, type A pumps cost from $650 to more than $34,000. Typically, a type A 10 hp (7.5 kW) sewage pump might cost $5000 for a horizontal and $4000 for a vertical model. A 10 hp (7.5 kW) type F ejector costs about $24,000, but to do the same work as the type A pump, a 30 hp (22 kW) $47,000 ejector is needed. A 10 hp (7.5 kW) type I, J, or K sludge pump costs about $12,000. Prefabricated stations (including pump) range from $13,000 to $100,000. The cost is $2700 for a pneumatic metering pump and $5400 for type K metering pump with positioner. The cost of stainless steel pumps is three to four times that of cast iron or bronze-fitted pumps. The purchase cost of a submersible pump is higher than that of one of the “dry pit” types, but the total installed cost is lower because there is no need for both a dry and a wet well (Figure 7.4k).

Partial List of Suppliers: There are a multitude of pump manufacturers throughout the world, many of whom make a variety of types. For a survey of pump manufacturers, with access to catalogs and mechanical details, the Web offers several excellent sources, such as GlobalSpec, http://flow-control.globalspec.com/ProductFinder/Flow_Transfer_Control/Pumps; and Thomas Global Register, http://www.thomasglobal.com/us/products18/pumps_suppliers.htm.

Allweiler Pump, down:

Allweiler Pump (A, B, E, G, I) (www.allweiler.de)
American Lewa (J, K) (www.amlewa.com)
American Turbine Pump Co. (B) (americanturbine.net)
Ashbrook-Simon-Hartley (I)
ASF Thomas Inc. (G) (www.thomaspumps.com)
Aurora Pump (A, C) (www.aurorapump.com)
Blackmer Pump (A, H, I) (www.blackmer.com)
Bran+Luebbe Inc. (J, K) (www.branluebbe.com)
BW/IP International Inc. (A), now Flowserve
Cat Pumps (J) (www.catpumps.com)
Clark-Cooper Div. of Magnetrol (J) (www.clarkcooper.com)
Cole-Parmer Instrument Co. (G) (www.coleparmer.com)
Cornell Pump Co. (A, B) (www.cornellpump.com)
Crane Pumps and Systems–Barnes Pumps (A, J, K) (www.cranepumps.com/barnes)
Dean Pump Div. Metpro Corp. (A) (www.deanpump.com)
Duriron Div. of Flowserve (A, C, D, H) (www.flowserve.com/pumps)
Eaton Corp. (H) (hydraulics.eaton.com)
Edson International (G, K) (www.edsonpumps.com)
Edwards High Vacuum International, now BOC Edwards (E) (www.bocedwards.com)
Edwards Inc. & Jones Inc./Willet (J) (www.edwardsandjones.com)
EMU Unterwasserpumpen Gmbh (A) (www.emu.de/english)
Fairbanks Morse Pump (A, B) (www.fmpump.com)
Flowserve (A, B–J, K) (www.flowserve.com/pumps/ index.htm)
FMI—Fluid Metering Inc. (J) (www.fmipump.com)
GE Osmonics (A, K) (www.gewater.com/equipment/pumps/index.jsp)
GE Ruska (J) (www.ruska.com)
Geho Pumps–Div. Weir Netherlands b.r. (J, K) (www.geho.nl/minerals/home.nsf)
Gorman-Rupp Industries Div. (A, C, D, E, H) (www.gormanrupp.com)
Goulds Pumps (A, J) (www.goulds.com)
Hale Fire Pump Co. (A) (www.haleproducts.com)
Hydroflo Pumps (A, B) (www.hydroflo pumps.com)
IMO Industries Inc. (I, J) (www.imo-pump.com)
Ingersoll Dresser Pump Div. of Flowserve (A, C) (www.idpump.com)
IR ARO Fluid Products (K) (www.arozone.com/Index.html)
ITT A-C Pump (A, K) (www.gouldspumps.com/ac_files)
ITT Bell & Gossett (A) (www.bellgossett.com)
ITT Jabsco Products (A, F) (www.jabsco.com)
ITT Marlow Pumps (AB) (www.marlowpumpsonline.net)
Jaeco Fluid Systems Inc. (J, K) (www.jaecofs.com/metering_pumps.html)
KNF Neuberger Inc. (J, K) (www.knf.com/usa.htm)
Komline-Sanderson (J) (www.komline.com/index.html)
LaBour– Taber Div. of Peerless Pump Co. (A) (216.37.51.16/labourtaber)
Lakeside Equipment Corp. (I) (www.lakeside-equipment.com)

Lear Romec Div. of Crane Co. (H) (www.learromec.com)
Linc Milton Roy (J) (www.lincpumps.com)
Liquiflo Equipment Co. (A, H) (www.liquiflo.com)
Lutz Pumps Inc. (A, K) (www.lutzpumps.com)
McFarland Pump Co. (J, K) (www.mcfarlandpump.com)
Micropump Corp. (A, G, J, K) (www.micropump.com)
Mono Pumps Ltd. (Monoflo in North America) (E) (www.mono-pumps.com)
MP Pumps Inc. (K) (www.mppumps.com)
Nagle Pumps (A) (www.naglepumps.com)
Netzch Mohnopumpe Gmbh (A)
Oberdorfer Pumps Inc. (E, H) (www.thomaspumps.com/pumpsprofile.htm)
Pacer Pumps (A) (www.pacerpumps.com)
Patterson Pump Co. (A, B) (www.pattersonpumps.com)
PCM Moineau (E) (www.pcmpompes.com)
Peerless Pump (A, B) (www.peerlesspump.com)
Plenty Mirrlees Pumps Div. of SPX Process Equipment (H, I) (www.plenty.co.uk/pumps)
Price Pump Co. (A, K) (www.pricepump.com)
Pulsafeeder Div. of IDEX Corp. (J, K) (www.pulsa.com)
QED Environmental Systems Inc. (L) (www.qedenv.com)
Robbins & Myers Inc. —Moyno (A, E, J) (www.robn.com)
Roper Pump Co. (H) (www.roperpumps.com)
Tuthill Corp.(H) (pump.tuthill.com)
Valcor Engineering Corp. (J) (www.valcor.com)
Vanton Pump & Equipment Co. (A, G) (www.vanton.com)
Viking Pump Div. of IDEX Corp. (C, H) (www.vikingpump.com)
Wallace & Tiernan (K) (www.wallace-tiernan.com)
Wanner Engineering Inc. (A, G, K) (www.wannereng.com)
Warman Pump Group (A, G, K) (www.warman.co.za)
Warren Pumps Inc. (H) (www.warrenpumps.com)
Waukesha Cherry Burrell Div. of SPX Process Equipment (H) (www.gowcb.com)
Weir Minerals Division (A, J, K) (www.weirminerals.com)
WEMCO Div. of Weir Clearliquid (A) (www.weirclearliquid.com)
Wright Pump Div. of IDEX Corp. (H) (www.idexcorp.com/groups/wright.asp)
Zenith Div. of Parker Hannifin Corp. (H, J) (www.zenithpumps.com)
Zimpro/Passavant Inc. (I) (www.usfilter.com)

Because pumping is the primary means of liquid transportation in most processes, pumps are important parts of control systems. The various pump control systems will be discussed in Chapter 8; the features and designs of variable-speed drives are covered in Section 7.10; and metering pumps have been discussed in Section 2.14 of the *Process Measurement and Analysis* volume of this fourth edition of the handbook. Therefore, the main emphasis of this chapter will be to describe the features and selection of conventional (centrifugal, reciprocating, screw) pumps and their applications. In addition, this section will also briefly discuss metering pumps for the benefit of those who do not have access to *Process Measurement and Analysis.*

ROTODYNAMIC OR CENTRIFUGAL PUMPS

Types A, B, C, and D of the feature summary (Table 7.4a) fall within this classification, which is the most common type of pump. In the form of tall, slender, deep well submersibles, they pump clear water from depths greater than 2000 ft (600 m). Horizontal centrifugals with volutes almost the size of a man can pump 9000 gpm (0.57 m^3/s) of raw sewage through municipal treatment plants. Few applications are beyond their range, including flow rates of 1–100,000 gpm (3.78 lpm to 6.3 m^3/s) and process fluids from liquefied gases through clear water to all but the densest sludge.

Radial-Flow

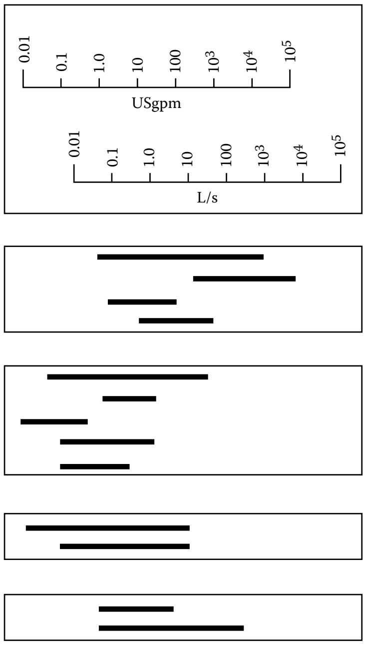
Radial-flow pumps are designed to throw the liquid entering the center of the impeller or diffuser out into a spiral volute or bowl. The impellers may be closed, semi-open, or open, depending on the application (Figure 7.4b). Closed impellers have higher efficiencies and are more popular than the other two types. They can readily be designed with nonclogging features. By using more than one impeller the discharge head characteristics can be increased, in proportion to the number of impellers. These pumps may be of horizontal or vertical design. Multiple stage designs with up to 99 impeller/volute assemblies on a single shaft are available, though not common. Flow can be throttled, but many pumps have a minimum

TABLE 7.4a

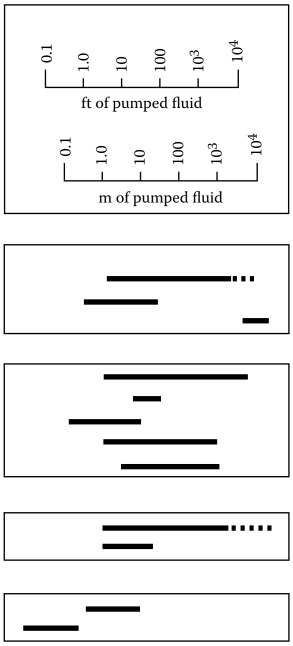
Pump Feature Summary

Type designation	Type of pump	Clear liquids—Low viscosity	Clear liquids—High viscosity	Thin slurries or suspensions	Raw or partially treated sewage and heavy suspensions	Viscous or thick slurries and sludges
Rotodynamic pumps						
A	Radial flow centrifugal	✓		✓	✓	✓
B	Axial- and mixed-flow centrifugal	✓		✓	✓	
C	Peripheral or regenerative turbine	✓				
D	Pitot or jet pumps	✓				
Positive displacement pumps—Rotary						
E	Progressing cavity	✓	✓	✓	✓	✓
F	Flexible-rotor	✓	✓	✓		
G	Peristaltic	✓				
H	Gear	✓	✓			
I	Rotary screws		✓		✓	✓
Positive displacement pumps—Reciprocating						
J	Reciprocating piston and plunger pumps	✓	✓	✓	✓	✓
K	Diaphragm pumps	✓	✓	✓	✓	✓
Miscellaneous						
L	Pneumatic ejectors and blow eggs				✓	
M	Air lift pumps			✓	✓	

For liquid pumped — *Capacity range*



Developed head range



flow below which they become bistable, flipping between zero and a much higher flow. In some cases, this may be as high as 75–80% of design rate.

Axial- and Mixed-Flow

Axial-flow (propeller) pumps, although classed as centrifugals, do not truly belong in this category because the propeller thrusts axially rather than throwing the liquid outward. Impeller vanes for mixed-flow centrifugals are shaped so as to provide partial throw, partial push of the liquid outward and upward. Axial-flow and mixed-flow designs can handle huge capacities but only at the expense of a reduction in discharge heads. They are constructed vertically. The head/flow characteristic is such that throttling the flow is usually undesirable, and bypassing or speed control is a better control strategy.

Peripheral or Regenerative Turbine

Peripheral turbine pumps (Figure 7.4c) are low-flow/high-head devices that require very little positive suction head. The fluid enters an annular space near the periphery of the rotor, which has a large number of slots. The fluid is carried around by the rotor, gaining pressure as it circulates in the rotor slots, and is discharged after traveling about 300° around the casing. A small single-stage impeller can provide up to 500 ft (150 m) head.

Pitot or Jet Pumps Pitot pumps (Figure 7.4d) provide extremely high head, up to 5000 ft (1500 m), at relatively low flow rate. The pump has an internal cylindrical casing that rotates at high speed, with a fixed pitot pickup inside. The discharge

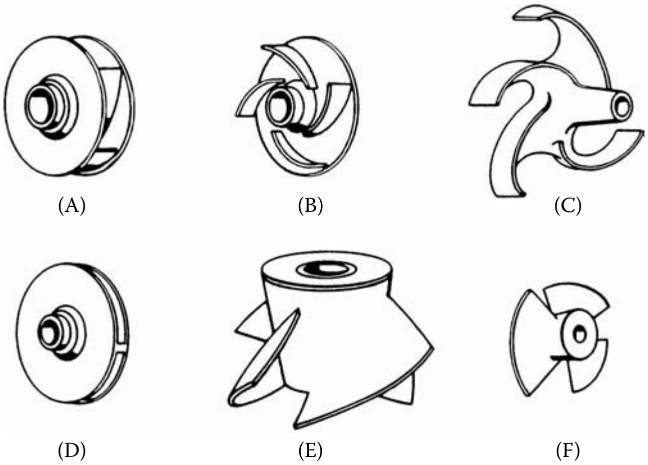


FIG. 7.4b

Types of centrifugal pump impellers: (A) closed impeller, (B) semi-open impeller, (C) open impeller, (D) diffuser, (E) mixed-flow impeller, (F) axial-flow impeller.

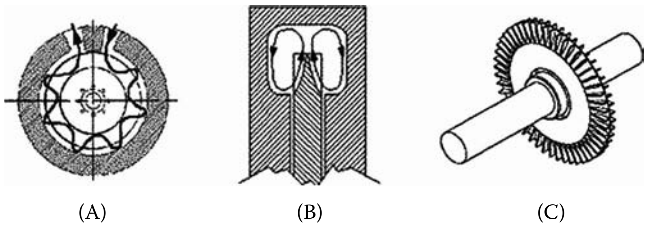


FIG. 7.4c

Peripheral turbine pump impeller: (A) view on shaft end, showing fluid path, (B) view of impeller and housing, showing fluid internal circulation, (C) impeller and shaft assembly. (Courtesy of Dynaflo Engineering Inc.)

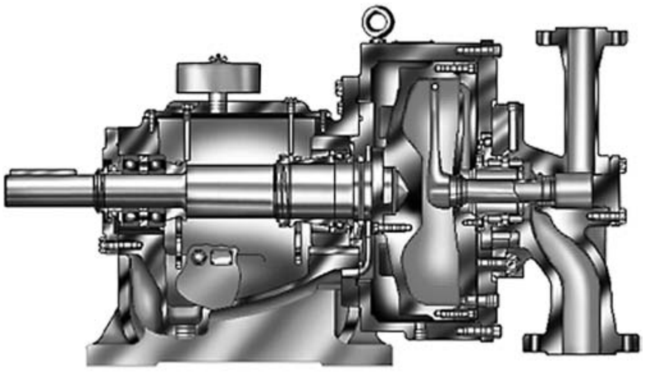


FIG. 7.4d
Roto-jet pitot pump sectional drawing. (Courtesy of Weir Clear Liquid Division.)

from the pitot exits on the impeller axis, coaxial with the suction connection.

Applications Most liquids and wastes can be pumped with centrifugal pumps. It is easier to list the applications for which they are not suited than the ones for which they are. They should not be used for pumping (1) very viscous industrial liquids or sludges (the efficiencies of centrifugal pumps drop to zero, and therefore various positive displacement pumps are used), (2) low flows against very high heads (except for deep well applications, the large number of impellers needed put the centrifugal design at a competitive disadvantage), and (3) low to moderate flows of liquids with high solids contents (except for the recessed impeller type, rags and large particles will clog the smaller centrifugals). For low flow and high head, the turbine and the jet pump designs may be competitive with positive-displacement types up to 400 hp (300 kW).

POSITIVE-DISPLACEMENT PUMPS

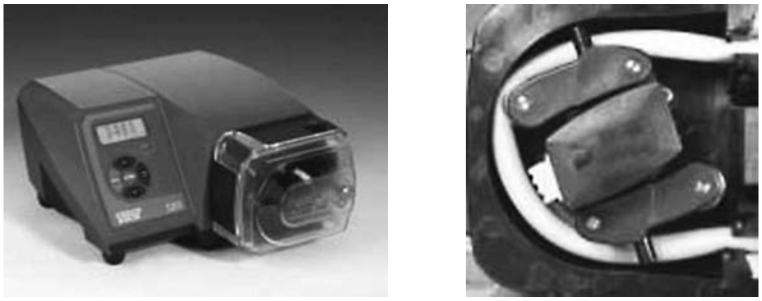
Rotary Pumps

Progressing Cavity Pumps Invented by R. Moineau in 1930, progressing cavity pumps have a helical metal rotor within a (usually) elastomer stator with a bihelical bore; the rotor is connected to the rotating drive through a sealed universal joint set or a flexible shaft enabling the rotor to nutate as it rotates. The resulting motion causes a series of trapped cavities to progress axially through the pump. Such pumps offer unusual capabilities in handling fragile products (small live fish can pass through them) and can be classed as semipositive displacement; the capacity is largely proportional to rotational speed; though slippage increases above a value depending on the tightness of fit of the rotor in the stator. The pumps must not be run dry, as this will rapidly damage the stator, but will act as vacuum pumps provided there is enough liquid to lubricate the system. They are therefore capable of suction lift equivalent to about 90% of the vapor pressure of the liquid. They find application in paste and slurry handling service, as well as small high-lift submersible and down-hole pumps. Variants with hopper/auger feed will handle filtercake and similar extremely high viscosity products. Multistage variants utilize a single, long rotor and a series of standard stator units. Other variants exist with an n-start helical rotor and $n + 1$-start stator, which have even less pulsation than the standard design. The progressive cavity pump with VSD can be one of the most useful tools for a control system engineer, offering a simple alternative to centrifugal pump, control valve, and check valve in a single package. A single-stage pump is capable of producing a differential of up to about 62 psid (430 kPa).

Flexible-Rotor Pumps A cylindrical metallic housing with diametric suction and discharge connections contains a multivaned elastomer rotor, eccentrically mounted (Figure 7.4e). As the rotor turns, the volume between the vanes increases to a maximum as the set of vanes pass the suction connection, trapping a volume of liquid, and then diminishes to a minimum as the trapped volume passes the discharge connection. The pump is reversible.

Peristaltic Pumps Widely used in laboratory applications, the peristaltic pump has an elastomer tube. held inside a cylindrical retainer and a set of spring-loaded rollers that trap a series of volumes of fluid in the tube as they rotate, transferring liquid from the suction to the discharge (Figure 7.4f). These commonly have multiple heads on a single variable-speed drive mechanism and will accept a variety of different bore tubes, enabling a variety of reagents and samples to be fed proportionately to analytical equipment. They are also used for industrial purposes, handling material such as nitric and hydrofluoric acids. Pumps will handle both liquids and gases,

FIG. 7.4e
Flexible rotor pump section. (Courtesy of Bombas Trief.)



(A) Laboratory peristaltic pump

(B) Industrial peristaltic pump

FIG. 7.4f
Peristaltic pumps: (A) laboratory, (B) industrial. (Courtesy of Watson-Marlow Bredel.)

and available capacities vary from fractions of a milliliter/minute to 1400 gph (90 lpm).

Gear Pumps

Gear pumps are available in a multitude of variations. Most designs have a pumping chamber with a pair of meshed gears; as the gears rotate, liquid is trapped between the housing and the gear teeth, being carried from the suction chamber to the discharge chamber. As the gear teeth engage, the liquid is forced out of the discharge port. Another common design is the “star and moon” with a planetary gear and a quarter-moon-shaped stationary component inside a driven ring-gear.

Reciprocating Pumps

Almost all reciprocating pumps are either metering or power pumps. The steam-driven pump is historically interesting but rarely used. Most frequently a piston or plunger is utilized in a cylinder, which is driven forward and backward by a crankshaft connected to an outside drive. Metering pump flows can readily be adjusted by changing the length and frequency of strokes of the piston. The diaphragm pump is similar to the piston type, except that instead of a piston, it contains a flexible diaphragm that oscillates as the crankshaft rotates. Diaphragm pumps commonly have a buffer hydraulic fluid to transfer the force to the diaphragm.

Smaller units may be solenoid driven or use pneumatic power.

Plunger and diaphragm pumps can feed metered amounts of chemicals (acids or alkalis for pH adjustment) or can also pump sludge and slurries. Plunger pumps commonly come as simplex (one head), duplex (two head) or triplex (three head). The triplex design, with the piston throws operating at 120° to each other, minimize pulsation in suction and discharge, and acceleration losses in the suction lines. Pulsation dampers are commonly fitted to further reduce these effects.

Variable-capacity hydraulic pumps utilize a variable-angle swashplate to alter the stroke of a set of pistons (commonly five) in a common assembly. Fitted with an integrated hydraulic servomechanism, these can be set up to provide constant discharge pressure/variable flow, or constant power (flow × pressure). This design can also be used as a reversible variable-speed hydraulic motor.

AIR PUMPS AND AIR LIFTS

This method of pumping (type L) employs a receiver pot, into which the wastes flow by gravity, and an air pressure system that transports the liquid to a treatment process at a higher elevation. A controller is usually included, which keeps the tank vented while it is filling. The level controller energizes a three-way solenoid valve when it is full to close the vent and open the air supply to pressurize the vapor space in the tank.

The air system may use plant air (or steam), a pneumatic pressure tank, or an air compressor directly. With large compressors, a capacity of 600 gpm (2.28 m^3/min) with lifts of 50 ft (15 m) may be obtained. The advantage of this system is that it has no moving parts in contact with the waste and thus no impellers to clog. Ejectors are normally more maintenance free and longer lived than pumps.

Condensate Pumps

A related device can be used to transfer steam condensate at subatmospheric pressure to a condensate return system, by injecting live steam above the condensate surface. When the vessel is empty, the vapor space is connected to the steam space of the heater to equalize pressure and allow the condensate to refill the vessel (Figure 7.4g).

Air Lifts

Air lifts consist of an updraft tube, an air line, and an air compressor or blower. Air is blown into the bottom of the submerged updraft tube, and as the air bubbles travel upward, they expand (reducing density and pressure within the tube), inducing the surrounding liquid to enter. Flows as great as 1500 gpm (5.7 m^3/min) may be lifted short distances in this way. Air lifts are of great value in waste treatment to transfer mixed liquors or slurries from one process to another.

DESIGN OF PUMPING SYSTEMS

In order to choose the proper pump, the conditions that must be known include capacity, head requirements, and liquid characteristics. To compute capacity, one should first determine the average flow rate for the system and then decide if adjustments are necessary. For example, when pumping wastes from a community sewage system, the pump must handle peak flows that are roughly two to five times the average flow, depending on the size of the community. Summer and winter flows and future needs may also dictate capacity, and the population trends and past flow rates should be considered in this evaluation.

Head Requirements

Head describes pressure in terms of height of fluid. It is calculated by the expression:

$$\text{head in feet} = \frac{\text{pressure (psi)} \times 2.31}{\text{specific gravity}} \qquad \mathbf{7.4(1)}$$

The discharge head on a pump is a summation of several contributing factors: static head, friction head, velocity head, and suction head.

Static head (h_d) is the vertical distance through which the liquid must be lifted (Figure 7.4h).

Friction head (h_f) is the resistance to flow caused by the friction in pipes. Entrance and transition losses may also be

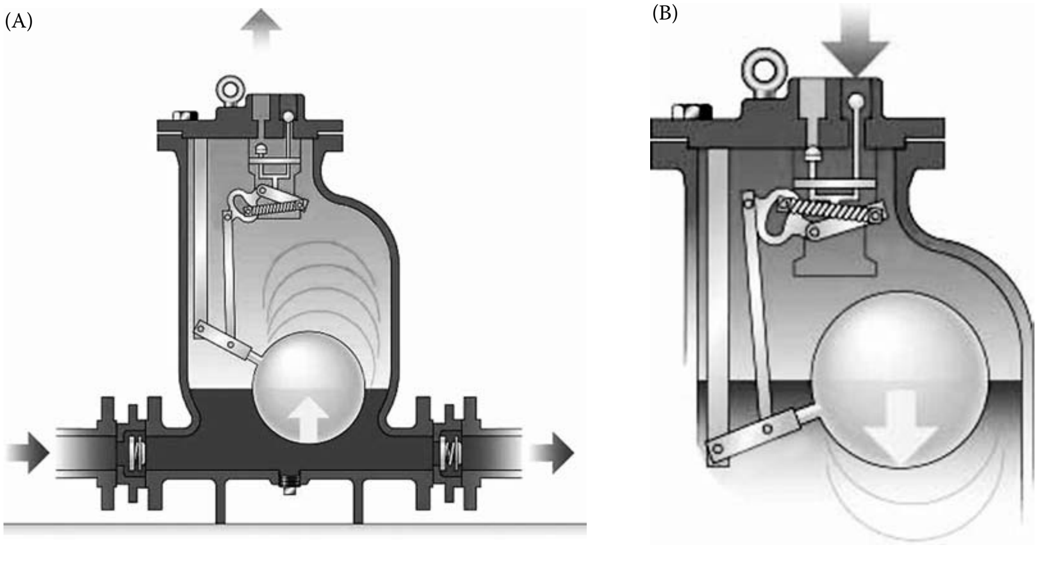


FIG. 7.4g
Automatic condensate pump operation: (A) filling, (B) emptying. (Courtesy of Spirax Sarco.)

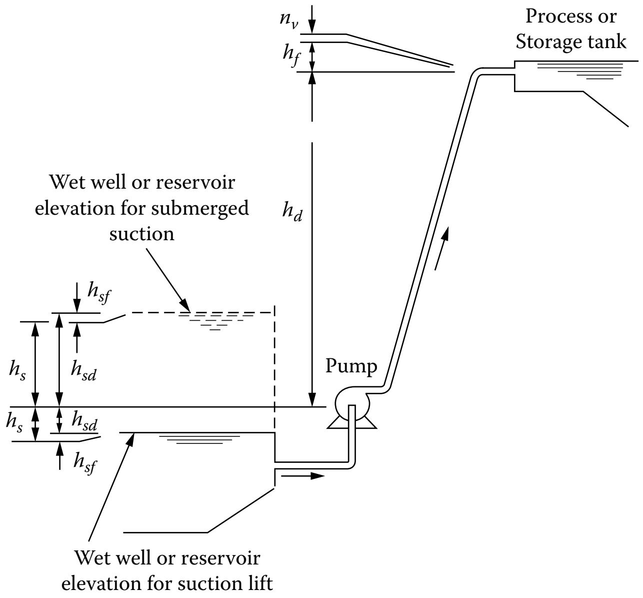


FIG. 7.4h

Determination of pump discharge head requirements. Legend: h_v = velocity head; h_f = friction head; h_d = static head; h_s = suction head; h_{sd} = suction side static head; h_{sf} = suction side friction head.

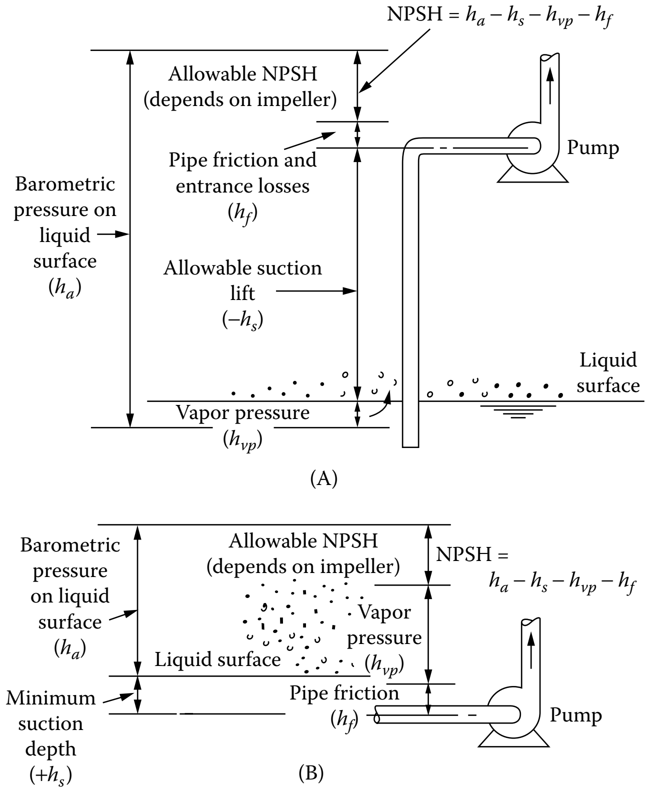


FIG. 7.4i

Role played by NPSH in determining allowable suction lift: (A) pump with suction lift, (B) pump with submerged suction but high vapor pressure (possibly hot water).

included. Because the nature of the fluid (density, viscosity, and temperature) and the nature of the pipe (roughness or straightness) affect the friction losses, a careful analysis is needed for most pumping systems, although for smaller systems, tables can be used.

Velocity head (h_v) is the head required to impart energy into a fluid to induce velocity. Normally this is quite small and may be ignored unless the total head is low.

Suction head (h_s), if there is a positive head on the suction side (a submerged impeller), will reduce the pressure differential that the pump has to develop. If the liquid level is below the pump, the suction lift plus friction in the suction pipe must be added to the total pressure differential required.

Total head (H) is expressed by

$$H = h_d + h_f + h_v \pm h_s \qquad \textbf{7.4(2)}$$

NPSH Calculation

The suction lift that is possible to handle must be carefully computed. As shown in Figure 7.4i, it is limited by the barometric pressure (which in turn is dependent on elevation and temperature); the vapor pressure (also dependent on temperature); friction and entrance losses on the suction side; and the net positive suction head (NPSH)—a factor that depends on the shape of the impeller and is obtained from the pump manufacturer.

In order for the pump to be able to “pull in” the pumped fluid, the net positive suction head available (NPSHA) for the particular installation must be greater than the NPSH that the pump requires. The NPSH values for the particular pump are obtained from the pump curve (Figure 7.4j), while the available NPSH is calculated according to the following equation:

$$\text{NPSHA} = h_a(+\text{ or }-)h_s - h_{vp} - h_f \qquad \textbf{7.4(3)}$$

where

h_a = the absolute pressure (in feet) at the surface of the source of the pumped liquid. If the source is atmospheric, h_a = 33.96 ft.

h_s = the static head of the installation, which is the vertical distance between the pump inlet and the surface of the liquid on the supply side. It is positive if the liquid level is above the pump inlet, and it is negative if it is below.

h_{vp} = the vapor pressure of the pumped fluid (in feet) at the operating temperature. The h_{vp} rises with temperature, and $h_{vp} = h_a$ when the liquid reaches its boiling point.

h_f = the suction side friction head in feet. This term increases with the square of flow and reflects the pressure drop through all pipes, valves, and fittings

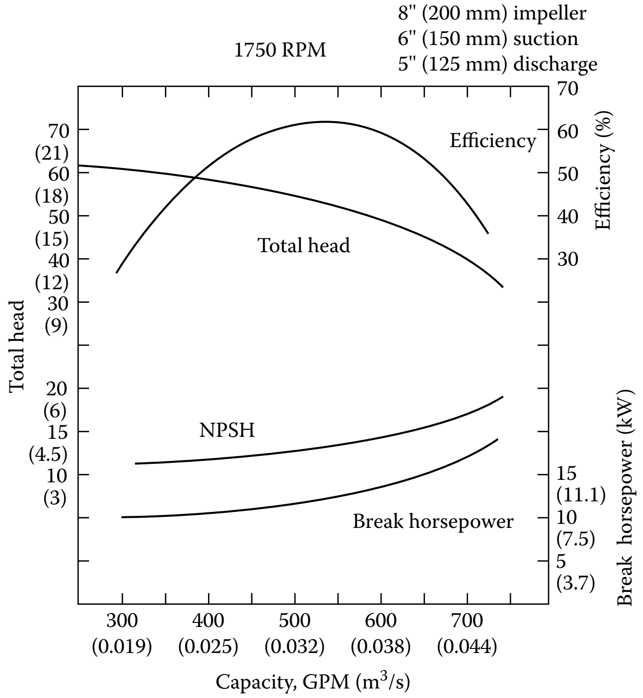


FIG. 7.4j
Typical pump curve for a single impeller.

on the suction side of the pump. When dealing with reciprocating pumps, the suction column is accelerated and decelerated with each stroke, and the head required to accelerate the suction column must be included.

If it is desired to convert NPSHA from feet to PSI, the NPSHA given in feet should be multiplied by the specific gravity of the fluid and should be divided by 2.31. (When converting to metric units, ft = 0.3048 m and PSI = 6.9 kPa.)

It is generally sufficient to calculate the available NPSH at maximum flow rate, because at that flow, the suction side friction head (h_f) is maximum and the NPSHA value is likely to be minimum. The NPSH required (NPSHR) of the pump rises with flow (Figure 7.4j). Therefore, if NPSHA exceeds NPSHR when the flow is maximum, it will exceed it by an even greater margin as the flow drops.

Specific Speed

The rotational speed of the impeller affects capacity, efficiency, and cavitation. Even if the suction lift is within permissible limits, cavitation can still occur, because additional static head is converted to velocity head as the fluid is accelerated in the pump.

The specific speed of the pump can be found using Equation 7.4(4).

$$\text{specific speed, } N_s = \frac{\text{RPM} \times \sqrt{\text{capacity (gpm)}}}{H^{3/4}} \qquad \textbf{7.4(4)}$$

Charts are available showing the upper limits of specific speed for various suction lifts. Caution: In metric units, specific speed values are NOT the same as in U.S. units, as the relationship is not dimensionally consistent.

Horsepower

The power required to drive the pump is called brake horsepower. It is found by solving Equation 7.4(5).

$$\text{BHP} = \frac{\text{capacity (gmp)} \times H\text{(ft)} \times \text{Sp. Gr.}}{3960 \times \text{pump efficiency}} \qquad \textbf{7.4(5a)}$$

In metric units,

$$\text{Power (kW)} =$$

$$\frac{\text{capacity (m}^3\text{/h)} \times H\text{(Nm/kg)} \times \text{density (kg/m}^3\text{).}}{3.670 * 10^5 \times \text{pump efficiency}} \qquad \textbf{7.4(5b)}$$

or

$$\text{Power (kW)} = \frac{\text{capacity (m}^3\text{/h)} \times H\text{(kPa)}}{3600 \times \text{pump efficiency}} \qquad \textbf{7.4(5c)}$$

Installation Considerations for Wastewater Pumping Stations

The typical designs for wastewater pumping stations are shown in Figure 7.4k. In selecting the best design for a particular application, the following factors should be considered:

1. Many gases are formed by domestic wastes, including some that are flammable. When pumps or other equipment are located in rooms below grade, the possibility of explosion or the build-up of these gases exists, and ventilation is extremely important. Similarly, such gases may be toxic (hydrogen sulfide) and asphyxiant (methane, carbon dioxide).
2. When pumping at high velocities or through long lines, water hammer can be a problem. Valves and piping should be designed to withstand these pressure waves. Even for pumps that discharge to atmosphere, check valves should be chosen so as to cushion the surge.
3. Bar screens and comminutors are undesirable because they require maintenance, but they may be necessary for small centrifugal pump stations where the flow might get clogged.
4. Pump level controls are not fully reliable because rags can short electrodes and hang on floats. Purged air systems (air bubblers) require less maintenance but need an air compressor that must operate continuously. Therefore, it is important to provide maintenance-free instrumentation.

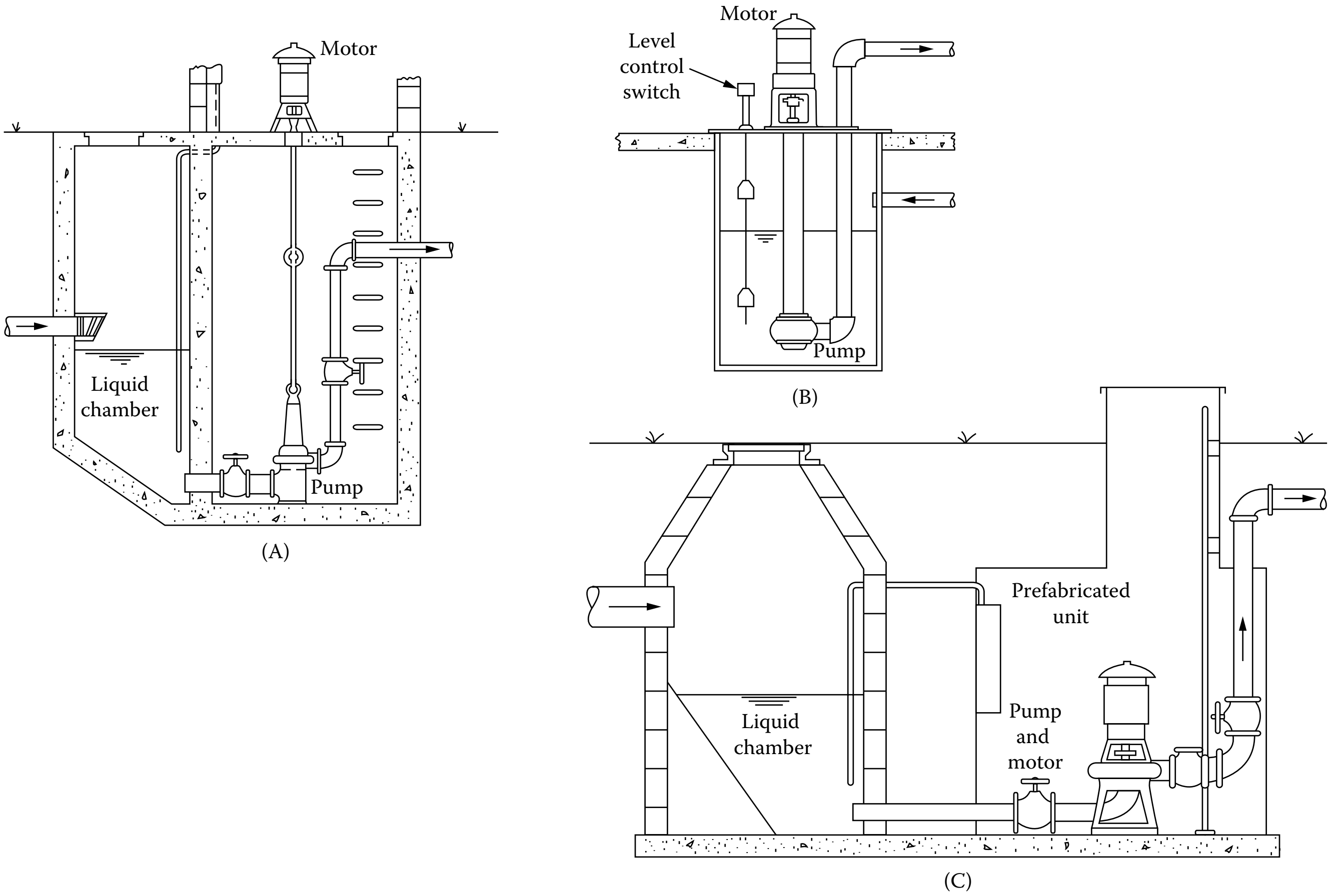


FIG. 7.4k
Pumping stations: (A) dry well design, (B) wet well design, (C) prefabricated pumping station.

5. Charts and formulas are available for sizing wet wells, but infiltration and runoff must also be taken into account.
6. Sump pumps, humidity control, a second pump with alternator, and a pump hoisting mechanism are desirable.
7. Most sewage utilities prefer the dry well designs for ease of maintenance.

METERING PUMPS

Flow control of liquids can be accomplished by means of pumps that incorporate the measurement and control element in a single unit. Metering pumps are designed to provide measurement and control of the process. For a measurement-oriented discussion of these pumps, refer to Section 2.14 in the *Process Measurement and Analysis* volume of this handbook.

Plunger Pumps

Plunger pumps are suitable for use on clean liquids at high pressures and low flow rates. A typical plunger pump is shown in Figure 7.4l. The pump consists of a plunger, cylinder, stuffing box, packing, and suction and discharge valves. Rotary motion of the driver is converted to linear motion by an eccentric. The plunger moves inside the cylinder with reciprocating motion, displacing a volume of fluid on each stroke.

Stroke length, and thus the volume delivered per stroke, is adjustable. The adjustment can be a manual indicator and dial, or for automatic control applications, a pneumatic actuator with positioner can be provided. Stroke adjustment alone offers operating flow ranges of 10:1 from maximum to minimum. Additional rangeability can be obtained by means of a variable-speed drive. A pneumatic stroke positioner used in conjunction with a variable-speed drive provides rangeability of at least 100:1. In the case of automatic stroke adjustment and variable speed, the pumping rate can be controlled by two independent variables, or the controller output can be "split-ranged" between stroke and speed adjustment.

The reciprocating action of the plunger results in a pulsating discharge flow, as represented in Figure 7.4m by the dotted simplex curve. For applications where these flow pulsations cannot be tolerated, particularly if a flow measurement is required, pumps can be run in duplex or triplex arrangements. With the duplex pump, two pump heads are

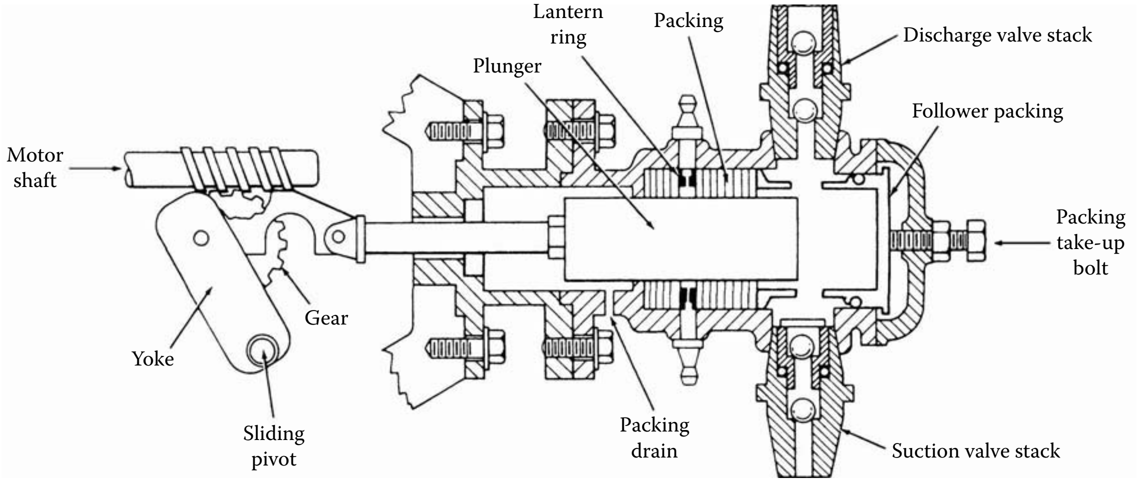


FIG. 7.4l
Plunger- or piston-type metering pump.

driven off the same motor, and the discharge strokes are phased 180° apart. With a triplex arrangement, three pump heads are driven by one motor and the discharge strokes are 120° apart. Both the duplex and triplex pumps provide a smoother flow than the single pump, as shown by the dashed and solid curves of Figure 7.4m.

For blending two or more streams, several pumps can be ganged to one motor. Stroke length adjustment can be used to control the blend ratio, and drive speed can control total flow. However, in this case rangeability is sacrificed for ratio control.

Pumping efficiency is affected by leakage at the suction and discharge valves. These pumps are therefore not recommended for fluids such as slurries, which will interfere with proper valve seating or settle out in pump cavities.

Diaphragm Pumps

Diaphragm pumps use a flexible diaphragm to achieve pumping action. The input shaft drives an eccentric through a worm and gear. Rotation of the eccentric moves the diaphragm on the discharge stroke by means of a push rod. A spring returns the push rod and diaphragm during the suction stroke. A typical pump is shown in Figure 7.4n.

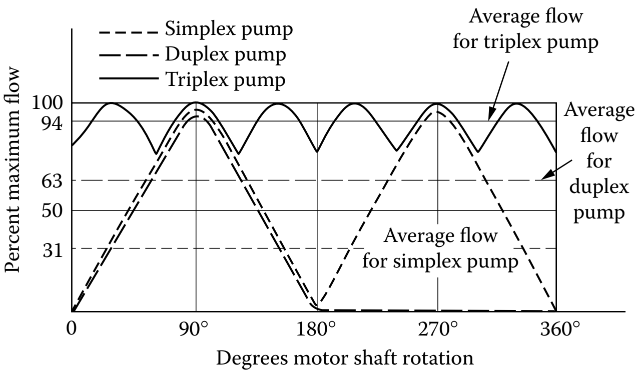


FIG. 7.4m
Flow characteristics of simplex and multiple plunger pumps.

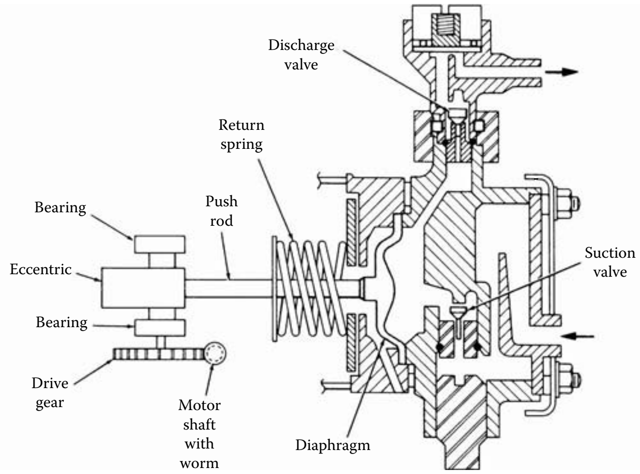


FIG. 7.4n
Diaphragm-type metering pump.

Operation of the diaphragm pump is similar to that of the plunger pump; however, discharge pressure is much lower due to the strength limitation of the diaphragm. Their principal advantage over plunger pumps is lower cost. Designs with two pumps driven by one motor can be used to advantage for increased capacity or to smooth out flow pulsations. By combining automatic stroke length adjustment with a variable-speed drive, operating ranges can be as wide as 20:1. These pumps can be used only on relatively clean fluids, because solids will interfere with proper suction and discharge valve seating or may settle out in the pump cavities.

The weakness of the diaphragm pump design is in the diaphragm, which is operated directly by the push rod. The diaphragm has to be flexible for pumping and yet strong enough to deliver the pressure. The strength requirement can

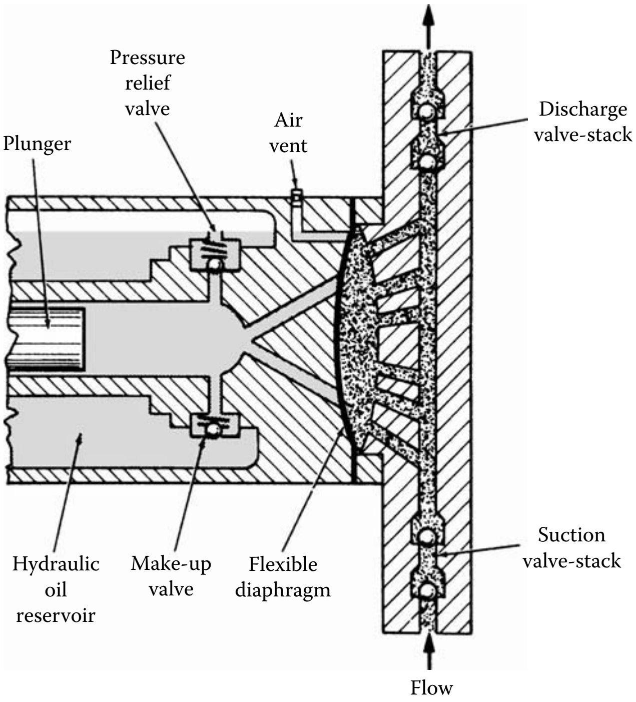


FIG. 7.4o
Diaphragm pump operated with hydraulic fluid.

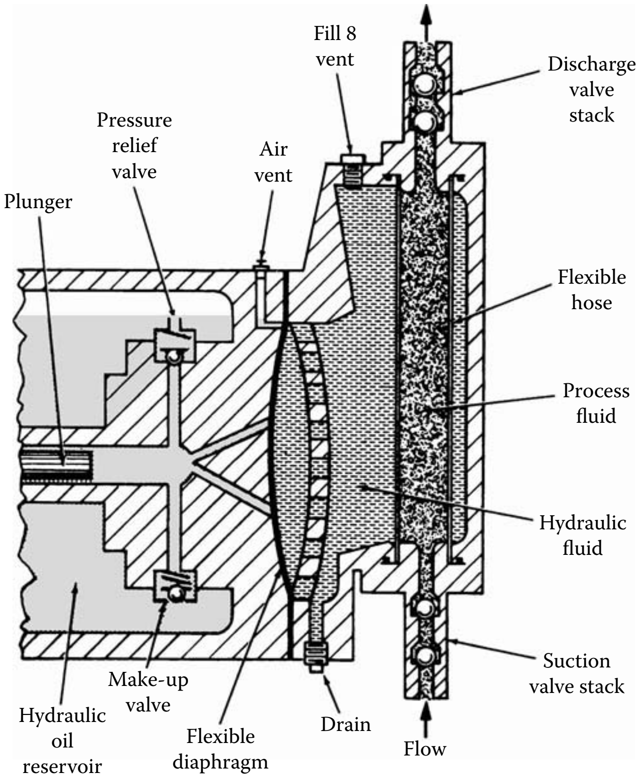


FIG. 7.4p
Diaphragm pump operated with hydraulic fluid and flexible hose element.

be reduced by using a hydraulic fluid to move the diaphragm, thereby eliminating the high differential pressures across it. This design consists basically of a plunger pump to provide hydraulic fluid pressure for diaphragm operation and the diaphragm pumping head (Figure 7.4o). The forces on the diaphragm are balanced, and discharge pressures comparable to plunger pumps are possible. The volume pumped per stroke is equal to the hydraulic fluid displaced by the plunger, and this volume is controlled by the stroke length adjustment as in the plunger pump.

A pump design using a flexible tube to achieve pumping action is shown in Figure 7.4p. Motion of the plunger displaces the diaphragm, which in turn causes the flexible tube to constrict, forcing fluid in the tube to discharge (similar to the operation of a peristaltic pump). This design is better suited for use on viscous and slurry liquids than the previously discussed types because the flow path is straight with few obstructions and no cavities; however, seating of the valves can still be a problem.

Pneumatic Metering Pumps

Pneumatically operated plunger-type (Figure 7.4q) and bellows-type metering pumps are also available for use when liquids in small quantities need to be injected at high pressure. The pneumatic timer is adjustable between 4 and 60 strokes per min, while the stroke length is also adjustable from 1/4 –1 in. (6–25 mm). When at the end of the stroke, the pressurized air-operated plunger has displaced the process fluid through the discharge check valve, the piston trips the vent valve, which resets the timer and allows the spring to return the piston to the starting position.

Pneumatic metering pumps are available in all stainless steel construction, require no lubrication, can be provided with plungers of 1/8 – 1/2 in. (3–12 mm) diameter, and can deliver flows from 0.1–60 gpd (0.4–225 lpd) at pressures up to 5000 PSIG (3.4 MPa). Until recently, natural gas was frequently used as the motive fluid for odorant injection pumps of this type. While this has the advantage that no other utilities are required, it releases methane to the atmosphere, and this is illegal in many places.

Installation Considerations

In order to ensure a properly working installation, a number of factors associated with the physical installation and with the properties of the fluid must be considered. Some factors that can contribute to a poor installation include:

1. Long inlet and outlet piping with many fittings and valves
2. Inlet pressure higher than outlet pressure
3. Pocketing of suction or discharge lines
4. Low suction head or suction lift

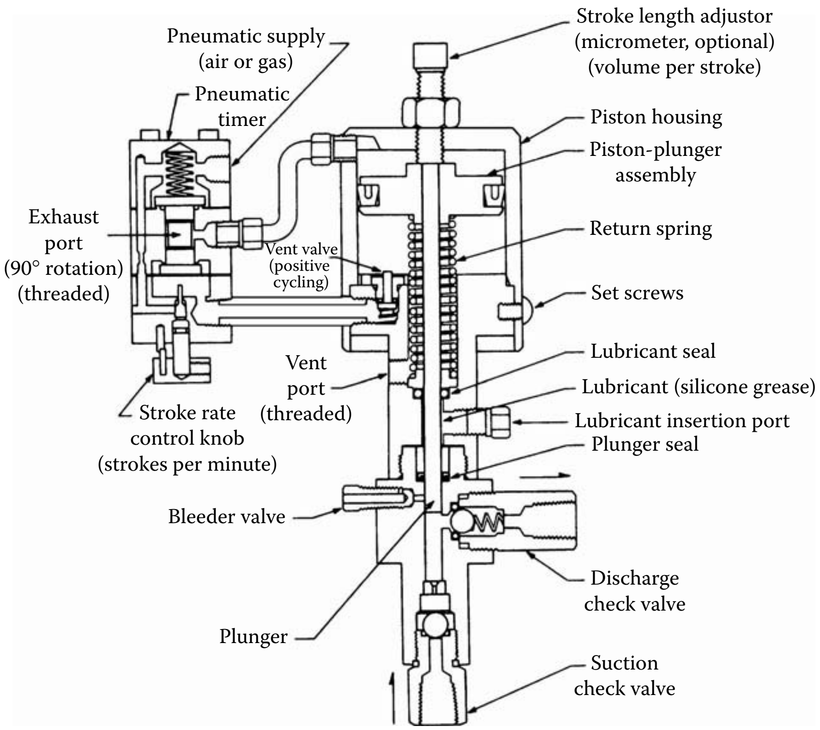


FIG. 7.4q

Pneumatically operated plunger-type metering pump. (Courtesy of Linc Mfg.)

A tortuous flow path in the pump suction or discharge can be troublesome when the fluid handled contains solids, is of high viscosity, or has a high vapor pressure, or if the suction head available is low. Generally, valves that offer a full flow path (such as ball valves) are preferred. Needle valves should be avoided. If the inlet pressure is higher than discharge, the fluid may flow unrestricted through the pump. Spring-loaded check valves at the pump are undesirable because the spring loading stops the ball check from rotating and from finding a new seating surface, for increased valve life. For such applications the installations shown in Figures 7.4r and 7.4s offer solutions.

In Figure 7.4r, the piping arrangement will supply the head to prevent through flow and siphoning. The height of the liquid should be varied depending on pump capacity and fluid velocity. This dimension varies between 2 and 10 ft (0.6 and 3 m) and increases with capacity and velocity. Figure 7.4s illustrates the use of a spring-loaded back-pressure valve to overcome the suction pressure. For this installation a volume chamber (gas-filled bladder) to dampen pulsations should be placed between the pump discharge and the valve.

Dissolved or entrained gases in the fluid can destroy metering accuracy, and in quite small volume they can stop pumping action entirely, as the gas volume is compressed before the fluid can exit through the discharge check valve. Figure 7.4t illustrates an installation design to vent entrained gases back to the fluid hold tank.

It is always desirable to locate the pump below and near the fluid hold tank. Under these conditions the fluid will flow by gravity into the pump suction and loss of prime is unlikely. If the pump cannot be located below the hold tank, other measures must be taken to prevent loss of prime.

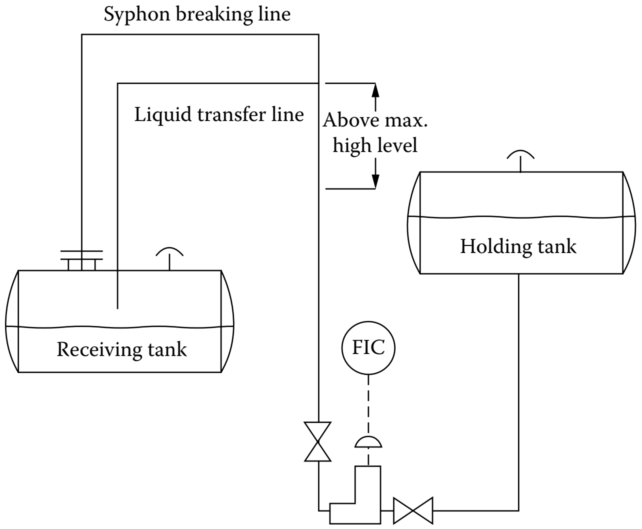


FIG. 7.4r

Piping arrangement to prevent through flow and siphoning.

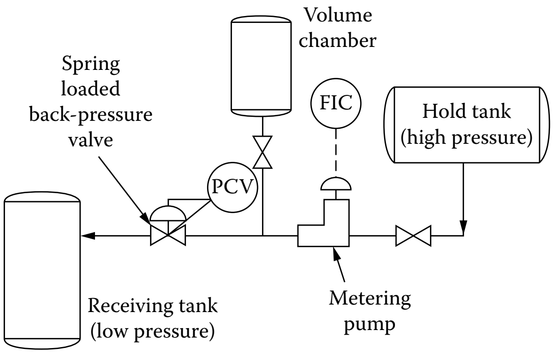


FIG. 7.4s
Metering pump with artificial head created by back-pressure valve.

NPSH and Pulsation Dampening

In order for the pump to operate properly, the net positive suction head must be above the minimum practical suction pressure of approximately 10 psia (69 kPa abs). The available net positive suction head is given by Equation 7.4(6).

$$\mathrm{NPSHA} = P - P_v \pm P_h - \sqrt{\left(\frac{lvGN}{525}\right)^2 + \left(\frac{lvC}{980Gd^2}\right)^2} \qquad \mathbf{7.4(6)}$$

where

P = feed tank pressure (psia)
P_v = liquid vapor pressure at pump inlet temperature (psia)
P_h = head of liquid above or below the pump centerline (psid)
l = actual length of suction pipe (ft)
v = liquid velocity (ft/s) at maximum piston speed (see Figure 7.4m)
G = Liquid specific gravity
N = number of pump strokes per minute
C = viscosity (centipoise)
d = inside diameter of pipe (in.)

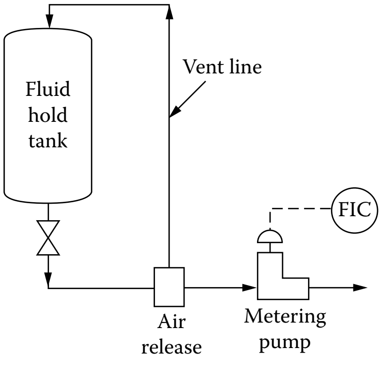


FIG. 7.4t
Elimination of entrained gases in metering pump installations.

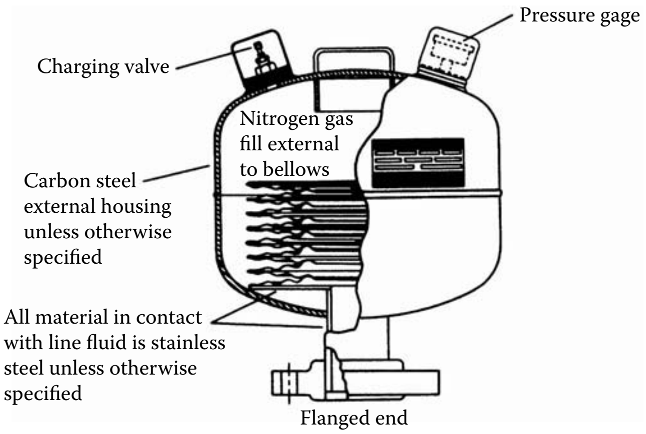


FIG. 7.4u
Pulsation dampener will suppress the pressure surges caused by positive displacement pumps. (Courtesy of the Meraflex Co.)

For liquids below approximately 50 centipoise (0.05 Pa·s), viscosity effects can be neglected, and Equation 7.4(6) reduces to

$$\mathrm{NPSHA} = P - P_v \pm P_h - \frac{(lvGN)}{(525)} \qquad \mathbf{7.4(7)}$$

The calculated value to NPSHA must be above the minimum suction pressure required by the selected pump.

In addition to multiple pumping heads, a pulsation dampener can be used on the pump discharge to smooth the discharge flow pulsations. The pulsation dampener is a pneumatically charged diaphragm chamber that stores energy on the pump discharge stroke and delivers energy on the suction stroke, thus helping to smooth the flow pulses. In order to be effective, however, the dampener volume must be equal to at least five times the volume displaced per stroke (Figure 7.4u), and the precharge pressure matched to the discharge pressure. Suction dampeners are also used to minimize the acceleration losses in the suction line if the NPSHA is close to the minimum, particularly for simplex and duplex designs; triplex designs are less sensitive. When handling flammable fluids, the piping code may call for the use of excess-flow valves in the suction line, to shut off the tank if the line ruptures. The use of these on a simplex or duplex reciprocating pump suction needs sizing for the maximum pump suction flow, not the average value.

OPPOSED CENTRIFUGAL PUMPS

The opposed centrifugal pump is not a control element but is an adaptation of a centrifugal pump to flow control. This method of control is particularly suitable for coarse, rapidly settling slurries at low flow rates. In such services the conflicting requirements of control at low flow and the need for a large free area to pass the solids may make it impossible

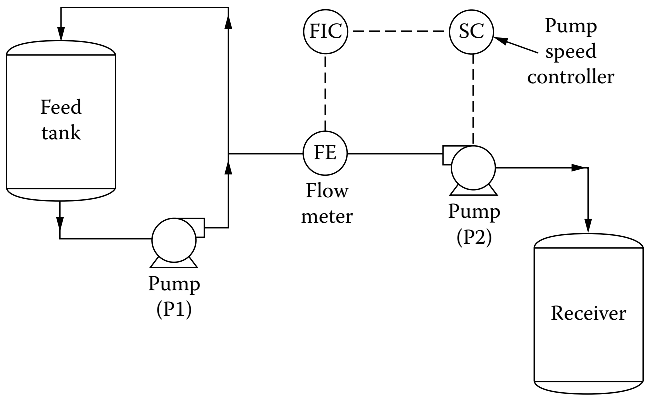


FIG. 7.4v
Opposed centrifugal pump as final control element.

to find a suitable control valve. A system that requires a small quantity of slurry to be fed to a receiving vessel under controlled conditions is depicted in Figure 7.4v. Pump P_1 continuously circulates the slurry from the feed tank at high velocity. A branch line from the discharge of P_1 is connected to the opposed centrifugal pump P_2. Pump P_2 is connected in opposition to the direction of slurry flow and provides a pressure drop to throttle flow. A variable-speed driver on pump P_2 throttles the pump pressure drop so as to keep the flow constant. At full speed the pressure difference across P_2 is sufficient to stop the branch line slurry flow completely. A magnetic flowmeter or some other suitable device can be used to measure the slurry flow. A VSD can be used to vary pump speed in response to the flow controller output signal.

A related approach utilizes a progressing-cavity pump as the restriction element, close-coupled to the circulating line and preferably discharging freely from the drive end. This does not need the flowmeter, as it is effectively positive displacement at low head, and flow is proportional to speed.

Bibliography

Adams, R. M., "Don't Overlook Simple Regulators for Complex Fluid Control Jobs," *InTech,* April 1983.

Baumann, H. D., "A Case for Butterfly Valves in Throttling Applications," *Instrument and Control Systems,* May 1979.

Baumann, H. D., "Control Valves vs. Speed-Controlled Pumps," Texas A&M Symposium, 1981.

Baumann, H. D., "How to Assign Pressure Drop Across Liquid Control Valves," *Proceedings of the 29th Annual Symposium on Instrument Engineering for the Process Industry,* Texas A&M University, January 1974.

Bower, J. R., "The Economics of Variable-Speed Pumping with Speed-Changing Devices," *Pump World.*

Carlson, R., "The Correct Method of Calculating Energy Savings to Justify Adjustable Frequency Drives on Pumps," Paper PCIC-99-26, IEEE Petroleum and Chemical Industry Conference, 1999.

Coughlin, J. L., "Control Valves and Pumps: Partners in Control," *Instruments and Control Systems,* January 1983.

Driedger, W., "Controlling Centrifugal Pumps," *Hydrocarbon Processing,* July 1995.

Fischer, A. K., "Using Pumps for Flow Control," *Instruments and Control Systems,* March 1983.

Gibson, I. H., "Variable-Speed Centrifugal Pump Control Guidelines," *Proceedings 6th World Congress of Chemical Engineering,* 2001.

Gottliebson, M., "Explore the Use of Variable-Speed Water Booster Pumps," *Water and Wastes Engineering,* May 1978.

Greene, A. M., *Pumping Machinery*, New York: Wiley, 1919.

Hall, J., "Motor Drives Keep Pace with Changing Technology," *Instruments and Control Systems,* September 1982.

Hall, J., "Pump Primer," *Instruments and Control Systems,* April 1977.

Holland, F. A., "Centrifugal Pumps," *Chemical Engineering,* July 4, 1966.

Holloway, F. M., "What to Look for in Nutering Pumps," *Instruments and Control Systems,* April 1976.

Hydraulic Institute, Hydraulic Institute Standards (www.pumps.org).

Hydraulic Institute, "Pump Life Cycle Costs: A Guide to LCC Analysis," 2005 (www.pumps.org).

Hydraulic Institute, ''Variable Speed Pumping: A Guide to Successful Applications," 2005 (www.pumps.org).

Irvine, G., and Gibson, I. H., "The Use of Variable Frequency Drives as a Final Control Element in the Petroleum Industry," 2000 IEEE Industry Applications Society Annual Meeting, World Conference on Industrial Applications of Electrical Energy, Rome, October 2000.

ISA Training CD, Centrifugal Pumps Series (5 parts) CSTL24, Research Triangle Park, NC: ISA.

ISA Training CD, Pumps and Actuators CSTL16C, Research Triangle Park, NC: ISA.

Janki, C., "What's New in Motors and Motor Controls?" *Instruments and Control Systems,* November 1979.

Karassik, I. J., Krutzsch, W. C., Fraser, W. H., and Messina, J.P., *Pump Handbook*, New York: McGraw-Hill, 1986.

Lipták, B. G., "Costs and Benefits of Boiler, Chiller, and Pump Optimization," in *Instrumentation in the Chemical and Petroleum Industries,* Vol. 16, Research Triangle Park, NC: ISA, 1980.

Lipták, B. G., "Save Energy by Optimizing Boilers, Chillers, Pumps," *InTech,* March 1981.

Liu, T., "Controlling Pipeline Pumps for Energy Efficiency," *InTech,* June 1979.

Merritt, R., "Energy Saving Devices for AC Motors," *Instruments and Control Systems,* March 1980.

Merritt, R., "What's Happening with Pumps," *Instruments and Control Systems,* September 1980.

Mulley, R., *Flow of Industrial Fluids–Theory and Equations,* Boca Raton, FL: CRC Press, 2004.

Papez, J. S., and Allis, L., "Consideration in the Application of Variable Frequency Drive for Pipelines," ASME Paper 80-PET-78.

Poyton, *Metering Pumps,* Marcel Dekker.

Rishel, J. B., "The Case for Variable-Speed Pumping Systems," *Plant Engineering,* November 1974.

Rishel, J. B., "Matching Pumps to System Requirements," *Plant Engineering,* October 1975.

Rishel, J. B., "Water System Head Analysis," *Plant Engineering,* October 1977.

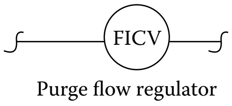
Schroeder, E. G., "Choose Variable Speed Devices for Pump and Fan Efficiency," *InTech,* September 1980.

Stewart, R. F., "Applying Adjustable Speed AC Drives," *Instruments and Control Systems,* July 1981.

Thurlow, C., "Pumps and the Chemical Plant," *Chemical Engineering,* May 24, 1965.

7.5 Regulators—Flow

G. F. ERK (1985, 1994) **B. G. LIPTÁK** (1995) **T. J. BAAN** (2005)



Purge flow regulator

Flow sheet symbol

Types:	A. Rotameters, purge flowmeters (variable area) B. Chromatographic flow controller C. Variable, spring-loaded, and flexible orifice flow regulator D. Thermal mass flow controller
Design Pressures:	A. Up to 600 PSIG (4.1 MPa) B. 500 PSIG (3.45 MPa) C. Industrial flow regulators are available for maximum working pressures up to 3600 PSIG (2.07 MPa) at 100°F (38°C). Oil flow controllers are available for up to 3750 PSIG (26 MPa) D. From 20 PSIG (138 kPa) minimum to 1000 PSIG (6.9 MPa) maximum
Design Temperatures:	A. Standard units for 200°F (93°C) B. Barstock units, brass 180°F (82°C), stainless steel 250°F (121°C) C. Standard units up to 400°F (204°C) D. Standard units from 41 to 149°F (5 to 65°C); special up to 302°F (150°C)
Materials of Construction:	A. Glass, aluminum, brass, carbon steel, stainless steel, or PTFE-PFA B. Brass or 316 stainless steel with a choice of elastomeric O-rings C. Typically brass, copper, ductile iron, or steel bodies with stainless steel internals D. 316 SS, corrosion-protected steel, nickel, Inconel, elastomer seals
Flow Rates:	A. For liquids from 0.01 cc/min and for gases from 0.5 cc/min and up. A standard 1/4 in. (6 mm) unit can have a 0.05–0.5 gpm (0.2–2 lpm) range on liquid or a 0.2–2 SCFM (0.3–3 m^3/hr) range on gas service. Off-the-shelf 2 in. meters for liquids 0.25–120 gpm (35–440 lpm), range for gas service 1.2 SCFM (2.1 m^3/hr) to 250 SCFM (420 m^3/hr) B. Chromatographic applications in the cc range C. Water regulators for standard HVAC applications are available with settings from 0.5 to 150 gpm (2 to 600 lpm), while oil flow regulators are available with ranges from 0–5 to 0–800 gpm (0–20 to 0–3200 lpm). Industrial units cover liquid ranges from 0.02 to 550 gpm (0.1 to 2000 lpm) and gas ranges from 0.2 to 6000 SCFM (6 slpm to 170 m^3/m) D. Full scale for gases from 10 sccm to 1000 slm (35 SCFM). Special units up to 1900 slm (500 SCFM)
Sizes:	A. Typically from 1/2 to 2 in. (3 to 50.8 mm) connection size B. 1/2 in. FNPT C. 0.5 in. to 6 in. (12 to 150 mm) D. 1/8 in. (3.2 mm) to 3/4 in (19 mm) compression fittings, 3/4 in. FNPT
Inaccuracy:	A,B. 2–5% of full scale C. 1.5% of full scale for industrial designs, 5% of full scale for HVAC designs D. 1% of full scale
Costs:	A. A typical 65 mm scale glass tube meter with a brass or aluminum built-in needle valve costs about $80. An all-stainless steel 150 mm scale rotameter with a 16-turn precision valve with 1/4 in. (6 mm) connections costs about $350. 65 mm and 150 mm purge flow regulators with built-in differential pressure controllers cost $600 and $650, respectively. B. Brass barstock $100, stainless steel barstock $150

C. HVAC units in smaller sizes and larger quantities are under $100
D. A standard thermal mass flow controller for gases costs $1100

Partial List of Suppliers:

Aalborg Instruments & Controls Inc. (www.aalborg.com) (A, D)
ABB Fischer & Porter (www.abb.com) (A)
Advance-Tech Controls Pvt. Ltd. (www.advancetechindia.com) (D)
Analyt-MTC Gmbh (www.analyt-mtc.de) (A, D)
ARO Corp. (www. aro.ingersoll-rand.com) (C)
Blue White Industries (www.bluwhite.com) (A)
Brooks Instrument (www.emersonprocess.com/brooks) (A,D)
Cole-Parmer Instrument Co. (www.coleparmer.com) (A, D)
Eaton Corp. (web.eaton.com) (C)
Fisher Scientific (www.fishersci.com) (A)
Flowmetrics Inc. (www.flowmetrics.com) (A)
Griswold Controls (www.griswoldcontrols.com) (C)
Hays Fluid Controls Inc. (www.haysfluidcontrols.com) (C)
W.A. Kates Co. (www.wakates.com) (C)
Ketema Inc., Schutte and Koerting Div. (A)
Key Instruments (www.keyinstruments.com) (A)
King Instrument Co. (www.kinginstrumentco.com) (A)
Krohne America Inc. (www.krohneamerica.com) (A,D)
Matheson Tri-Gas Gas (www.matheson-trigas. com) (A)
MKS Instruments Inc. (www.mksinst.com) (D)
Mott Metallurgical Corp. (www.mottcorp.com) (B)
Mykrolis Corp. (formerly Tylan General) (www.mykrolis.com) (D)
Neoperl Inc. (www.neoperl.com) (C)
Numatics Inc. (www.numatics.com) (B)
Omega Engineering Inc. (www.omega.com) (A)
Pierburg Instruments (www.pierburginstruments.com) (B)
Porter Instrument Co. Inc. Pty. Ltd. (www.porterinstrument.com) (A, D)
Pryde Measurement (www.pryde.com.au) (A, D)
Scott Specialty Gases (www.scottgas.com) (A)
Sierra Instruments (www.sierrinstruments.com) (D)
Taylor Valve Technology (www.taylorvalve.com) (C)
Wallace & Tiernan Inc. (www.usfwt.com) (A)

INTRODUCTION

For the general subject of flow measurement, please refer to Chapter 2 of the first volume of this handbook. In this section a number of self-contained flow regulators/controllers are described. Their common feature is that the flow sensor, the flow controller, and the flow control valve (in other words, the whole control loop) are combined into a single unit.

The design variations covered in this section include the purge flow regulators, which detect the flow by using a rotameter (variable area flowmeter), the various flow regulator valves used in chromatography, the various adjustable-orifice flow regulators that reduce their internal opening to flow when their inlet pressure rises and thereby keep the flow relatively constant, and the thermal mass flow controllers used as accurate regulators of gas flows.

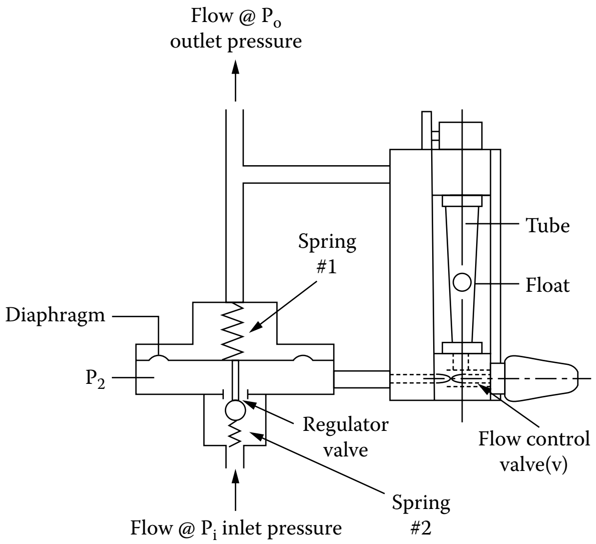


FIG. 7.5a
The main components of the illustrated purge flow regulator are a glass tube rotameter, an inlet needle valve, and a differential pressure regulator. (Courtesy of Krone Inc.)

PURGE FLOW REGULATORS

When a variable area flowmeter is combined with a differential pressure regulator (Figure 7.5a), it becomes a self-contained flow controller. The purge flow is fixed by adjusting springs #1 and #2 for a particular pressure difference, usually in the range of about 60–80 in. (150–200 cm) of water. This constant pressure drop ($P_2 - P_o$) is then maintained across the flow control valve (V). The configuration in Figure 7.5a maintains the outlet pressure (P_o) constant by compensating any variation in the inlet pressure P_i by changing the regulator valve opening.

Other purge flowmeter designs are also available that work in a reverse configuration by keeping the inlet pressure P_i constant and allowing the outlet P_o to vary. In these designs the constant pressure drop across the valve (V) is maintained to equal ($P_i - P_2$) instead of ($P_2 - P_o$) being kept constant. The gas flows through purge flow controllers are usually adjustable in a range of 0.2–2 SCFH (6–60 slph). The error or inaccuracy is usually 5% of full scale over a range of 10:1. The standard pressure and temperature ratings are 150–300 PSIG (1–2 MPa), and 212–572°F (100–300°C).

Figure 7.5b illustrates how a needle valve can be placed at the inlet or the outlet of a variable area flowmeter and the pressure difference controlled across this assembly. The main applications of purge flow regulators are to introduce small gas or liquid purge streams into pressure taps, where for reasons of corrosion or plugging protection it is desirable to do so. As the interest in these applications is only to make sure that there is some flow, the typical ±5% of full-scale error in these units is acceptable.

FLOW REGULATORS FOR CHROMATOGRAPHS

Another self-contained flow regulator valve design is shown in Figure 7.5c. It is designed to accurately adjust and maintain small gas and liquid flows. This design utilizes a nonrising stem and features a positive, direct mechanical means of adjusting a sliding tapered needle that virtually prevents sticking caused by foreign matter in the fluid stream. These barstock valves are available in straight line or 90° configurations, and are particularly suitable for the precise control requirements in chromatography.

VARIABLE-ORIFICE FLOW REGULATORS

These are mature devices that have not changed much during the last decades. One of their simplest versions is a thick rubber orifice (Figure 7.5d). It has such a geometry that it is deflected as the pressure drop increases in such a way that with a rising pressure drop, the opening to the flow is reduced.

Such flexible orifices, which inversely vary their openings with the pressure of the water supply, are used in HVAC distribution and other water balancing applications, such as in showers, appliances, filters or sprinklers. These units can be mounted inside 1/2 and 3/4 in. fittings. Different orifices are available for approximately maintaining crude flows at 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 9, and 10 gpm, while the inlet pressure varies between 15 and 125 PSIG (103 and 870 kPa). The cost of these fittings is a few dollars.

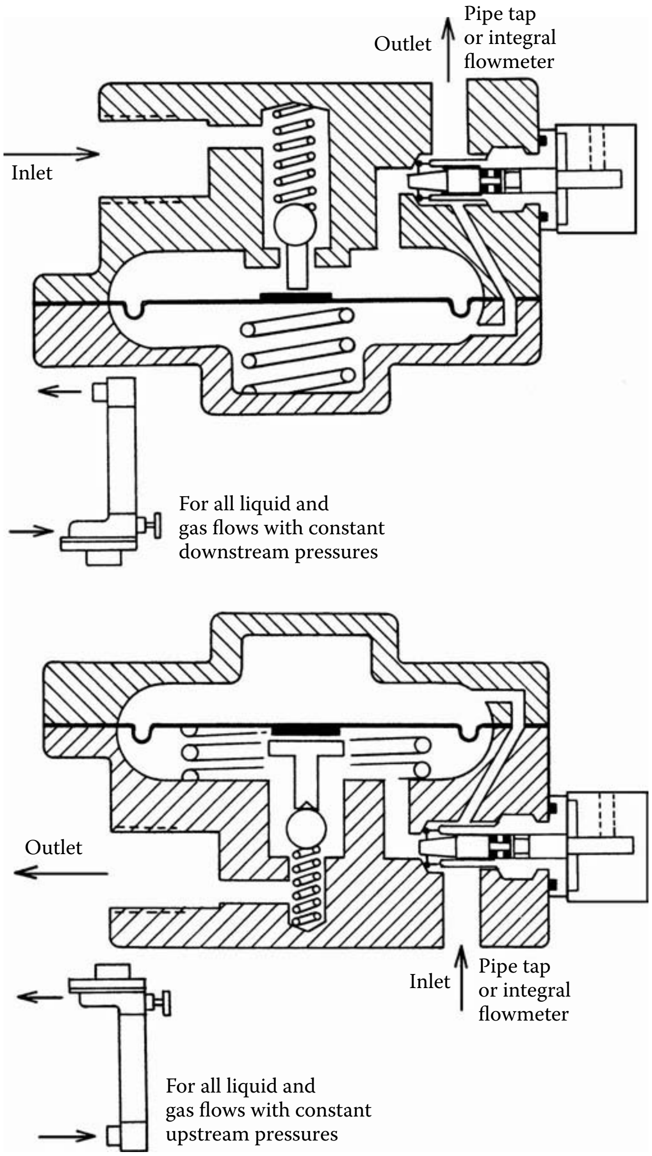


FIG. 7.5b
Purge flow rate can be regulated by maintaining the differential pressure across a variable area flowmeter constant. The d/p regulator is located at the inlet of the rotameter (top) if the downstream pressure is constant, and at the outlet (bottom) if the upstream pressure is constant.

HVAC Balancing Flow Regulators

The purpose of the flow regulator shown in Figure 7.5e is to balance the water flow to the various heating/cooling coils in large HVAC systems, where the water pressure varies on the different floors of a building. In order to keep the flow constant while the inlet pressure varies, it is necessary to vary the orifice opening as a function of pressure.

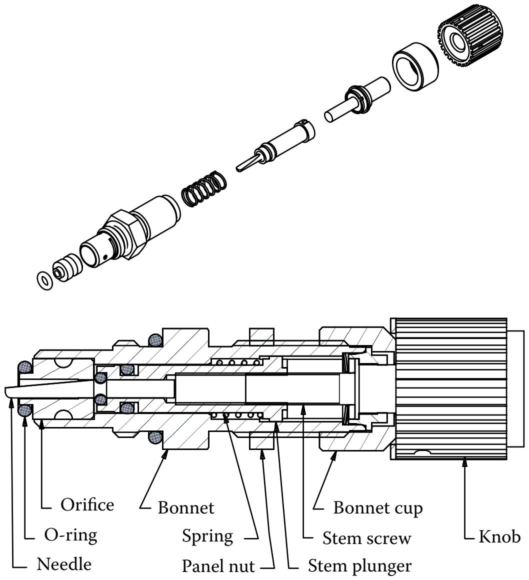


FIG. 7.5c
Nonrising stem-type valve cartridge. (Courtesy of AALBORG Instruments.)

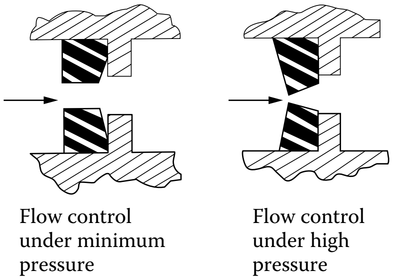


FIG. 7.5d
Flexible-orifice-type flow regulator, which is used to balance the flow of shower and sprinkler heads or in HVAC applications.

The inner flow element of the regulator illustrated in Figure 7.5e is pushed against a spring by the flowing water. When the water pressure is low (minimum pressure drop), the spring forces the element out and exposes the maximum number of flow openings to the water. As water pressure rises, the element is pushed in against the spring and therefore fewer orifices are available for the water to pass through. This increases the pressure drop across the regulator, while keeping the flow rate relatively constant.

Units are available in sizes up to 3 in. (75 mm) and with flow settings up to 200 gpm (600 lpm). The spring should be selected to match the probable pressure drop variation in the system. Springs are available with 14, 32, 57, and 128 psid (96, 220, 393, and 883 kPa) ranges.

Another flexing action flow regulator consists of an automatic control valve with an elastic diaphragm-orifice combination. This design permits passage of small particles, sludge, thread chips, and rust impediments and minimizes the need

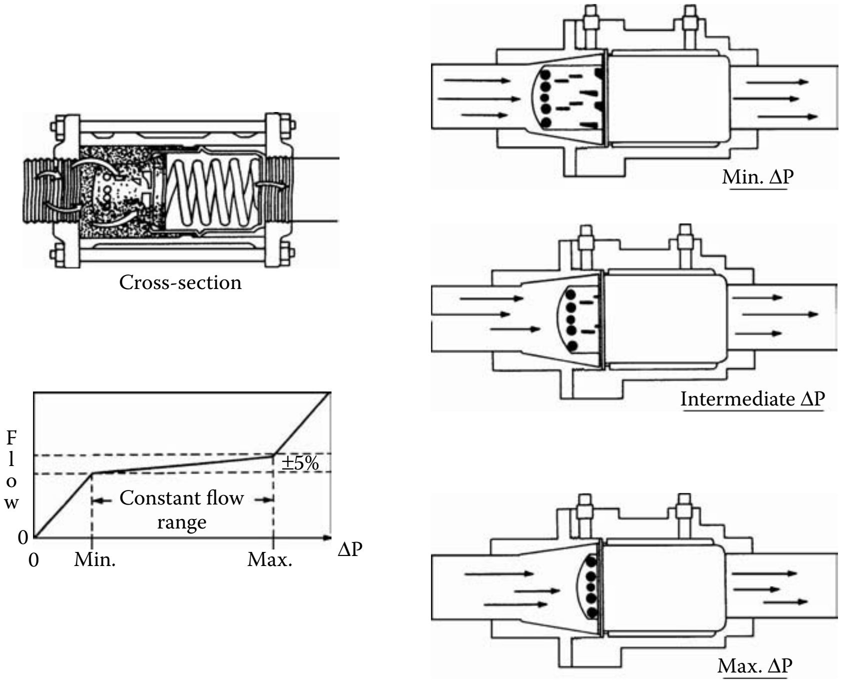


FIG. 7.5e
As inlet pressure rises, the orifice element is pushed in against the spring, thereby exposing fewer openings to the flow and as a result increasing the pressure drop so as to maintain the flow relatively constant. (Courtesy of Griswold Controls.)

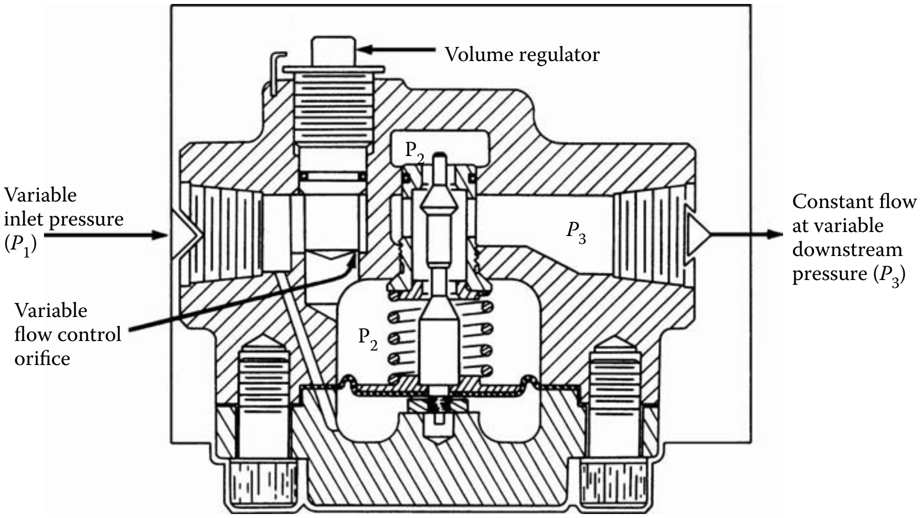


FIG. 7.5f

Oil flow regulator keeps the pressure difference ($P_1 - P_2$) constant across the flow control orifice and thereby allows for the maintenance of constant flow even when the inlet or outlet pressures vary. (Courtesy of Taylor Valve Technology, Inc., formerly Taylor Tools.)

for strainers. The operation is quiet and the flow is maintained within a 5–10% error.

Oil Flow Regulator

Figure 7.5f depicts a flow controller, which is used to regulate the flow of high-pressure oil. In this design, flow is maintained constant by keeping the pressure drop across a manually adjustable internal orifice constant. This way either the upstream or downstream pressures can vary without affecting the flow.

The internally controlled differential pressure ($P_1 - P_2$) can be as low as 15 psid (103 kPa) and the flow will still be maintained approximately constant when the overall differential ($P_1 - P_3$) is several thousand psid. These units are available in sizes from 0.5–4 in. (12–102 mm) and can control flow ranges from 0–5 to 0–400 gpm (0–20 to 0–1514 lpm). These units have been designed for high-pressure operation and are suited for working pressures up to 3750 PSIG (26 MPa).

Industrial Flow Regulators

Self-contained flow-rate controllers usually combine an adjustable orifice and an automatic internal regulating valve in a single body. Differential pressure across the adjustable orifice is automatically maintained constant, regardless of the orifice opening. Therefore, the total flow rate is directly proportional to the area of the orifice.

The orifice and the automatic differential pressure regulator are usually enclosed in a single housing with a calibrated flow-rate indicating dial. A regulator that is provided with an external needle valve with the differential pressure controller is also available. This provides some flexibility in adjusting flow rates remotely.

A flow-rate controller of this type is shown in Figure 7.5g. Here, the metering orifice is formed by an arcuate slot in a cylinder (orifice sleeve) and a second slot is provided in an inner (orifice) cylinder around which the first (orifice sleeve) can be rotated. When the two slots match or coincide, the orifice has its maximum open area; when the outer sleeve is rotated, the open length of the slot and the open area decrease in proportion to the angle rotation of the sleeve. When the outer sleeve is rotated to 160°, the open orifice area is reduced to 0.

This automatic differential pressure regulator includes an impeller, spring, valve sleeve, and valve tube. The impeller disc reciprocates in the upper part of the orifice sleeve, driven by the balance between the forces due to the differential pressure across the orifice ($P_U - P_D$) and the spring. The impeller is rigidly connected to the valve sleeve. It closes or opens the valve ports as dictated by the force balance, maintaining the orifice differential pressure and holding the orifice flow at the valve determined by the differential pressure and the orifice area.

Advantages and Limitations Advantages of the self-contained flow-rate controllers are as follows:

1. Provides reasonably accurate flow-rate control by means of a self-operated single unit.
2. Does not require straight pipe runs upstream or downstream of the unit.
3. No utility requirements to operate the unit (no electric power or pneumatic or hydraulic fluids).
4. Maximum speed of response, because control occurs at the point where the measurement is made.

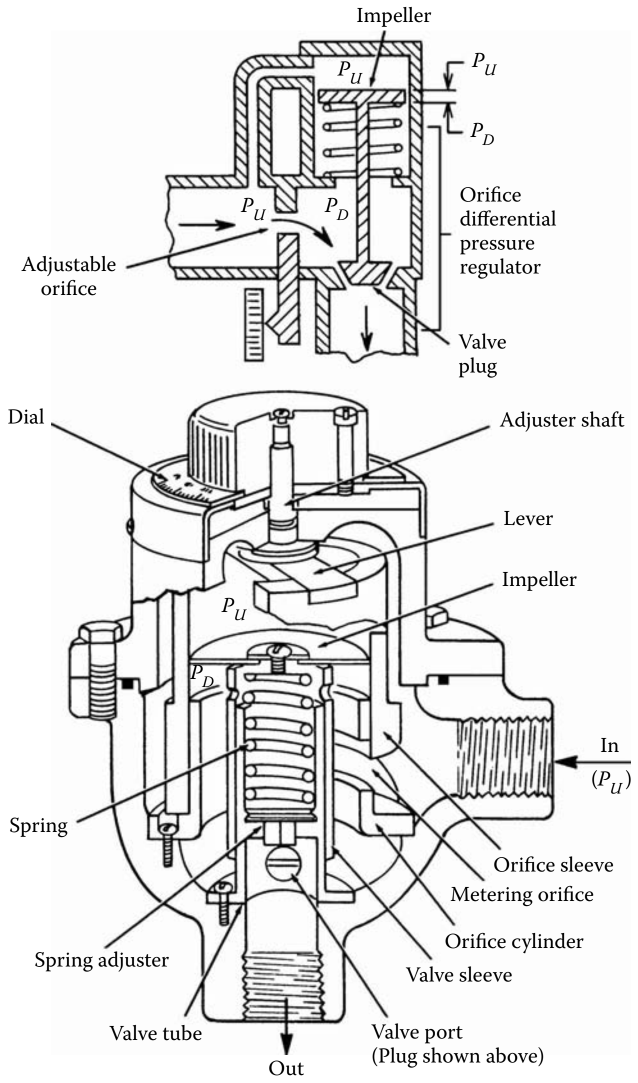


FIG. 7.5g
Industrial flow regulator.

5. Flow scale is linear and, therefore, adjustment rangeability is better than with square root scales.
6. Simple design and few parts with little calibration and adjustment requirements.
7. Low oscillation or hunting phenomena in liquid service.
8. Turndown ratio can be up to 25:1 or better.
9. Large pressure drops have minimal effects.
10. On clean water services, both the first and maintenance costs are low.

Disadvantages include that the regulator requires a fairly high minimum pressure drop to work properly and that it can only be used on clean fluids.

Typical applications include the controlling of the flow of sealant fluids to the rotating seals on centrifugal pumps, compressors, turbines, and other rotating machinery. Another application is in pressure filtration, where a self-contained flow-rate controller will keep the filter effluent flow rate constant as the pressure drop across the filter increases with increasing filter cake build-up. Other applications can involve blending, batch control, and hydraulic oil flow control, where constant flow rates to various components must be maintained.

Gas Flow Applications For accurate flow-rate control in gas flow applications, the variations in the upstream pressure and temperature should be minimized. To determine the capacities of a particular regulator, manufacturers provide gas flow equivalent graphs. These graphs show values in gpm of water. All capacity data are based on water for purposes of standardization and convenience.

Water is normally used as the test fluid for factory calibrations. Graphs are also available for air, hydrogen, natural gas, nitrogen, and oxygen. For other gases, a factor (C) determined from its specific gravity is applied to the capacity data given for air. Hence, at a given pressure and temperature, the gas flow = (air flow) × C, where

$$C = \frac{1}{\sqrt{\text{S.G. of gas}}} \tag{7.5(1)}$$

The error in regulating the flow is usually less than 1.5% of set point. In some applications it is desirable to set the regulated flow rate remotely. Pneumatic and electric actuators are available for this purpose.

THERMAL MASS FLOWMETERS

For the regulation of small gas flows, complete mass flow control packages using thermal flow detectors are available. Thermal flowmeters are discussed in detail in Section 2.13 of the *Process Measurement* volume of this set of handbooks, and therefore are only briefly discussed here.

In order to make the heat-transfer-type flowmeter also suitable for the measurement of larger flow rates, the bypass design illustrated in Figure 7.5h has been introduced. The flow sensor in the bypass consists of two self-heated resistance thermometers wound around a thin-walled tube.

When there is no flow, the two heaters raise the tube temperature by about 70°C above ambient in a symmetrical manner (solid line in the upper part of Figure 7.5h) and there is no temperature difference between the two sensors. As the gas starts to flow, the molecules will carry the heat downstream (dotted line on Figure 7.5h) and the two sensors will register a temperature difference, which is directly proportional with the mass flow of the gas.

The thermal flowmeter tubes in these bypass units are small capillary tubes, usually under 0.125 in. (3 mm) diameter. They ensure laminar flow over the full operating range

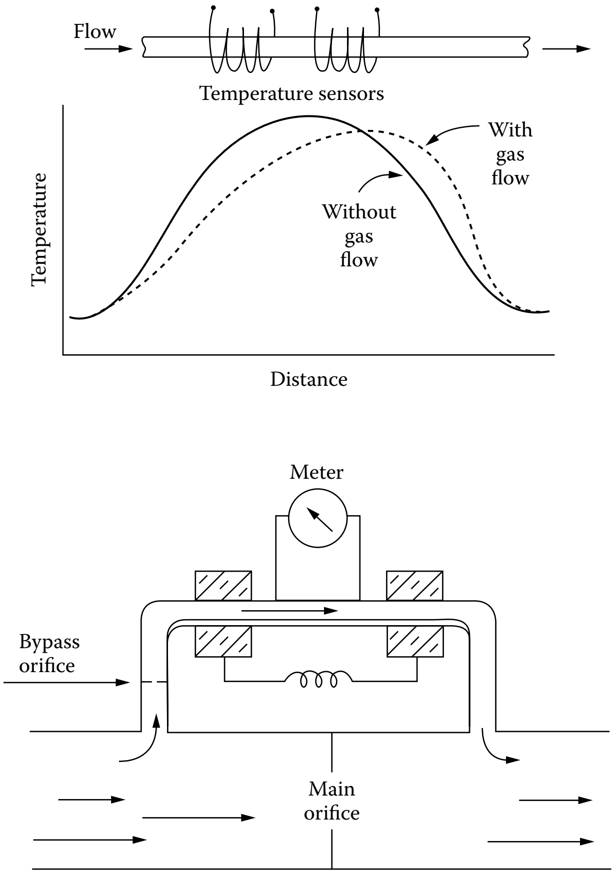


FIG. 7.5h
The operation of a bypass-type thermal flowmeter. (Courtesy of Mykrolis Corporation, formerly Tylan General Corp.)

of the meter. Their small size is advantageous in minimizing the electric power requirement and also in increasing their speed of response but necessitates the use of upstream filters to protect them against plugging. Some of the units also require up to 45 PSIG (3 bars) pressure drop in order to develop the laminar flow conditions at which they operate.

These units are also available as complete control loops, with sensor, controller, and automatic control valve and LCD readouts, all assembled into a single unit (Figure 7.5i). The cost of these units is competitive, and therefore if it is sufficient to control the flow of small gas streams within an error of ±1% of full scale, these units are a good selection.

These mass flow controllers are usually equipped with electromagnetic, fast response control valves, but slow response thermal-expansion-type control valves can also be used, if the fast response would result in over- and undershooting of the set point. Thermal-expansion-type valves are operated without seals and therefore without friction or wear. In this design, the controller heats the stem of a thermal expansion valve and the expanding stem moves the valve plug toward the seat, thereby closing the valve. Because of this nature of the valve actuator, it takes 5–30 sec to bring the flow to set point.

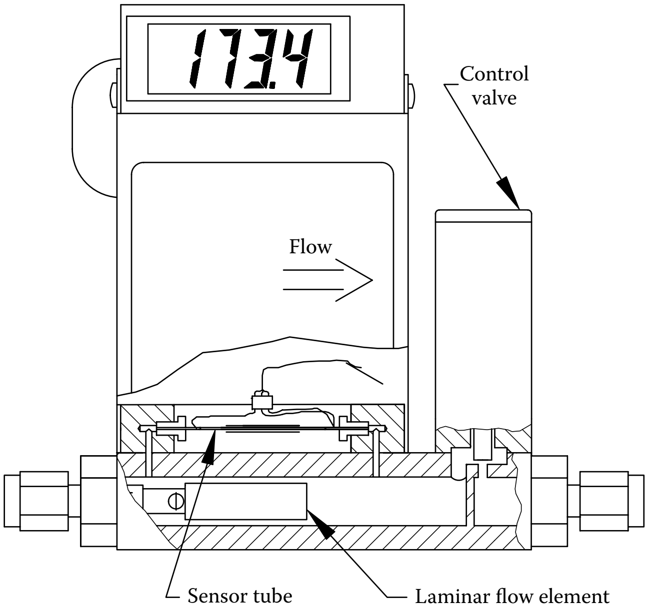


FIG. 7.5i
Bypass-type thermal mass flow controller with built-in readout. (Courtesy of AALBORG Instruments.)

In the microprocessor-driven mass flow controllers, firmware enables these digital devices to perform an array of functions that are unavailable with analog controllers. Intelligent mass flow controllers incorporate automatic zeroing, totalizers with set, start/stop, read, start at preset flow, and stop at preset volume capabilities.

Typically, up to ten primary gas flow calibrations may be stored in their memory. In addition, 256 gas coefficients are provided to convert between calibrating and other gases. High/low flow alarms, up to ten user-programmable flow ramping patterns, self-diagnostics, choice of engineering units, PID auto-tune, and dry contact closures are now standard in digital devices.

"Smart" mass flow controllers typically have one of the following digital interfaces: RS-232, RS-485, Device Net, and PROFIBUS. Because of the requirements of the semiconductor industry, thermal mass flow controllers are now optionally available with clean room assembly and all metallic O-ring features. Another new development is the increasing use of IP65 (NEMA4/12) water- and dust-resistant enclosures.

Bibliography

Adams, R. M., and DeChiara, W. A., "Don't Overlook Simple Regulators for Complex Fluid Control Jobs," *InTech,* April 1983.

Baker, W. T., "Flowmeters Using Temperature," *Measurements and Control,* September 1991.

Furness, R., "Flowmeter Terminology and Practical Operating Limits for Flowmeters," *Flow Control Network,* September 2001.

Gray, D. E., et al., "Mass Flow Controller Real-Time Process Gas Delivery," *Microelectronics Manufacturing Technology,* August 1991.

Hayward, A. J., "Choose the Flowmeter Right for the Job," *Processing Journal,* 1980.

Kates, W. A., "Direct-Acting Flow Rate Regulators," Paper #86-NY60, presented at ISA Conference, New York, 1960.

Kates, W. A., "Self-Actuated Flow Rate Controls," Paper #69-536, presented at the ISA Conference, Houston, TX, 1969.

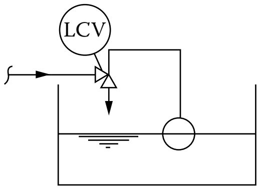
Laskaris, E. K., "The Measurement of Flow," *Automation,* 1980.

Sheriff, D., "Mass Flow Control," *Measurements and Control,* September 1990.

Smith, P., and Zappe, R. W., "Valve Selection Handbook: Engineering Fundamentals for Selecting the Right Valve Design for Every Industrial Flow Application," *Gulf Publishing,* December 12, 2003.

Welch, J. V., "Trends in Low Gas Flow Metering," *InTech,* February 1991.

7.6 Regulators—Level



Flow sheet symbol

B. G. LIPTÁK (1969) **G. F. ERK** (1985, 1995)
M. D. CHAPMAN, B. G. LIPTÁK (2005)

Pressure in Vessel:	Up to 500 PSIG (3.5 MPa) with float type; atmospheric for diverter and altitude valves
Maximum Pressure Drop through Valve:	100 psid (0.7 dMPa) for float type; 50 psid (0.35 MPad) for diverter or altitude valves
Design Temperature:	Up to 450°F (232°C) with float type; ambient for altitude valves; up to material selected on diverter valves
Materials of Construction:	Iron, steel, stainless steel for float type; unlimited for diverter valves; cast iron for altitude valves
Range:	Normally less than 12 in. (300 mm) for float type; diverter valves essentially on/off; altitude valves practically on/off (some regulate)
Valve Sizes:	1/2 to 42 in. (12.5 to 1050 mm) for float and diverter type; 2 to 42 in. (50 to 1050 mm) for altitude valves
Cost:	Float-type control valves are basically globe valves in which the spring-and-diaphragm actuator has been replaced by a float operator. Therefore, their costs are similar to the costs of globe valves, which for carbon and stainless steel valves are given in Table 6.19b. In cast iron, a 3/4 in. (18 mm) valve with a float costs about $300, while a 30 in. (750 mm) altitude valve with multiple control functions can cost over $50,000
Partial List of Suppliers:	CLA-VAL Co. (www.cla-val.com) Emerson Process Measurement (www.emersonprocess.com/fisher) Masoneilan Div. of Dresser (www.masoneilan.com) Ross Valve Manufacturing Co. Inc. (www.rossvalve.com) Siemens Energy & Automation Inc. (www. sea.siemens.com) Yokota Manufacturing Co. (www.aquadevice.com)

INTRODUCTION

Regulators are self-contained mechanical devices, while controllers require an external (pneumatic, hydraulic, or electric) energy source. Float-type level regulators have been around for a long time and have not changed much in the last couple of decades. The use of these devices is limited to clean liquid services and to applications where the offset that is inherent in the operation of any proportional-only regulator can be tolerated.

In the designs that are described in this section, the energy source for the regulator's operation is either the static and kinetic energy of the flowing stream or the buoyant force generated by the fluid inside the process vessel.

Three basic level regulator designs will be described: the float level control valves, the solid-state diverter valves, and the altitude valves.

FLOAT-TYPE LEVEL REGULATORS

In the float-type level control valves, mechanical linkages transmit the movement of the level float to the final control element (control valve) and thereby serve to regulate the flow to or from the vessel in such a manner that the level is maintained relatively constant. In order to evaluate the advantages and disadvantages of such a device one needs to consider the following:

1. Offset due to load changes, hysteresis, dead band, and span
2. Float power limitations
3. Stuffing box arrangements on pressurized vessels
4. Installation requirements

The Phenomenon of Offset

When process load (flow) changes, a plain proportional controller is not capable of keeping the controlled variable (level) on set point, because it first needs an error to develop, before it can change the valve opening. Therefore, as illustrated in Figure 7.6a, when the load changes, a permanent offset (deviation from set point) will occur.

The proportional band or gain of the control loop shown in Figure 7.6a can be defined as the level change that will result in fully stroking the inlet valve. As shown, the valve stroke from open to close is 1 in. (25 mm) and the level change required to stroke the valve is 7 in. (175 mm). The ratio *A*/*B* is the gain of the regulator. If $A = B$, the gain is 1 or the proportional band is 100%. If a small change in level results in a large change in valve opening ($B/A < 1$), the regulator has a high gain (narrow proportional band) and if $B/A > 1$ (the gain is less than 1), the proportional band is over 100%.

Proportional sensitivity (K_c) can be defined as the percentage change in flow rate through the valve that will result from a 1% change in liquid level in the tank. In the illustrated example, a 7 in. (175 mm) change in level will move the valve from fully closed (no flow) to fully open (maximum flow of 10 gpm or 40 lpm). Therefore, the proportional sensitivity is:

$$K_c = \frac{q}{h} = \frac{10}{7} = 1.4 \text{ gpm/in.} \qquad 7.6(1)$$

If the flow controller in the discharge pipe from the vessel maintains an outflow rate of 5 gpm (19 lpm), at steady-state conditions, the flow into the vessel has to match that flow rate. Therefore, if the process load is 5 gpm, the level in the tanks at steady state is

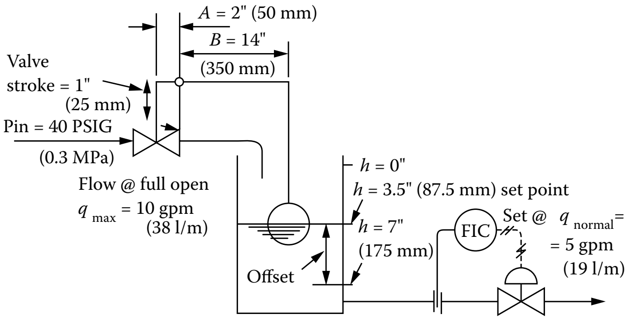


FIG. 7.6a

Illustration showing why in a float-type level regulator an offset is unavoidable if a load change occurs.

$$h = \frac{q \text{ normal}}{K_c} = \frac{5}{1.4} = 3.5 \text{ in.} \qquad 7.6(2)$$

When a load change occurs, either because the operator changed the FIC set point or because the liquid supply pressure at the inlet of the regulator valve changed, the result will be a permanent offset in the tank's level. Assuming, for example, that the header pressure is reduced from 40 PSIG (276 kPa) to 10 PSIG (69 kPa), and therefore the maximum flow through the control valve when fully open drops from 10 gpm (40 lpm) to 5 gpm (20 lpm), the new proportional sensitivity of the process becomes

$$K_c = \frac{q}{h} = \frac{5}{7} = 0.7 \text{ gpm/in.} \qquad 7.6(3)$$

Therefore, in order to obtain the previous outflow of 5 gpm (20 lpm) under these new conditions, the float of the level regulator has to sink until the valve is fully open. This occurs at

$$h = \frac{q \text{ nomal}}{K_c} = \frac{5}{0.7} = 7 \text{ in.} \qquad 7.6(4)$$

Therefore, the level of 3.5 in. that was previously maintained cannot be kept under these new load conditions and a permanent offset error of 3 (87.5 mm) will result in the controlled level. The amount of offset can be reduced by increasing the gain (reducing *B*/*A*) but it cannot be fully eliminated.

If the proportional band was reduced from 700 to 100% ($B/A = 1$), the proportional sensitivity would be increased by a factor of seven ($K_c = 4.9$), and the permanent offset would be reduced to 1/2 in. (12.5 mm). On the other hand, with a wider proportional band (*B*/*A* increased), the permanent offset would also rise.

Span, Dead Band, and Hysteresis

In all instruments that have boxes, pivots, and mechanical linkages, before motion can occur, the "break-away torque" has to be overcome, and therefore they all have some dead band. In addition, in such devices the direction of motion will influence the force-to-motion relationship, which will cause hysteresis. Both hysteresis and dead band are substantial in float-type level control valves.

The span of these blind, self-contained level control devices is adjustable, but the maximum control range is normally less than 12 in. (300 mm).

Power Generated by the Float

The buoyant force generated by the float is a function of the float volume and of the fluid density. The power that is

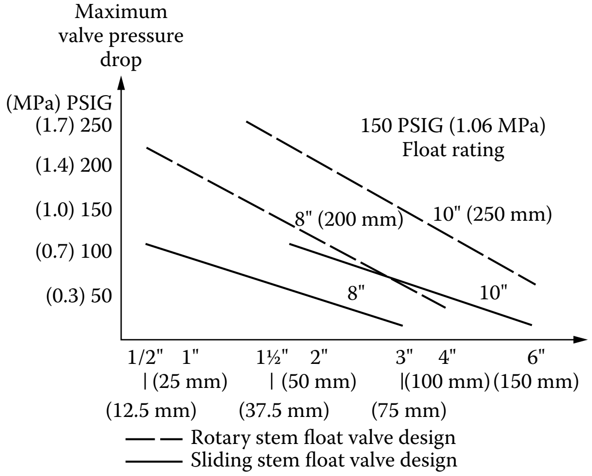


FIG. 7.6b
The maximum pressure difference against which the float-type level regulator can close is a function of both the valve size and the stem design.

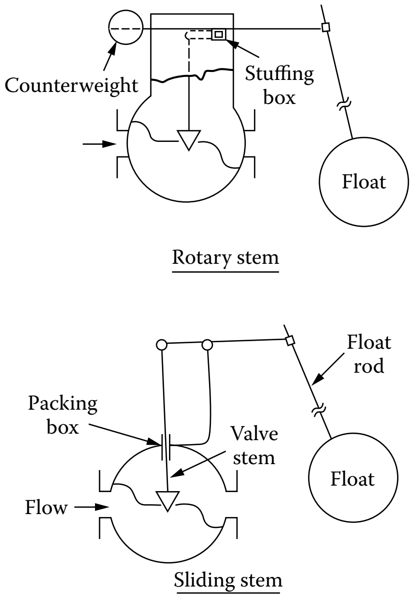


FIG. 7.6c
Of the total power generated by the float, the rotary stem design leaves more to overcome the process fluid forces at the valve plug because it has less friction than the sliding stem design.

available to operate the float or to close the valve against the static pressure in the supply header is substantially less than the buoyant force, because the float weight has to be subtracted from the gross buoyant force. The friction losses in the stuffing boxes, pivots, and mechanical linkages also reduce the net power available.

In addition, the total float volume cannot be used for sizing because the float is only partially submerged when it operates the valve. For example, an 8 in. (200 mm) diameter float with 150 PSIG (1.04 MPa) pressure rating, when totally submerged in water, has a gross buoyancy of 9.5 lb_f (42.3 N). After subtracting the float weight, the net total buoyancy is 6 lb_f (26.7 N). If the float design pressure is 500 PSIG (3.45 MPa), the net buoyancy is reduced to 4.5 lb_f (20 N). This net force is then amplified through the mechanical levers, while some of it is used to overcome the friction of the components.

Figure 7.6b shows the relationship between the size of the valve and the maximum pressure drop against which the valve is able to close. As one would expect, a larger float is required to operate the same size valve against a higher pressure drop.

As shown in the figure, there is a substantial difference between the limitations of the rotary and sliding stem valve designs. The rotary stem design shown in Figure 7.6c has less friction than the sliding stem design because of the relatively low friction of its lubricated stuffing box and therefore can handle higher pressure drops.

Specific Gravity and Temperature Effects

Specific gravity (SG) is the ratio of the process fluid density to that of 60°F (15.6°C) water at sea level. The buoyant force generated by a float is a function of the difference between the specific gravity of the float (SG_f) and the specific gravity of the process fluid (SG_p). The point at which the process fluid begins to lift the float and therefore change the valve opening is also affected by the float's specific gravity in relation to the process fluid's specific gravity.

The float will sink in the process fluid if its specific gravity is higher than that of the fluid ($SG_f > SG_p$) and therefore such float cannot be used at all. The specific gravity of the float must always be lower than that of the process fluid ($SG_f < SG_p$). The float will generate more buoyant force as the difference ($SG_f - SG_p$) rises.

So, for example, a float with a specific gravity of 0.6 in water with a specific gravity of 1.0 has to be submerged to about 60% of its volume before it will become buoyant and begin to lift. If that float is 8 in. (200 mm) in diameter, the water level has to rise to 4.8 in. (122 mm) before the float begins to generate a buoyant force to move the valve stem. In comparison, if the same 8 in. (200 mm) diameter float was used in diesel fuel oil with an SG of 0.82, the float will need to be submerged to a depth of 5.9 in. (150 mm) above the bottom of the float before buoyant force is generated.

The process fluid temperature will also affect the float buoyancy, because temperature affects the specific gravity of the fluid. Using the same 8 in. (200 mm) diameter float as an example, it was estimated above that cold water at a specific gravity of 1.0 will start to generate buoyant force when the level has risen to 4.8 in. (122 mm). In contrast, if the water is heated to 176°F (80°C) the specific gravity drops to 0.97, and therefore the float will only start to generate buoyancy force when the level has risen to 5 in. (127 mm) above the bottom of the float.

One can approximate the depth to which the float has to be submerged before it starts to generate a buoyant force as

$$\text{Zero buoyant force level} = (D)(\text{SG}_f/\text{SG}_p) \qquad \textbf{7.6(5)}$$

where:
D = float diameter

Stuffing Boxes

Displacer and float-type level transmitters and controllers are described in the *Process Measurement* volume of this handbook. Sections 3.7 and 3.8 in that volume cover units that are isolated from the process vessels by means of frictionless seals. The seals on the float-type level control valves are not frictionless. The rotation caused by float motion is transmitted through a stuffing box to the mechanical levers that operate the valve.

The stem friction in the stuffing box is reduced by the use of a lubricated fitting, which can be refilled while the unit is in service. The design of the fitting is such that it lubricates the total region of metal-to-metal contact. Ball bearings are normally provided in these stuffing boxes to absorb radial and thrust loads, and thereby reduce stem friction. When repacking is needed, it can be done by breaking the packing gland union on the stuffing box and by replacing the package without removing the unit from the vessel.

With operating temperatures below 30°F (–1.1°C) or above 450°F (232°C), radiation fins are provided between the vessel and the packing box to cool the assembly by the air at ambient temperature.

Installation

The installation of float control valves on open vessels is shown in Figure 7.6a, and the regulator can be as simple as the float valve shown in Figure 7.6d.

When the vessel is pressurized, one option is to use a flange-mounted unit, where the float is inserted into the vessel through a nozzle (Figure 7.6e). In this configuration, the vessel has to be drained before the float regulator assembly can be removed for repairs or maintenance.

In this configuration, the stuffing box is perpendicular to the float arm. The proper balance of linkage, turnbuckle, float, and inner valve weights is achieved by adjusting the counterweights. The connecting rod between the float assembly and the valve should be of minimum length and needs to be carefully aligned to avoid binding during float travel.

In Figure 7.6f, the stuffing box is screwed directly into the flat vessel wall. The float has to be installed from the inside, and the vessel must be drained for servicing.

The external cage can also be utilized, as shown in Figure 7.6g. Here the float level regulator chamber is provided with isolation valves so that the instrument can be

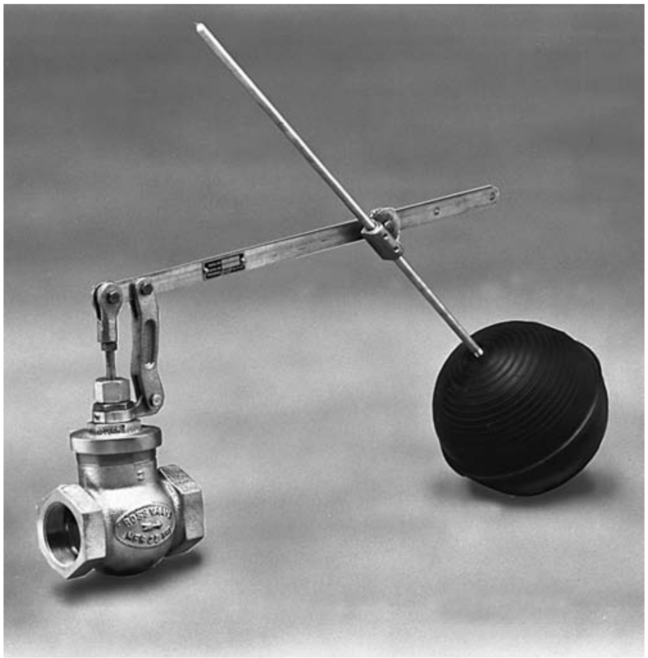


FIG. 7.6d
Simple float level regulator for open tank installation. (Courtesy of Ross Valve Mfg. Co.)

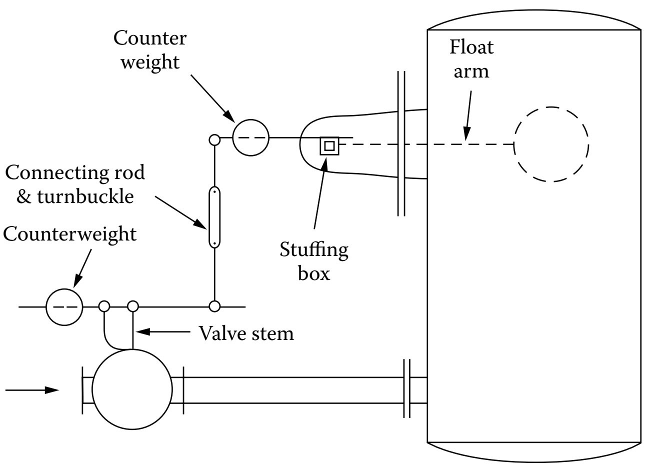


FIG. 7.6e
Flange-mounted float assembly.

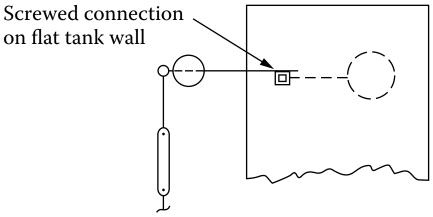


FIG. 7.6f
Direct mounting of float into coupling on tank wall.

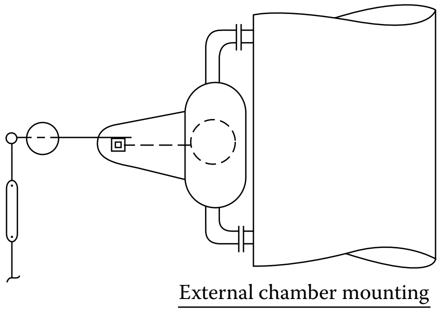


FIG. 7.6g
The mounting of a float level regulator on a pressurized tank in an external chamber.

removed from the tank and serviced without disruption of the process.

The maximum recommended length for the connecting rod with the turnbuckle is about 8 ft (2.4 m). If the vertical distance between float assembly and valve is greater than that, a mechanical lever system cannot be used and the use of a regular level control loop has to be installed.

The float assembly with rotating stem stuffing box can also be provided with a pneumatic pilot. Its main advantage is the increase in the available power to operate the valve and the ease of sensitivity adjustment. As shown in Figure 7.6h, these pilot-operated valves can be very large, up to 48 in. (1.22 m). Such a valve weighs 21,000 lb.

The float-type level switches and transmitters are discussed in Section 3.8 of the first volume of this handbook. They are available in electric or pneumatic designs Their sensitivity adjustment is simpler than of the mechanical unit, because the power required to operate the valve need not be furnished by the float. The float assembly can also be used for level alarming. In that design, a switch is actuated by the rotating stem instead of by mechanical levers.

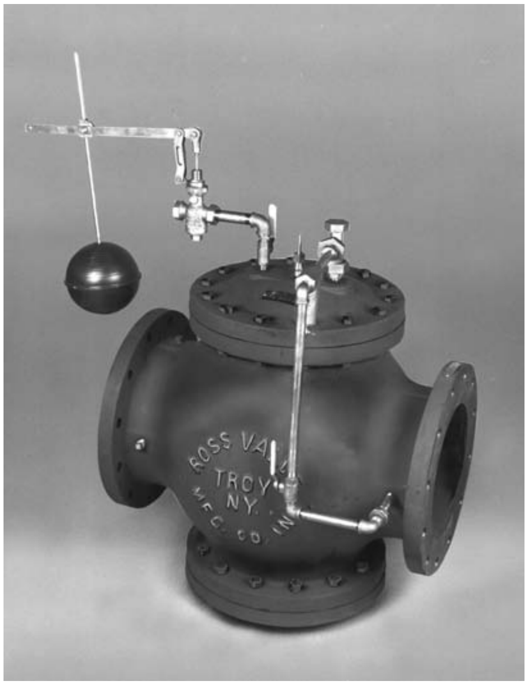


FIG. 7.6h
Pilot-operated regulator valves are not limited by the buoyant force of the float and can be very large. (Courtesy of Ross Valve Mfg. Co.)

Conclusions

Float-type level regulators can be used where the quality of control is not critical and the narrow span, offset, dead band, and hysteresis that are inherent in such regulators can be tolerated. Because of the large number of moving parts, pivots, levers, and the stuffing box, the device requires regular maintenance and should be used only on clean fluids where material build-up will not interfere with the operation of the moving parts.

The use of these level controllers is limited because they are blind, are not frictionless, and the physical relationship between valve and float is restricted. The economy of this design must also be balanced against the limitations of float power and the limitations in the available materials of construction. The industrial applications of float-type level controllers are therefore quite limited. On the other hand, they are still used on open water tanks, such as cooling tower basins, where the lack of readily available or economical external power sources does not limit the use of this design.

DIVERTER VALVES

This solid-state level regulator has no moving parts and utilizes fluidics principles. This valve works uses the Coanda effect, namely that a fluid stream will attach itself to the nearest side wall. Figure 7.6i illustrates such an on/off level control valve, where one wall of the diverter is closer to the incoming fluid stream than the opposite sidewall. On the closer side wall, a level sensing dip pipe is connected to the control port of the diverter.

When the level in the vessel is below the dip tube opening, air from the atmosphere enters the dip tube. The air is entrained into the control port, causing the inlet jet to detach

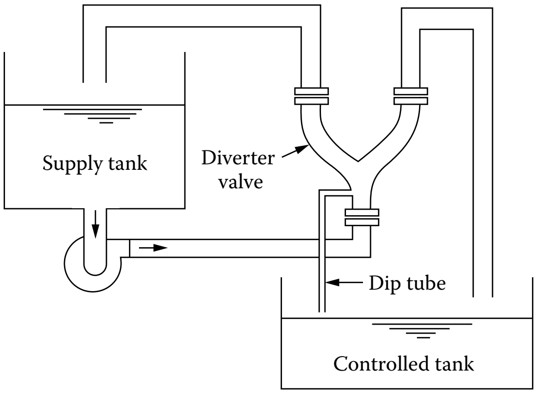


FIG. 7.6i
On/off level control using a solid-state diverter valve.

itself from the adjacent side wall and to attach itself to the opposite side wall, thereby diverting the full flow to the controlled vessel. As the level in the controlled vessel rises, it will eventually cover the end of the dip tube, thereby cutting off the air flow to the control port. As soon as this occurs, the flow stream will attach itself to the "favored" side wall and will divert the full flow back to the supply tank for recirculation. Having no moving parts, the on/off cycling of the valve has no harmful effects.

It is the shape of the flow passage that makes the valve function. With no plugs or packing, as in conventional valves, the unit is essentially maintenance-free. The valve is not affected by vibration or temperature (operating or ambient) and requires no external power source for operation. The flow stream is diverted in one tenth of a second, and the valve can be manufactured out of metal, plastic, or ceramic material. The design is streamlined without shoulders or pockets and so is recommended for use not only on clean liquids, but also on fluidized solids and slurries. It should be noted, however, that for proper operation the dip tube and the control port must be open; any material build-up in them will interfere with the functioning of the device.

When the solid-state diverter valve is considered for control application, the effect of back-pressure must be carefully considered. Outlet pressures other than atmospheric should be evaluated as to their effect on tight shut-off and on the safe-failure position.

The solid-state diverter valves can also be used for throttling instead of on/off control. In such a design, there are two ports for each leg of the valve, and the quantity of air entering the ports determines the ratio between the streams leaving through the two legs. For throttling control, the back-pressure in the controlled leg cannot be more than half of the inlet pressure to the valve in order not to spill into the unused leg.

The solid-state diverter valve is a reliable on/off control device for atmospheric vessels containing hot or cold, corrosive or noncorrosive clean liquids. Its application in pressurized vessels or with hard-to-handle process fluids is not recommended.

ALTITUDE VALVES

Altitude valves are installed in supply lines to elevated basins, tanks, or reservoirs for the purpose of holding the reservoir level constant and to prevent overflow. No external energy source is required because the power for operation is gained from the pressure of the process fluid.

One of the simplest versions of this design is shown in Figure 7.6j. The purpose of this valve is to admit water into the elevated tank until a preset level is reached and to close at that point to prevent overflow. The hydraulic head in the elevated tank is balanced by the setting spring of the pilot. When the tank water level exceeds the setting of the spring, the pilot diaphragm is moved down, closing the drain port and opening the main line pressure to apply its force to the top piston. Because the top area of the piston is greater than the bottom, the main valve closes.

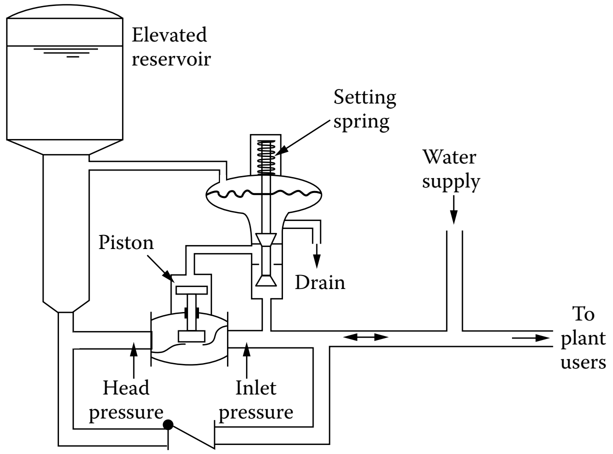


FIG. 7.6j
Typical altitude valve installation. When the reservoir level is high, the pilot opens the water supply to the top of the piston in the main valve, which closes the water supply to the reservoir.

If the head in the tank decreases, the setting spring moves the diaphragm up, closing the main line pressure connection and opening the drain port on the pilot. This results in removing the pressure from the large end of the piston, thereby opening the main valve to supply additional water to the head tank.

Where this type of altitude valve is used, discharge from the tank is handled by a swing check valve in a bypass line around the altitude valve or by a separate line from the tank to the plant. With the check valve as shown in the figure, water will flow from the head tank to the plant whenever the demand for water used exceeds the supply available. When the plant demand falls back below the capacity of the supply, the check valve closes and the water is made up in the tank under the control of the altitude valve.

Altitude valves are multipurpose devices, and with the addition of pilots, they can satisfy a number of requirements, in addition to the simple on/off level control described above.

All altitude valves close on high tank level to prevent overflow, but they can have other features as well:

1. The valve can have a delayed opening so that it opens only when the tank level has dropped below the control point by a set amount.
2. Instead of on/off action, the valve can be closed gradually to eliminate pressure shock (water hammer).
3. Instead of opening on low level, the valve can stay closed until the water supply pressure drops below the head pressure of the tank. With this unit, there is no need for the check valve in the bypass as shown in Figure 7.6j, because the valve is designed for two-way flow, allowing the water to return from the tank when the distribution or supply pressure drops below that of the tank head.

4. The valve can be modified to delay its opening until the supply or distribution pressure has dropped to an adjustable preset point.
5. By using another combination of pilots, the valve will open on low level, but only to the extent needed to maintain constant inlet pressure to the valve (which is the distribution pressure to the plant).
6. The unit can be provided with an additional check valve feature. In this arrangement, the valve operates as described, except that it closes if the distribution pressure drops to a predetermined low point due to line breakage or other reasons.
7. When used on aerator basins with high supply pressures, the altitude valve can also act as a pressure reducer. In this case, the valve will open on low level only to the extent needed to provide a permissible outlet pressure for the aerator nozzles.

As far as accessories are concerned, all altitude valves can be provided with solenoid pilots to close the valve on a remote electrical signal. Limit switches can also be furnished to give remote indication of valve position (whether open or closed).

The use of altitude valves is recommended for locations where only hydraulic power is available to operate large water valves controlling the level in storage reservoirs. They are not applicable to level control in pressurized vessels. The materials of construction of these units restrict their applications to water service. Due to the substantial number of components and moving parts, they require periodic cleaning and maintenance. Protection against freezing of stationary sensing lines is also required. In line sizes greater than 12 in. (300 mm), there is some economic advantage to using butterfly valves with hydraulic pilots.

Bibliography

API Guide for Inspection of Refinery Equipment, Chapter 15, "Instruments and Control Equipment."

API Recommended Practice 550, "Manual on Installation of Refinery Instruments and Control Systems," Part I, Section 2, Level.

Bacon, J. M., "The Changing World of Level Measurement," *InTech*, June 1996.

Belsterling, C. A., "A Look at Level Measurement Methods," *Instruments and Control Systems,* April 1981.

Boyes, W., "The Changing State of the Art of Level Measurement," *Flow Control*. February 1999.

Carsella, B., "Popular Level-Gauging Methods," *Chemical Processing*, December 1998.

Cheung, T. F., and Luyben, W. L., "Liquid-Level Control in Single Tanks and Cascades of Tanks with Proportional-Only and Proportional-Integral Feedback Controllers," *Industrial and Engineering Chemistry Fundamentals,* February 1979, p. 15.

Entwistle, H., "Survey of Level Instruments," 1991 ISA Conference, Anaheim, CA, Paper #91-9484, 1991.

Felton, B., "Level Measurement: Ancient Chore, Modern Tools," *InTech,* August 2001.

Glenn, L. E., "Tank Gauging—Comparing the Various Technologies," 1991 ISA Conference, Anaheim, CA, Paper #91-0471, 1991.

Hall, J., "Level Monitoring: Simple or Complex," *Instruments and Control Systems,* October 1979.

Hall, J., "Measuring Interface Levels," *Instruments and Control Systems,* October 1981.

Johnson, D., "Checking Level: Not Glamorous, Sometimes Dangerous, but Necessary," *Control Engineering*, August 2001.

Koeneman, D. W., "Evaluate the Options for Measuring Process Levels," *Chemical Engineering*, July 2000.

Lawford, V. N., "How to Select Liquid-Level Instruments," *Chemical Engineering,* October 15, 1973.

"Level Measurement and Control," *Measurements and Control,* April 1993.

Morris, H. M., "Level Instrumentation from Soup to Nuts," *Control Engineering,* March 1978.

Paris, T., and Roede, J., "Back to Basics," *Control Engineering,* June 1999.

Parker, S., "Selecting a Level Device Based on Application Needs," *Chemical Processing 1999 Fluid Flow Manual,* 75–80, 1999.

Paul, B. O., "Seventeen Level Sensing Methods," *Chemical Processing*, February 1999.

Sastry, V. A., "Self-Tuning Control of Kamyr Digester Chip Level," International Symposium on Process Control, May 2–4, 1977.

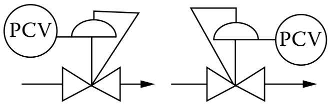
Sholette, W., "Pick the Proper Level Measurement Technology," *Chemical Engineering Progress,* October 1996.

Taylor, D., "Economical Method of Glass Level Control Utilizing a Laser Beam Generator and a Solid State Detector," *Proceedings of IAS Annual Conference,* 1977.

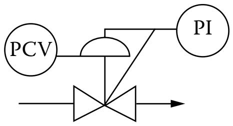
Yearous, H. B., "Sensors for Liquid/Solid Level Controls," *Plant Engineering,* March 8, 1979, pp. 109–112.

7.7 Regulators — Pressure

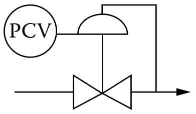
R. L. MOORE (1970) **D. M. HANLON** (1985)
B. G. LIPTÁK (1995) **J.E. JAMISON** (2005)



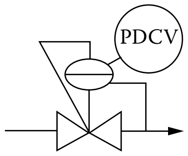
Pressure reducing and back pressure regulators



Airset, regulator with outlet pressure relief and indicator



Regulator with external pressure tap



Pressure differential regulator with internal and external pressure taps

Flow sheet symbols

Regulator Types:	A. Weight-loaded B. Spring-loaded C. Piloted 1/4 in. (6.25 mm) air regulator D. Internally piloted E. Externally piloted
Sizes:	A. 1/2 to 6 in. (12.5 to 150 mm) B. 1/4 to 4 in. (6.25 to 100 mm) C. 1/4 in. (6.25 mm) D. 1/2 to 6 in. (12.5 to 150 mm) E. 3/4 to 12 in. (9.35 to 300 mm)
Design Inlet Pressure:	B. Up to 6000 PSIG (41.4 MPa) D and E. Up to 1500 PSIG (10.35 MPa) A and C. Up to 500 PSIG (3.45 MPa)
Minimum Regulated Outlet Pressure:	B and E. Down to 2 PSIG (13.8 kPa) A and D. Down to 0.5 PSIG (3.45 kPa) C. Down to 0.1 PSIG (0.69 kPa)
Droop or Offset:	B. 5 to 80% E. 2 to 10% A and D. 1 to 2% C. 0.5%
Regulator Specification Form:	See Table 7.7r at the end of this section

Cost:	A. $150 to $2000 B. $50 to $1000 C. $100 D. $150 to $4000 E. $150 to $6000
Partial List of Suppliers:	Advanced Pressure Products (www.pmiapp.com) Ametek Inc. (www.ametek.com) Anderson Greenwood & Co. (www.andersongreenwood.com) Armstrong International Inc. (www.armstrong-intl.com) Atwood & Morrill Co. (www.petropages.com/vendors/v7566.htm) Barksdale Controls Div. of IMO Industries Inc. (www.flw.com/barksdale) Bellofram Corp. (www.marshbellofram.com) Brooks Instrument (www.fischerporter.com) A. W. Cash Valve Manufacturing Co. (www.cashacme.com) Cashco Inc. (www.cashco.com) Circle Seal Controls Inc. (www.circle-seal.com) Control Components Inc. (www.ccivalve.com) CPV Manufacturing Inc. (www.cpvmfg.com) Eaton Corp. (www.aeroquip.com) Fairchild Industrial Products Co. (www.flw.com/fairchild) Fisher Controls International Inc. (www.emersonprocess.com) Furon Co. Gilmore Valve Co. (www.gilmorevalve.com) Go Inc. (www.goreg.com) Honeywell Inc. (www.honeywell.com) ITT Conoflow (www.conoflow.com) Jordan Valve (Div. of Richards Industries) (www.jordanvalve.com) Kaye & MacDonald Inc. (www.kayemacdonald.com) Keane Controls Corp. (www.tyco.com) Keystone International Inc. (www.business.com/directory/) Leslie Controls Inc. (www.lesliecontrols.com) Masoneilan (www.masoneilan.com) Matheson Gas Products Inc. (www.thinkenergy.com) Moore Products Co., a Division of Siemens (www.sea.siemens.com) Norgren (www.norgren.com) Parker Hannifin Corp. (www.parker.com) Plast-O-Matic Valves Inc. (www.plastomatic.com) Richards Industries—Valve Group Inc. (www.jordanvalve.com) Robertshaw Controls Co. (www.robertshaw.com) Spence Engineering Co. (www.spenceengineering.com) Spirax Sarco Inc. (www.spirexsarco.com) Tescom, Pressure Controls Div. (www.tescom.com) Tokheim Corp. (www.tokheim.com) U.S. Para Plate/Servo-Dome Regulators (www.usparaplate.com) Valtek Inc. (www.flowserve.com) Veriflo Corp. (www.parker.com) Watson McDaniel Co. (www.watsonmcdaniel.com) Watts Regulator Co. (www.wattsreg.com) Wilkins Regulator Co. (www.trumbull.com) *Note:* The most popular regulators are manufactured by Fisher Controls, Cashco, Spence Engineering, and Anderson Greenwood.

INTRODUCTION

The pressure regulator is a complete pressure control loop, incorporating a sensor, a controller, and a valve. It is called a regulator (and not a controller) because it is mechanical and self-contained, requiring no external energy source.

In this section, first a comparison is drawn between regulators and controllers. This is followed by a description of the operating characteristics of pressure regulators, giving particular emphasis to the subjects of droop, noise, and stability. Next, the specific designs and their variations will be described, including such specialized units as the safety shutoff regulators used in gasoline station applications.

REGULATORS VS. CONTROL VALVES

In many applications, pressure can be controlled by either a regulator or a full control loop. In making the choice, the design engineer should consider their relative merits, which are listed in Table 7.7a.

Particularly in the smaller sizes, regulators usually cost less than a control loop consisting of a control valve, transmitter, and controller. Regulators are less expensive to buy, install, and maintain. But when the application requires a larger valve, the economics begin to change and, in sizes over 4–6 in. (100–150 mm), might favor of control valves.

Regulators have a built-in controller and do not require an air supply. This results in savings in both purchase and installation costs. As a consequence, regulators are not subject to air-supply failure. In applications where the fail-safe feature is essential or in locations that are remote from a source of compressed air, this is an important consideration. On the other hand, diaphragm failure in a regulator usually results in the opening of the valve, which can be unsafe.

Control Valves

Control valves are used with an external controller. The controller can be provided with one, two, or three control modes and can operate in either the automatic or the manual mode. Figure 7.7b illustrates a simple, proportional-only controller.

TABLE 7.7a
Comparison between Regulators and Control Valves

Regulators	*Control Valves*
Lower cost (in small sizes and ordinary materials of construction)	Higher cost
Smaller size	Larger and heavier
Lower installation cost	Higher installation cost due to size and weight
No air supply required	Requires air connections
Built-in controller	External controller
Limited materials of construction	Practically no limit on materials of construction
No remote control; set point adjustment is at regulator location	Local or remote control; remote/manual control possible
Single mode (proportional only) with associated droop	Varied control modes possible
Most fail open; relief valve is recommended	No limitation in fail-safe
Limited accessories	Wide variety of accessories
Limited interchangeability	Possibility of interchangeability among services
Fewer applications	More (and more complex) applications

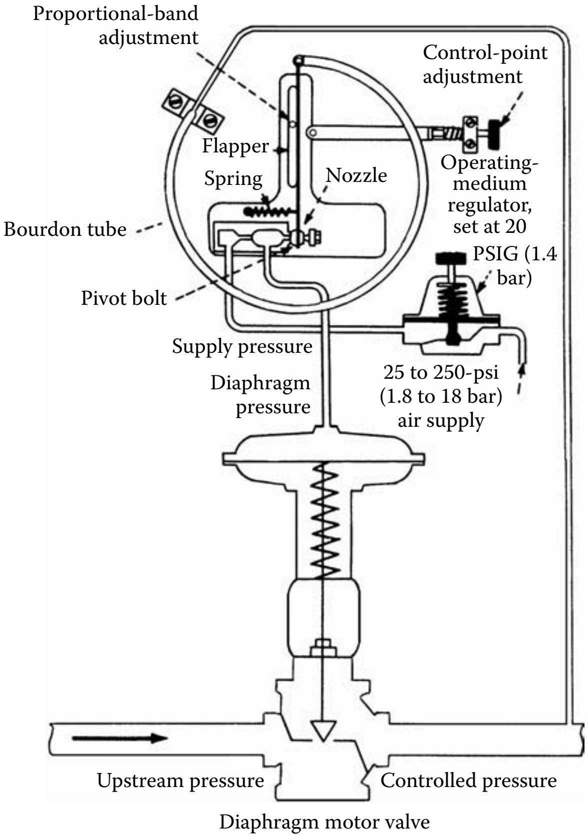


FIG. 7.7b
The main components of a proportional-only, pneumatic pressure control loop.

The valve itself can be made from castable metal, and it may be lined. Control valves can be provided with the failure position, which makes the operation of the process the safest. Valve accessories include positioners, limit switches, manual handwheels, and solenoids. These can be interchangeable with the accessories on valves in the plant, thereby reducing spare-parts requirements.

In a regulator, the set point is integral, and remote set point adjustment is usually not possible. All regulators are single mode (proportional only) and therefore they all have a set point droop, which results in an offset if the load (throughput) changes. The available regulator materials of construction are limited, as is interchangeability with other services. Accessories are not available.

In contrast with regulators, regular control loops are more expensive to purchase, install, and maintain. The control valves are less compact, require more space for installation, and require a source of compressed air for their operation.

REGULATOR APPLICATIONS

On steam-driven pumps, pump governors frequently regulate the suction or discharge pressures by modulating steam flow. On purging applications, differential-pressure regulators are

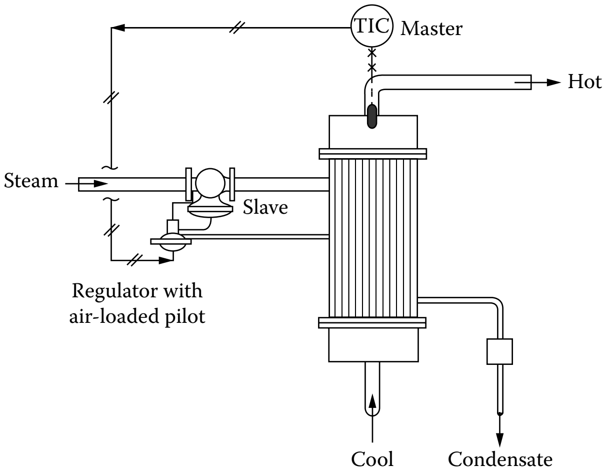


FIG. 7.7c
Because of its speed of response, a pressure regulator provides an ideal slave controller in a cascade loop.

used to keep the purging media at a pressure higher than that of the process. This same regulator is also used in air bubbler-type level measurement systems. A differential-pressure reducing valve can also be used to provide constant water flow for wet-bulb humidity detection.

Oil-burner back-pressure regulators are installed in the oil return line. Their settings are modified according to the steam pressure in the boiler. The regulator stem is discontinuous to permit both low- and high-fire pressure adjustments.

Pressure regulators have been successfully used as cascade slaves receiving their set points from temperature control masters. Figure 7.7c illustrates such an installation on a steam-heated water heater. The air-loaded pressure regulator is extremely fast and corrects for load changes or steam-supply pressure variations almost instantaneously. This is exactly the kind of performance that is required from a cascade slave.

In this arrangement, the work that the temperature controller (master) has to do is much reduced because its task is reduced to compensate for the nonlinearity of the heat transfer curve by adjusting the set pressure.

Gas Industry Applications

In addition to regulators available for general service, the gas utility industry uses an extensive family of regulators designed specifically for gas service. The types of regulators for these services are tabulated in Table 7.7d.

There are basically two types of pressure regulators: the flow type and the dead-end type (Figure 7.7e). The flow type can reduce the flow to zero, but cannot allow for flow reversal. This is acceptable for most applications except for the dead-ended air pressure control services. For them, if the outlet pressure is too high then the air has to be released to the atmosphere.

Pressure regulators perform well on gas control applications. Stable control is provided because the regulator response is usually much faster than that of the process, and the deviation from set point (droop or offset) can be kept small, because the gain is high (20 or more) and therefore the amount of offset is low. These performance characteristics will be discussed in more detail later.

TABLE 7.7d
Regulators Used in the Gas Industry

Application Pressures	*Inlet Pressures*	*Outlet*
Appliance	8 in. H_2O (2 kPa)	3–4 in. H_2O (0.75–1 kPa)
Household	2–100 PSIG (13.8–690 kPa)	5–8 in. H_2O (1.25–2 kPa)
Community of users, 60 PSIG (414 kPa)	Up to 200 PSIG (1380 kPa)	3 in. H_2O to 60 PSIG (0.75–414 kPa)
Transmission line take-off, PSIG in two stages	Up to 1200 PSIG (8.28 MPa)	100–1000 PSIG (690 kPa–6.9 MPa)
Transmission line take-off to individual user	Up to 1500 PSIG (10.35 MPa)	3–600 PSIG (20.7–4140 kPa)

A frequent problem in the gas industry is the need to accurately regulate to very low pressure levels while using a relatively high supply pressure. Shutting off against such pressure requires a relatively large force. Because the regulated pressure is low, it must be amplified by using both large diaphragm areas and mechanical amplification.

Liquid Pressure Regulation

The use of pressure regulators on liquid pressure control applications is not as likely to provide good control. This is because on noisy (very fast) liquid processes, the regulator is not likely to be fast enough to eliminate noise and, in fact, might amplify it. Generally it is better to control such processes with low gain (wide proportional) and long integral mode settings. Such settings are not available in regulators, as regulators are high gain, basic proportional controllers.

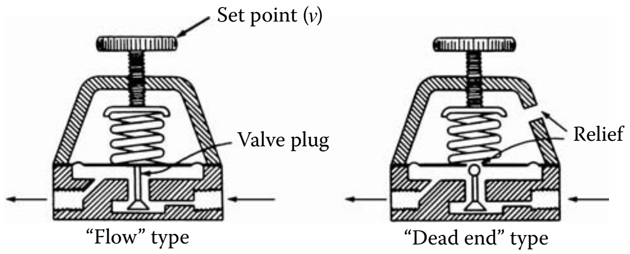
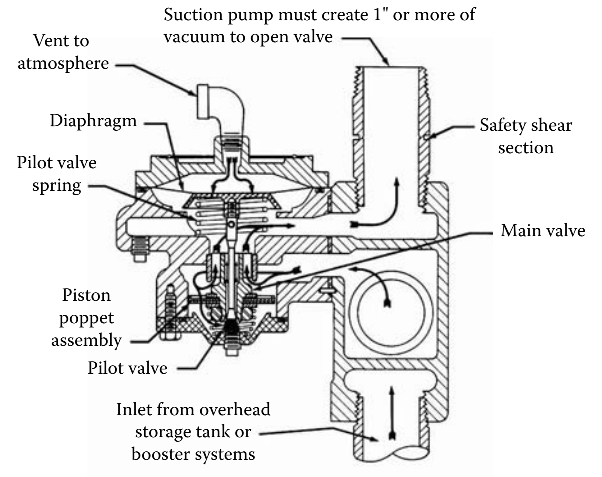
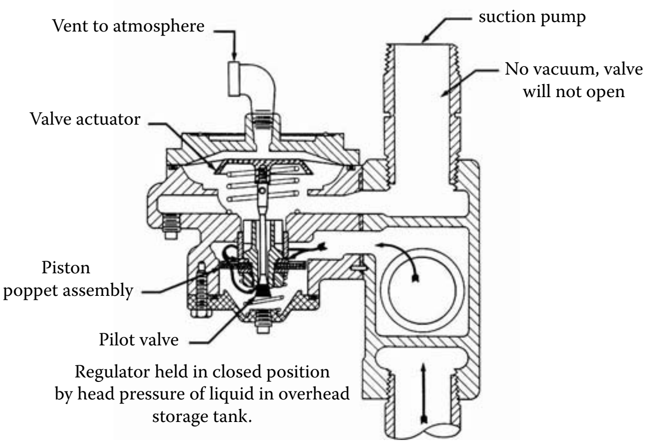


FIG. 7.7e
Pressure regulators can be either the flow type or the dead-end type. Dead-end units are used on air applications where the relief of excess pressure to the atmosphere is also needed.



Open position



Closed position

FIG. 7.7f

Safety regulators serve to protect gasoline dispensers at gas stations. The regulator is installed below the pump in the dispenser cabinet. It can open only if the gasoline pump is on, because the pump's suction generates a vacuum. If the safety shear section is broken, the vacuum is lost and the regulator shuts off, preventing a spill. (Courtesy of Tokheim Corp.)

Another reason why the use of regulators is usually not a good choice for liquid pressure control is that regulators do not provide tight shut-off. Therefore overpressuring of the system due to leakage can occur when the supply regulator is not fully closed, while there is no demand for the controlled fluid.

Gas Station Safety Regulators

A variety of pressure regulators have been designed to meet specific safety concerns in different operations. One such concern is the spilling of gasoline at gas stations if a truck accidentally backs into a dispenser unit and thereby breaks the gas pipe. The safety regulator shown in Figure 7.7f has been designed to protect against such an occurrence. The regulator is installed in the ground, below the gas dispenser unit, on the suction side of the gas pump.

When the gas pump is off, the regulator is closed. When the gas pump starts, its suction pulls a vacuum below the diaphragm of the regulator and opens it. If at any time the safety shear section of the outlet of the regulator (suction of the gas pump) is damaged, the vacuum is lost and the atmospheric pressure closes the regulator. When the gasoline is supplied from overhead tanks, the blocked-in gasoline contained in the piping between the shut-off valve at the tank and the regulator can be exposed to thermal expansion. In order to protect from mechanical damage, thermal expansion relief valves must be installed.

Pressure Surge Relief Valves

The pressure surge relief valve assembly shown in Figure 7.7g is designed to open and regulate flow and velocity in crude oil pipeline applications when the upstream pipeline pressure reaches a specified set point. The pressure surge relief control consists of a main valve body assembly for relieving flow, a hydraulic cylinder for actuator thrust, and a control manifold assembly to control modulation of the main valve in relation to the surge pressure level above the set point. The main valve trim limits the flow velocity through the valve that should yield smooth and quiet operation. The pipeline fluid and its pressure level are used to control the surge relief function.

The control element is a three-way pilot-operated hydraulic valve that senses the upstream pressure. The three-way valve directs fluid pressure to either the top or bottom of the hydraulic actuator to modulate the main valve movement. A spring assembly above the hydraulic actuator, which provides main valve seating load when closed, can ensure a Class VI leakage requirement. The spring is sized to also permit the main valve to open when pressure of the pipeline exceeds the set point. If the pipeline pressure is less than the relief pressure, the three-way valve blocks the pipeline fluid.

When the set point pressure is exceeded, the three-way valve vents the top of the hydraulic actuator and pressurizes the bottom to open the main valve. The fluid from the top of the hydraulic actuator is vented to the main valve bonnet, which is open to the downstream pressure via a balanced plug. The valve characteristics are nearly linear; the flow through the valve is nearly in linear proportion to the valve stroke (Figure 7.7h). The capacity of the three-way valve is such that the main valve will fully open in about 4 sec. The speed of closing can be adjusted by changing the flow rate into the top of the hydraulic cylinder.

PRESSURE REGULATOR DESIGNS

The self-contained pressure regulator is simple, dependable, rugged, and inexpensive. Energy from the flowing fluid operates it. The pressure acts on the diaphragm, which moves the

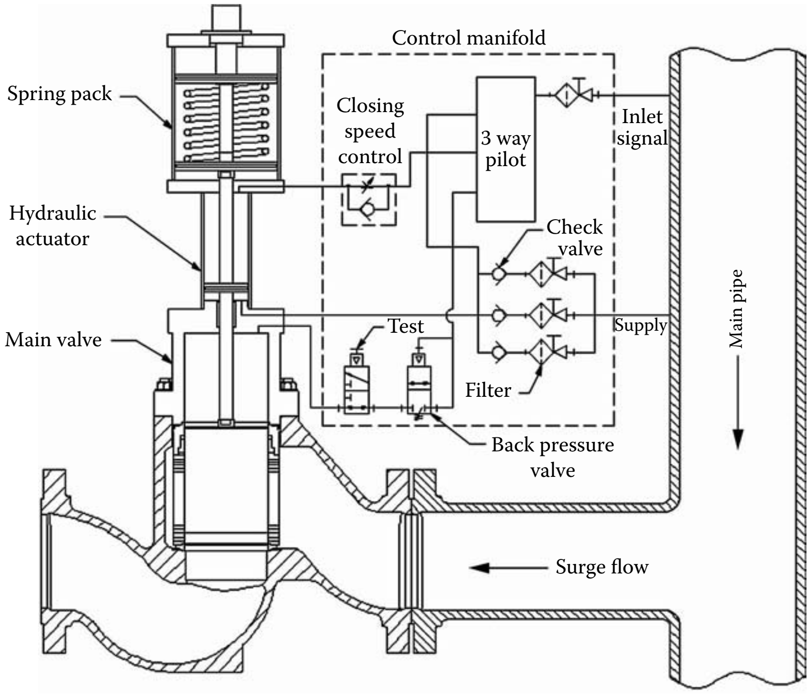


FIG. 7.7g
Schematic for a modulating crude oil pipeline application of a surge pressure relief valve (Courtesy of Control Components.)

valve plug (by compressing the spring). The initial spring compression sets the pressure at which the valve begins to open. For each pressure on the diaphragm, there is a corresponding seat position, or valve opening.

Figure 7.7i shows that the diaphragm is the "brain" of the regulator. It compares the set point that is converted into a spring force to the regulated pressure that is converted into a force by the diaphragm itself. It adjusts the valve opening to reduce the error between the two. The diaphragm comprises a feedback device, an error-detecting mechanism, and an actuator. Figure 7.7i shows a spring-loaded pressure-reducing valve. In other regulator designs, pressure regulators can also be loaded by weights or by gas pressure.

Regulators are also available to control back-pressure, differential pressure, or vacuum and can control the pressure of gases, vapors, and liquids. The self-contained regulator has no adjustments in terms of capacity (i.e., reduced valve trim) or in terms of tuning (proportional band and reset adjustments). It has a fixed proportional band of about 5%, or a gain of about 20. There are many different regulator types and designs available. Descriptions of the most common designs will be given later.

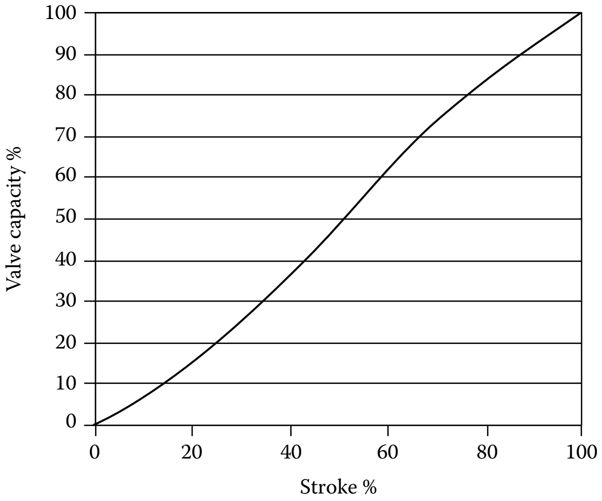


FIG. 7.7h
The characteristics (stroke vs. C_v) of this pressure relief valve are inherently linear. (Courtesy of Control Components Inc.)

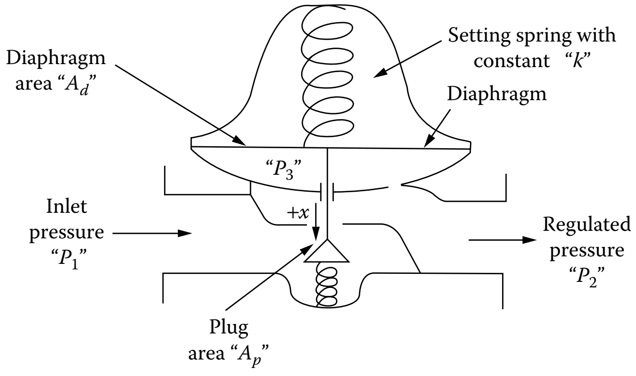


FIG. 7.7i
The basic components and operation of a spring-loaded pressure-reducing valve.

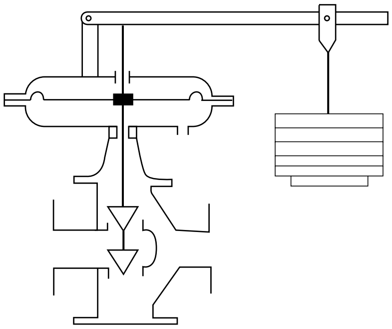


FIG. 7.7j
Weight-loaded pressure regulator with the external pressure tap below the diaphragm.

Weight-Loaded Regulators

The force in a spring changes as the spring is compressed. In a weight-loaded regulator, however, a weight provides a constant actuating force on the diaphragm, thus minimizing droop. Set-point changes are accomplished by adding weight or by changing the weight position (Figure 7.7j). Weight-actuated regulators are used to regulate gas pressures at 1 PSIG (6.9 kPa) or below when load changes are slow. It is not recommended for service where mechanical shock or vibration are present or for regulation of incompressible fluids.

Spring-Loaded Regulators

Spring-actuated regulators are used more than any other type because of their economy and simplicity. They also have an extremely fast dynamic response. Disadvantages include high droop, awkward set-point adjustment, and sensitivity to shock and vibration.

The "air regulator" is a common type of spring-loaded regulator. It is a 1/4 in. (6.25 mm) air pressure regulator used to reduce instrument air pressure to a level compatible with pneumatic instruments. The term "air set" is also applied to these regulators. Bleed-type regulators have the ability to relieve excess regulated pressure (Figure 7.7e). This design is recommended for dead-end (no flow) service. Various manufacturers supply air regulators having capacities from 10–60 SCFM (280–1680 lpm). These regulators are usually provided with an integral filter, hence the name "filter regulator."

Figure 7.7k shows one design of a spring-actuated regulator. The valve is normally closed. Turning the set-screw compresses the spring and opens the valve. Increasing downstream pressure acts beneath the diaphragm, thus raising it and closing the valve. Spring compression is adjusted to provide the desired downstream pressure at the demand flow.

When vacuum is to be regulated, the spring force is applied in the opposite direction than in case of the regulation of positive pressures. Figure 7.7l shows a spring-actuated vacuum pressure regulator.

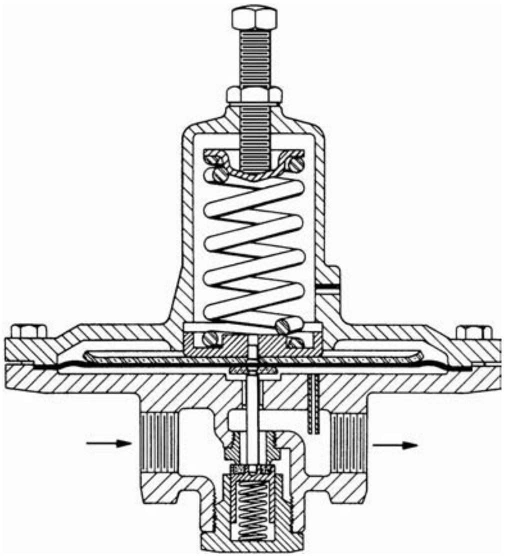


FIG. 7.7k
Spring-loaded pressure regulator.

Figure 7.7m illustrates a differential-pressure regulator that can be used to maintain a certain pressure difference between two streams.

Pilot-Operated Regulators

The pilot-loaded regulator is a two-stage device. The first stage contains a spring-actuated pilot regulator that controls pressure on the diaphragm of the main regulating valve. The second stage contains a spring-actuated regulator that controls the main valve stem. One advantage is that the actuating fluid operating the first-stage (pilot) regulator is at the upstream pressure. This provides a higher level of force to

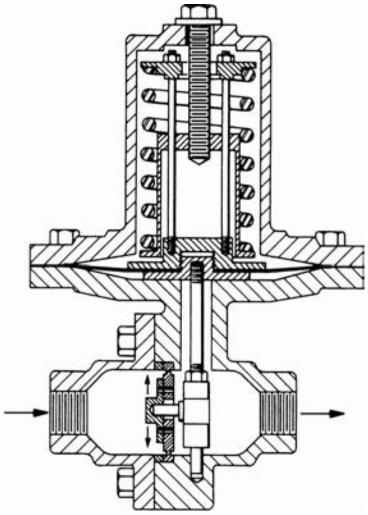


FIG. 7.7l
Spring-loaded vacuum regulator. (Courtesy of Jordan Valve.)

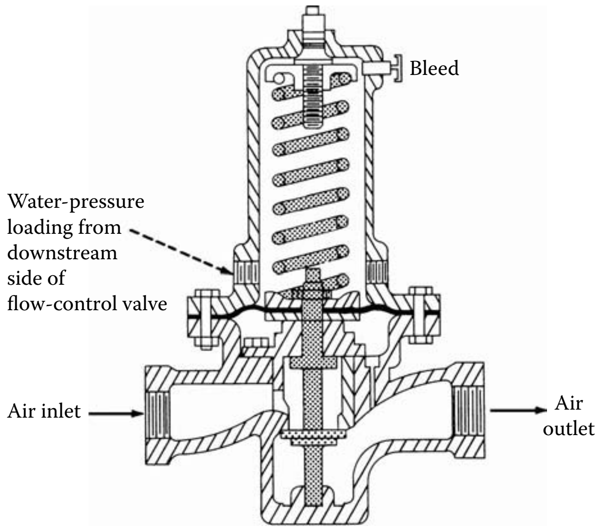


FIG. 7.7m
This differential-pressure regulator maintains a fixed pressure difference between a water pressure reference and the air pressure at its outlet.

the actuating mechanism. Travel is very short and that reduces droop. More accurate regulation is possible with this design.

Figure 7.7n shows that the pilot regulator can be external or internal. It controls pressure to a diaphragm operating the main valve. In some designs a piston can be used in place of the diaphragm. Pilot-operated regulators use the difference between upstream and downstream pressures for actuation. Therefore, a minimum differential is required for proper operation.

Pilot-operated regulators provide accurate regulation for a wide range of pressures and capacities. They are more expensive than spring-loaded regulators but also provide more capacity. Because the small passages and ports can become plugged, these regulators can only be used with clean fluids. Their dynamic response is slower than with the spring-loaded designs.

Externally piloted regulators are available in a wide range of sizes. One advantage of this design is that the pilot regulator can be mounted at a distance from the main valve, thereby keeping the control pilot in an accessible location away from possible pipeline shock and vibration. Such location of the remote pilot also simplifies its maintenance.

Figure 7.7o illustrates an internally piloted regulator operated by a remote set-point air signal. This design offers the convenience of remote set-point adjustment at distances up to 500 ft (150 m), but the requirement for compressed air makes it more expensive to operate than the use of spring-loaded regulators.

Table 7.7p provides some data that can help the design engineer in deciding when to use spring-loaded and when to select pilot-loaded regulators for a particular application.

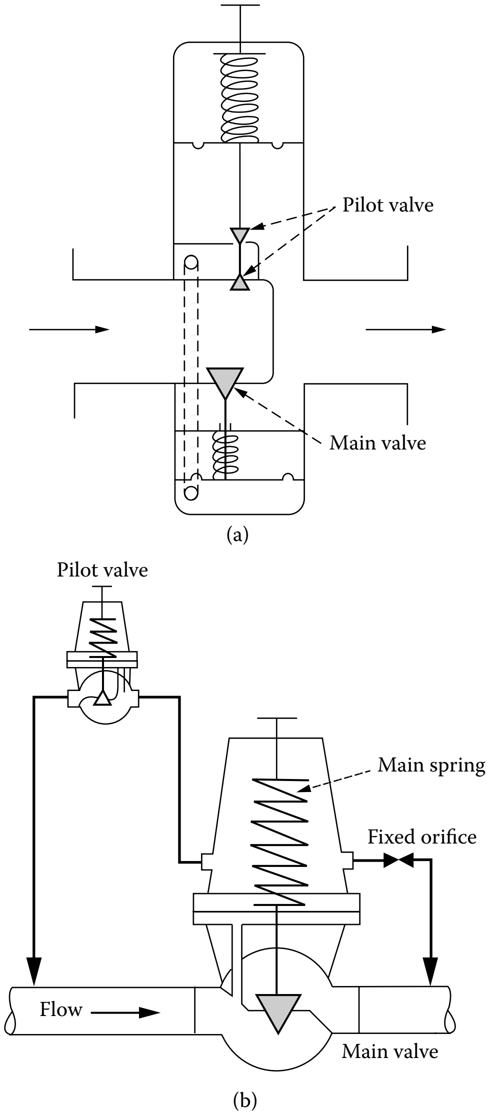


FIG. 7.7n
Pilot-operated regulators can be provided with integral or external pilots.

REGULATOR CHARACTERISTICS AND SIZING

Seating and Sensitivity

As shown in Figure 7.7i, the balance of forces on the valve stem is

$$kx = A_p(p_1 - p_2) - A_d p_3 \qquad \mathbf{7.7(1)}$$

where

k = spring constant (lb$_f$/in. or N/m)
x = valve stem movement (in. or mm)
A_p = port area (in.^{2} or mm^2)

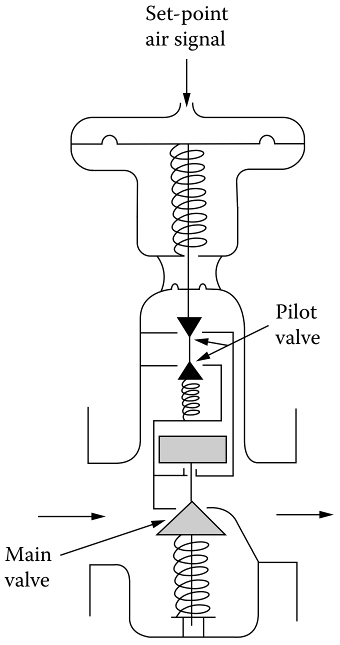


FIG. 7.7o
Remote set point provided by pneumatic loading of an internally piloted regulator valve.

p_1 = inlet pressure (PSIG or Pa)
p_2 = outlet pressure or regulated pressure (PSIG or Pa)
p_3 = diaphragm back-pressure (PSIG or Pa)
A_d = diaphragm area (in.^{2} or mm^2)

Inserting into Equation 7.7(1) the typical values of a typical regulator application will likely result in a relatively large product of $A_p(p_1 - p_2)$ as compared to $A_d p_3$. In other words, the actuating force is comparatively low, and a balanced plug is desirable because it eliminates the term $A_p(p_1 - p_2)$.

TABLE 7.7p
Regulator Selection for Various Service Conditions

Recommended Regulator Type	*Magnitude of Pressure Reduction*	*Supply Pressure Variations*	*Load Variations*
Spring-loaded	Moderate	Small	Moderate
Spring-loaded	Moderate	Large	Moderate
Pilot-operated	Moderate	Large	Large
Pilot-operated	Large (in one stage)	Moderate	Large
Pilot-operated first stage	Large (in two stages)	Large	Large (in first stage)
Spring-loaded or pilot-operated	Large (in two stages)	Large	Moderate (in second stage)

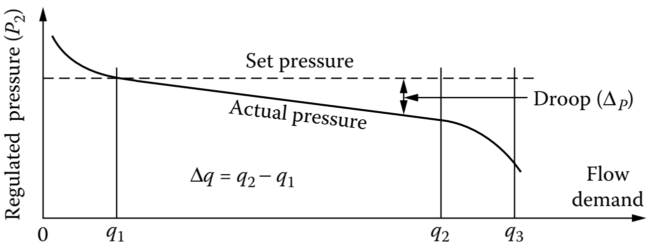


FIG. 7.7q
The actual value of the regulated pressure equals the set pressure only at minimum flow (q_1), and once the flow rises above it, the actual controlled pressure drops. This "droop" or offset is permanent and increases with flow.

Figure 7.7j shows an essentially balanced, double-seated plug valve design, which is limited to continuous service because it cannot provide tight shut-off.

Figure 7.7k shows a sliding gate mechanism that is also balanced, but this design is not suited for applications with high pressure drops. Analysis of the regulator as a feedback mechanism shows that the balanced plug reduces sensitivity to variations in supply pressure.

Intermittent service requiring tight shut-off necessitates using a single-seated valve. Balancing pistons, diaphragms, and ingenious seating configurations can be used to provide a balanced plug in a single-seated design, but these devices are more complex and costly. In valves with port diameters of 1 in. (25 mm) or less, the ratio of diaphragm area to port area can be made large enough so that balancing is not required for moderate pressure drops.

Droop or Offset

The regulator is a complete, self-contained feedback control loop with proportional-only control action. Thus, the regulated pressure will be offset by changes in the load disturbance variables (flow demand, in the case of a pressure-reducing valve). The offset in regulated pressure with changing flow is called droop. It can be expressed as follows:

$$\Delta p = \frac{(kx/q)(p_1^2 - p_2^2)}{kxp_2 + A_d(p_1^2 - p_2^2)} \Delta q \qquad \textbf{7.7(2)}$$

where

q = throughput and the other letters are as defined in connection with Equation 7.7(1).

Equation 7.7(2) shows that the droop or offset is a function of the operating pressures and of the regulator design parameters such as spring rate k, valve lift x, and diaphragm area A_d.

Size and economic considerations limit the length of the spring (for low spring rate) and also limit the diaphragm area. Low lift (x) is necessary to reduce diaphragm fatigue as well as minimize droop. Figure 7.7q illustrates the phenomenon of droop.

Equation 7.7(2) applies to the flow range of q_1 to q_2 in Figure 7.7q. This is the operating span of the regulator. While ideally linear, the operating span takes on many shapes in practice because of valve plug flow characteristics and varying effective diaphragm areas. The valve is fully open at q_2 and acts as a fixed orifice restriction from q_2 to q_3. The span from 0 to q_1 is dominated by flow-generated pressure forces on the plug, and q_1 is considered to be the minimum controllable flow rate. The minimum flow rate is a function of plug design and is typically 5–10% of maximum capacity q_2.

Therefore, a pressure regulator cannot control pressure at a set value. All it can do is travel on its operating curve (shown in Figure 7.7q) and provide a pressure that this curve dictates as a function of flow. In order to provide tight shutoff, extra force is needed and therefore the pressure difference on the diaphragm must rise. Consequently, at near-zero flow the regulated pressure will rise.

What the manufacturers call the "set point" of the regulator is in fact is the pressure at minimum flow (q_1). To call this value the set point is not correct, because this "set point" varies with flow (Figure 7.7q). A more accurate designation would be to call this point the "zero droop" (zero offset) point or to call it the upper end of the pressure throttling range, because as the flow increases the actual controlled pressure drops.

From Figure 7.7q, the reader should conclude that the maximum regulator capacity is not at full valve opening (q_2), but at the minimum pressure, which the process can tolerate as the droop increases with flow. Reliable data on droop vs. flow is therefore essential to check if regulator performance will be satisfactory.

Minimizing Droop Regulator designs that minimize droop are available. Many of them place the feedback sensing line at a point of high velocity, either in the throat of a slight restriction or in the middle of the flowing fluid (Figure 7.7k). The latter design uses the aspirating effect.

Droop compensation can also be provided by a moving valve seat, which is called a pressure-compensating orifice, that moves with upstream pressure. The roll-out diaphragm that reduces its effective area with spring compression also acts to reduce droop. These devices allow the specification of the droop curve for limited flow spans.

Noise

High velocities of compressible fluids are primarily responsible for noisy regulator installations, a continuing source of complaint. Aerodynamic noise caused by high-velocity gas flow in the valve body has been identified as the principal source of irritating "screaming." The volume of the gas increases in proportion to the ratio of inlet to outlet pressure. Gas velocity increases in proportion to the volume increase if the cross-sectional area of the regulator outlet port does not also increase.

Increased velocity raises the pressure drop until sonic velocity is reached. A tapered pipe expander downstream from the valve will increase the sonic velocity (Mach 1) to supersonic velocities (Mach 2 to Mach 3). Velocities above sonic result in noise, because a major portion of the energy conversion takes place downstream of the valve plug. The static pressure build-up in the downstream piping can far exceed its pressure rating and can cause "choking" (reduced capacity) of the regulator due to the downstream static pressure.

Sonic velocity is known to occur when downstream pressure is less than 50% of upstream pressure. Because changes in flow path direction also contribute to noise, a maximum gas velocity of 200 ft/s (60 m/s) (below ground, 500 ft/s, or 150 m/s) has been recommended. Gas velocity in regulators is difficult to calculate because of interaction between velocity, pressure drop, and cross-sectional area. Many manufacturers provide tables of maximum capacities for quiet operation.

Calculating Velocity In order to avoid noise, the gas velocity should be kept under the sonic velocities at the regulator outlet. Because velocity is difficult to calculate (see Section 6.14), sonic velocity can be avoided by limiting pressure reduction to less than the critical ratio of 2:1.

Higher pressure ratios can be handled if several regulators are installed in series. Two to three stages are common to reduce noise and to improve regulation. In these installations a slightly less than critical reduction is made in the second of the two stages, or in the second and third of three stages, with the remainder of the reduction taking place across the first stage.

Eliminating changes in flow direction will also reduce noise. Therefore, it is recommended to use as much straight pipe as practical on both sides of the regulator. The use of mufflers can also be considered.

Water hammer can also cause noise in regulators on liquid service when the valve is opened or closed too quickly. When a valve closes quickly, a pressure wave is generated as the moving water column hits the closing plug. This pressure wave travels upstream at sonic speed and is reflected by the first solid surface with which it comes in contact. If, by the time this pressure wave returns, the valve is closed, vibration results.

Water hammer is eliminated by slowing valve closure or by reducing the distance (L) between the reflecting surface and the valve so that the valve closure time (t) is more than $2L$ divided by the speed of sound.

Cavitation

Cavitation is caused by localized vaporization as the pressure at the vena contracta drops below the vapor pressure of the flowing fluid. As the pressure recovers after the vena contracta, these vapor bubbles collapse and can cause vibration damage and valve-metal erosion.

The energy for the increase in flow velocity at the valve's vena contracta is obtained from the pressure (potential energy) of the flowing fluid, causing a localized pressure reduction. If this localized pressure falls below the vapor pressure of the fluid, it causes temporary vaporization.

Cavitation damage always occurs downstream of the vena contracta, where the pressure recovery occurs in the valve. It is in this area where the vapor bubbles collapse. Destruction is due to the implosions that generate the extremely high-pressure shock waves in the substantially incompressible stream (Figure 6.1w). When these waves strike the solid metal surface of the valve or downstream piping, the damaged metal surface will give a cinder-like appearance. Cavitation is usually coupled with vibration and makes a sound like that of rock fragments or gravel flowing through the valve.

Cavitation can be eliminated by altering either the valve, the process, or the installation. The various anticavitation valve designs such as the labyrinth (Figure 6.1y) are usually not available for regulators. Therefore, cavitation in regulators must be eliminated by changing either the process or the installation.

Anticavitation Designs A slight reduction of operating temperature can usually be tolerated from a process point of view, and this might be sufficient to lower the vapor pressure sufficiently to eliminate cavitation. Similarly, increased upstream and downstream pressure, with ΔP unaffected, can also relieve cavitation. Valves subject to cavitation should be installed at the lowest possible elevation in the piping system. Moving the valve closer to the pump will also serve to increase both upstream and downstream pressures.

If cavitating conditions are unavoidable from the process condition's point of view, then it is actually preferable to have not only cavitation but also some permanent vaporization (flashing) through the valve. This can usually be guaranteed by a slight increase in operating temperature or by decreasing the outlet pressure. Flashing eliminates cavitation by eliminating pressure recovery.

Another way to eliminate cavitation is to install two or more regulators in series. Cavitation problems can also be alleviated by absorbing some of the pressure drop in breakdown orifices, chokes (Figure 6.1cc), or partially open block valves upstream of the pressure-reducing valve. The amount of cavitation damage is related to the sixth power of flow velocity or to the third power of pressure drop. This is the reason why reducing ΔP by a factor of two, for example, will result in an eightfold reduction in cavitation destruction.

In some high-pressure let-down stations, it might not be possible to eliminate cavitation accompanied by erosion or corrosion, or both. In such installations, one ought to consider the use of inexpensive choke fittings (shown in Figure 6.1cc) instead of control valves. The fixed chokes can be of different capacities and be isolated by full-bore on/off valves, providing a means of matching the process flow with the opening of the required number of chokes. If *n* chokes are installed, this will allow for operating at (2^n-1) flow rates. If, for example, the chokes discharge into the vapor space of a tank, this will minimize cavitation damage because the bubbles will not be collapsing near any metallic surfaces.

Sizing and Rangeability

Oversizing is the most common error in regulator selection. The droop characteristic makes a larger valve attractive, because a greater capacity is obtainable for the same droop. The larger valve also reduces noise because of its larger passages. These apparent advantages are offset by higher cost, severe seat wear, and poor regulation.

The limitation on sizing is rangeability. Rangeability varies from 4 to 1 for a steam regulator, which cannot be operated close to its seat because of wire drawing, to over 50 to 1 for an air regulator. Figure 7.7q illustrates rangeability (q_2/q_1). Minimum flow, q_1, is 5–10% of q_2, depending on the seat configuration.

Maximum flow is not necessarily q_2 but is determined by maximum acceptable droop. A typical rangeability is 10 to 1. An oversized regulator may not be able to control at the minimum flow because it falls outside its rangeability.

Regulators are sized on the basis of tabulated data or valve coefficients (C_v) provided by the manufacturer. Size is chosen to accommodate the maximum flow at minimum pressure drop. The valve should ideally operate at 50–60% open under normal conditions. Catalog information must be used judiciously, because capacity and droop may be specified at different points. Rated capacity might result in too high a velocity, or the (external) pressure feedback tap might have been located at a different point during testing than in the application.

Rangeability is increased by using two regulators in parallel. In such installations the pressure setpoint of one regulator is set 10% higher than the other, so one regulator will be wide open while the other modulates under high load conditions. As demand is reduced, the second regulator will close and the first will modulate. Leakage in the second regulator must be a minor portion of the capacity of the first regulator.

Stability

Stability of the regulator installation depends on the open loop gain. For the regulator shown in Figure 7.7i, open loop gain is defined as

$$K_o = \frac{A_d(p_1^2 - p_2^2)}{kxp_2} \qquad \textbf{7.7(3)}$$

where K_o is the open loop gain. Comparing this equation with the one given earlier for droop, it is apparent that any changes that might decrease droop would increase the open-loop gain and decrease stability, as expected in a feedback mechanism. Because the regulator has no controller mode adjustments, the maker must choose design parameters that will provide an adequate compromise between droop and stability. It can thus be expected that the open-loop gain of some applications will be too high, resulting in a noisy and cycling pressure regulator.

Guidelines for a stable installation are few. Difficulties are generally found after installation. Because the regulator has no adjustments it is costly to correct problems. The following steps can be considered in stabilizing an installed regulator: (1) Relocation of the pressure-sensing tap. (2) Redesign of the downstream piping to provide more volume. (3) Restricting the pressure feedback line, either by a needle valve in an external line or by filling an internal line and redrilling it to provide a smaller hole.

Safety

Diaphragm rupture is the most common failure in a regulator. Most regulators fail full open upon diaphragm failure, which is usually an unsafe condition. A relief valve on the regulated pressure side is recommended if it is unsafe to allow the increase in regulated pressure to the level of the supply pressure.

In installations where it is imperative that the user continue to be supplied even on a regulator failure, two regulators can be placed in series. In such installations, the second regulator should have a set point that is 10% higher than the first, so that it will remain wide open under normal circumstances. If the first regulator fails, the second will take over pressure regulation.

Installation

Regulator installation is generally easier than regulator selection. The following installation suggestions will help ensure satisfactory regulator performance:

1. Steam regulators should be preceded by a separator and a trap.
2. All regulators should be preceded by a filter or strainer.
3. A valve bypass is recommended where it is necessary to service the regulator while continuing to supply users.
4. Use straight lengths of pipe upstream and downstream to reduce noise.
5. External feedback lines should be 1/4 in. (6.25 mm) pipe or tubing.
6. Locate the pressure feedback tap at a point where it will not be affected by turbulence, line losses, or sudden changes in velocity. A distance of 10 ft (3 m) from the regulator is recommended.

Bibliography

Adams, R. M., "Don't Overlook Simple Regulators," *InTech,* April 1983.
Baumann, H. D., "Why Limited Velocities in Reducing Valves?" *Instruments and Control Systems,* September 1965.
Beard, C. S., *Final Control Elements,* Philadelphia: Rimbach Publications Division, 1969.
Beard, C. S., and Marton, F. D., *Regulators and Relief Valves,* Philadelphia: Chilton Book Co.
Comber, J., and Hockman, P., "Pressure Monitoring: What's Happening?" *Instruments and Control Systems,* April 1980.
"Control Valves—Regulator/Safety/Relief/Check," *Measurements and Control,* June 1993.
Dickinson, W., "Don't Overlook Pressure Regulators," *Instruments and Control Systems,* May 1983.
Driskell, L. R., *Control Valve Selection and Sizing,* Research Triangle Park, NC: ISA, 1983.
Fleming, A., "Regulator Station Design," *Instruments and Control Systems,* April 1961.
Fluid Controls Institute, "Definitions of Regulator Capacities," FCI-58-1, April 1966.
Fluid Controls Institute, "Guide to Material Selection for Industrial Regulators," FCI-65-1.
Hall, J., "Monitoring Pressure with Newer Technologies," *Instruments and Control Systems,* April 1979.
Kubitz, R., "Valves in Gas Regulators," *Instruments and Control Systems,* April 1966.
Law, S., "Measurement and Instrumentation: Automating District Regulators," *Gas Industries,* January 2003.
Ledeen, H., "A New Look at Valve Regulation," *Instruments and Control Systems,* August 1965.
Lipták, B. G., "A Close Look at Pressure Regulators," *Chemical Engineering,* April 13, 1987.
Moore, R. L., "The Use and Misuse of Pressure Regulators," *Instrumental Technology,* March 1969.
1992–94 Catalog of U.S. & Canadian Valves & Actuators, Washington, D.C.: Valve Manufacturers Association of America.
North American Sourcebook of Valves, Actuators & Control Manufacturers, 6th ed., Washington, D.C.: Valve Manufacturers Association of America, 2004.
O'Connor, J. P., "12 Points to Consider when Selecting Pressure Regulators," *Plant Engineering,* March 1958.
Perrine, E. B., "Pressure Regulators and their Application," *Instruments and Control Systems,* September 1965.
Quentzel, D., "A Balance Factor for Pressure Reducing Valves," *Control Engineering,* May 1965.
"Selecting and Sizing Pressure Reducing Regulators," in *Fisher Regulator Handbook,* 2nd ed., Fisher Controls Regulator Division, McKinney, TX, 2001.
Spitzer, D. W., "Remotely Set Pressure Regulators," *InTech,* November 1992.
Watts, G. R., "Pressure Regulators or Control Valves?" *Instrumental Technology,* June 1980.
Wilmer, J., Orzechowski, B., and Mackenzie, D., "Component and Design Optimization, Process Measurement and Control for Colloidal CMP Slurries," ISA EXPO 2003, Houston, TX, October 2003.
Yedidiah, S., "Effect of Seat Configuration on Pressure Operated Control Valves," *Design News,* October 11, 1967.

TABLE 7.7s *ISA Form 7.7s*

	PRESSURE CONTROL VALVES PILOTS and REGULATORS			SHEET ____ OF ____	
				SPEC. NO.	REV.
	NO	BY	DATE	REVISION	
				CONTRACT	DATE
				REQ. P.O.	
				BY	CHK'D	APPR.

GENERAL	1.	Tag No.					
	2.	Service					
	3.	Line					
	4.	Line Size/Sched. No.					
	5.	Function					
BODY	6.	Type of body					
	7.	Body Size	Port Size				
	8.	Guiding	No. of Ports				
	9.	End Conn. & Rating					
	10.	Body Material					
	11.	Packing Material					
	12.	Lubricator	ISO. Valve				
	13.	Seal Type					
	14.	Trim Form					
	15.	Trim Material					
	16.	Seat Material					
	17.	Required Seat Tightness					
	18.	Max. Allow Sound Level dBA					
ACTUATOR/ PILOT	19.	Type of Actuator					
	20.	Pilot					
	21.	Supply to Pilot					
	22.	Self Cont.	Ext. Conn.				
	23.	Diaphragm Material					
	24.	Diaphragm Rating					
	25.	Spring Range					
	26.	Set Point					
	27.						
ACCESSORIES	28.	Filt. Reg.	Supply Gage				
	29.	Line Strainer					
	30.	Housing Vent					
	31.	Internal Relief					
	32.						
	33.						
SERVICE	34.	FLOW UNITS		LIQUID	STEAM	GAS	
	35.	Fluid					
	36.	Quant. Max.	C_v				
	37.	Quant. Oper.	C_v				
	38.	Valve C_v	Valve F_L				
	39.	Norm. Inlet Press.	ΔP				
	40.	Max. Inlet Press.					
	41.	Max. Shut Off	ΔP				
	42.	Temp. Max.	Operating				
	43.	Oper. sp. gr.	Mol. Wt.				
	44.	Oper Visc.	%Flash				
	45.	% Superheat	%Solids				
	46.	Vapor Press.	Crit. Press.				
	47.	Predicted Sound Level dBA					
	48.	Manufacturer					
	49.	Model No.					

Notes:

Instruments, Primary

Reprinted by permission. Copyright © 1981, Instrument Society of America. From ISA Standard S20, "Specification Forms for Process Measurement and Control Instruments, Primary Elements and Control Valves."

ISA SPECIFICATION FORM (TABLE 7.7s)

Compiling the information necessary to specify a regulator is best done with the aid of a tabulation sheet. Many large companies have their own customized forms. A general-purpose form standardized by the Instrumentation, Systems and Automation Society, formerly the Instrument Society of America (ISA) is provided below with the instructions for completing ISA Form S20.51. The ISA SP20 Committee is currently updating all of the S20 Instrument Specification Forms.

Instructions for ISA Form S20.51

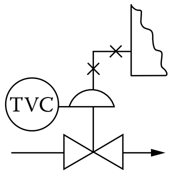
1–4. Identification and service/location. It is assumed that each tag number is for a single valve.
5. Pressure reducing, back-pressure control, or differential pressure regulator.
6. Globe, or Manufacturer's Standard (MFR. STD.).
7. Body connection size and inner valve size.
8. Guiding may be top, top and bottom, skirt, or MFR. STD. Select single or double port, if applicable.
9. Specify screwed (NPT), flanged, or weld end, and flange rating, such as 150 lb ANSI.
10–11. Specify materials.
12. Write in "yes" or use checkmark if required.
13. Quick open, equal percent, linear, etc.
State Characteristic:

L = Linear
LV = Linear V Port
EP = Equal Percentage
EPT = Equal Percentage Turned
EPB = Equal Percentage Balanced
Q = Quick Opening
Or use your own code and identify in notes.

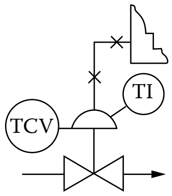
14. Refers to seal between body and top works, such as diaphragm, stuffing box, etc.
15. Refers to seat, plug, stem; in general, all internal wetted parts.
16. Use only to specify soft seat; otherwise material will be same as trim specified in line 14.
17. Use if required.
18. Max. allowable sound level dBA 3 ft from pipe and 3 ft downstream of the valve outlet.
19. Actuator may be spring type or springless pressure balanced.
20. The pilot is an integral or external auxiliary device that amplifies the force available through an operating medium, usually air.
21. Give pressure available and specify medium.
22. Refers to valve pressure sensing system. Specify whether controlled pressure is sensed internally or by means of an external line requiring an additional piping connection.
23–24. Specify diaphragm material and pressure or temperature limits, if applicable.
25. Range over which pressure setting can be made.
26. Specification of set pressure does not apply to factory setting. This must be called for specifically, if required.
27. Specify filter regulator, with or without gage, if required for air supply to pilot. Write "yes" or use checkmark.
28–29. Specify if strainer is to be furnished with valve. Write "yes" to check off, or give style or model number.
30–33. Options available in gas regulators. On line 30, specify "bug-proof" if required. Specify accessories on lines 32–33.
34. State liquid, steam, gas units (gpm, lb/hr, ft^3/min., etc.)
35. Name of fluid and state, whether vapor or liquid, if not apparent.
36. State maximum quantity required by process and corresponding C_v.
37. State operating quantity required by process and corresponding C_v.
38. The manufacturer shall fill in the valve C_v and F_L (liquid pressure) recovery factor without reducers or other accessories.
39. Operating inlet pressure and pressure differential with units (psia, PSIG, inches H_2O, or Hg). Note at this point that one might consider how minimum conditions will fit the sizing.
40. Maximum inlet pressure if different from normal.
41. State the maximum pressure drop in shut-off position to determine proper actuator size. This is actual difference in inlet and outlet pressure stated in psi, inches of H_2O or H_g, etc.
42. State °F or °C.
43. State operating specific gravity and molecular weight.
44. State operating viscosity and its units. State flash at valve outlet, i.e., of max. flow that will be flashed to vapor because of the valve pressure drop.
45. In the case of vapors, state the superheat, and in the cases of liquids, state the solids, if present.
46. Note vapor pressure of fluid as well as the critical pressure.
47. Give manufacturer's predicted sound level dBA.
48. Complete when available.

7.8 Regulators—Temperature

L. S. DYSART (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Blind temperature regulator



Temperature regulator with direct indicator

Flow sheet symbols

Size:	From 1/4 to 2 in. (6.25 to 50 mm) standard, special up to 6 in. (150 mm)
Features:	Two- or three-way designs with direct or reverse action, only proportional mode of control and the only adjustment capability being the changing of set point. Bulb pressure may exceed 1000 PSIG (6.9 MPa)
Ranges:	Overall: –25 to 480°F (–31.7 to 248.9°C) Individual: 1) 10 to 65°F (–12 to 18°C) 2) 40 to 100°F (4 to 38°C) 3) 70 to 145°F (21 to 63°C) 4) 110 to 170°F (43 to 77°C) 5) 130 to 190°F (54 to 88°C) 6) 170 to 240°F (77 to 116°C) 7) 220 to 275°F (104 to 135°C) 8) 270 to 340°F (132 to 171°C)
Spans:	Standard from 40 to 70°F (22.2 to 38.8°C); special up to 100°F (55.6°C)
Materials of Construction:	Body in iron, bronze, steel, or stainless steel; bulb in copper, steel, stainless steel, or coated with PVC, Teflon, and so on, tube in copper or stainless steel
Tubing Length:	Usually 10 ft (3 m) with 50 ft (15 m) being the maximum
Base Cost:	\$500 for standard unit in 1/2 in. (12.5 mm) size; \$1200 for standard unit in 2 in. (50 mm) size. For larger sizes and carbon or stainless steel materials, refer to the costs of globe valves given in Figure 6.19b in Chapter 6.
Partial List of Suppliers:	ABB Inc.-Instrumentation (www.abb.com/us/instrumentation) Dwyer Instruments (www.dwyer-inst.com/htdocs/valves) Enercorp Instruments Ltd. (www.process-controls.com/enercorp/) Eurotherm Controls Inc. (www.eurotherm.com) Fenwal Inc. (www.gaumer.com) Honeywell Sensing and Controls (www.honeywell.com/sensing) Jordan Controls (www.jordancontrols.com) Jumo Process Control Inc. (www.jumousa.com) Leslie Controls Inc. (www.lesliecontrols.com) Ogontz Controls Co. (www.ogontz.com) Powers Process Control (www.powerscontrols.com)

Robertshaw IPD, an Invensys Co. (www.robertshawindustrial.com)
Spence Engineering Co. (www.spenceengineering.com)
Sporlan Valve Co. (www.sporlan.com)
Trerice (www.trerice.com)

INTRODUCTION

Because they require no external power supplies (electricity, air, and so on), regulators are described as "self-actuated." They actually "borrow" thermal energy from the controlled medium to obtain the required forces. Temperature regulators are self-contained mechanical devices that provide fixed, narrow band proportional control action. They are referred to as temperature control valves (TCVs). Their operation is similar to that of a thermostat throttling a temperature control valve, except that the throttling control feature is part of their actuator.

This means that they are not capable of maintaining a set point, because in order to change the valve opening, first the measured temperature has to change. Therefore, all they can do is follow an operating line that relates the valve opening (load) to the controlled temperature (measurement). This phenomenon has been discussed in detail in connection with pressure regulators (Figure 7.7q) and thermostats (Figure 7.9c) and therefore will not be repeated here.

The difference between the temperature setting and the actual process temperature is called the "offset" of the regulator. Because of this offset, the resulting temperature control is not very accurate, and therefore these regulators are used only in such secondary applications as HVAC.

The temperature is detected by measuring the thermal expansion of the filling material in the bulb, which is inserted into the controlled process. Table 7.8a lists the thermal expansion coefficients of a variety of substances at 20°C (64°F). These values vary with temperature.

TABLE 7.8a
Thermal Expansion Coefficients of Different Substances (Values are given in cubic units, giving the incremental change of volume of a unit volume in response to a 1°C increase in temperature)

Substance	*Cubic Expansion* ($\times 10^{-4}$)	*Substance*	*Cubic Expansion* ($\times 10^{-4}$)
Acetone	14.9	Mercury	1.8
Air	36.7	Nickel	0.13
Aluminum	0.23	Nitrogen	36.7
Brass	0.19	Oxygen	36.7
Carbon dioxide	37.3	Platinum	0.089
Chloroform	12.8	Quartz	0.006
Copper	0.15	Steel	0.12
Ethyl alcohol	11.2	Tin	0.27
Ethyl ether	16.6	Toluene	11.2
Glass (regular)	0.085	Tungsten	0.035
Glass (Pyrex)	0.036	Turpentine	9.7
Glycerine	5.0	Water	2.1
Hydrogen	36.6	Xylene	11.3
Invar	0.009	Zinc	0.26

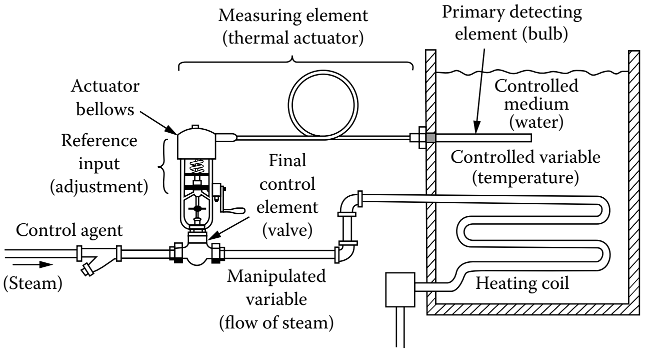


FIG. 7.8b
The main components of a self-actuated temperature regulator.

As shown in Figure 7.8b, the temperature regulator consists of three main components: 1) the temperature detecting element (bulb), 2) the measuring element (thermal actuator) and the reference input (adjustment), and 3) the final control element (valve). These three components will be discussed in that order.

Control Quality

Only the proportional mode is available in all temperature regulators. The proportional band is not adjustable in the field but is determined by various design factors, and it generally varies in direct proportion to the valve size and lift. For example, a typical 1 in. (25 mm) vapor pressure actuated regulator, using a double-seated valve, has a mid-range throttling range (full open to full closed) of 7°F (3.9°C), and a similar 2 in. (50 mm) regulator has a throttling range of 11°F (6.1°C).

The relationship between proportional band, gain, and throttling range is explained in Section 7.9, in connection with Figure 7.9c. Because thermostats and regulators do not have a set point but only an operating line, therefore the controlled temperature changes as the valve opens or closes. The maximum error—the maximum offset—is half the throttling range.

Response time of temperature regulators is much longer than that of pneumatic or electronic control loops. TCV time constants vary widely with design factors such as actuator bellows size (area), valve size (lift), transmission tube length, bulb size, and material (heat conductivity). The time constant of a typical regulator is likely to be in the range of 30–90 sec. Although the time constant can be reduced to under 10 sec, this can only be done under favorable conditions.

TYPES OF TEMPERATURE REGULATORS

Temperature regulators are usually depicted on a flowsheet by a TCV symbol, emphasizing the fact that most of these units are blind. A regular temperature control loop (depicted as TIC or TRC on the flowsheet) is superior not only because it indicates the measurement (which the TCV can also do, but providing indication is a special feature), but also because TICs are not limited to plain proportional control and therefore are capable of maintaining their set points without offset.

Two types of TCV regulators are available: direct-actuated and pilot-actuated. They are distinguished by the way in which the valve (the final control element) is actuated.

Direct- and Pilot-Actuated

In the direct-actuated type, the power unit (bellows, diaphragm, and so on) of the thermal actuator is directly connected to the valve plug and develops the force and travel necessary to fully open and close the valve (Figure 7.8c). Direct-actuated regulators are generally simpler, lower in cost, and more truly proportional in action (with somewhat better stability).

In the pilot-actuated design, the thermal actuator moves a pilot valve, internal or external (Figure 7.8d). This pilot controls the amount of pressure energy from the control agent (fluid through valve) to a piston or diaphragm, which in turn develops power and thrust to position the main valve plug. The pilot may be internal or external. When external, independently acting multiple pilots are also available.

Compared to direct-actuated TCVs, pilot-actuated regulators have smaller bulbs, faster response, and narrower proportional band, and they can handle higher pressures through the valve. Pilot-actuated regulators can also handle interrelated functions through use of multiple pilots, such as temperature plus pressure plus electric interlocks.

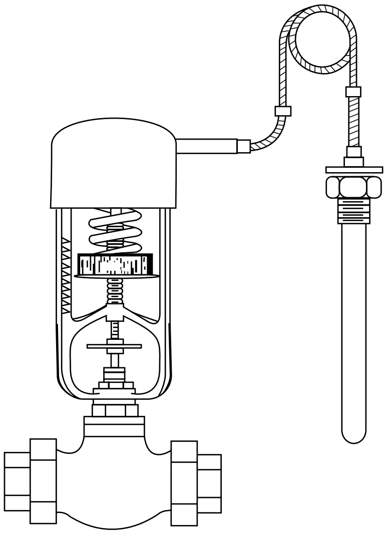


FIG. 7.8c
Direct-actuated temperature regulator.

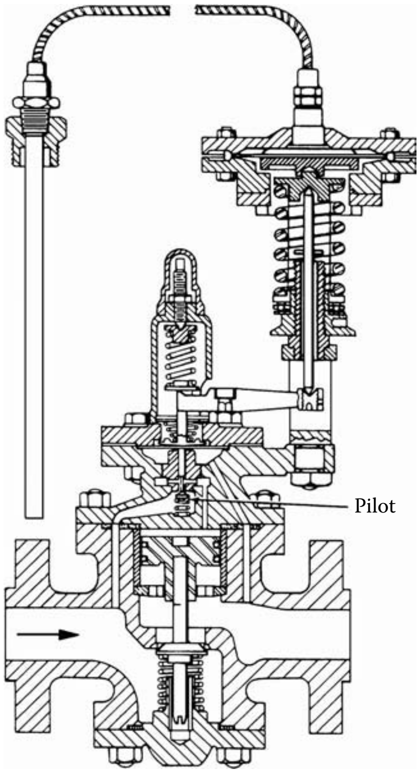


FIG. 7.8d
Pilot-actuated temperature regulator.

There are two styles of temperature regulators: self-contained and remote-sensing. Which one is appropriate depends on the location and structure of the measuring element (thermal actuator).

Remote or Internal Element

In remote-sensing regulators, the bulb (the primary detecting element) is separate from the power element (bellows and so on) of the thermal actuator and is usually connected to it by flexible capillary tubing (Figures 7.8b and 7.8c). This type of TCV measures and regulates the temperature of the process by throttling the heating or cooling medium (the control agent).

Self-contained regulators contain the entire thermal actuator within the valve body, and the actuator serves as the primary detecting element (Figure 7.8e). Thus, the self-contained style can sense only the temperature of the fluid that is flowing through the valve. That fluid is both the controlling agent and the controlled medium. The device regulates the temperature of the fluid by regulating the fluid's flow. Self-contained regulators are generally provided with liquid expansion or fusion-type thermal elements.

One common application for an internal sensor-type temperature regulator is illustrated in Figure 7.8f. Here the goal is to keep the tracing tubes full of steam while draining off only the condensate. The internal sensor-type TCV valve guarantees this, because if it is set at 200°F (93°C), it will open

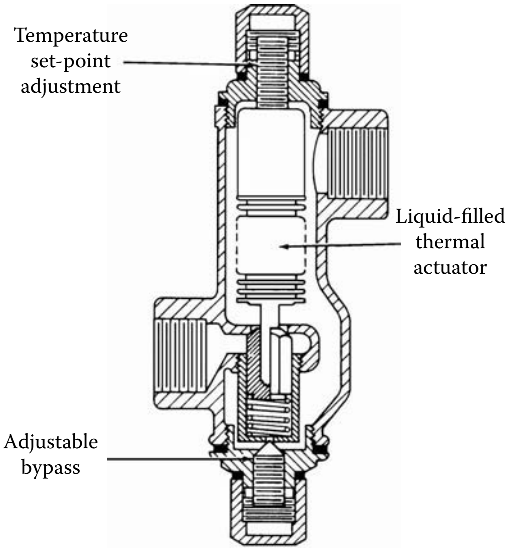


FIG. 7.8e
Internal element type self-contained temperature regulator.

only when it detects a lower temperature, thereby releasing condensate, and will close when the temperature rises above that setting, thereby preventing the release of steam.

The self-contained style is simpler, frequently lower in cost, and "packless" (has no stem sliding through the valve body "envelope"), but it is limited to such uses as regulating the temperature of coolant (water, and so on) that is leaving the engine, compressor, and exothermic process.

Accessories and Special Features

Temperature indication is available with most vapor pressure actuated regulators (Figure 7.8o). It usually consists of a pressure gauge sealed in the thermal system and calibrated to correspond to the particular range (charging liquid) used. Inaccuracy is about ±2% of indicated span.

Fail-safe design is available as a special version of the vapor pressure regulator. Here failure (loss of fill) of the thermal actuator produces downward movement of the valve stem; i.e., the valve goes to the "safe" position, producing shut-off of the heating medium or full flow of the cooling medium, depending on valve action. Adjustable ranges are short (approximately 30°F, or 16.7°C), and valve actuating force is rather low, because the actuator works on the subatmospheric portion of the nonlinear vapor pressure curve of the filling medium.

Manual handwheels are available in a few regulator versions (Figure 7.8o). These are useful to manually override the element, to provide a minimum valve position, or for emergency operation.

Combination temperature/pressure control is available in the pilot-actuated regulator. This is accomplished in two ways. In the external pilot version (Figure 7.8d), two separately adjusted pilot units are piped in series in the pilot pressure line. The temperature pilot modulates the flow of heating medium (control agent) to regulate the process temperature, while the pressure pilot limits the discharge pressure from the main valve to its setting under maximum load conditions.

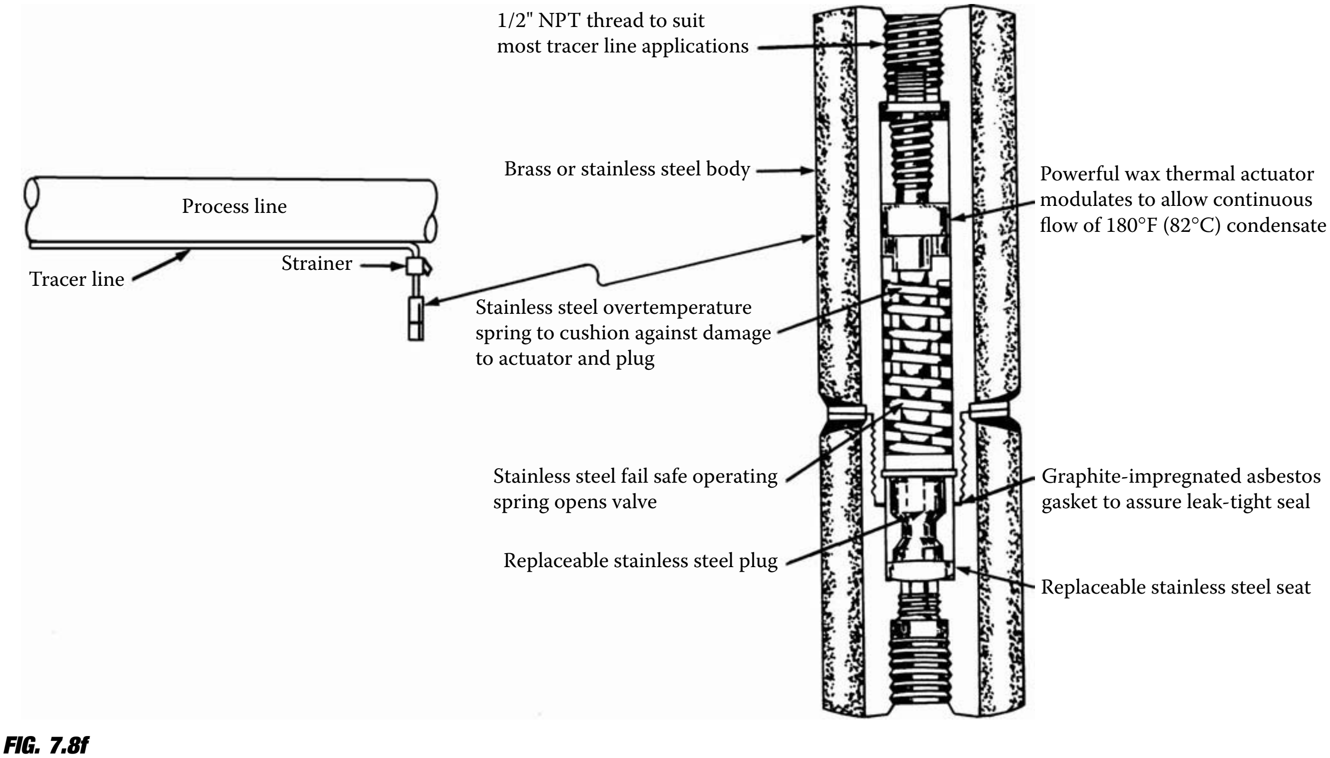


FIG. 7.8f
A temperature regulator can be used in place of a steam trap. It will guarantee that the tracer line is always full of steam and only the condensate is drained, because it opens at a low enough temperature where steam cannot exist. (Courtesy of Ogontz Controls Co.)

In the internal pilot version, the two pilot elements are cascaded; i.e., the temperature element mechanically changes the setting of the pressure element from zero at no load to the maximum pressure for which the pressure pilot is set at full load.

THERMAL SYSTEMS

There are four classes of "filled system" thermal actuators that are used in temperature regulators. They all develop power and movement proportional to the measured temperature, and hence temperature regulators are proportional controllers. For a detailed discussion of filled thermal systems, refer to Section 4.6 in the *Process Measurement* volume of this handbook.

Gas-filled systems (SAMA Class III), bimetal, and "differential expansion" thermal elements are usually not used in temperature regulator valves.

Bulbs, Wells, Fittings

The temperature-sensitive element, the bulb, comes in many sizes and shapes to handle the many different applications. It is good practice to use the largest bulb that will do the job. This will cut down on ambient temperature errors and permit smaller spans and longer capillaries.

Plain bulbs are used where the measured medium is not under pressure and will not harm the bulb material. If this is not the case then a separate well to protect the bulb from the process medium is needed. This, of course, will slow down the response time even more than the 4–7 sec speed of response range.

Speed of Response The speed of response generally doubles with the doubling of bulb diameter and tends to be the fastest with vapor or gas and the slowest with liquid filling, with mercury and Class IID vapor filling in the middle. When thermowells are added to the bulb, speed of response slows as follows: 12–24 sec for 0.25 in. (6 mm), 20–35 sec for 3/8 in. (9 mm), 25–50 sec for 0.5 in. (13 mm), and 40–75 sec for 3/4 in. (19 mm) diameters. Figure 7.8g illustrates a filled bulb installed in a thermowell.

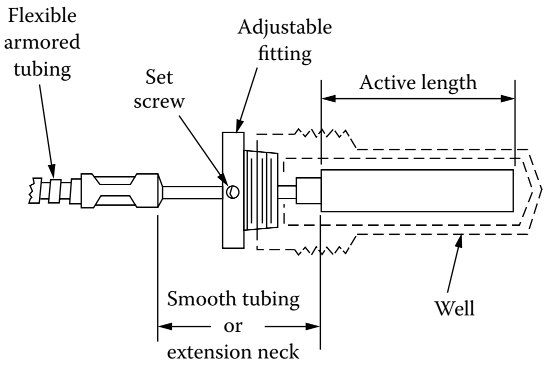


FIG. 7.8g
The main components of a thermal bulb installation with adjustable fitting and well.

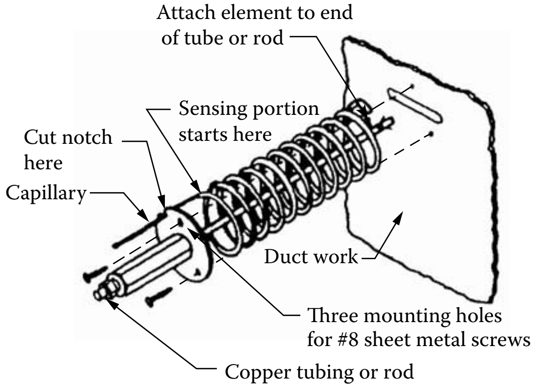


FIG. 7.8h
Averaging capillary-type sensing element for air duct installation. (Courtesy of Johnson Controls.)

Higher speed of response can be obtained with a long, thin, bendable bulb used to sense the average temperature in large areas. Another way of presenting a long bulb is by coiling it. For use in low-velocity gas flow temperature measurement, the coil is set at the factory and cannot be uncoiled (Figure 7.8h).

The primary sensing element of temperature regulators is usually called the bulb in remote sensing designs. The following forms, sizes, fittings, and materials are available.

Bulb Forms and Sizes The cylindrical bulb form is usually standard and is suitable for most liquid control requirements (Figure 7.8i). For faster response in air or gas, a cylindrical form with metal fins is utilized. Other increased surface forms

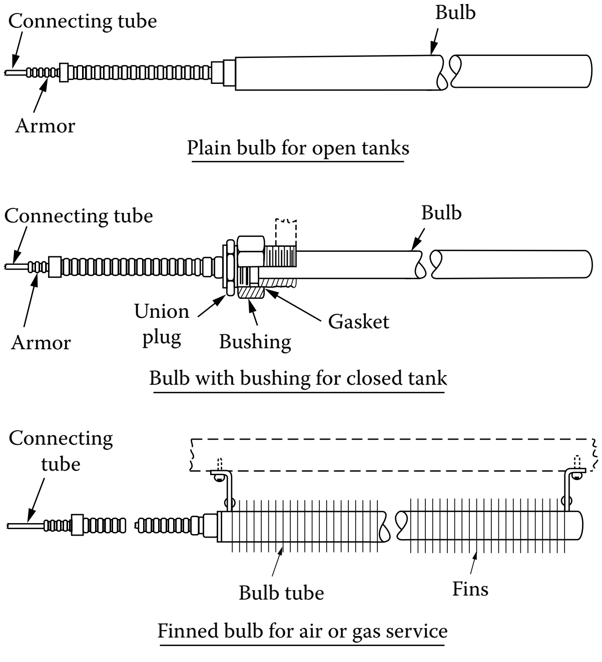


FIG. 7.8i
Thermal bulb designs for open tank (top), closed tank (center), and air service installations.

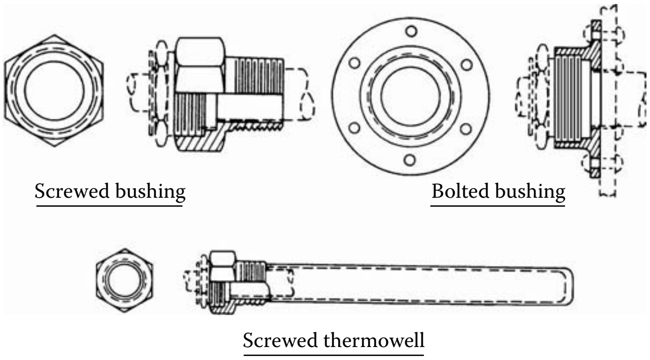


FIG. 7.8j
Thermal bulb design variations.

are the precoiled bulb (for liquid expansion and hot chamber classes) and the tubular bulb. The tubular bulb is made from a substantial length of flexible tubing, to be formed and supported in the field, for example, in and across a duct or wrapped around piping. Other forms include those with dead (inactive) extensions and flexible extensions.

Because of widely varying regulator designs and bulb volume requirements, no bulb dimension standards have been developed. Bulb sizes vary between the extremes of 1/2 in. (12.5 mm) diameter by 9 in. (225 mm) long and 1 1/2 in. (37.5 mm) diameter by 36 in. (900 mm) long. Most bulbs range from 3/4 × 9 in. to 1 1/4 × 24 in. (18.7 × 225 mm to 31.25 × 600 mm) long. Some manufacturers will vary the diameter and length combination to fit special needs.

Fittings and Pressure Ratings For open tank use, no fittings are needed. For pressure vessel or pipeline installations, the use of the union-type screwed bushing is most common (Figure 7.8j). These are generally ground joint unions but are sometimes gasketed. Wells, or sockets, are also available for higher pressure ratings, corrosion protection, and to maintain the integrity of the pressure vessel or system during service. Sanitary fittings are also available.

Bulbs are used in built-up assemblies and are available in copper or in other materials at an extra cost. Table 7.8k gives a list of the available bulb materials and their temperature and pressure ratings.

Capillary Tubing The tubing that transmits the pressure from the bulb to the actuator bellows is usually covered with armor (spiral or braided) for mechanical strength. A tube with 3/16 in. (4.69 mm) OD and approximately 0.050 in. (1.25 mm) wall thickness is common.

The relatively fragile thin-wall capillary should be protected by a flexible armored stainless steel or by PVC-covered bronze tubing. An extension neck to the bulb prevents the tubing from being immersed directly in the measured medium. Bendable smooth steel tubing might also be used for capillary protection.

The standard tubing material is copper with brass armor, but stainless steel (tube or armor), lead-coated copper, and PVC- or Teflon-coated copper or stainless steel are also available.

Standard tubing length is 10 ft (3 m). Longer or shorter lengths are also provided, but lengths above 50 ft (15 mm) are not recommended because of increased mechanical hazards and slow response.

Vapor-Filled System

Still extensively used are the vapor fillings, but these too are limited in their usefulness. The vapor-filled system is limited

TABLE 7.8k
Construction Materials and Other Features of Thermal Bulbs

			Rating	
Bulb Material	*Assembly*	*Fittings*[5]	*Pressure*[1], *PSIG (MPa)*	*Temperature, °F (°C)*
Copper, brass	Brazed	Brass	700 (4.8)	300 (148.9)
Stainless steel[2]	Welded	Stainless steel	1100 (7.6)	600 (315.6)
Steel	Welded	Steel, stainless steel	1000 (6.9)	600 (315.6)
Lead[3]	Fusion	Lead	50 (0.35)	250 (121.1)
Monel	Welded	Monel	1000 (6.9)	600 (315.6)
PVC-coated[4]	Fusion	None	—	160 (71)
Teflon-coated	Shrinking	None	—	300 (148.9)

1. Average, based on 1 in. (25 mm) diameter bulb. Larger diameters have lower ratings.
2. Type #316 is available.
3. Some lead bulbs are sheathed over copper.
4. Coatings are usually over copper, but may be over other materials.
5. Wells for sockets are available in all metals. The use of copper bulbs with other well materials is usual.

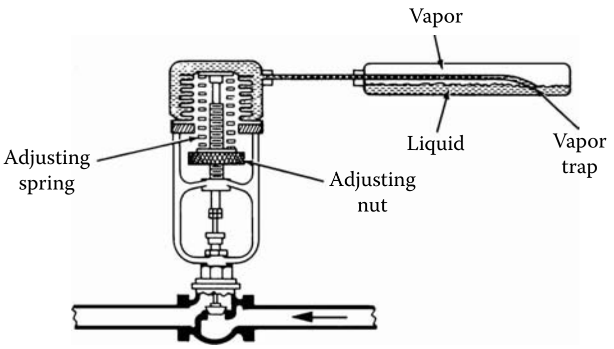


FIG. 7.8l
The main components of a self-actuated temperature regulator using a vapor-filled actuator and thermal element.

due to its nonlinearity and its potential for problems caused by cross-ambient operation or by errors due to elevation.

The operation of the vapor pressure class (SAMA Class II) is illustrated in Figure 7.8l. The thermal actuator, having been evacuated to eliminate the contaminating effect of air and other gases and vapors, is partially filled with a volatile liquid. Liquids used are chemically stable at temperatures well above the range used.

The bellows and tubing usually contain liquid only, if the process temperature is above ambient, although with ambient conditions *above* the set point of the regulator they may contain vapor only—*never both* under conditions of normal operation. If the bulb contains the liquid-vapor interface, the pressure within the actuator is the vapor pressure of the fill liquid at the bulb temperature. This vapor pressure increases with rising temperature and acts on the bellows against the spring force of the adjustment to produce the power required to reposition the valve plug.

Figure 7.8m shows the temperature to vapor-pressure relationship of the various charging liquids used, each having a different vapor pressure curve and atmospheric boiling point. Different ranges are achieved by using different liquids and not by changing springs.

Ranges of Adjustment Ranges of adjustment from –25 to 480°F (–31.7 to 248.9°C) are commonly available, and common spans are in the order of 40 to 60°F (22 to 33°C). In some special cases, involving small short-stroke valves, longer spans up to 100°F (55.6°C) can be furnished. Other

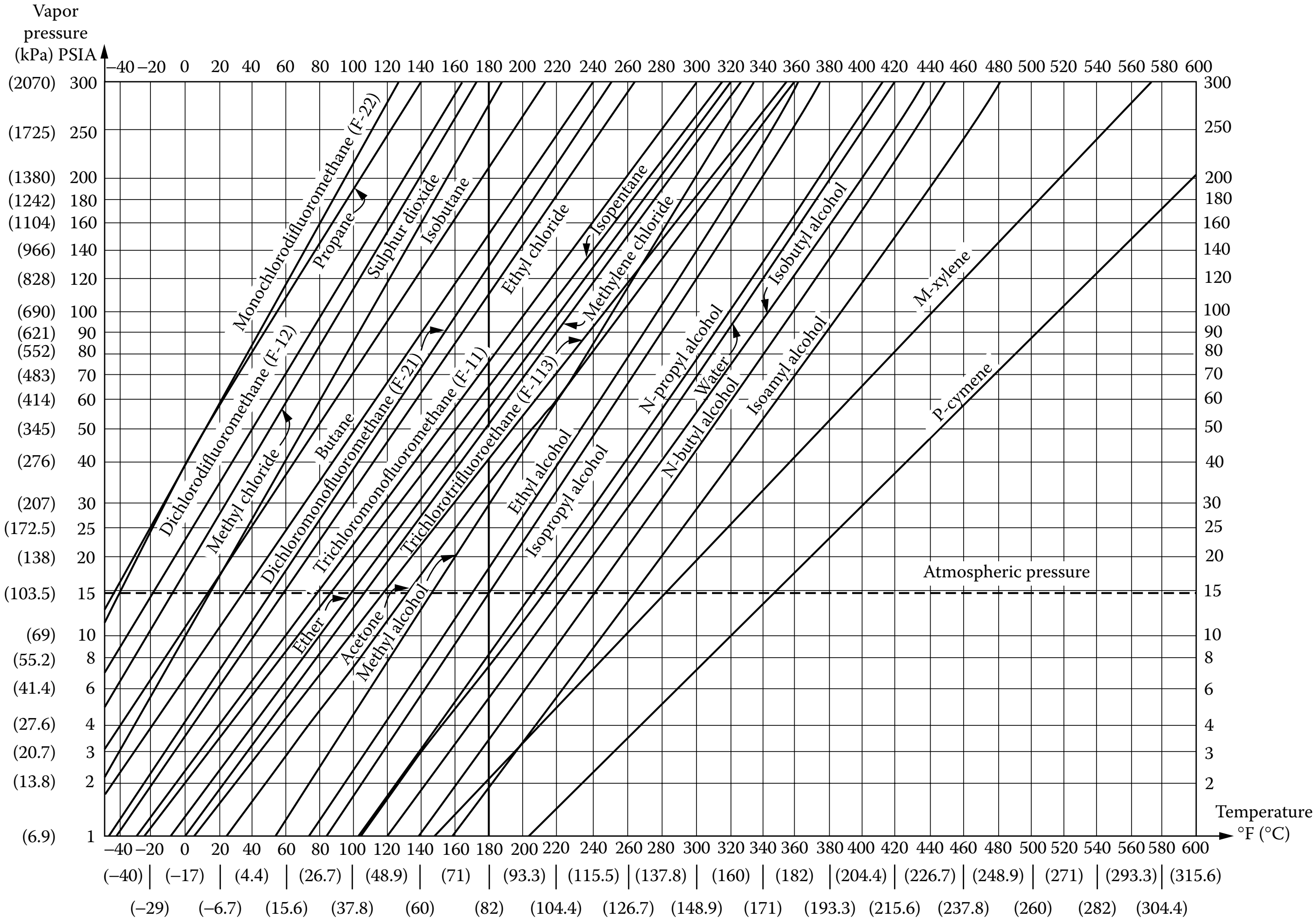


FIG. 7.8m
The relationships between vapor pressure and temperature for a variety of filling liquids.

special designs have shorter ranges of 25 to 30°F (13.9 to 16.7°C).

Adjustment of vapor pressure actuated regulators is accomplished by changing the initial thrust of the adjusting spring (Figure 7.8h). Turning the adjusting nut toward the bellows to compress the spring increases the spring thrust. This requires an increased pressure to start valve movement and thus effects a higher temperature setting. Lowering the adjustment to decompress the spring permits motion of the stem with lower actuator pressure, thereby lowering the set temperature.

The rate of the adjusting spring (force required per inch of spring compression) is an important factor in determining the length of adjustment span as well as the proportional band of a regulator. The stiffer the spring, the longer the span or range, but also the larger the proportional band (lower the gain).

Refrigerant Controls A specialized application of vapor-filled TCVs is the control of heat exchangers that are cooled directly by refrigerant. As the cooling effect is obtained from the evaporation of the refrigerant, these heat exchangers operate the same way as do the evaporators on chillers. As shown in Figure 7.8n, the refrigerant control valve is operated by a diaphragm actuator that compares two pressures: P_1 is the pressure in the temperature bulb, and P_2 is the refrigerant vapor outlet pressure in the equalizer line. The valve is balanced when P_1 equals the sum of the force produced by P_2 and that of the spring.

If the cooling load on the cooler exceeds the refrigerant supply, the refrigerant vapor outlet temperature rises, causing an increase in the vapor pressure inside the thermal bulb (P_1). When P_1 exceeds P_2 plus spring force, the diaphragm in Figure 7.8n moves down, which in turn moves the push rod down and opens the valve. As the refrigerant valve opens, more refrigerant is evaporated, and therefore more cooling is done by the exchanger. As a result, the vapors leaving the evaporator become cooler, P_1 drops, and the force balance on the diaphragm is reestablished.

Inversely, if the cooling load on the evaporator drops, the refrigerant flow becomes excessive for that load, and therefore the temperature at the bulb drops. This results in a drop in P_1 and the closing of the valve until the force balance on the diaphragm is reestablished.

The pressure drop variation across the evaporator tends to assist in the above-described operation. This is the case because when the required refrigerant flow is rising, the pressure drop across the evaporator is also rising, and consequently P_2 is dropping. The drop in P_2 has the same effect as the rise in P_1; that is, they both open the valve. Inversely, when the refrigerant flow is dropping, P_2 is rising, and thereby it assists in closing the valve.

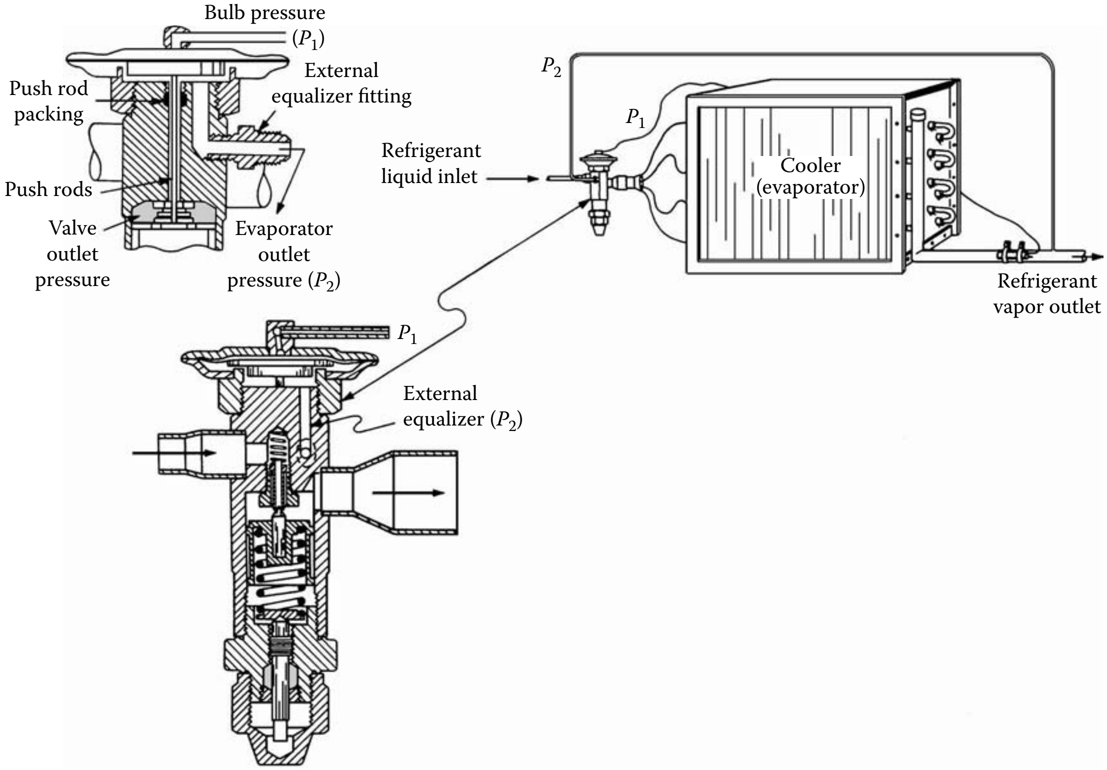


FIG. 7.8n
Temperature regulators provided with refrigerant-filled bulbs admit as much refrigerant liquid into the evaporator as the load demands. (Courtesy of Sporlan Valve Co.)

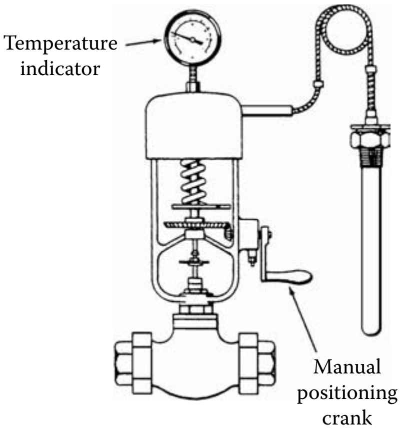


FIG. 7.8o
Accessories such as indicator and handwheel can be provided with the temperature regulators.

The temperature regulator described in Figure 7.8n, therefore, matches the refrigerant flow to the varying coolant demand of the process, while maintaining some superheat in the refrigerant vapor leaving the evaporator. If the cooler is properly sized, this guarantees that the refrigerant liquid will never spill into the outlet line.

Special Features Weight-and-lever adjustment, once very common, is seldom used today because of problems of position, location, and general vulnerability of the moving lever. Here, the force produced by the weight acting on an adjustable length lever or "movement" arm opposes the thrust produced by the thermal actuator. Increasing the length of the lever raises the setting. Because the force of the weight-and-lever system has no "rate" (being almost constant for a given setting), the proportional band of a weight-adjusted regulator is narrower than for a spring-adjusted design.

Overtemperature protection is standard in most regulators of the vapor pressure class. It consists of a special spring-loaded and compressible "overrun" section of the actuator stem. When increasing temperature produces a force beyond that for normal control action, this overrun compresses. The added movement of the actuating bellows drains the remaining liquid from the bulb, and pressure then follows the gas law instead of the vapor pressure curve. This provides protection up to 100°F (55.6°C) above the top end of the range.

Temperature indication is available as a special feature (Figure 7.8o). Generally, this is simply a compound pressure gauge, sealed in the thermal system and responding to the changing pressure of the fill. The dial is calibrated to correspond to the temperature-pressure curve of the particular fill (change) and, therefore, "reads" the bulb temperature. In some designs the thermometer has its own sensing elements, with a separate bulb contained in or attached to the regulator bulb. Improved characteristics are claimed for this design. In either case, the "thermometer" is in the order of approximately ±2% inaccuracy. Thermometers are furnished with both Fahrenheit and Celsius calibration.

Bulb Size Actuator bellows design (diameter, area, length, etc.) is the major factor in determining the characteristics of the vapor pressure regulator. A small bellows, having limited power and stroke, results in a compact actuator, successful only for small valves and on moderate to low pressures.

Increasing bellows size makes direct actuation of larger valves possible at higher pressure levels. However, because the bellows movement requires the transfer of some liquid from the bulb to the bellows (equal to the volumetric displacement), larger bellows require larger bulbs. In general, larger bellows result in shorter adjustable ranges but also narrower proportional bands. Some manufacturers offer a choice of designs based on bellows size in order to offer different characteristics.

Bulb size of the vapor pressure regulators is governed by the volumetric displacement of the bellows (area × valve travel) and varies with actuator design and with valve design and size. It is also affected by the possible need to operate under "cross-ambient" condition, i.e., with ambient temperature either above or below the regulator's set point.

This requirement produces a bulb substantially larger than that used where the ambient temperature will always be either above or below the range of the regulator. *The use of vapor-filled thermal elements on cross-ambient applications is generally not recommended.* Bulb sizes of this class range from 1/2 in. (12.5 mm) OD by 6 in. (150 mm) OD by 36 in. very small units to 1 1/2 in. (37.5 mm) OD by 36 in. (900 mm) long in large valves with cross-ambient fill.

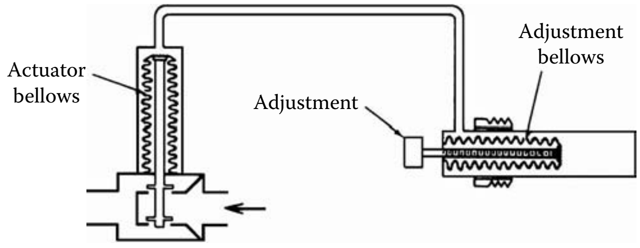
Bulb construction for vapor pressure regulators includes a "vapor trap" (Figure 7.8l). This serves to give the best possible response by ensuring liquid transfer of pressure changes from bulb to bellows. Because the end of the vapor trap tube must always be in liquid, it requires that the installation of the bulb be made in accordance with markings on the bulb. For installations where the tube connection end of the bulb is below the horizontal, the extended vapor trap must be omitted.

Advantages of vapor pressure regulators are (1) ambient temperature has no effect on calibration, and (2) there is a wide choice of designs, characteristics, ranges, and special features available.

Liquid-Filled System

The use of mercury filling has been discontinued in most industrial applications because of health concerns. The use of both mercury and liquid fillings have also lost ground because of the expense associated with compensating for the ambient effects on the capillary and because of the errors caused by elevational differences between bulb and readout.

The operation of the liquid expansion class (SAMA IB) is illustrated in Figure 7.8p. The thermal actuator is completely filled with a chemically stable liquid. Changing temperature



Uniform liquid expansion per degree temperature rise at sensing bulb compresses valve actuating bellows to control flow of heating or cooling medium

FIG. 7.8p
Liquid-filled thermal actuator and sensing element.

at the bulb produces a volumetric change, which causes the small-area bellows to move in a corresponding direction, thereby moving the valve. The force and action are positive and linear, limited only by the components of the element, chiefly the actuator and adjustment bellows.

Ranges are from 0 to 300°F (–17.8 to 148.9°C) with spans generally from 30 to 50°F (16.7 to 27.8°C). The special 100°F (55.6°C) spans are sometimes furnished with very small, short-stroke valves. The different ranges (not field-changeable) are achieved by filling procedure and sealing temperature. Hydrocarbon fills are generally used. Mercury is no longer used as a filling fluid for regulators because of cost and hazards.

Adjustment is achieved by moving the actuator bellows relative to the valve plug (Figure 7.8b) or by using a separate adjusting bellows (Figure 7.8p). This adjustment bellows transfers liquid into the actuator bellows to lower the set point, or vice versa. The adjustment may be located at the bulb or at the valve, or it may be separate from both.

The proportional band is generally wider than with other classes of actuators (lower gain), and it is a direct function of bulb volume to actuator bellows area.

Overtemperature protection, which permits overtravel of the bellows, allows an uncontrolled temperature to overrun the set point by up 50°F (27.8°C). Ambient temperature effect on the liquid in the tubing, bellows, and adjustment is minimized by design, but wide ambient swings do introduce errors.

Packless construction (no moving stem through the body envelope) is a feature of this class of regulator. However, they must not be used in applications involving corrosive fluids. Temperature indication is not generally available.

Advantages of the liquid expansion class regulators include a positive actuating force, linear movement, and packless design.

Hot Chamber System

The hot chamber class of regulator has no corresponding SAMA designation. This thermal actuator is partially filled with a volatile fluid (Figure 7.8q) and has some characteristics common to the liquid expansion class. Rising temperature at the bulb forces liquid into the actuator bellows (the "hot chamber"). There, the heat of the control agent, always at a temperature substantially above the regulator range, flashes the liquid into a superheated vapor. The pressure increase in the bellows causes the valve to move against a return spring.

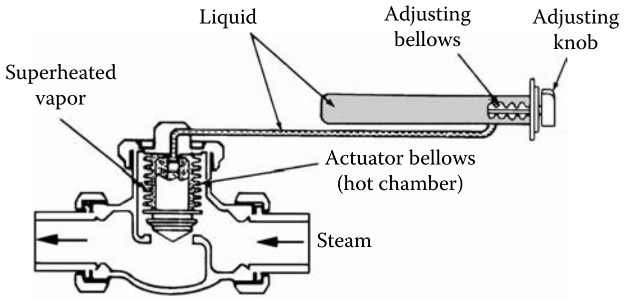


FIG. 7.8q
The sensing element and the actuator of a hot chamber-type temperature regulator.

Hot chamber regulators are available in sizes from 1/2 to 1 1/2 in. (12.5 to 37.5 mm). Because of practical considerations, only direct-acting designs are available (close on rising temperature), and they are limited to use on steam at pressures up to 75 PSIG (518 kPa). Theoretically, other designs are also possible.

Ranges are from 30 to 170°F (–1.1 to 76.7°C) (to 200°F, or 93°C, in special cases) with spans from 30 to 60°F (16.7 to 33.3°C). Adjustment is achieved by the use of adjusting bellows that are usually located at the bulb but may be separate and remote.

Ambient temperature effects are minimized by design, but wide ambient swings are nevertheless detrimental. Effects of steam supply pressure variations are partially self-compensating.

Bulbs (thermostats) take several forms. For steam space heating systems, the wall-mounting types or duct-mounting types are common, while for liquid heating the cylinder types with pressure-tight bushings are utilized.

Advantages of the hot chamber class include calibrated knob and dial adjustment, good response and good proportional characteristics, packless design, and compact construction.

Fusion-Type System (Wax-Filled)

The fusion class of regulator does not have a SAMA equivalent. In the least common class, the compact element is filled with a special wax containing a substantial amount of copper powder to improve its rather poor heat-transfer characteristics (Figure 7.8r). A considerable and positive volumetric increase occurs in the wax as temperature rises through its fusion or melting range. This volumetric change produces a force used for valve actuation through the ingenious arrangement of a sealing diaphragm that acts to compress a rubber plug that "squeezes out" the highly polished stem (Figure 7.8r). This

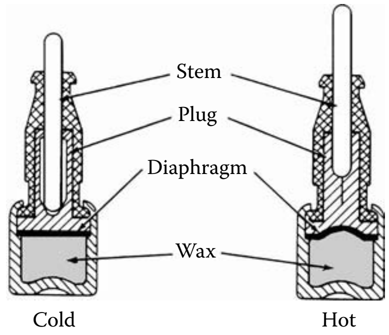


FIG. 7.8r
Wax-filled, fusion-type thermal element.

element has more hysteresis and dead band than the other classes.

Different waxes, such as natural waxes, hydrocarbons, silicones, and mixes of these are available that provide a variety of operating (control) spans between 100 and 230°F (55.6 and 127.8°C).

No adjustment of set point is provided, because the fusion (melting) span of the wax is narrow, from 6 to 14°F (3.3 to 8.3°C). A heavy return spring, which is always present in the regulator, forces the rubber components of the element to follow the wax fill on falling temperature.

Remote bulb designs are not offered because they would provide no advantages over the other three classes. Thus, fusion class regulators are available only in "self-contained" styles (Figure 7.8s).

Overtemperature protection up to 50°F (27.8°C) above the operating span is provided by the flat temperature expansion curve of the wax above its melting range.

Valve sizes are from 1 to 6 in. (25 to 150 mm), with the majority being of the three-way valve design in sizes from 2 to 6 in. (50 to 150 mm).

Advantages of fusion class regulators lie in their compact size and low cost of the thermal element, even where multiple elements are required to stroke larger valves.

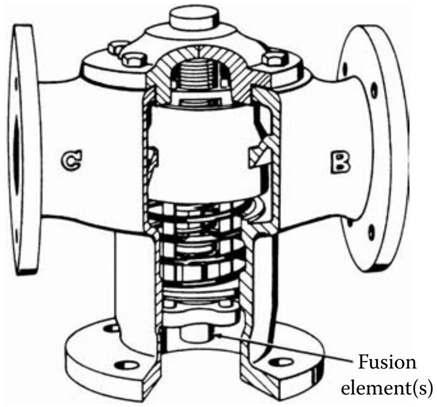


FIG. 7.8s
Self-contained temperature regulator with wax-filled (fusion-type) element.

THE REGULATOR VALVE

Further discussion here is limited to some special forms of valves and to factors relating to selection and use due to force limitation of the thermal actuator. For a detailed discussion of the features of globe valves, refer to Section 6.19.

Valve action refers to the relationship of stem motion to plug and seat position. A direct-acting valve closes as the stem moves down into the valve. A reverse-acting valve opens as the stem moves down. A three-way valve combines these, one port (or set of ports) opening and the other port(s) closing on downward movement.

A direct-acting valve on heating service will reduce the flow rate of the heating medium if the controlled temperature increases. Therefore, heating control is accomplished by a direct-acting regulator, in combination with a direct-acting valve.

Cooling control is most frequently achieved by a reverse-acting regulator with a reverse-acting valve to increase the flow of coolant on rising temperature. However, on constant coolant flow systems, a direct-acting regulator with its valve in a bypass line is occasionally used to vary the flow through the heat exchanger.

A three-way valve provides positive control of such a cooling system by giving a full range of flows through the two legs of the cooling circuit (exchanger and bypass). This gives good temperature control, but wastes energy. Mixing of two media at different supply temperatures to control the mixed temperature is also accomplished with three-way valves.

Single- and Double-Seated Valves

Single-seated valves are desirable for minimum seat leakage in the closed position. The pilot-actuated regulators always use single-seated valves (Figure 7.8t). On direct-actuated types, the use of single-seated valves is limited to a 2 in. (50 mm) maximum for two reasons: (1) the required full-open lift (approximately 1/4 of valve size) and (2) the closing force required (full port area times maximum pressure drop across the valve). A single-seated valve with piston-balanced plug (Figure 7.8t) eliminates the closing force problem but requires slightly higher full-open lift.

Double-seated valves are most common on direct-actuated regulators in sizes up to 4 in. (100 mm). The reasons for this include their low full-open life (approximately 1/8 of valve size) and minimum closing force requirement. The unbalanced area of double-seated valves is the difference between the two port areas, which is approximately 6% of the larger of the two — hence, the term "semibalanced" is used for double-seated valves (Figure 7.8u).

Three-Way Valves

Three-way valves used with temperature regulators are of three basic designs (Figure 7.8v). Smaller valves (to 2 in., or

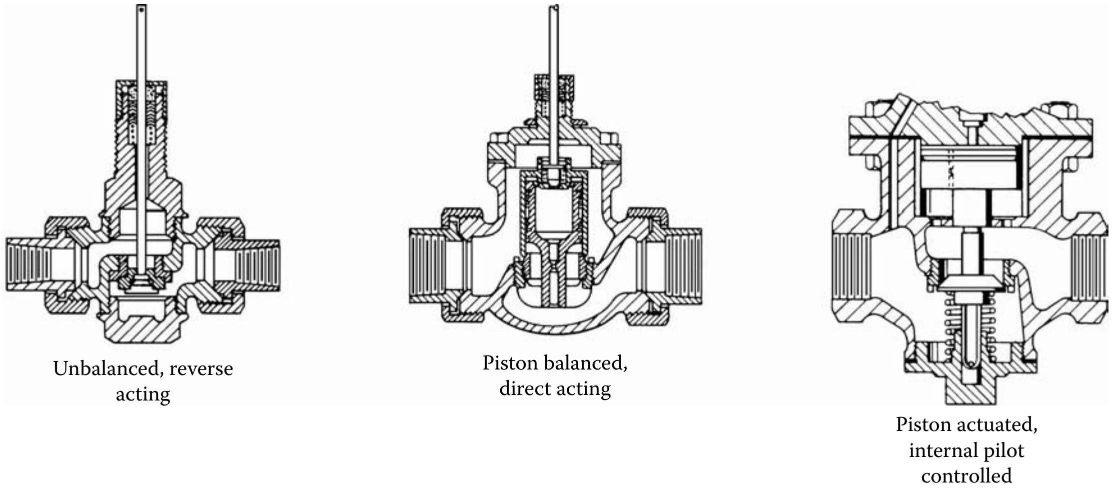


FIG. 7.8t
Three designs of single-seated, two-way regulator valves.

50 mm) are of the "unbalanced" single-seated construction with a common plug moving between two seats. In sizes from 2 to 6 in. (50 to 150 mm) there are two designs. A sleeve-type plug (requiring a body seal) moves between two seats and provides near balance of closing forces, but it requires typical single-seated lift. A semibalanced design, essentially two double-seated valves for each flow path with connected plugs, gives both low lift and low seating force. This construction is somewhat more expensive, however.

Plug guiding is considered to be a less important factor with regulators than with control valves, because the applications involve relatively low pressures and velocities. Therefore, many direct-actuated single-seated valves are only stem guided. Most double-seated valves are seat ("wing") guided, but some top- and bottom-guided designs are also in use. Pilot-actuated valves are generally well guided, as are the piston-balanced valves.

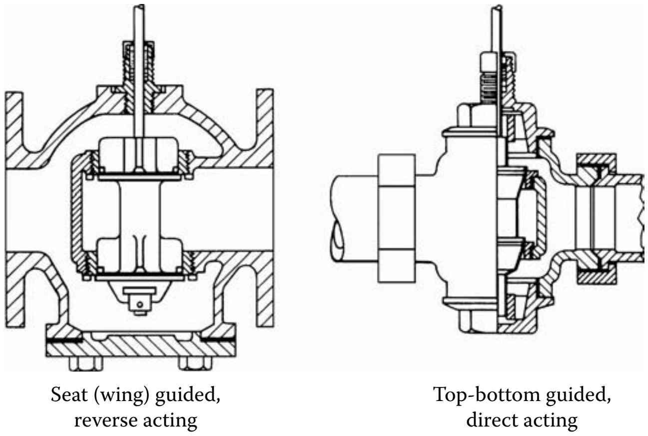


FIG. 7.8u
Double-seated, two-way regulator valves with direct (right) and reverse (left) action.

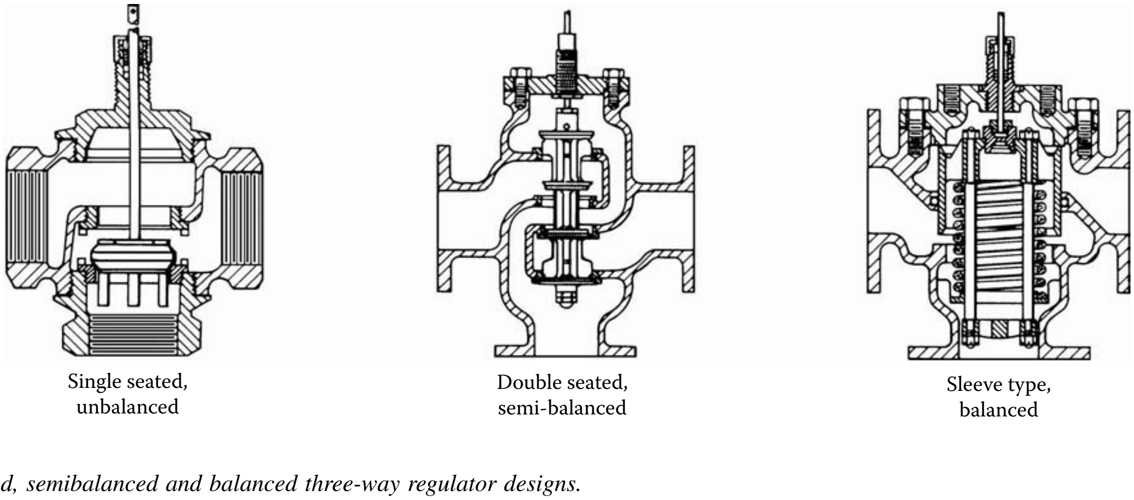


FIG. 7.8v
Unbalanced, semibalanced and balanced three-way regulator designs.

Valve Features

Flow Characteristics Most regulator valves are of the quick-opening type in order to use the shortest practical lifts. On special basis, low-lift plugs are available in some small single-seated valves and also in a few double-seated valves. As was explained in Sections 6.1 and 6.7 in Chapter 6, the ideal valve characteristic for temperature control loops is equal percentage. This is another reason why TCV regulators are not used on critical temperature control applications.

Materials and Connections Standard materials used in body construction are bronze in sizes below 2 in. (50 mm) and cast iron in larger sizes. Steel and stainless steel are considered to be special body materials. The standard bronze valve trim is used with pressure drops below 50 psid (345 kPa), while stainless steel is recommended for greater drops.

On low-pressure and -temperature services, composition discs can be used for tight shut-off. Stellite or hardened alloy trim for extreme service is available as special, though some pilot-actuated designs offer it as standard. Monel and other special alloys are available for specialized services, such as seawater applications.

To lessen the chance of installation damage, union ends are usual for 1/2 to 1 1/2 in. (12.5 to 37.5 mm) valves (in some cases 2 in., or 50 mm). Flanged ends are standard in sizes 2 in. (50 mm) and larger.

Pressure Rating The pressure rating of a temperature regulator is frequently determined more by limitations in the thermal actuator power (seating force) than by the design and materials of the valve body and trim. Therefore, while catalogs will usually state the maximum temperature and pressure rating for the valve, they should also give a lower recommended maximum pressure limit for a direct-actuated regulator. This will usually decrease as valve size increases, and it should be given full cognizance.

With pilot-actuated valves and piston-balanced valves (direct actuated), the actuator limitation is not a factor. However, regulators are seldom available or recommended for pressures above 250 PSIG (1.7 MPa).

Valve Capacity and Leakage The valve capacity data published by manufacturers is reasonably accurate, having been determined by recognized test procedures (Section 6.6 in Chapter 6). Most manufacturers publish capacity tables for common fluids, from which the full-open capacity figures can be read directly for various supply pressures and pressure drops. In a few cases, the information takes nomographic form. Some list the valve capacity factor C_v for each valve size and type.

Regulators should not be considered as positive shut-off devices. However, in "as new" condition, leakage across seats in the closed position for metal-seated valves may be expected to be 0.02% of full capacity for single-seated and 0.5% for double-seated regulators. Single-seated valves with composition discs can be nearly "dead tight" or "bubble tight."

CONCLUSIONS

Because the characteristics of a regulator cannot be altered in the field, greater care must be taken in the proper evaluation and selection than would be the case when using a complete temperature control loop with adjustable control mode or modes.

In general, temperature regulators are most successful on control applications having high "capacitance" (slow processes), such as heated or cooled storage tanks, process ovens, and dryers, or on applications with small load changes, such as metal treating (cleaning and plating) tanks.

Proper installation includes the selection of a good location for the bulb, where process temperature changes can be sensed quickly, but away from potential damage by moving equipment. Regulators in sizes 2 in. (50 mm) and greater should be installed upright in horizontal lines. Strainers should be installed for the protection of all regulators.

Regulator maintenance is minimal. Direct-actuated regulators with packed valve stems may require occasional lubrication and infrequent replacement of packing and cleaning or polishing of stem. Pilot-actuated regulators with piston-actuated valves (Figure 7.8d) may require programmed cleaning to remove "glaze" and deposits that would cause the piston to stick. Depending on the service, occasional examination and servicing of valve seats may be required to maintain needed tightness.

Relative to using full-temperature control loops, the advantages and limitations of temperature regulators can be summarized as follows: Advantages include low installed cost, ruggedness, simplicity, low maintenance, and no need for an external power source. The limitations include the fixed gain and the resulting offset, the local set point, the limited and narrow ranges, the slow response, the limitations in size and operating pressures, and requirements for a relatively large bulb.

Bibliography

Adler, C. B., "Reliability Aspects of Temperature Measurement" 2001 ISA Conference in Chicago.

Carlson, D. R., "Temperature Measurement in Process Control," *InTech,* October 1990.

Deamarais, R., and Breuer, J., "How to Select and Use the Right Temperature Sensor," *Sensors,* October 2001.

Drenberg, C. H., "Automatic Valve Control," *Chemical Processing,* August 1980.

Farr, W. E., "Save Dollars on Plant Heaters," *Chemical Processing,* May 10, 1980.

Farr, W. E., "Temperature Control Valves Replace Traps," *Fond Processing,* July 1980.

"Filled System Thermometers," *Measurements and Control,* September 1992.

Hashemian, H. M., et al., "Assuring Accurate Temperature Measurement, *InTech,* October 1989.

Jutila, J. M., "Temperature Instrumentation," *Instrumentation Technology,* February 1980.

Kerin, T. W., and Katz, E. M., "Temperature Measurement in the 1990s," *InTech,* August 1990.

Magison, E., "Selecting the Right Temperature Sensor," *InTech,* October 25, 2001.

Suczkstorf, D., "Temperature-Activated Valves Save Steam," *Plant Services,* July 1980.

Volbrecht, A., and Gordon, W., "Temperature Measurement: Making Sense of it All," *Sensors,* June 1998

Warmoth, D., "Temperature Measurement and Control," *Measurement and Control,* February 1986.

When, S., "Liquid-in-Glass Thermometers, Still Well and Indicating," 1991 ISA Conference, Paper #91-0308.

7.9 Thermostats and Humidistats

B. G. LIPTÁK (1985, 1995, 2005)

Types:	A. Conventional HVAC thermostats B. Advanced, digital, intelligent thermostats C. Conventional HVAC humidistats D. Advanced, digital humidistats
Standard Ranges:	A. 45 to 85°F (7 to 29°C) for heating and from 55 to 105°F (13 to 40°C) for cooling services B. 40 to 90°F (5 to 32°C) to 23 to 122°F (-10 to 50°C) or higher C. 15 to 95% RH D. 0 to 100% RH
Inaccuracies:	Sensor Error: A. ±1°F (0.5°C) B. ±0.5°F (0.28°C), even better with RTD-based digital units C. ±3 to 5% RH D. ±2% RH (some w/1% resolution) Offset Error: A. ±1 to 5°F (0.5 to 2.8°C) B. Zero with integral C. ±2 to 10% RH D. Zero with integral
Air Capacities:	Relay types (high volume): 10 to 30 SCFH (0.3 to 0.84 m^3/hr) Non-relay types (low-volume): 1 SCFH (0.028 m^3/hr)
Costs:	A. Room thermostat $50 to $100. Low-volume, proportional, pneumatic room thermostats cost $75 to $125; high-volume, proportional, dual room thermostats for heating/cooling and with automatic changeover cost $175 to $250; averaging, proportional thermostats for duct mounting cost about $150; on/off electric thermostats can be obtained for $100 or less B. Microprocessor-based on/off home thermostats cost $100 to $200, programmable microprocessor-based energy-saver thermostats cost $200 to $400; C. Proportional, pneumatic room humidistats cost $200 to $250; duct-mounted, proportional, pneumatic humidistats cost about $300; on/off electric humidistats can be obtained for about $100, HVAC transmitter $600 to $800. D. For the costs of higher-quality humidity instruments see Section 8.32 in the *Process Measurement* volume of this handbook (Note: Installed total costs can double the hardware costs listed above.)
Partial List of Suppliers:	(For Type D see Section 8.32 in the *Measurement* volume.) Ametek Thermox (C) (www.thermox.com) Barber-Colman Co. (A, B, C) (www.barber-colman.com) Dwyer Instrument, Mercoid Div. (A) Emerson Electric Co. (A, B) (www.emersonprocess.com) Eurotherm/Barber Coleman (A, B) (www.eurotherm.com) Foxboro-Invensys (C, D) (www.foxboro.com) General Eastern Instruments (C) (www.geinet.com) Honeywell Inc. (A, B, C) (www.honeywell.com/sensing)

Johnson Controls Inc. (A, B, C) (www.johnsoncontrols.com)
Jumo (A, B, C) (www.jumousa.com)
Ohmic Instrument Co. (C) (www.ohmicinstruments.com)
Panametrics Inc. (C, D) (www.panametrics.com)
Powers Controls (A/B) (www.powerscontrols.com)
Princo Instruments Inc. (A, C) (www.princoinstruments.com)
Robertshaw/Invensys (A/B) (www.robertshaw.com/cli-fam-robTherm.htnl)
Rotronic Instrument Corp. (C) (www.rotronic-usa.com)
Staefa/Siemens Building Automation (A/B) (www.sbt.siemens.com/hvp/Components/catalog/thermostats.asp)
Vaisala Oyj. (C) (www.vaisala.com)
Watlow Controls (C) (www.watlow.com)
Westinghouse Electric Corp. (A, B) (www.westinghousepc.com)
White Rogers (A, B) (www.whiterogers.com)
Yokogawa Corp. of America (C) (www.yokogawa.com/us)

INTRODUCTION

There is an overlap between the measurement and the control of room temperature and humidity. The first volume of this handbook on *Process Measurement and Analysis* covers these topics from the measurement perspective in Sections 4.11 and 8.32. In this section, the emphasis will be placed on the control capabilities of these devices.

Thermostats and humidistats have been developed to serve the heating, ventilating, and air conditioning (HVAC) industry. While energy costs were low, the consideration of performance was secondary to the purchase price of these devices. When energy costs increased, new, better-quality units were developed, but the conventional HVAC devices still remain in use in existing systems and in new installations, where the designers are less sensitive to the quality of performance. In this section, the conventional quality units and the more advanced designs are both described.

On/off and throttling devices are both referred to as thermostats or humidistats. Therefore, both versions will be discussed, but more emphasis will be placed on the throttling-type designs. First the narrow proportional band controllers, the stats, will be described. After that, the various thermostat and humidistat designs will be covered.

ACCURACY OF THERMOSTATS AND HUMIDISTATS

Conventional

In conventional HVAC thermostats and humidistats, the total error is the sum of the sensor error and of the offset error. Even after individual calibration, the sensor error of conventional thermostats cannot be reduced to less than ±1°F (±0.55°C) for thermostats or to ±5% relative humidity (RH) for humidistats. Added to the sensor error is the offset error, which increases as the throttling range increases. This, for conventional thermostats, can be as high as ±5°F (±2.8°C) and ±20% RH for humidistats. If the two errors are additive, the total errors of conventional units can approach ±6°F (±3.3°C) and ±20–25% RH, respectively.

In conventional HVAC electrical thermostats or humidistats, the total error is the sum of three components. The first error component is the element error, which is the same as the sensor error in pneumatic designs. The second error component is the mechanical differential set in the thermostat switch, which does not modulate the final control element, but turns it on and off. The differential (or bandwidth) can vary for the thermostats from 0.5 to 10°F (0.3 to 5.6°C) and for humidistats from 2 to 50% RH.

The third error component is due to the fact that turning on a heat or humidity source does not instantaneously return the space to within the control differential. Instead, conditions will continue to deviate further for a while. The total error resulting from these three contributing factors can be as high as for HVAC pneumatics, although if the final control elements are sized large enough and if the control differential is set small, electrical units are likely to outperform their pneumatic counterparts.

Advanced

If space conditions are to be accurately maintained, such as in clean rooms where the temperature must be controlled within ±1°F (±0.55°C) and the humidity must be kept within ±3% RH, neither of the above designs is acceptable. In such cases it is necessary to use a resistance temperature detector (RTD)-type or a semiconductor transistor-type temperature sensor and a proportional-plus-integral thermostat controller, which will eliminate the offset error. This can be most economically accomplished through the use of microprocessor-based thermostats that communicate with sensors over data highways.

The performance of advanced, microprocessor-based digital units is much better than that of conventional thermostats. The error in an RTD-based digital thermostat is around 1% of span. Similarly, the resolution of a digital hygrometer is around 1%.

THERMOSTATS

Conventional thermostats are usually uncalibrated devices, and their manufacturers usually do not guarantee their accuracy. This is a limitation, because it is possible to have some

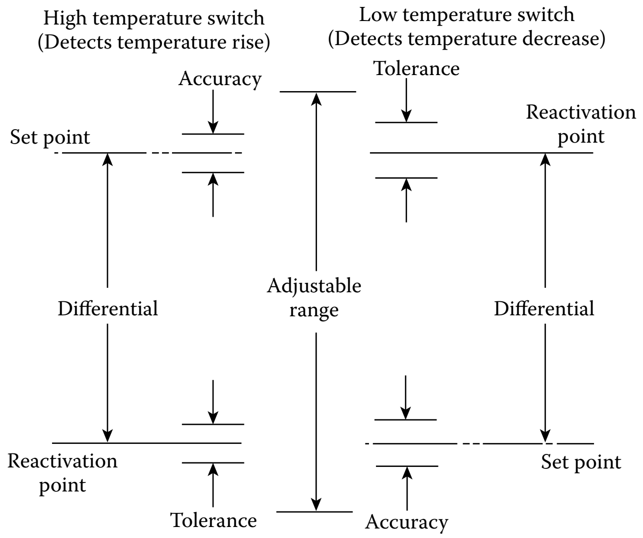


FIG. 7.9a

A cooling thermostat operates as a high-temperature switch, and a heating thermostat functions as a low-temperature switch.

thermostats with as much as 5–10°F (3–6°C) error in their measurement. In the last decades, a new generation of thermostats have been introduced that guarantee to limit their error to 1°F (0.6°C) or less.

ElectroMechanical Designs

There is little difference between a two-position thermostat and a temperature switch. A thermostat that is controlling the heating of a space operates like a low-temperature switch, and a cooling thermostat operates as a high-temperature switch (Figure 7.9a).

The movement produced by a bimetallic temperature sensor or by the expansion of a humidity-sensitive element can be used to build two-position controllers. The electric thermostats in Figure 7.9b illustrate the various differential gaps in two-position control. The thermostat on the left operates load contacts directly from the bimetallic element.

The set point is adjusted by setting one contact or by repositioning the whole bimetal element. The gap in this arrangement is theoretically zero, although in actual practice there is a very small gap due to contact "stickiness."

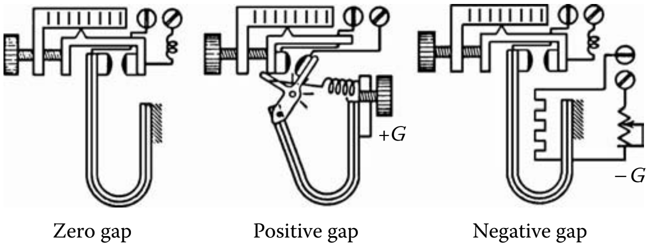


FIG. 7.9b

Design variations of electrical two-position thermostats.

A positive differential gap is often necessary in two-position control in order to save wear on the control apparatus. The mechanism shown in Figure 7.9b employs a toggle action to produce a positive gap. The gap is adjustable by the tension of the spring, because this determines the suppressive (hysteresis) force acting on the element.

A negative differential gap is often used on domestic thermostats in order to reduce an abnormally long period of oscillation. An auxiliary heater operated by the load contacts turns on a small heater within the thermostat when the contact is made. Thus, the bimetal element opens the load contact before the actual room temperature attains the same point. The load contact make point is normal. This is sometimes termed "anticipation."

There are many forms of two-position thermostat- or humidistat-type controllers for both domestic and industrial applications. These offer many arrangements of load contacts, set point, and differential gap. Generally, a solenoid valve or an electrical heating element, motor, or other power device is operated by the controller.

For floating controller action, a neutral zone is usually necessary. This is obtained by a second contact position in the controller and is often obtained with two independently set on/off control mechanisms.

Single-speed floating control can be achieved by means of the same equipment as for two-position control with an electric-motor-operated valve. The difference lies only in the speed of operation of the motor or valve actuator. In two-position control the valve stroke is 120 sec or less. In single-speed floating control, the valve stroke is usually 120 sec or more. An electrical interrupter is sometimes employed in conjunction with the motor to decrease the speed of opening and closing the valve.

Electrical/Electronic Design Features

The most common thermostat is the low-voltage unit that uses external interposing relays to control heating or cooling equipment (typically 24 VAC). Other models include:

Line voltage units that directly control AC circuits (120 or 240 VAC).
Heating only or heating/cooling models.
Digital or analog indication of room temperature.
Bimetallic, filled system or electronic sensing elements.
Direct or reverse acting.
Local or remote set point.
Local set points may be external or internal and may be key protected.
Limited control range thermostats allow the occupant of an office to move the set point to any value desired, but will disregard any setting that exceeds the limit value.
Programmable set-back models with 24-hour or 7-day programs.

Microprocessor-Based Units

A *smart thermostat* is usually a microprocessor-based unit with an RTD-type or a transistorized solid-state sensor. It is usually provided with its own dedicated memory and intelligence, and it can also be provided with a communication link (over a shared data bus) to a central computer. Such units can minimize building operating costs by combining time-of-day controls with intelligent comfort gap selection and with maximized self-heating.

Microprocessor-based units continue to incorporate new features. These are programmable devices with memory and communication capability. They can be monitored and reset by central computers using pairs of telephone wires as the communication link. Microprocessor-based units can be provided with continuously recharged backup batteries and with accurate electronic room temperature sensors. They can also operate without a host computer (in the "stand-alone" mode). In this case the user manually programs the thermostat to maintain various room temperatures as a function of the time of day and other considerations.

Control by Phone

In the year 2000, the first residential gateway for the HVAC industry's new open communications standard for residential environmental control was introduced. This gateway enables secure, remote access to residential HVAC systems via touch-tone telephone, standard phone line, and pass code. Some systems offer either "dial-in" or "dial-out" functions. Owners can call in to check the temperature, change the status of their HVAC system, and perform other thermostat-related functions.

These telephone access modules enable monitoring of the building's temperature and its heating/cooling equipment. It can report when a furnace filter needs replacing or an electronic air cleaner's cells need cleaning. And it can call up to three phone numbers to immediately alert the owner, contractor, or others of problems, such as freezing temperatures or an extended power outage.

The Proportional-Only Controller

A thermostat or humidistat is a simplified controller, having only one control mode—proportional—that is narrow and fixed. The pressure of the output signal from a pneumatic "stat" is a near straight-line function of the measurement, described by the following relationship:

$$O = K_c(M - M_0) + O_0 \qquad \textbf{7.9(1)}$$

where

O = output signal

K_c = proportional sensitivity (K_c can be fixed or adjustable depending on the design)

M = measurement (temperature or relative humidity)

M_0 = "normal" value of measurement corresponding to the center of the throttling range

O_0 = "normal" value of the output signal, corresponding to the center of the throttling range of the control valve (or damper)

Throttling Thermostats are distinguished from other controllers in that they provide proportional action only and that their proportional band (gain) is narrow and fixed. In this section, the operation of thermostats will be described on the basis of their pneumatic designs, but all concepts and features that are discussed are equally applicable to analog or digital electronic thermostats, also.

A conventional pneumatic room thermostat might have a set point range of 55–85°F (13–29°C) and a *fixed sensitivity* (gain) of 2.5 psi/°F (0.3 bar/°C). Assuming that this thermostat operates a control valve having a 9–13 PSIG (0.6–0.9 bar) spring range, we can convert the fixed sensitivity of 2.5 psi/°F into a percent-proportional band value that is better understood by instrument engineers.

Sensitivity (or gain) is stated as the ratio between a change in measurement and a change in the corresponding controller output. A 1°F (0.56°C) change in measurement is 3.3% of the measurement range of 30°F (55–85°F, or 13–29°C), while the 2.5 psi change in controller output is 62% of the controller output range of 4 psi (9–13 PSIG). If a 3.3% change in input results in a 62% change in output, that is a gain of 62/3.3 = 19. Converting this typical thermostat gain to proportional band, we get:

$$P = \frac{100}{G} = \frac{100}{19} = 5.3\% \qquad \textbf{7.9(2)}$$

If a thermostat is controlling a valve or a damper, one needs to determine K_c, M_0, and O_0 in Equation 7.9(1) in order to draw the straight-line relationship between measurement and output signal. Figure 7.9c illustrates the behavior of a control loop consisting of a direct acting thermostat, having a fixed K_c of 2.5 PSIG/°F (0.3 bar/°C), a "normal" measurement value of M_0 = 72°F (22ºC), and a damper spring range of 8–13 PSIG (0.55–0.9 bar). This combination results in a normal output signal: O_0 = 10.5 PSIG (0.73 bar).

Does it Have a Set Point? It is a common error to call M_0 in Figure 7.9c the set point of that thermostat. Stats do not have setpoints in the sense of having a predetermined temperature value to which they would seek to return the condition of the controlled space. (One must add integral action in order for a controller to be able to return the measured variable to a set point after a load change.)

M_0 does not represent a set point; it only identifies the space temperature that will cause the cooling damper in Figure 7.9c to be 50% open. This can be called a "normal" condition, because relative to this point the thermostat can both increase and decrease the cooling air flow rate as space temperature

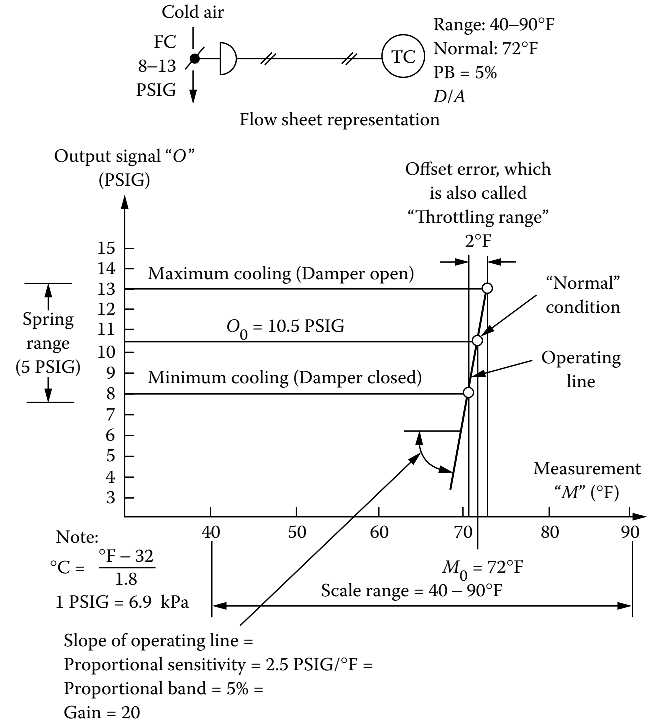


FIG. 7.9c
Operating characteristics of a narrow and fixed proportional band, pneumatic thermostat.

changes. If the cooling load doubles, requiring the damper to be fully open, this cannot take place until the controlled space temperature has first risen to 73°F (23°C). As long as the cooling load remains that high, the space temperature must also stay up at the 73°F (23°C) value.

Similarly, the only way this thermostat can reduce the opening of the cooling damper below 50% is to first allow the space temperature to drift down below 72°F (22°C). Therefore, stats do not have set points, but they have throttling ranges, and if a throttling range is narrow enough, this gives the appearance that the controller is keeping the variable near set point, when in fact the narrow range allows the variable to drift within limits.

Offset Error This controller has a fixed proportional band of 5%; it is is very sensitive, has a small "offset," and, as its gain is fixed, it cannot be tuned. Let us examine the consequences of these characteristics separately. A sensitive controller is suited for the control of very slow, large-capacity processors. Temperature control in the HVAC industry usually fits that description, and therefore thermostats with fixed, high gains can give acceptable results, if space temperature can change only very slowly. On the other hand, such thermostats cannot control spaces with fast dynamics (short time constants) and will cycle or lose control when applied to such service.

The "offset" inherent in all proportional controllers is not a serious drawback because the resulting error is small, due to the narrowness of the proportional band. Assuming that the thermostat output is set at 11 PSIG (0.76 bar) when the error is zero (thermostat is on set point), we can calculate the maximum error due to offset. When there is no error, the control valve is 50% open (output is 11 PSIG, or 0.76 bar).

As the load changes, an error must be allowed to develop in order for the valve opening to change. Assuming that the valve has been correctly sized and is large enough to handle all expected loads, the maximum error due to offset will be that deviation that is required to move the valve from half to full opening. If the thermostat's sensitivity is 2.5 PSIG/°F, this 2 PSIG (0.14 bar) change in thermostat output will occur when the deviation from set point is 0.8°F (0.4°C).

Therefore, this is the size of the permanent offset error for a gain of 19:1. There is an inverse linear relationship between gain and offset error, such that if the gain is cut in half, the offset error will be doubled. From the above discussion it might be concluded that the conventional room thermostat is a good selection for the HVAC-type applications and that more expensive instruments, such as PID controllers, would not necessarily improve the overall performance.

Throttling Range, Gain, or Sensitivity The control response of thermostats is described by the slope of the operating line in Figure 7.10c. This slope is described in different ways depending on the industry and manufacturer involved. People in the process control fields tend to be more familiar with terms such as "gain" and "proportional band," while people working in the HVAC field are more used to the terms "proportional sensitivity" or "throttling range."

All of these terms describe the same slope shown in Figure 7.9c and they are also defined below:

The ***throttling range*** is the gap within which the space conditions are allowed to drift as the final control element is modulated from fully closed to fully open (2°F or about 1°C in Figure 7.9c). Figure 7.9d describes the throttling range of a 3–13 PSIG (0.2–0.9 bar) thermostat.

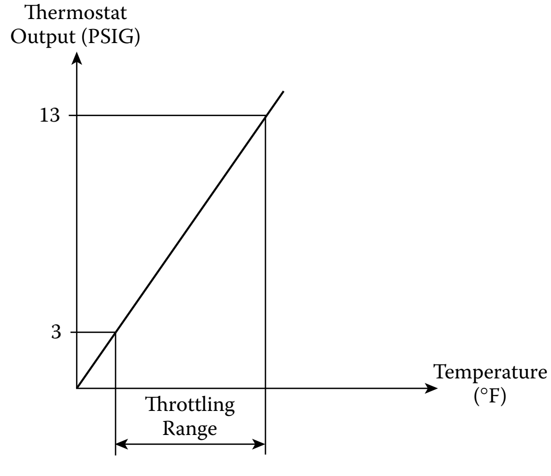


FIG. 7.9d
The throttling range of a 3–13 PSIG (0.2–0.9 bar) thermostat.

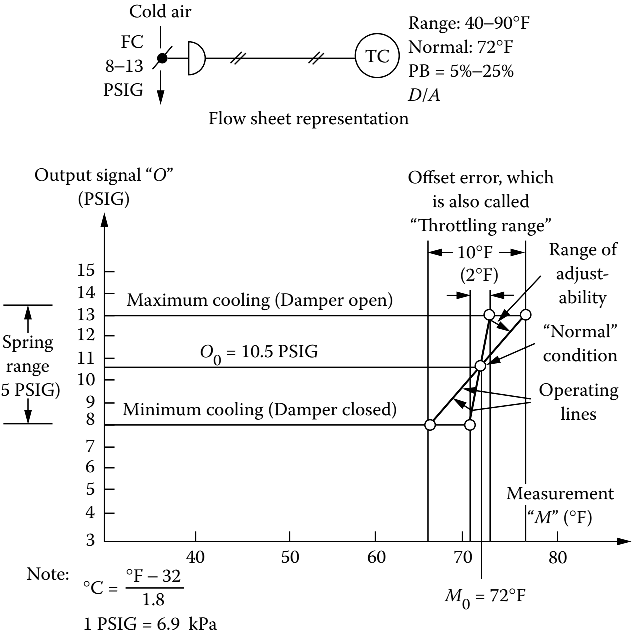
Proportional sensitivity is the amount of change in the output signal pressure that results from a change of one unit in the measurement (2.5 PSIG/°F in Figure 7.9c).

Gain is the ratio between the sizes of the changes in control output and measurement input. In case of Figure 7.9c, a 100% change in output (5 PSIG, moving the damper from fully open to fully closed) will result from a 2°F (1°C) change in measurement, which represents only 4% of the thermostat span of 40–90°F (6.6–32°C). Therefore, the gain of this thermostat is 25.

Proportional band (PB) is related to gain as follows: PB = 100/G = 100/25 = 4%. Therefore, the proportional band of the thermostat in Figure 7.9c is 4%.

Having defined the "normal" conditions (M_0 and O_0) and also the slope of the operating line, the behavior of the thermostat is now fully defined.

Gain or Sensitivity Adjustment The simplest thermostats are manufactured with fixed sensitivities. Such a unit was illustrated in Figure 7.9c, having a fixed proportional sensitivity of 2.5 PSIG/°F (0.3 bar/°C). Control flexibility is improved in those designs where the proportional sensitivity is adjustable. Figure 7.9e illustrates a thermostat with a 5:1 range of gain adjustment, which is typical to standard thermostats. Lowering the proportional sensitivity tends to improve stability, but it also increases the offset error due to the widening of the throttling range.



Range of adjustability:
Offset error or throttling range: 2 to 10°F (1.1 to 5.6°C)
Proportional sensitivity: 2.5 to 0.5 PSIG/°F (31 to 6.3 kPa/°C)
Proportional band: 5% to 25%
Gain: 20 to 4

FIG. 7.9e
The features and characteristics of a thermostat with adjustable sensitivity.

Thermostat Action and Spring Range

The thermostat in Figure 7.9c is said to be direct acting (D/A) because its output signal is increased as the measurement rises. If, instead of the Fail Closed damper, a Fail Open damper were used, a reversal in the thermostat action would be required, because now maximum cooling would take place when the thermostat output signal is low. Hence, in that case a reverse-acting (R/A) thermostat would be used, and the left-to-right slope of the operating line in Figure 7.9c would change to right-to-left.

Spring Ranges The operating line in Figure 7.9c varies with the spring range of the final control element. The typical spring ranges that are available are listed below:

Fail Closed HVAC Quality Dampers

A—3 to 7 PSIG (0.2 to 0.5 bar)
B—5 to 10 PSIG (0.35 to 0.7 bar)
C—8 to 13 PSIG (0.55 to 0.9 bar)

HVAC Quality Valves

D—Fail Open, 4 to 8 PSIG (0.9 t 0.55 bar)
E—Fail Closed, 9 to 13 PSIG (0.6 t 0.9 bar)
F—Three-way, 7 to 11 PSIG (0.5 to 0.76 bar)

Speed or Blade Pitch Positioners (Fail Closed)

G—3 to 15 PSIG (0.2 to 1.0 bar)

Figure 7.9c illustrated the operating line of a thermostat with a 2.5 PSIG/°F (0.3 bar/°C) proportional sensitivity in combination with a final control element having a type *C* spring. If the range spring of the final control element is changed, this would also change the operating lines, as illustrated in Figure 7.9f.

From Table 7.9g, it is obvious that the range spring selection can substantially affect the throttling range and the apparent proportional band. The wider the throttling range, the larger the control error (offset or drift away from the "normal" measurement value). On the other hand, the wider the throttling range, the more stable (less cycling) the final control element is likely to be.

Therefore, the trade-off is between lower stability or higher offset error. By narrowing the spring ranges to 4 or 5 PSIG (0.3 or 0.35 bar), the conventional HVAC control manufacturers have selected lower stability as the more acceptable of the two undesirable options. The process control industry has made just the opposite choice, as shown in Figure 7.9f by the 12 PSIG (0.8 bar) spring line (type *G*). This selection stems from the desire to maximize stability while being unconcerned with offset error because it was eliminated by the addition of the integral mode.

Thermostat Design Variations

Figure 7.9h illustrates the design of the simplest thermostat that is manufactured. This fixed-proportional controller uses

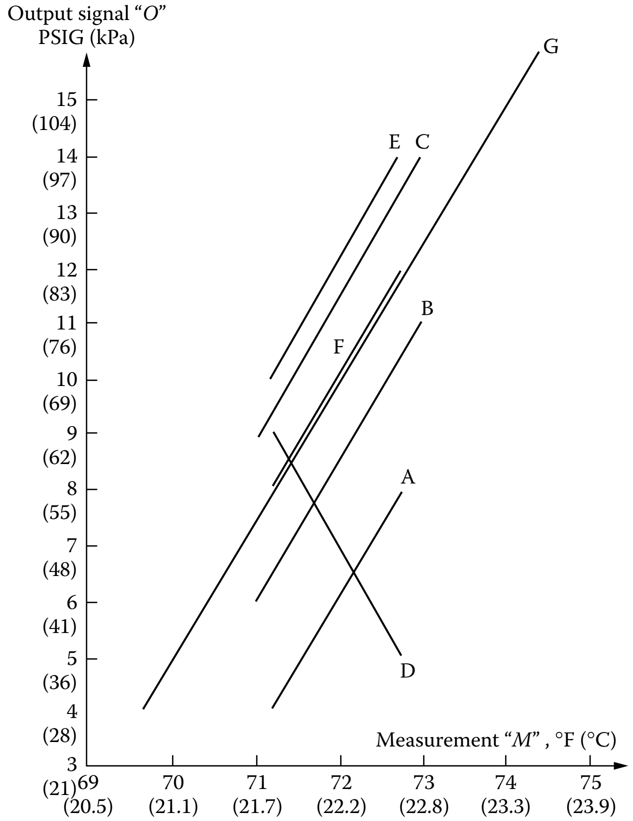


FIG. 7.9f

Illustrating the effect of damper or valve spring range on the gain or throttling range of the thermostat.

a flapper arrangement, which moves in front of a nozzle as the temperature detected by a bimetallic element changes relative to the manually adjusted "normal" value. Air at about 20 PSIG (140 kPa) is supplied through a restriction to the open nozzle, the back-pressure of which is inversely proportional to the distance between nozzle and flapper.

Because the A/B ratio is fixed in Figure 7.9h, the proportional sensitivity (K_c = change in output pressure per °F measured) is also fixed. Figure 7.9q describes a humidistat design in which the proportional sensitivity is adjustable.

Due to the small restriction and small nozzle openings in Figure 7.9h, the resulting output airflow is rather small, around 1.0 SCFH (0.028 m^3/hr). This type of design is also referred to as "nonrelay" or "low volume" design. The 1.0 SCFH air capacity can be insufficient to fill connecting tubing and operate final control elements. In order to increase the air capacity of a thermostat, a booster or repeater relay can be added.

Such "relay type" or "high-volume" thermostats will usually provide an output airflow of 10–30 SCFH (0.3–0.84 m^3/hr), which is sufficient for most applications. In the case of the relay-type design (Figure 7.9q), the nozzle back-pressure, instead of operating a final control element directly, is sent to the bellows chamber and acts against the bellows. Because the bellows has some stiffness (spring gradient), the pilot valve plug is positioned between the inlet and outlet ports in accordance with the nozzle back-pressure. A low nozzle back-pressure positions the valve plug to the left, throttles the exhaust port, and causes a high output pressure.

The advantages of adding the pilot are (1) the actuating signal vs. output pressure relationship can be made linear, and (2) the capacity for airflow can be considerably increased.

Special-Purpose Thermostats

Application engineers can choose from a fairly large variety of design features when specifying thermostats, because they can not only be electromechanical, pneumatic, or electronic, but they can also

- be indicating or blind
- be direct or reverse acting
- automatically switch their actions in response to a pneumatic or electronic signal
- have bimetallic, filled, or electronic sensing elements
- have local or remote set points
- have their set points under key, concealed, or externally adjustable.

Advanced Features If the set point of a pneumatic thermostat is adjusted remotely by an air signal, each 1 PSIG (0.07 bar)

TABLE 7.9g

Thermostat Gain (Proportional Sensitivity) and Throttling Range as a Function of Its Spring Range

Range Spring Type	PB (%)	GG	Proportional Sensitivity PSIG/°F (bar/°C)	Throttling Range °F (°C)	"Normal" Output PSIG (bar)	Spring Range PSIG (bar)
A	3.2	31	2.5 (0.3)	1.6 (0.9)	5.0 (0.35)	3–7 (0.2–0.5)
B	5.0	20	2.5 (0.3)	2.0 (1.1)	7.5 (0.52)	5–10 (0.35–0.7)
C	5.0	20	2.5 (0.3)	2.0 (1.1)	10.5 (0.72)	8–13 (0.55–0.9)
D	3.2	31	2.5 (0.3)	1.6 (0.9)	6.0 (0.4)	4–8 (0.3–0.55)
E	3.2	31	2.5 (0.3)	1.6 (0.9)	11.0 (0.76)	9–13 (0.6–0.9)
F	3.2	31	2.5 (0.3)	1.6 (0.9)	9.0 (0.6)	7–11 (0.5–0.76)
G	9.6	10	2.5 (0.3)	4.8 (2.7)	9.0 (0.6)	3–15 (0.2–1.0)

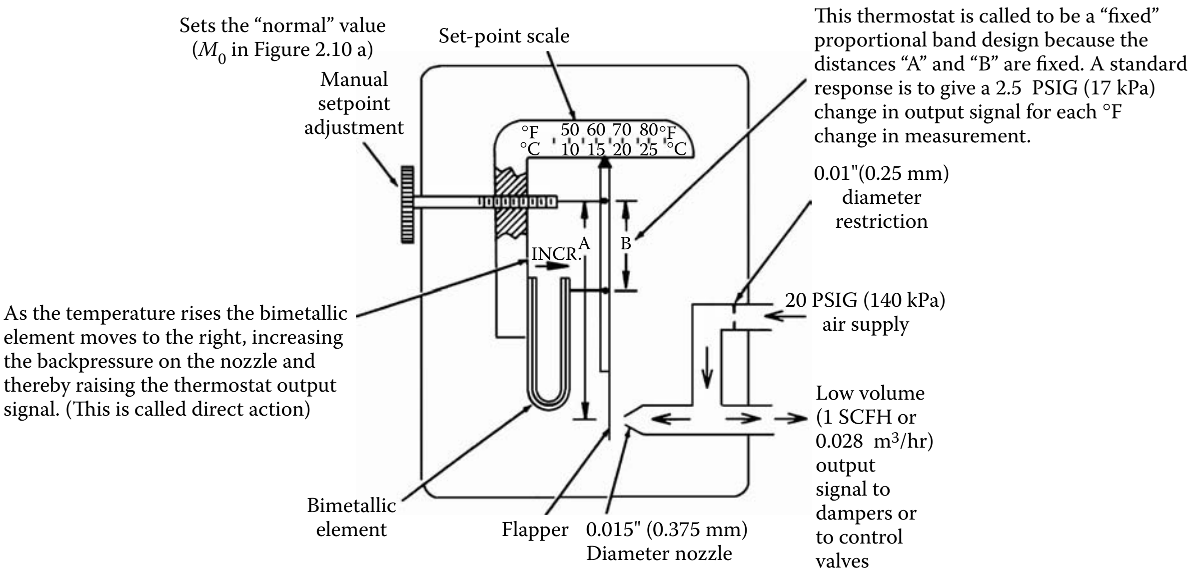


FIG. 7.9h

Direct-acting bimetallic thermostat with manual set point and fixed proportional band. The output airflow is limited because no booster relay is included. For this reason, it called a "low-volume" thermostat.

change in set point pressure will move the set point by an adjustable preset amount. The range of this adjustment is usually from 0.15–1.4°F (0.1–0.8°C) per 1 PSIG (0.07 bar). If the set point is to change as a function of the time of day, a timer can automatically operate a solenoid and thereby switch the set-point signal.

Some of the more recently developed and more advanced thermostat features include:

Dual Set Points Dual set point thermostats will switch their settings in response to a change in the air supply pressure. Both set points can be manually adjustable, with the day setting made by external thumbwheel and the night setting concealed internally.

One design variation is a dual thermostat that has two set points, one corresponding to the minimum, the other to the maximum allowable space temperature. These thermostats can use their dual outputs to operate a variable air volume (VAV) damper to cool the space and a reheat coil to heat it (Figure 7.9i).

In other control systems the dual outputs of the thermostat can be used to throttle the hot or the cold water to condition the air being supplied to the room (Figure 7.9j).

Night–Day or Set-Back These thermostats will operate at different "normal" temperature settings for day and night. They are provided with both a "day" and "night" setting dial, and the change from day to night operation can be made automatic for a group of thermostats.

The *pneumatic day–night* thermostat uses a two-pressure air supply system, the two pressures often being 13 and 17 PSIG (89.6 and 117 kPa) or 15 and 20 PSIG (103.35 and 137.8 kPa).

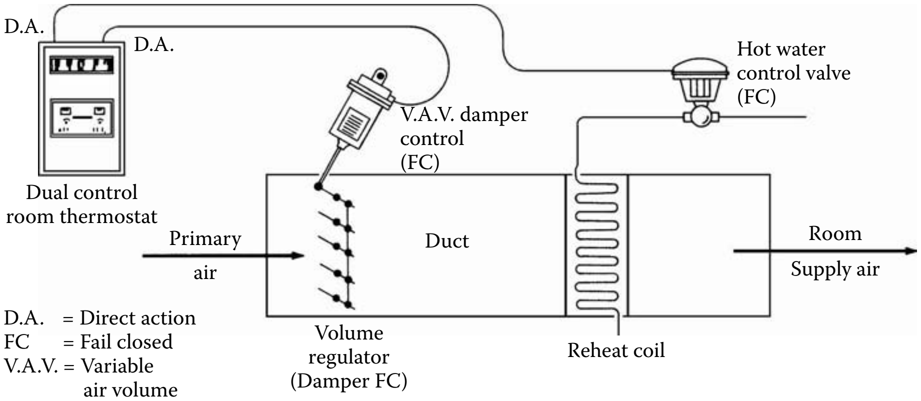


FIG. 7.9i

Dual thermostat for controlling a variable air volume (VAV) box with reheat. (Courtesy of Powers Controls.)

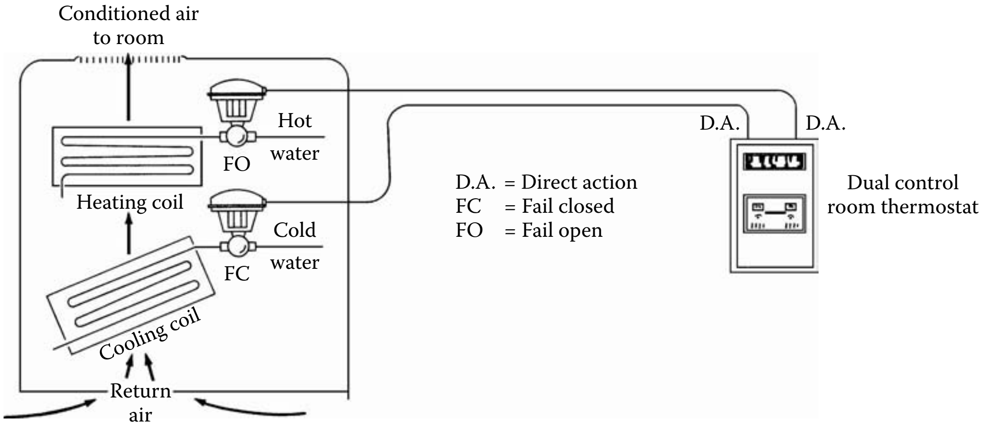


FIG. 7.9j
Dual thermostat installed to control a heating and a cooling coil. (Courtesy of Powers Controls.)

Changing the pressure at a central point from one value to the other actuates switching devices in the thermostat and indexes them from day to night or vice versa. Supply air mains are often divided into two or more circuits so that switching can be accomplished in various areas of the building at different times. For example, a school building may have separate circuits for classrooms, offices and administrative areas, the auditorium, and the gymnasium and locker rooms. In some of the electric designs, dedicated clocks and switches are built into each thermostat.

Heating–Cooling or Summer–Winter This thermostat can have its action reversed and, if desired, can have its set point changed by means of indexing. It is used to actuate controlled devices, such as valves or dampers, that regulate a heating source at one time and a cooling source at another. It is often manually indexed in groups by a switch, or automatically by a thermostat that senses the temperature of the water supply, the outdoor temperature, or another suitable variable.

In the heating–cooling design there frequently are two bimetallic elements, one being direct acting for the heating mode, the other being reverse acting for the cooling mode. The mode switching is done automatically in response to a change in the air supply pressure, similarly to the operation described for the day–night thermostats.

Limited Control Range These thermostats allow the occupant of an office to move the set point to any value desired, but will disregard any setting that exceeds the limit value. For example, in heating applications, the limit could be 74°F (23°C). In this case the space temperature will be limited to a maximum of 74°F, regardless of the setting by the occupant. In the cooling season the minimum limit value of 75°F (24°C) can be set for cooling. This is done internally without placing a physical stop on the setting knob.

Zero Energy Band Control The idea behind *zero energy band* (ZEB) control is to conserve energy by not using any when the room is comfortable. Zero energy band thermostats will provide heating when the zone temperature is below 72°F (22°C) and will provide cooling when it is above, say, 78°F (25.6°C) (these are adjustable settings), and in between, they will neither heat nor cool.

One way to achieve this goal is to use the dual thermostat illustrated in Figures 7.9i and 7.9j. If in these installations a single set-point thermostat was used instead of a dual one, energy would be wasted, because as soon as heating was stopped, cooling would be initiated (Figure 7.9k). This is wasteful, as no energy needs to be added or removed from a space that is comfortable as it is (within the ZEB band). This approach can reduce the yearly operating cost by up to about 33%.

ZEB control can be accomplished in one of two ways. The dual set-point approach is shown on the right of Figure 7.9l. The single set-point and single output approach is illustrated on the left side of Figure 7.9l.

In the single set-point, spilt-range configuration, the cooling valve fails closed and is shown to have an 8–11 PSIG (0.55–0.76 bar) spring range, while the heating valve is selected to fail open and has a 2–5 PSIG (0.14–0.34 bar) range. Therefore, between 5 and 8 PSIG (0.34 and 0.55 bar) both valves are closed, and no pay energy is expended while the thermostat output is within this range. The throttling range is usually adjustable from 5–25°F (3–13°C). Thus, if the ZEB is 30% of the throttling range, it can be varied from a gap size of 1.5°F (.85°C) to 7.5°F (4.2°C) by changing the throttling range (or gain).

ZEB control can also allow buildings to become self-heating by transferring interior heat to the perimeter. (For a detailed discussion of self-heating, see Section 8.2.)

Slave or Submaster This thermostat has its set point raised or lowered over a predetermined range, in accordance with variations in the output from a master controller. The master controller can be a thermostat, manual switch, pressure controller, or similar device. For example, a master thermostat measuring outdoor air temperature can be used to adjust a submaster thermostat controlling the water temperature in a heating system. Master–submaster combinations are sometimes designated as single-cascade action.

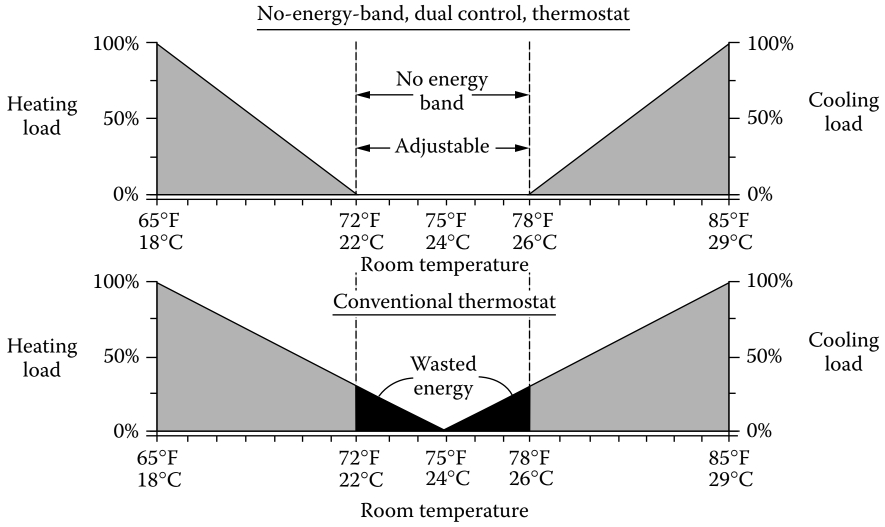


FIG. 7.9k
Energy can be saved by neither cooling nor heating a space if its temperature is within a zone that is comfortable.

When such action is accomplished by a single thermostat having more than one measuring element, it is known as compensated control.

Split-Range Control *Multistage* thermostats are designed to operate two or more final control elements in sequence. While the split-range approach is a little less expensive than the dual set-point scheme (shown on the right of Figure 7.9l), it is also less flexible and more restrictive. The two basic limitations of the split-range approach are:

1. The gap width can only be adjusted by also changing the thermostat gain; maximum gap width is limited by the minimum gain setting of the unit.
2. In this design, the heating valve must fail open, which is undesirable from an energy conservation point of view.

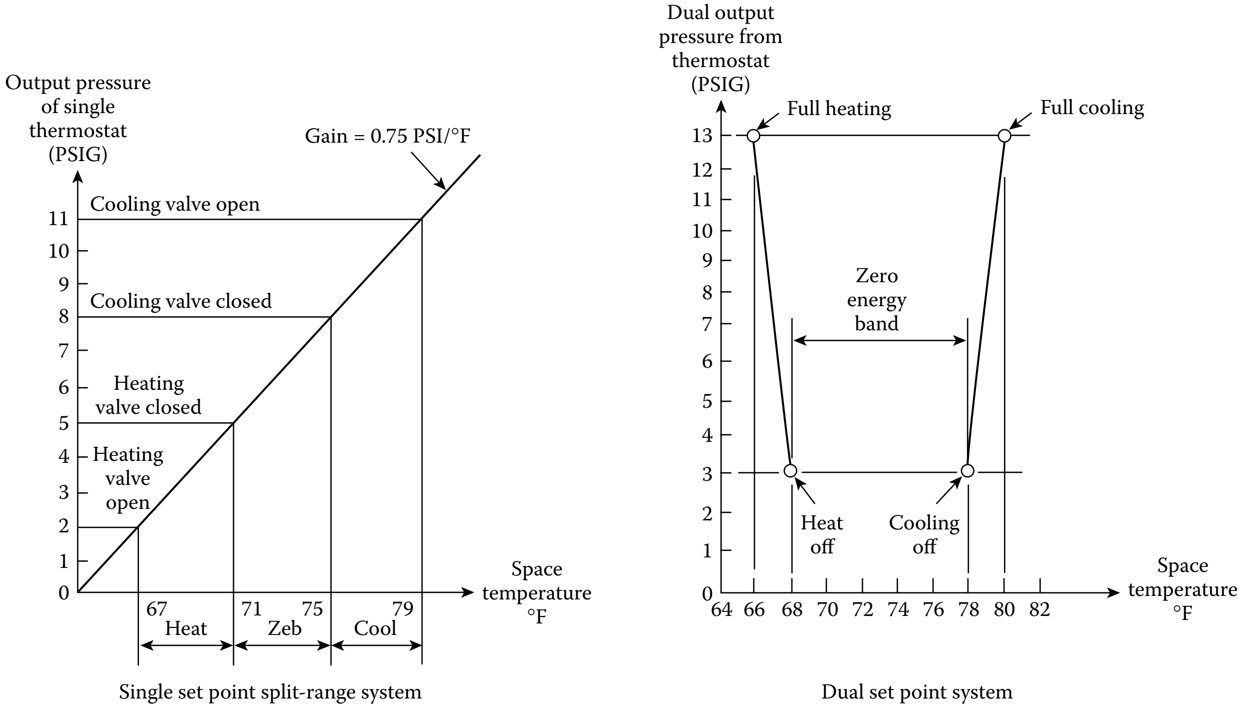


FIG. 7.9l
The zero energy band (ZEB) can be obtained by single set-point split-range design (left) or by dual set-point system (right).

These limitations are removed when a dual set point, dual output thermostat is used. Here both valves can fail closed and the bandwidth is independently adjustable from the thermostat gain. The gains of the heating and cooling thermostats are also independently adjustable. In Figure 7.9l, the heating thermostat is reverse acting and the cooling thermostat is direct acting.

For a detailed discussion of controlling and optimizing HVAC systems, refer to Section 8.2.

HUMIDISTATS

Basically the same design variations are available for humidity control as have already been discussed in connection with thermostats. By replacing the temperature sensor element, such as a bimetallic spring, with a humidity-sensitive sensing element, such as a polymer, a humidistat can be obtained.

A *wet-bulb* thermostat is also often used for humidity control with proper control of the dry-bulb temperature. A wick, or other means for keeping the bulb wet, and rapid air motion to ensure a true wet-bulb measurement are essential.

A *dew-point* thermostat is a device designed to control from dew-point temperatures.

Just like thermostats, humidistats can also be on/off, pneumatic, electric, electronic, or microprocessor-based digital. For the design features of these and for a discussion of the nature of fixed and narrow proportional control and the resulting offset, the reader is referred to the previous paragraphs.

Thermometers and hygrometers can also be combined into single instruments, and their microprocessor-operated versions can also be provided with memory, so they can recall past temperature and humidity records and can allow remote access to such data.

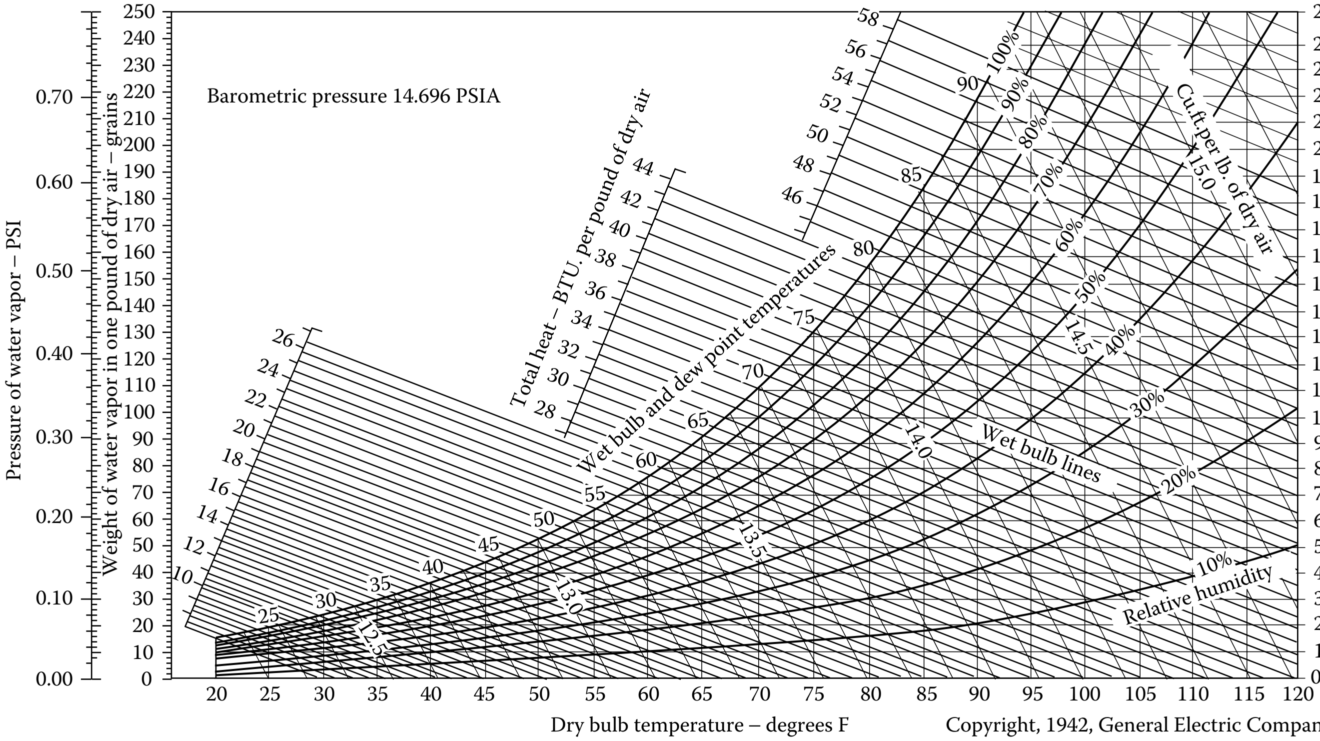
Relative Humidity Sensors

Humidity refers to the water vapor contained in the air at a particular temperature. Warm air has a greater capacity for water vapor than does cold air. Relative humidity is the mole fraction of moisture in air to the mole fraction of moisture in a saturated mixture at the same temperature and pressure. In other words, RH is the ratio of how much water vapor is in the air vs. the maximum it could contain at the particular pressure and temperature.

Relative humidity is the ratio of the actual partial pressure of the water vapor to the saturation vapor pressure at a particular temperature. Therefore, in a room that has 50% RH at 20°C, the RH will drop to 37%, if the temperature rises to 25°C. (Note that temperature has no effect on the dew point temperature, which in the above example stays constant at 9.3°C.)

Relative humidity can be calculated using the wet and dry bulb temperature readings, or relative humidity can be read from a psychrometric chart such as the one shown in Figure 7.9m.

The sensor in a humidistat can be moisture-sensitive hair, cellulose, synthetic polymer fiber, surface conductivity, or resistance or capacitance elements. Other, more accurate



Copyright, 1942, General Electric Company

FIG. 7.9m

One can read the relative humidity of air from the above psychrometric chart if one knows any two of the air properties shown in the chart.

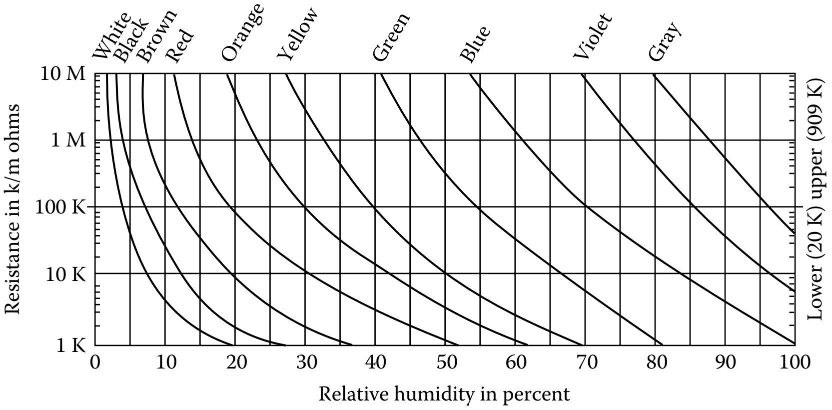


FIG. 7.9n
The chart describes the RH-to-resistance relationship of ten color-coded Dunmore sensor elements. (Courtesy of Ohmic Instruments Co.)

humidity and moisture sensors are also available and are discussed in detail in Section 8.32 in the first volume of this handbook.

Cellulose Hygrometers Cellulose strips or other shapes are used to measure humidity. Like human hair and synthetic materials, cellulose changes its dimensions as the water vapor concentration varies. This elongation has been used to build dial and digital indicators and also relatively inexpensive HVAC-type thermostats, transmitters, and recorders. The operation of the cellulose element instruments involves the detection of dimensional change of the element.

Solution Resistance Elements (Dunmore Cells) The Dunmore sensor consists of a wire grid on a insulating substrate that is coated with lithium chloride solution. Lithium chloride is hygroscopic and therefore takes up moisture from the air. The resulting resistance of the sensor is an indication of the relative humidity in the air. One disadvantage of this otherwise widely used, inexpensive, and good sensor is its narrow range.

Figure 7.9n illustrates the ranges of ten color-coded elements that are used by one manufacturer to cover the full range of 0–100% RH. Wide-range Dunmore sensors can be configured by placing several narrow-range elements in a single housing and connecting them to a microprocessor-based readout.

Polystyrene Surface Resistivity (Pope Cells) The Pope cell is similar to the Dunmore element, but instead of lithium chloride solution it uses polystyrene as a substrate for the conductive wire grid. In this design the polystyrene is treated with sulfuric acid, which produces a thin hygroscopic layer on its surface.

Changes in humidity cause large changes in the impedance of this layer, and because the operating portion of the sensor is only its surface, its speed of response is fast — only a few seconds. An AC-excited Wheatstone bridge circuit can be used to measure the impedance. The output signal is nonlinear and can cover a range of 15–99% RH. This type of humidity probe can be mounted on the wall or in ducts (Figure 7.9o).

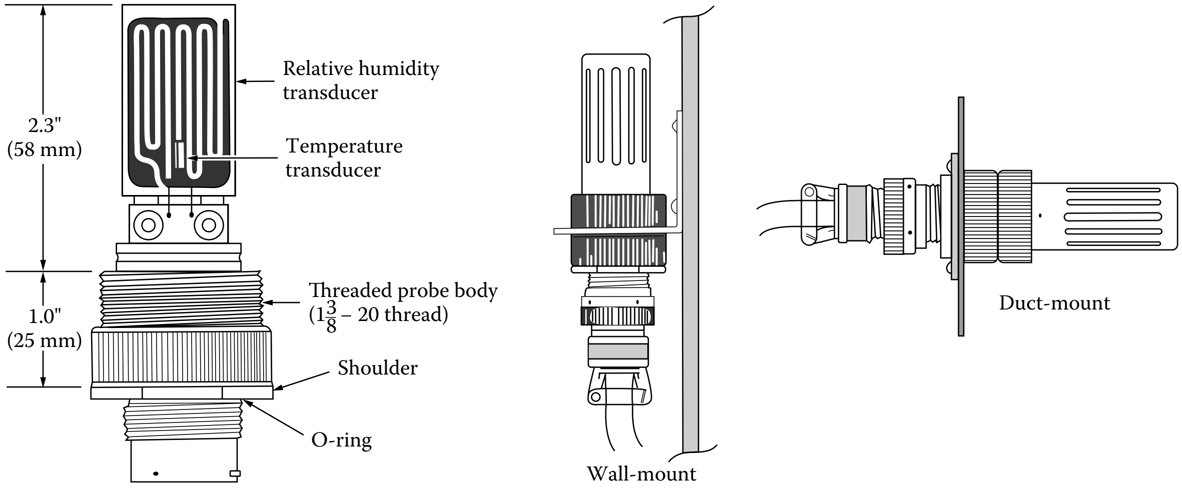


FIG. 7.9o
Polystyrene surface conductivity element. (Courtesy of General Eastern, which used to be Phys–Chem Scientific Corp.)

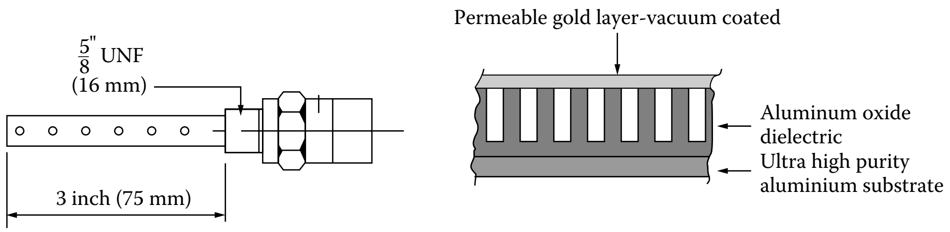


FIG. 7.9p
Thin-film aluminum oxide capacitance element packaged as a probe. (Courtesy of Michell Instruments Ltd.)

Thin-Film Capacitance The largest number of manufacturers use some form of a capacitance measurement to detect humidity or dew point in air.

Metal Oxide Sensor One variation is to form a capacitor by depositing a layer of porous aluminum oxide on a conductive aluminum substrate and coating the oxide with a thin film of condensation from evaporated gold. The aluminum substrate and the gold film serve as the electrodes of the capacitor (Figure 7.9p).

When exposed to air, the water vapors penetrate through the gold layer into the aluminum oxide dielectric and are absorbed by it. The amount of water absorbed determines the capacitance registered by the sensor. Some suppliers "dope" the aluminum oxide dielectric with lithium chloride to extend its range down to dew points of –94°F (–70°C). The disadvantages of lithium chloride doping include the need for individual calibration, the slowing of the response, and the potential for shorting due to condensation.

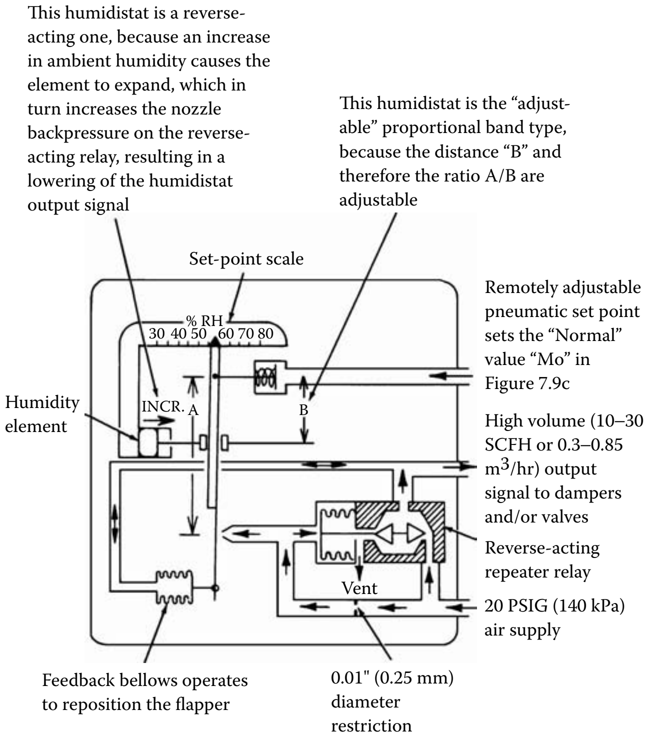
The accuracy of the aluminum oxide moisture sensor is low; each sensor requires a separate calibration curve, which is nonlinear; and the unit must be periodically recalibrated to compensate for aging and contamination. These sensors are designed for low dew-point measurements, but can be disabled if exposed to high humidity or if wetted. The advantages of these sensors include their small size, their probe-type packaging, and the fact that their measurement range is wide and is well suited for the detection of low dew points. As a consequence of these characteristics, they are widely used in such applications as HVAC, where cost is the primary concern and accuracy of measurement is not critical.

Polymer Sensor Design variations of the capacitance-type humidity sensor include the replacement of aluminum oxide dielectric with hygroscopic polymer dielectrics. These units consist of a capacitive polymer sensor bonded to a resistive temperature sensor. The polymer sensor detects humidity directly, while dew point is calculated by a microprocessor, which also reads the temperature. These sensors are claimed to provide better linearity and allow the use of longer lead wires, because their low-frequency operation reduces the effects of stray capacitance. Other advantages include resistance to contamination by various chemical vapors.

Polymer sensors are immune to condensed water and therefore can have a wider range than metallic oxide ones. These units are available as transmitters with intelligent electronics and are capable of detecting low dew points while requiring reduced maintenance due to their stability.

Humidistat Design Features

As shown in Figure 7.9q, the pneumatic humidistats are similar in design to thermostats, except that the bimetallic measuring element is replaced by a humidity-sensitive element. A common humidity-sensitive element is cellulose acetate butyrate. This substance expands and contracts with changes in relative humidity, and the resulting movement can be used to operate the flapper of the pneumatic humidistat shown in Figure7.9q or it can open and close the contacts of electric humidistats.


This humidistat is a reverse-acting one, because an increase in ambient humidity causes the element to expand, which in turn increases the nozzle backpressure on the reverse-acting relay, resulting in a lowering of the humidistat output signal

This humidistat is the "adjustable" proportional band type, because the distance "B" and therefore the ratio A/B are adjustable

Remotely adjustable pneumatic set point sets the "Normal" value "Mo" in Figure 7.9c

High volume (10–30 SCFH or 0.3–0.85 m^3/hr) output signal to dampers and/or valves

Feedback bellows operates to reposition the flapper

0.01" (0.25 mm) diameter restriction

FIG. 7.9q
Reverse-acting humidistat with pneumatic set point and adjustable proportional band. The use of the repeater relay increases the humidistat output flow capacity, and therefore this design is also referred to as "high volume."

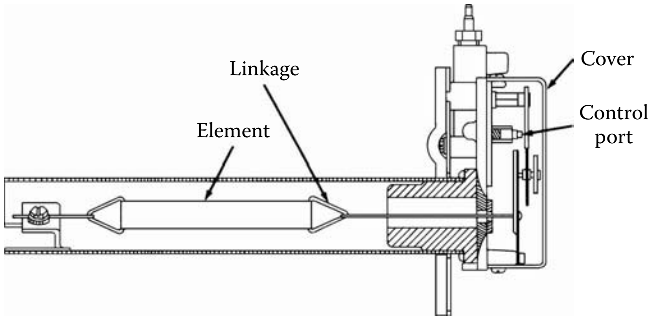


FIG. 7.9r
Duct-mounted humidistat. (Courtesy of Johnson Controls Inc.)

The choice of humidistat features and characteristics is similar to those of thermostats.

1. They can be direct or reverse acting.
2. They can be high or low volume (relay type or non-relay type).
3. The "normal" value can be set manually or by remotely adjustable pneumatic signal.
4. The proportional sensitivity determined by the A/B ratio can be fixed or adjustable.
5. Feedback bellows can be added to minimize the effects of variations in supply pressure or temperature, to reduce output leakage, and to increase the range of adjustability of the proportional sensitivity setting.

Humidistats are also installed in air ducts to control upper and lower limits of the relative humidity of the air. One such duct-mounted device is illustrated in Figure 7.9r. This unit is usually set for either 85% RH or for 35% RH.

Bibliography

Air-Conditioning and Refrigeration Institute (ARI) Standard 750-76 (1976), "Thermostatic Refrigerant Expansion Valves," ARI, 1976.

American Society of Testing and Materials (ASTM) Standard B106-68 (1972), "Standard Method of Test for Flexibility of Thermostat Metals," ASTM, 1972.

ANSI: Z21.23-1975, "Gas Appliance Thermostats," American Gas Association (AGA) standard, 1975.

ASHRAE Handbook and Product Directory, 1980 Systems, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 1980.

ASHRAE Standard 17-75 (ANSI: B60.1-1975), "Method of Testing for Capacity Rating of Thermostatic Refrigerant Expansion Valves,"ASHRAE, 1975.

ASTM Standard B389-65 (1975), "Standard Method of Test for Thermal Deflection Rate of Spiral and Helical Coils of Thermostat Metal," ASTM, 1975.

Carlson, D. R., "Temperature Measurement in Process Control," *InTech,* October 1990.

Eckman, D. P., *Automatic Process Controls,* New York: John Wiley & Sons, 1958.

"Filled Systems Thermometers," *Measurements and Control,* September 1992.

Kerin, T. W., and Katz, E. M., "Temperature Measurement in the 1990s," *InTech,* August 1990.

Lipták, B. G., "Reducing the Operating Costs of Buildings by the Use of Computers," *ASHRAE Transactions,* Vol. 83, Part I, 1977.

National Electrical Manufacturers' Association (NEMA) Standard DC13-1973, "Residential Controls — Integrally Mounted Thermostats for Electric Heaters," NEMA, 1973.

NEMA Standard DC3-1972, "Residential Controls — Low Voltage Room Thermostats," NEMA, 1972.

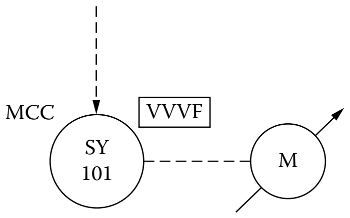
Spethmann, D. H., "Importance of Control in Energy Conservation," *ASHRAE Journal,* February 1975.

Underwriters' Laboratories (UL) Standard UL-521 (ANSI: Z220.1-1971), "Standard for Fire-Detection Thermostats," UL, 1971.

Warmoth, D., "Temperature Measurement & Control," *Measurements and Control,* February 1986.

7.10 Variable-Speed Drives

A. BRODGESELL (1970) **R. D. BUCHANAN, J. B. RISHEL** (1985)
B. G. LIPTÁK, R. H. OSMAN (1995) **I. H. GIBSON** (2005)



Flow sheet symbol

Types of Drives:	A. Mechanical drives B. Hydraulic, hydroviscous, and fluid coupling drives C. Magnetic particle, fluid, and eddy current drives D. Variable voltage E. Variable-speed DC, thyristor DC F. Variable-frequency AC F1. Voltage source, pulse-width modulated (PWM) F2. Current source F3. Load commutated inverter (LCI) on synchronous motor F4. Power recovery versions of F1, F2, or F3
Abbreviations Used:	ASCI—Autosequentially commutated current-fed inverter GTO—Gate-turn-off thyristor IGBT—Insulated gate bipolar transistor LCI—Load commutated inverter PU—Per unit PWM—Pulse width modulated RMS—Root mean square SCR—Silicon-controlled rectifier VFD—Variable-frequency drive VVVF—Variable-voltage variable-frequency drive
Size Range:	A. 1 to 100 hp (0.75–75 kW) B. Up to 4000 (up to 3MW) C. Air-cooled up to 900 hp (670 kW); water-cooled up to 18,000 hp (13.5MW) D. 10 to 100 hp (7.5–75 kW) E. From under 1 to 500 hp (to 375 kW) F1. From under 1 to 500 hp (to 375 kW) F2. From under 100 to over 1000 hp (75 to 750 kW) F3. From 100 to 20,000 hp (75 kW to 15MW); generally above 1000 hp (750kW)
Speed Turndown:	A. 6:1 B. 3:1 to 40:1 C. 5:1 to 10:1 D. Low E. 10:1 with standard and 100:1 with tachometer feedback F. 3:1
Efficiency at 70% of Full Speed:	A. 50% B. 55% C. 58% D. 52% E. 80% F. 78%
Inaccuracy of Speed Control:	2 to 5% in standard (can do better with feedforward) and 0.1 to 1% with tachometer feedback designs

Variable Speed Drive Costs (Excluding the Motor):
E. For a 200 hp (150 kW) motor, a single converter, and thyristor DC drive: $13,000 F1 and F2. For a 200 hp (150 kW) induction motor: $20,000
F3. $200/hp ($270/kW) at around 1000 hp (750 kW) size and $100/hp ($130/kW) at around 5000 hp (3.75MW) size
(For the costs of other sizes and designs, see Tables 7.10r and 7.10u.)

Partial List of Suppliers:
ABB Inc. Automation Technologies Drives and Motors (www.abb-drives.com)
Allen-Bradley Div. of Rockwell Automation (www.ab.com/drives)
ASIRobicon Inc. (www.asirobicon.com)
Baldor Electric Co. (www.baldor.com)
Cutler-Hammer Div. of Eaton Corp. (www.eatonelectrical.com)
Danfoss Inc. North America Motion Controls (www.namc.danfoss.com)
Emerson Control Techniques (www.emersonct.com/index.htm)
Eurotherm Drives (www.eurothermdrives.com)
General Electric/GE-Fuji Electric Co. (www.geindustrial.com/cwc/home)
Hitachi Ltd. (www.hitachi.us)
Lenze Corp. (www.lenzeusa.com)
Mitsubishi Electric Corp. (www.meau.com)
Omron IDM (www.idmcontrols.com)
Reliance Electric Div. of Rockwell Automation (www.reliance.com)
Robicon Corp. (www.robicon.com/products/acdrives/index.html)
Saftronics Inc. (www.saftronics.com)
Siemens Energy & Automation (www.sea.siemens.com/drives/default.html)
Square D Div. of Schneider Electric (www.squared.com)
TB Woods Inc. (www.tbwoods.com)
TECO-Westinghouse Motor Company (www.tecowestinghouse.com/Products/drives.html)
Telemechanique Div. of Schneider Electric (www.modicon.com/Default.htm)
Toshiba America (www.tic.toshiba.com/products.php)
Unico Inc. (www.unicous.com)
US Drives Inc. (www.usdrivesinc.com)
Yaskawa Electric America Inc. (www.magnetekdrives.com)

INTRODUCTION

This section begins with a brief explanation of the role of variable-speed drives in improving the efficiency of transporting materials. After that, the different variable-speed drive designs are described, starting first with the older electromechanical designs and then proceeding to the more modern electrical and solid-state drives.

TRANSPORTATION EFFICIENCY

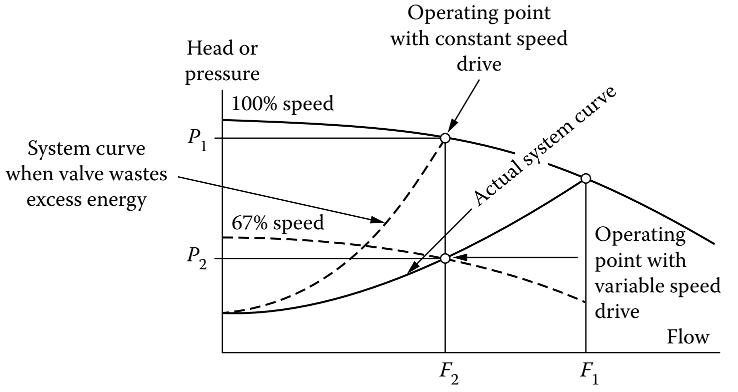
Constant or variable-speed drives can be applied to the transportation of liquids and gases. The amount of discharge pressure needed in these transportation systems is a function of the flow rate demanded by the process, the load. This relationship between the required flow and pressure is called the system curve. This curve rises with load (Figure 7.10a), while the pump curve drops with increasing flow. When constant speed drives are used, the pump, fan, or compressor curve is fixed, and therefore it cannot coincide with the system curve at more than one point.

As shown in Figure 7.10a, a control valve or damper needs to be introduced to burn up (waste) the excess energy, which should not have been introduced in the first place. Variable-speed drives eliminate this waste by shifting the pump or fan curve to cross the system curve, thereby eliminating waste by introducing only as much transportation energy as is needed to meet the load (Figure 7.10a). This subject is covered in Chapter 8, where the control of variable-speed pumps, compressors, fans, and turbines is covered in some detail.

CHARACTERISTICS OF VARIABLE-SPEED DRIVES

All drives for centrifugal pumps, regardless of type, must have certain characteristics to make them acceptable for operating centrifugal pumps. Some of these characteristics are as follows:

1. Broad speed turndown without damage to the motor or variable speed drive. The operating speed range for most variable speed pumps is from 50 to 100% of the synchronous speed. However, through misadjustment



Speed, %	*Flow, %*	*Horsepower required, %*
100	100	100
90	90	73
80	80	51
70	70	34
60	60	22
50	50	13
40	40	6
30	30	3

FIG. 7.10a

When pumping with a constant speed drive, energy is wasted, as the unnecessarily introduced pumping energy is burned up in a flow control valve. In contrast, when the flow rate is reduced by lowering the speed, such energy is not introduced in the first place.

or momentary system dynamics, the controls may cause the pump to run at zero speed. The drive must be able to function with zero pump speed without damage or actuation of circuit protection devices such as thermal overloads for a significant period of time.

2. Control repeatability, so that fluid system controls can achieve desired pump speeds at all loads.
3. Reliability. The variable-speed drive and motor must be designed to fit field operating conditions that vary appreciably in centrifugal pump applications. Ambient temperature in the field is often outside the "nominal" rating of the motor. Ambient temperature of the drive electronics may also be other than anticipated by the designer.
4. Serviceability. The personnel normally employed for maintenance of motors and pumping equipment must be capable of servicing the electronic, electrical, or mechanical equipment involved on each specific application.

ELECTROMECHANICAL DRIVES

Electromechanical drives include those that utilize an electrical motor to drive a mechanical, speed-changing device that is connected to the centrifugal pump. These include the following:

1. V-belt drives
2. Hydraulic couplings
3. Hydroviscous couplings
4. Eddy current couplings

There are a great number of methods for speed adjustment of rotating machines. These methods fall into two categories: speed adjustment of the prime mover and speed adjustment through a transmission connecting the driver to the driven machine.

Within each of these two groups several degrees of sophistication are possible, ranging from manually actuated stepwise speed changes to continuously variable automatic speed changers. Each method of speed control offers certain advantages and disadvantages (Table 7.10b) that must be weighed against the design criterion to permit a proper selection.

Mechanical Variable-Speed Drives

Stepped Speed Control Mechanical methods of speed adjustment offer a number of gear and pulley devices for both stepped and continuously variable-speed control. Stepped speed control methods provide setting a number of speeds very accurately. However, they are not readily adaptable to automatic process control.

The stepped pulley system shown in Figure 7.10c is one of the earliest methods of speed adjustment. Its advantages are low cost and simplicity, but belt slippage contributes to inefficiency, high maintenance, and reduced speed control. Two factors to be considered in the design of a stepped pulley system are the proper ratio of pulley diameters to obtain the desired speeds and proper pulley dimensions to maintain belt tension for all positions.

Pulley dimensions for a system such as in Figure 7.10c can be determined from the relationships in Equations 7.10(1) and 7.10(2).

$$\frac{\pi}{2}(R_1 + r_1) + \frac{(R_1 - r_1)^2}{4d} = \qquad \mathbf{7.10(1)}$$

$$\frac{\pi}{2}(R_2 + r_2) + \frac{(R_2 - r_2)^2}{4d}\frac{S_2}{S_1} = \frac{R_2}{r_2} \qquad \mathbf{7.10(2)}$$

The variables in the above equations are defined in Figure 7.10c.

TABLE 7.10b

Features and Applications of Various Rotary Drives

	Type of Drive				
Service and Feature	*Electric Motor*	*Eddy Current or Magnetic Couplings*	*Mechanical Stepped-Speed Transmissions*	*Continuously Variable Mechanical Drives*	*Hydraulic Drives*
Very wide range of speeds	✓				
Few speed steps with remote control	✓				
Few speed steps with local and manual output	✓		✓		
Very high output power at variable speed	✓	✓			
Vibration at load or driver		✓			
Small or moderate load with narrow speed range				✓	
Accurate speed control	✓	✓			✓
Shock loads, frequent overloads					✓
Speed reversal required	✓		✓		✓

Gear transmissions offer high efficiency of power transmission at precise stepped speed control. For speed changes of gear train drives, either clutches or brakes are required or the change must be made at rest. Epicyclic gears, such as planetary gears, offer the most compact unit, operating quality, and high efficiency. However, auxiliary clutches and brakes are required, and the cost is therefore higher than that of other gear drives. The complexity of control arrangements increases rapidly with the number of speeds required, making the planetary gear drive impractical above four or five speeds.

Continuously Variable-Speed Drives The cone pulley systems in Figure 7.10d is a natural evolution of the stepped pulley system shown in Figure 7.10c. Cone pulleys are designed similarly to stepped pulleys. A series of diameters is calculated equidistant on the pulley axis. The diameter end points are joined to form the pulley contour.

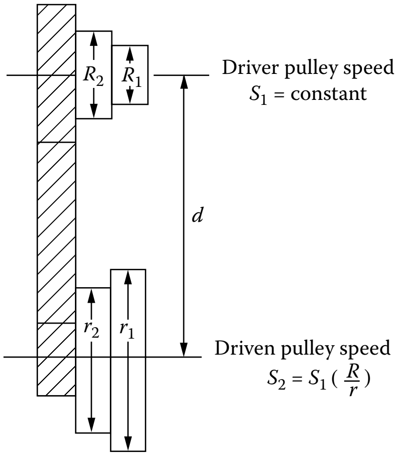


FIG. 7.10c
Early method of speed adjustment: the stepped pulley system.

The cone pulley is inherently inefficient, because contact surface speed varies across the belt, causing slippage. Belts must be kept narrow to reduce slippage and wear, and thus the capacity for power transmission is reduced. Belt guides at the pulleys are required to hold the belt in position. These guides must move simultaneously for speed changes.

Variable-Pitch Pulleys Cone pulleys were the forerunners of the more sophisticated variable-pitch pulley systems, which permit continuous, automatic speed adjustments over wide ranges. Speed adjustment in all of these systems is obtained by means of sliding cone face pulleys or sheaves, whose effective diameter can be changed.

A simple variable-pitch sheave is shown in Figure 7.10e. Two flanges are mounted on a threaded hub. The flanges are set to the desired spacing and locked in place with a set-screw. Speed adjustments are accomplished by changing the flange spacing, producing in effect a pulley of different diameter. This method is very economical, but it requires stationary

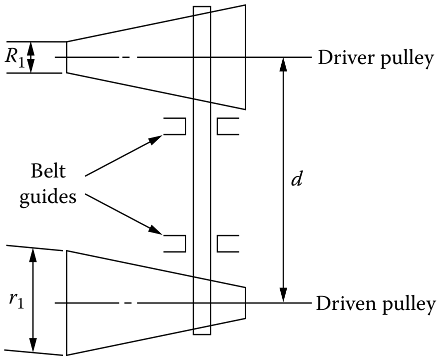


FIG. 7.10d
Continuous variation of speed can be provided by the cone pulley system.

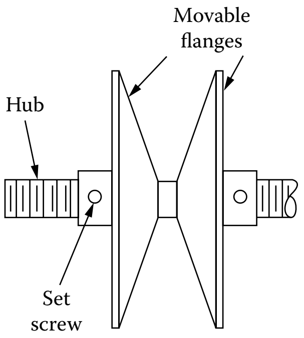


FIG. 7.10e
Automatic speed adjustments over wide ranges can be provided by the variable-pitch sheave.

speed adjustment and special adjustable motor bases to maintain belt tension.

In place of the adjustable motor base, a spring-loaded flat-face idler pulley can be used to maintain belt tension when center-to-center distance must be held constant. Alignment of driving and driven sheaves is also critical, because belt wear will be severe with poor alignment. Also, speed adjustment with this drive is limited because the set-screw must engage a flat surface.

A number of designs are available with in-motion speed adjustment. The designs vary from manual speed adjustment, through a crank or handwheel, to automatic actuators. Generally these drives use one sheave with adjustable pitch and one spring-loaded sheave that automatically adjusts itself to maintain belt tension (Figure 7.10f).

Speed adjustment is obtained by means of a mechanical linkage that moves one of the flanges of the driving sheave. The opposite flange on the driven sheave is spring loaded and moves to maintain belt tension and alignment. Radial motion of the belt along the flange faces in response to speed changes is facilitated by the rotary motion of the sheaves. When speed adjustments are made at rest, the belt cannot move radially and the adjusting linkage may be damaged. On automatic actuators, an interlock should be provided to vent air off the diaphragm whenever the drive is stopped and inhibit speed changes at rest.

Variable-pitch sheave drives are available as package units including motor. Speed adjustment ranges of 4:1 are common, but higher ranges to 10:1 are possible. Horsepower ranges to 100 hp (75 kW) are also possible. For very narrow speed ranges, drives to 300 hp (225 kW) are available.

Hydraulic Variable-Speed Drives

In hydraulic variable-speed drives, a pump with fixed or variable displacement drives a hydraulic motor, which itself can have fixed or variable displacement. The pump is driven by an electric motor at fixed speed. Output speed is controlled by changing pump or motor displacements.

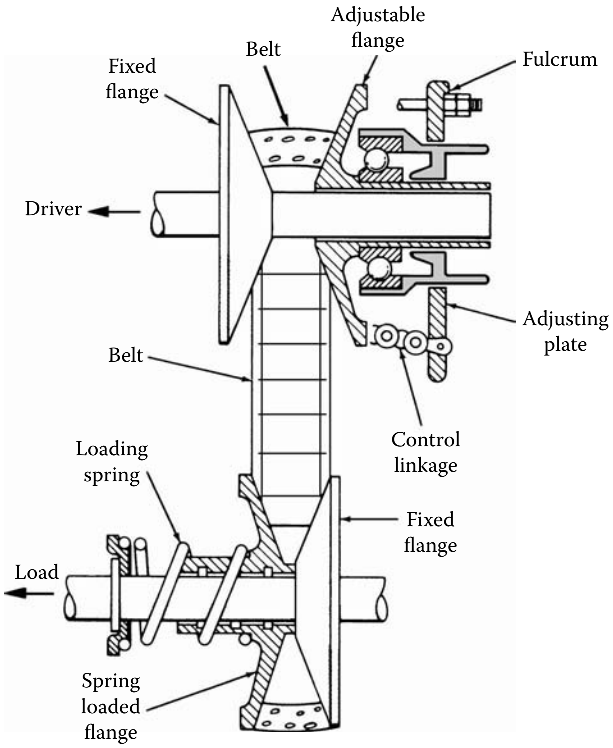


FIG. 7.10f
The main components of a continuously adjustable mechanical speed transmission system.

Pumps and motors are virtually identical, differing only in the location of the power input and output. Several designs of variable-displacement units are possible, including axial piston, radial piston, and vane types. An axial piston unit is shown in Figure 7.10g.

Of course, combinations of pumps and motors with variable or fixed displacement are possible for speed control. Pumps with variable displacement combined with motors with fixed displacement yield a variable-horsepower, fixed-torque drive. Motor speed reversal is possible by reversing the pump stroke.

Combinations of fixed-displacement pump with variable-displacement motors produce drives with fixed horsepower and variable torque. Speed reversal in this combination, however, requires the use of a valve to reverse the supply and return connections at the motor. A variable-displacement pump in conjunction with a variable-displacement motor has characteristics intermediate between the two previously discussed combinations. Range of speed control, however, is widest for this combination.

Pumps and motors with fixed displacement can be used for speed control by controlling the amount of fluid delivered to the motor. This can be accomplished by means of a pressure relief valve on the pump discharge and throttling valve in the oil line to the motor. Motor reversal can also be accomplished by means of a four-way valve, reversing the oil supply

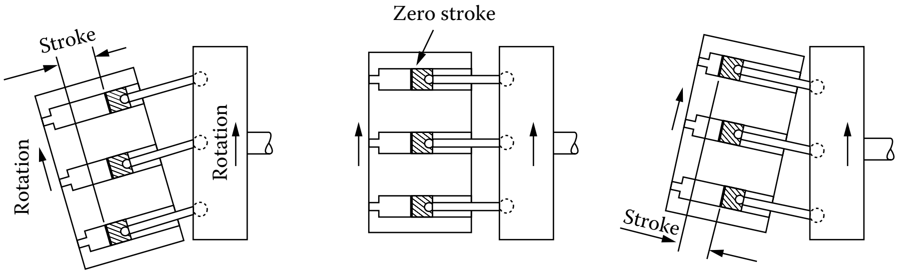


FIG. 7.10g
Axial piston pump provides the variable displacement in this hydraulic variable-speed drive.

and return lines at the motor. These methods, however, are comparatively inefficient and are not used with automatic control.

Hydraulic drives are available as pump-motor packages, or the motor can be mounted remotely and connected to the pump by hydraulic tubing. Displacement is adjustable through push-pull rods or handwheels, and both methods are adaptable for operation by hydraulic (usually integral) and pneumatic actuators or by electric motors for automatic speed control. Speed ranges are adjustable up to about 40:1 at power ratings up to 4000 hp (3 MW).

Several attributes of hydraulic variable-speed drives are clear advantages over the other types of drives discussed. Hydraulic drives offer the fastest response in acceleration, deceleration, and speed reversal. They are generally better suited than other types of drives for high shock loads, frequent speed-step changes, and reversals.

A disadvantage associated with hydraulic drives is frequent fluid leaks. Besides requiring maintenance, leakage of hydraulic fluid creates safety hazards that can preclude their use. Nonflammable hydraulic fluids are available, but the choice of compatible fluid and equipment may be limited. For offshore applications, once-through hydraulic systems are sometimes used. These are normally water-based fluids with low-marine-toxicity additives.

Hydraulic systems normally operate at pressures of 1500–3000 psi (10–20 MPa), but higher pressure equipment is available and can save weight. To avoid cavitation damage, it is usual to maintain the entire return system at a slight positive pressure in the order of 70 psi (500 kPa).

Fluid Couplings Fluid couplings consist of two coupling halves, one driven by a standard electrical motor and the other connected to the centrifugal pump. Oil is circulated in the coupling to regulate pump speed. A splitter or diverter assembly applies the oil to the coupling or bypasses it around the coupling. The pump shaft speed increases as the amount of oil supplied to the coupling is increased, thus producing a simple control system for varying pump speed.

The fluid coupling has historically been one of the most popular variable-speed devices for centrifugal pumps because of its relatively low first cost, high reliability, and ease of maintenance. It is a slip-type device, like the eddy current coupling, and has a lower efficiency curve than the variable frequency, direct current, and wound rotor regenerative types of variable speed drives.

Hydroviscous Drives Hydroviscous drives are similar to fluid couplings in configuration and operation. Instead of an oil-filled coupling, one or more disk assemblies are used, with driving and driven members pressed together by oil pressure. Increasing the oil pressure increases the pump speed; likewise, reducing the oil pressure reduces the pump speed.

The hydroviscous drive is, therefore, a slip-type device, like the eddy current coupling, and has a lower efficiency curve than most electrical-type drives. This drive often requires a more complex control system than other drives because of variations in oil pressure caused by oil viscosity or temperature.

Magnetic Variable-Speed Drives

Eddy Current Couplings The eddy current coupling was for many years the most acceptable drive for centrifugal pumps, particularly for large vertical turbine pump applications. It is rugged in design and available in both horizontal and vertical configurations. The eddy current coupling uses a standard, constant-speed motor that drives the eddy current coupling, which consists of a drum connected to the electric motor and electromagnetic pole-type rotor assembly, which is connected to the centrifugal pump, inside of and free from the drum.

The speed control system regulates the amount of flux that exists between the drum and rotor assemblies. The amount of slip or speed difference between motor and pump increases and decreases with the flux density in the coupling. Being a slip-type, variable-speed device, the eddy current coupling has a lower efficiency than the variable frequency, direct current, and wound rotor regenerative types of variable speed drives.

In most electrical machinery, eddy currents are detrimental to operating efficiency, and great pains are taken to eliminate them. In the eddy current coupling, however, these

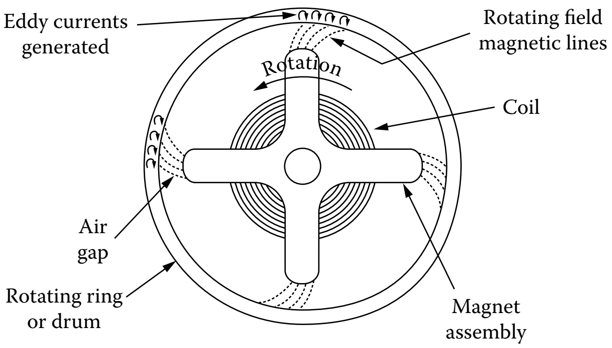


FIG. 7.10h
The main components of an eddy current coupling.

currents are harnessed and are the basis for the operation of the coupling. The eddy current coupling is a nonfriction device where input energy is transferred to the output through a magnetic field.

The eddy current coupling consists of a rotating magnet assembly separated from a rotating ring or drum by an air gap. In addition, a coil is wound onto the magnet assembly (Figure 7.10h), or, on larger units, the coil is stationary on the coupling frame.

There is no mechanical contact between the magnets and drum. When the magnet assembly is rotated, the drum remains stationary until a DC current is applied to the coil. Relative motion between magnet assembly and drum produces eddy currents in the drum, whose magnetic field attracts the magnet assembly. Attraction between the two magnetic fields causes the drum to follow the rotation of the magnet assembly. The attraction between the two rotating members is determined by the strength of the coil's magnetic field and by the difference in speed between the two members. Thus, by controlling the coil excitation, the amount of slip and, hence, the output speed can be controlled.

Slip and Slip Loss Being a slip device, the eddy current coupling must of necessity develop slip and reject the slip power in the form of heat. The amount of slip loss can be determined from Equation 7.10(3):

$$P_s = P_L \frac{S_s}{S_0} \qquad \textbf{7.10(3)}$$

where
P_s = slip loss power
P_L = load power
S_s = slip speed (rpm)
S_0 = output speed (rpm)

Slip devices always generate heat, so eddy current couplings are either air or water cooled. Small air-cooled units below 5 hp (4 kW) can be designed to dissipate all of the rated power. Air-cooled units above 300 hp (225 kW) in size can dissipate only about 25% of the rated power capacity. For this reason, air-cooled units are not recommended where cooling capacity would be greater than about 20% of rated power.

Water-cooled units are designed to dissipate the rated horsepower continuously. In addition, water-cooled units show a small efficiency advantage over air-cooled ones, particularly on "water-in-the-gap" types, where the water contributes slightly to the torque capability. Generally, water-cooled units are preferred to air-cooled types except when lack of coolant precludes their use or where very low slip losses are encountered.

Eddy current couplings require only a low percentage of transmitted power for excitation. Typically, a 3 hp (2.2 kW) unit will require 50 W of excitation, while a 12,000 hp (9 MW) unit will require 20 kW. Eddy current couplings are readily adaptable to silicon-controlled rectifier (SCR) (Section 7.2), providing speed control within 1% accuracy, over 10–100% load change.

Size and Efficiency Efficiency of the eddy current coupling is acceptable in the large sizes and at full speed torque, as will be shown later. At lower speeds, efficiency drops considerably, while efficiencies as high as 95% are possible at full excitation and torque. Because there is no contact between the input and output shafts of the coupling, the unit will not transmit vibrations. On prime movers that exhibit some torsional vibration, the use of an eddy current coupling can be of advantage because it will attenuate these vibrations.

Integral combinations of motor, coupling, and excitation are available in small sizes of 1 hp (0.75 kW) or less. Air-cooled couplings range in size up to 900 hp (670 kW), but the larger sizes are practical only where a relatively low cooling capacity is required. Liquid-cooled units can be as large as 18,000 hp (13.5 MW) capacity. Speed-control units range from simple open-loop control to precise, automatic closed-loop control with tachometer-generator speed feedback.

Magnetic Particle Coupling The magnetic particle coupling offers another solution to adjustable speed drives. Basically the coupling consists of two concentric cylinders separated by an air gap and a stationary excitation coil surrounding the cylinders. Ferromagnetic particles fill the gap between the concentric cylinders.

When a controlled amount of current is used to energize the coil, the particles form chains along the magnetic lines of flux connecting the cylinder surfaces. The shear resistance of the magnetic particles is proportional to the coil excitation and provides the basis for power transmission from input to output.

The output torque of this unit is always equal to input torque, regardless of speed. This fact allows the output torque to be set at standstill by controlling the coil excitation. Whenever the torque capacity of the coupling is exceeded, slip will

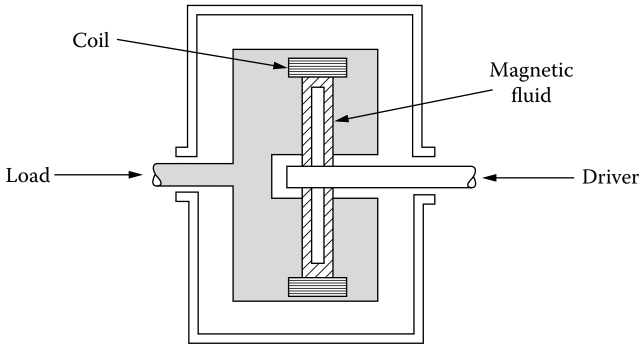


FIG. 7.10i
Disc-type magnetic fluid clutch.

result with accompanying heat liberation. Whenever the coupling is used for speed control, selection of a coupling with adequate cooling capacity is of great importance. Air-cooled units are available with large heat dissipation capacity; however, water cooling is more effective and therefore preferred.

Magnetic particle couplings can be used effectively for speed control, because they can provide a constant torque output independent of speed. However, a closed-looped control system is required for effective speed control in order to cancel the effect of changing torque on slip speed.

Magnetic Fluid Clutches The magnetic fluid clutch is similar in operating principle to the magnetic particle coupling, but either a magnetic dry powder or magnetic powder suspended in a lubricant is utilized. A typical disc-type magnetic fluid clutch is shown in Figure 7.10i.

While the operating principle is similar to that of magnetic particle coupling, the clutch can be used only for low-power applications in the order of a few horsepower. Recharging of the clutch due to fluid deterioration is a definite disadvantage; however, recharging can be fairly easily accomplished during routine maintenance.

VARIABLE VOLTAGE

One of the least desirable drives for centrifugal pumps is the variable voltage drive utilizing NEMA D, high-slip, induction motors. This is because of their poor efficiency, limited speed turndown, and need for special motors. This drive is no longer in common use.

POLE-CHANGING AC MOTORS

AC motors, while not readily usable for continuous speed control, can be made reversing or furnished with several fixed speeds. The squirrel cage induction motor is best suited from the standpoint of reliability, and it can be furnished with up to four fixed motor speeds.

Speed changes are accomplished through motor starter contacts that reconnect the windings to yield a different number of poles. In the resulting pole motor, the sections of the stator winding are interconnected through the motor starter to provide different motor speeds. On separate winding motors, a different winding with the required number of poles is energized for each motor speed. The starter contactors, however, must be interlocked to prevent two or more contactors from closing simultaneously. In current practice, more than two-speed designs are unusual, and variable-frequency drives have become common.

SOLID-STATE VARIABLE-SPEED DRIVES

In the past three decades, the use of solid-state variable-speed drives has increased dramatically. There are several reasons for this trend. The energy shortages of the 70s were a powerful incentive to find new ways to save energy. Because a third of all the electric power generated in the United States is consumed by AC motors, a major focus was on means of cutting the power consumed by those motors.

The main use of the AC motor is as a prime mover for pumps and fans. The traditional practice of flow control was to use a throttling valve or damper. These methods are notoriously wasteful (Figure 7.10a). Efficient flow control can be achieved by changing the speed of the pump or fan and thereby eliminating the valve entirely, but with an AC motor that could only be achieved (while maintaining efficiency) by two-speed motors, rotary frequency changers, or inverter motor drives.

The increase in the cost of electric power brought the payback period down to the point where the use of AC drives became cost-effective. During this same period, technological advances improved the performance, increased the reliability and efficiency, and lowered the cost of AC and DC solid-state motor drives. The main developments that caused this rapid advance were the improvements in semiconductor switching devices, large-scale integrated circuits, and the microprocessor.

The engineers designing the drives immediately adopted these advances and incorporated them into their products. As a result, the price of small AC variable-speed drives (below 20 hp/15 kW) has dropped by about a factor of three. The combined effect of better and cheaper drives and the need to save energy is responsible for rapid growth of the solid-state drive industry.

Underlying Semiconductor Technology

All solid-state variable-speed drives are based on semiconductor switching devices. The important ones are the rectifier diode, the power transistor, the thyristor, the gate-turn-off thyristor (GTO), and the insulated gate bipolar transistor (IGBT). All except the diode have the property of being

TABLE 7.10j
Comparison of Semiconductor Switches

Device Type	Blocking Voltage	Current Capability	Switching Time
Thyristor	6000 V	5000 A	50–300 μsec
GTO thyristor	4500 V	1000 A	10–50 μsec
Transistor	1400 V	500 A	3–10 μsec
IGBT	1400 V	400 A	100–300 nsec

controlled by a low power signal that changes their state from blocking to conduction.

Thyristors can be turned on by gate control but not turned off at the gate. The GTO, transistor, and IGBT have both turn-on and turn-off capability. The ability of these devices to handle higher currents and voltages has improved steadily, with the result that drives could be built for ever-larger power ratings.

Table 7.10j illustrates the primary switching properties of devices available in production quantities in 1992. Today, drives are available that operate directly at 7200 V, although the most common drive voltages in the United States are 230 and 460 V, three-phase.

Although the thyristor was the first semiconductor switch used in drives, it has been supplanted in the power range below 500 hp (370 kW) by power transistor- and IGBT-based drives. The GTO has been used at all drive power levels, but it too is being displaced by the transistor and IGBT. GTO-based drives are primarily limited to drives above 1 MW and those in which space and weight is at a premium, such as propulsion system drives. Due to the voltage limitations of the semiconductor switches, transistor drives are available at 600 V and below.

Drive Circuit Topologies

All drives have an input stage that converts the fixed voltage and constant-frequency AC mains to DC. If the input devices are controllable, like thyristors, then the DC output of this stage is adjustable. If the input devices are diodes, which conduct whenever forward biased, then the output DC is a fixed proportion of the input AC voltage. After the input stage, the details of the circuit depend on the type of motor being controlled.

The simplest power circuit applies to DC motors, because their speed may be controlled by adjusting the DC armature voltage. AC drives are more complex, because the output voltage must be adjustable-frequency and adjustable-voltage AC. This necessitates another stage of power conversion to invert the DC from the input stage into AC for the motor. In AC drives the DC power from the input stage is filtered by a bank of electrolytic capacitors or an inductor.

Variable-Speed DC Motors

Direct current motors lend themselves extremely well to speed control, as demonstrated by the variety of speed-load and load-voltage characteristics obtainable by means of parallel, series, or separate excitation. A typical set of speed-load characteristics for shunt, series, and compound motors is shown in Figure 7.10k.

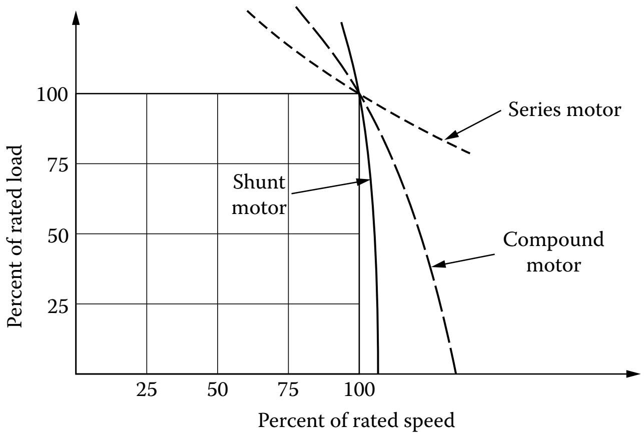


FIG. 7.10k
Speed-load characteristics of DC motors.

The inherent speed regulation, or constancy of speed under varying load conditions, of the shunt motor is shown graphically in the diagram. The speed change of the motor is only 5% for a load change from zero to full load. Although the methods of speed control are applicable to all types of DC motors, the superior speed regulation of the shunt motor accounts for its wider use on control applications.

The speed of a DC motor is a function of armature voltage, current, and resistance; the physical construction of the motor; and the magnetic flux produced by the field winding. The equation relating these variables to speed is:

$$S = \frac{V - IR}{K\Phi} \qquad \textbf{7.10(4)}$$

where
S = motor speed
V = armature terminal voltage
I = armature current
R = armature resistance
Φ = magnetic field flux
K = a constant for each motor depending on physical design

Three methods of speed control are suggested by Equation 7.10(4): adjustment of field flux Φ, adjustment of armature voltage V, and adjustment of armature resistance R.

Field Flux Adjustment The first of these, adjustment of field flux, involves varying the field current. This can be accomplished either by means of a field rheostat or by the current being controlled electronically by a set-point signal through an energy-throttling device such as an SCR. Open-loop control

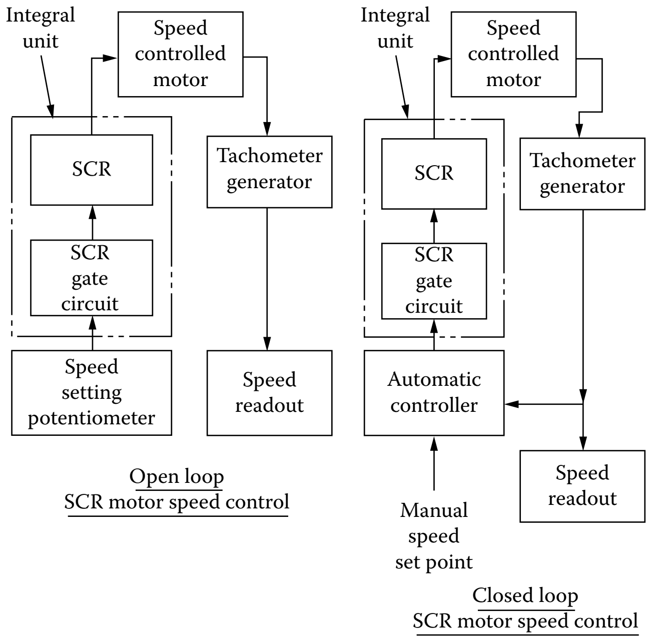


FIG. 7.10l
SCR-based motor speed controls can either be open or closed loop.

should be utilized only when the load is constant and precise speed control is not essential.

Closed-loop control is generally used in conjunction with electronic devices and offers better control than open loop, but at higher cost (Figure 7.10l). Either type of control requires a measurement of the controlled speed; in closed-loop control the measured speed is compared with the set point, and the field current is adjusted automatically to bring the difference to zero.

For a discussion of speed sensors refer to Section 7.19 of the *Process Measurement* volume of this handbook. In open-loop control a speed readout is required to permit set adjustments by the operator for the desired speed.

Adjustment of field flux yields a motor with constant horsepower. The allowable armature current is approximately limited to the motor rating in order to prevent overheating. The effects of changing flux and changing speed effectively cancel each other so that the allowable output horsepower—the product of armature current and induced armature voltage—is constant. Torque, however, varies directly with field flux, and therefore this type of speed control is suited for applications involving increased torque at reduced speeds. The speed range possible with field flux adjustment is approximately 4:1.

Armature Voltage Adjustment Armature voltage control can also be used to control motor speed. With armature voltage control, the change in speed from zero to full load is almost entirely due to the full-load armature resistance drop, and this speed change is independent of the no-load speed. Consider two motors with identical speed changes from zero to fully loaded, but one operating at 100 rpm and the other at 1000 rpm at zero load. In terms of percent speed change, a 10 rpm variation may be unacceptable for the lower-speed motor but may be insignificant for the higher-speed motor.

One method used for armature voltage control is the use of a motor-generator set to supply a controlled voltage to the motor whose speed is to be regulated. An AC motor drives a DC generator whose output voltage is variable and supplies the variable-speed motor armature. An obvious disadvantage of this method is the initial investment in three full-size machines; this method is so versatile, however, that it is often used.

Armature voltage, of course, can also be controlled electronically by one of the devices discussed in Section 7.2. These are more commonly used for speed control due to the lower investment and greater efficiency as compared to the motor-generator set. In the motor with controlled armature voltage, both the allowable armature current and field flux remain constant. The driver therefore has a constant torque output, as opposed to the constant horsepower output of the field-controlled motor.

The armature voltage method of speed control yields speed ranges in the order of 10:1. By combining field flux control and armature voltage control, speed ranges of 40:1 are obtainable. The base speed of the motor is set at full armature voltage and full field flux; speeds above base are obtained by field flux control, speeds below base by armature voltage control. Speed ranges greater than 40:1 are obtainable through the addition of special motor windings and SCR controls.

Armature Resistance Adjustment Adjustment of armature resistance is another method of speed control suggested by Equation 7.10(4), and it can be used to obtain reduced speeds. An external, variable resistance is inserted into the armature circuit (Figure 7.10m). Speed regulation with this method and its variants is very poor, however, and this type of speed control is not commonly used. An added disadvantage of the armature resistance method is the decrease in efficiency due to the power consumption in the resistor.

Starting Circuits DC motors must be protected from high in-rush currents during starting. In Figure 7.10n, the starting current through the armature is limited by resistors $R1$, $R2$,

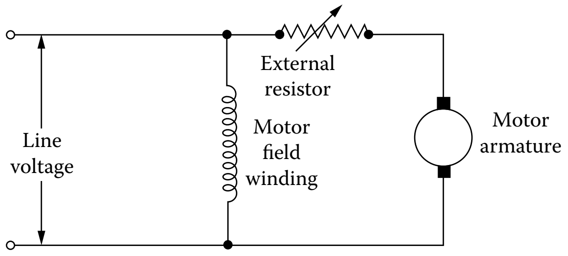


FIG. 7.10m
Shunt motor is controlled by varying the armature resistance.

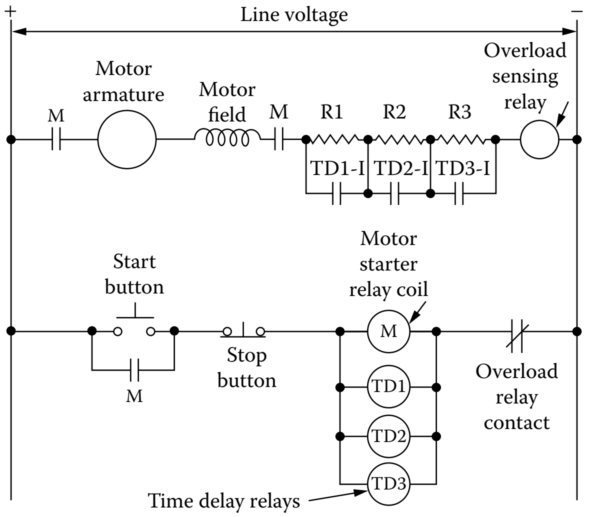


FIG. 7.10n
The starting circuit of a nonreversing DC motor.

and *R*3. When the start button is depressed momentarily, relay M and timer delay relays TD1, TD2, and TD3 are energized. The two M contacts close instantaneously, providing power to the motor. After a time delay, contact TD1-1 closes, shunting resistor *R*1. Contacts TD2-1 and TD3-1 close successively until the armature is directly on the line. The number of resistors is set by the torque and current limitations of the motor and the desired smoothness of startup.

Braking, Control, and Reversing The motor will also act as a generator, converting mechanical energy to electrical energy. This feature can be utilized to advantage where rapid stopping of the motor is required. In this configuration, when power is disconnected from the motor, a resistor is automatically connected across the armature and the mechanical energy of the rotating member is dissipated as heat in the resistor.

DC motors can be run in reverse rotation by changing the polarity of the armature voltage. This can be accomplished by means of two contactors—one forward, one reverse. These must be either mechanically or electrically interlocked to inhibit closing both contactors simultaneously and to prevent applying reverse line voltage to the motor prematurely.

Normally, DC power for operation of the motors is not readily available. Rectifier tubes and solid-state devices such as SCRs are used extensively to convert the incoming power to DC and simultaneously to throttle the current or voltage delivered to the motor in response to an external control signal. A discussion of these devices is given in Section 7.2.

Pump Drives Direct current motors have become more popular for pump drives because of the development of silicon-controlled rectifiers. The SCRs and direct current motor have resulted in a variable-speed drive of modest first cost and high efficiency. The significant objection to direct current motors has been maintenance of commutators and brushes. Direct current motors have been used on elevators in public buildings for many years.

Unlike centrifugal pumps, elevator loads are of the constant torque type. Brush life varies from 6 weeks to 18 months on elevator applications. At present, no extensive experience is available on brush and commutator wear and maintenance on centrifugal pump applications in the 1–100 hp (75 kW) range. If trained maintenance personnel are available at the point of application, the direct current drive can provide a highly efficient, variable-speed drive for centrifugal pumps. Cost of brushes and brush maintenance must be included in any economic evaluation. At the present, direct current motors are not as readily available as AC induction motors.

Thyristor DC Drives

The thyristor DC drive is the oldest form of solid-state drive and is still a viable choice, although the drawbacks of the DC motor, including high cost, size, and sparking, have not been overcome. These drives can be built over a wide power range, but due to difficulties in commutation in the machine they are usually limited to less than 700 VDC output. Thyristor DC drives are available off the shelf with or without the motor from fractional to 500 hp (375 kW). Fractional horsepower drives typically are powered from 240 VAC single phase, while the range above 5 hp (3.7 kW) is furnished for 240 or 480 VAC three-phase.

The behavior of the shunt DC motor makes it relatively easy to control. Torque is proportional to the product of current and flux, while voltage is proportional to the product of speed and flux. Figure 7.10o is a block diagram of a single converter thyristor DC drive suitable for two-quadrant operation. The thyristor converter is equipped with a fast current regulator, which effectively controls the motor torque. Enclosing the current loop is a speed regulator that amplifies the speed error and generates the current reference for the inner current loop. Speed feedback is provided by a tachometer for precision (0.1–1% accuracy and 100:1 turndown) or

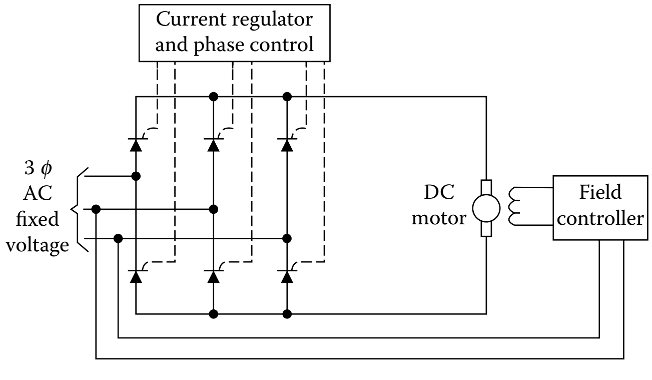


FIG. 7.10o
Single converter thyristor DC drive.

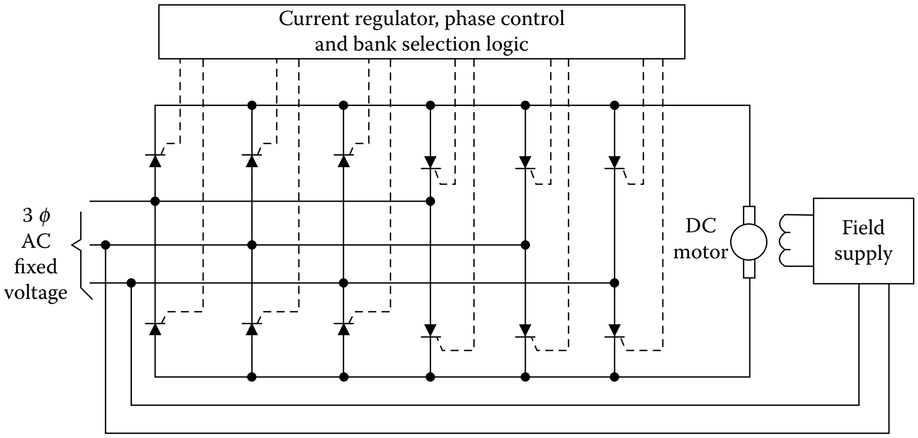


FIG. 7.10p
Dual converter thyristor DC drive.

by the motor terminal voltage for less demanding applications (2–5% accuracy and 10:1 turndown).

Because the converter can produce current in one direction only, a fast-reversing drive requires a second converter in inverse parallel to the first (see Figure 7.10p). Only one converter at a time is in conduction, depending on the desired direction of the motor torque. Although it is also possible to control direction with reversing contactors in the armature circuit, nowadays that feature is done with dual converters or field reversing.

Field-reversing drives use the property of the machine that the torque direction is controlled jointly by armature current direction and flux polarity. For applications in which the fastest reversing is not necessary, a small dual converter on the field is an economical approach to reversing drive applications (see Figure 7.10q).

Approximate costs of thyristor DC drives are listed in Table 7.10r. These costs include only the power conversion equipment, not the motor. Although the cost of the DC drive electronics is less than that of the AC variable-frequency drive, when the much lower cost of the AC motor is included, an AC drive and motor package usually costs less than the DC package of the same rating. This is almost always true if a weather-protected motor or explosionproof motor is required.

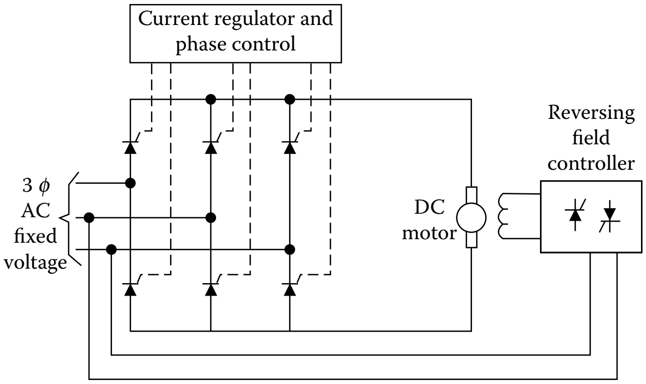


FIG. 7.10q
Three-phase field-reversing thyristor DC drive.

Variable-Frequency Drives

Often called *adjustable frequency,* these drives convert alternating current to direct current and back to alternating current at frequencies from 0 to 120 Hz. Of solid-state construction, these drives provide a highly reliable means of varying pump speed. They have wire-to-shaft efficiencies as high as 95% at full speed and 70–75% efficiencies at 40% speed when driving new, high-efficiency motors. The motors are standard induction types, found in stock in most major cities. (Load commutated inverter, or LCI, is applied to synchronous motors.)

The primary objections to these drives were their high first cost and complex designs. The development of power transistors and other electronic advances has reduced the cost of these drives, and the ongoing training of field personnel in electrical and electronic service has made variable frequency more acceptable in fields that have, traditionally, used mechanical, variable-speed drives. Variable-frequency drives are available up to 5000 hp with broad speed turndown ranges.

TABLE 7.10r
2004 Cost of Thyristor DC Drives

Power Rating	*Single Converter*	*Dual Converter*
100 hp (75 kW)	$5,000	$6,500
200 hp (150 kW)	$7,000	$8,000
500 hp (375 kW)	$15,000	$18,000
1000 hp (750 kW)	$20,000	$27,000

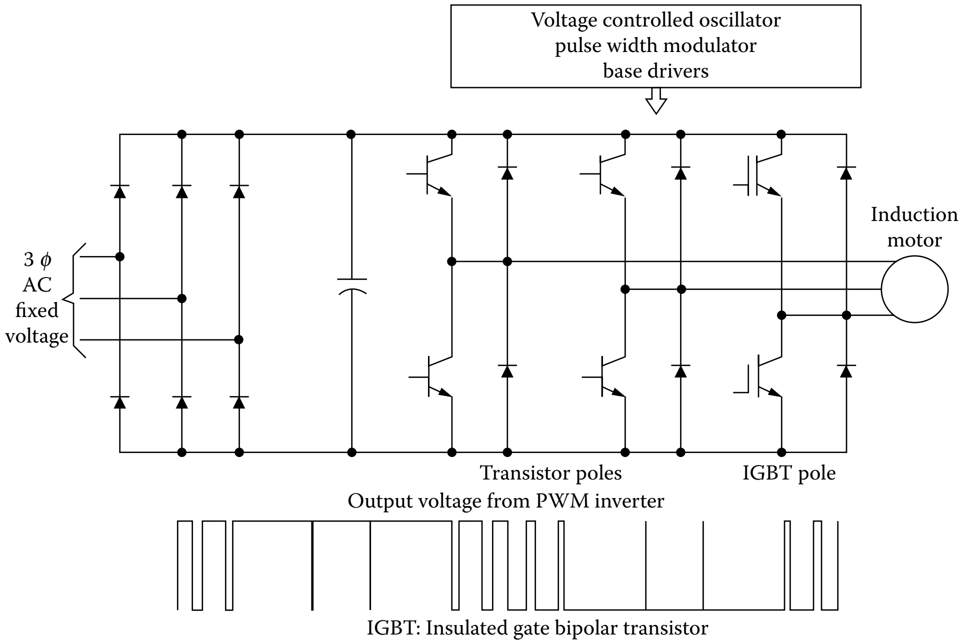


FIG. 7.10s
Pulse-width modulated inverter drive using transistors.

For the AC variable-frequency drive, the power electronics is more complex than for the DC drive, but the advantages of the AC motor over the DC are so pronounced that AC drives are now more economical in most applications and will ultimately supplant DC drives. More sophisticated approaches such as the use of field-oriented control have permitted AC drives to match the performance of DC drives.

There are two basic approaches to AC drives, which are distinguished on the basis of how the DC power from the input stage is handled. If the DC link filter is a capacitor and the inverter output consists of pulses of voltage from the link, then it is a voltage-fed drive. This category includes all drives with diode bridge inputs, such as pulse-width modulated (PWM) drives.

On the other hand, if the link filter is an inductor and the inverter output consists of pulses of current, then it is a current-fed drive. In this group we find the conventional autosequentially commutated current-fed inverter (ASCI), the load commutated inverter, and almost all other induction motor drives that operate at 2.3 kV and above. It is worth noting that different types of drives generally use different switching devices, as the different properties lend themselves to certain types of circuits.

Transistors and IGBTs are used in PWM drives; thyristors are used in DC drives, ASCIs, and LCIs. GTOs can be used in any of the drive circuits. These four types of drives (DC, PWM, ASCIs, and LCIs) dominate the solid-state variable-speed drive market. There is no one “best” drive; all of them have a different mix of features such that one needs to define the application requirements carefully in order to choose the best drive for the particular service.

Pulse-Width Modulated Drives Figure 7.10s illustrates the power circuit of a PWM drive using transistors. It is a very simple circuit, with an input diode bridge, fixed voltage DC link, and six output switches. Control of the output voltage and frequency is exercised entirely by the modulation strategy for the transistors. As the output voltage waveform indicates, the DC link voltage is applied to the motor in short pulses by turning the transistors on and off.

The duration of the pulses and their spacing controls the harmonic spectrum of the output. The objective is to have a large fundamental and as small an amount of harmonics as possible. The inverter determines the amplitude and frequency of the motor voltage, but the motor current is determined by the motor parameters and the load on the motor. Modern PWM drives switch rapidly (greater than 1 kHz for transistors; greater than 10 kHz for IGBTs), so that the low-order harmonics are essentially eliminated and a smooth sinusoidal motor current is obtained.

The best modulation strategies require very complex combinatorial logic or fast real-time computer processing, features that are now readily available in PWM drives but were not when they first appeared on the market. Transistor PWM drives are available in fractional through about 500 hp (400 kW) at rated input voltages of 120, 240, 480, and 600 VAC; GTO PWM drives are available in even higher voltage and power ratings.

Current Source Variable-Frequency Drives The basic power circuit of the current source inverter is shown in Figure 7.10t. The input conversion is performed by a thyristor bridge. This is controlled with a current regulator to function as an adjustable

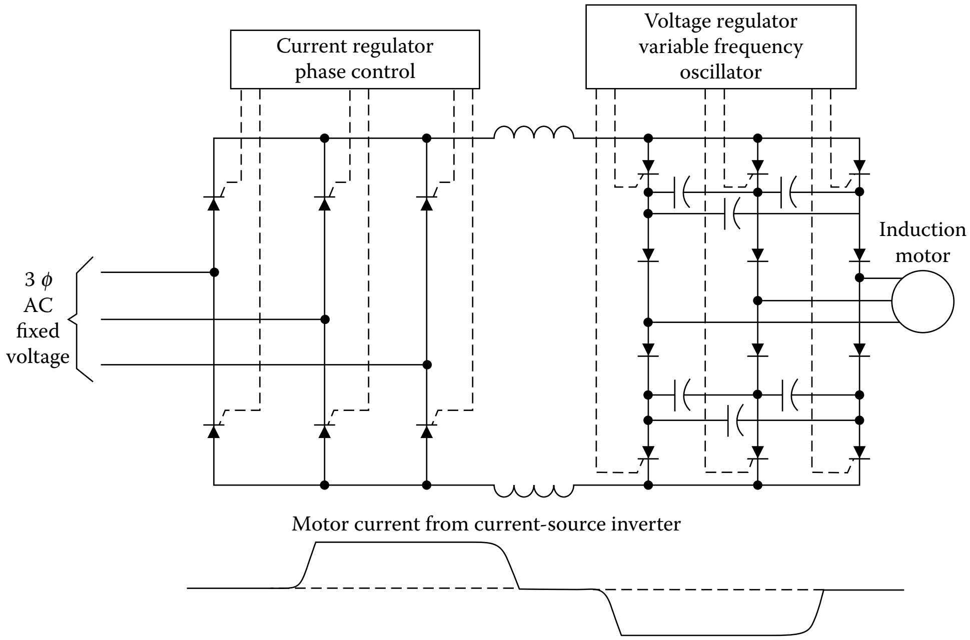


FIG. 7.10t
Autosequentially commutated current-fed inverter (ASCI).

current source. The voltage ripple of the converter is filtered by the choke in the DC link so that a smooth current flows into the output inverter stage. This current is steered in sequence into the motor windings by the thyristors so that the motor receives a trapezodial current waveform.

Because the thyristors cannot be turned off by gate control, commutation is forced by the energy stored in the capacitors. This energy is exchanged between the machine inductance and the capacitors. It is prevented from discharging into the machine by the diodes in series with the thyristors. In this case the inverter establishes the motor current, but the motor voltage is determined by its parameters and load.

Unlike the PWM inverter, the output amplitude is controlled by the current regulator on the input converter, and the output stage controls only the frequency. Because this type of drive is constructed from thyristors, it can be economically built in very large power ratings. A cost comparison of ASCI and PWM variable-frequency drives is shown in Table 7.10u. Once again, cost of the machine is not included.

Load Commutated Inverters Yet another common type of drive is the load commutated inverter, whose power circuit is shown in Figure 7.10v. This is also a current-fed topology, as evidenced by the similarity to Figure 7.10t. The input converter is a thyristor bridge equipped to function as a controlled current source. Smooth DC current flows through the choke into the thyristor inverter, where it is switched into the motor windings. But in this case commutation is performed by the machine.

By operating the machine at a leading power factor, the thyristors are naturally commutated because an off-going device experiences reverse voltage from the machine. However, a synchronous machine is required because an induction motor cannot be operated at a leading power factor. Although it is possible to build LCIs over a wide power range—100–20,000 hp (75 kW–15MW)—they are generally used in applications above 1000 hp (750 kW), where the cost differential between a synchronous and induction motor narrows. The cost of a medium voltage LCI (excluding motor and exciter) ranges from $150 per hp ($200 per kW) down to $75 per hp ($100 per kW), as the rating increases from 1000 to 5000 hp (750 kW to 3.7 MW).

Drives with Power Regeneration The ability to receive mechanical energy from the load and return it to the AC line is inherent in the ASCI and LCI, and the DC dual converter. The

TABLE 7.10u
2004 Cost of Variable-Frequency Induction Motor Drives

Power Rating	*PWM Including Reactor*	*Current-Fed ASCI*
10 hp (7.5 kW)	$1,700	
20 hp (15 kW)	$2,500	
50 hp (37 kW)	$5,300	
100 hp (75 kW)	$7,500	$8,000
200 hp (150 kW)	$11,000	$12,000
500 hp (375 kW)		$18,000
1000 hp (750 kW)		$32,000

Note: The reactor price adds 50–90% to the base price of the variable-frequency drive electronics.

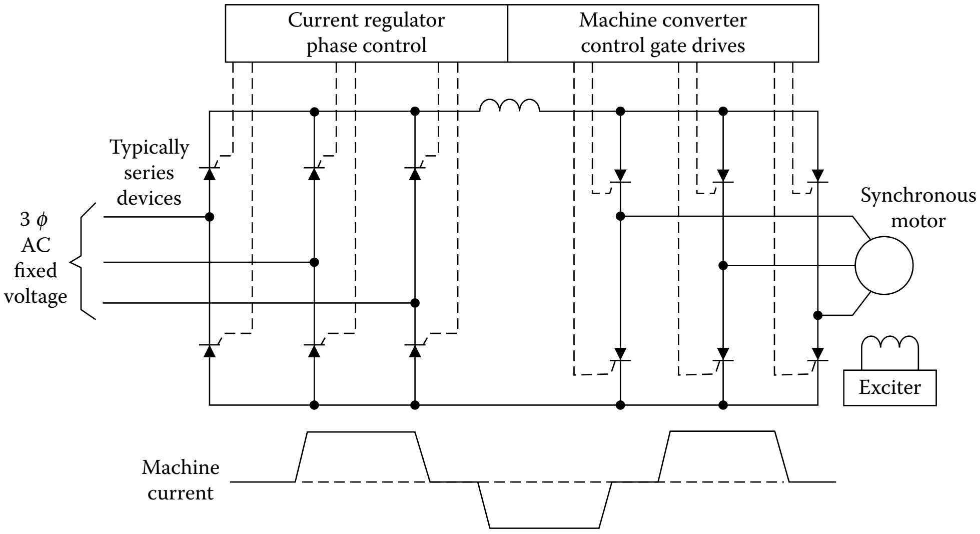


FIG. 7.10v
Load commutated inverter (LCI).

PWM drive requires an additional input stage to achieve this property. This capability is essential for a four-quadrant drive.

The wound rotor motor has been used for many years to vary the shaft speeds of cranes, machine tools, and similar heavy equipment. The energy loss due to heating of secondary resistance in these earlier applications has been recovered through the use of rectifiers and inverters and has resulted in efficient drives for centrifugal pumps. The higher cost of the wound rotor motor and drive makes this design cost-effective only for larger pumps in the range of 500–10,000 hp (375 kW–7.5MW).

Comparison of Solid-State Drives

The specific requirements of the application determine which type of drive is best; they all have somewhat different properties. It is difficult to make quantitative comparisons because drives of the same type may behave differently due to detail design differences such as size of suppressors and rating of devices. Some general comments follow.

Input Power Factor The diode bridge input configurations like PWM drives usually have near-unity displacement power factor, regardless of speed and load. However, the total power factor can be much lower due to input harmonic currents if there is inadequate inductance in the DC link or on the line side of the bridge. Drives with an input thyristor converter like DC drives, ASCIs, and LCIs have an input displacement power factor proportional to PU speed (for DC drives) or PU speed times motor power factor in the case of current-fed AC drives.

Physical Size and Weight The DC drive is the most compact design as it has fewest components; PWM drives are slightly larger; and the current-fed units are the largest and heaviest due to the magnetic components.

Immunity to Shorts and Ground Faults The current-fed designs (ASCI and LCI) have inherent immunity to load shorts and ground faults due to their current regulators and high-impedance link chokes. PWM drives rely on detecting such faults and quickly turning off the transistors. DC drives are frequently equipped with transformers to isolate the drive from a ground-referenced AC source.

Input Harmonic Currents All solid-state motor drives draw harmonic currents from the source. The harmonic spectrum of all the six-pulse drives that were mentioned results in at least 30% total harmonic current. The LCI, ASCI, and DC drives that have a highly inductive DC link draw odd harmonics excluding those divisible by three. The magnitude of the currents is approximately 1/h times fundamental where h is the harmonic number.

The same harmonics are present in the input currents of a diode bridge, but it is harder to quantify the current distortion because that depends on the values of the AC line reactance and link choke, if present. For low inductance PWM designs, the total harmonic current can easily exceed 50%.

Motor Effects All VFDs have a nonsinusoidal output. This results in a higher RMS motor current for the same motor output as compared to line (sinusoidal) operation. Consequently, more motor heating occurs, necessitating careful motor selection or de-rating.

The extra current varies from 3% for LCI and ASCI drives down to well under 1% for PWM drives. However, the extremely high amplitude and high frequency modulation of PWM drives raise eddy current losses and rotor losses in the machine. A rule of thumb is that the total motor losses increase by 5–15% (compared to line operation) when used with a drive. In the case of DC drives, the output voltage is not smooth but is rich in harmonics. However, the armature inductance in modern DC motors filters this so that the current is acceptably smooth and no motor de-rating is necessary.

Efficiency of Solid-State Drives

All the drives mentioned here have efficiencies at full speed and full load of 95% or more. Drives rated at higher power have slightly higher efficiencies than do lower power units because the fixed-power-consumption components, such as control circuits and cooling fans, constitute a larger proportion of the smaller power rating. This is also true for higher voltage ratings, because the conduction losses of the semiconductors arise from a nearly constant forward voltage drop.

The focus of efficiency should be on the energy savings due to the use of the drive (as compared to throttling devices) rather than small differences between types of drives. Furthermore, there are no industry standards on efficiency measurement or calculation for drives, which means that precision comparisons between manufacturers' claims are not very meaningful. To some extent there is a trade-off between efficiency and protective circuits, as components like suppressors and line reactors degrade efficiency but improve the reliability of the drive.

Recent VFD Developments

Over the past 40 years, since the VFD was first developed, the weight of the drive electronics has dropped rapidly. The first Danfoss VLT5 2 hp (1.5 kW) drive weighed 128 lb (58 kg). The current equivalent VLT5000 weighs just 17.6 lb (8 kg), a sevenfold reduction.

Current-generation equipment offers serial communication capability (ModbusPlus, LONWorks, Profibus, Device-Net, etc.) and is frequently provided with PID control capability with 4–20 mA I/O, providing access to a variety of internal parameters.

When the drive motors are fitted with an encoder, position feedback allows a whole series of individual drives to be synchronized, matched in angular position, or adjusted as though they had a mechanical gearbox between them.

For asynchronous motors, $T \sim \phi \times I_L$ normally applies, where (T) is the torque developed, I_L is the rotor current, and ϕ is the air gap flux of the machine. To optimize torque from the motor, the air gap flux of the machine ($\phi \sim V/f$) should be kept constant (Figure 7.10w). This means that if the line frequency (f) is changed, the line voltage (V) must be changed proportionally.

For heavy starts (screw conveyors) and to optimize the stalling torque, an extra (start) voltage (Vo) is required. When loaded and in the low speed range (f < 10 Hz), the voltage loss is clearly seen on the active resistance of the stator winding (particularly in small motors), leading to a specific weakening of the air gap flux (ϕ).

When operating above normal supply frequency (many small four-pole motors are capable of operation up to two-pole speeds), the torque available becomes inversely proportional to speed, and the motor is limited to design power, regardless of speed.

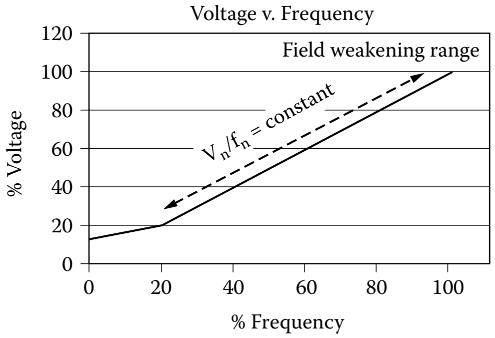


FIG. 7.10w
The motor torque is optimized by keeping the ratio of line voltage to line frequency constant.

EFFICIENCY OF VARIABLE-SPEED DRIVES AND PUMPS

The following three efficiencies should be evaluated for any application of variable-speed pumps:

1. Wire-to-shaft efficiency of the variable-speed drive and the motor. Obviously, this is the ratio of the useful energy applied to the pump shaft divided by the energy applied to the drive-motor combination.
2. Pump efficiency.
3. Wire-to-water efficiency of the drive, motor, and pump. This is the ratio of the useful energy applied to the water divided by the energy applied to the pump-motor-drive combination.

Following is a list of terms that will be used to describe the efficiencies of various pump-motor-drive combinations. All of the energy terms are in horsepower to simplify this discussion; they can be converted into watts (HP = 746 w).

Chp = Energy input to an electrical drive
Ce = Efficiency of an electrical drive
Ihp = Energy input to an electric motor
Me = Efficiency of an electric motor
Dhp = Energy input to a mechanical-type drive
De = Efficiency of a mechanical-type drive
Php = Energy input to a centrifugal pump
Pe = Efficiency of a centrifugal pump
Whp = Water horsepower or useful energy applied to the water
Ws = Wire-to-shaft efficiency (often called line-to-shaft efficiency)
Ww = Wire-to-water efficiency

Figure 7.10x describes the configuration of an electrical-type variable-speed drive and pump.

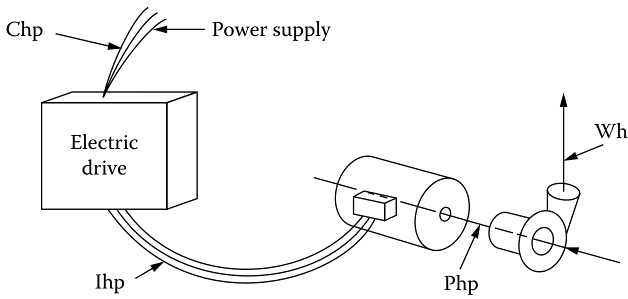


FIG. 7.10x
The configuration of a pump provided with an electrical variable-speed drive.

Following are the equations for this combination. The efficiencies are entered into these equations as fractions.

$$\text{Whp} = \text{water horsepower} = \frac{\text{GPM} \times \text{pump TDH (in ft of water)}}{3960} \quad \textbf{7.10(5)}$$

$$\text{Php} = \text{pump brake horsepower} = \frac{\text{Whp}}{\text{Pe}} \quad \textbf{7.10(6)}$$

$$\text{Ihp} = \text{motor input horsepower} = \frac{\text{Php}}{\text{Me}} \quad \textbf{7.10(7)}$$

$$\text{Chp} = \text{energy input to electrical drive} = \frac{\text{Ihp}}{\text{Ce}} \quad \textbf{7.10(8)}$$

$$\text{Ws} = \text{wire-to-shaft efficiency} = \frac{\text{Php} \times 100\%}{\text{Chp}} = \text{Ce} \times \text{Me} \times 100\% \quad \textbf{7.10(9)}$$

$$\text{WW} = \text{wire-to-water efficiency} = \frac{\text{Whp}}{\text{Chp}} = \text{Ce} \times \text{Me} \times \text{Pe} \times 100\% \quad \textbf{7.10(10)}$$

Figure 7.10y describes the configuration of the electromechanical drive and pump along with the various energy terms that are applicable to this combination.

The following equations are similar but not necessarily equal to those for electrical drives. Equations 7.10(5) and 7.10(6), for water horsepower and pump brake horsepower, are the same.

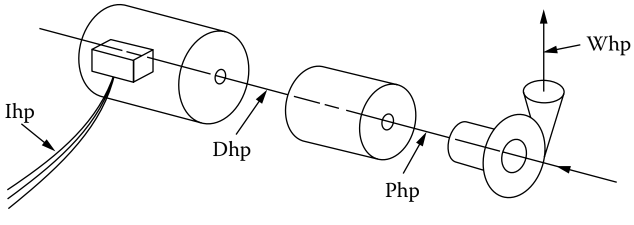


FIG. 7.10y
The configuration of a pump provided with an electomechanrical variable-speed drive.

$$\text{Dhp} = \text{mechanical drive input horsepower} = \frac{\text{Php}}{\text{De}} \text{(same as the electric motor shaft horsepower)} \quad \textbf{7.10(11)}$$

$$\text{Ihp} = \text{motor input horsepower} = \frac{\text{Dhp}}{\text{Me}} \quad \textbf{7.10(12)}$$

$$\text{Ws} = \text{wire-to-shaft efficiency} = \text{Me} \times \text{De} \times 100\% \quad \textbf{7.10(13)}$$

$$\text{Ww} = \text{wire-to-water efficiency} = \text{Me} \times \text{De} \times \text{Pe} \times 100\% \quad \textbf{7.10(14)}$$

Evaluation of VSD Efficiencies

The above equations demonstrate that the wire-to-shaft efficiency for any drive-motor combination is dependent upon the efficiency of the electric motor and the drive, whether the drive is electrical or mechanical. As shown in Equations 7.10(9) and 7.10(13), the efficiency of both the motor and the drive must be determined to achieve the true wire-to-shaft efficiency. These component efficiencies should be determined throughout the anticipated operating speed range, not just at full-speed condition.

The high-efficiency electric motor has created the opportunity for achieving higher wire-to-shaft efficiencies for variable-speed drives, but it is imperative that equivalent motors be used when comparing the efficiency of one type of drive with that of a different type. If high-efficiency motors are used, their efficiencies should have been secured from tests conducted in accordance with NEMA Standard MG1-12.53a, which is based on Institute of Electrical and Electronics Engineers (IEEE) Standard 112, Method B.

Figure 7.10z provides a general comparison of the efficiencies of various types of variable-speed drives for centrifugal pumps. The curves shown in this figure should not be used for energy calculations for a specific application. Rather, efficiencies for drives and motors under consideration should be certified by the manufacturers of that equipment.

As shown in Figure 7.10z, most electric drives are more efficient than electromechanical drives at reduced speeds. The exception is the variable-voltage electric-type drive, which is generally less efficient than mechanical drives. The mechanical drives are usually less efficient than the electric drives, because most of them utilize slip between the input and output shafts of the mechanical drive. This results in a slip loss that is directly in proportion to the amount of slip. The equation for slip loss is

$$\text{slip horsepower loss} = \frac{\text{slip rpm} \times \text{Php}}{\text{output rpm}} \quad \textbf{7.10(15)}$$

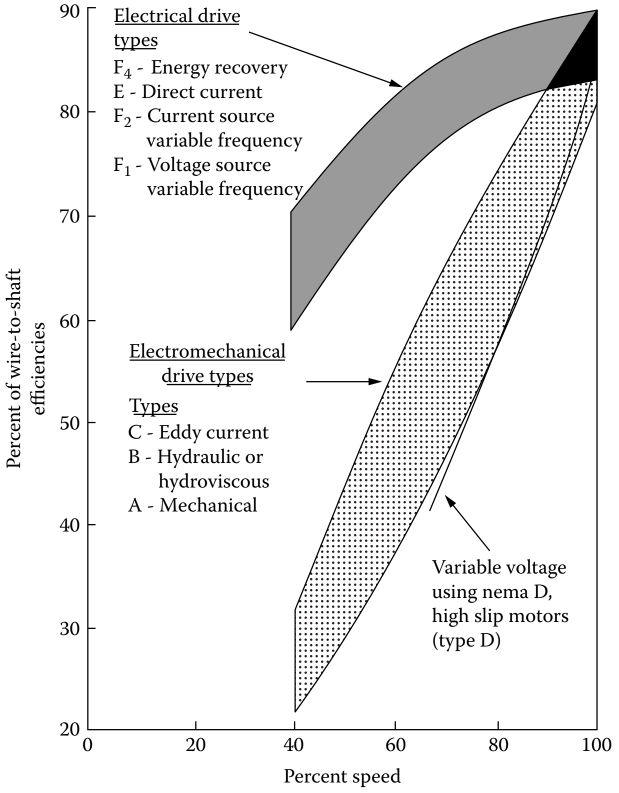


FIG. 7.10z

The wire-to-shaft efficiencies of various variable-speed centrifugal pump drives.

The wire-to-shaft efficiency of the electromechanical drive is

$$
\begin{aligned}
\mathrm{Ws} &= \frac{\mathrm{Php}}{\mathrm{Ihp}} \\
&= \frac{\mathrm{Php} \times \mathrm{Me} \times 100\%}{\mathrm{Php} \times \dfrac{\text{slip rpm}}{\text{output rpm}} + \mathrm{Php} + \text{circulation losses}} \\
&= \frac{\mathrm{Php} \times \mathrm{Me} \times 100\%}{\dfrac{\text{input rpm}}{\text{output rpm}} \times \mathrm{Php} + \text{circulation losses}} \qquad \mathbf{7.10(16)}
\end{aligned}
$$

Circulation losses are those losses incurred by the constant rotation of the input shaft, along with energy for specific uses such as the loss imparted by oil pumps to circulate oil in fluid couplings and hydroviscous drives. Circulation losses must always be stated by mechanical drive manufacturers.

The overall system efficiency index (SEI) of a variable-speed pump installation is determined as follows:

$$\mathrm{SEI} = (\mathrm{Ep} \times \mathrm{Em} \times \mathrm{Ev} \times \mathrm{Eu})10^{-6} \qquad \mathbf{7.10(17)}$$

where

Ep = Pump efficiency
Em = Motor efficiency
Ev = Variable-speed drive efficiency
Eu = Efficiency of utilization = (Qa)(Ha)/(Qd)(Hd)
Ha = Actual head
Hd = Design head
Qa = Actual flow
Qd = Design flow

CONCLUSIONS

The above brief descriptions provide general information on the various types of variable-speed drives. Actual selection of the drive for a specific application requires an evaluation of (1) wire-to-shaft efficiency, (2) first cost, (3) reliability, (4) serviceability, (5) maintenance costs, (6) need for special motor, (7) speed range, and (8) control repeatability.

No variable-speed drive should be selected for a centrifugal pump without careful evaluation of the process load itself to determine the feasibility of variable speed. This, along with consideration for the eight factors listed above, should result in the selection of the optimum drive. For more details on pump applications and pump controls, refer to Chapter 8.

Bibliography

Barnes, M., *Practical Variable Speed Drives and Power Electronics,* Oxford: Newnes, 2003.

Blaschke, F., "The Principle of Field Orientation as Applied to the New Transvektor Closed-Loop Control System for Rotating Field Machines," *Siemens Review,* May 1972.

Bose, B. K. (Ed.), *Adjustable Speed AC Drive Systems,* New York: IEEE Press, 1981.

Bose, B. K., *Power Electronics and AC Drives,* Englewood Cliffs, NJ: Prentice Hall, 1986.

Bower, J. R., "The Economics of Variable-Speed Pumping with Speed-Changing Devices," *Pump World.*

Brumbach, M. E., *Electronic Variable Speed Drives,* Clifton Park, NY: Delmar Learning, 2001.

Doll, T. R., "Making the Proper Choice of Adjustable-Speed Drives," *Chemical Engineering,* August 9, 1982.

Fischer, K. A., and Leigh, D. J., "Using Pumps for Flow Control," *Instruments and Control Systems,* March 1983.

Gabriel, R., Leonhard, W., and Nordby, C., "Field-Oriented Control of a Standard AC Motor Using Microprocessors," *IEEE Transactions on Industry Applications,* March/April 1980.

Gottliebson, M., "Explore the Use of Variable Speed Water Booster Pumps," *Water and Wastes Engineering,* May 1978.

Gray, C. B., *Electrical Machines and Drive Systems,* Essex, England: Longman Scientific and Technical, 1989.

Hall, J., "Motor Drives Keep Pace with Changing Technology," *Instruments and Control Systems,* September 1982.

The Hydraulic Institute, "Pump Life Cycle Costs: A Guide to LLC Analysis," 2005 (www.pumps.org).

The Hydraulic Institute, "Variable Speed Pumping: A Guide to Successful Applications," 2005 (www.pumps.org).

IEEE, *Adjustable Speed Drives: Self-Study Course,* New York: IEEE, 1998.

Irvine, G., and Gibson I. H., "The Use of Variable Frequency Drives as a Final Control Element in the Petroleum Industry," 2000 IEEE Industry Applications Society Annual Meeting, World Conference on Industrial Applications of Electrical Energy, Rome, October 2000.

ISA, Motion Systems & Control, Online Collection of Technical Papers, 2003.

Janki, C., "What's New in Motors and Motor Controls?" *Instruments and Control Systems,* November 1979.

Jarc, D. A., and Robechek, J. D., "Static Induction Motor Drive Capabilities for the Petroleum Industry," IEEE Conference, September 8–10, 1980.

Leonhard, W., *Control of Electrical Drives,* New York: Springer-Verlag, 1985.

Liu, T., "Controlling Pipeline Pumps for Energy Efficiency," *InTech,* June 1979.

Merritt, R., "Energy Saving Devices for AC Motors," *Instruments and Control Systems,* March 1980.

Merritt, R., "What's Happening with Pumps?" *Instruments and Control Systems,* September 1980.

Morgan, R. A., "Large AC Motor Drives Gain Attention," *Instruments and Control Systems,* February 1981.

"Motor Controllers," *Measurements and Control,* February 1986.

Murphy, J. M. D., and Turnbull, F. G., *Power Electronic Control of AC Motors,* Oxford: Pergamon Press, 1988.

Polka, D., *Motors & Drives: A Practical Technology Guide,* ISA, 2003.

Pollack, J. J., "Adjustable Rate Pumping," *Plant Engineering and Maintenance Conference/East,* December 10, 1979.

Rajashekara, K., Kawamura, A., and Matsuse, K., *Sensorless Control of AC Motor Drives: Speed and Position Sensorless Operation,* New York: IEEE, 1996.

Rishel, J. B., "Water System Head Analysis," *Plant Engineering,* October 1977.

Schauder, C. D., Choo, F. H., and Roberts, M. T., "High-Performance Torque Controlled Induction Motor Drives," IEEE-IAS Conference Record, 1982 Annual Meeting, pp. 563–568.

Schroeder, E. G., "Choose Variable Speed Devices for Pump and Fan Efficiency," *InTech,* September 1980.

Sen, P. C., *Thyristor DC Drives,* New York: John Wiley & Sons, 1981.

Solar, D. N., "Applications of Inverters," IEEE Conference, April 14, 1980.

Spitzer, D. W., "Variable Speed Drives: Principles and Applications for Energy Cost Savings," ISA, 2003.

Stefanides, E. J., "Electronic Drives Give Energy-Efficient Flow Control," *Design News,* October 6, 1980.

Stewart, R. J., "Applying Adjustable Speed AC Drives," *Instruments and Control Systems,* September 1982.

Wasko, C. R., "500 Hp 120 Hz Current-Fed Field-Oriented Control Inverter for Fuel Pump Test Stand," IEEE-IAS Conference Record, 1986 Annual Meeting, pp. 314–320.

Control and Optimization of Unit Operations

8

8.1
AERATION AND DO CONTROLS 1484

8.2
AIRHANDLER AND BUILDING CONDITIONING CONTROLS 1507

8.3
BATCH CONTROL DESCRIPTION, TERMINOLOGY, AND STANDARD S88 1528

8.8
CHEMICAL REACTORS: BATCH SEQUENCING 1640

8.9
CHEMICAL REACTORS: BASIC CONTROL STRATEGIES 1664

8.10
CHEMICAL REACTORS: CONTROL AND OPTIMIZATION 1697

8.11
CHEMICAL REACTORS: SIMULATION AND MODELING 1711

8.12
CHILLER CONTROL 1720

8.13
CHILLER OPTIMIZATION 1729

8.14
CLEAN-ROOM CONTROLS AND OPTIMIZATION 1753

8.15
COMPRESSOR CONTROL AND OPTIMIZATION 1763

8.16
COOLING TOWER CONTROL 1794

8.17
COOLING TOWER OPTIMIZATION 1805

8.18
CRYSTALLIZER CONTROLS 1811

8.19
DISTILLATION: BASIC CONTROLS 1820

8.20
DISTILLATION: CALCULATIONS OF RELATIVE GAINS 1855

8.21
DISTILLATION: OPTIMIZATION AND ADVANCED CONTROLS 1866

8.22
DRYER CONTROLS 1904

8.23
EVAPORATOR CONTROLS 1914

8.28
HEADER-SUPPLIED DISTRIBUTION CONTROL 1993

8.29
HEAT EXCHANGER CONTROL AND OPTIMIZATION 2004

8.30
INERT GAS BLANKETING CONTROLS 2025

8.31
ORP CONTROLS 2032

8.32
pH CONTROL 2044

8.33
POWER PLANT CONTROLS: COGENERATION AND COMBINED CYCLE 2057

8.34
PUMP CONTROLS 2084

8.35
PUMP OPTIMIZATION 2110

8.36
Rolling Mill Controls 2116

8.37
SEPARATION CONTROLS, AIR 2123

8.41
WELL-SUPPLIED UNDERGROUND GAS STORAGE CONTROLS 2194

8.1 Aeration and DO Controls

P. KESKAR (2005)

INTRODUCTION

Aeration is mostly used in wastewater treatment facilities to introduce oxygen into the wastewater. Oxygen is necessary for biological treatment of carbonaceous matter (secondary treatment) and for oxidation of ammonia into nitrite and nitrate (nitrification). The transfer of oxygen is accomplished using two basic aeration methods. One method is the use of mechanical aerators (mechanical mixing), and the other method is the delivery of air through a network of pipes and diffusers located at the bottom of treatment basins (diffused aeration).

Aeration control is a very important part of any wastewater plant control system design. The volume of air required depends upon the biomass oxygen demand of the wastewater, which is continuously changing. An inadequate supply of air (oxygen) can inhibit secondary treatment and nitrification processes, causing major process problems. Excessive air supply wastes energy and can dramatically increase plant-operating costs. It is estimated that aeration energy consumption can be 50–90% of the total energy demand of an activated sludge process.

A well-thought-out aeration control scheme could optimize the overall process and significantly reduce energy cost. Aeration control schemes include two basic building blocks, namely the measurement of the process variable, which in this case is dissolved oxygen (DO) concentration in the wastewater, and implementation of control adjustments using control algorithms and final control elements, like control valves on air lines or surface aerator speeds via variable-frequency drives (VFDs). The control strategies can be as simple as manual control of final control elements based on DO concentrations or advanced control algorithms utilizing feedback/feedforward and cascade control.

Success of any control strategy depends upon careful evaluation of all elements of the control loop, including the field instrument measuring the process variable (DO), the algorithm, the final control element, and other related mechanical process equipment utilized in the process (e.g., blowers, diffusers, and surface aerators).

The following paragraphs in this section will discuss the process equipment, field instrumentation, control strategies, and current trends in implementation of aeration and DO control.

MECHANICAL AERATION

Mechanical aeration is accomplished by transferring atmospheric oxygen to the liquid by surface agitation. Mechanical aerators are classified as plate, updraft, downdraft, and brush types.

The plate-type aerator employs a circular plate equipped with radial blades and creates a large amount of turbulence by causing a peripheral hydraulic jump.

The updraft-type aerator employs a surface impeller, which draws liquid upward and violently outward at the surface. The surface turbulence causes oxygen transfer. The updraft-type aerators are popular because of their low cost, high oxygen transfer efficiency, and good mixing. The disadvantages include high maintenance and operational costs due to large number of units, frequent icing in cold weather, uneven DO distribution, and not enough oxygenation for nitrification. The updraft-type aerators are particularly suitable for smaller plants in warmer climates that do not require nitrification.

The downdraft-type aerator employs an impeller in a vertical tube to force liquid from the surface, down through the tube to the bottom of the tank. Air is entrained in the liquid as it is forced down into the tube.

The brush-type aerator rotates around a horizontal shaft equipped with a series of projections. Oxygen transfer is accomplished by surface turbulence. The brush-type aerator is used extensively in oxidation ditch application. Typical advantages are high oxygen transfer efficiency, low capital cost, and low operational cost. However, this type of aerator is not suitable for cold climates (due to icing problems), requires critical level controls, and does not provide enough oxygenation capacity for nitrification.

Process Configurations

Two basic types of process configurations are used for mechanical aeration systems.

Plug Flow Configuration Figure 8.1a shows the plug flow configuration. Rectangular tanks with large length-to-width ratios and containing several aerators approach plug flow conditions. In this configuration the tank essentially operates like several smaller tanks connected in series, and the result is more like plug flow rather than complete mixing.

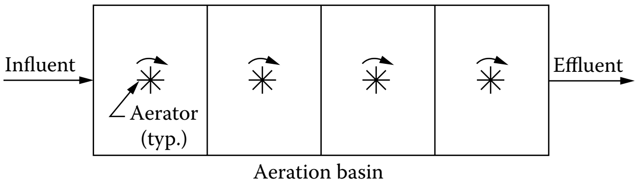


FIG. 8.1a
Plug flow configuration of a mechanical aeration system.

In this configuration control of aeration capacity is accomplished in two ways:

Speed Control of Aerators Simultaneous speed control of all aerators based on DO concentration may not be the best way to control oxygen transfer. Aerators can be classified in several groups, e.g., those near the influent end, those in the middle, and those at the effluent end. Speed of each group of aerators is controlled using a dedicated DO probe to maintain the DO concentration within an acceptable band.

Level-Based Control Adjustment of oxygen transfer capacity can be applied to the entire tank by raising or lowering of the effluent weir to change submergence of the aerators. The weir position is changed to keep the DO levels within acceptable band. The location of the DO probe for weir control is best left up to the operator. DO probe receptacles should be provided at several locations to allow placement of DO probe at the most optimum location following a period of trial and error.

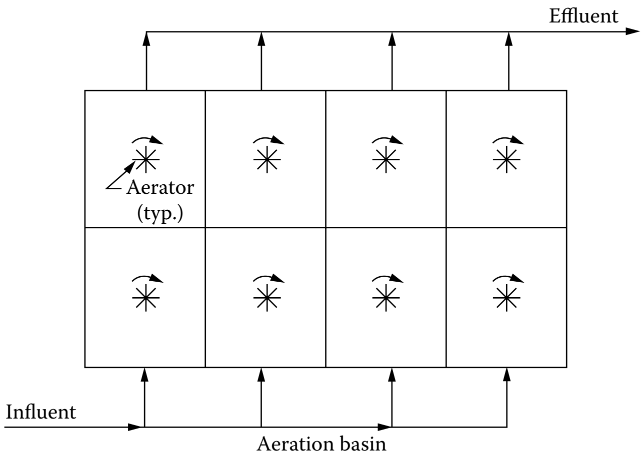


FIG. 8.1b
Completely mixed configuration of a mechanical aeration system.

Completely Mixed Configuration Figure 8.1b shows a completely mixed configuration of aerators. For control purposes, the four aerators near the influent end can be grouped in one group for simultaneous speed control, and the four aerators near the effluent end can be gathered in another group for collective speed control. Other groupings may be considered for achieving better control; i.e., the four aerators at each end can be divided into two groups, making a total of four groups for collective speed control. Dedicated DO probe can be provided for each group for speed control.

Aerator Control Strategies

On/Off Control Figure 8.1c shows a typical on/off aerator control strategy. In this control scheme, multiple single-speed aerators in a zone are controlled based on DO measurement in that zone. The sequence in which the aerators are to be

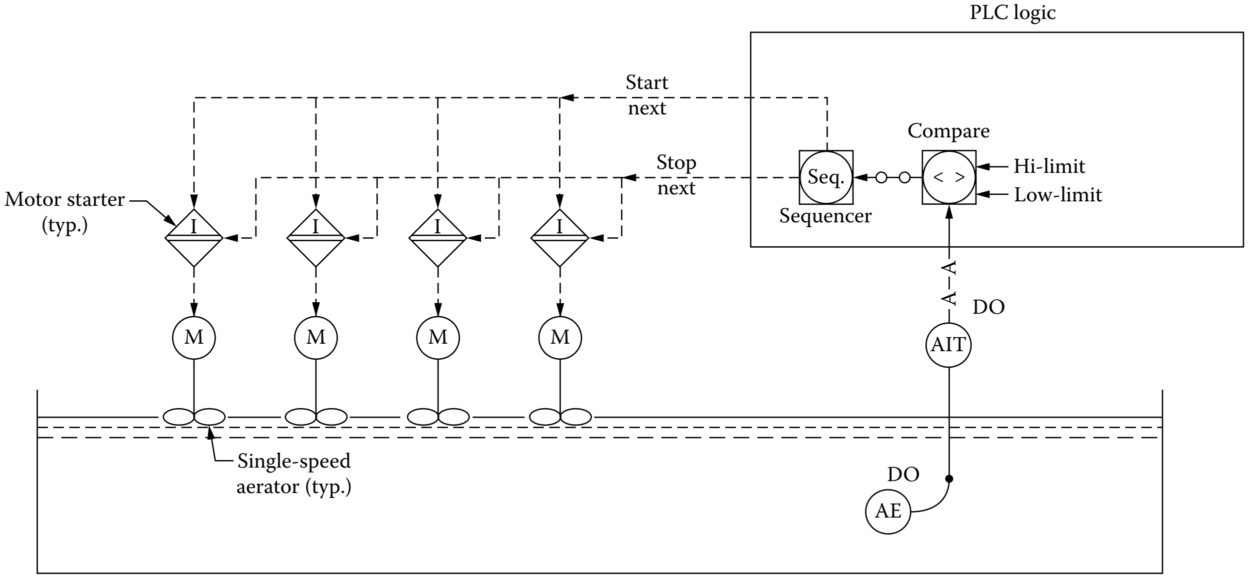


FIG. 8.1c
On/Off aerator control system.

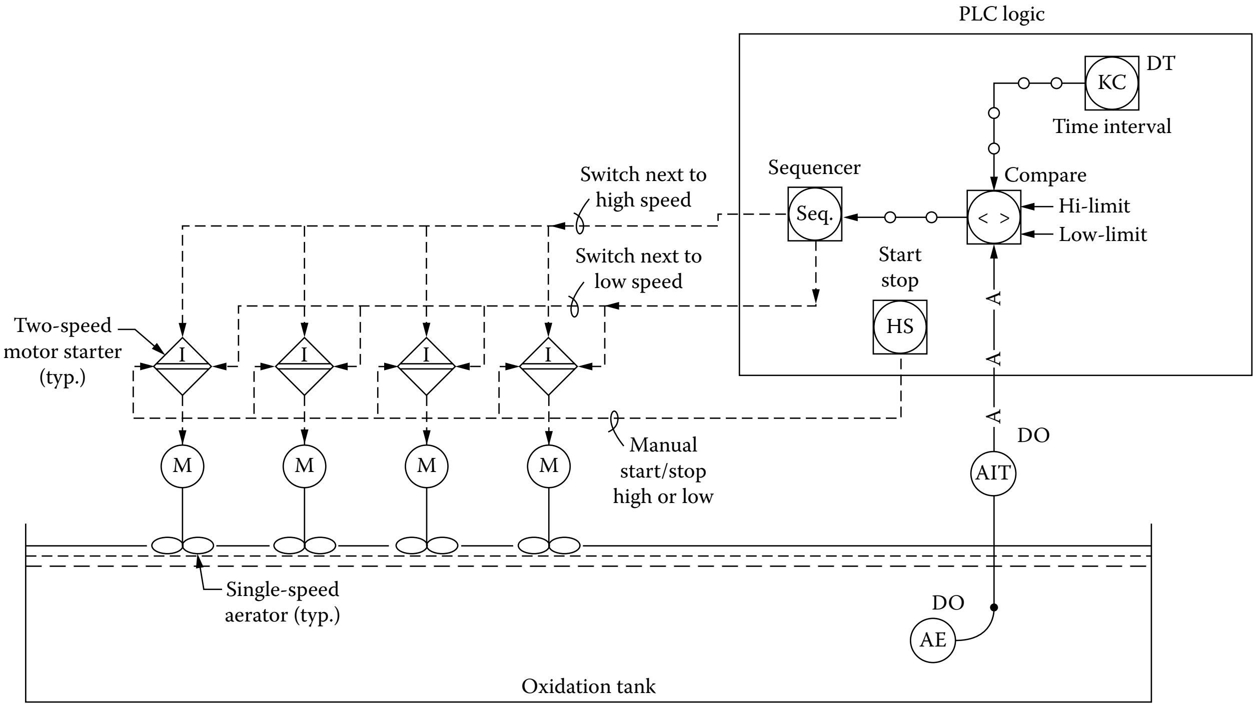


FIG. 8.1d
Dual-speed aerator control system.

started or stopped based on DO level is loaded in a sequence table part of the PLC logic. The operator can change the sequence via operator interface.

A DO sensor/analyzer sends a 4–20 mA signal to the PLC. The DO signal is linearized and is compared against the high and low DO set points. If the DO is below the low limit set point for an adjustable time period, the sequence logic sends a start command to the "next" aerator to start. As long as the DO level is below the set point, the sequencer will start the next aerator until either all aerators have been started or the DO level is within the high and low DO set points.

If the DO signal is above the high set point for an adjustable time period, the sequencer issues a stop command to the next aerator. No aerator is stopped or started by the sequencer as long as the DO level is within the band set by the high and low DO set points. Excessive cycling of the aerators can be avoided by increasing the DO control band or by adjusting the time delay before a start or stop command is issued to the aerators.

Dual-Speed Control Dual-speed aerators are used to achieve a better control on aeration and save energy. Figure 8.1d shows a typical dual-speed aerator control system. Dual-speed motor starters are used to run the motors at two different speeds. As discussed before, a DO sensor/analyzer is used to measure DO in the wastewater and transmit a 4–20 mA signal to PLC.

PLC logic performs the following functions:

At the end of each adjustable preset time interval of DT, as determined by an interval timer logic, check the measured DO signal against the preset high and low DO limits. If at the end of each time interval the measured DO level is less than preset, start the next aerator at low speed in accordance with the sequence loaded in the sequencer.

When all aerators in the zone are running at low speed and the DO level is still below preset, switch the first aerator to high speed.

On subsequent time intervals, if the DO level remains below the low limit, continue switching the next aerator to high speed as determined by the sequencer. If the DO level continues to be below the low limit, eventually all aerators will be running at high speed and the DO level will be at the maximum achievable limit. On the other hand, during any time interval, if the DO level is above the preset high limits during subsequent time intervals, the sequencer will switch the next aerator to low speed one at a time. The aerators can be manually started at high or low speeds or stopped via commands from the operator interface.

Figure 8.1e shows a hypothetical graph of typical aeration capacity variation vs. time that can be expected using dual-speed control. The aeration capacity is at minimum when all aerators are running at low speed. The aeration capacity is increased by a step of DA during a time interval DT when an aerator is switched from low to high speed. Maximum aeration capacity is reached when all aerators are running at high speed.

Variable-Speed Control Variable-speed drives can be used to continuously vary the speed of the aerator based on DO

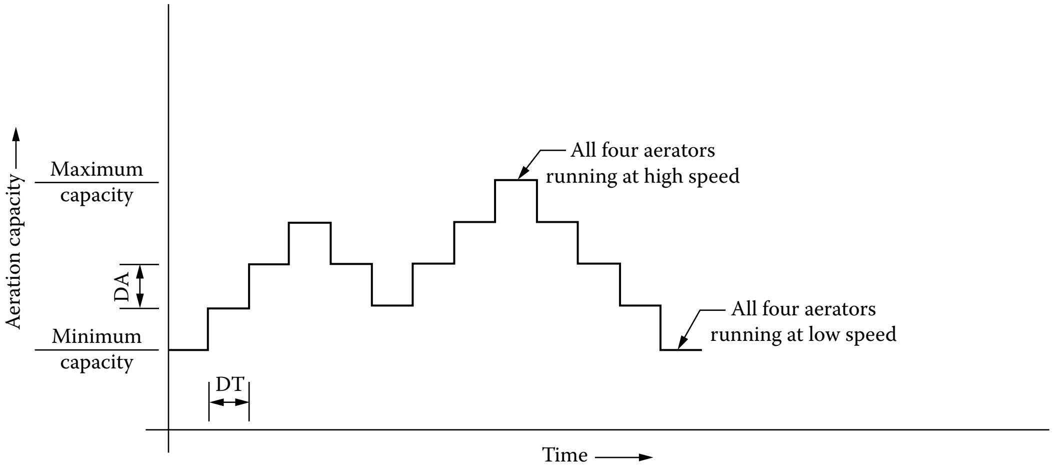


FIG. 8.1e
The above relationship can be expected between aeration capacity and time, when using dual-speed aeration controls.

measurement. The speed can be either manually adjusted from the operator interface based on the DO readings or controlled automatically using a PID algorithm part of the PLC logic. The DO measured variable is compared against the desired set point to generate an error signal, and the controller output and speed of the aerator is modulated to minimize the error. Manual speed control is accomplished by putting the controller in manual and adjusting controller output to desired value.

Figure 8.1f shows a typical variable-speed aerator control loop. It should be noted that the aerators are constant torque machines, and the kW power requirement varies linearly with the speed; therefore, energy savings at reduced speeds are not as dramatic as in case of centrifugal blowers.

Variable-Impeller Depth Control Aeration transfer capacity of a turbine-type aerator can be controlled by varying the submergence depth of the impeller. Figure 8.1g shows a typical variable-impeller depth control system. Two cascaded control loops are used for this control system. A faster internal control loop is used to adjust the vertical position of the impeller to maintain the position set point.

Position transmitter ZT provides the measured variable and PID controller ZIC compares the actual position with the position set point and adjusts the vertical position to minimize the error. The position set point is modulated proportional to the output of PID controller AIC (external slower control loop), whose output is based on error between the measured DO and the desired DO set point.

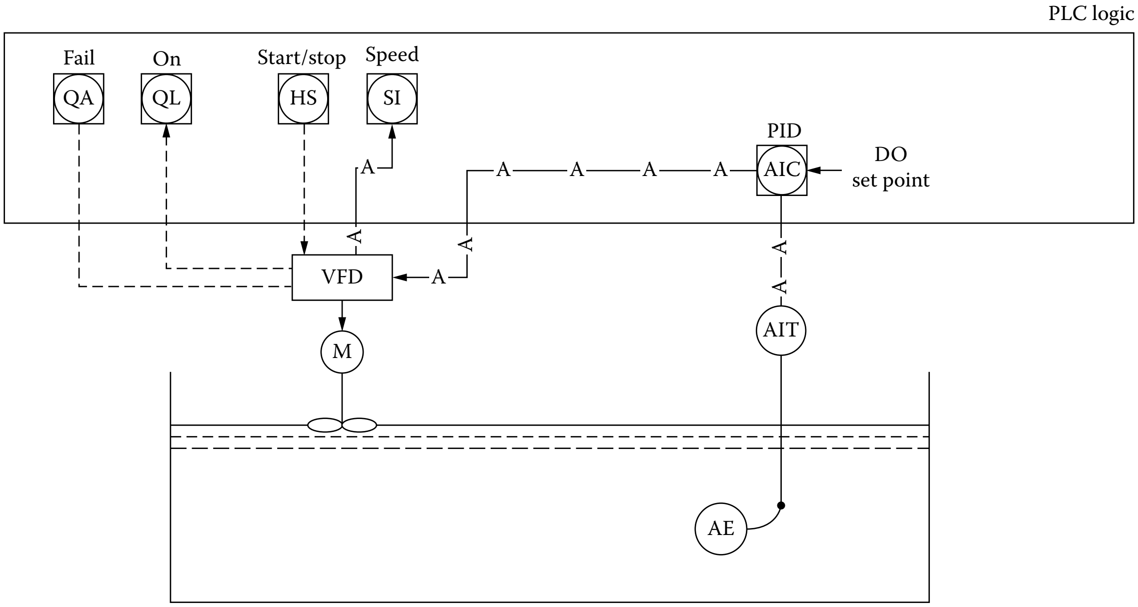


FIG. 8.1f
Variable-speed aerator control system.

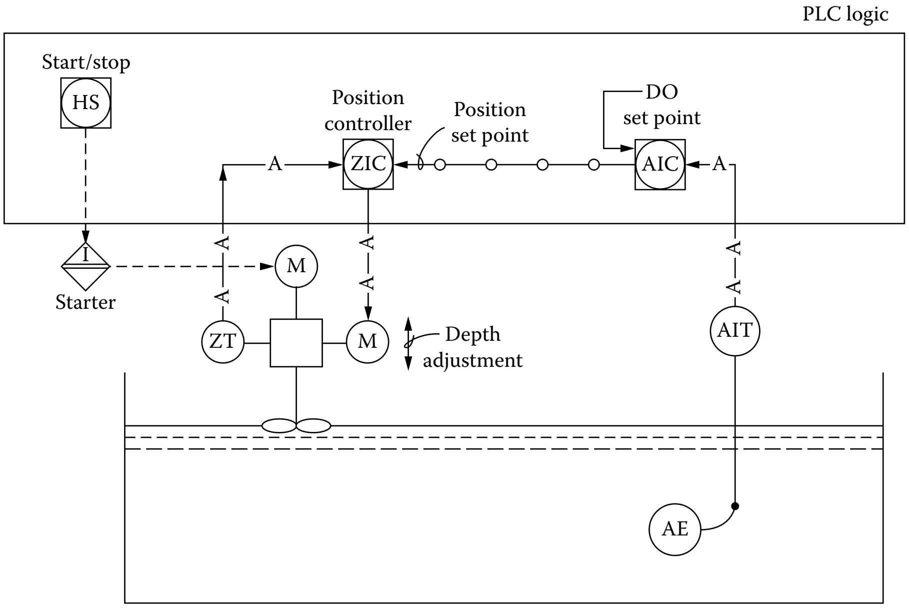


FIG. 8.1g
Variable-impeller depth control system.

Variable-Level Control Another method to control the oxygen transfer rate of a turbine-type aerator is by controlling the level of wastewater in the oxidation tank, thereby controlling the submergence of the turbine impeller. The liquid level is controlled by raising or lowering a weir. The cascaded control loops, similar to the variable impeller depth control loop described above, are used.

The faster inner loop (ZIC) controls the position of the weir, and the position set point is modulated by the external slow DO control loop (AIC). Adjusting the weir position adjusts the liquid level, thus adjusting the submerged impeller depth.

Diffused Aeration Diffused aeration is accomplished by injecting air from an external source into the aeration basin through a diffuser. A diffuser can be as simple as a perforated pipe or a complex device. Air is pushed through the diffuser using air blowers.

Several types of configurations and aeration control strategies are used for diffused aeration of wastewater. Proper application of blowers, blower capacity control, and control valve selection can lead to an energy-efficient aeration system. This subsection provides general background on types of diffusers and blowers used, methods of blower capacity control, blower surge control, and the types of control valves used for air distribution. Measurement of DO and typical control strategies are discussed in later subsections of this section.

Aeration Diffusers

There are two basic types of diffusers used in diffused aeration, namely the coarse bubble diffusers and the fine bubble diffusers. Three factors are important in selection and application of diffusers, namely the head loss through the diffusers, the oxygen transfer efficiency, and the layout and arrangement of diffusers. Excellent discussion of diffusers and their application is available in the literature referenced at the end of this section. A brief discussion is provided here as an overview.

Coarse Bubble Diffusers These diffusers release large air bubbles into the wastewater. These diffusers are further divided into four general categories: orifice, valve, shear, and shallow submergence diffusers. The selection of the type of diffuser depends upon application, head loss characteristics, and oxygen transfer efficiency of the diffuser. The main advantage of the valve-type orifice diffuser is that it prevents backflow of liquid in the air header on loss of air pressure.

Shear-type diffusers are difficult to attach to removable pipe headers, and therefore these types of diffusers are used only in applications where they can be attached to the bottom of the tank. The shallow submergence diffusers are installed at shallow depths of about 3 ft below the liquid surface. The main advantage of shallow submergence-type diffusers is the low air pressure required, which allows the use of low-pressure blowers, reducing blower capital cost. In general, the coarse bubble diffusers are less efficient compared to fine bubble diffusers.

Fine Bubble Diffusers The fine bubble diffusers produce fine air bubbles, which produce less turbulence in the wastewater, as compared to the coarse bubble diffusers. Diffuser devices are generally made from porous ceramic materials or elastometric membranes. Fine bubble diffusers are available in flat tubes, hollow plates, or dome-type construction. The head loss characteristics for different types of fine bubble diffusers are not much different but their oxygen transfer efficiencies may differ significantly.

The fine bubbles created by porous ceramic and elastometric membrane-type diffusers mounted on a piping grid at the bottom of the aeration basin provide the best oxygen transfer efficiency.

Aeration Blowers

There are two types of blowers commonly used in aeration applications, namely the positive displacement (PD) blowers and the centrifugal blowers. There is plenty of literature available on the subjects of blowers and their applications in the references listed at the end of this section. Only a brief overview is included in this paragraph. Selection of an appropriate blower type should be based on capital cost, operating cost, efficiency, and the flow/capacity turndown capability. Complete life-cycle costs should be analyzed for various blower types/configuration before a final selection is made.

Positive Displacement Blower The rotary, two-impeller, PD blower is a constant volume, variable-pressure machine. At the rated speed, this type of blower provides constant flow at widely varying discharge pressures. Figure 8.1h shows the typical operating characteristics of a PD blower. Low capital costs and its ability to operate at widely varying pressures are the advantages of a PD blower. The disadvantages include the difficulty of throttling the air flow rates to meet the varying demands, high maintenance cost, and noisy operation.

The capacity control of a PD blower can be achieved either by blowing off excess air through a blow-off valve or by varying the speed of the blower using a variable-frequency drive. The first method of blowing off the excess air during low air demand period wastes energy and is not recommended.

Reducing air flow by reducing the blower speed is a better way of controlling air flow of a PD blower, because it conserves energy by reducing blower horsepower requirement. However, it should be noted that the blower efficiency drops significantly at lower speed and the energy savings are not as dramatic, as in the case of the centrifugal blower, because the horsepower vs. speed relationship is linear. A turndown of 100–50% capacity is possible by using variable-frequency drives to vary the speed of a PD blower.

The PD blower is used typically in small aeration systems requiring aeration flows of 1000 SCFM or less. PD blowers were used extensively before 1960; however, the development of high-efficiency centrifugal machines with single- and dual-vane capacity control has limited the use of PD blowers to smaller systems with variable head requirements.

Centrifugal Blowers Figure 8.1i shows a typical operating characteristic of a centrifugal blower. The curve *B*1-*B*2 is the blower-operating curve. Curve *S*1-*S*2 represents the system curve that includes the static head and the frictional resistance. The intersection of the two curves (point "*O*") is the operating point of the blower. At the operating point, the blower is running at rated speed to deliver the flow at the pressure dictated by the system curve, without any speed control or valve throttling.

It should be noted that the blower curve is almost horizontal at point *B*1 and shown as a dotted line to the left of point *B*1. "*B*1" is the "surge condition" point of the blower, which indicates unstable blower operating point; the blower should not be operated at or to the left of the point *B*1. In all practical applications, blower capacity control is required to meet the varying demands of the aeration process. The blower discharge flow can be controlled by using one of the three methods: (1) discharge throttling by using a control valve in the discharge line, (2) changing the speed of the blower, and

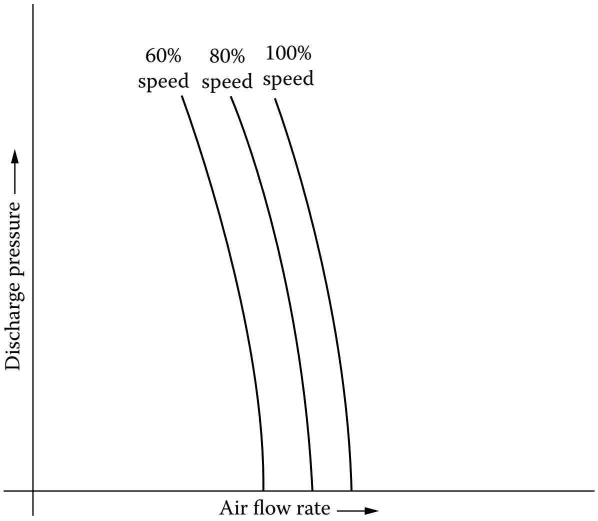


FIG. 8.1h
Characteristic curves of a typical rotary positive displacement blower.

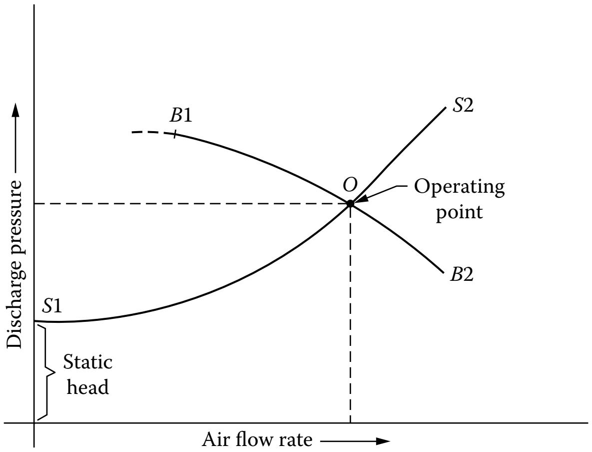


FIG. 8.1i
Characteristic curves of a typical centrifugal blower.

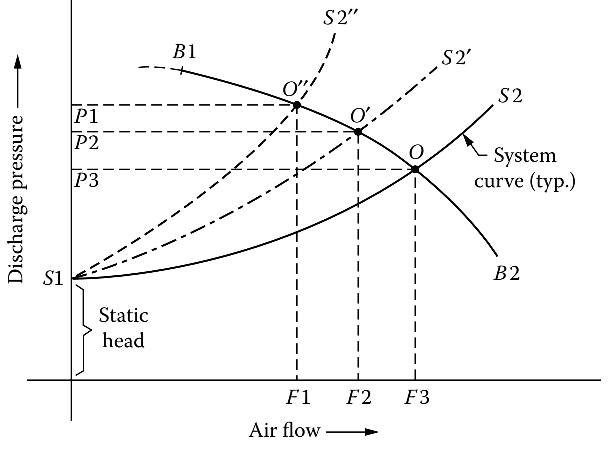


FIG. 8.1j
The effect of discharge valve throttling on the outlet pressure and flow of a centrifugal blower.

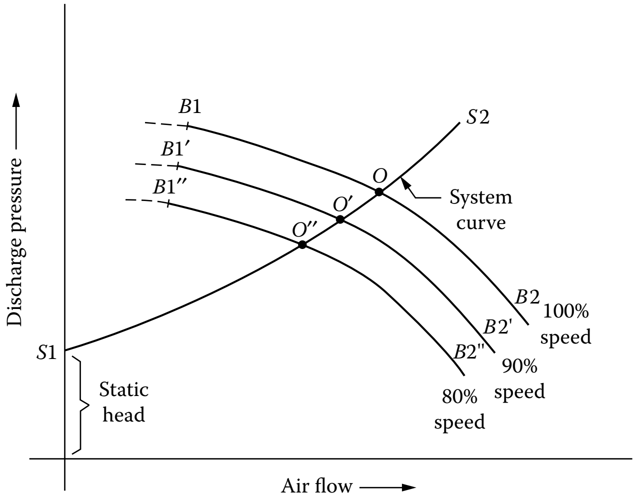


FIG. 8.1k
The effect of suction valve throttling on the outlet pressure and flow of a centrifugal blower.

(3) throttling the inlet using a throttling valve or inlet guide vanes.

Discharge Throttling The effect of discharge throttling on the blower-operating characteristic is shown in Figure 8.1j. Curve $B1$-$B2$ is the blower-operating curve. $S1$-$S2$ is the system curve representing system static head and resistance without any throttling. Point "O" is the operating point without any throttling. System curves $S1$-$S2'$ and $S1$-$S2''$ represent increased resistance for two separate throttled positions of the discharge valve. Note that as the valve is throttled to reduce flow, the operating point "O" moves to the left on the operating curve to point O' and O'', respectively, increasing the header pressure.

As the valve is throttled and the operating point starts moving toward the surge point $B1$, care must be taken so that the operating point never reaches the surge point $B1$. Several methods of surge control are discussed later in this section. Although discharge valve throttling is an effective method of blower capacity control over a limited range, it suffers from two problems.

First, the throttling of valve adds artificial resistance in the system, wasting valuable pressure and energy across the discharge valve, and second, the method is susceptible to reach surge conditions if the pressure and flow are not carefully controlled. To avoid surge conditions, blower output should be kept above 60% of rated output. Because throttling of discharge valve wastes energy, other methods of controlling blower output are more attractive.

Blower Speed Control Blower output can also be regulated by changing the blower speed. Figure 8.1k shows the effect of blower speed control on the blower operating curves. The operating points O, O', and O'' are for 100, 90, and 80% blower speeds, respectively. It should be noted that lower speeds will cause lower system pressure, and any fluctuation in speed will cause fluctuation of flow and pressure conditions.

There is potential for significant energy savings due to the cubic relationship between speed and horsepower requirements for a centrifugal blower. Variable-speed control of blowers is most commonly used with the multistage centrifugal blower applications. Multistage blowers tend to lose efficiency at lower speeds.

Blower Inlet Throttling Blower inlet throttling can be accomplished by using either a throttling control valve or inlet guide vanes (IGVs) at the inlet of the centrifugal blower. The inlet-throttling valve to modulate the blower capacity is used with the multistage centrifugal blowers. However, in this method of capacity control, the blowers tend to lose efficiency at lower capacities due to the pressure drop generated by the inlet-throttling valve.

The IGV are used for capacity control of single-stage centrifugal blowers. The design of the single-stage blowers incorporates IGVs used for controlling the flow and head conditions without a need for speed adjustment of the impeller. In the past, single-stage blowers were controlled utilizing a single set of vanes positioned in the front of the impeller (IGV). Most of today's single-stage blowers utilize a dual vane control system. The two types of controls are briefly described below.

Single vane control utilizes a series of guide vanes positioned in front of the impeller for air flow control. These vanes are utilized to throttle the amount of air allowed into the blower's impeller. By throttling (restricting) the incoming inlet air, an increased pressure differential is built across the blower, resulting in a decrease of discharge air volume. This, in turn, reduces the amount of work required by the mechanical process and reduces the loading on the electric motor.

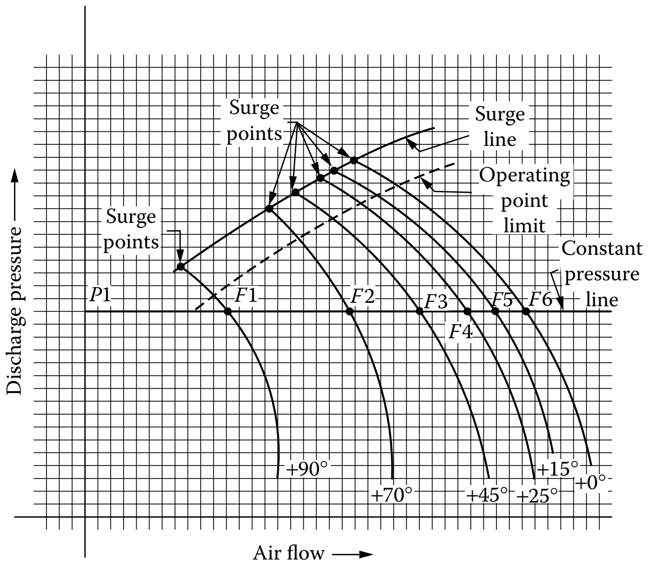


FIG. 8.1l
The characteristic operating lines and surge points when a constant discharge pressure is maintained by setting the inlet guide vane of a centrifugal blower at various angles.

This load reduction translates into a reduction in amp draw, which is directly related to operational expense through the consumption of electric power.

Although less than in the throttling valve of a multistage, the single vane control causes the single-stage blowers to lose efficiency at partial capacity due to the pressure drop generated by the inlet vanes. Figure 8.1l shows how blower capacity control is accomplished by modulating inlet guide vane angles. Again, the surge points are noted for each vane angle position. Care must be taken to avoid flows and pressure conditions corresponding to the surge points. Surge control methods are discussed later in this section.

Dual vane control utilizes two sets of control vanes, one fore and one aft of the blower's impeller. In this control design, the aft or discharge guide vanes are used to control the aerodynamic shape or flow pattern of the air, as it is released from the tip of the impeller blade. By increasing or restricting this airflow path, the amount of flow coming off the impeller wheel is controlled. In concert with this series of flow control vanes, the blower also utilizes onboard instrumentation (differential pressure transmitter and inlet air temperature transmitter), a PLC processor, and an additional set of inlet efficiency optimizing control vanes to achieve optimal airflow efficiency through the blower.

The three variables of machine flow (position of discharge guide vanes), inlet air temperature (this is related to air density of mass), and the differential pressure measured across the machine are utilized to compute the optimal position of the inlet guide vanes in relationship to these three input variables. This process thereby keeps a balanced impeller condition by putting no more air into the impeller than is required by the discharge flow demand.

This dual vane control technology dramatically increases the efficiency of the mechanical blower process. Because of the dual vane control system, the single-stage unit can modulate capacity from 100% to approximately 40%, maintaining the high efficiency practically constant throughout the entire turndown. Today, single-stage blowers with dual vane control systems are the most efficient units on the market. For installation greater than 100–150 hp and higher, with cost of power of $0.035 per kWh and higher, single-stage dual vane blowers have the lowest life cycle cost. The surge points and related control strategies are discussed in the following paragraphs of this section.

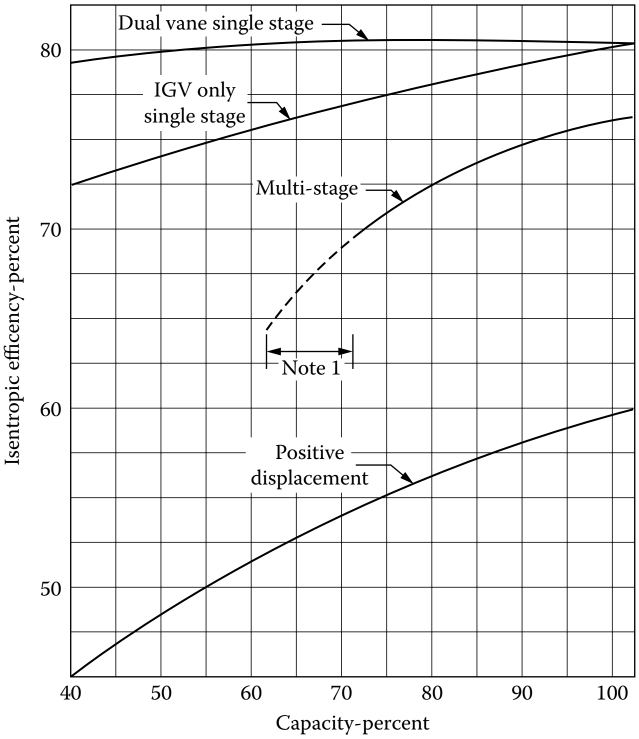
Comparison of Blowers Over the last 75 years, blower design has shown a continuous improvement in optimizing efficiency and operating costs. This improvement in efficiency and lower operation and maintenance cost translates into higher capital cost. Table 8.1m shows an approximate comparison of relative cost and efficiency for various types of blowers used in the aeration process.

Before 1960, positive displacement blowers were widely used in aeration applications. Between 1960 and 1980, multistage blowers became popular due to their higher efficiencies. During the 1980s and 1990s, the single-stage blower with a single inlet vane (IGV) was used extensively because of its higher efficiency, better turndown capability, and the fact that these blowers did not require variable-speed drives.

The present state of the art is the use of single-stage dual vane blowers. Although higher in capital cost, it has the advantage of maintaining high efficiency throughout the entire turndown range of 100–45% of capacity. Figure 8.1n shows isentropic efficiency curves for different types of blowers. It can be seen why single-stage single vane and dual vane blowers are preferred for high-volume aeration applications. The multistage blowers are not only less efficient at 100% capacity but rapidly lose efficiency as the inlet butterfly valve is closed to reduce capacity. Turndown is often limited to 70–80% of capacity. Although low in capital cost, their lower efficiency and high operational cost limit their use in high-volume applications. The positive

TABLE 8.1m
Comparison of Relative Cost and Efficiency for Various Types of Blowers

Type of Blower	*Relative Cost*	*Isentropic Efficiency (%)*
1. Positive displacement blower	1.0	50–60
2. Multistage centrifugal blower with inlet throttling or variable-speed drive	1.5	65–75
3. Single-stage single inlet vane (IGV) centrifugal blower	2.0	72–80
4. Single-stage dual vane centrifugal blower	2.5	80–85



Note:

1. Dashed multistage curve indicates limited turndown

FIG. 8.1n

The isentropic efficiencies of four blower designs.

displacement blowers are used in smaller applications of up to 3000 SCFM (100 hp). As seen in Figure 8.1n, PD blowers have very poor efficiency at 100% capacity, and the efficiency drops rapidly as speed control is used to reduce capacity. Although lowest in efficiency, PD blowers remain popular in small applications due to their low cost.

It is important to note that both capital cost and operational cost must be considered during the blower selection process. A present-worth analysis based on life-cycle costs should be performed to select the best type of blower for a given application.

Surge Control of Blowers Centrifugal blowers have a minimum flow point below which the blower performance becomes unstable. This unstable operating point is called surge condition, and when this condition is reached, wide oscillations in pressure and flow conditions can occur, causing damage to the blower. The surge condition is prevented by always maintaining flow greater than the safe minimum required. This is accomplished by providing a blow-off valve on the discharge line and blowing off excess flow through this valve during low-flow conditions. Figure 8.1o shows typical characteristic curves for a variable-speed centrifugal blower and the function of a surge control algorithm.

The graph of minimum flow surge points for various pressure and speed conditions is called the surge line. The surge line represents a parabolic pressure flow relationship. The purpose of a surge control algorithm is to keep the blower operating to the right of the surge line at all times. The dotted line in Figure 8.1o shows the minimum flow blower operating points maintained by the surge control algorithm. Figure 8.1.1 also shows the surge points and the operating point limit set by the surge control algorithm for a single-stage blower using inlet guide vanes. Blower manufacturers generally provide the data for the surge line as part of their package to ensure that the blower will never be allowed to go into surge.

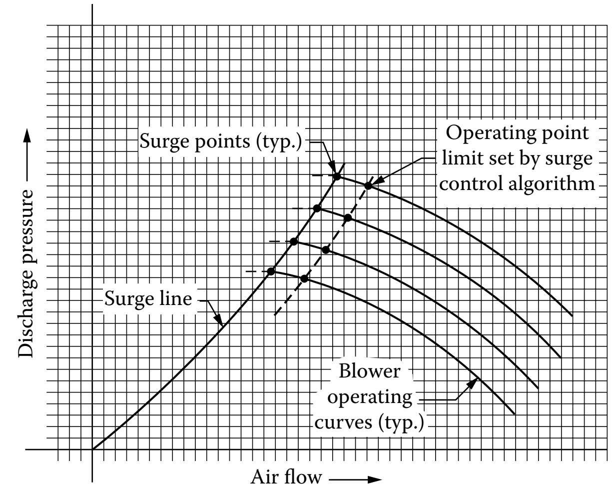


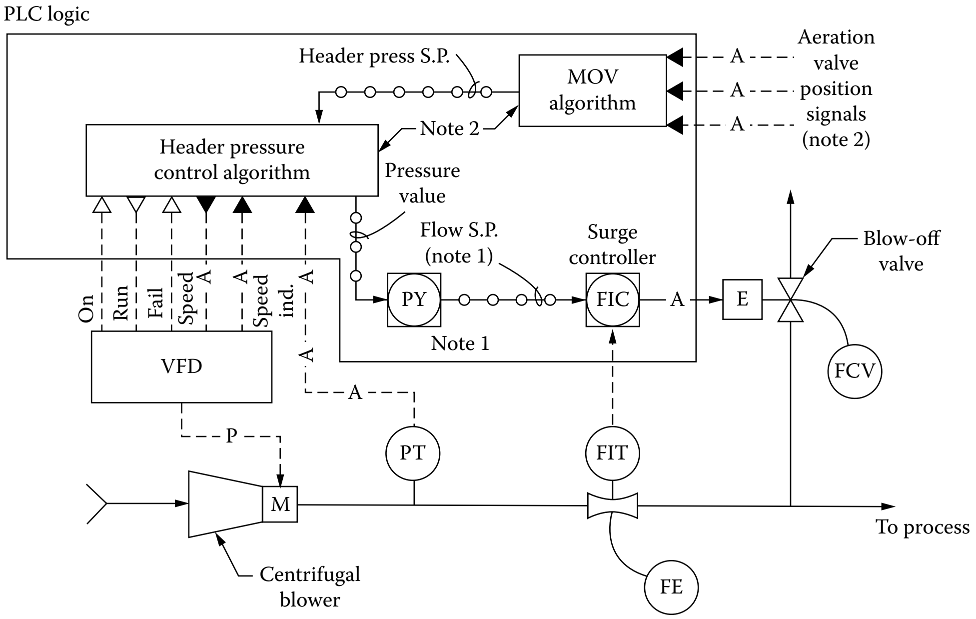
FIG. 8.1o

The characteristic curves and surge points of a variable-speed blower at various speed settings.

Several different control strategies are used to prevent surge. Figure 8.1p shows a typical surge control strategy for a variable-speed centrifugal blower. Blower discharge pressure, flow, and aeration cell valve positions are measured and transmitted to the PLC as analog inputs. Based on aeration valve position signals, the most-open valve (MOV) algorithm computes the header pressure set point.

The header pressure algorithm maintains the pressure set point by controlling the blower speeds. Both the MOV and header pressure algorithms will be discussed in the Aeration System Control subsection. The "PY" block in Figure 8.1p receives the pressure value from the header pressure algorithm and computes the minimum permissible flow set point using either a look-up table or an equation (dotted line in Figure 8.1o). The flow set point computed by the "PY" block is used by the surge controller (FIC) to control the blow-off valve to maintain the flow set point that avoids surge conditions.

Figure 8.1q shows another control strategy that can be used for surge control of the centrifugal blower that uses inlet guide vane throttling for control of blower capacity. In this example, the blower motor current and the inlet vane position are continuously measured and transmitted to the PLC.



Notes:

1. Minimum permissible flow S.P. corresponding to a pressure value to prevent surge is computed using either a look-up table or an equation, part of "PY" block.
2. Detailed strategies for most-open valve (MOV) algorithm and header pressure control are described in Figure 8.1w and the subsection Aeration System Control.

FIG. 8.1p
Surge protection control strategy for a variable-speed centrifugal blower, which prevents the blower flow from dropping below an allowable limit.

When inlet vanes are modulated by the header pressure algorithm, to reduce blower output, the motor current draw corresponding to surge point also reduces. The relationship between the inlet vane position and the surge point current draw is nonlinear. This nonlinear relationship can be programmed in the PLC either as a look-up table or an equation. For a measured inlet vane position, the desirable minimum motor current draw is computed ("ZY" block) and used as a set point for the current control algorithm, part of the surge controller (IIC).

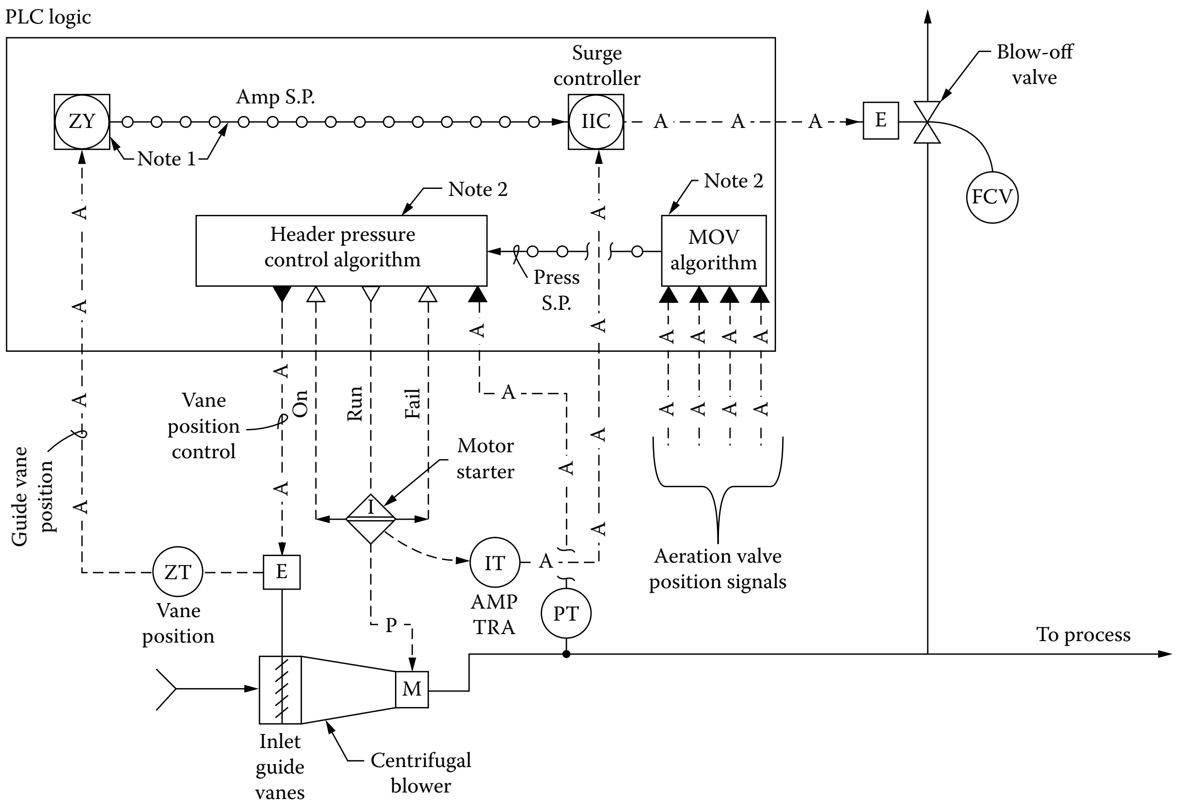
The error between the set point and the actual motor current measurement is used by the PID algorithm (IIC) to open the air blow-off valve to increase the motor ampere draw to avoid the surge point. As was noted earlier, the header pressure and MOV control algorithms will be described in the Aeration System Control subsection. In both Figures 8.1p and 8.1q, the blow-off valve is a modulating-type control valve.

Air Distribution Control Valves

Air distribution control valves are important building blocks of any aeration control system. Selection, sizing, and application of the correct type of control valves for air distribution are very important for a properly operating aeration control system. An oversized control valve can reduce the effective air flow turndown capability, create flow oscillations and difficulty in maintaining flow or pressure set points. However, an undersized control valve can cause excessive pressure drops, energy losses, and increased noise levels. The subject of control valves, their selection, and sizing is discussed in detail in Chapter 6 of this handbook; however, a brief discussion is included here for the sake of continuity.

Typical valve-sizing calculations should include checking for choked flow, determination of the expansion and compressibility factors, and the calculation of valve capacity coefficient C_v. The valve-size selection should also include air velocity and noise considerations.

Selection of valve flow characteristics is an important consideration in designing a stable control system. The most common flow characteristics are linear, equal percentage, and quick opening. For aeration flow application, the equal-percentage characteristics are the most suitable. As the name suggests, the equal-percentage characteristic increases the valve flow capacity by the same percentage for each equal increment of travel. Equal-percentage butterfly valves are commonly used in the aeration flow and pressure control applications.



Notes:

1. Minimum permissible blower amp set point corresponding to a guide vane position to prevent surge is computed either using a look-up table or an equation, part of ZY block.
2. Detailed strategies for most-open valve (MOV) algorithm and header pressure control are described in Figure 8.1w and the subsection Aeration System Control.

FIG. 8.1q
Surge protection control strategy for an inlet vane controlled centrifugal blower, using the inlet guide vane position and the motor current drawn to detect the approach of a surge condition.

Automatic valve actuators serve to position the valve in accordance with the control signal received. The commonly used actuators in aeration applications include the pneumatic spring and diaphragm and the electric motor-driven designs.

Pneumatic Actuators A pneumatic actuator in conjunction with a three-way solenoid valve is used in an on/off application, whereas an actuator with a positioner allows modulation of the valve in response to a control signal. The electronic 4–20 mA control signal is converted to 3–15 psi pneumatic signal applied as input to the positioner.

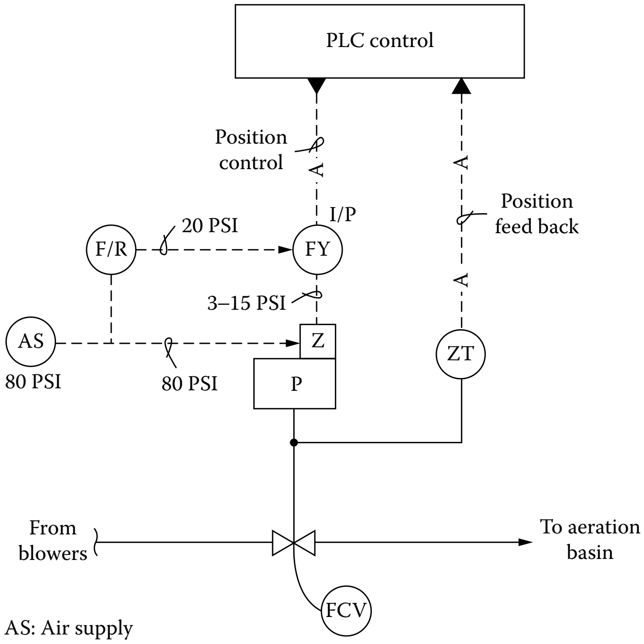
Figure 8.1r shows typical control and instrumentation devices associated with pneumatic diaphragm-type valve, actuator, and positioner. For applications where fail open- or fail close-type operation is desired, the pneumatic spring-and-diaphragm-type actuator has a distinct advantage because the failure modes are easily implemented.

Electric Actuators The electric motor-driven actuators can be on/off or modulating. The actuator consists of an electric motor and a reversing starter/contactor that allows the motor to run in forward or reverse direction, thus opening or closing the valve. Modulating valves can be throttled by a 4–20 mA electronic signal.

The actuators for smaller valves operate on 120 V, single-phase AC power, utilizing single-phase "capacitor start" motors, and for the larger valves, the actuators operate on 480 V, three-phase, AC power supply using three-phase motors. The valve operational torque requirements determine the type of motor and power supply.

Typically a modulating valve with electric actuator consists of the following devices:

A local-off-remote switch. When this switch is in local position, the valve can be manually controlled using either actuator-mounted open–close pushbuttons or a local potentiometer to



AS: Air supply
F/R: Air filter/pressure regulator
FY: Current to pressure converter
ZT: Position transmitter
Z: Pneumatic positioner
P: Pneumatic actuator
FCV: Flow control valve

FIG. 8.1r
The components of a control loop consisting of a PLC that is controlling a pneumatic control valve with positioner and stem position feedback.

adjust the valve position. When the switch is in remote position, the valve is controlled by a remote 4–20 mA signal generated from the PLC-based control system.

Limit switches indicate fully open or closed position of the valve; they are often wired to the PLC for status indication.

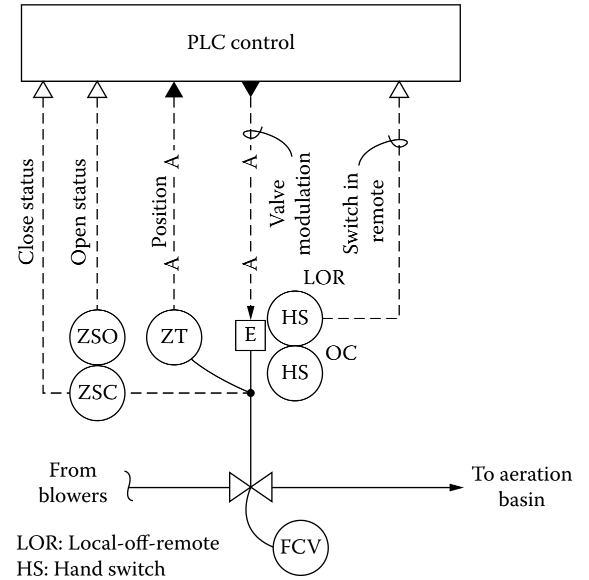
A position transducer transmits an analog 4–20 mA signal proportional to the valve position to be wired to the PLC-based control system.

Figure 8.1s shows the components of a loop consisting of an electrically operated throttling valve, which is controlled by a PLC.

Materials of construction for the valve body, trim, and liner must be specified to suit the application. The best way to specify a control valve is to use a standard ISA data sheet with all the blanks filled appropriately.

MEASUREMENT OF DO CONCENTRATION

Section 8.43 in the first volume of this handbook provides an in-depth discussion of the features and capabilities of all the oxygen detector designs that can be used with liquid samples. Accurate measurement of DO concentration is very important to the success of any aeration control strategy.



LOR: Local-off-remote
HS: Hand switch
OC: Open-close
ZT: Position transmitter
ZSO: Limit switch-valve open
ZSC: Limit switch-valve closed
E: Electric actuator
FCV: Flow control valve

FIG. 8.1s
The components of a control loop consisting of a PLC that is controlling an electrically actuated control valve, which is provided with stem position feedback.

Typically, a DO sensor generates an electronic low-level signal proportional to the DO level in the process fluid, which is converted to a 4–20 mA analog signal by the transmitter/analyzer for connection to the control system.

Galvanic Cell

In a galvanic-type DO sensor, the reduction of oxygen at the cathode and production of electrons at the anode generates a current. The magnitude of the current flowing between the cathode and anode is proportional to the amount of DO present in the process fluid. The electrodes are either made of silver alloy and zinc or platinum and lead. There are two types of designs available in the galvanic-type sensors.

In the first type of design, the electrodes are enclosed in a thin membrane to isolate and protect them from the process fluid. The membrane-type sensors have to be periodically cleaned to keep them operational, and they require electrolyte solution. Some manufacturers supply integral automatic cleaning systems with the electrodes for ease in maintenance. The periodic automatic cleaning is accomplished using a high-pressure air jet to scour the membrane of the accumulated deposits. The sensor is electrically connected to the analyzer/transmitter for generating 4–20 mA analog output proportional to DO. The high-pressure air supply and associated control is incorporated in the analyzer/transmitter enclosure. The control logic, part of the analyzer, holds the analog output during the cleaning cycle to avoid interruption of the signal to the control system.

In the second type of design, the galvanic electrodes are directly immersed in the process fluid without any membrane. The cleaning of electrodes is accomplished by a motor-driven diamond grindstone to provide a continuous cleaning action on the metal electrodes. This self-cleaning action on the electrodes addresses the maintenance problem. The electrodes are protected from large suspended solids using a sample chamber. The sample chamber continuously oscillates to pump fresh sample to the measuring electrodes.

Polarographic Cell

These types of sensors use oxygen current measurement technique, wherein a "polarizing voltage" is applied across a selected electrode, where the oxygen reduction occurs via an electrical reaction called *polarographic*. The polarographic current is proportional to the oxygen concentration in the process fluid.

The sensor is a three-electrode polarographic sensor consisting of a silver reference, silver anode, and a gold cathode. The electrodes are covered by a 50μ hydrophobic membrane, which is preinstalled in a membrane head assembly. The membrane and membrane head assembly are constructed of fluorinated ethylene propylene and polyoxymethylene, respectively. The sensor body is constructed of stainless steel and polyoxymethylene. The sensor includes automatic temperature compensation to compensate for changes in process temperature.

Again, the presence of membrane requires extra care in maintenance and the need for electrolyte solutions.

Galvanic and Polarographic Cell Limitations

The galvanic and polarographic DO sensing techniques have been used for the last 50 years. As discussed before, the key components used in these types of sensors are anodes, cathodes, membranes, membrane cleaning devices, and electrolyte solutions. Although these types of sensors have a long track record of successful applications, they pose the following challenges related to operation and maintenance.

Some of the limitations and concerns include:

1. Anodes are eventually consumed and require periodic replacement and calibration.
2. Calibration kits consisting of disposable calibration bags are required to perform "saturation method" air calibrations.
3. Electrolyte solutions require periodic replacement and are subject to contamination.
4. Gases like hydrogen sulfide can contaminate electrodes and electrolytes, requiring replacement.
5. Sensor membranes become coated with grease and dirt and require regular cleaning. Although some manufacturers provide mechanisms like air-blast cleaning for periodic automatic cleaning, these mechanisms also need additional maintenance.
6. The sensors are subject to interference from anything that produces small voltages.

Luminescent DO Sensors

Luminescent DO sensors are relatively new in the DO sensor technology. These types of sensors have been developed to minimize the maintenance problems associated with the galvanic- or polarographic-type sensors. Figure 8.1t shows the basic operating principle of a luminescent-type DO sensor.

The luminescent sensor is coated with a luminescent material. Blue light from a light-emitting diode (LED) is transmitted to the sensor surface. The blue light excites the luminescent material. As the material relaxes, it emits red light. The time from when the blue light was sent and the red light is emitted is measured by the photo diode. The more oxygen that is present, the shorter the time it takes for the red light to be emitted. This time is measured and correlated to the oxygen concentration. Between the flashes of blue light, a red LED is flashed on the sensor and used as a reference.

Although not enough actual operational data is available at the time of this writing, the limited experience indicates that this type of sensor has a good potential for success. The luminescent-type DO sensor has the following advantages over the traditional sensors:

1. Membrane cleaning and replacement is eliminated because the technique does not use membranes. Because membranes are not used, there is no need for electrolytes.
2. There are no anodes or cathodes, and this eliminates the need for their replacement.

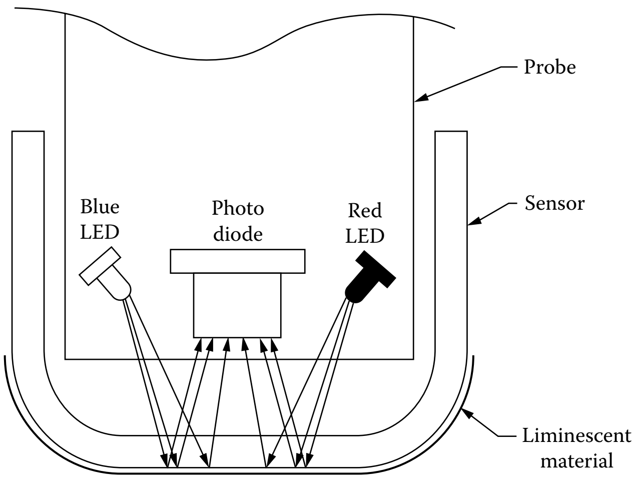


FIG. 8.1t
The main components of a luminescent dissolved oxygen sensor.

3. Sensing technique is less susceptible to interference from spurious signals.
4. The sensor can be mounted at any desired submersion depth.

MISCELLANEOUS FIELD INSTRUMENTS

Other important measurements required in any aeration control system are air flow, header pressure, and air temperature measurements. Air flow to each aeration cell, part of the aeration basin, is measured to be used as a process variable in the air flow control algorithm discussed later in this section. The newer thermal dispersion-type flow meter is successfully used in air flow measurement applications.

A pitot tube or a venturi tube can also be used for air flow measurements, but the pitot tube is less accurate than the thermal dispersion-type flow meter. The venturi tube gives satisfactory results, but is expensive and requires long straight pipe runs upstream and downstream of the meter. Orifice plates are not suitable for flow measurement on low-pressure systems. Pressure transmitters are required for measurement of header air pressure to be used for control of blower capacity.

Both air temperature measurement and differential pressure measurement across the blower are used by blower manufacturers to develop algorithms for precise control of dual vane single-stage centrifugal blowers.

CONTROL OF DIFFUSED AERATION

Parallel and Cascade Control of Blowers

All major aeration applications require multiple blower units connected to a common header. Blowers are turned ON or OFF and the capacities controlled to suit the aeration cells. Today, many of the high-volume aeration processes use single-stage blowers because of their efficiency and ease of control by modulating inlet guide vanes. Two types of single-stage blowers are generally used, namely single stage with IGV only and single stage with dual vanes (IGVs and variable diffuser vanes). Depending on the type of blower used, two types of capacity control strategies are used:

In *Parallel capacity control*, multiple blower capacities are simultaneously modulated to suit the air demand. As explained in the next paragraphs, due to the drooping "surge line," the single IGV blowers are restricted to use only parallel capacity control, which allows capacity control without the possibility of approaching surge conditions.

Cascade capacity control is defined as varying the capacity of one on-line blower while capacities of other blowers are held at either maximum or minimum. Because of the flat "surge line" of the dual vane blowers, the cascade control can be used for these blowers without the possibility of reaching surge conditions.

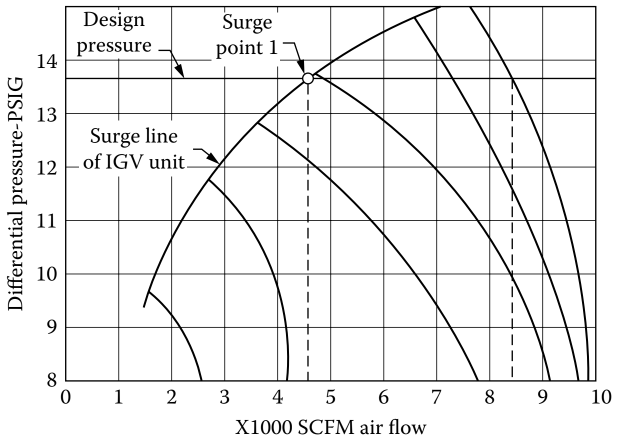


FIG. 8.1u
The surge and operating lines of blowers whose capacity is controlled by inlet guide vanes.

Single-Stage IGV Blowers in Parallel Capacity control of single IGV blowers is accomplished by modulating the inlet guide vanes to throttle the inlet airflow. Closing the IGVs to reduce the capacity increases pressure drop across the IGVs and thus reduces both the discharge pressure and the surge pressure point of the unit.

Figure 8.1u shows the air flow vs. differential pressure curves for the IGV blower with the characteristic curves corresponding to the various positions of the inlet guide vanes. As shown in the figure, as the capacity is reduced by the closure of the vanes, the surge pressure is also reduced. This results in a drooping surge line, which is the reason why these blowers must operate in parallel.

To explain the limitations of parallel operation, assume as shown in Figure 8.1u that the design pressure is 13.7 PSIG, and the maximum flow is 8700 SCFM. Let us further assume the unit is designed to turndown to a minimum of 45% capacity or 3400 SCFM, which is below the surge limit. Hence, at high ambient temperature conditions (100°F), the unit may go into surge before maximum turndown is reached.

This is illustrated in Figure 8.1u, where the surge point 1 is at 4400 SCFM or at 51% of maximum capacity. Thus, in case of an operation where a single blower was meeting the process load until its full capacity (100%) is reached and at that point a parallel blower is started, the resulting total loading would be 50%, while the surge limit is 51%, hence the system would go into surge. On the other hand, if the on-line unit(s) were reduced in capacity and the off-line unit brought on-line at this same reduced capacity, all units would thereby have equal surge pressures, and one unit would not go into surge before the others.

The parallel capacity control has the following limitations:

All on-line units vary capacity at the same time, which means small changes in vane setting creates large flow changes and finer airflow variation cannot be accomplished.

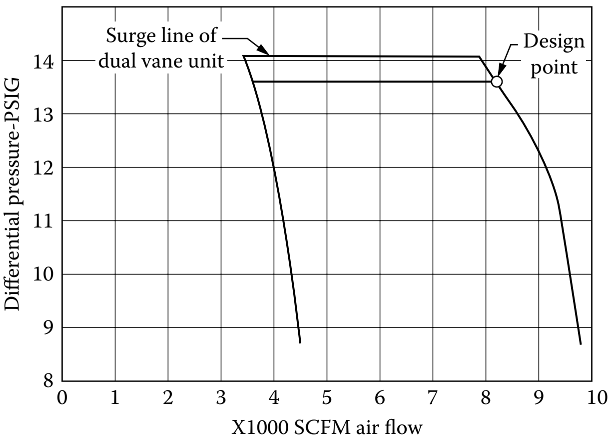


FIG. 8.1v
The surge and operating lines of blowers whose capacity is controlled by dual vanes.

Parallel control requires more elaborate instrumentation to keep all units at the same vane position over long operating periods. Vane position feedback signals or motor current draw are used for keeping the units from "floating" away from each other over long periods of continuous operation.

Single-Stage Dual Vane Blowers in Cascade Single-stage dual vane blowers were developed in the mid 1980s. These designs utilize inlet guide vanes with variable-diffuser vanes. These types of blowers not only have the highest efficiency, but also change the surge curves of single-stage blowers.

As shown in Figure 8.1v, the surge curve is flat (as opposed to drooping in the case of IGV blowers) over the entire turndown range. The surge pressure is the same, regardless of the vane position. This flat surge curve allows cascade control where only one blower capacity is modulated, while others are held constant at maximum or minimum capacity. A typical example of cascade control sequence to maintain discharge header pressure (to meet flow demand) for a three-blower system is described below. In this description, 100% capacity refers to full capacity of single unit.

First Blower The following is the sequence of system response as the process load (demand for air) rises:

On-line at 45% capacity
If, in order to maintain the header pressure set point (S.P.), more air is required:
Increase first compressor to 100% as required to maintain header pressure S.P.
If, in order to maintain the header pressure set point, more air is required:

Start Second Blower

During 2-min prelube cycle, first unit reduces to 45% capacity.
Second unit comes on-line at 45% capacity, resulting in a total capacity of 90%.
Hold second unit at 45%, increase first unit up to 100% as required to maintain header pressure S.P. When first unit reaches 100%, this results in a total of 145% capacity.
If, in order to maintain the header pressure set point, more air is required:
Hold first unit at 100% capacity, increase second unit to 100% as required to maintain header pressure S.P. When second unit reaches 100%, this results in a total of 200% capacity.
If, in order to maintain the header pressure set point, more air is required:

Start Third Blower

During 2-min prelube cycle, the first unit remains at 100%, the second unit reduces to 45% capacity, and the third unit comes on-line at 45% capacity, resulting in a total capacity of 190%.
If, in order to maintain the header pressure set point, more air is required:
Hold first unit at 100% capacity, hold third unit at 45%, increase second unit to 100% as required to maintain header pressure S.P. When second unit reaches 100%, this results in 100% + 100% + 45% = 245%.
If, in order to maintain the header pressure set point, more air is required:
Hold first and second units at 100%, increase third unit to 100% as required to maintain header pressure S.P.
On decreasing air demand or increasing header pressure, a sequence similar to above but in reverse order can be followed to reduce speed to 45% and shut down blowers one at a time.

In this example, cascade control allows precise control of air flow from 45% of the single unit's capacity to the maximum of 300% of the single unit capacity. The control is simplified because exact vane position of the modulating unit is not as important as in parallel control. Because of their higher efficiency and more precise capacity control, the dual vane single-stage blowers are becoming popular.

Aeration System Control

Figure 8.1w shows major control loops for a typical aeration process. This example shows three single-stage blowers (*B*-1, *B*-2, and *B*-3) that are discharging into a common air header. Three separate air lines branch off from the common header and connect to the diffusers in three separate aeration cells. Each aeration cell includes DO sensors and transmitters (AE/AT-1, -2, -3) for the measurement and transmission of DO concentration data. Each of the air lines includes a flow

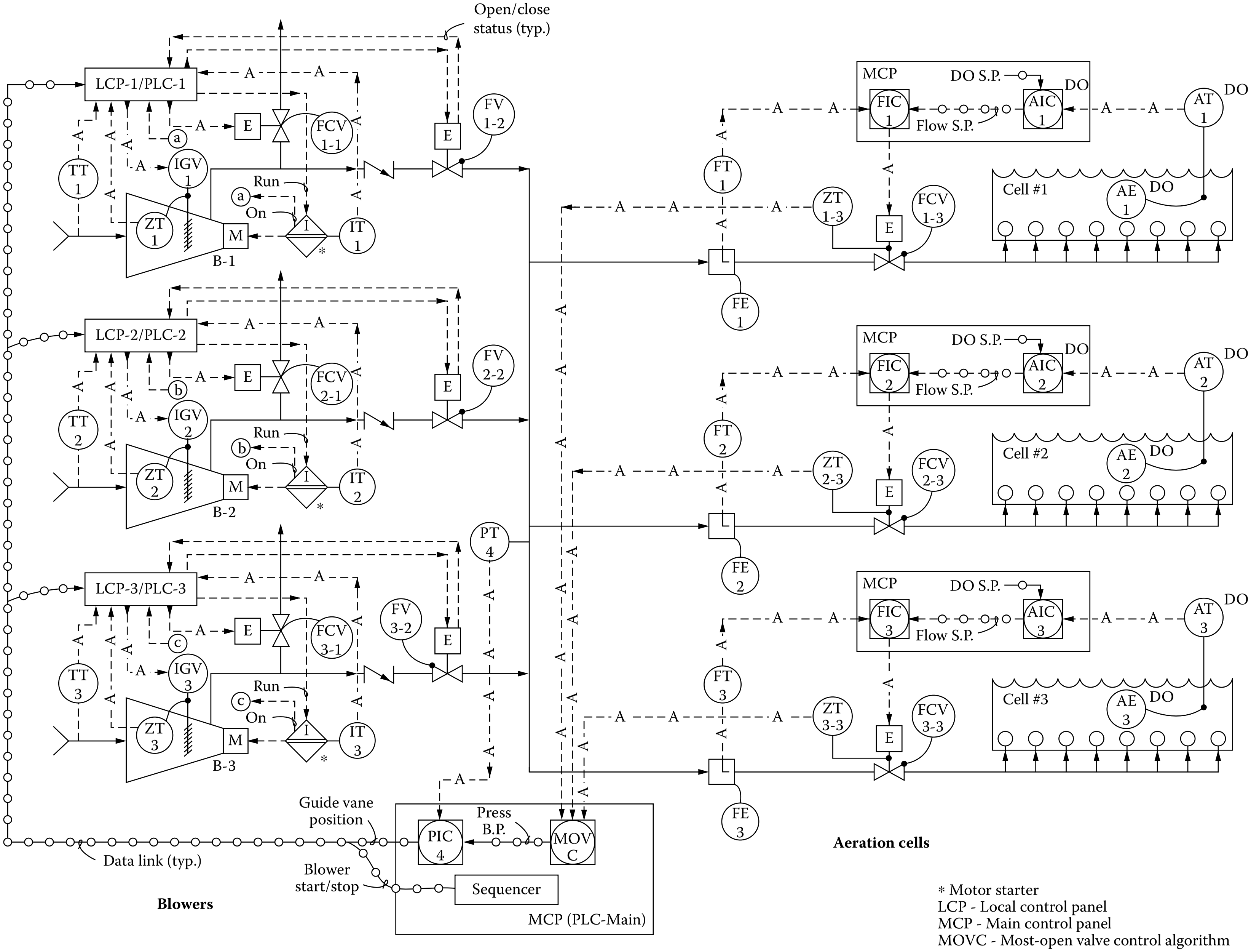


FIG. 8.1w
The overall control system of an aeration process.

meter (FE/FT-1, -2, -3) and a control valve (FCV-1-3, -2-3, -3-3) for measurement and control of air flow into each cell.

The common air header includes a pressure transmitter (PT-4). The discharge air header pressure controller (PIC-4) controls the blower capacities. Each blower discharge line includes a blow-off valve (FCV-1-1, -2-1, -3-1) and a discharge valve (FV-1-2, -2-2, -3-2). The blower can be multistage with inlet throttling valve or single stage with single IGV type or dual vane, but an IGV type is shown.

The position of IGV is transmitted to the local control panel using position transducers (ZT-1, -2, -3). Blower motor current for each blower is also transmitted to the local control panel using current transducers (IT-1, -2, -3). Each blower is equipped with a local PLC-based control panel for control and monitoring of the blower (LCP-1, -2, -3).

A PLC-based main control panel (MCP) is provided for overall control of the aeration process, including DO control, air flow control, air header pressure control, and blower control. A data link allows communication between the main PLC and the local PLC panels.

Major Control Loops Referring to Figure 8.1w, the overall aeration control system can be divided into three basic subsystems:

- Aeration cell control loops
- Header pressure and "most-open valve" control
- Control of blowers to maintain header pressure

Aeration Cell Control Loops (Loops 1, 2, 3) Control of air flow to each aeration cell is accomplished using a dedicated cascade flow/DO control loop as shown in Figure 8.1w. For cell #1, the inner loop or faster loop is a flow control loop, which modulates the flow control valve (FCV-1-3) to maintain the flow set point. Flow transmitter FT-1 provides the measured variable signal to PID flow controller FIC-1, which generates an output to modulate FCV-1-3 to maintain the air flow set point. However, the flow set point must be continuously adjusted to suit the oxygen demand of the aeration cell.

This continuous flow set point adjustment is provided by the slower outer DO control loop. DO transmitter AT-1 provides the DO signal proportional to DO concentration in the aeration cell #1. The DO controller AIC-1 compares the DO signal with the desired DO set point and generates an output that is linearized to provide a flow set point to the inner flow control loop, which maintains the desired air flow to the aeration cell. Thus, the aeration cell control loop consists of two subcontrol loops; the outer DO loop is cascaded with the inner flow control loop for air flow control.

The DO loop is a much slower loop and the response of the DO probe must be averaged over relatively long periods of time. A dead band can be provided on the flow set point such that the set point is not changed unless the change is greater than the preset dead band. This will minimize excessive change in the flow set point. Control loops for other cells (#2 and #3) are similar.

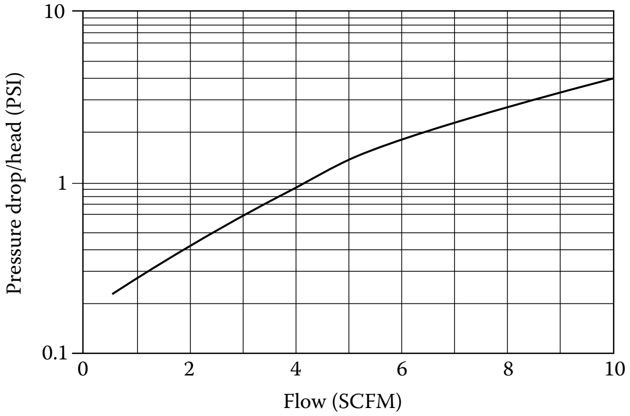


FIG. 8.1x
The relationship between flow and pressure drop in a fine bubble diffuser.

Minimum and maximum flow limits are usually set to maintain minimum mixing requirements and to maintain air flow within efficient operating ranges of the aeration cell diffusers and piping. In general, the speed of the aeration flow control loops should be approximately one tenth of the speed of the blower control loops in order to maintain system stability.

All flow and DO controllers are PID algorithms programmed in the PLC, part of the main control panel.

Header Pressure and Most-Open Valve Control Air header pressure is used as the process variable to control blower capacity. The air header pressure set point must be defined to provide enough pressure at maximum flow to overcome all frictional losses through piping and head loss through the diffusers. Figure 8.1x shows the head loss characteristics of a typical fine bubble diffuser.

Doubling the air flow through the diffuser can more than double the head loss, depending upon where the air flow is located on the pressure drop curve. Thus, for controlling an aeration system, it is very important to vary the air header pressure as a function of flow. Using Bernoulli's equation, the following relationship can be used to define the header pressure set point as a function of aeration flow.

$$SP = K_1 + K_2 \ (Q_T/K_3)^2 \qquad \mathbf{8.1(1)}$$

where

SP = header pressure set point
Q_T = total flow set point (sum of flow set points of all cells)
K_1, K_2, K_3 = constants

The problem in using this equation is that it does not account for the actual system pressure losses. Equation 8.1(1) is generally characterized and constants are evaluated during start-up by conducting tests. No correction is applied for gradual increase in differential pressure across the diffusers as the system gets dirty, diffusers clog, or an abnormal obstruction occurs in the piping to one of the cells.

Another method to relate the header pressure set point to the aeration flow is to define a relationship between the header pressure set point and the most-open air flow control valve position for the aeration cells. Equation 8.1(2) shows a typical relationship:

$$SP = K_1 + K_2\,(VP) \qquad \mathbf{8.1(2)}$$

where

SP = header pressure set point
VP = most-open valve position
K_1, K_2 = constants

Equation 8.1(2) is a linear equation that is a reasonable approximation to the actual relationship, which should follow Bernoulli's equation.

The constants K_1 and K_2 can be determined theoretically during aeration system design and further refined during start-up by conducting tests. During the design phase, the pressure drop to each aeration cell is characterized for minimum and maximum design flows. The flow control valve to each aeration cell (FCV-1-3, -2-3, and -3-3) should be chosen to provide pressure control between 30 and 80% open.

During process start-up, a test is conducted to determine header pressure at minimum and maximum flows. The values of header pressure and most-open valve position (VP) for minimum and maximum flow are substituted in Equation 8.1(2), resulting in two equations (one corresponding to minimum flow and the other corresponding to maximum flow) and two unknowns (constants K_1 and K_2). These equations are solved to determine the values of the two constants K_1 and K_2.

During operation, when the position of the most-open valve is changed either to adjust the air flow to maintain the DO set point or to maintain the flow under increased frictional or diffuser head loss, the header pressure set point is adjusted using Equation 8.1(2). In other words, the header pressure set point is continuously adjusted based on the position of the most-open valve.

For a review of the operation of the most-open valve algorithm refer to Figure 8.1w. Let us assume that cell #1 diffuser is clogged, restricting flow. Controller FIC-1 will cause the valve FCV-1-3 to fully open, trying to maintain the required flow set point. Equation 8.1(2) will increase the header pressure set point due to increased most-open valve (FCV-1-3) position VP. The increased header pressure will clear out the obstruction in the cell #1 diffuser and as soon as the diffuser is unclogged, FIC-1 will bring FCV-1-3 to its normal position, bringing the header pressure set point back to normal value.

The most-open valve algorithm is allowed to quickly and automatically fix the clogged diffuser situation, which otherwise would have remained unnoticed for a long period leading to process upset. A properly designed and tuned most-open valve algorithm allows all air flow control valves to find a natural equilibrium. Generally slight changes in the most-open valve position will cause all other valves to slightly open or close as a result of the algorithm using Equation 8.1(2).

Referring to Figure 8.1w, the most-open valve algorithm (MOV-C) is resident in the main PLC. It receives the valve position signals from all the valves, determines the most-open valve position, and generates a header pressure set point value using Equation 8.1(2). A dead band can be provided on the pressure set point that will not allow change in pressure set point unless the required change exceeds the dead band.

The header pressure PID control algorithm is also resident in the main PLC and generates output to control the blower capacities to maintain the pressure set point. The blower control is described in the next paragraph.

Control of Blowers to Maintain the Header Pressure Figure 8.1w shows three blowers (*B*-1, *B*-2, and *B*-3) connected to the common header part of the aeration system. The blowers shown in this example are single-stage centrifugal blowers. These blowers can be either multistage with inlet throttling valve or single-stage single vane (IGV) or dual vane blowers. Each blower has a PLC-based local control panel (LCP-1, -2, or -3) that controls the blower operation.

The local control panel receives air temperature and inlet vane position analog signals and generates analog outputs for controlling the inlet guide vanes and the blow-off valve. The local panel also generates digital outputs (dry contacts) to start/stop the blower, interfaced with motor starter and open/close outputs for control of blower discharge valve (FV-1-2, -2-2, -3-2). Each panel receives digital inputs (contact closure) from the motor starter for ON status and the open/close status from discharge valve. All safety control logic including surge control is part of the local panel controls.

As discussed in the previous paragraph the algorithm for header pressure control using the most-open valve algorithm is programmed in the main control panel PLC. The blower sequencer is also programmed in the main PLC. The local panel PLCs (PLC-1, PLC-2, or PLC-3) receive blower start/stop commands, IGV position commands, and header pressure values from the main PLC via data link.

The sequencer logic decides when to start or stop the next blower, and the start/stop commands are relayed to the local panels via the data link. As described in the earlier section Parallel and Cascade Control of Blowers, either parallel or cascade control strategy is programmed in the main PLC, depending on whether the blowers are multistage or single or dual vane single-stage blowers.

Using the parallel or cascade control strategy, the sequencer issues commands for starting or stopping the next blower, and the header pressure control algorithm either modulates multiple blower IGVs (parallel control) simultaneously or single blower IGV one at a time (cascade control), depending upon the type of blower system selected. Both parallel and cascade control strategies have been described in the earlier section Parallel and Cascade Control of Blowers.

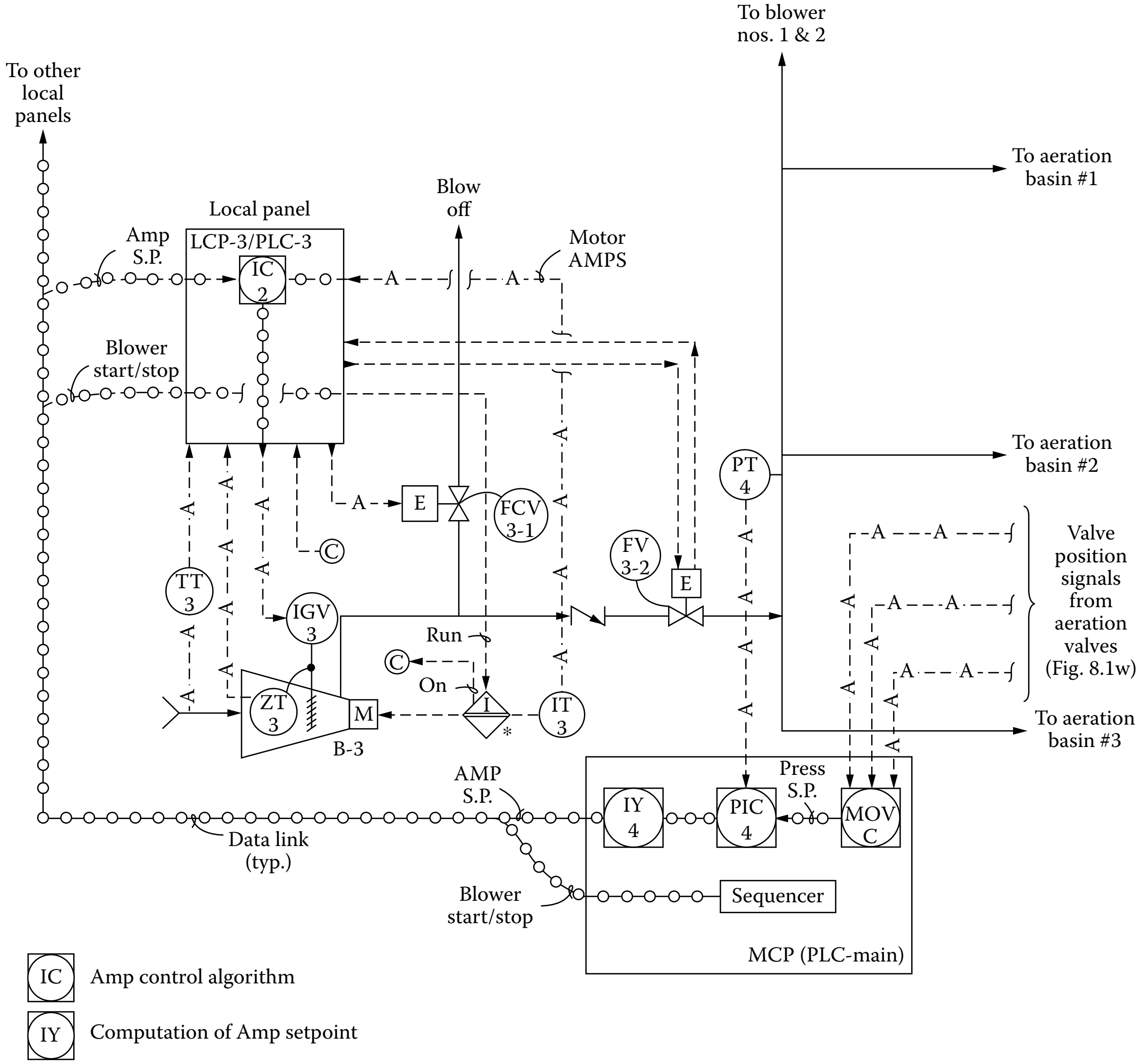


FIG. 8.1y
Controlling the header pressure by throttling the current to the blower motor.

The surge control strategy for each blower is programmed in the local PLC, part of the dedicated local control panel for each blower. The control strategies for surge control for different types of blowers were discussed in the earlier subsection Surge Control of Blowers. The local PLC logic includes a look-up table that provides the desired blower ampere draw set point that is required to avoid surge conditions for a given inlet guide vane position. The ampere controller algorithm looks at the measured blower current and compares this value with the desirable set point.

If the measured current is less than the desired set point, the algorithm generates an output to open the blow-off valve to bring the motor current draw to the desired value that is within the safe operating region. Thus, in this example, blower ampere draw is controlled to prevent the blower from going into surge condition.

Throttling Motor Current to Maintain Header Pressure In some applications, the inlet guide vanes for the single-stage blowers, or the inlet valve in the case of multistage blowers, is modulated to maintain a blower motor current set point. Referring to Figure 8.1y, the output of the pressure controller PIC-4 is converted into a blower current (amperes) set point using a linear equation, using the computation block IY-4. The computed ampere set point is transmitted to the local PLC panel (LCP-3), via the data link. The amp controller (IC-3) is programmed in the local individual blower panel PLC.

The current (amp) controller receives the measured value of motor current, compares it to the desired amp set point, and generates an output to modulate the inlet guide vane or valve to bring the motor current to the set point. The other control loops for aeration basin flow control, DO control, and most-open valve control strategies remain the same as shown in Figure 8.1w.

NITROGEN PROFILE ANALYZERS FOR WASTEWATER

The classical method of aeration control has been to maintain a desired DO concentration set point using DO measurement and control algorithms discussed before in this section. One problem with this approach is that the DO set point is usually set very conservatively to ensure excess of DO in the wastewater. Maintaining surplus DO in the aeration basin effluent to an arbitrary high set point is a convenient way to ensure that all ammonia has been converted to oxidized nitrogen, i.e., the wastewater has been fully "nitrified."

The difference between the oxygen required to achieve process objectives (full nitrification) and the actual oxygen delivered to meet a DO set point represents the potential energy savings that can be realized by measuring ammonia, nitrate, and nitrite concentrations on-line and using these values to control the aeration process. The nitrification process is a two-step process involving two types of microorganisms. The first type converts ammonia into nitrite (NO_2), and the second type converts the nitrites into nitrates (NO_3). Complete nitrification is achieved when all ammonia is converted into nitrates.

If the objective of the aeration process is full nitrification, the aeration energy costs can be optimized by direct measurement of ammonia and nitrates at several points and using this data to control the aeration process. Higher ammonia and lower nitrate concentrations indicate greater oxygen demand, and lower ammonia and higher nitrate concentrations indicate lower aeration demand.

The aeration process can be either manually or automatically controlled using on-line ammonia/nitrate concentration monitoring to achieve complete nitrification. This approach, in some cases, can lead to optimum aeration and reduced energy costs as compared to classical strategies based on DO measurement and control. DO measurement can still be used as a check on the overall success of the aeration process. The technique of measurement of ammonia, nitrite, and nitrate concentrations at various sample points is called "nitrogen profiling."

Nitrogen Profile Analyzers

The nitrogen profile analyzers receive filtered wastewater samples from multiple sampling points in a sequential manner and perform UV spectrum analysis to determine ammonia, nitrite, and nitrate concentrations in each sample. Typically, a single high-capacity pump is used to draw samples from multiple sampling points. Each sample point is controlled by a motor-operated ball valve on the suction side of the pump, controlled via signals from the sequencer, part of the analyzer. Sample obtained from each sampling point is filtered and placed in an accumulator.

Sample from the accumulator is transferred to a flow cell using a small pump part of the analyzer. The analyzer uses reagents and UV spectrum analysis to determine the ammonia, nitrite, and nitrate concentrations. The sequencer automatically controls the sample collection and analysis of each sample. Either analog 4–20 mA signals or digital data via Modbus communication link are available, corresponding to each parameter. Figure 8.1z shows a simplified block diagram of a typical nitrogen profile analyzer.

Mechanical Aerator Control Application

Figure 8.1aa shows a typical installation of mechanical aerators in the aeration basin. The influent flow is split into the East and West basins. Each basin is divided into four cells. Cell #1, #3, #5, and #7 include dual-speed 125/81 hp aerators. Cell #2 and #6 include 200 hp adjustable-speed aerators, and cell #4 and #8 have 75 hp adjustable-speed aerators.

A single nitrogen profile analyzer is used for on-line measurement/display of ammonia, nitrite, and nitrate concentrations at two locations in each basin (East and West). The nitrogen profile data is transmitted to the plant SCADA system and displayed on the graphic screens. Plant operators look at the nitrogen profiles to make adjustments to the aerator operation. For example, at minimum ammonia and maximum nitrate levels, all aerators are running at low speeds. As the ammonia levels rise and nitrate levels fall, the two dual-speed aerators #1 and #5 are switched to 100% speed.

If the ammonia levels still keep rising, the speeds of 200 hp aerators #2 and #6 are gradually increased to meet the oxygen demand. The last two aerators #4 and #8 are ramped up if, in a highly unlikely event, the South sampling point (located close to aerators #4 and #8) shows higher ammonia concentrations. Most of the time the North sampling point is used to validate the results of the control actions. The aerators are ramped down in reverse sequence on falling ammonia levels.

As an additional check, DO levels are measured in aeration cells #4 and #8 to ensure adequate DO level. The plant operators indicate that before installation of the nitrogen profile analyzer, their target DO levels in the basin effluent were 1.0 mg/l. However, following the installation of nitrogen profile analyzer and using the nitrogen profile for aeration control, the DO levels dropped to 0.3 mg/l, translating in energy savings while ensuring complete nitrification.

Diffused Aeration Control Application

A nitrogen profile analyzer can also be used for diffused aeration control. Referring to Figure 8.1w, the DO control loop can be replaced with an ammonia or nitrate control loop. Operator could select either ammonia or nitrate control mode. The DO controllers (AIC-1, -2, -3) are replaced with ammonia or nitrate controllers. The output of ammonia/nitrate controller is used to adjust the set point of the flow control loop that modulates the air control valve to each aeration cell. A single analyzer can be used to analyze samples from different aeration cells.

When ammonia concentration is chosen for control, the controller compares ammonia concentration in the sample with the set point and increases the air flow controller

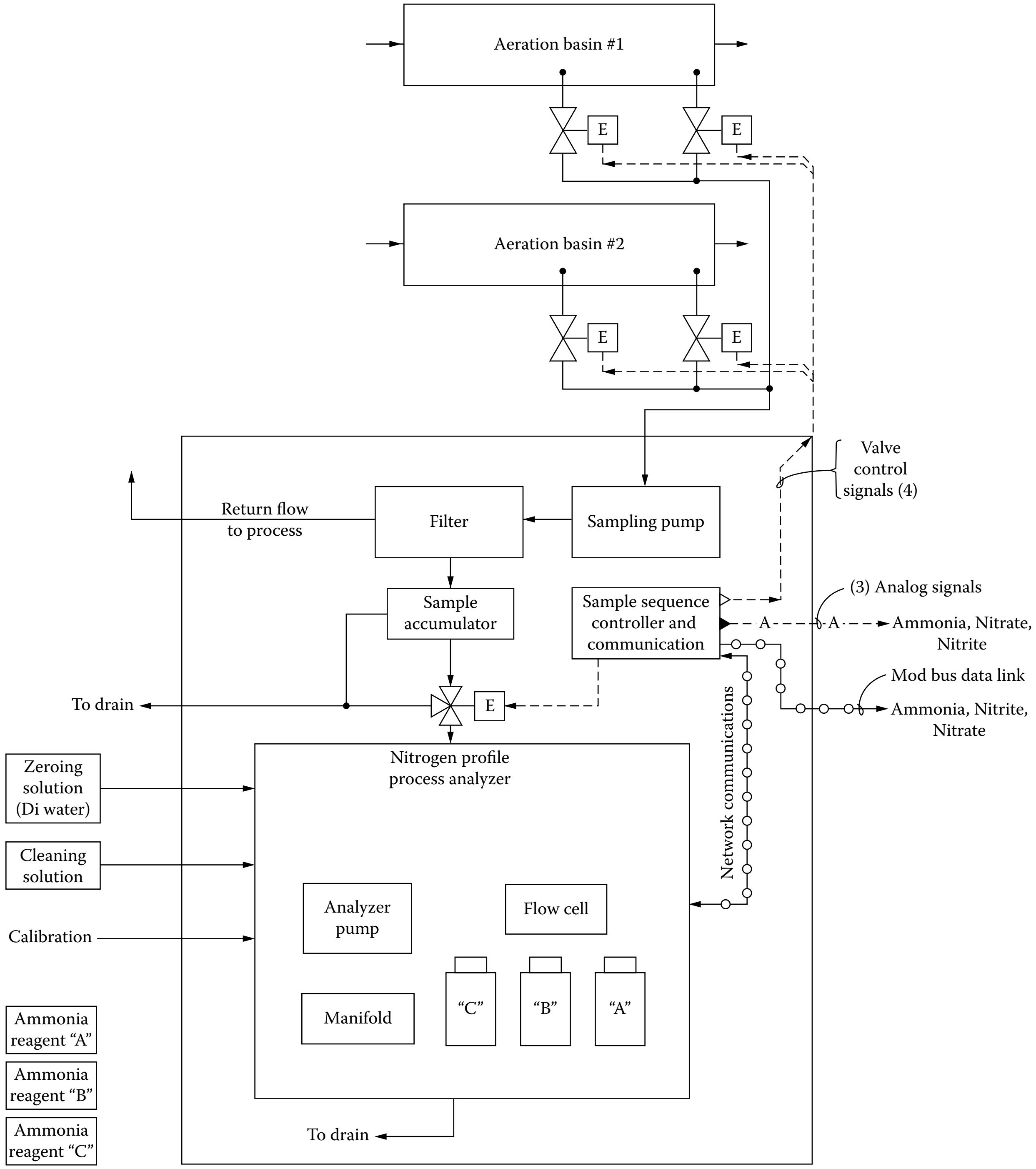
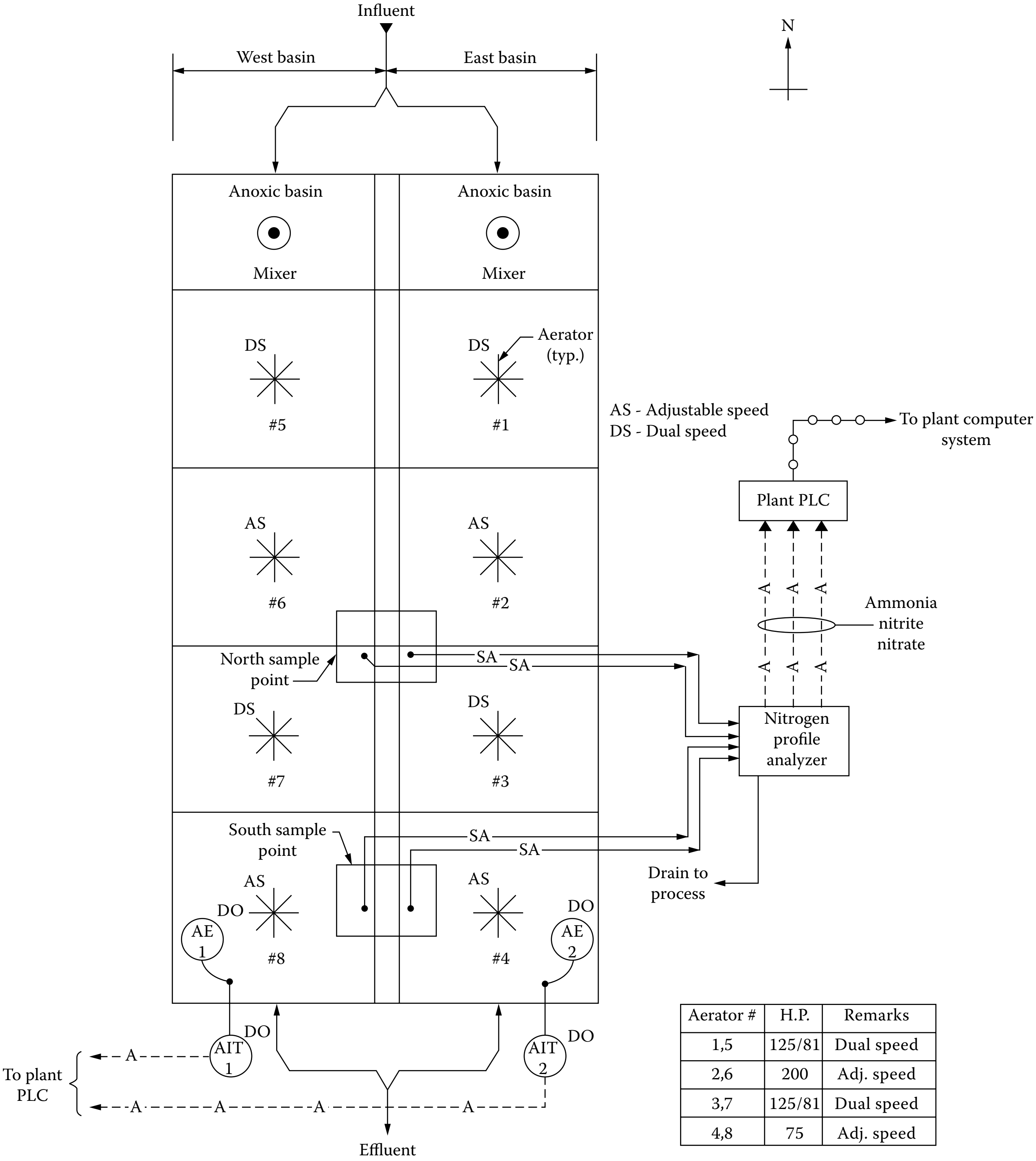


FIG. 8.1z
A simplified block diagram of a nitrogen profile analyzer.

(FIC-1, -2,- 3) set point as the ammonia concentration increases above set point, and reduces the air flow set point as the ammonia concentration starts dropping below the set point.

If nitrate concentration is chosen for control then the action of the nitrogen controller is opposite to that of ammonia control, i.e., the set point of air flow controller is reduced when the nitrogen concentration is above the set point and the set point of the air flow controller is increased when the nitrate concentration is below the desired set point. It is possible that ammonia control mode may work better in one cell while nitrate control mode may work better in another cell (Figure 8.1aa).

The most-open valve control strategy for header pressure control and the individual blower control is similar to that described in the earlier subsection Aeration System Control and shown in Figures 8.1w and 8.1y. Although nitrogen profile



Aerator #	H.P.	Remarks
1,5	125/81	Dual speed
2,6	200	Adj. speed
3,7	125/81	Dual speed
4,8	75	Adj. speed

FIG. 8.1aa
Aeration basin with mechanical aerators, provided with a nitrogen profile analyzer.

technology is gaining support among plant operators, its cost and additional maintenance should be analyzed before making a final decision.

Ultimately, the decision of using nitrogen profiling or DO control will largely depend upon the size of the plant, cost analysis, process objectives, and the comfort level of the operators.

ACKNOWLEDGMENTS

The help and support of the following colleagues are gratefully acknowledged:

Tim Ptak, Senior Technologist, CH2M HILL
Steven Byce, Chief Operator, Gainesville Regional Utilities

Allesandro Lequio, Turblex Inc.
Bernie Beemster, President, Applied Spectrometry Associates

References

1. Flanagan, M. J., and Bracken, B. D., EPA –600/2-77-032, "Design Procedures for Dissolved Oxygen Control of Activated Sludge Processes," Cincinnati, OH: U.S. EPA, 1977.
2. Hill, R., et al., *Automated Process Control Strategies,* Alexandria, VA: Water Environment Federation, 1997.
3. Mace, G. R., and Mathews, S., "Upgrade of Aeration Diffusers, Blowers, and Instrumentation Saves upwards of 50% Power: Case Study of two wastewater Treatment Plants," South Carolina Environmental Conference, 1999.
4. Vinton, R. H., and Mace, G. R., "Most-Open Valve Control and Cascade Control of Multiple Compressors Improve Aeration Efficiently and Cut Costs," Water Environment Federation, National Meeting, October 1997.
5. "Advantages of Cascade vs. Parallel Control," Turblex Publication No. 44.
6. Mace, G. R., "High Efficiency Single-Stage Air Compressors and Systems Instrumentation Reduce Power in the Biological Wastewater Aeration Process," Energy Efficiency Forum, Denver, CO, August–September 1998.
7. Pelletier, R. A. and Sloan, D. S., "Orlando's Nitrogen Profile Program: A New Approach to Activated Sludge Process Control," 72nd Annual Water Environment Federation Conference and Exposition, New Orleans, LA, October 1999.
8. "Project Report and Data Summary, Evaluation of On-Line Ammonia, Nitrite, and Nitrate Analysis for Process Control and Energy Management," Applied Spectrometry Associates, April 1998.
9. Application Summary No. 96, Nitrification Process Control, Applied Spectrometry Associates, Inc.
10. Application Summary No. 114, Aeration Control, Applied Spectrometry Associates, Inc.

8.2 Airhandler and Building Conditioning Controls

B. G. LIPTÁK (1985, 1995, 2005)

INTRODUCTION

The main components of HVAC control systems include 1) the various comfort sensors, such as thermostats and humidistats (Section 7.9 in Chapter 7) and pressure sensors (Chapter 5 in Volume 1), 2) the control systems for heat and coolant supply systems, the boilers (Section 8.6), chillers (Sections 8.12 and 8.13), and the cooling towers (8.16 and 8.17), 3) the air and water transportation controls, including the fans and blowers (Section 8.25) and pumping stations (Section 8.34), and 4) the final control elements, including the dampers (Section 7.1), control valves (Chapter 6), and variable-speed drives (Sections 7.10). For more information on the above topics the reader is referred to the noted sections.

This section will concentrate on the control and optimization of the total space conditioning system. This will be approached by first discussing the process being controlled and its various operating modes as the seasons change. Once the "personality" of the process has been described, the control of the various comfort-related variables (temperature, humidity, and air quality) will be discussed. The emphasis will be placed on systems in which air is the final carrier of heat or cooling into the conditioned spaces, although brief mention will also be made of the more traditional, but still used, water-based systems.

In the second half of this section, the emphasis will be on the optimization of the total process by such methods as making the buildings self-heating and by eliminating the chimney effects.

THE AIRHANDLER

The airhandler is the basic unit operation of space conditioning. It is used to keep occupied spaces comfortable (Figure 8.2a) or unoccupied spaces at desired levels of temperature and humidity. In addition to supplying or removing heat or humidity from the conditioned space, the airhandler also provides ventilation and fresh air makeup. Depending on the type of space involved, from 75,000–300,000 BTU/year (19,000–76,000 cal/year) are required to condition 1 ft^2 (0.092 m^2) of office space. Depending on the energy sources used, this corresponds to a yearly operating cost of a few dollars per square foot of floor space.

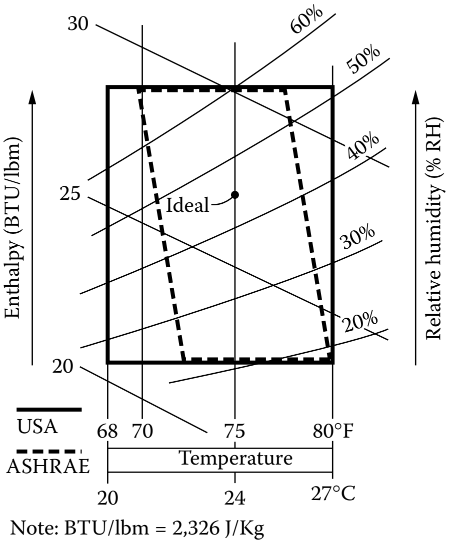
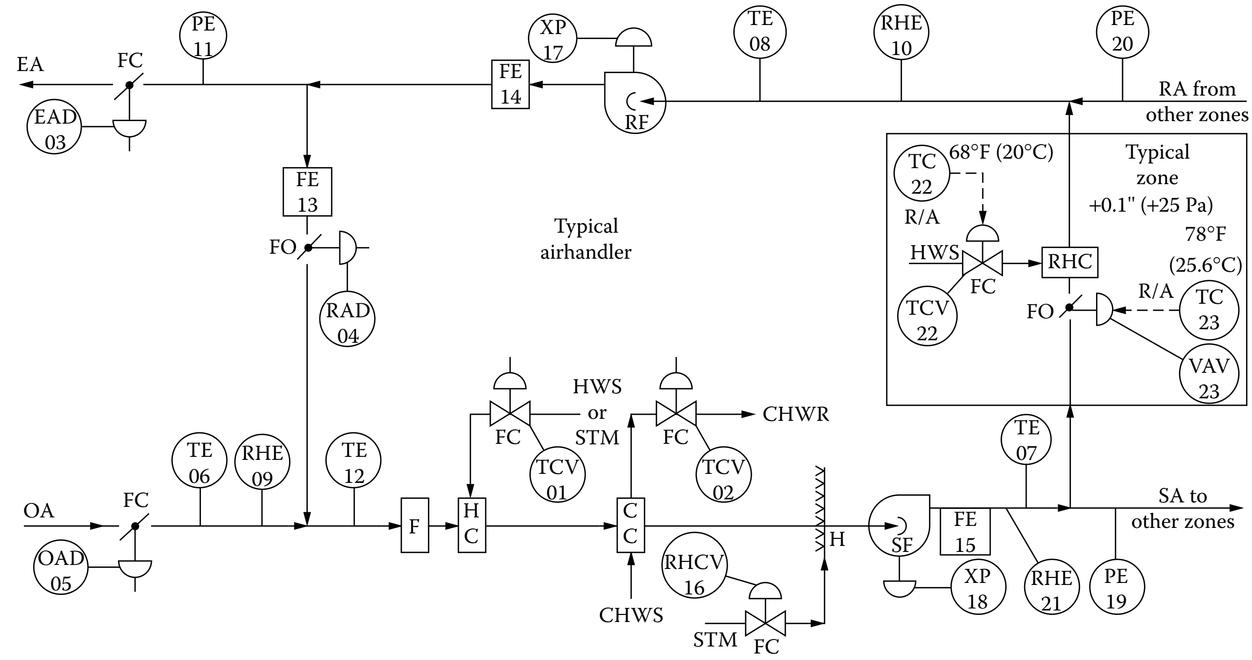


FIG. 8.2a
"Comfort zones" are defined in terms of temperature and humidity.[1]

Whereas other unit operations have benefited substantially from the advances in process control, airhandlers have not. Airhandlers today are frequently controlled the same way as they were 20 or 30 years ago. For this reason, airhandler optimization can result in much greater percentages of savings than can the optimization of almost any other unit operation. Optimization can sometimes cut the cost of airhandler operation in half — a savings that can seldom be achieved in any other type of unit operation.

Some of the optimization goals and strategies include the following:

- Let the building heat itself
- Use free cooling or free dying
- Benefit from gap control or zero energy band (ZEB)
- Eliminate chimney effect
- Optimize start-up timing
- Optimize air makeup (CO$_2$)
- Optimize supply air temperature
- Minimize fan energy use
- Automate the selection of operating modes
- Minimize reheat
- Automate balancing of air distribution



CC = Cooling coil
CHWR = Chilled water return
EA = Exhaust air
EAD = Exhaust air damper
F = Filter
FC = Fail closed
FE = Flow element

FO = Fail open
H = Humidifier
HC = Heating coil
HWS = Hot water supply
OA = Outside air
OAD = Outside air damper
PE = Pressure element

RA = Return air
RAD = Return air damper
RF = Return fan
RHC = Reheat coil
RHCV = Relative humidity control valve
RHE = Relative humidity element
SA = Supply air

SF = Supply fan
STM = Steam
TCV = Temperature control valve
TE = Temperature element
VAV = Variable air volume damper
XP = Positioner for fan volume control, such as a blade pitch positioner

FIG. 8.2b
A typical major airhandler has these components and controls.

Airhandler Components

The purpose of heating, ventilation, and air conditioning (HVAC) controls is to provide comfort in laboratories, cleanrooms, warehouses, offices, and manufacturing spaces. Supply air is the means of providing comfort in the conditioned zone. The air supplied to each zone must provide heating or cooling, raise or lower humidity, and provide air refreshment. To satisfy these requirements, it is necessary to control the temperature, humidity, and fresh-air ratio in the supply air.

Figure 8.2b illustrates the main components of an airhandler. The term *airhandler* refers to the total system, including fans, heat-exchanger coils, dampers, ducts, and instruments. The system operates as follows: Outside air is admitted by the outside air damper (OAD-05) and is then mixed with the return air from the return air damper (RAD-04). The resulting mixed air is filtered (F), heated (HC) or cooled (CC), and humidified (H) or dehumidified (CC) as required. The resulting supply air is then transported to the conditioned zones (groups of offices) by the variable-volume supply fan station. Variable volume means that the air flow rate generated by the fan(s) is variable.

In each zone, the variable air volume damper (VAV-23) determines the amount of air required, and the reheat coil (RHC) adjusts the air temperature as needed. The return air from the zones is transported by the variable-volume return-air fan station. If the amount of available return air exceeds the demand for it, the excess air is exhausted by the exhaust air damper (EAD-03). The conditioned spaces are typically pressurized to about 0.1 in. H_2O (25 Pa), relative to the barometric pressure on the outside. This pressurization results in some air leakage through the walls and windows, which varies with the quality of construction. Therefore, the air balance around the system is:

$$\text{OA} = \text{EA} + \text{pressurization loss} \qquad \textbf{8.2(1)}$$

Under "normal" operation, the airhandler operates with about 10% outside air. In the "purge" or "free cooling" modes, RAD is closed, OAD is fully open, and the airhandler operates with 100% outside air.

As can be seen, the HVAC process is rather simple. Its process material is clean air, its utility is water or steam, and its overall system behavior is slow, stable, and forgiving. For precisely these reasons, it is possible to obtain acceptable HVAC performance using inferior-quality instruments that are configured into poorly designed loops. Yet, there is an advantage in applying state-of-the-art process control to the HVAC process, because it can provide a drastic reduction in operating costs, attributable to increased efficiency of operation. Some of the more efficient control concepts are described in the paragraphs below.

Operating Mode Selection

The correct identification and timing of the various operating modes can contribute to the optimization of the building. The *normal* operating modes include start-up, occupied, night, and purge.

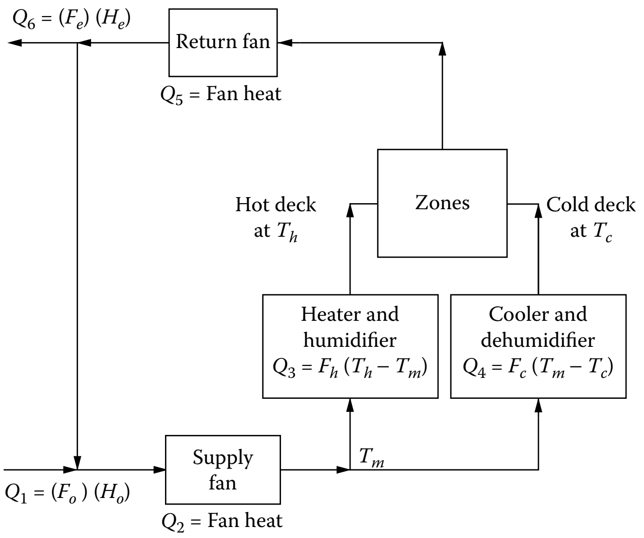
Optimizing the time of *start-up* will guarantee that the minimum required cost is invested in getting the building ready for occupancy. This is done by automatically calculating the amount of heat that needs to be transferred and dividing it according to the capacity of the start-up equipment. A computer-optimized control system will serve to initiate the unoccupied (night) mode of operation; it will also recognize weekends and holidays and, in general, provide a flexible means of time-of-day controls.

The *purge* mode is another convenient tool of optimization. Whenever the outside air is preferred to the return air, the building is automatically purged. In this way, "free cooling" can be obtained on dry summer mornings, or "free heating" can be provided on warm winter afternoons. Purging is the equivalent of opening the windows in a home. In computer-optimized buildings, an added potential is to use the building structure as a means of heat (or coolant) storage. In this case, the purge mode can be automatically initiated during cold nights prior to hot summer days, thereby bringing the building temperature down and storing some free cooling in the building structure.

Summer/Winter Mode Reevaluation Another important mode selection involves switching from summer to winter mode and vice versa. Conventional systems are switched according to the calendar, whereas optimized ones recognize that there are summer-like days in the winter and winter-like hours during summer days. Seasonal mode switching is therefore totally inadequate.

Optimized building operation can be provided only by making the summer/winter selection on an enthalpy basis: If heat needs to be added, it is "winter"; if heat needs to be removed, it is "summer," regardless of the calendar. In those airhandlers that serve a variety of zones, it is essential to first determine if the unit is in a "net" cooling (summer) or "net" heating (winter) mode before the control system can decide if free cooling (or free heating) by outside air can be used to advantage.

Figure 8.2c illustrates the heat balance evaluation that is required to determine the prevailing overall mode of operation. This type of heat balance calculation, which must be reevaluated every 15 to 30 minutes, can be implemented only through the use of computers.



F_c = Cold deck flow
H_o = Outside air enthalpy
F_e = Exhaust air flow
T_c = Cold deck temperature
F_h = Hot deck flow
T_h = Hot deck temperature
H_e = Exhaust air enthalpy
T_m = Mixed air temperature
Net airhandler load = $Q_0 = Q_1 + Q_2 + Q_3 - Q_4 + Q_5 - Q_6$

FIG. 8.2c
When the net airhandler load is negative, summer mode is required; when it is positive, winter mode is required.

Emergency Mode In addition to the above operating modes, the airhandler can also be placed in an *emergency* mode, if fire, smoke, freezing temperature, or pressure conditions require it. Table 8.2d lists the status of each fan, damper, and valve in each of the operating modes. In a computer-optimized control system, both the mode selection and the setting of the actuated devices is done automatically.

When a smoke or fire condition is detected by sensors S/F-4 or S/F-8 in Figure 8.2e, the fans stop, the OADs and RADs close, the EAD opens, and an alarm is actuated. The operator can switch the airhandler into its purge mode, so that the fans are started, OAD and EAD are opened, and RAD is closed. If the smoke/fire emergency requires, the fire command panel (FC in Figure 8.2e) can be used by firefighters. From this panel, the fire chief can operate all fans and dampers as needed for safe and orderly evacuation and protection of the building.

In another emergency condition, a freezestat switch on one of the water coils is actuated. These switches are usually set at approximately 35°F (1.5°C) and serve to protect from coil damage resulting from freeze-ups. Multistage freezestat units might operate as follows:

- At 38°F (3°C): close OAD
- At 36°F (2°C): fully open water valve
- At 35°F (1.5°C): stop fan

If single-stage freezestats are used, they will stop the fan, close the OAD, and activate an alarm.

Yet another type of emergency is signaled by excessive pressures in the ductwork on the suction or discharge sides of the fans, resulting from operation against closed dampers

TABLE 8.2d
The Status of Various Actuated Devices during Various Operating Modes

Operating Mode or Emergency Condition	*Supply Fan*	*Return Fan*	*Outside Air Damper*	*Exhaust Air Damper*	*Return Air Damper*	*Coil Control Valves*	*Alarm*
Off	—	—	C	C	O	C	—
On	On	On	◀—— Modulating ——▶				—
Warm-up	On	On	C	C	O	O(HC)	—
Cool-down	On	On	C	C	O	O(CC)	—
Night	◀—— Cycled to maintain required nighttime temperature ——▶						
Purge	On	On	O	O	C	Modulating	—
PSH-2	—	Off	—	C	—	—	Yes
PSL-3	—	Off	—	C	—	—	Yes
S/F-4	Off	Off	C	O	C	C	Yes
TSL-5	Off	—	C	—	O	C	Yes
PSL-6	Off	—	C	—	O	—	Yes
PSH-7	Off	—	C	—	O	—	Yes
S/F-8	Off	Off	C	O	C	C	Yes

or from other equipment failures. When this happens, the associated fan is stopped and an alarm is actuated.

Fan Controls

The standard fan controls are shown in Figure 8.2f. Each zone shown in Figure 8.2b is supplied with air through a thermostat-modulated damper, also called a variable air volume box (VAV-23).

The VAV box openings in the various zones determine the total demand for supply air. The pressure in the supply air (SA) distribution header is controlled by PIC-19, which modulates the supply air fan station to match the demand (Figure 8.2f). When the PIC-19 output has increased the fan capacity to its maximum, PSH-19 actuates and starts an additional fan. Inversely, as the demand for supply air drops, FSL-15 will stop one fan unit whenever the load can be met by fewer fans than the number in operation. The important point to remember is that in cycling fan stations, fan units are started on pressure and are stopped on flow control. The operating cost of such a fan station is 20–40% lower than if constant-volume fans with conventional controls were used (Figure 8.2g).

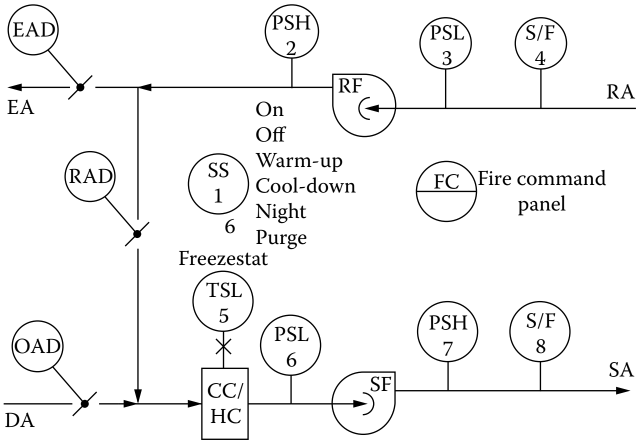


FIG. 8.2e
The safety and operating mode selection instruments used on an airhandler. Most abbreviations used on this figure have already been defined in connection with Figure 8.2b; S/F = smoke and fire detector, SS = selector switch, FC = fire command panel.

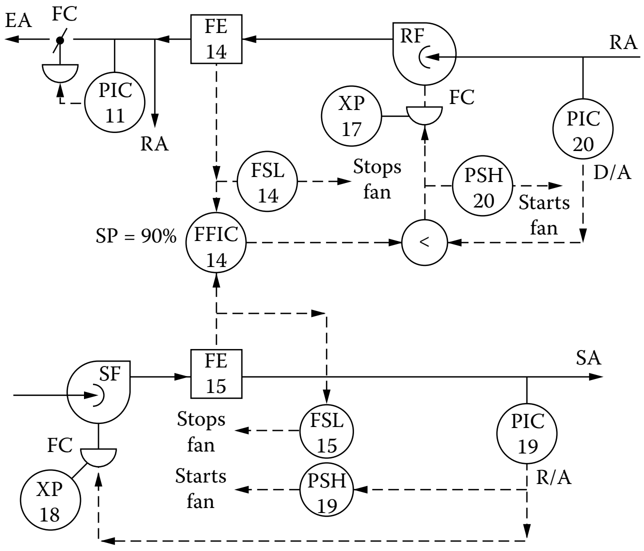


FIG. 8.2f
Variable-volume fan controls operate as shown here.

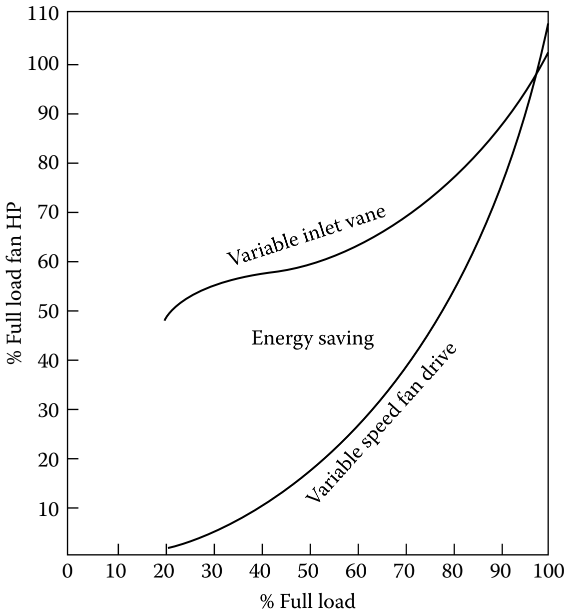


FIG. 8.2g
Using variable-speed fans can save significant amounts of energy. (Courtesy of Dana Corp.)

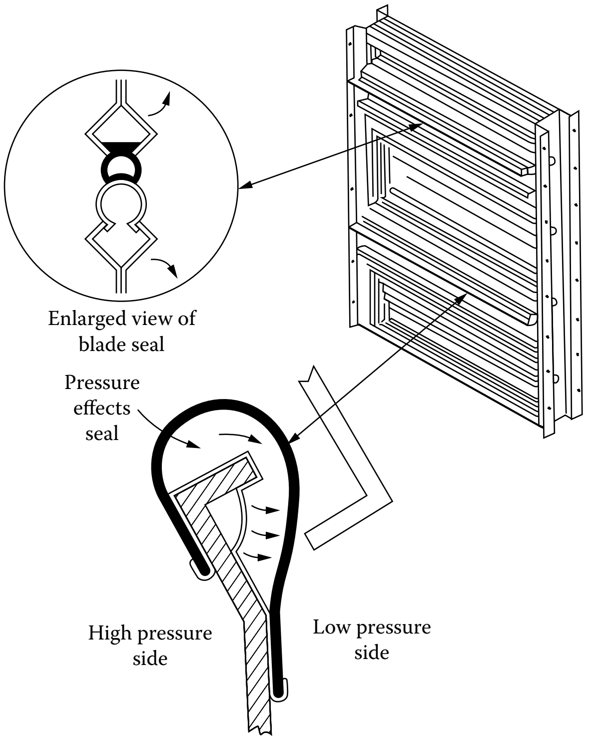


FIG. 8.2h
Low-leakage damper designs increase the efficiency of HVAC systems.

Because the conditioned zones are pressurized slightly, some of the conditioned air will leak into the atmosphere, creating pressurization loss. Being able to control the pressurization loss is one of the advantages of the control system described in Figure 8.2f. The flow ratio controller FFIC-14 is set at 90%, meaning that the return-air fan station is modulated to return 90% of the air supplied to the zones. Therefore, pressurization loss is controlled at 10%, which corresponds to the minimum fresh-air makeup requirement, resulting in a minimum-cost operation.

Because the conditioned zones represent a fairly large capacity, a change in supply air flow will not immediately result in a need for a corresponding change in the return air flow. Thus, PIC-20 (Figure 8.2f) is included in the system to prevent the flow-ratio controller from increasing the return air flow rate faster than required. This dynamic balancing eliminates cycling and protects against collapsing the ductwork under excessive vacuum. Closure of the exhaust-air damper by PIC-11 indicates that the control system is properly tuned and balanced and is operating at maximum efficiency. Under such conditions, the outside air admitted into the airhandler exactly matches the pressurization loss, and no return air is exhausted.

To maximize the benefits of such an efficient configuration, the dampers must be of tight shut-off design. When exposed to a pressure difference of 4 in. H_2O (996 Pa), a closed conventional damper will leak at a rate of approximately 50 cfm/ft^2 [15.2 (m^3/min)/m^2]. In the HVAC industry, a 5 cfm/ft^2 [1.52 (m^3/min)/m^2] leakage rate is considered to represent a tight shut-off design. Actually, it is cost-effective to install tight shut-off dampers with leakage rates of less than 0.5 cfm/ft^2 [0.15 (m^3/min)/m^2], because the resulting savings over the life of the buildings will be much greater than the increase in initial investment for better dampers (Figure 8.2h).

Dampers and Pressure Distribution In order for dampers to give good control, a fair amount of pressure drop should be assigned to them. They should be sized for a ΔP of about 10% of the total system drop. On the other hand, excessive damper drops should also be avoided, because they will increase the operating costs of the fans. A good sizing basis for outside and return air fans is to size them for 1500 fpm (457 m/min) velocity at maximum flow.

In locations where two air streams are mixed, such as when outside and return airs are ratioed (RAD-04 and OAD-05 in Figure 8.2b), it is important that the damper ΔP be relatively constant as the ratio is varied. Figure 8.2i shows that parallel blade dampers give a superior performance in this service.

Figure 8.2j illustrates the pressure levels in the various portions of typical airhandlers. It can be seen that the kind of pressure drops that would be required by opposed blade dampers (Figure 8.2i) are simply not available. Therefore, if such dampers were installed, the airhandler would be starved for air (the dampers could not pass the design flow) whenever the ratio was near 50:50 (percent).

Figure 8.2j also shows that in traditional airhandlers more fan energy is used than necessary. This is because the return air fan is sized to generate the pressure needed to exhaust the

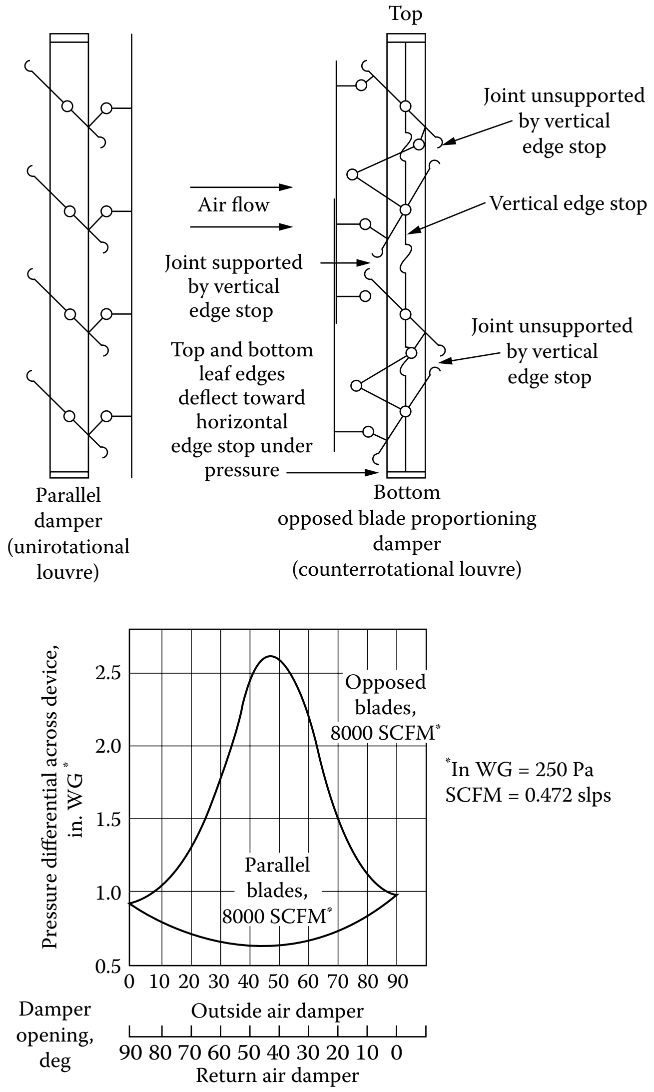


FIG. 8.2i
When outside and return air dampers are throttled to vary their ratio at constant total flow, the required pressure drop varies with damper design (see upper portion of figure). The lower portion of this figure shows the results of American Warming and Ventilating Co. tests (per AMCA Standard 500) of pressure drops across parallel-blade and opposed-blade outside and return air damper sets.[2]

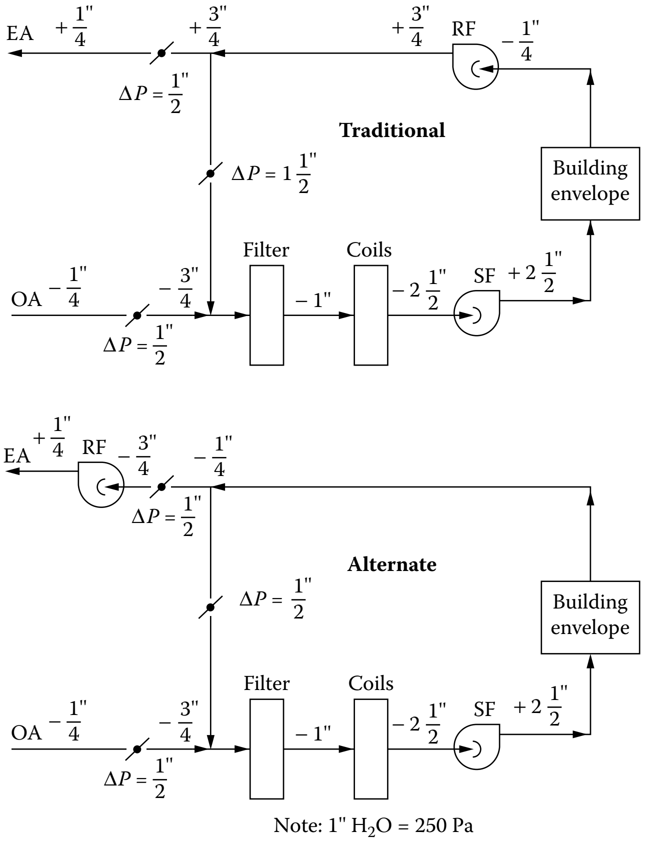


FIG. 8.2j
Damper pressure drops and the typical pressure levels in the various segments of airhandlers.

air from the building. A consequence of this is that the pressure drop of 1$^1/_2$ in. H_2O (375 Pa) across the return air damper is three times greater than what is necessary ($^1/_2$ in. H_2O, or 125 Pa).

The alternate system shown in the lower portion of Figure 8.2j eliminates this waste of fan energy. Here, only the supply fan (SF) operates continuously, which reduces the pressure drop across the return air damper to $^1/_2$ in. H_2O (125 Pa). The return fan (RF) is started only when air needs to be relieved, and its speed is varied to adjust the amount of air to be exhausted. Relocating RF also removes its heat input, which, in the traditional system, represents an added load on the cooling coil.

Temperature Controls

Space temperatures are controlled by thermostats. The traditional thermostat is a proportional-only controller (see also Section 2.2). The pressure of the output signal from a pneumatic "stat" is a near straight-line function of the measurement, described by the following relationship:

$$O = K_c(M - M_o) + O_o \qquad \textbf{8.2(2)}$$

where

O = output signal
K_c = proportional sensitivity (K_c can be fixed or adjustable, depending on the design)
M = measurement (temperature)
M_o = "normal" value of measurement, corresponding to the center of the throttling range
O_o = "normal" value of the output signal, corresponding to the center of the throttling range of the control valve (or damper)

Another term used to describe the sensitivity of thermostats is *throttling range*. As shown in Figure 8.2k, this term refers to the amount of temperature change that is required to change the thermostat output from its minimum to its

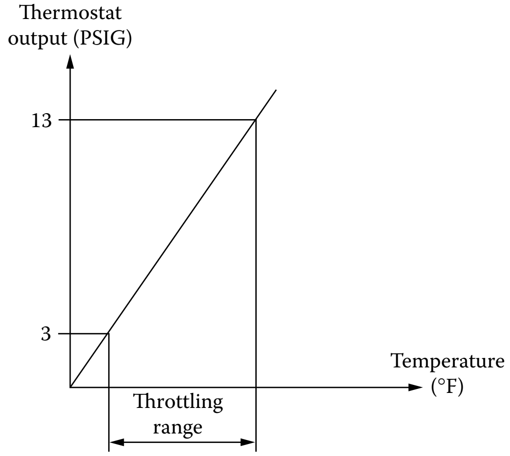


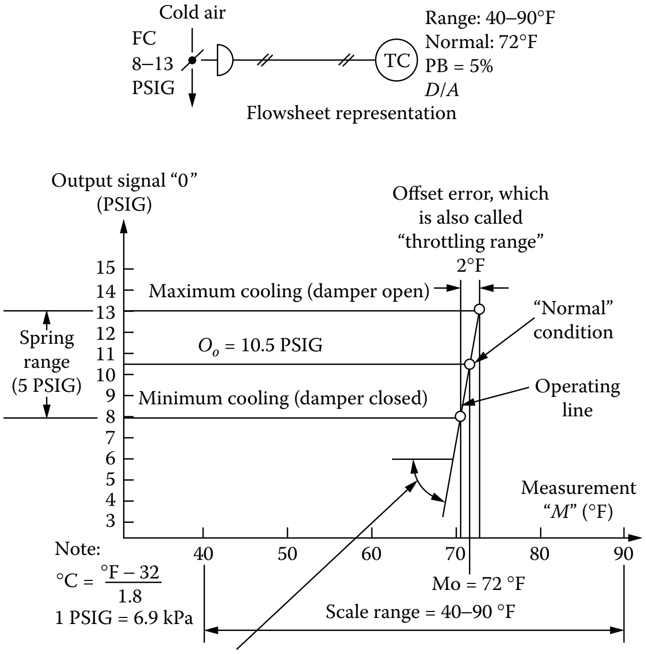
FIG. 8.2k
Throttling range can be defined as the temperature change required to change the thermostat output from its minimum to its maximum value.

maximum value, such as from 3 to 13 PSIG (21 to 90 kPa). The throttling range is usually adjustable from 2 to 10°F (1 to 5°C).

One important point to remember is that thermostats do not have set points, in the sense of having a predetermined temperature to which they would seek to return the controlled space. (Integral action must be added in order for a controller to be able to return the measured variable to a set point after a load change.) M_o does not represent a set point; it only identifies the space temperature that will cause the cooling damper in Figure 8.2l to be 50% open. This can be called a "normal" condition, because relative to this point the thermostat can both increase and decrease the cooling air flow rate as space temperature changes.

If the cooling load doubles, the damper will need to be fully open, which cannot take place until the controlled space temperature has risen to 73°F (23°C). As long as the cooling load remains that high, the space temperature must also stay up at the 73°F (23°C) value. Similarly, the only way this thermostat can reduce the opening of the cooling damper below 50% is to first allow the space temperature to drop below 72°F (23°C). Thus, thermostats have throttling ranges, not set points. If a throttling range is narrow enough, this gives the appearance that the controller is keeping the variable near the set point, when in fact the narrow range only allows the variable to drift within limits.

Special-Purpose Thermostats *Day-night, set-back,* or *dual room thermostats* will operate at different "normal" temperature values for day and night. They are provided with both a "day" and a "night" setting dial, and the change from day to night operation can be made automatic for a group of thermostats. The pneumatic day-night thermostat uses a two-pressure air supply system, the two pressures often being 13 and 17 PSIG (89.6 and 177 kPa) or 15 and 20 PSIG (103.35 and 137.8 kPa). Changing the pressure at a central point from one value to the other actuates switching devices in the thermostat and indexes them from day to night or vice versa.



Slope of operating line = Proportional sensitivity = 2.5 PSIG/°F = Proportional band = 5% = Gain = 20

FIG. 8.2l
A fixed proportional band thermostat has a fixed throttling range and no setpoint.

Supply air mains are often divided into two or more circuits so that switching can be accomplished in various areas of the building at different times. For example, a school building may have separate circuits for classrooms, offices and administrative areas, the auditorium, and the gymnasium and locker rooms. In some of the electric designs, dedicated clocks and switches are built into each thermostat.

The *heating–cooling* or *summer–winter thermostat* can have its action reversed and, if desired, can have its set point changed by means of indexing. This thermostat is used to actuate controlled devices, such as valves or dampers, that regulate a heating source at one time and a cooling source at another. It is often manually indexed in groups by a switch, or automatically by a thermostat that senses the temperature of the water supply, the outdoor temperature, or another suitable variable.

In the heating–cooling design, there are frequently two bimetallic elements, one being direct acting for the heating mode, the other being reverse acting for the cooling mode. The mode is switched automatically in response to a change in the air supply pressure, much as the day–night thermostats operate.

The *limited control range thermostat* usually limits the room temperature in the heating season to a maximum of 75°F (24°C), even if the occupant of the room has set the thermostat beyond these limits. This is done internally, without placing a physical stop on the setting knob.

A *slave or submaster thermostat* has its set point raised or lowered over a predetermined range, in accordance with variations in the output from a master controller. The master controller can be a thermostat, manual switch, pressure controller, or similar device. For example, a master thermostat measuring outdoor air temperature can be used to adjust a submaster thermostat controlling the water temperature in a heating system. Master–submaster combinations are sometimes designated as single-cascade action. When action is accomplished by a single thermostat having more than one measuring element, it is referred to as *compensated control.*

Multistage thermostats are designed to operate two or more final control elements in sequence.

A *wet-bulb thermostat* is often used for humidity control, as the difference between wet- and dry-bulb temperature is an indication of moisture content. A wick or other means for keeping the bulb wet and rapid air motion to ensure a true wet-bulb measurement are essential.

A *dew-point thermostat* is a device designed to control humidity on the basis of dew point temperatures.

A *smart thermostat* is usually a microprocessor-based unit with RTD-type or transistorized solid-state sensor. It is usually provided with its own dedicated memory and intelligence, and it can also be equipped with a communication link (over a shared data bus) to a central computer. Such units can minimize building operating costs by combining time-of-day controls with intelligent comfort gap selection and maximized self-heating.

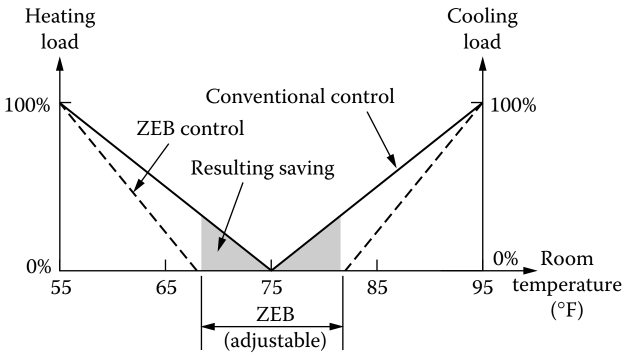


FIG. 8.2m
Zero energy band (ZEB) control is designed to save energy by not using any when the room is comfortable.

Zero Energy Band Control A recent addition to the available thermostat choices is the zero energy band design. The idea behind ZEB control is to conserve energy by not using any when the room is comfortable. As illustrated by Figure 8.2m, the conventional thermostat wastes energy by continuing to use it when the area's temperature is already comfortable. The "comfort gap" or "ZEB" on the thermostat is adjustable and can be varied to match the nature of the particular space.

ZEB control can be accomplished in one of two ways. The single set point and single output approach is illustrated on the left side of Figure 8.2n. Here the cooling valve fails closed and is shown to have an 8–11 PSIG (55–76 kPa) spring range, while the heating valve is selected to fail open and

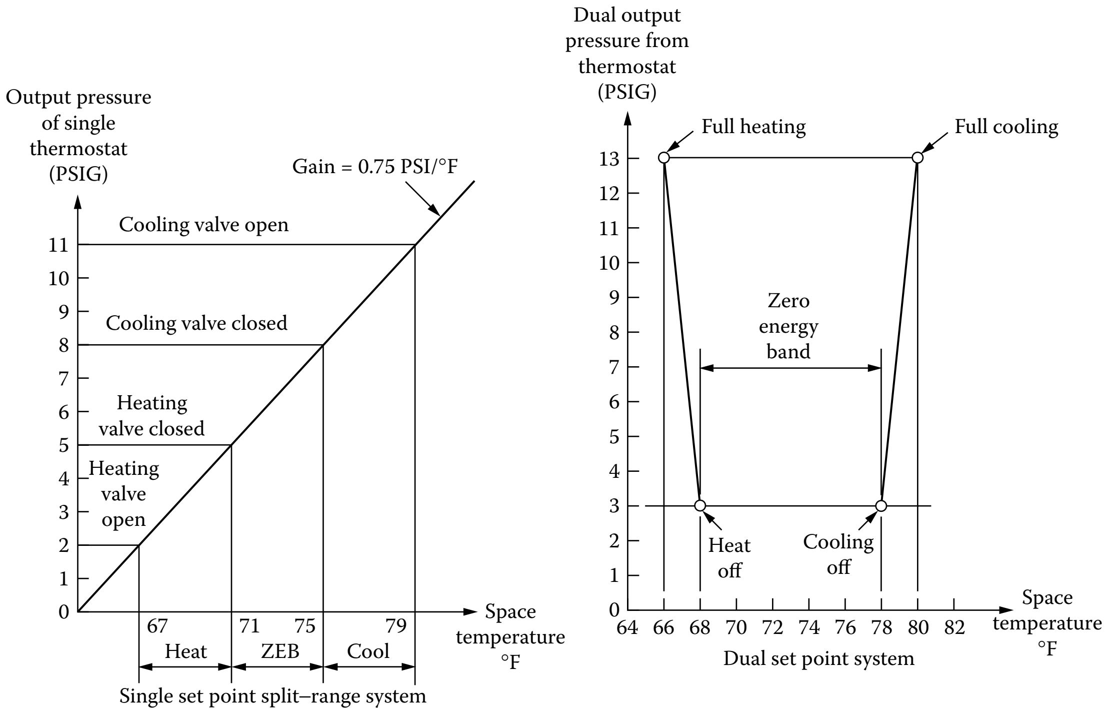


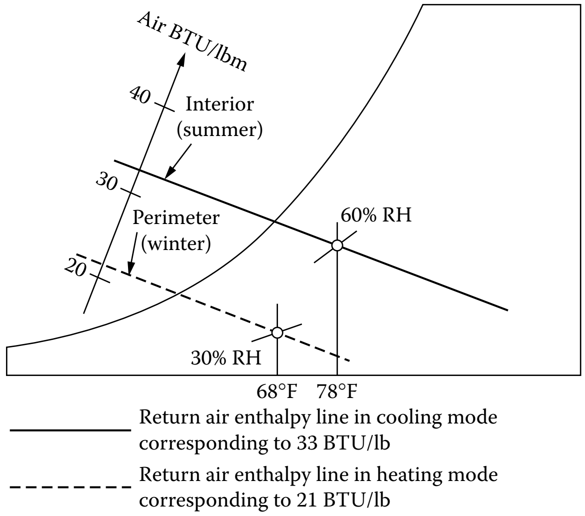
FIG. 8.2n
ZEB control schemes include the single setpoint split range approach (shown on left) and the dual set point approach (shown on right).

has a 2–5 PSIG (14–34 kPa) range. Therefore, between 5 and 8 PSIG (34 and 55 kPa), both valves are closed; no pay energy is expended while the thermostat output is within this range. The throttling range is usually adjustable from 5 to 25°F (3 to 13°C). Thus, if the ZEB is 30% of the throttling range, it can be varied from a gap size of 1.5°F (0.85°C) to 7.5°F (4.2°C) by changing the throttling range (or gain).

Although the split-range approach is a little less expensive than the dual set-point scheme (shown on the right of Figure 8.2n), it is also less flexible and more restrictive. The two basic limitations of the split-range approach are: 1) The gap width can be adjusted only by also changing the thermostat gain; maximum gap width is limited by the minimum gain setting of the unit. 2) The heating valve must fail open, which is undesirable in terms of energy conservation.

These limitations are removed when a dual set point, dual output thermostat is used. Here both valves can fail closed, and the bandwidth and the thermostat gain are independently adjustable. The gains of the heating and cooling thermostats are also independently adjustable. In Figure 8.2n, the heating thermostat is reverse-acting and the cooling thermostat is direct-acting.

The most recent advances in thermostat technology are the microprocessor-based units. These are programmable devices with memory capability. They can be monitored and reset by central computers, using telephone wires or other communication links. Microprocessor-based units can be supplied by continuously recharged backup batteries and accurate room temperature sensors. They can also operate without a host computer (in stand-alone mode). In this case, the user manually programs the thermostat to maintain various room temperatures as a function of the time of day and other considerations.



Note: BTU/lbm = 0.555 Kcal/Kg

FIG. 8.2o
The building can be made self-heating, because the return air can transport about 10 BTU/lbm to the perimeter spaces, which in the winter, because of the windows, do require heat.

Gap Control and the Self-Heating Building The winter and summer enthalpy settings of a building are illustrated in Figure 8.2o. The concept of gap control is simple: When comfort level in a zone is somewhere between acceptable limits, the use of “pay energy” is no longer justified. Allowing the zones to float between limits instead of maintaining them at arbitrarily fixed conditions can substantially reduce the operating cost of the building.

The savings come from two sources. First, there is a direct trade-off between the selected acceptable limits of discomfort and the yearly total of required degree days of heating and cooling. Second, there is the added side benefit that the building becomes self-heating during winter conditions. This occurs because during winter conditions, gap control automatically transfers the heat generated in the inside of the building to the perimeter areas, where heating is needed. This can result in long periods of building operation without the use of any pay energy.

Gap control can be looked upon as an override mode of control that is superimposed on the operation of the individual zone thermostats. As such, it can be implemented by all levels of automation, but the flexibility and ease of adjustment of the computerized systems make them superior in those applications in which the gap limits of the various zones are likely to change frequently.

Figure 8.2a illustrates the concept of comfort envelopes. Any combination of temperature and humidity conditions within such envelopes is considered to be comfortable. Therefore, as long as the space conditions fall within this envelope, there is no need to spend money or energy to change those conditions. This comfort gap is also referred to as zero energy band, meaning that if the space is within this band, no pay energy of any type will be used. This concept is very cost-effective.

When a zone of the airhandler in Figure 8.1b is within the comfort gap, its reheat coil is turned off and its VAV box is closed to the minimum flow required for air refreshment. When all the zones are inside the ZEB, the HW, CHW, and STM supplies to the airhandler are all closed and the fan is operated at minimum flow. When all other airhandlers are also within the ZEB, the pumping stations, chillers, cooling towers, and HW generators are also turned off.

With larger buildings that have interior spaces that are heat-generating even in the winter, ZEB control can make the building self-heating. Optimized control systems in operation today are transferring the interior heat to the perimeter without requiring any pay heat until the outside temperature drops below 10 to 20°F (2.3 to 6.8°C). In regions in which winter temperature does not drop below 10°F (–2.3°C), ZEB control can eliminate the need for pay heat altogether. In regions farther north, ZEB control can lower the yearly heating fuel bill by 30–50%.

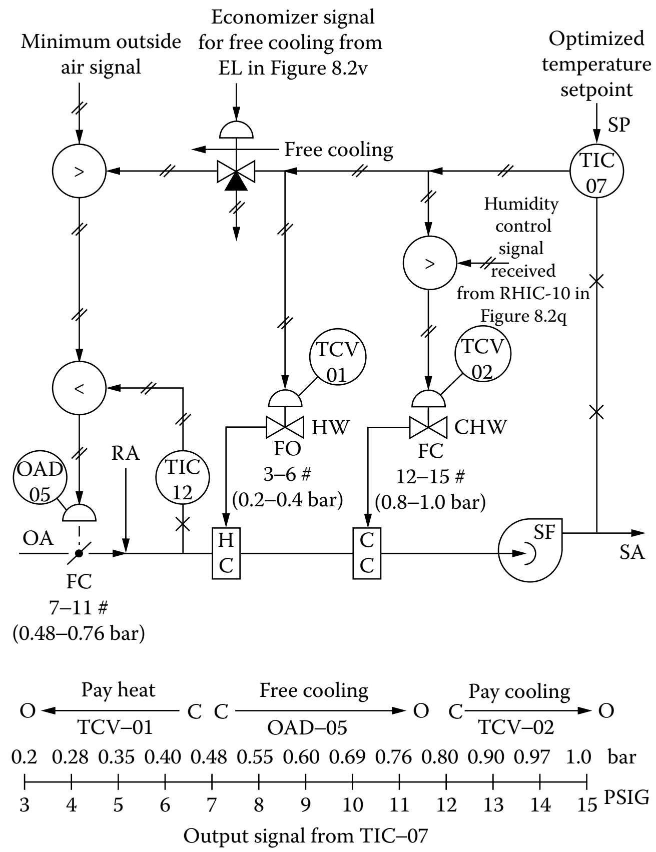


FIG. 8.2p
Illustration of a fully coordinated, pneumatic, split-range temperature control system. Such controls can reduce the yearly operating costs by more than 10%.[8]

Supply Air Temperature Control A substantial source of inefficiency in conventional HVAC control systems is the uncoordinated arrangement of temperature controllers. Two or three separate temperature control loops in series are not uncommon. For example, one of these uncoordinated controllers may be used to control the mixed air temperature, another to maintain supply (SA) temperature, and a third to control the zone-reheat coil. Such practice can result in simultaneous heating and cooling and, therefore, in unnecessary waste. Using a fully coordinated split-range temperature control system, such as that shown in Figure 8.2p, will reduce yearly operating costs by more than 10%.

In this control system, the SA temperature set point (set by the temperature controller, TIC-07) is continuously modulated to follow the load. The methods of finding the correct set point will be discussed under Optimizing Strategies. The loop automatically controls all heating or cooling modes. When the TIC-07 output signal is low — 3–6 PSIG (20.7–41.3 kPa) — heating is done by TCV-01. As the output signal reaches 6 PSIG (41.3 kPa), heating is terminated; if free cooling is available, it is initiated at 7 PSIG (48.2 kPa). When the output signal reaches 11 PSIG (75.8 kPa) — the point at which OAD-05 is fully open — the cooling potential represented by free cooling is exhausted, and at 12 PISG (82.7 kPa), "pay cooling" is started by opening TCV-02. In such split-range systems, the possibility of simultaneous heating and cooling is eliminated. Also eliminated are interactions and cycling.

Figure 8.2p also shows some important overrides. TIC-12, for example, limits the allowable opening of OAD-05, so that the mixed-air temperature will never be allowed to drop to the freezing point and permit freeze-up of the water coils.

The minimum outdoor air requirement signal guarantees that the outside air flow will not be allowed to drop below this limit.

The economizer signal allows the output signal of TIC-07 to open OAD-05 only when "free cooling" is available. (A potential for free cooling exists when the enthalpy of the outdoor air is below that of the return air.)

Finally, the humidity controls will override the TIC-07 signal to TCV-02 when the need for dehumidification requires that the supply air temperature be lowered below the set point of TIC-07.

Humidity Controls

Humidity in the zones is controlled according to the moisture content of the combined return air (see Figure 8.2q). The

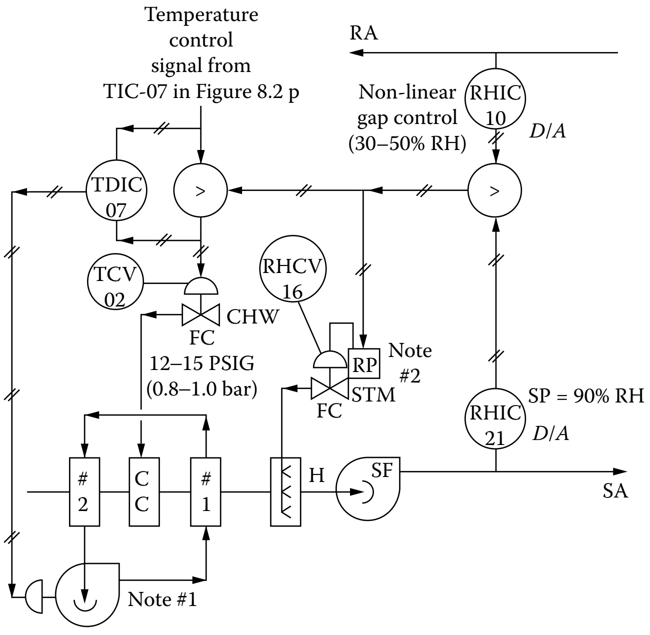


FIG. 8.2q
Humidity is controlled in the combined return air. Note 1: *When the need for dehumidification (in the summer) overcools the supply air and therefore increases the need for reheat at the zones, this pump-around economizer loop is started. TDIC-07 will control the pump to "pump around" only as much heat as is needed.* Note 2: *This reversing positioner functions as follows:*

Input from RHIC-10	*Output to RHCV-16*
3 PSIG = 0%	100% (open)
9 PSIG = 50%	0% (closed)

process controlled by RHIC-10 is slow and contains large dead-time and transport-lag elements. In other words, a change in the SA humidity will not be detected by RHIC-10 until some minutes later. During the winter, it is possible for RHIC-10 to demand more and more humidification. To prevent possible saturation of the supply air, the RHIC-10 output signal is limited by RHIC-21. In this way, the moisture content of the supply air is never allowed to exceed 90% RH.

For best operating efficiency, a nonlinear controller with a neutral band is used at RHIC-10. This neutral band can be set to a range of humidity levels — say, between 30% and 50% RH. If the RA is within these limits, the output of RHIC-10 is at 50%, and neither humidification nor dehumidification is demanded. This arrangement can lower the cost of humidity control during the spring and fall by approximately 20%.

The same controller (RHIC-10) controls both humidification (through the relative humidity control valve, RHCV-16) and dehumidification (through the temperature control valve, TCV-02) on a split-range basis. As the output signal increases, the humidifier valve closes, between 3 and 9 PSIG (20.7 and 62 kPa). At 9 PSIG (62 kPa), RHCV-16 closes and remains so, as the output signal increases to 12 PSIG (82.7 kPa). At this condition, TCV-02 starts to open.

Dehumidification is accomplished by cooling through TCV-02. This chilled water valve is controlled by humidity (RHIC-10) or temperature (TIC-07). The controller that requires more cooling will be the one allowed to throttle TCV-02.

Subcooling the air to remove moisture can substantially increase operating costs if this energy is not recovered. The dual penalty incurred for overcooling for dehumidification purposes is the high chilled water cost and the possible need for reheat at the zone level. The savings from a pump-around economizer can eliminate 80% of this waste. In this loop, whenever TDIC-07 detects that the chilled water valve (TCV-02) is open more than would be necessary to satisfy TIC-07, the pump-around economizer is started. This loop in coil #1 reheats the dehumidified supply air, using the heat that the pump-around loop removed from the outside air in coil #2 before it entered the main cooling coil.

In this way, the chilled-water demand is reduced in the cooling coil (TCV-02), and the need for reheating at the zones is eliminated. Although Figure 8.2q shows a modulating controller setting the speed of a circulating pump, it is also possible to use a constant-speed pump operated by a gap switch.

Outdoor Air Controls

Outdoor air is admitted to satisfy the requirements for fresh air or to provide free cooling. Both control loops are shown in Figure 8.2r.

The minimum requirement for fresh outdoor air while the building is occupied is usually 10% of the airhandler's capacity. In most advanced control systems, this value is not controlled as a fixed percentage but as a function of the number of people in the building or of the air's carbon dioxide content. In most conventional systems, the minimum outdoor air is provided by keeping 10% of the area of the outdoor air damper always open when the building is occupied.

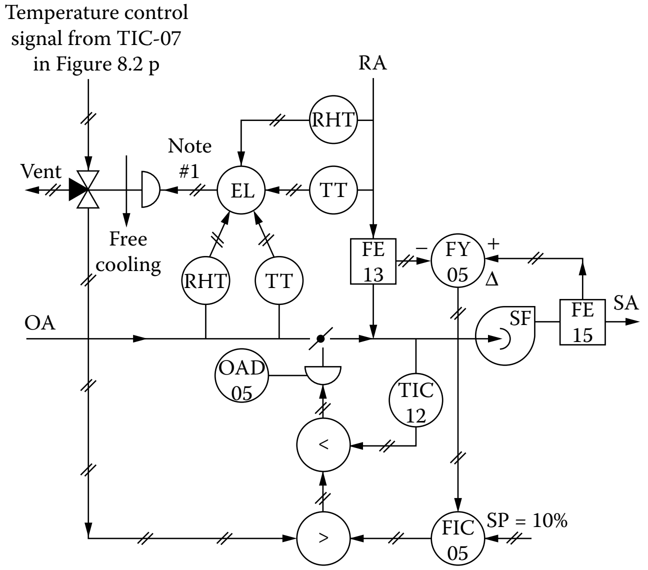


FIG. 8.2r
Outside air control loops provide fresh air or free cooling. Note: *The enthalpy logic unit (EL) compares the enthalpies of the outside and return airs and vents its output signal if free cooling is available. Therefore, the economizer cycle is initiated whenever Hoa < Hra.*

This method is inaccurate, because a constant damper opening does not result in a constant air flow. This flow varies with fan load, because changes in load will change the fan's suction pressure and will therefore alter the ΔP across the damper. This conventional design results in waste of air-conditioning energy at high loads and insufficient air refreshment at low loads.

The control system depicted in Figure 8.2r reduces operating costs while maintaining a constant minimum rate of air refreshment, which is unaffected by fan loading. Direct measurement of outdoor airflow is usually not possible because of space limitations. For this reason, Figure 8.2r shows the outdoor air flow as being determined as the difference between FE-15 and FE-13. FIC-05 controls the required minimum outdoor airflow by throttling OAD-05.

CO_2-Based Ventilation In conventional installations the amount of outdoor air admitted is usually based on one of the following criteria:

- 0.1–0.25 cfm/ft^2 (30–76 lpm/m^2) of floor area
- 10–25% of total air supply rates
- About 5 cfm (25 lps) volumetric rate per person

These criteria all originated at a time when energy conservation was no serious consideration; therefore, their aim was to provide simple, easily enforceable rules that will guarantee that the outdoor air intake always exceeds the required minimum. Today the goal of such systems is just the opposite: It is to make sure that air quality is guaranteed at minimum cost. As the floor does not need oxygen—only people do—some of the above rules make little sense.

There is a direct relationship between savings in building operating costs and reduction in outdoor air admitted into the building. According to one study in the United States,[4] infiltration of outdoor air accounted for 55% of the total heating load and 42% of the total cooling load. Another survey[5] showed that 75% of fuel oil consumed in New York City schools was devoted to heating ventilated air. Because building conditioning accounts for nearly 20% of all the energy consumed in the U.S.,[6] optimized admission of outdoor air can make a major contribution to reducing our national energy budget. This goal can be well served by CO_2-based ventilation controls.

The purpose of ventilation is not to meet some arbitrary criteria, but to maintain a certain air quality in the conditioned space. Smoke, odors, and other air contaminant parameters can all be correlated to the CO_2 content of the return air.[7] This then becomes a powerful tool of optimization, because the amount of outdoor air required for ventilation purposes can be determined on the basis of CO_2 measurement, and the time of admitting this air can be selected so that the air addition will also be energy efficient. With this technique, health and energy considerations will no longer be in conflict, but will complement each other.

CO_2-based ventilation controls can easily be integrated with the economizer cycle and can be implemented by use of conventional or computerized control systems. Because the rate of CO_2 generation by a sedentary adult is 0.75 cfh (27 lph), control by CO_2 concentration will automatically reflect the level of building occupancy.[8] Energy savings of 40% have been reported[9] by converting conventional ventilation systems to intermittent CO_2-based operation.

Economizer Cycles The full use of free cooling can reduce the yearly air-conditioning load by more than 10%. The enthalpy logic unit (EL) in Figure 8.2r will allow the temperature-controller signal (TIC-07 in Figure 8.2p) to operate the outdoor air damper whenever free cooling is available. This economizer cycle is therefore activated whenever the enthalpy of the outdoor air is below that of the return air.

Free cooling can also be used to advantage while the building is unoccupied. Purging the building with cool outdoor air during the early morning results in cooling capacity being stored in the building structure, reducing the daytime cooling load.

The conventional economizers—such as the one shown in Figure 8.2r—are rather limited devices for two reasons. First, they determine the enthalpy of the outdoor and return air streams, using somewhat inaccurate sensors. Secondly, although they consider the free cooling potential of the outside air, they disregard all the other possibilities of using the outdoor air to advantage.

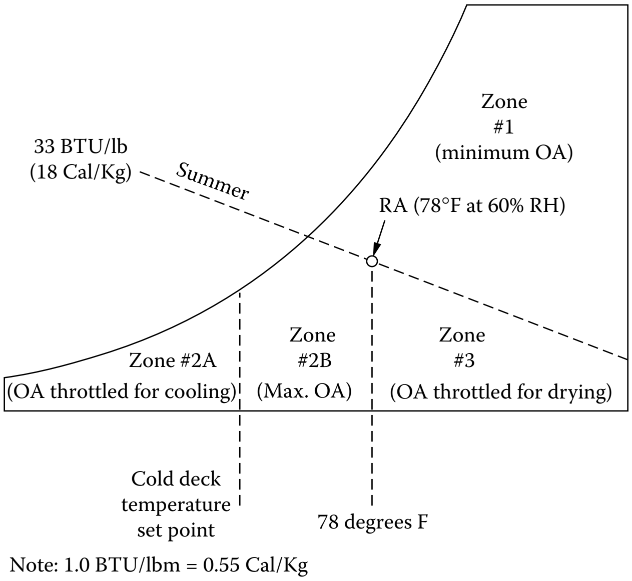


FIG. 8.2s
Free cooling and drying can often be obtained in the summer, depending on the zone where the outside air falls relative to the return air.

Advanced, microprocessor-based economizers overcome both of these limitations. They use accurate sensors and the psychrometric chart to evaluate all potential uses of outside air, not only free cooling. Figure 8.2s illustrates the various zones of operation, based on the relative conditions of the outside and return airs.

If the enthalpy of the outside air falls in zone 1 (that is, if its BTU content exceeds 33), no free cooling is available; therefore, the use of outside air should be minimized in the summer. In the winter or fall, it is possible that the enthalpy of the outside air on sunny afternoons will exceed the return air enthalpy, which in the winter is about 21 BTU/lbm (11.6 cal/kg). Under such conditions, "free heating" can be obtained by admitting the outside air in zone 1.

If the condition of OA corresponds to zone 2 (BTU < 33 and temperature < 78°F), free cooling is available. If the condition of OA corresponds to zone 3 (BTU < 33 and temperature > 78°F), free dehumidifying (latent cooling) is available.

When there are both cold and hot air ducts in the building (dual duct system), the control system in zone 2 will function differently depending on whether the outside air temperature is above or below the cold deck temperature. If it is above that temperature (zone 2B), maximum (100%) outside air can be used; if it is below that temperature (zone 2A), the use of outside air needs to be modulated or time-proportioned.

Therefore, in zone 2A, where the outside air is cooler than the cold deck temperature, free cooling is available, but only some of the total potential can be used. The OA damper

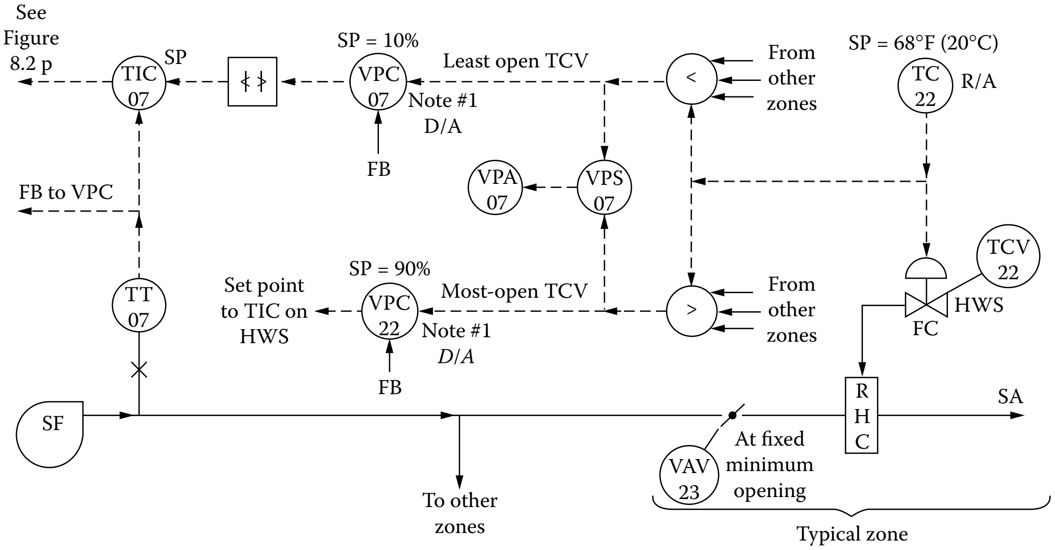


FIG. 8.2t

Optimization of air and water temperatures during heating (winter) mode of airhandler operation is designed to distribute the heat load efficiently between the main heating coil and the zone reheat coils. Note: *Valve position controllers (VPCs) are provided with integral action only for stable floating control. The integral time is set to be 10 times the integral time of the associated TIC. External feedback is provided to eliminate reset wind-up when VPC output is overruled by the setpoint limit on TIC-07.*

therefore must be either positioned or cycled to admit enough cold outside air to lower the mixed air temperature to equal the cold deck temperature set point. This way the need for "pay cooling" is eliminated.

When the outside air is in zone 3 (hot but dry), the control system should admit some of it to reduce the overall need for latent cooling. The amount of outside air admitted under these conditions should be controlled so that the resulting mixed-air dew point temperature will equal the desired cold deck dew point. This method takes full advantage of the free drying potential of the outside air.

As can be seen from the above, the use of the economizer cycles is similar to opening the windows in a room to achieve maximum comfort with minimum use of "pay energy."

OPTIMIZING STRATEGIES

The main goal of airhandler optimization is to match the demand for conditioned air with a continuous, flexible, and efficient supply. This requires the floating of both air pressures and temperatures to minimize operating costs while following the changing load.

In traditional HVAC control systems, the set point of TIC-07 in Figure 8.2p is set manually and is held as a constant. This practice is undesirable, because for each particular load distribution, the optimum SA temperature is different. If the manual setting is less than that temperature, some zones will become uncontrollable. If it is more than optimum, operating energy will be wasted.

Temperature Optimization in the Winter

In the winter, the goal of optimization is to distribute the heat load between the main heating coil (HC in Figure 8.2p) and the zone reheat coils (RHC in Figure 8.2t) most efficiently. The highest efficiency is obtained if the SA temperature is high enough to meet the load of the zone when the zone load is minimum. In this way, the zone reheat coils are used only to provide the difference between the loads of the various zones, whereas the base load is continuously followed by TIC-07.

In Figure 8.2t the set point of TIC-07 is adjusted to keep the least-open TCV-22 only about 10% open. First, the least-open TCV-22 is identified and its opening (in the valve position controller, VCP-07) is compared with the desired goal of 10%. If the valve opening is more than 10%, the TIC-07 set point is increased; if less, the set point is lowered. Stable operation is obtained by making VPC-07 an integral-only controller with an external feedback to prevent reset wind-up.

In Figure 8.2t it is assumed that the hot water supply (HWS) temperature is independently adjustable and will be modulated by VPC-22 so that the most-open TCV-22 will always be 90% open. The advantages of optimizing the HWS temperature include:

- Minimizing heat-pump operating costs by minimizing HWS temperature
- Reducing pumping costs by opening all TCV-22 valves in the system
- Eliminating unstable (cycling) valve operation, which occurs in the nearly closed position, by opening all TCV-22 valves

In Figure 8.2t, a valve position alarm (VPA-07) is also provided to alert the operator if this "heating" control system is incapable of keeping the openings of all TCV-22 valves between the limits of 10% and 90%. Such alarms will occur if the VPCs can no longer change the TIC set point(s), because their maximum (or minimum) limits have been reached. This condition will occur only if the load distribution was not correctly estimated during design or if the mechanical equipment was not correctly sized.

If the HWS temperature cannot be modulated to keep the most-open TCV-22 from opening to more than 90%, then the control loop depicted in Figure 8.2u should be used. In this loop, as long as the most-open valve is less than 90% open, the SA temperature is set to keep the least open TCV-22 at 10% open (zone A). When the most-open valve reaches 90% open, control of the least-open valve is abandoned and the loop is dedicated to keeping the most-open TCV-22 from becoming fully open (zone B). This, therefore, is a classic case of herding control, in which a single constraint envelope "herds" all TCV openings to within an acceptable band and, thereby, accomplishes efficient load following.

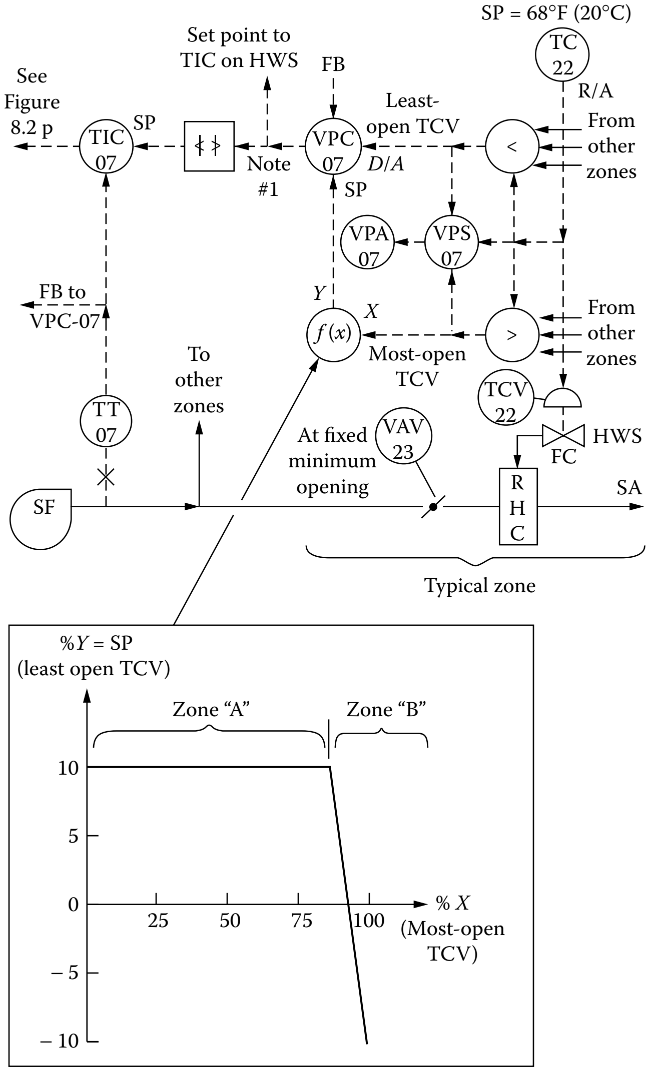


FIG. 8.2u

This alternative method of air supply temperature optimization in the winter should be used when the HWS temperature cannot be modulated. Note: *Valve position controllers (VPCs) are provided with integral action only for stable floating control. The integral time is set to be 10 times the integral time of the associated TIC-07. External feedback is provided to eliminate reset wind-up when TICVPC-07 output is overruled by the set-point limit on TIC-07.*

Temperature Optimization in the Summer

In the cooling mode during the summer, the SA temperature is modulated to keep the most-open variable volume box (VAV-23) from fully opening. Once a control element is fully open, it can no longer control; therefore, the occurrence of such a state must be prevented. On the other hand, it is generally desirable to open throttling devices such as VAV boxes to accomplish the following goals:

1. Reduce the total friction drop in the system
2. Eliminate cycling and unstable operation (which is more likely to occur when the VAV box is nearly closed)
3. Allow the airhandler to meet the load at the highest possible supply air temperature

This statement does not apply if air transportation costs exceed cooling costs (for example, undersized ducts, inefficient fans). In this case, the goal of optimization is to transport the minimum quantity of air. The amount of air required to meet a cooling load will be minimized if the cooling capacity of each unit of air is maximized. Therefore, if fan operating cost is the optimization criterion, the SA temperature is to be kept at its achievable minimum, instead of being controlled as in Figure 8.2v.

If the added feature of automatic switchover between winter and summer modes is desired, the control system depicted in Figure 8.2w should be used.

When all zones require heating, this control loop will behave exactly as does the one shown in Figure 8.2u; when all zones require cooling, it will operate as does the system shown in Figure 8.2v. In addition, this control system will operate automatically with maximum energy efficiency during the transitional periods of fall and spring. This high efficiency is a result of the exploitation of the self-heating effect.

If some zones require heating (perimeter offices) and others require cooling (interior spaces), the airhandler will automatically transfer this free heat from the interior to the perimeter zones by intermixing the return air from the various zones and moving it through the 10° zero energy band (between the settings of TC-22 and TC-23. When the zone temperatures are within this comfort gap of 68°F (20°C) to 78°F (26°C), no pay energy is used and the airhandler is in its self-heating, or free-heating, mode. This is an effective

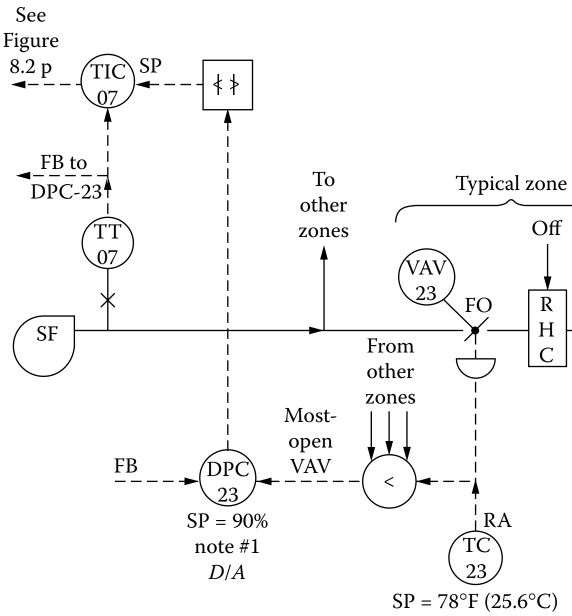


FIG. 8.2v
This method of air supply temperature optimization in summer (cooling mode) should be used if fan operating costs are less than cooling costs. Note: *The damper position controller (DPC) has integral action only, with its setting being 10 times the integral time of TIC-07. External feedback is provided to eliminate reset wind-up.*

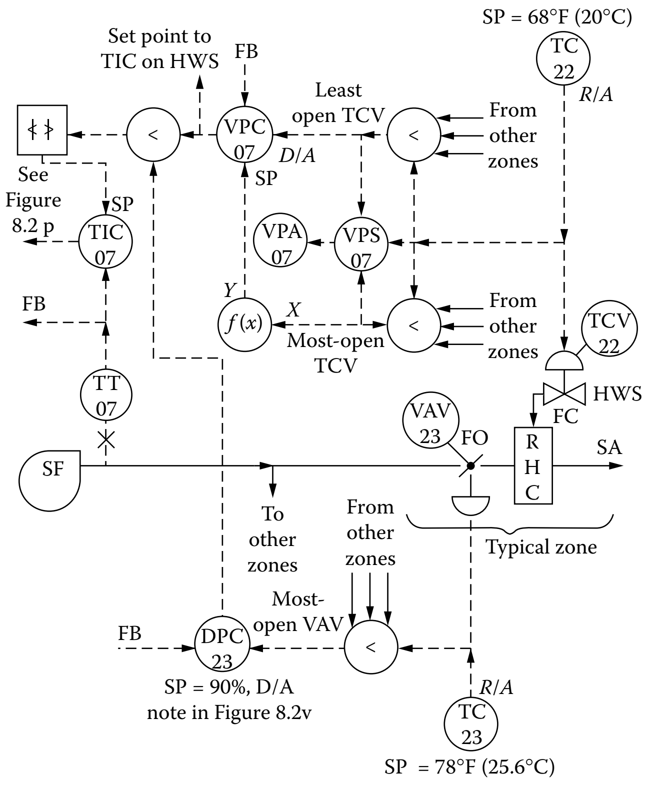


FIG. 8.2w
This control system optimizes the water and air temperatures in both summer and winter. See also Figures 8.2u and 8.2v.

means of reducing operating costs in buildings. The savings can amount to more than 30% during the transitional seasons.

When the temperature in one of the zones reaches 78°F (26°C), the air supply temperature set point will be lowered by DPC-23 and the air-side controls will be automatically switched to cooling (as depicted in Figure 8.2v). If, at the same time, some other zone temperatures drop below 68°F (20°C), requiring heating, their heat demand will have to be met by the heat input of the zone-reheat coils only. This mode of operation is highly inefficient because of the simultaneous cooling and reheating of the air.

Fortunately, this combination of conditions is highly unlikely, because under proper design practices, the zones served by the same airhandler should display similar load characteristics. The advantage of the control loop in Figure 8.2w is that it can automatically handle any load or load combination, including this unlikely, extreme case.

AutoBalancing of Buildings

In computerized building control systems, the optimization potentials are greater than those that have been discussed up to this point. When all zone conditions are detected and controlled by the computer, it can optimize not only the normal operation but also the start-up of the building.

The optimization of airhandler fans is directed at two goals simultaneously. The first goal is to find the optimum value for the set point of the supply air pressure controller (PIC-19 in Figure 8.2f). Generally, the supply air pressure is at an optimum value when it is at the lowest possible value, while all loads are satisfied. As the supply pressure is lowered, the fan operating cost is reduced, but with lowered supply pressures the VAV boxes serving the individual zones (VAV-23 in Figure 8.2b) will have to open up so that the airflow to the zones will not be reduced. Therefore, the optimum setting for PIC-19 is that pressure at which the most open VAV box is nearly 100% open, while all other VAV boxes are less than 100% open.

The second goal of optimization is to automatically rebalance the air distribution in the building as the load changes. If the VAV boxes (VAV-23 in Figure 8.2b) are not pressure independent (are not able to maintain constant air flow when the supply pressure changes), manual rebalancing is required every time the load distribution changes. Naturally this is a very labor-demanding and inefficient operation. The optimization strategy described below serves the multiple purposes of automatic rebalancing and finding the optimum set points for the supply air pressure and temperature.

Figure 8.2x illustrates an airhandler that is serving several zones. The abbreviations used in that figure and in the algorithm tables that follow are listed below:

AI-1 to AI-N: Analog inputs (zone temperatures)
AI-AT: Analog input (air supply temperatures)
AI-RT: Analog input (return air temperature)
AO-1 to AO-N: Analog output (zone VAV opening)

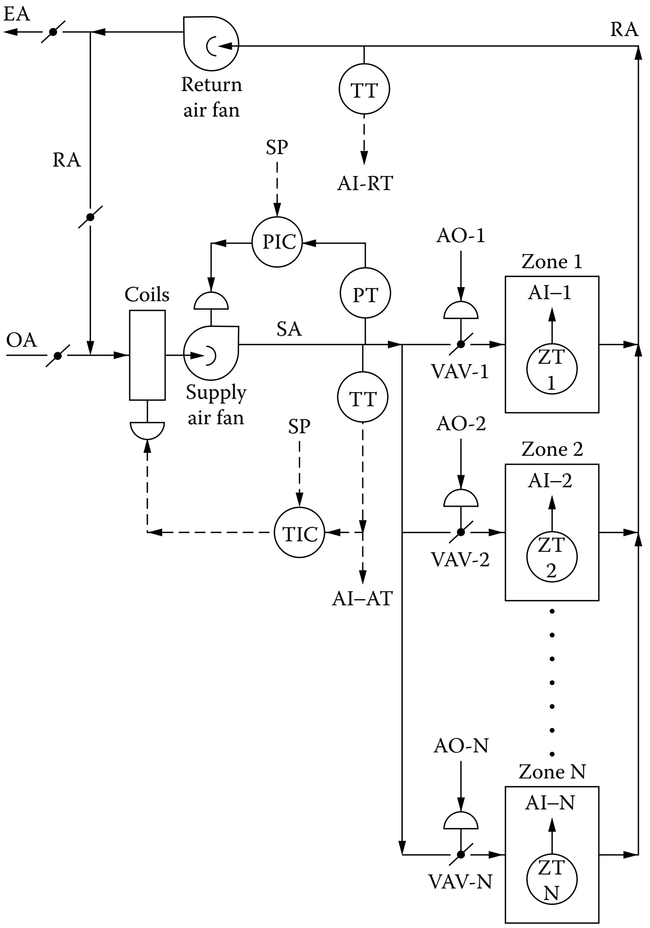


FIG. 8.2x

Airhandler optimization and auto-balancing can be handled efficiently by computer.

AT: Supply air temperature
EA: Exhaust air
OA: Outside air
PIC: Pressure controller (supply air)
PT: Pressure transmitter
RA: Return air
RT: Return air temperature
SA: Supply air
SP: Set point
TH: Upper limit of comfort zone (Figure 8.2a) (maximum allowable zone temperature)
TIC: Temperature controller (air supply)
TL: Lower limit of comfort zone (Figure 8.2a) (minimum allowable zone temperature)
VAV-1 to VAV-N: Variable air volume boxes
ZT-1 to ZT-N: Zone temperatures
ZT5-1 to ZT5-N: ZT 5 min after start-up
ZT10-1 to ZT10-N: ZT 10 minutes after start-up
XMIN: Minimum VAV opening required for ventilation
XSET-1 to XSET-N: Initial VAV opening after “start-up”
XMAX-1 to XMAX-N: Maximum limit on VAV opening during normal operation

Start-Up Algorithm

All VAV boxes are set to their minimum openings required for ventilation purposes (XMIN), such as 25%. Therefore, at the time of start-up, AO-1 through AO-N are all set for 25%. PIC is set to mid-scale; therefore, the start-up value of its SP = 50%. TIC is set for (TL+25)°F in the heating mode and for (TH-25)°F in the cooling mode.

After 5 min of operation, the zone temperatures are detected (ZT5-1, ZT5-2, and so on), and after 10 min of operation they are detected again (ZT10-1, ZT10-2, and so on). At the end of the first 10 min of operation, the supply air temperature is also measured as AT10 and the return air temperature as RT10.

Once the above readings are obtained, they are entered into a table such as Table 8.2y, which serves as the basis for determining the required start-up openings of each of the VAV boxes (XSET-1, XSET-2, and so on). The purpose of this table is to select the initial opening for each VAV box in a logical manner. Therefore, if the zone temperature after 10 minutes of operation is already within 5°F of reaching the comfort zone, the VAV box can be left at its minimum opening.

If comfort is not yet within 5°F, a higher opening is needed. The initial VAV opening is increased on the basis of the zone’s performance during the previous 5 minutes. The larger the temperature change experienced by the zone during the previous 5 minutes, the sooner it will reach the comfort zone and therefore the smaller the opening that is required. By this logic, the VAV boxes on those zones that are furthest from comfort and that are moving most slowly toward comfort will be given the highest openings.

Normal Algorithm for VAV Throttling

The initial VAV opening for each zone (XSET), which is determined by the methods above, is then used as the maximum limit on the VAV opening (XMAX) during the first 5 minutes of normal operation. The value of XMAX is reevaluated every 5 minutes, as shown in Table 8.2z. The logic here is to increase the maximum limit on VAV opening (XMAX) to any zone in which the VAV has been open on its maximum limit for 5 minutes. Similarly, this logic will lower the XMAX limit if the VAV damper was at its XMIN during the previous 5 minutes. If a VAV damper has been throttled somewhere in between these two limits (XMAX and XMIN), its limit will not be altered. The change increment of 10% shown in Table 8.2z is adjustable for maximum flexibility.

The algorithm described above and illustrated in Table 8.2z guarantees that changes in load distribution will not result in starving some zones; the building will be automatically rebalanced in an orderly manner. The value of XMAX from Table 8.2z and the permanent values of XMIN, determined by ventilation requirements, are used to reevaluate the individual VAV openings every 2 minutes, as described by the algorithm in Table 8.2aa.

TABLE 8.2y
Algorithm to Determine Start-Up Openings of Individual VAV Boxes (XSET)

Input Conditions			*Output*
Operating Mode	*Approach between Zone and Supply Temperatures*	*Amount of Temperature Change During Last 5 Minutes of Start-Up*	*Initial Value of XSET to be Used for AO-1 to AO-N (%)*
Heating (AT > RT)	(TL − ZT10) > 5°F	(ZT10 − ZT5) < 0.5°F	100
		(ZT10 − ZT5) 0.5–1°F	90
		(ZT10 − ZT5) 1–1.5°F	80
		(ZT10 − ZT5) 1.5–2°F	70
		(ZT10 − ZT5) 2–3°F	60
		(ZT10 − ZT5) 3–4°F	50
		(ZT10 − ZT5) 4–5°F	40
		(ZT10 − ZT5) > 5°F	30
	(TL − ZT10) < 5°F	Disregard	25
Cooling (AT < RT)	(ZT10 - TH) > 5°F	(ZT5 − ZT10) < 0.5°F	100
		(ZT5 − ZT10) 0.5–1°F	90
		(ZT5 − ZT10) 1–1.5°F	80
		(ZT5 − ZT10) 1.5–2°F	70
		(ZT5 − ZT10) 2–3°F	60
		(ZT5 − ZT10) 3–4°F	50
		(ZT5 − ZT10) 4–5°F	40
		(ZT5 − ZT10) > 5°F	30
	(ZT10 - TH) < 5°F	Disregard	25

Note: °C = (F − 32)/1.8.

The main optimizing and auto-balancing feature of this algorithm is that whenever a zone is inside the comfort gap, its VAV opening is reduced to XMIN. This reduces the load on the fans and also provides more air to the zones experiencing the highest loads.

Optimization of Air Supply Pressure and Temperature

Optimization means that the load is met at minimum cost. The cost of operating an airhandler is the sum of the cost of air transportation and conditioning. These two cost factors tend to change in opposite directions; minimizing the cost of one will increase the cost of the other. Therefore, it is important to monitor both the transportation and the conditioning costs continuously and to minimize the larger one when optimizing the system. Computerized control systems allow these costs to be readily calculated on the basis of utility costs and quantities.

TABLE 8.2z
Reevaluation of Value of XMAX

Input Conditions		*Output*
Has VAV been Continuously Open to its XMAX During Last 5 Minutes?	*Has VAV been Continuously Throttled to its XMIN During the Last 5 Minutes?*	*Incremental Change in Value of XMAX at the End of 5-Minute Period*
Yes	Yes	Leave XMAX = XMIN
	No	Increase by 10%
No	Yes	Decrease by 10%
	No	Leave as is

TABLE 8.2aa
Algorithm to Determine Analog Outputs, Setting the Openings of VAV Boxes

Input Conditions		*Output*
Operating Mode	*Control Criteria*	*Required VAV Opening: AO-1 to AO-N is to be Equal*
Heating (AT > ZT)	ZT < (TL − 1)	XMAX
	(TL − 1) < ZT < (TL + 1)	No change
	ZT > (TL + 1)	XMIN
Cooling (AT < ZT)	ZT > (TH + 1)	XMAX
	(TH − 1) < ZT < (TH + 1)	No change
	ZT < (TH − 1)	XMIN

TABLE 8.2bb

Optimization of Supply Air Pressure and Temperature, When Fan Costs Exceed Conditioning Costs (Frequency = 5 min)

			Incremental Ramp Adjustment in the Set Points of	
VAV Status	*Airhandler Mode*	*Is TIC SP at Its Limit?*	*TIC*	*PIC*
None at 100% for 15 minutes continuously	Heating (AT > RT)	Yes, at max.	−2°F	N.C.* (at min.)
		No	−1°F	N.C. (at min.)
	Cooling (AT < RT)	Yes, at min.	+2°F	N.C. (at min.)
		No	+1°F	N.C. (at min.)
Not more than one at 100% for more than 30 minutes continuously	Heating (AT > RT)	Yes, at max.	N.C. (at max.)	N.C.
		No	N.C.	N.C.
	Cooling (AT < RT)	Yes, at min.	N.C. (at max.)	N.C.
		No	N.C.	N.C.
More than one at 100% for more than 30 minutes continuously	Heating (AT > RT)	Yes, at max.	(at max.)	+0.25 in. H_2O
		No	+1°F	N.C.
	Cooling (AT < RT)	Yes, at min.	(at min.)	+0.25 in. H_2O
		No	−1°F	N.C.
More than one at 100% for more than 60 minutes continuously	Heating (AT > RT)	Yes, at max.	(at max.)	+0.5 in. H_2O
		No	+2°F	N.C.
	Cooling (AT < RT)	Yes, at min.	(at min.)	+0.5 in. H_2O
		No	−2°F	N.C.

*N.C. = No change is made at the end of that 5-minute period.

For example, if the transportation cost exceeds the conditioning cost, the optimization goal is to minimize fan operation. This is achieved by conditioning the space with as little air as possible. The quantity of air transported can be minimized if each pound of air is made to transport more conditioning energy; that is, if each pound of air carries more cooling or heating BTUs. Therefore, when the goal is to minimize fan costs, the air supply pressure is held as low as possible, and the air supply temperature is maximized in the winter and minimized in the summer. Fan costs tend to exceed conditioning costs when the loads are low, such as in the spring or fall, or when the economizer cycle is used to provide free cooling.

Table 8.2bb describes the algorithm used to achieve this goal. When none of the VAV boxes (Figure 8.2x) are fully open, indicating that all loads are well satisfied, the air pressure (PIC set point) is kept at a minimum, and the air temperature (TIC set point) is lowered in the winter and raised in the summer. When more than one VAV boxes are fully open, the air supply temperature is increased in the winter (lowered in the summer). When its limit is reached, the algorithm will start raising the PIC set point.

Table 8.2cc describes the algorithm used when the conditioning costs are higher than the fan operating costs. This is likely to be the case when the loads are high, such as in the summer or the winter. Under such conditions, the supply pressure is maximized before the supply air temperature is increased in the winter or lowered in the summer. When none of the VAV boxes in Figure 8.2x are fully open, the PIC set point is lowered, while the TIC set point is at or near minimum in the winter (maximum in summer). When more than one VAV box is fully open, the PIC set point is increased to its maximum setting. When that is reached, the supply temperature starts to be increased in the winter (decreased in the summer).

The algorithms described above provide the dual advantages of automatic balancing and minimum operating cost. They eliminate the need for manual labor or for the use of pressure-independent VAV boxes, while reducing operating cost by about 30%. They also provide the flexibility of assigning different comfort envelopes (different TL and TH values) to each zone. Thereby, as occupancy or use changes, the comfort zone assigned to the particular space can be changed automatically.

ELIMINATION OF CHIMNEY EFFECTS

In high-rise buildings, the natural draft resulting from the chimney effect tends to pull in ambient air at near ground elevation and to discharge it at the top of the building.

TABLE 8.2cc
Optimization of Supply Air Pressure and Temperature, When Conditioning Costs Exceed Fan Costs (Frequency = 5 min)

			Incremental Ramp Adjustment in the Set Points of	
VAV Status	*Airhandler Mode*	*Is PIC Set Point at Its Maximum?*	*TIC*	*PIC*
None at 100% for 15 minutes continuously	Heating (AT > RT)	Yes	–1°F	–0.5 in. H_2O
		No	N.C.* (at min.)	–0.25 in. H_2O
	Cooling (AT < RT)	Yes	+1°F	–0.5 in. H_2O
		No	N.C. (at max.)	–0.25 in. H_2O
Not more than one at 100% for more than 30 minutes continuously	Heating (AT > RT)	Yes	N.C.	N.C. (at max.)
		No	N.C.	N.C.
	Cooling (AT < RT)	Yes	N.C.	N.C. (at max.)
		No	N.C.	N.C.
More than one at 100% for more than 30 minutes continuously	Heating (AT > RT)	Yes	+1°F	(at max.)
		No	N.C.	+0.25 in. H_2O
	Cooling (AT < RT)	Yes	–1°F	(at max.)
		No	N.C.	+0.25 in. H_2O
More than one at 100% for more than 60 minutes continuously	Heating (AT > RT)	Yes	+2°F	(at max.)
		No	N.C.	+0.5 in H_2O
	Cooling (AT < RT)	Yes	–2°F	(at max.)
		No	N.C.	+0.5 in H_2O

* N.C. = No change.

Although eliminating the chimney effect can lower the operating cost by approximately 10%, few systems with this capability are yet in operation.

Figure 8.2dd shows the required pressure controls. The key element of this control system is the reference riser, which allows all pressure controllers in the building to be referenced to the barometric pressure of the outside atmosphere at a selected elevation. Using this pressure reference allows all zones to the operated at 0.1 in. H_2O (25 Pa) above that reference pressure (PC-7) and permits this constant pressure to be maintained at both ends of all elevator shafts (PC-8 and -9).

If the space pressure is the same on the various floors of a high-rise building, there will be no pressure gradient to motivate the vertical movement of the air, and as a consequence, the chimney effect will have been eliminated. A side benefit of this control strategy is the elimination of all drafts or air movements between zones, which also minimizes the dust content of the air. Another benefit is the capability of adjusting the "pressurization loss" of the building by varying the setting of PC-7, -8, and -9.

Besides reducing operating costs, the use of pressure-controlled elevator shafts increases comfort because drafts and the associated noise are eliminated.

Figure 8.2dd also shows the use of cascaded fan controls. The set points of the cascade slaves (PC-2 and PC-5) are programmed so that the air pressure at the fan is adjusted as the square of flow and the pressure at the end of the distribution headers (cascade masters PC-1 and PC-6) remains constant. This control approach results in the most efficient operation of variable-air volume fans.

If the building is maintained at a constant pressure that equals the pressure at ground elevation plus 0.1 in. H_2O, this will result in higher pressure differentials on the higher floors, as the barometric pressure on the outside drops. Therefore, air losses due to out-leakage and pressure differentials on the windows will both rise. If the pressure reference is taken at some elevation above ground level, these effects will be reduced on the upper floors, but on the lower floors the windows will be under positive pressure from the outside and air infiltration will be experienced.

CONCLUSIONS

The airhandler is just one of the industrial unit operations. The process of air conditioning is similar to all other industrial processes. Fully exploiting state-of-the-art instrumentation and control results in dramatic improvements. There are few other processes in which the use of optimization and of instrumentation know-how alone can halve the operating cost of a process.

The control and optimization strategies described in this chapter can be implemented by pneumatic or electronic

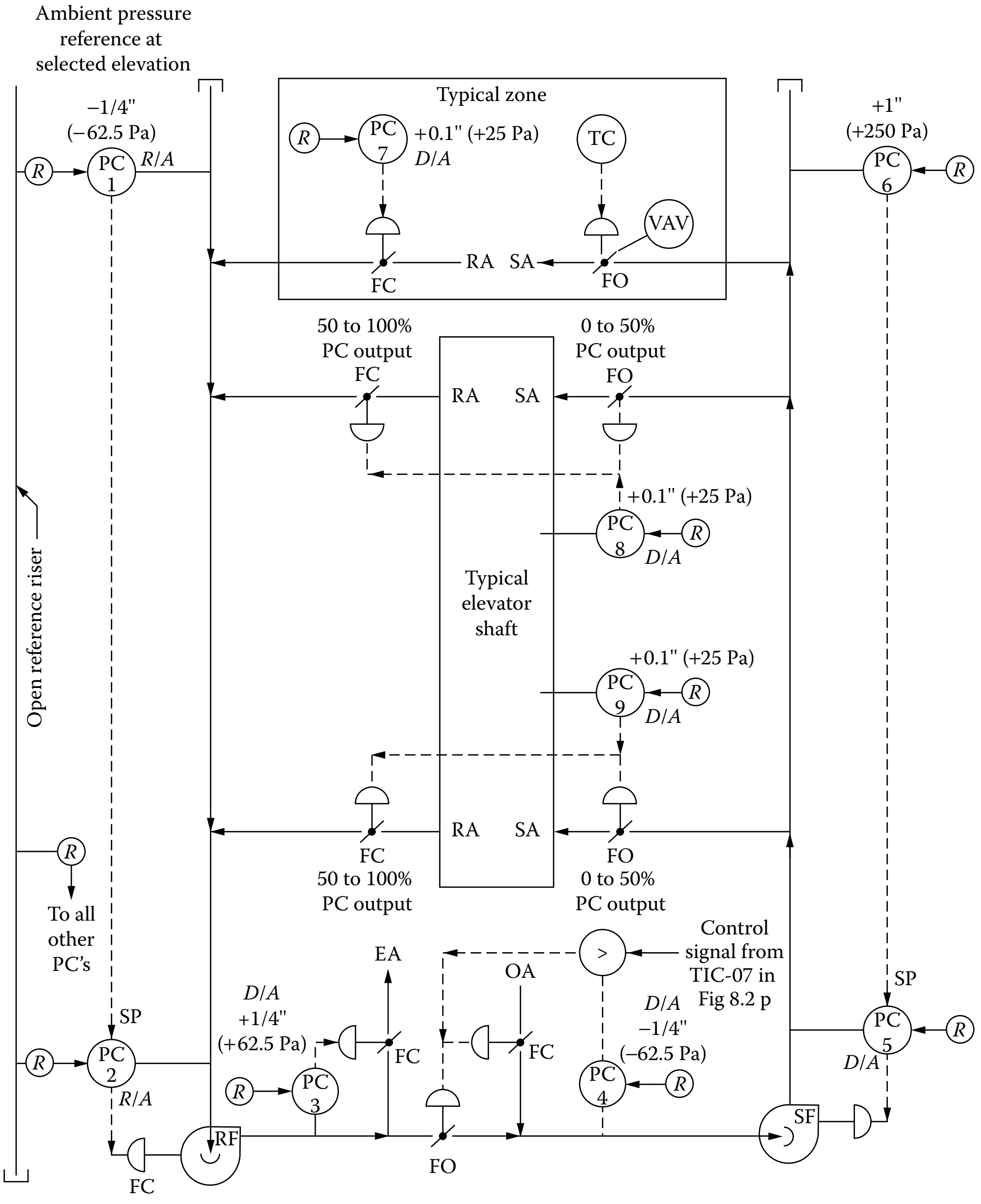


FIG. 8.2dd
Chimney effects in high-rise buildings can be eliminated by using the proper pressure controls.

instruments and can be controlled by analog or digital systems. The type of hardware used in optimization is less important than the understanding of the process and of the control concepts that are to be implemented. The main advantage of digital and computerized systems is their flexibility and convenience in making changes, without the need to modify equipment or wiring.

References

1. Lipták, B. G., "Applying the Techniques of Process Control to the HVAC Process," *ASHRAE Transactions,* Paper No. 2778, Vol. 89, Part 2A, 1983.
2. Avery, G., "VAV Economizer Cycle," *Heating, Piping, Air Conditioning,* August 1984.
3. Lipták, B. G., "Reducing the Operating Costs of Buildings by the Use of Computers," *ASHRAE Transactions,* 83:1, 1977.
4. Kovach, E. G., "Technology of Efficient Energy Utilization: The Report of a NATO Science Committee Conference Held at Les Arcs, France, 8–12 October, 1973," Scientific Affairs Division, NATO, Brussels, Belgium.
5. Liu, S. T., et al., "Research, Design, Construction and Evaluating an Energy Conservation School Building in New York City," NBSIR.
6. Stanford Research Institute, "Patterns of Energy Consumption in the United States," prepared for the Office of Science and Technology, Washington, D.C.
7. Nomura, G., and Yamada, Y., "CO_2 Respiration Rate for the Ventilation Calculation," *Transaction of Japanese Architectural Society Meetings,* October 1969.

8. Department of Defense, "Environmental Engineering for Shelters," TR-20-Vol. 3, Department of Defense, Office of Civil Defense, May 1969.
9. Kusuda, T., "Intermittent Ventilation for Energy Conservation," NBS report of ASHRAE Symposium in Dallas, TX, February 1976.

Bibliography

American Institute of Architects, *The Architects Handbook of Professional Practice,* with CD, 13th edition, New York: John Wiley & Sons, Inc.

ASHRAE Standard 62-2001, "Ventilation for Acceptable IAQ," ASHRAE, 2001.

Bell, A. A., Jr., *HVAC Equations* New York: McGraw-Hill, 2000.

Cooper, F. G., "Low Pressure Sensing and Control," *InTech,* September 1992.

Daryani, S., "Design Engineer's Guide to Variable Air-Volume Systems," *Actual Specifying Engineer,* July 1974.

DHO/Atlanta Corp., "Conserving Energy," Powered Induction Unit (Data Sheet #1), Atlanta, GA: DHO/Atlanta Corp., 1974.

Haines, R. W., and Wilson, C. L., *HVAC Systems Design Handbook*, New York: McGraw-Hill, 2003.

Luciano, J. R., "Energy Conservation Techniques for Hospital Operating Rooms," *ASHRAE Journal,* May 1983.

Nordeen, H., "Control of Ventilation Air in Energy Efficient Systems," ASHRAE Symposium #20, Paper #3.

Ozawa, K., Noda, Y., et al., "A Tuning Method for PID Controller Using Optimization Subject to Constraints on Control Input," ASHRAE Annual Meeting, Kansas City, 2003.

Rosaler, R. C., *HVAC Maintenance and Operations Handbook*, New York: McGraw-Hill, 1997

Shih, J. Y., "Energy Conservation and Building Automation," ASHRAE Paper #2354.

Spielvogel, L. G., "Exploding Some Myths about Building Energy Use," *Architectural Record,* February 1976.

Stillman, R. B., "Systems Stimulation," engineering report prepared for IBM-RECD, 1971.

Turk, A., "Gaseous Air Cleaning," *ASHRAE Journal,* May 1983.

Woods, J. E., "Impact of ASHRAE Ventilation Standards 62-73 on Energy Use," ASHRAE Symposium #20, Paper #1.

8.3 Batch Control Description, Terminology, and Standard S88

B. A. JENSEN (1995, 2004)

FEATURE SUMMARY

Batch processing and automation are discussed in this section both from a management point of view and from the perspective of the types of software structures required to implement it. Other sections in this chapter also deal with the subject of batch control. Section 8.4 concentrates on batch processes and their automation, while Sections 8.8 to 8.10 are devoted to the automation of chemical batch reactors.

In 1988, the Instrumentation, Systems and Automation Society (ISA) formed a standards committee to provide guidelines for the design and specifications of batch control systems as used in the process industries. Building upon previous work of the German NAMUR (Normenarbeits-gemeinschaft fur Mess und Regeltechnik in der Chemischen Industrie) and the Purdue Workshop TC-4 committee, the first part of the multipart standard ("Models and Terminology") defined terminology specific to batch control systems to facilitate understanding between manufacturers and users.[1] It also defined a standard batch control architecture that outlined a hierarchical structure relating control equipment and data communications needed for the physical areas involved in batch control and a functional model that showed the relationships between the control activities of recipe management, production scheduling, information management, batch management, unit supervision, and process control (sequential and regulatory control) required in batch control. This evolved into an international standard: IEC 61512-1.[2] As such, the other sections are also turning onto IEC standards including "Part 2: Data Structures and Guidelines for Languages,"[3] IEC 61512-2 standard in 2003.[4] "Part 3: General and Site Recipe Models and Representation" was approved by ISA in March 2003,[5] while "Part 4: Production Records,"[6] is expected shortly.

In addition, a nonprofit, professional organization called the World Batch Forum (WBF) was established in 1994 to promote the exchange of information related to the management, operation, and automation of batch process manufacturing. This association of end users, vendors, consultants, and academics hosts an annual conference with formal presentations and technical papers dedicated to advance batch processing knowledge and technology, with a strict, noncommercial agenda. Other forums include the European Batch Forum (EBF) and Japanese Batch Forum (JBF), which facilitate the same agenda.

Batch automation and batch control is unique depending upon the perspective used to describe it. Unlike continuous processes, a batch process has a finite beginning, middle, and end. Though it has been argued that continuous processes can be thought of as batch processes with very long batch cycle times, the uniqueness of batch control has to do with what procedure is performed, what formulations are used, and on what type of equipment.

Batch processes are event-driven processes that vary with time. Charging, heating, reacting, agitating, cooling, and discharging are examples of sequential events in time requiring corresponding control actions. In the design of a batch control system, time-based process conditions and transition phenomena must be handled. Attention to abnormal events and the interface to the operator may actually take more of the design process than that of the actual automation.

Batch processing can be viewed through three perspectives. One is from a process point of view. The second is an equipment view by which products are processed. The third is a product-based view, or recipe-based view. To that end, the discussion will first begin on the process point of view.

BATCH PROCESS CLASSIFICATION

A batch processing model must consider that a batch can have more than one end product and that production is possible in various plants or plant areas. Thus, batch processes can be classified according to either the number of products made or the physical structure of the process.

Recipe Point of View

A recipe is a batch entity that contains "the necessary set of information that uniquely defines the production requirements for a specific product."[1] That includes the header (purpose, source and recipe version product identification, creator, and issue date) the procedure (the strategy and set of instructions and action needed to carry out the production of a batch), formula (process input values, process parameter variables, and process output data), equipment requirements, and miscellaneous information, such as material safety data sheets (MSDS) and any other material that may be related to the batch.

Batch processes can be classified according to the number of chemicals, substances, or items produced, as follows: single procedure/single formula, single procedure/multiple formula, or multiple procedure/multiple formula. A *single-procedure/single-formula* batch process produces the same product in each batch. The same operations are performed in the same sequence, using the same percentages of the raw materials, though the batch sizes may change. Many of these types of plants can be converted to continuous processes. The *single-procedure/multiple-formula* batch process produces different grades of products that are similar but differ in formula quantities. The same operations are performed in each batch, but the quantity of raw materials or processing conditions are varied. The procedure is the same but the formulas are changed. *Multiple-procedure/multiple-formula* batch processes produce products by utilizing different methods of production or control. The procedures performed, the amounts of raw materials used, the processing conditions encountered, and the equipment used may vary with each batch. This is the most difficult of the three batch-type processes to automate.

Equipment Point of View

Batch facilities can also be characterized according to the physical structure of the process facility. Three basic types of batch structures are series (single-stream) structures, parallel (multistream) structures, and networked structures. A *series structure* is a group of batch equipment through which a batch passes sequentially. It could be a single batch unit, such as a reactor, or several processing units in sequence. In the parallel structure several batches can be undergoing the same (or different) operations at one time. A hybrid of the two structures is a series/parallel batch process, also known as a networked-type structure whereby the batch can pass through any equipment type. This requires the highest degree of sophistication in control equipment to achieve effective batch control.

Product Point of View

The product is the overall objective of any batch process. This view takes into account repeatability and quality of the batches produced. It is information intensive. It is concerned with batch and unit recipe cycle times, cycle time frequency, batch and unit recipe performance ratings, product quality, and other performance-based metrics. Questions are answered, such as, When was batch B-4290 made? What equipment was used? What products were in progress at 14:00, 2 May 2004? How was the product made in 2002? How long does it usually take to make this product? Which reactor typically makes this product the quickest? What other batches were in progress when B-4290 was running? What products have the greatest variability in their production? A fully developed automation system with a historical recipe-based manufacturing execution system is required to support this view.

Table 8.3a lists a variety of factors that can be considered in distinguishing batch processes.

TABLE 8.3a
Variations in Batch Processes and Their Features

Features	*Choices*
Types of recipes	Single procedure/single formula Single procedure/multiple formula Multiple procedure/multiple formula
Topology of plant	Numbers of stages Series/parallel/networked Interconnections: fixed or flexible connections Shared or exclusive use resources allocations
Type of equipment	Single-purpose Multipurpose
Movement of batches	Fixed paths/varying paths from batch to batch predefined in a recipe Dynamic unit assignments coupled with continuous processes Automatic reblending/rework
Control activity levels	Safety interlocking Basic control Unit supervision Batch management Production planning and scheduling Information management Recipe management
Sequence requirements	Number of control actions, inputs, and outputs are small Number of independent sequences and parallel actions are small
Amount of operator intervention	As part of the normal execution includes verification of required manual actions Overriding the recipe in case of abnormal events with return to automatic execution
Exception handling procedures	Permissives for advancing through unit procedure Interlocks continuously checked, system suspends or is held on detection of abnormality Interlocks independent of operations and phases; same interlocks throughout batch; no jumping to other phases Interlocks cause jump to a safe state; interlocks are control-step independent Interlocks are a function of operations and phases; both shutdowns and jumps involved; may be a number of abnormal transitions and states
Requirements for batch data collection	Receiving and storing information on individual batches Producing output information on one or more batches Producing batch reports Maintaining a batch history archive

BATCH AUTOMATION

Automation of batch production is composed of three functional levels:

1. Batch planning: The sequence of production is planned and scheduled taking into account the available resources, including raw materials, personnel, and equipment.
2. Batch control: Production flows and production steps are described. Recipes are processed and executed in this level.
3. Real-time monitoring and control: Basic equipment control is performed at this level. This classical process automation includes safety interlocking input/output processing, and routine sequential, regulatory, and discrete control.

The control activities of an entire batch control system are shown in a hierarchical manner from the sensors and elements to the business planning level in Table 8.3b. In many organizations the lines between the levels may be quite blurred. Additionally, the levels may be compressing though the functions and activities are certainly valid.

Batch planning activities involve process/product management, production/batch management and production planning and scheduling, batch control activities involve batch management and unit supervision, and basic equipment control activities involving sequential/regulatory/discrete control of physical control devices, such as sensors and actuators, and safety interlocking. The following paragraphs describe the control activities as they pertain to batch processes, from the lowest level on up.

Batch automation is planned from top down and implemented from the bottom up.[7]

TABLE 8.3b
Control Activity Model

Level	*Function*	*Activity*
Planning	Process/product management	Production planning, inventory planning, general recipe management, etc.
	Production management	Recipe management, production scheduling, batch history management, etc.
Batch control	Batch management	Recipe generation/selection, batch execution supervision, unit activities coordination, log and report generation, etc.
	Unit supervision	Unit allocation management, unit coordination, etc.
Monitoring and control	Process control	Sequential/regulatory/discrete control: device, loop, and equipment module control, predictive control, model-based control, process interlocking, etc.
	Safety interlocking	—

Process/Product Management

Process/product management is the highest level of control activity. This is where corporate planning for the business is made and is linked to the various operating units. Activities like material and resource planning (MRP), inventory planning, and accounting take place.

Process/product management is an activity that accepts inputs, such as customer orders, and based upon a manufacturing strategy develops a production plan. The production plan can cover such topics as:

What is to be produced?
How much is to be produced?
When is it needed?
Where is it to be produced?
How is it to be packaged?

The output of the production plan becomes an input to the production schedule of the individual plant. Process/product management entails a broad range of corporate planning and provides the basis for management's relationship with its operating units.

Production Management

Production management is made up of three control activities: recipe management, production scheduling, and batch history management.

Recipe Management A recipe is the complete set of data and operations that define the control requirements of a particular type or grade of product. A recipe is composed of the following types of information: (1) header, (2) equipment requirements, (3) formula, and (4) procedure. Headers provide information about the source and version of the recipes, such as recipe and product identification, author, issue, and date and any other pertinent batch information such as MSDS or hazard or toxicity issues specific to the product or processing activities and the like. Equipment requirements specify the type and size of equipment needed, such as glass lining required and 20,000-liter vessel required. Formulas are sets of parameters, such as types and quantities of ingredients, durations, and process condition set points, that distinguish the products defined by procedures. Procedures define and order the actions to be performed and the associated control requirements necessary for making a class of products in the batch process.

The procedure defines the generic strategy for producing a batch product. A procedure is made up of unit procedures.

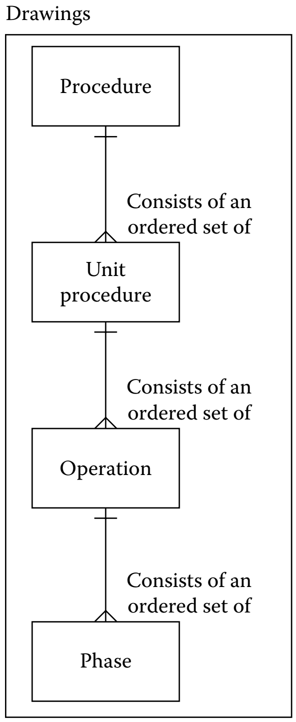


FIG. 8.3c
Procedure model.

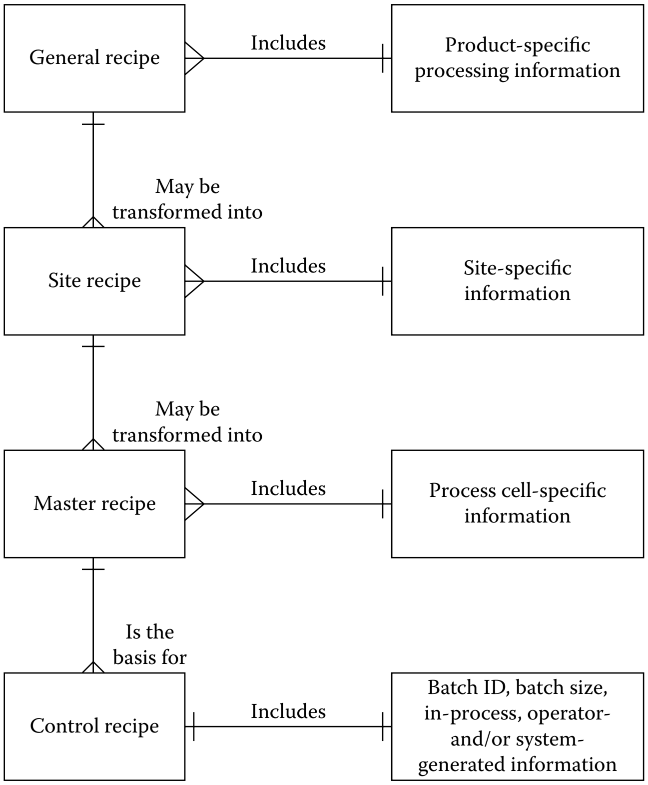


FIG. 8.3d
Recipe model.

Unit procedures are an ordered set of operations that causes a continuous production sequence to occur in a unit. An operation is an ordered set of phases that defines a major processing sequence. And operations are made up of phases that accomplish a process-oriented task. Phases may include the control steps or control instructions that execute the base-level control. The relationship is diagrammed in Figure 8.3c.[1]

The actual executable logic of the control is linked via the recipe information to a particular unit where the unit recipes, operations, or phases are run in the order determined by the recipe.

In the current generation of control systems, the recipe management function maintains a set of master recipes for various products and families of products. Specific information of the batch equipment and units that can run each operation within the procedure are contained within the control recipe. A master recipe is constructed from the site recipe using the formulas, procedures, and equipment-specific information. The master recipe is selected and accessed by the batch management activity, which converts it to a control recipe. This control recipe is the batch-specific recipe that is ready to run. The hierarchy of recipes is shown in Figure 8.3d.[1]

Utilizing process and product knowledge, a process analysis is performed and basic phases are determined. These basic phases, along with product knowledge from the laboratory chemist, are used to construct the general (corporate-wide) recipe. Plant knowledge (for example, raw material availability) from the plant site engineer is used to transform the general recipe to a site-specific recipe. Equipment knowledge (for example, what vessels and piping are available at the plant) is used to transform this site recipe into a master recipe. This master recipe is used as the basis for a control recipe when a batch is ready to be produced. The activities involved during recipe management are summarized in Figure 8.3e.

Production Scheduling Schedules serve as guides for production requirements in terms of availability of equipment, personnel, raw materials, facilities, equipment, and process capacity. The schedule generally has many of the following objectives:

1. Minimize the processing time
2. Minimize the deviation from a master production plan
3. Optimize the production of products within quality guidelines
4. Minimize energy costs
5. Minimize the usage of raw materials
6. Minimize rerun

Production scheduling accepts inputs such as the production plan and based upon a scheduling activity develops a production schedule that typically specifies batches/amounts to be produced, target trains/lines to be used, time targets,

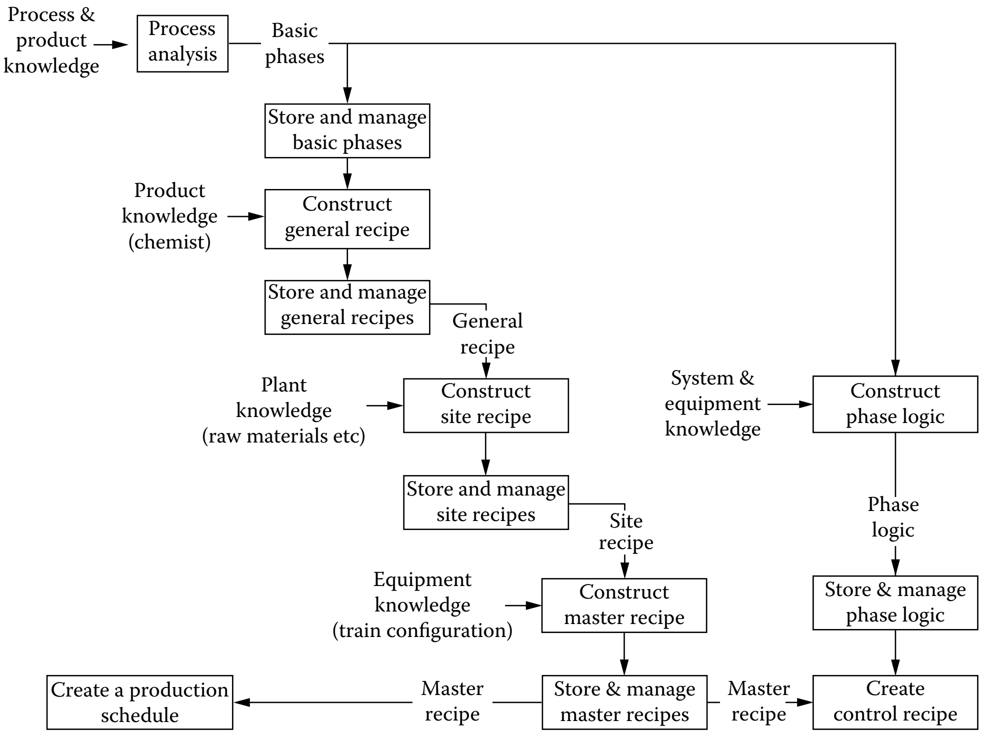


FIG. 8.3e
Recipe activities.

product dispositions, and resource constraints. In essence, the master production schedule answers what to schedule, when to schedule, and how much to schedule.

The production schedule further reduces the production plan, which was developed in the company management layer (process/production management) and directly drives the production of individual batches. The responsibility of the production scheduler is to develop a detailed time-based plan of activities to achieve the production targets set by the production plan. It needs to dynamically allocate a new schedule at any time. It should be feasible to reallocate or create a schedule automatically via some algorithm or manually via user intervention.

Schedulers can be implemented via any number of ways. Linear programs, expert systems, or other multivariable techniques have been used successfully.[8] The scheduler needs to provide a procedure or method for batch sizing and is the logical place where lot or even batch ID assignments are made. A production scheduling model is shown in Figure 8.3f.

As shown by the model, a dynamic scheduler accepts user inputs, master recipe information, and updates from

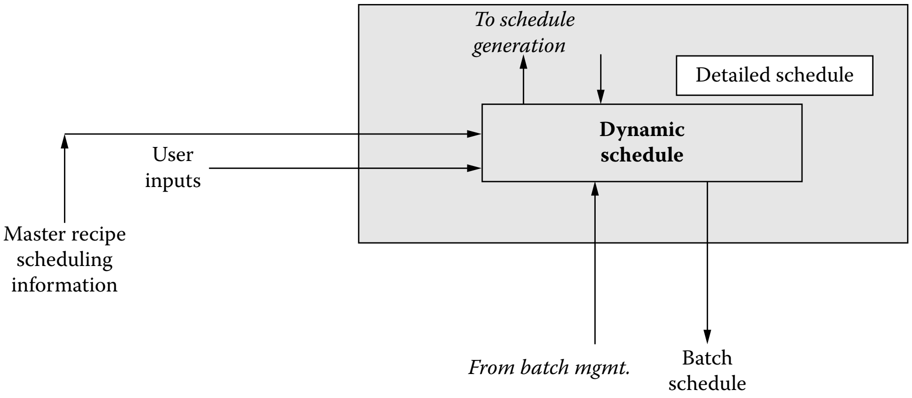


FIG. 8.3f
Production scheduling model.

process management to develop a batch or production schedule. The dynamic scheduler must be able to:

Organize a new schedule at any time
Provide for interactive scheduling
Allow for manual intervention
Determine the availability of resources
Provide for a method to carry out the schedule
Provide a procedure or method for the lot sizing along with a means to organize autonomous orders with this lot size
Determine the feasibility of the schedule

Information about times and availability of resources are key inputs to the scheduler. Because the basic data about the unit operations, master recipes, key times, resources, quantities, priority, and orders and their operations are required, it seems reasonable to conclude that the dynamic scheduler is a real-time activity. The dynamic scheduler requires non-real-time data from production planning and recipe management as well as real-time feedback data from the batch manager and lower-level control activities. The dynamic scheduler can contain some kind of optimizing capability. Batch management provides the updates of information to the scheduler in real time and also provides the status of material and equipment.

The user needs to be advised of the schedule situation continuously. Thus, reporting and status information are important. Information provided to the user can include the control recipe, the lot number, the product being produced, the amount being produced, the train/line being used, the status of the recipe, the mode of operation, priority, start times, and end times. The mode of the lot in the queue determines whether the recipe is to start automatically, semiautomatically, or manually. In the manual mode, stepping through the batch sequence is done by the plant operator via specific commands. In semiautomatic mode, the batch sequence is initiated by the operator and requires operator intervention when proceeding from one phase or operation to another. In the automatic mode, once the sequence is initiated, it can be repeated a predetermined number of times without any operator intervention. The mode of the batches in the queue also determines whether any information in the queue is operator alterable, and whether the control recipe has been bumped or interrupted by another control recipe of higher priority.

Batch History Management Batch history management is the subject of Part 4 of the ISA S88 standard. It has the following main functions:

1. Receiving and storing information from other parts of the overall batch control system on the individual batches
2. Producing output information from single batches, from several batches, or as an overview of multiple batches
3. Producing batch reports on the basis of output information
4. Maintaining production records and batch history archive by supporting data reduction, backup, and deletion features

Although many data logging and reporting techniques for batch processes are similar to those for continuous systems, production records have some significant requirements that may be different. Two needs are batch tracking (some may use the term lot tracking) and the batch end report.

A batch historian must collect and maintain integrated, identifiable sets of dissimilar data. Batch tracking is the collection of this data. It is generally event-triggered and typically contains the following related data:

Continuous process data (flows, temperatures, and pressures)
Event data (operator actions, alarms, notes)
Quality data (lab analysis, inspection notes)
Recipe formula data (quantities desired and used, set points, times)
Calculated data (totalizations, material usage, accounting data)
Manual entries with audit trail (location of change, operator of record)
Stage, batch, lot identification
Time/date stamps on all data

A batch end report may typically include a copy of the master and/or control recipe that was used to make the batch. This may not be identical to the original recipe because of operator modifications, equipment problems, and so on. Events such as alarms, operator instructions, and equipment status changes should also be logged. This log can be designed so that it will retain the total operational sequence chronologically with date and time. A trend chart can also be retained. A recipe expresses the desired approach by which a batch is to be made, while the batch report provides a record of how the batch was actually made. Batch management takes care of the recording and collecting of batch end reports, which are then archived to some other medium. Batch reports are a statutory requirement in some applications (for example, in the pharmaceutical industry); however, because the information is so valuable, it is being demanded in many other batch applications. A simplified batch history management model is shown in Figure 8.3g.

Batch history management is an activity that is not bound to the actual execution time of the batches and not bound to the equipment on which the batches are produced. It involves the process of sorting out the production records and batch end reports, because even a perfect batch may be undeliverable without batch records. Advances in relational databases allow the bridging of data between the process control of current batches and the histories of previous batches. The ability to use standard query language (SQL)-like calls to

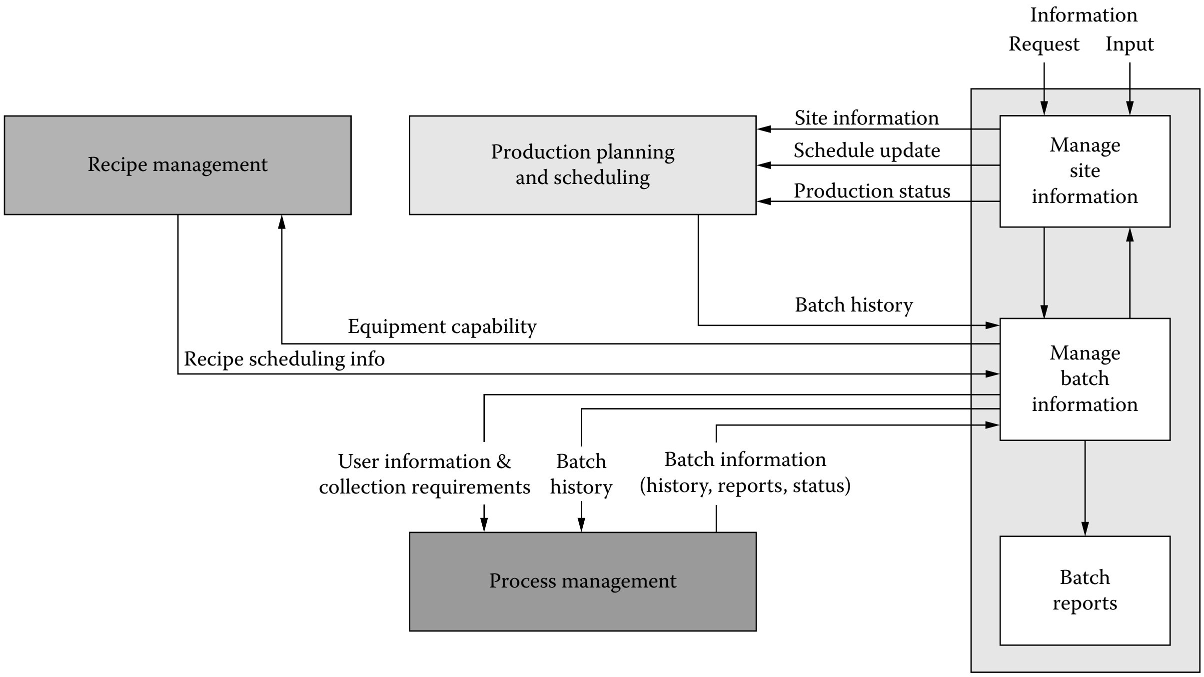


FIG. 8.3g
Batch history management.

access batch history allows new avenues to analyze and report batch histories. Other analysis techniques, such as statistical process control and statistical quality control (SPC and SQC), can be applied at this level.

Batch Management

Batch management interfaces with the user, recipe management, production scheduling, unit management, sequential control, and regulatory and discrete control in performing its functions of (1) recipe selection, transformation, and editing to manage batches; (2) initiation and supervision of batch processes; (3) management of batch resources; and (4) acquisition and management of batch information. A batch management model[1] is shown in Figure 8.3h.

Batch management includes basic functions to managing batches by using control recipe information, batch information, and equipment information to:

1. Select a master recipe from recipe management and transform it to a control recipe that can be used to run a batch
2. Assign a batch identification code
3. Time-stamp the control recipe when the batch identification is entered
4. Verify the information in a control recipe, such as its completeness and its ability to execute on the selected units
5. Maintain the control recipe until the batch is completed

Additionally, the batch management function:

1. Initiates control recipe execution based on time or event
2. Assigns and releases units, and updates their status
3. Distributes the parts of the control recipe to unit supervision
4. Starts the batch based upon start conditions and the detailed schedule, whether upon event or time
5. Regulates the distribution of operation or phases for execution
6. Reports and time-stamps events for information management
7. Allows users to alter normal processing
8. Maintains batch status information
9. Allows batch to be suspended, removed, and later recalled

Managing batch resources takes the detail schedule and master/control recipe information and provides for:

1. Dynamically predicting start and end times for batches and operations and tagging batch events
2. Maintaining the dynamic schedule of all batches, including their current state
3. Dynamically detecting resource requirements and resource availability
4. Updating production scheduling with modifications to the schedule

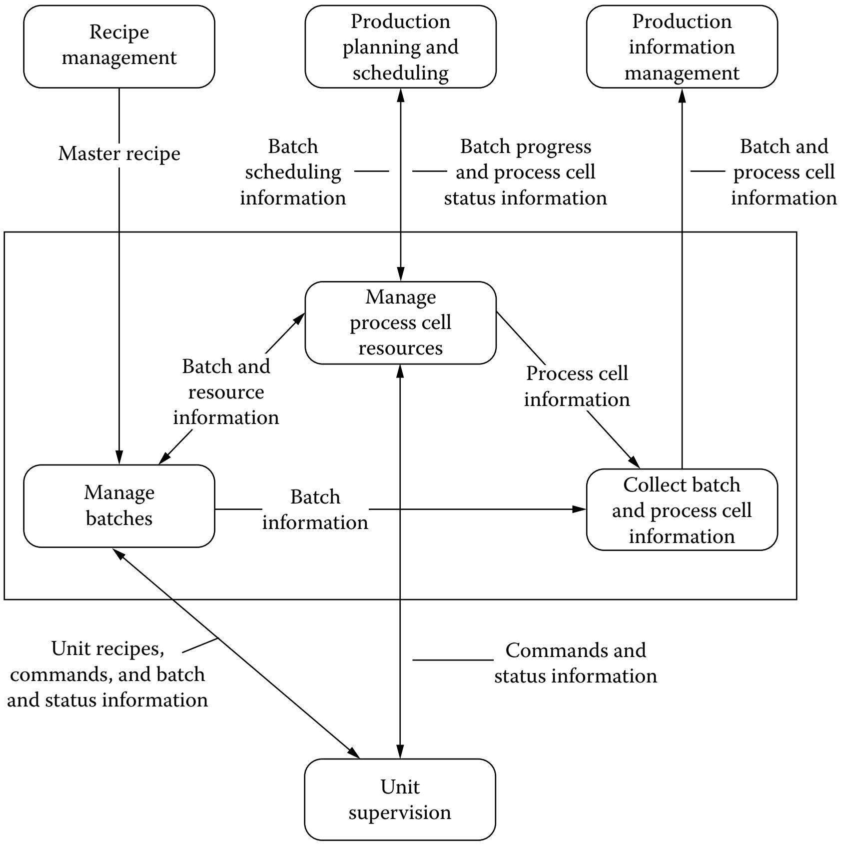


FIG. 8.3h
Batch management model.

Additionally, managing batch resources involves:

1. Receiving the detailed schedule
2. Dynamically detecting conflicts between resources needed and those that are available
3. Granting "permission" to unit to establish resource link
4. Arbitrating multiple requests for resources
5. Tracking and maintaining all batches, ready to run, in-progress, completed, or aborted

Finally, batch management provides functions to acquire and manage batch information, such as:

1. Collecting and reporting batch information by batch, operation, phase, time, or event
2. Providing in-progress or complete batch reports
3. Archiving and retrieving batch data in batch history

A batch report can include recipe data, snapshot process data, or batch data in response to the recipe or on an ad hoc basis necessary for documenting the batch. This includes operator modifications, historical trends, reports, and other information that may be available to the operator.

Unit Supervision

A unit consists of a physical grouping of equipment as well as the unit control functions required to carry out the execution of a batch. A process unit consists of a group of mechanical equipment; each piece performs, in a somewhat independent manner, a portion of the chemical process. Examples of process units are filters, batch reactors, heat exchangers, and distillation columns. Process control involves the actions required to perform the unit operations. Examples are charging, heating, cooling, agitating, reacting, discharging, and washing.

Unit supervision interfaces with the user, batch management, sequential control, regulatory control, and discrete control in performing its functions of (1) communication with other units, equipment modules, and control modules; (2) acquisition of resources; (3) unit procedure execution; and (4) exception handling. The unit supervision model as currently defined by the ISA standards committee is shown in Figure 8.3i.

Unit supervision requires certain information from batch management. The most important information is the recipe information required to run the unit. This recipe information contains the executable logic as well as formula information.

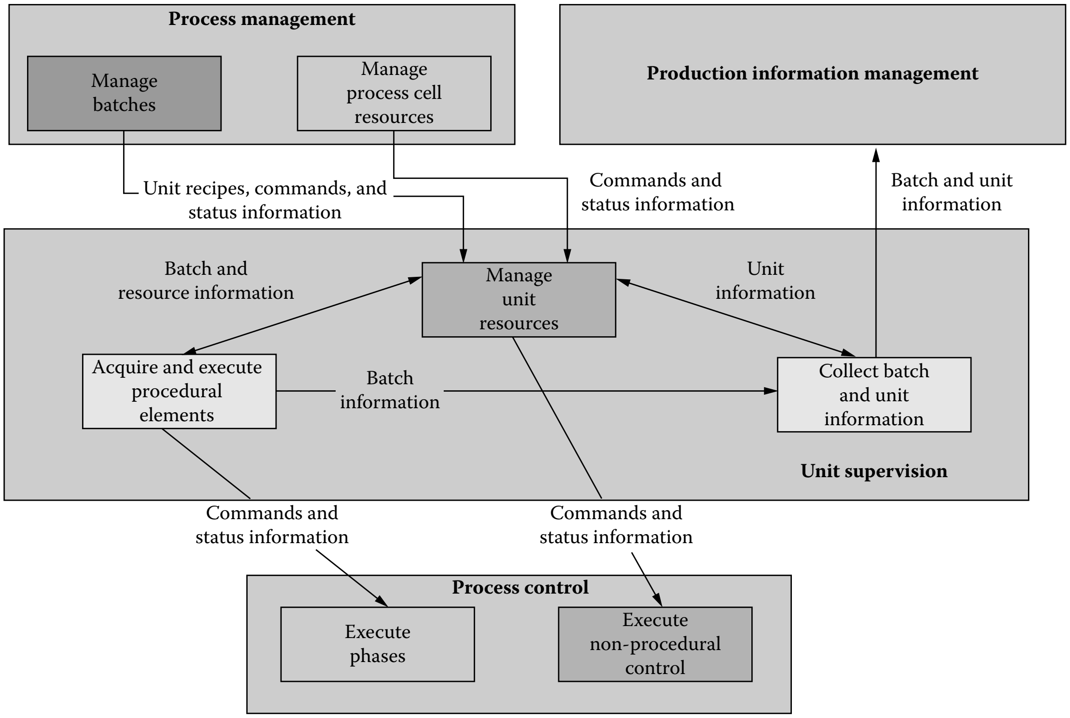


FIG. 8.3i
Unit supervision model.

Unit supervision performs the unit procedures, operations, and phases in the order they are to be run; performs exception logic when some abnormal condition arises during execution; and downloads parameters from the formulas to the control modules.

When more than one processing unit affects the status or use of a resource, the resource is designated as a common resource. Common resources almost always exist with parallel and series/parallel batch process structures. A common resource may either be required exclusively or can be shared. Unit supervision coordinates with other units to account for exclusive-use and shared-use resources.

Exclusive-use resources can be used by only one unit at a time. This means that some mechanism must be in place to prevent more than one batch from trying to use the resource at the same time. Also, the batch scheduling system must take this exclusive resource into consideration. Another problem is associated with distribution of risk in control systems. If a controller module containing the exclusive-use resource fails with no backup, all units utilizing the resource are affected. Shared-use resources are associated with units that can simultaneously use the common resource. A unit should not be able to deactivate the resource while other units are using it, and the capacity of the resource should not be exceeded by multiple users.

The engineering of the equipment into units greatly affects how shared and exclusive resources are handled, which translates into the overall modularity of the batch control. As a batch processes through one unit to the next, the use of the equipment providing the transfers must be coordinated.

Process Control

Sequential, regulatory, and discrete control functions interface directly with elements and actuators to cause changes in the process. Discrete control is concerned with maintaining the process states at a target value chosen from a set of known stable states. Regulatory control serves to maintain the measurements of a process as close as possible to their respective set point values during all events, including set point changes and disturbances. Sequential control sequences the process through a series of distinct states as a function of time.

These types of control functions are implemented using control modules and equipment modules. A control module *device* is an item of process equipment that is operated as a single entity and that may have multiple states or values. Discrete states are initiated (using hardware and software) to control discrete devices such as solenoid valves, pumps, and agitators. A control module *loop* is a combination of elements and control functions that is so arranged that the signals pass between elements for the purpose of measurement or control of the process variable. A proportional, integral, derivative (PID) control algorithm is a common control loop function. An example of *equipment module* control is the sequential control of dehydrator bed control valves in order to put one bed on-line while the other bed is being regenerated according to a time schedule. Most of today's process control systems utilize function blocks as the tools to describe and implement control modules.

Process interlocking and advanced control, in the forms of feedforward, predictive, or model-based control, are additional control functions that reside at this level and serve to achieve a higher level of automation to obtain additional benefits.

As opposed to continuous processing, additional control algorithms and control methodology are normally used in batch processing. Functions such as time-based PID (heat soak ramps), charging algorithms, sequencers, and timers are often required. Techniques such as enabling/disabling control functions based upon a phase state, enabling/disabling alarms on devices and loops, and employing antireset windup protection on PI or PID loops are commonly used. Batch processes tend to be device-oriented, while continuous processes are predominantly loop-oriented.

Safety Interlocking

Safety interlocks ensure the safety of operating personnel, protect the plant equipment, and protect the environment. These types of interlocks are initiated by equipment malfunction and usually cause shutdown. Often, a separate system is used in implementing the safety interlocks. This system includes the necessary redundancy and fault tolerance, and is independent from the other control functions. An example of a safety interlock is the stopping of a centrifugal compressor if its gear oil pump has failed, thus preventing mechanical damage.

Safety interlocking serves a different purpose from process interlocking or permissive interlocking. Process interlocking can be safety-related, but it is primarily associated with the process. An example of a process interlock is to stop charging a material if the agitator is not running. A permissive interlock establishes an orderly progression of sequences. An example would be to not allow the feeding of an extruder before the barrel temperature has reached a minimum value.

ENGINEERING

The effort to automate a batch facility is done to achieve operational benefits. Generally, benefits can be achieved through (1) increased production, (2) reduced utility costs, and (3) higher product quality. Many manufacturers use reduction of off-spec materials due to operator error, which affects all three of the above criteria, as justification for automation.

For batch reactors, benefits acquired through increased production through automation can be achieved via reduced batch cycle times, minimized turnaround time between batches, and better scheduling of reactors for like products. Decreased cycle times for each reactor are certainly measurable and are easy to quantify. A reactor cycle can be thought of as having four separate steps:

1. *Charge times:* Ingredients must be metered accurately and be added to the mixers or reactors as fast as possible. Dribble charging and other anticipatory methods are used to accurately charge the correct amount of an ingredient.
2. *Reaction times:* Good temperature control and accurate pressure detection are essential in order to produce consistent reactions batch after batch. Liquid chromatographs measuring molecular weight distributions and many other analyzers have been used to determine the end of the batch reaction.
3. *Dump time:* Reactor contents are emptied. Level devices or NMR devices have been used here to determine when the batch has been dumped completely.
4. *Turnaround time:* After a batch is dumped, the time until the next batch can be charged should be minimized.

Benefits of automation acquired through reduced utility costs can be achieved by minimizing heating medium usage and minimizing cooling medium usage. Benefits of automation acquired through higher product quality are much more difficult to define. They would normally translate to higher customer satisfaction and less reject material. These benefits can be achieved by achieving tighter control of the batch reactors and higher accuracy and consistency in charging the raw materials.

Once the architecture of the plant is known, the control strategies can be planned. An example of a batch plant is shown in Figure 8.3j.

In his paper presented at the World Batch Forum technical proceedings in 1998, Christie[9] provided his nine points for batch automation design:

1. Understand the process before generating the design.
2. Don't implement until you have designed.
3. Get the user involved in the design.
4. Document the design.

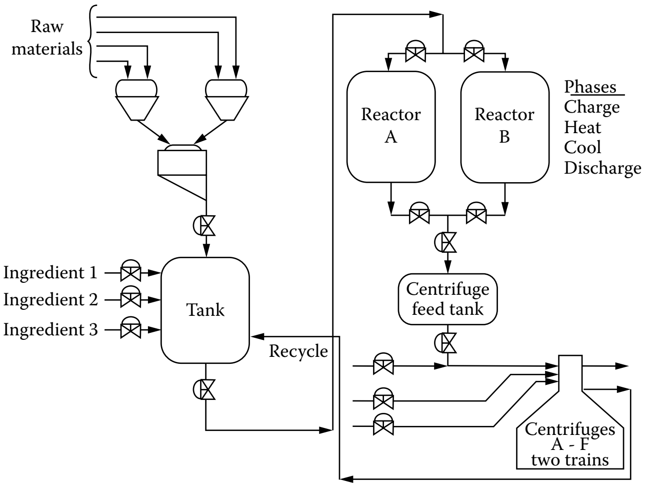


FIG. 8.3j
Batch process example.

5. Agree on the recipes.
6. I/O assignments are important.
7. Don't underestimate the state transition matrix.
8. Pay particular attention to exception handling, the hard part of batch.
9. Batch reports are not an add-on.

Inherent in the message is that the design based upon S88 concepts allows easier implementation and a high degree of modularity, translating into easier modifications and lower maintenance costs. Today's process automation suppliers have embraced the S88 standard concepts into their system architectures, facilitating more modular design and implementation.

With that, Fleming and Schreiber in their 1998 World Batch Forum paper on batch processing design[10] counsels the designer to consider process segmentation so as to:

Identify the process cell first
Identify the units after the boundary of the process cell has been identified
Identify the equipment modules in the process cell and units
Identify control modules in all other modules
Verify equipment module boundaries against defined procedural elements
Verify unit boundaries against defined procedural elements
Finally, verify the process cell boundary against defined procedural elements

The key to design is to segment the process properly so that the controls can be equally identified and segmented. Poor segmentation results in a batch control system that is difficult to support and enhance. A batch control system that fails to exploit the inherent flexibility of the plant and recipe development that requires the assistance of control system developers will result in an implementation that cannot be flexible, resulting in inflexibility of the organization to respond to product needs.[11]

Sequential control functions can be diagrammed to express the logic of batch control. Such diagrams include paradigms such as ladder diagrams (change-oriented), matrix diagrams (state-oriented), flow diagrams (flow-oriented), petri-nets (data/flow-oriented), Gantt charts (state-oriented), and sequential function charts (GRAPHCETS). Many of these tools are listed as possible recipe formats.[12,13] Most all process automation systems claim to be S88 based. In that regard, most, then, utilize function blocks for control modules and sequential function charts for procedures, unit procedures, or operations as their engineering tools to implement batch control automation.

Figure 8.3k is an example of a batch procedure for the batch process shown in Figure 8.3j. It shows the time sequence of activities in making a specific product. Between the active-sequence phases, idling, holding, and waiting states can occur. These states can allow information exchanges with other batch process units or can receive directives from operators. Fail-safe, emergency, or exception handling states can also be defined along with these phases.

Formula information from the recipe is loaded into controllers at the proper time when executing the phases. Formula information is a list of parameters, such as temperature set points, flow set points, quantities or totalization set points for ingredients, transition times, controller modes, and whatever else is needed by the phases.

Finally, at the end of the design and implementation aspects of automation design comes testing of the batch control

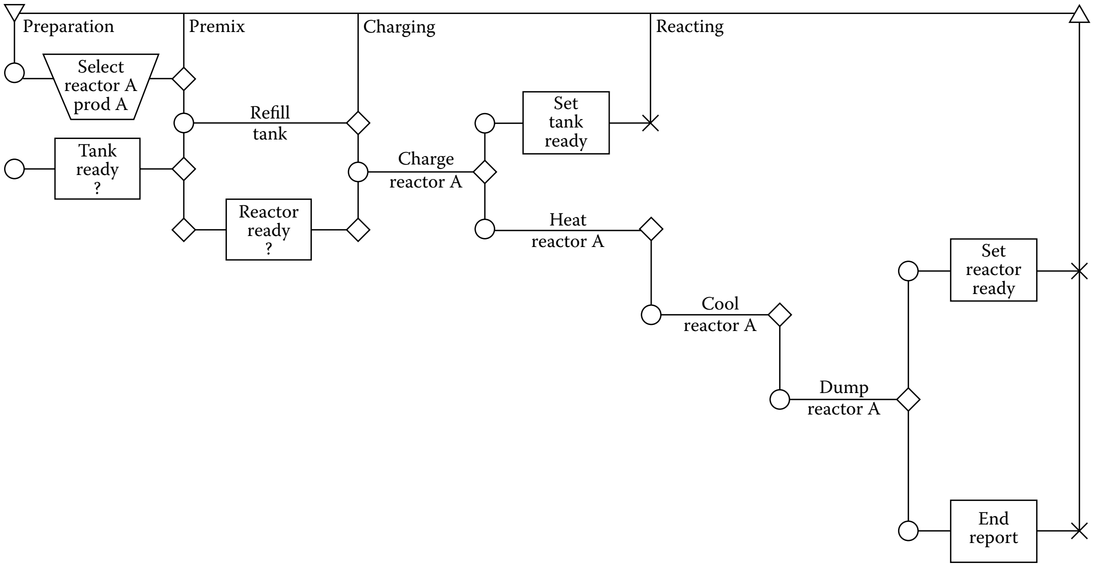


FIG. 8.3k
Batch procedure time sequence.

and automation. Pillai's[14] phase testing guidelines state to clearly define the modules, clearly define the testing strategy, clearly define the test plan, perform the tests, document the test results, and initiate change control.

When the plant configuration is known, the steps necessary to engineer the control of the process can be defined. Figure 8.31 shows the engineering steps performed by the control engineer. However, many other kinds of disciplines

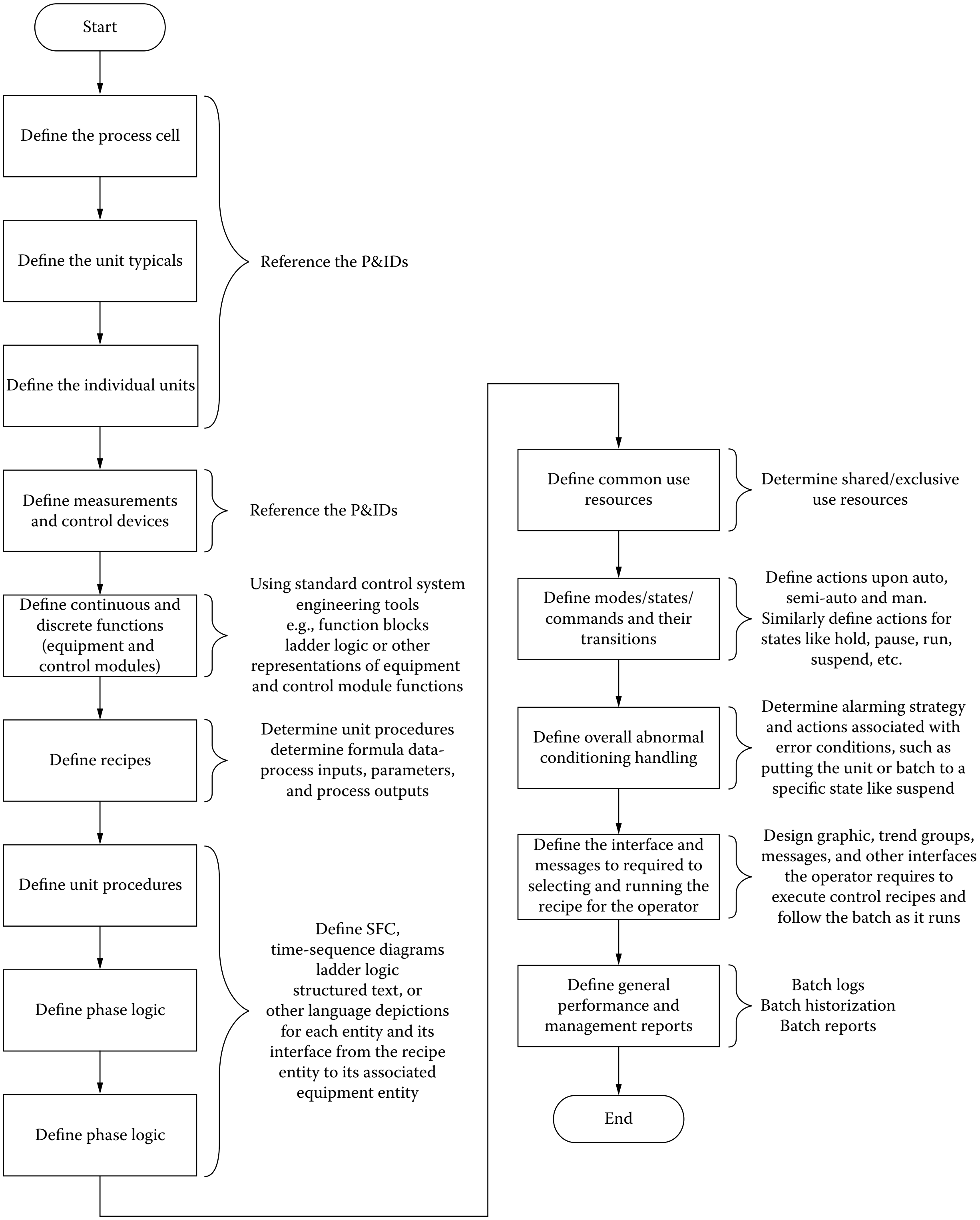


FIG. 8.3l
Batch process control flowchart.

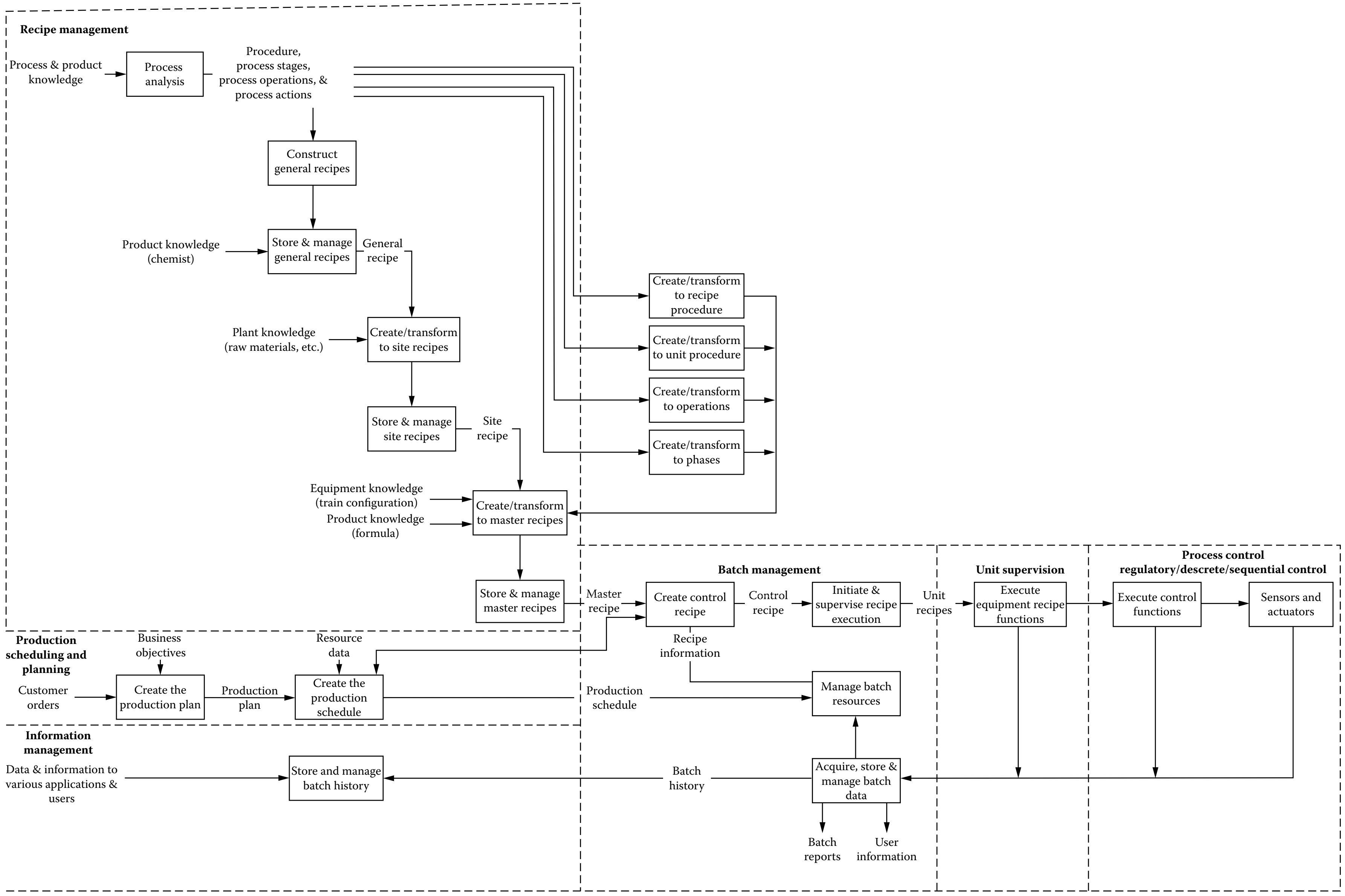


FIG. 8.3m
Consolidated control activities.

are involved in the entire process. Figure 8.3m illustrates the engineering activities required at various levels by people with different backgrounds and qualifications.

Process and product knowledge from central engineering provides the input to construct basic phases for the procedure of making a product. This knowledge combined with product knowledge of the research or development chemist is used to construct general recipes. Plant knowledge provides the information to convert the general recipe to a site-specific recipe. Plant engineers also have the equipment knowledge necessary to construct phase and operation logic and to transform the site recipe into a master recipe by knowing the specific trains/lines on which the product can be made. Process control engineers then are able to use both system knowledge and process knowledge to create the control recipe, which is system-specific and is designed to execute a batch. The control engineers also use their knowledge to design, configure, and implement the lower-level control functions required to provide process control. Management information systems (MIS) people implement process/product management and turn business objectives and customer orders into a production plan that is used to create a production schedule. Batch history archives may be maintained by MIS or plant computer system personnel. Therefore, overall batch control implementation requires the talents of many disciplines.

CONCLUSIONS

Execution of a batch process is a sequence of processing stages, operations, or actions. These stages, operations, or actions are independent process-oriented events within the overall batch operation, such as charging, heating, reacting, agitating, cooling, and discharging. These stages, operations, and actions translate to unit procedures, operations, and phases, which are defined by boundaries that define safe or logical points where one can charge an ingredient or direct a finished product to a different downstream vessel. The set of sequences under which these are performed constitutes the batch operation.

The user must be considered at each level of the control activity. The operator interface allows the operator to interrupt automated commands received from higher activity levels and to enter information or to command the process directly. Data can enter the system at any level. Also, an output can be generated at any level where it is needed and then be transmitted to the next higher or lower levels.

The models defined by the ISA batch control standards committee are designed to be valid for batch processing facilities, regardless of the level of automation involved. The control activity levels are designed to be collapsible in case those functions are not applicable. From plants whose recipes are on paper and whose control is done by manual valves to fully automated paperless batch manufacturing facilities, the descriptions and terminology of the batch process should be understandable by all who participate in it, regardless of the industry or the products involved. The goal is a modular, maintainable, flexible batch automation implementation.

References

1. ISA S88.00.01, "Batch Control, Part 1: Models and Terminology," Research Triangle Park, NC: Instrument Society of America (ISA), 1995.
2. IEC 61512-1, "Batch Control, Part 1: Models and Terminology," International Electrotechnical Commission (IEC), 1997–08.
3. ANSI/ISA-88.00.02, "Batch Control Part 2: Data Structures and Guidelines for Languages," Research Triangle Park, NC: Instrument Society of America, 2001.
4. IEC 61512-2-2002, "Batch Control, Part 1: Models and Terminology," International Electrotechnical Commission, 2002.
5. ANSI/ISA-88.00.03, "Batch Control Part 3: General and Site Recipe Models and Representation," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society (ISA), 2003.
6. ISA-88.00.04 Draft 3b, "Batch Control Part 4: Production Records," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 2003.
7. Christie, D., "The Top-Down Approach to Successful Process Control Projects," *Control* October 1989.
8. Reklaitis, G. V., "Scheduling Approaches for the Batch Processing Industries," Proceedings of the World Batch Forum, Newton Square, PA, May 1995.
9. Christie, D., "A Methodology for Batch Control Implementation — Real World Lessons," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.
10. Fleming, D. W., and Schreiber, P. E., "Batch Processing Design Example," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.
11. Hancock, J., and Hopkinson, P., "A Case History of the Implementation of an S88-Aware Batch Control System," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.
12. ISA-TR88.0.03, "Possible Recipe Procedure Presentation Formats," Research Triangle Park, NC: Instrument Society of America, 1996.
13. Emerson, D., "What Does a Procedure Look Like? The ISA S88.02 Recipe Representation Format," *Proceedings of the World Batch Forum,* San Diego, CA., April 1999.
14. Pillai, V., "S88 Phase Testing and Change Management," *Control*, March 1999.

Bibliography

Hedrick, J. L., and Severns, G., "Planning Control Methods for Batch Processes," *Chemical Engineering,* April 18, 1983, pp. 69–78.

Rosenof, H. P., and Ghosh, A., *Batch Process Automation,* New York: Van Nostrand Reinhold, 1987.

"Batch Control Systems," *ISA Transactions*, Vol. 28, No. 3, 1989.

Faccenda, J. F., and Baker, T. E., "An Integrated AI/OR Approach to Blocked Operations Scheduling," *Proceedings of the NPRA Computer Conference,* Seattle, WA, October 1990.

Fisher, T. G., "Batch Control Systems: Design, Application, and Implementation," Instrument Society of America, 1990.

Hogenson, D. C., "Automatic Scheduling of Batch Operations," *InTech,* April 1990, pp. 30–32.

Kompass, E. J., Witlock, S. K., and Williams, T. J., "Generic Control for Batch Manufacturing," *Proceedings of the Sixteenth Annual Advanced Control Conference,* sponsored by the Purdue Laboratory for Applied Industrial Control, Purdue University, and *Control Engineering,* West Lafayette, IN, September 1990.

Feldmeyer, E., Alsup, M., and Burns, H., "Recipe Interface Requirements for Flexible Batch Control Systems," *Proceedings of the Industrial Computing Conference,* Vol. 1, Instrument Society of America, Anaheim, CA, October 1991.

Ghosh, A., "An Approach to a Flexible and Easy to Use Batch Control System," *Proceedings of the Industrial Computing Conference,* Vol. 1, Instrument Society of America, Anaheim, CA, October 1991.

Hornbeck, D., "A New Approach to Batch Process Control," *Proceedings of the Industrial Computing Conference,* Vol. 1, Instrument Society of America, Anaheim, CA, October 1991.

Larson, K., "Flexibility Takes Center Stage in Batch Process Automation," *Control,* November 1991, pp. 24–33.

Massey, L. E., "The Development of Modular Batch Automation Techniques," *Proceedings of the Industrial Computing Conference,* Vol. 1, Instrument Society of America, Anaheim, CA, October 1991.

Nowicki, P. L., "Transforming Recipes is the Key to Flexible Batch Processing," *Proceedings of the ISA/91*, Vol. 46, Part 2, Instrument Society of America, Anaheim, CA, October 1991.

Alsup, M., "Recipe Configuration for Flexible Batch Systems," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Tom, T. H., "DCS Selection for Flexible Batch Automation," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Bullotta, R., "Designing Effective User Interfaces for Batch Processes," *Proceedings of the World Batch Forum,* Phoenix, AZ, March 1994.

Offenbacher, F. M., and Smith, E., "Integrated Production Scheduling," *Proceedings of the World Batch Forum,* Phoenix, AZ, March 1994.

Van de Pol, P., "Real-time Production Scheduling in a Multipurpose Batch-Oriented Environment," *Proceedings of the World Batch Forum,* Phoenix, AZ, March 1994.

Webb, M., "Computer System Implementation, Batch Standards, and Validation," *Proceedings of the World Batch Forum,* Newton Square, PA, May 1995.

Rosenof, H., "Dynamic Scheduling for a Brewery," *Proceedings of the World Batch Forum,* Newton Square, PA, May 1995.

Chappell, D. A., "Recipe Requirements for Dissimilar Manufacturing Trains: The Ultimate Complexity," *Proceedings of the World Batch Forum,* Newton Square, PA, May 1995.

Deitz, D., and Snyder, B., "Automating a Modern Bio-Manufacturing Plant: A Case Study," *Proceedings of the World Batch Forum,* Toronto, ON, May 1996.

Gudaz, J., "Batch in the Context of Field-Based Control," *Proceedings of the World Batch Forum,* Toronto, ON, May 1996.

Nomikos, P., "Detection and Diagnosis of Abnormal Batch Operations," *Proceedings of the World Batch Forum,* Toronto, ON, May 1996.

Korkmaz, B., and Parapar, R., "Equipment Flexibility in a Pilot Plant Operation," *Proceedings of the World Batch Forum,* Toronto, ON, May 1996.

Clark, P., Grove, P., Liu, B., and Joglekar, G., "Using Simulation to Verify the Decision Logic of Batch Control Systems," *Proceedings of the World Batch Forum,* Toronto, ON, May 1996.

Uhlig, R. J., "Batch Control in a Resin Plant," in *ISA Practical Guide Series for Batch Control,* Research Triangle Park, NC: ISA, 1996.

Nisenfeldh, A. E., and Leegwater, H. (Eds.), *Batch Control,* Research Triangle Park, NC: ISA, 1996.

Berber, R., "Control of Batch Reactors: A Review," *Trans. IChemE.,* 74(A), 3–20, 1996.

Sano, Y., Watanabe, H., and Yoshida, M., "Design Approach to Developing the Batch Control System," *Proceedings of the World Batch Forum,* Houston, TX, April 1997.

Tom, T. H., "Handling Product Related Exceptions Using the S88 Concept of Modular Batch Automation," *Proceedings of the World Batch Forum,* Houston, TX, April 1997.

Choi, J. Y., Mercure, P. K., and Rhinehart, R. R., "Optimization of a Batch Polymerization Reactor," *Proceedings of the World Batch Forum,* Houston, TX, April 1997.

Lawrence, F. B., and Robinson, E. P., "Optimization Procedures for Scheduling Batch Packaging Operations," *Proceedings of the World Batch Forum,* Houston, TX, April 1997.

Koning-Bastiaan, H., "Process Model and Recipe Structure, the Conceptual Design for a Flexible Batch Plant," *Proceedings of the World Batch Forum,* Houston, TX, April 1997.

Owen, J. M., "Evangelizing S88.01: Converting the Masses," *Proceedings of the World Batch Forum,* Houston, TX, April 1997.

Wooley, K., and Christie, D., "Achieving Increased Batch Plant Productivity by Application of a Multi-Level DCS Solution," *Proceedings of the ISA,* Houston, TX, October 1998.

Kawano, K., "Applications of Batch Progress Prediction by Rolling Scheduling to Operation Support System," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Emerson, D., and Himono, R., "Automated Scheduling in a Polymer Plant," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Akesson, K., and Tittus, M., "Modular Control for Avoiding Deadlock in Batch Processes," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Poulsen, L., "Paperless Batch Production — A Practical Example from the Pharmaceutical Industry," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Musier, R., Sundaram, S., and Vardy, J., "The Recipe Model: New Advances in Using Life-Cycle Principle in the Batch Industry R&D and Manufacturing Environment," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Korkmaz, B., and Pillai, V., "Unit Supervision — A Key Component in Process Management and Process Control," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

LeBlanc, L., and Malenfant, M., "How S88 Provides Consistency in Design and Terminology beyond Batch Processing," *Proceedings of ISA,* 1998.

Jensen, B., and Lagonikos, P., "Automation Philosophy of a Pilot Plant for Active Pharmaceutical Intermediates/Ingredients," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Brooks, T., "Batch-Based Historians," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Korkmaz, B., Pillai, V., and Srinivasan, R., "Guidelines for Exception Handling Design for Batch Processes," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Zarichniak, S., "The Inherent Reusability of Batch Phases and Devices," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Crowl, T., and Wolin, D., "The Paperless Batch Plant," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Kandadamy, S., and Ibrahim, E., "Automation of a Lube Oil Additives and Blending Plant Using an S88.01-Consistent Batch Software — a Case Study," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Kerrick, S., Smith, S., and Charpientier, L., "A Most Unusual Approach to Implementing S88 Recipes," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Nishitani, H., and Niwa, T., "Modeling & Simulation of Pipeless Batch Plants," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Ito, H., and Emerson, D., "Integrated MMI for Batch Process Operations," *ISA TECH,* 1999.

Chen, L. W., Salisbury, R., and Kamal, S. Z., "Scheduling in Batch Process Plant Perspective," IMS, 1999.

Adler, D., "Trends in Manufacturing," Interkama Technical Conference, Dusseldorf, Germany, 1999.

Liefeldt, A., Lohl, T., Stobbe, M., and Engell, S., "Simulation and Scheduling of Recipe-Driven Batch Processes Based on a Single Data Model," Interkama Technical Conference, Dusseldorf, Germany, 1999.

Koning-Bastiaan, H., Schreil, C., and Roy, C., "Automated Execution of Bio Tech Batch Manufacturing," Interkama Technical Conference, Dusseldorf, Germany, 1999.

Parshall, J. H., and Lamb, L. B., "Applying S88: Batch Control from a User's Perspective," Research Triangle Park, NC: ISA 1999.

Brandl, D., "Using Corporate Recipes to Accelerate Time to Market," Aspenworld 2000 Technical Session, Orlando, FL, February 2000.

Brandl, D., "Batch Automation Is Coming of Age," *Chemical Processing,* September 2000.

Huzmezan, M., Gough, B., Kovac, S., and Le, L., "Advanced Control of Batch Reactor Temperature," *Proceedings of the ISA Expo*, 2000.

Drillenburg, C., "Tracking and Tracing on an ISA S88 Foundation," *Proceedings of the ISA Expo,* 2000.

Benton, A., "Batch Control Application Frameworks and Reuse," *Proceedings of the World Batch Forum,* Atlantic City, NJ, April 2000.

Fisher, T. G., "Batch Control: Applying the S88.01/IEC 61512-1 Standards," *Proceedings of the World Batch Forum,* Atlantic City, NJ, April 2000.

Brandl, D., Emerson, D., and Ward, B., "Batch Schedules and Production Schedules: Which Should You Use?" *Proceedings of the World Batch Forum,* Atlantic City, NJ, April 2000.

Roy, C., and Wyshak, G., "Implementing Automated Batch Control of a Biotechnology Manufacturing Facility Using S88.01 Concepts: A Case Study," *Proceedings of the World Batch Forum,* Atlantic City, NJ, April 2000.

Himono, R., Mano, K., and Takahata, K., "An Integrated System with Batch Functions and Front-End Scheduling Based on S88: Application to Beverage Plant," *Proceedings of the World Batch Forum,* Atlantic City, NJ, April 2000.

Koning-Bastiaan, H., "Production Information Management," *Proceedings of the World Batch Forum,* Atlantic City, NJ, April 2000.

Stapper, H., and Verhagen, A. M., "Benefits of Advanced Engineering Methodologies for the Design and Support of Batch Projects in the Pharmaceutical Industry," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Rickard, J., "Combining Optimal Off-Line Scheduling with On-Line Supervisory Control," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Hauff, T., "Batch Manufacturing: Status and Future Challenges," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Ghosh, A., "Maximizing the Potential of Batch Process Control," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Morse, C., "S88 Case Study for Improved Production Flexibility, Safety, and Environment," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Craig, L., "S88 Process Modularization," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Vanhove, G., "Tracking and Tracing on an ISA S88 Foundation," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Amkreutz, R., and van Beurden, I., "Safety in Batch Production," *Proceedings of the ISA Expo,* 2001.

Jensen, B. A., "Six Sigma and S88 Unite for Batch Automation Productivity Improvement," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Christie, D. A., "Automatic Adaptation of Batch Logic to Changing Field Automation," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Wilson, P., "Batch Application Migration," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Branch, T., and Ray, T., "Batch Automation Project Increases Production," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Crowl, T. E., and Heckmanski, J., "Design Methods to Defer Costs on Batch Projects," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Kovacs, K., "Designing Batch Systems for e-Manufacturing," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Patnaik, S., "Master Batch Record Management," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Johnson, L., and Roy, C. "Use of Web Technologies in Batch Management," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Givens, M., and McDonald, A., "Repeated S88 Success Yields Cost Reductions at Large Consumer Products Company," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Christie, D., "Life's a Batch," *InTech,* July 2001.

Blanchard, I., "Batch Control Systems Current and Future," *Proceedings of the World Batch Forum*, Woodbridge Lake, NJ, April 2002.

Årzén, K.-E., and Olsson, R., "Exception Handling in S88 Using Grafchart," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2002.

Marklew, C., "Electronic Batch Records and Flexible Batch Reporting," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2002.

Staus, R., "Improved Temperature Control in Batch Production Systems," *Proceedings of the ISA,* 2002.

Holy, R., and Pozivil, J., "Batch Control System Project for a Pharmaceutical Plant," *ISA Transactions,* 2002.

Sanchez, A., Rotstein, G., Alsop, N., Bromberg, J. P., Gollain, C., Sorensen, S., Macchietto, S., and Jakeman, C., "Improving the Development of Event-Driven Control Systems in the Batch Processing Industry: A Case Study," *ISA Transactions*, 2002.

Lorenzo, D., and Pérez, N., "Applying Model Predictive Control to a Batch Process," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Pettus, N., and Wolf, D., "Automated Batch Scheduling and MES Integration Using a Hierarchical Based, Batch Control Software Architecture," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Arnold, J., and Brandl, D., "Automating the Manufacture of Highly Energetic Organics Using the S88 Model," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Hunter, M., "Real-Time Batch Control and Materials Management in Perfume Production," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Clark, S. C., "S88 Design and Implementation Case Study for a Complex Bulk Pharmaceutical Batch Process," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Brill, E., and Yulevitch, O., "Targeting Batch Outcome Using Predictive Control," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Parks, J., and Yuki, T., "Using Alarm and Event Analysis to Achieve Batch Productivity Improvements," *Control Solutions,* February 2003.

Bonvin, D., and Srinivasan, B., "Optimal Operation of Batch Processes via the Tracking of Active Constraints: A Case Study," *ISA Transactions,* Research Triangle Park, NC: 2003.

Gough, B., Kovac, S., Devito, L., and Quick, D., "Model Predictive Control of Batch Temperature," *Proceedings of the World Batch Forum,* Chicago, IL, May 2004.

Fu, C., "Functions to be Considered in Batch Material Transfer Controls," *Proceedings of the World Batch Forum,* Chicago, IL, May 2004.

Brun, T. A., "Transfer Lines as Unit in an S88 Frame," *Proceedings of the World Batch Forum,* Chicago, IL, May 2004.

Marjanovic, O., Lennox, B., Lovett, D., and Sandoz, D., "Statistical Process Monitoring of Industrial Batch Processes," *Proceedings of the World Batch Forum,* Mechelon, Belgium, October 2004.

Martin, E., and Morris, J., "Monitoring Multi-recipe Batch Manufacturing performance," *Proceedings of the World Batch Forum,* Mechelon, Belgium, October 2004.

Mecchretto, S., "Integrated Batch Processing: A Recipe fpr Advanced Manufacturing in the Process Industries," *Proceedings of the World Batch Forum,* Mechelon, Belgium, October 2004.

8.4 Batch Processes and Their Automation

A. GHOSH (1995, 2005)

INTRODUCTION

Batch automation and batch control are discussed in this section from both a process characteristics and a modeling point of view. This discussion covers the various control functions, distributed control systems (DCS) capabilities, and reliability aspects. Other sections in this chapter also deal with the subject of batch control. Section 8.3 describes terminology, and Section 8.8 is devoted to the control of batch reactors.

In a continuous process, such as the distillation of crude oil or the manufacture of bulk chemicals and fertilizers, the product is manufactured on a continuous basis. In batch processes, used in the food, pharmaceutical, and fine chemical industries, products are manufactured in batches. Batch processes are sequential, where the control functions (called phases), such as charging, mixing, heating, cooling, and testing, are performed in an ordered fashion. Each phase may require many process steps, such as the opening and closing of valves, starting and stopping of pumps, and setting and resetting of control loops. In addition to the normal step-by-step control actions, batch process control requires many other functions; for example, responding to abnormal or failure conditions, keeping batch records, maintaining recipes, and scheduling batches.

Batch process control is a complex task rather than a difficult one. A difficult problem is one that is mathematically or scientifically difficult to solve and usually has a single solution. A complex problem, on the other hand, is logistically more challenging. A batch process is characterized by numerous interrelationships and constraints, mutually exclusive objectives, and many possible solutions. Most real-life problems tend to be complex. The solution of a complex problem involves its progressive decomposition into simpler functional modules and their proper integration.

This section deals with batch control problems and their solutions in general. Specific details on batch reactor control can be found in Section 8.8.

BATCH CONTROL STANDARDS

The ANSI/ISA-88 (IEC 61512) batch control standard[1–3] is providing significant benefits to users and suppliers of batch control systems worldwide. The standard is in three published parts, while the fourth part is under development (Table 8.4a).

Part 1: Models and Terminology

Part 1 defines standard terminology and a number of models for batch control. The key models are procedure, physical, and control activity (Figure 8.4b). The terminology, structures, and concepts used in the standard are affecting everyone in the batch control business. Most major suppliers have adopted standard terminology and have designing batch control systems with a modular set of functions and hierarchy based on the control activity model. This modularity allows for easier integration of third party-packages to do functions such as production planning, scheduling, and production information management. It also makes a control system easier to integrate with production management and business planning systems.

TABLE 8.4a
Batch Control Standards

Batch standard	Year published	US standard	International standard	Scope
Part 1	1995	ANSI/ISA-88.00.01	IEC 61512-01	Models & terminology
Part 2	2001	ANSI/ISA-88.00.02	IEC 61512-02	Data structures & language guidelines
Part 3	2003	ANSI/ISA-88.00.03	IEC 61512-03	General & site recipes

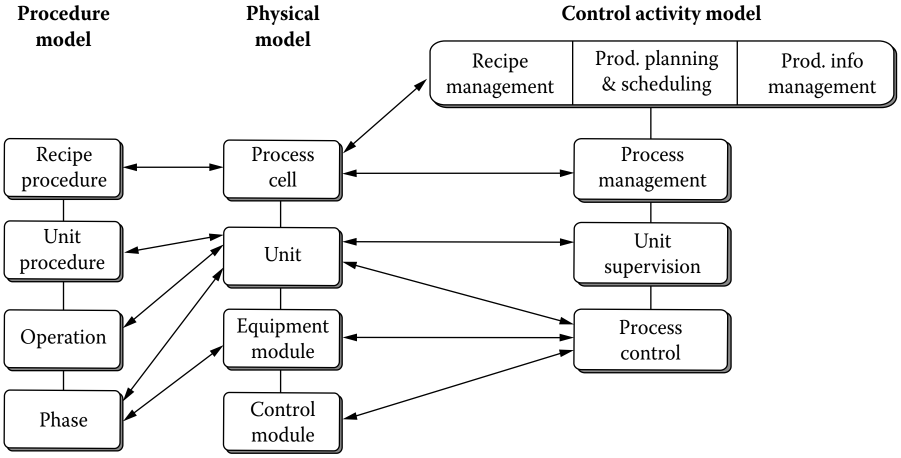


FIG. 8.4b
Batch standard models. (Note: Arrows show relationships, not data flow.)

The main benefits of Part 1 of the batch control standard are:

- Improved communication between suppliers and users of batch control
- Easier identification of end-user needs
- Straightforward recipe development
- Reduced cost of automating batch processes
- Reduced life-cycle engineering effort

Part 2: Data Structures and Language Guidelines

The ISA-88 Part 2 standard is in three parts: data models, information exchange tables, and procedure function charts. The data model section provides formal representation of entities specified in Part 1 of the standard, such as recipe, equipment, planning and scheduling, and information management using Universal Modeling Language (UML) notation. The information exchange section uses SQL relational tables to specify exchange requirements between recipes, process equipment, schedules, and batch production.

The final part of the standard deals with a graphical representation of procedures, such as master and control recipes and using Procedure Function Chart (PFC) notation (Figure 8.4c). PFC is somewhat similar to Sequential Function Chart (SFC) notation as defined in IEC 61131-3 standard.[4] PFC notation addresses procedural control and execution, while SFC notation was developed primarily for state machines. PFC notation meets the requirements of recipe procedures better than SFCs.

PFC notation is intuitive and easy to follow. Users familiar with SFC notation will find many similarities between the two. However, explicit specification of functions such as process equipment allocation and synchronization of phases makes PFC easier to follow than SFC notation.

Providing tools for information exchange, as specified in the standard, allows suppliers to design more modular batch control systems. This will reduce the high cost of maintenance and version upgrades. It is felt that PFC notation will soon become the common way for representing procedural elements in a recipe.

Part 3: General and Site Recipes

Recipes are of four types: general, site, master, and control (Figure 8.4d). A general recipe contains generic information for the manufacture of a product and does not include equipment or site-specific information. This is the first recipe that may be generated when a new product has been developed in a pilot plant.

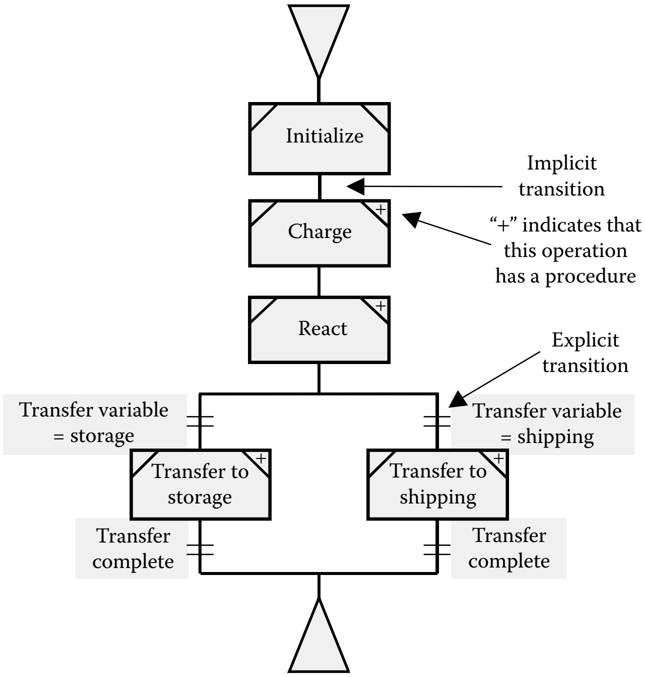


FIG. 8.4c
Unit procedure using procedure function chart (PFC) notation

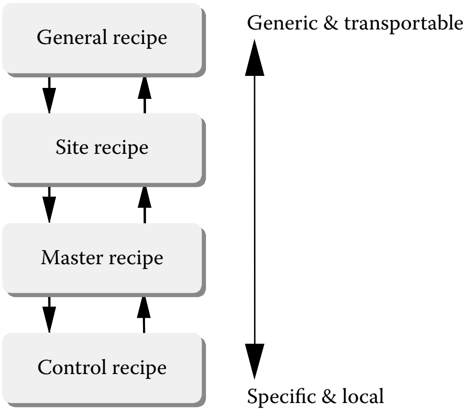


FIG. 8.4d
Recipe model.

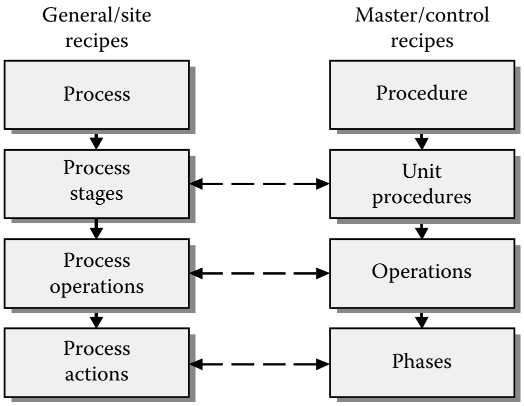


FIG. 8.4e
Mapping between general/site recipes and master/control recipes.

In the increasing complexity of global manufacturing, many process manufacturers find it challenging to maintain a single definition of a product in different manufacturing facilities. General and site recipes provide an equipment-independent means of describing batch manufacturing processes, accelerating both time to market and time to volume production. A standardized general recipe meets this challenge as a central repository for product manufacturing information. General and site recipe functions were identified in Part 1 of the ISA-88 batch control standard, however, little was specified about them in either Part 1 or 2.

Part 3 of the batch standard, which deals with these functions, was published in 2003.

A general recipe is an enterprisewide recipe for a manufactured product that serves as the basis for both site and master recipes. Chemists and chemical engineers with intimate knowledge of both the product's chemistry and processing requirements are generally responsible for creating it. It identifies raw materials, their relative quantities, the required processing, and the order of processing. It may define the processing capabilities required, such as cooling or heating, or the generalized equipment requirements, such as glass-lined reactors, but does not define the specific equipment that may be used to manufacture the product. A general recipe may also serve as input for corporate production planning and standard costing procedures. It is usually the parent of site and master recipes.

A site recipe has the same structure as a general recipe, but the information in a site recipe is tailored for each target location. A site recipe may be modified for the local language, units of measure, regulations, and raw material variability. A site recipe may include only a part of a general recipe that is actually implemented on the site. For example, a single product may have intermediate materials manufactured at one site that are then shipped to a second site for final processing. In that case, each site recipe would be derived from only the portion of the general recipe actually required for the processing to be done at that site.

General and site recipes contain the same categories of information, such as header, formula, procedure, equipment requirements, and other information, as in master and control recipes. The procedure elements of general and site recipes map well with master and control recipes (Figure 8.4e).

Part 4: Production Records

This part of the batch control standard defines a logical data model and means of data exchange for production records containing information about batches and other production segments. The standard is under development.

The Benefits of Standards

The benefits of the ANSI/ISA-88 (IEC 61512) batch control standard reach beyond batch process control. Although the standard is primarily designed for batch processes, it is also being applied successfully in various manufacturing industries. This is because the standard is not just for software, equipment, or procedures. It is a way to conceptualize your production processes that helps you better design your plants and manufacture your products.

The data models and information exchange tables provide significant advantage in the automation and modularization of any manufacturing process. The shift from monolithic centralized control architecture to a modular architecture allows a process to be more flexible in manufacturing different products and allows for easier maintenance and upgrade of control systems. It also facilitates the exchange of information between different systems. Other areas where batch control standards and their underlying concepts may be used include:

- Start-up and shut-down of continuous processes
- Material handling
- Changing products and product grades
- Alarm and exception handling
- Product packaging

The PackML working group within the Open Modular Architecture Controls (OMAC) Packaging Workgroup started using the batch control standard as the basis for defining a more flexible use of packaging systems, which will ensure consistency and quality.

DEFINITION OF BATCH TERMS

Following are explanations of common batch control terms, simplified for easier understanding. For more rigorous definitions, refer to batch control standards.[1–3]

BATCH: A quantity of material produced by the single execution of a batch process. A batch also refers to the intermediate materials during the manufacturing process.

CONTROL MODULE: A set of equipment and functions that can carry out basic control. For example, a control loop consisting of one or more sensors, actuators, and control functions is a control module.

CONTROL RECIPE: An equipment-specific recipe that defines the production of a single batch. It is usually derived from a master recipe.

EQUIPMENT MODULE: A group of equipment that can carry out a finite number of specific minor processing activities. An equipment module may include one or more control modules. It is typically centered around a piece of process equipment, such as a weigh tank, heat exchanger, filter, or weighing scale.

EXCEPTION HANDLING: Procedures and functions that deal with conditions that are outside the normal or desired behavior of a process.

FORMULA: A part of the recipe that include process inputs, process parameters, and process outputs.

GENERAL RECIPE: A type of recipe that expresses equipment- and site-independent processing requirements. It is the highest level of recipes.

LOT: Products produced by a set of similar batches, usually using the same master recipe.

MASTER RECIPE: A recipe for producing a batch of products using a particular set of process equipment.

OPERATION: A procedure that controls the execution of a number of phases in a batch.

PHASE: A set of logic steps that completes a major processing function, such as charge, mix, heat, and reaction. A batch is usually in a stable state at the end of a phase.

PROCESS CELL: A set of equipment required for production of one or more batches. It usually consists of one or more units.

PROCESS INPUTS: Identity and quantity of raw materials and other resources required to make a batch. Other resources include energy and personnel requirements.

PROCESS OUTPUTS: Identity and quantity of products or energy produced at the end of a batch.

PROCESS PARAMETERS: Variables, such as temperature, pressure, and time, which are set points and comparison values that are needed for the production of a batch.

RECIPE: A set of procedures and formula variables that specify the production of a batch. There are four types of recipes: general, site, master, and control.

SITE RECIPE: A recipe that includes site-specific information, such as local language and locally available raw materials.

UNIT: A major piece of process equipment with its associated equipment modules. Mixers, storage tanks, and reactors are examples of units. The associated equipment modules include pumps, valves, heat exchangers, and agitators that are closely associated with the major process equipment. Units operate relatively independently of one another.

UNIT RECIPE: A part of a recipe that defines a part of batch production requirements within a unit. It usually includes a number of operations and phases.

MODELS

This subsection discusses briefly the main models that are specified in Part 1 of the batch control standard. As stated before, these models help modular decomposition of batch control problems and allow suppliers and users to maintain a common view of batch control (Figure 8.4b).

Physical Model

This model shows the physical hierarchy in a process industry. In this model, commands flow from higher to lower levels, and information flows from lower to higher levels. This model generally fits well with batch manufacturing processes. Each manufacturing area may be divided into a number of process cells, where each process cell has the necessary equipment and resources to complete a batch. A process cell consists of a number of units, such as storage tanks, mixing tanks, or reactors.

A unit may also include a number of process equipment or equipment modules, such as agitators, heating systems, pumps, valves, and the like. However, an equipment module may exist independently and may be used by multiple units. A control module, which is either a loop or a device, can be a part of an equipment module or can be directly associated with a unit. A control module may include measuring elements, such as thermocouples and orifice plates, and control elements that manipulate process variables, such as control and on/off valves.

A unit may not acquire another unit but may request its service (for example, to supply a measured amount of ingredient or to accept its product). The control of a unit is carried out by the unit supervision function, which is responsible for acquiring common resources, carrying out inter-unit communications, and performing the step-by-step control actions for the execution of a phase.

Control Activity Model

The control activity model shows the hierarchy of batch control functions. This model shows the hierarchical nature of batch control and thus allows a batch control problem to be broken up into multiple levels for analysis, specification, and design. The need and functional importance of these levels may vary between applications. It should be noted that the desired speed of response of a control system changes as we move from the lower to higher levels. Thus, the lower three levels, process management, unit supervision, and process control, act in real time while transactional-type processing is more appropriate for the upper levels. Each of the functions in the control activity model may be resolved into a number of subfunctions.

Procedure Model

A master recipe, which may be derived from a general or site recipe, contains specifics of the units or unit types that are used in a production plant. Master recipes are stored and maintained in a control system and are used for generating control recipes. A control recipe is used for the actual production of a batch and is generated at the same time a batch is created. The control recipe is generally deleted at the completion of a batch.

A master or control recipe specifies the data and procedure for manufacturing a batch of a given product. The data set is called a formula, which specifies process inputs, process parameters, and process outputs. Process inputs specify the quantity of raw materials and other resources required to make a batch. Other resources may include energy and personnel requirements. Process parameters specify the variables, such as temperature, pressure, and time, which are set points and comparison values that are needed for the production of a batch. Process outputs specify the quantity of products or energy produced at the end of a batch.

Additionally, a recipe contains a header and equipment requirements. The header may contain information such as product and grade identifiers, originator, and date of issue. The equipment requirements specify the type and size of process units, as well as other equipment-related constraints such as the materials of construction.

A recipe procedure may consist of a number of unit procedures. A unit procedure specifies the order of the functions (operations) that are to be carried out within a unit to manufacture a batch of product. An operation specifies the order of phases, such as charge, heat, mix, and store, that are carried out within a unit.

BATCH PROCESS CELL

A cell in a batch process may be classified in a number of ways. One way is by product variation, such as:

- Single product of same grade
- Single product in multiple grades
- Multiple products in multiple grades

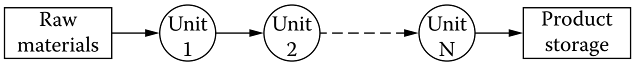


FIG. 8.4f
Single-stream structure.

A process cell that is set up to produce a single product of the same grade requires only one order of processing operations. There, a single recipe is required along with one set of formula variables. When a batch size is changed, the raw materials are scaled accordingly. In many such applications, the formula variables may be embedded in instructions within the phases. A process that produces a single product in many different grades requires a single recipe with multiple sets of formula variables. For manufacturing many products in multiple grades in the same process cell, multiple sets of recipes are required. There, the order of the operations and phases and the formulas vary from product to product.

Physical Structure of a Plant

Another way to classify batch processes is by their physical structure, such as:

- Single stream
- Parallel stream
- Multiple-path

In a single-stream structure, the units are ordered serially in a single train (Figure 8.4f). A batch moves from one unit to another in the predetermined serial order.

A parallel-stream structure is like a multiple single-stream configuration in which there may be common raw material and storage areas (Figure 8.4g), but otherwise each stream is isolated from the others.

In a multiple-path structure (Figure 8.4h), the movement of a batch is not along any fixed path. The choice of a downstream unit is based on the availability of a unit of the type required. The selection is made at the beginning of a batch or as required by the operator, or by a predefined algorithm.

A piece of equipment or service that is used by more than one unit is called a common resource. Common discharge pumps or common steam headers for multiple units are examples of common resources. A resource that can be

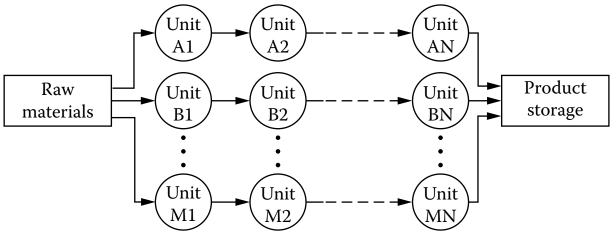


FIG. 8.4g
Parallel-stream structure.

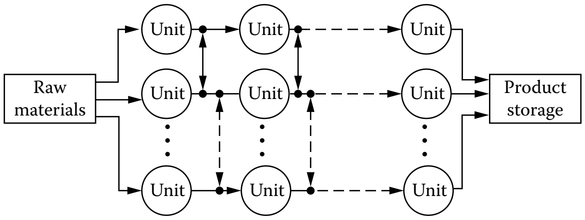


FIG. 8.4h
Multipath structure.

used by only one unit at a time is called an exclusive-use resource. A resource that can be used by several units at a time is called a shared-use resource.

In the case of a shared-use resource, there may be limitations on its capacity to serve. Thus, the common steam header may not be able to heat more than a certain number of units at the same time. In the case of an exclusive-use resource, book and release mechanisms prevent its use by more than one unit at a time. A shared-use resource requires mechanisms to protect it from exceeding its capacity.

EQUIPMENT FOR BATCH AUTOMATION

The types of control equipment commonly used for batch control are:

1. Programmable logic controllers (PLCs)
2. Personal computers (PCs)
3. Distributed control systems (DCS)

In a programmable logic controller the traditional programming environment is ladder logic or other languages specified in the IEC 61131-3 standard.[4] A ladder logic environment is well suited for safety interlocks, as these functions do not change with the state of a batch or the product being manufactured. For sequential control functions, which require step-by-step control actions, ladder logic can be obscure and difficult to follow. In addition, maintaining ladder logic for complicated processes can be expensive.

Traditionally, PLCs excelled in logical control but were weak in their continuous control capabilities. However, these characteristics are changing with the addition of new functions and programming environments, which include high-level procedure-oriented languages, block structures, and sequential function chart representations (Figure 8.4i). In addition, personal computers are also increasingly used as front-end devices to serve as programming and human interfaces.

A personal computer may also be used as a stand-alone controller where it is directly connected to input/output multiplexers. Small batch systems may be controlled in this fashion. The Windows operating system is often used, but capabilities of these systems are generally limited because of the size and throughput and also the limitations of the operating system.

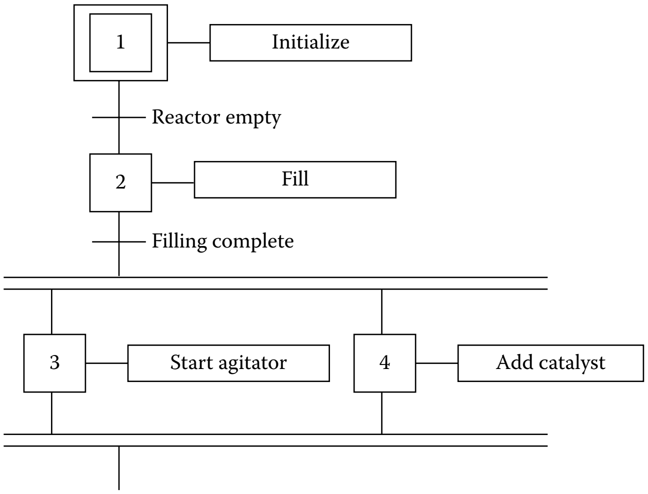


FIG. 8.4i
Illustration of a sequential function chart.

In a DCS system, the control is distributed into a number of controller modules. There are a number of significant advantages of a DCS system over a centralized control system, such as increased availability, ease of incremental expansion, and ease of interfacing with foreign devices. The capabilities of a contemporary DCS include standard communications protocols, industry-standard operating systems, and open information access throughout the system (Figure 8.4j).

Such a system allows easy partitioning of a batch control task horizontally and vertically. The horizontal partitions, based on physical model, are process cells, units, subunits, loops, and devices. The vertical partitions, based on control activity, are process management, unit supervision, sequential/regulatory control, discrete control, and safety interlocking.

BATCH CONTROL FUNCTIONS

Interlock Functions

Interlock functions enforce plant and personnel-related safety and are generally not dependent on the product or the state of the batch under manufacture. Thus, these interlocks, once set, are kept active all the time and are changed only when changes to the plant configuration or personnel safety considerations warrant it. These functions generally override other interlocks that may be active only during certain process phases or conditions. A simple example of such an interlock is if a pump ($P1$) has two outlet valves ($V1$ and $V2$) that are feeding two tanks. In that case, the pump should be switched off if both the valves are closed. This can be expressed by the Boolean equation

$$P1_{\text{off}} = V1_{\text{closed}} \text{ .AND. } V2_{\text{closed}} \qquad \textbf{8.4(1)}$$

Another way to specify interlock functions is by ladder logic. Because of the importance of interlocks in terms of

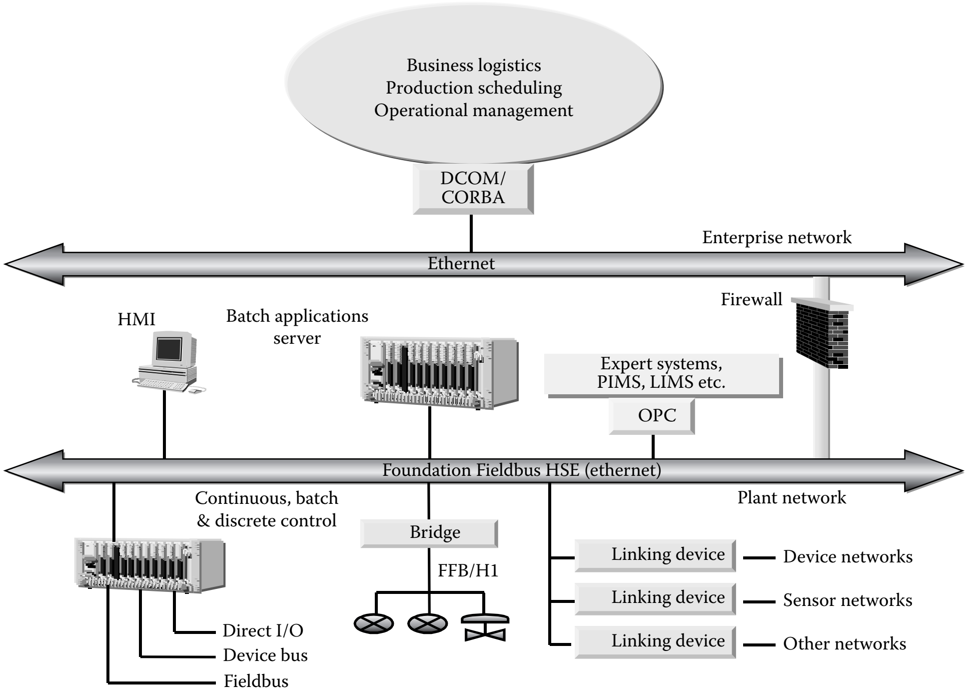


FIG. 8.4j
The main components of a typical DCS system for a batch control application.

safety, many users implement them using distinct hardware modules, either within or outside a DCS system.

Regulatory Control

Regulatory control serves to keep process variables as close as possible to their set points despite process and load disturbance. In a batch process, regulatory control is used extensively for controlling process variables, such as maintaining steady flow during charging, maintaining the agitator speed in a tank while mixing, or ramping up the temperature at a predetermined rate before reaction.

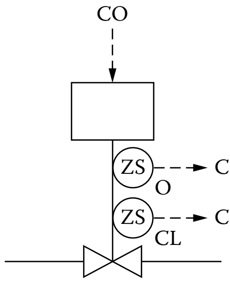
Regulatory control is often done using PID algorithms, through in recent years many modifications to this algorithm have been proposed for improved performance. In a DCS system, the usual method for configuring control loops is by interconnecting inputs and outputs of PID and other blocks. In a batch control system the activation, deactivation, setting of controller constants, and set points are usually carried out by steps in the phase logic.

Discrete Control

Discrete control is a term used for controlling process equipment that has only a limited number of stable states. On/off valves or manifolds are examples of such equipment. The on/off valve (Figure 8.4k) has only two states, which are manipulated by a contact output (CO).

The states of the two limit switches (ZS) specify the state of the valve. In the case of a manifold (Figure 8.4l), the three states are no flow, Tank A to Tank B, and Tank A to Tank C; all other possible states are considered abnormal and set off an alarm indicator. In a DCS system, the discrete control is usually done by a control block designed for this purpose, such as device block or a valve block. Some systems allow users to create blocks with unique functions.

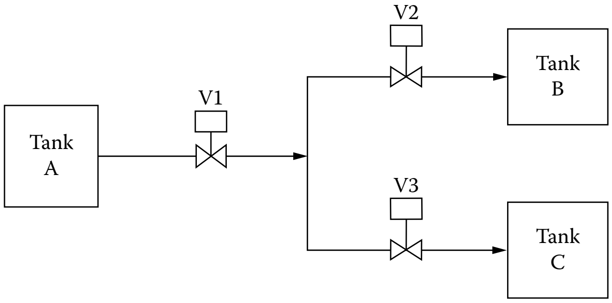
In addition, some DCS systems allow ladder logic or Boolean logic for configuring discrete control. When a large amount of discrete control is required, PLCs are commonly used. In a batch control environment, discrete control functions are usually directed by steps in phase logic.



State	CO	CI1	CI2
Open	On	On	Off
Closed	Off	Off	On

CO - Contact output
CI - Contact input
ZS - Limit switch

FIG. 8.4k
The operation of an on/off valve.



State	V1	V2	V3
No flow	Closed	Closed	Closed
Tank A to tank B	Open	Open	Closed
Tank A to tank C	Open	Closed	Open

FIG. 8.4l
The discrete control states of a manifold.

Sequential Control

Sequential control functions perform real-time control at the equipment level to move a process through a succession of distinct states. Opening a valve and starting a pump to transfer material from one tank to another and then closing the valve and stopping the pump after the transfer is complete is an example of sequential control. So, also, is the setting up of regulatory control loops or sending alarm messages to operators.

The sequential control required to manufacture a batch is usually divided up into a number of phases, such as mixing, heating, and reaction. A phase consists of a number of sequential control steps and can manipulate equipment within a unit boundary. When multiple units have to work in a synchronized fashion, for example, in the transfer of material from one unit to another, each unit will have its own phase. These phases may communicate with each other to set up the required synchronization via the unit supervision function.

Batch control may be divided into three broad categories (Figure 8.4m). The function that specifies the standard control actions is called normal logic. The function that checks the plant and process conditions on a periodic basis is called monitor function. Periodic checking of the state of the pump while a transfer of material is taking place is an example of a monitoring function. Unlike a safety interlock, which is active most of the time, a monitor function is activated by normal logic as and when required.

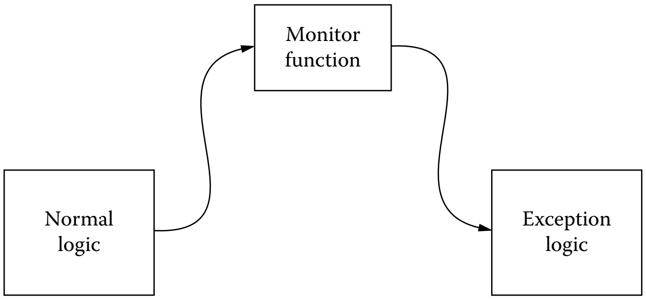


FIG. 8.4m
Sequential control functions. A. Monitor function is activated and deactivated by a normal logic. B. Monitor function calls an exception logic when a failure condition is detected.

"Minispec" For Weighing A Raw Material Using The Weigh Tank:

Get the amount and type of raw material from the control recipe.
Open appropriate raw material tank outlet valve.
Open weigh tank inlet valve 100%.
Start charging pump.
Wait until 90% of the raw material received.
Throttle weigh tank inlet valve to 10% open.
Wait until 100% of the raw material is received.
Close weigh tank inlet valve.
Stop charging pump.
Close the raw material tank outlet valve.
Store the weigh tank amount for batch record.
Inform mixer that weigh tank is ready to charge.
End of "Minispec."

FIG. 8.4n
Example of structured English as a means of specifying sequential logic functions.

When a monitoring function detects a failure condition, it invokes the appropriate exception logic. Exception logic, as the name implies, specifies control functions that are required to take care of failure conditions. Exception logic can be simple or elaborate. Sending an alarm to the operator and waiting until the device is fixed manually is an example of simple exception logic. Shutting down a malfunctioning equipment and starting up an available spare or initiating an emergency shut-down sequence for the whole unit are examples of more elaborate exception logic.

There are many ways of specifying sequential control functions, such as flow charts, state charts, Boolean equations, ladder logic, sequential function charts, and the like. Among these, the sequential function charts (Figure 8.4i) and structured English form (Figure 8.4n) are gaining increased acceptance.

In a DCS system, the normal and exception logic steps are usually implemented in a high-level procedural language or in a tabular state chart format. In a PLC environment, ladder logic and function blocks are commonly used. Block-oriented structures for sequence control are becoming increasingly common in DCS environments, where each phase is specified by a sequence block. Interblock communications are provided by connecting the inputs and outputs of these blocks.

The sequence logic steps are usually specified in a high-level Pascal-like language. Monitor blocks provide monitoring functions in a block-structured environment; otherwise, tables or Boolean equations are used. Block orientation provides modular structure and the ease of interblock communications.

Unit, Batch, and Recipe Management

The unit supervision function performs the control functions in a unit. Each unit has a set of phases (like charge, mix, and heat) that includes normal logic, exception logic, and monitor functions. The recipe specifies the order of these phases and the formula variables for a grade of product. Unit management also performs interunit communications and acquires the services of common resources when required.

The process management function is responsible for the manufacture of a batch of product. It allows the selection of a batch identifier (name) and a master recipe, either manually by an operator or automatically by a higher-level function. It generates a control recipe from the specified master recipe and maintains it during the course of a batch production. It directs the manufacturing of a batch by acquiring units as required and executing unit procedures, operations, and phases in the order specified by the control recipe. It also supplies the formula variables to the pertaining phases via unit supervision function.

Process management allows the scaling of a batch by proper proportioning of relevant formula variables. Process management is also responsible for creating and maintaining batch records during the production of a batch.

As stated before, there are four types of recipes: general, site, master, and control. The recipe management function provides the facilities for generation and maintenance of the general, site, and master recipes. The creation of a control recipe from a master recipe is a process management function.

Unit Modes The operating conditions of a process unit are broadly divided into two different modes: manual and automatic. In the manual mode, the procedural elements for the unit are executed manually. An operator may pause the progression of a batch but may not force transitions from one step to the next. While in this mode, the process inputs may still be monitored by the control system.

In the automatic (auto) mode, the control system controls the production of a batch in the unit. The transitions within phase logic are carried out without manual interruption as appropriate conditions are met. In some situations, operator may pause the progression, but may not force the transition from one step to another.

Unit States Equipment entities such as units and equipment modules may have many different states. The number of possible states of an entity varies with its complexity and its application. A simple on/off valve usually has only a limited number of states, such as open and close when working normally. In the case of a unit a much larger number of possible states have been specified in the batch standard[1] (Figure 8.4o).

Initially, a unit is in an idle state, which transitions to running state at the start of a batch. The running state continues as long as the normal batch production continues. Exception conditions can cause transition to a state, such as pause, held, stopped, or aborted depending on the type and seriousness of the exception condition. The state transition could be automatic or manually initiated. Usually there are transient states, such as pausing, holding, stopping, and aborting before the unit reaches paused, held, stopped, or aborted state. The unit may return to running state from a paused or held state. However, such transitions are not usually possible from states, such as stopped or aborted, which may require discarding the batch and starting anew.

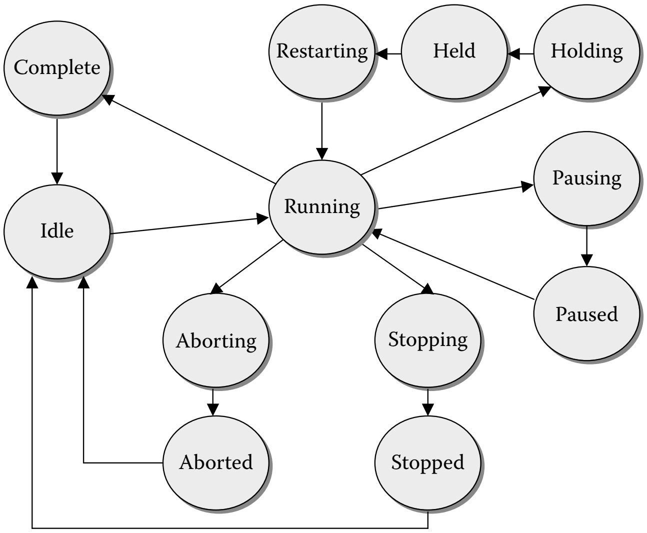


FIG. 8.4o
Typical unit states.

As stated, transitions may be automatic if specified in procedures, such as steps phases or by safety interlock functions. However, an operator can also cause transitions by manual commands. Transitions from paused or held to the running state usually require manual permissives. Detailed analysis of conditions that may lead to state transitions need to be carried out when the normal operations of a unit are specified.

FROM ANALYSIS TO IMPLEMENTATION

As stated before, the analysis, specification, design, and implementation of a batch control system are all complex tasks. That is largely due to the multistate nature of a batch process, which makes it inherently more complicated than a single-state continuous process. Batch processes generally need more sophisticated human interface than continuous processes.

Sharing of equipment and services can increase complexities, and proper booking and releasing mechanisms are required where equipment or a service can only be used for a single batch. Taking care of exception conditions can get very complicated for a critical process where different actions are needed for different alarm conditions and their combinations.

However, the complexities vary significantly from one batch process to another. They are typically based on plant topology (e.g., single stream, parallel stream, or multiple path) and the number of different products manufactured using the same equipment. In addition, complexities vary with the criticality of the process, the amount of exception handling, and the amount and the type of operator intervention required. The amount and complexity of exception handling for a batch process is also of great significance; unfortunately, this point is often ignored at the initial design stage. Finally, complexity depends also on the levels of control activity functions and the requirements of batch tracking and reporting.

Methods of Analysis

The analysis of a batch system requires resolving the problem into the various control levels and then, within each level, decomposing it into functional modules. The two common methods of analysis are structured analysis and object-oriented analysis.

Structured analysis involves starting from a high-level view and decomposing this into more detailed function.[5–7] The main components in this method are data flow diagrams, data dictionary, and minispecs. A data flow diagram consists of one or more bubbles, each representing a function, and arrows representing information or material flow (Figure 8.4p).

The storage of information and material is shown within parallel lines. Each bubble may be decomposed into a number of bubbles, each representing a subfunction. When a function cannot be decomposed into a set of subfunctions, then a minispec is generated for that function in structured English, as was illustrated in Figure 8.4n. All data names used in these diagrams are defined in data dictionaries (Figure 8.4q).

In object-oriented analysis, all functions are defined as objects.[8,9] An object mirrors a real-life entity or a function and contains the required procedure and data for carrying out that function. The key concepts of object orientation are:

- Encapsulation
- Inheritance
- Message passing
- Late binding

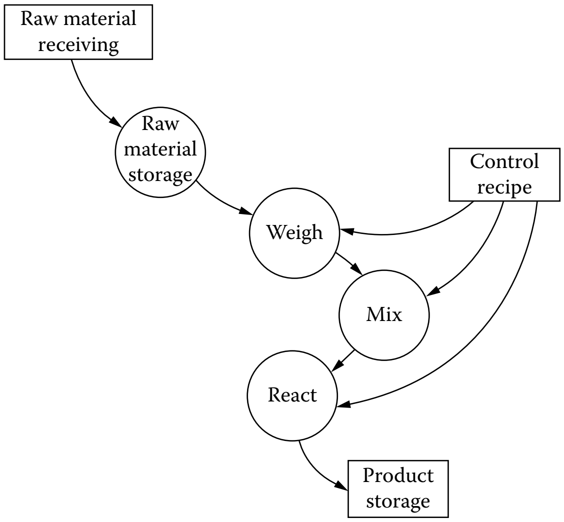


FIG. 8.4p
Example of a data flow diagram.

Control recipe =
Header + Equip Requirement + Formula + Procedure
Procedure = {Operation}
Operation = {Phase}
Phase = [Weigh, Charge, Mix, React, Cool, Transfer]
Formula = [Amount, Mix Time, React Temp, Ramp Rate, Prod Density]

FIG. 8.4q
The data names used are defined in a data dictionary such as this one.

Encapsulation, or information hiding, allows for the creation of an object with its set of data and procedures (Figure 8.4r). An object can be asked to do a certain function, but the user does not require the knowledge of the object's internals. Similar objects can be specified as a class so that their data structure and procedures need to be defined only once.

Inheritance allows a subclass of objects to inherit the data structure and procedures of its super-class (Figure 8.4s). However, a subclass object can have a data structure or procedures of its own, which are in addition to or instead of those inherited. Message passing is telling an object what to do without specifying how to do it. Late binding allows resolving the address of an object at run time rather than at compile time. This allows the ease of adding, deleting, or moving objects without affecting the rest of the system.

Project Application Specification

Batch control design and implementation problems often arise because of the lack of clear and detailed definition up front. This can be addressed by putting in enough effort at the beginning of a project by generating a project application specification (PAS). This document is in three parts:[10]

- Requirement specification
- System functional design
- System acceptance criteria

The requirement specification is the detailed narrative of the requirements, with little consideration for the specifics of the control system to be used. The requirement specification is ideally generated by the user organization before the selection of a particular system. This document can then be used as a bid specification for the control system suppliers to generate quotations.

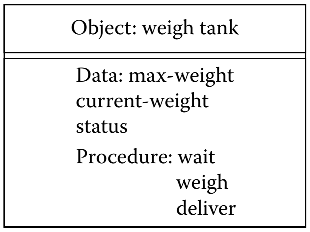


FIG. 8.4r
Object representation of a weigh tank.

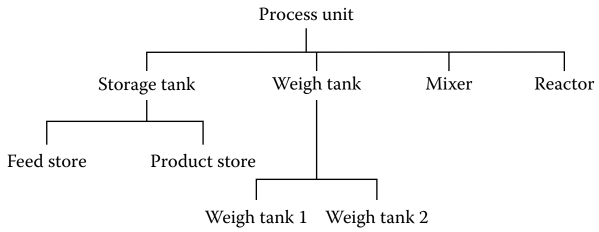


FIG. 8.4s
Example of an object class structure.

The system functional design, which is generated only after a specific control system has been selected, specifies the detailed design based on the requirement specification and the capabilities and constraints of the system chosen. The design part of the document is generated by the group responsible for integrating the system and doing the application engineering (system developer) and must be approved by the user before the actual implementation (configuration and coding) starts.

The final section of the PAS is the system acceptance criteria, which defines the tests that are to be carried out during and after the system's construction to ensure its proper functioning. The system acceptance criteria include detailed procedures for the systematic testing of all the significant functions. This section should preferably be written jointly by the user and the system developer.

A PAS is a single consistent document for the procurement and implementation of a batch control system. It is a "living" document, which requires updating and modification to accommodate changes in requirements. Appropriate mechanisms for suggesting and approving modifications need to be in place during the execution of a project. A PAS requires considerable effort and financial investment up front, but the return on investment is very significant in terms of the implementation efficiency and the quality of the product.

The solution of a batch control problem requires not only the selection of a suitable control system but also the creation or assembling of appropriate strategies for control. Object-oriented environments are changing the way a control system is configured and programmed. In the future, the competitive advantages of control companies will come more from exploiting new software techniques rather than from the hardware employed.

RELIABILITY AND AVAILABILITY

The reliability of a batch control system is of greater significance than is the reliability of a continuous control system. On failure, the fallback system needs to know the exact state of a batch in order to be able to continue the production or to bring the process to a safe condition. Reliability is defined as the probability a piece of equipment is performing its required function for a specified time interval under stated condition:

$$R(t) = e^{-\lambda t} \qquad \textbf{8.4(2)}$$

where
$R(t)$ = reliability
λ = failure rate
t = time

This definition, however, does not take into account that a piece of equipment can be repaired and put back to service. The definition of availability takes care of that:

$$A = \text{MTBF}/(\text{MTBF} + \text{MTTR}) \qquad \textbf{8.4(3)}$$

where
A = availability
MTBF = mean time between failures
MTTR = mean time to repair

The above equation shows that the availability of a given system can be increased significantly by merely reducing its mean time to repair.

The degree of reliability needed for a batch control system varies with the criticality of the process under control, but more importantly, it varies with the different levels of control (Table 8.4t). Thus, the reliability at the device control level is more critical than at the recipe control level.

In a DCS, where these control levels are controlled by different hardware modules, this need for reliability is generally taken care of by fault-tolerant pairs as required. In this arrangement, two identical processors are configured as a married pair, where both the processors run in parallel using the same inputs, and the outputs of one of the modules is used for control at any given time. The outputs of both the processors are periodically compared to ensure that they are synchronized and are in good health. If they disagree, then diagnostics are invoked to allow the good partner to continue the control, and appropriate messages are generated to fix the other.

In a dual-processor arrangement, the failure of the primary module causes the backup to take over. Timely detection and exact synchronization of the two processors before the failure are important for an effective takeover. In a triple modular redundant (TMR) system, these problems are avoided. In such an arrangement (Figure 8.4u), three or more identical processors run in parallel using the same set of inputs. The outputs are fed to a voting circuit, which allows the immediate detection of a faulty module.

The theoretical availability of a TMR system is lower than that of a fault-tolerant dual pair. However, it is claimed that a fault-tolerant pair with less-than-perfect failure detection and takeover mechanism is actually less reliable.[11]

TMR systems are available for process control but are generally more expensive because of the need for three processors. The dual fault-tolerant pair arrangement is currently more common in the DCS environment.

In evaluating the reliability of a batch control system, both hardware and software reliabilities should be taken into consideration. In recent years, the hardware has become more

TABLE 8.4t
Requirements of Fault Tolerance

Control Level	*Fault Tolerance Requirements*	*Ways of Achieving Fault Tolerance*
Sensing elements	+	Minimal for most, crucial for a limited number. Manual backup for those not crucial. Double- or triple-redundant sensors for the most crucial
Safety interlocks	+ + +	Mostly by dedicated hardware modules, within or independent of the control system
Continuous control	+ +	Backup controller or manual control stations
Device control	+ + +	Backup controller
Sequence control	+ + +	Backup controller or triple-redundant system
Batch management	+ +	Backup controller where required
Recipe management	+	Backup controller where required
Scheduling management	X	Manual backup
Information management	+ +	Redundant bulk storage and backup processors where needed
Human interface	+	Multiple human interface equipment (printers, CRTs, and keyboards)
Interprocessor communications	+ +	Redundant communication channels

Key: X: Little or no requirement
+: Some requirement
+ +: Significant requirement
+ + +: Crucial requirement

reliable and less expensive. This in turn has increased the demands for features and capabilities of the control system and thus has made the software effort much larger and more complex. The software reliability problem is not similar to the hardware reliability problem, because software does not degrade with use. Unreliability in software is caused by the presence of programming "bugs," which may stay undetected even after rigorous testing.

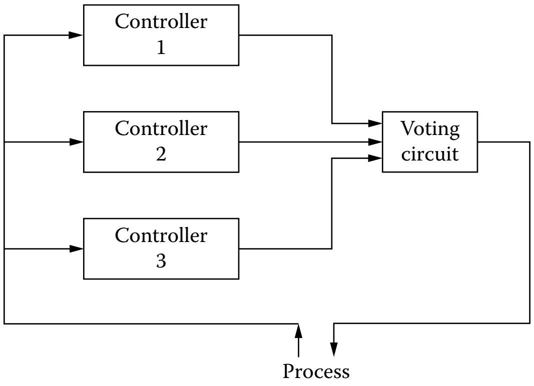


FIG. 8.4u
TMR system with voting circuit.

The reliability of a piece of software is dependent on the probability of its performance without fault for a given period of time and on the time to correct the fault after its detection. While testing a piece of software, an error detection curve may be drawn (Figure 8.4v), which can provide some indication of its reliability.

The standard software modules, such as the operating system, control packages, and utilities, are generally more reliable as they are used in multiple systems. When an error is reported, the system supplier usually generates a correction for all the installed systems. However, this is not the case with application-specific software, as they tend to be unique for each system. In addition, changes in control requirements require additions and modifications to existing software with

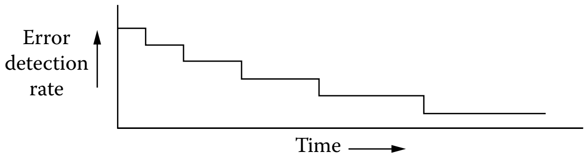


FIG. 8.4v
Software error detection curve.

the possibility of introducing new bugs. The ways for increasing the reliability and modifiability of software should be considered early in a project. These ways include:

1. Detailed analysis of control problem, breaking it into small manageable modules
2. Specification and design of the modules with clear definition of their interrelations
3. Proper design reviews
4. Modular programming using off-the-shelf software wherever possible and generating common routines for similar functions
5. Desk checking and walk-throughs for generated code
6. Rigorous testing of intramodule and intermodule functions with appropriate simulations

The object-oriented approach in specification, design, and implementation of software is gaining increased acceptance. This forces a modular approach and reduces the duplication of common code to a minimum. Object-oriented programming environments also allow for maintaining libraries of objects, which may be used in many different projects, thus reducing rework.

CONTROL SYSTEM SELECTION

For batch control system users, the most important criterion for supplier selection is the software and hardware functionalities. Additional criteria in order of importance are services capability, knowledge of the industry and its processes, local support, the cost of application services, and cost of hardware and software.

The selection of an appropriate control system for a manufacturing plant is largely dependent on the type and complexity of the application. The complexity of an application is dependent on many factors such as the plant's topology, the number of different products and grades, the criticality of the application, and the size of the plant.

Some batch processes are critical, because a control failure may result in a hazardous situation, equipment damage, or loss of very high-value product. The most common hazardous processes involve very fast exothermic reactions in which all corrective actions need to be taken automatically to ensure a safe operation.

The manufacture of PVC is an example of a fast exothermic process, and pharmaceutical and biotech manufacturing processes involve very high-value products. Critical applications generally require a high level of fault tolerance, a high level of safety interlocking, and extensive exception handling logic to take care of abnormal conditions. This combination of requirements significantly increases the complexity of the control solution.

PC-, PLC-, or DCS-based batch control systems perform similar functions, but some of their characteristics such as capacity, fault tolerance, and human interface, vary significantly. PC-based systems provide a lower-cost alternative to either a PLC- or DCS-based system, but offer little or no fault tolerance, and thus are not generally well suited for critical applications.

PLCs are well suited for logic control, permissives, and fast interlocking. PLCs also perform sequence control quite well and some continuous control. Some PLCs also provide good fault tolerance, making them suitable for critical applications. However, PLCs are generally not as good as DCS-based systems in the areas of expandability and flexibility. Modular programming is more difficult in PLCs, where only ladder logic is used for configuration. PLC-based systems are well suited for small- to medium-sized batch control applications and are generally not used for complex batch applications or those applications that require a high degree of flexibility (Figure 8.4w).

DCSs are well suited for continuous, sequential, and some basic logic control. They are generally not as good as PLCs at safety interlocking and other high-speed logic functions. Most DCS controllers can be configured in a redundant manner, with automatic switchover for a high degree of fault tolerance. DCSs also offer more flexible programming facilities than PLCs and easier-to-use human interfaces. However, the key strengths of many DCSs are their scalability, expandability, and the ability to interface seamlessly with PLCs and other types of controllers and business systems.

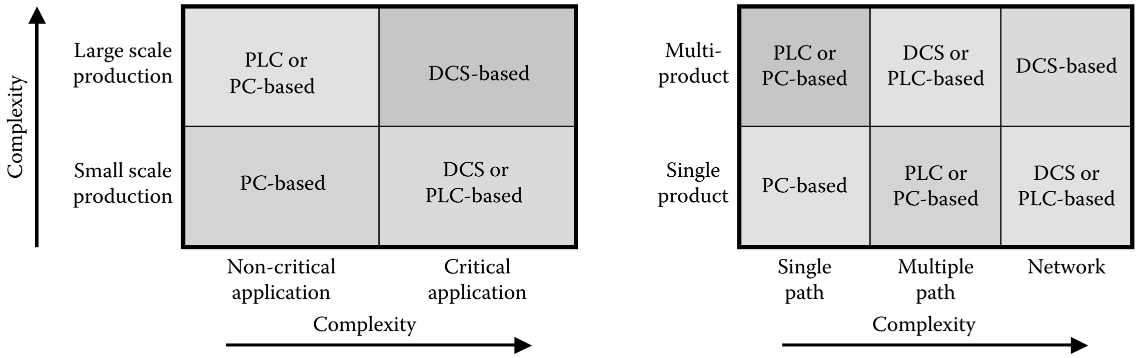


FIG. 8.4w
Recommended batch control systems based on process complexity.

These attributes make DCS-based batch control systems technically suitable for a wide range of batch applications. However, they are generally more expensive than PC or PLC systems, which make them less attractive for small- and medium-sized applications. Thus, DCS-based systems are most ideally suited for flexible, large, complex applications.

The lines that separate DCS- and PLC-based systems are starting to blur. A new type of system, called a hybrid system, is now being targeted for batch and discrete control applications. A typical hybrid system has a DCS architecture but uses controllers that are similar to PLCs. The functionality and scalability of hybrid systems are similar to that of the DCSs.

Examples of these new hybrid control systems include the DeltaV from Fisher-Rosemount, PlantScape from Honeywell, and ProcessLogix from Rockwell. It is also important to note that PC-based systems are beginning to acquire many DCS attributes.

To meet users' requirements in batch process automation, suppliers must exhibit many important characteristics. These include experience in batch process control, specific industry applications, enterprisewide integration, project management, and applying appropriate regulatory requirements.

Successful batch process control solutions have been elusive, and it is important to work with suppliers with proven records of accomplishment. Experience in your industry and specific application area is also very important in the selection process. Knowledge of specific industry segment characteristics, such as regulatory requirements, can be critical to the success of the project.

ACRONYMS

ANSI: American National Standards Institute
DCS: Distributed Control System
IEC: International Electrotechnical Commission
ISA: Instrumentation, Systems, and Automation Society
MTBF: Mean Time Between Failures
MTTR: Mean Time to Repair
OMAC: Open Modular Architecture Controls
PAS: Project Application Specification
PC: Personal Computer
PFC: Procedure Function Chart
PID: Proportional, Integral, and Derivative
PLC: Programmable Logic Controller
SFC: Sequential Function Chart
SQL: Sequential Query Language
TMR: Triple Modular Redundant
UML: Universal Modeling Language

References

1. ANSI/ISA-88.00.01-1995, "Batch Control Part 1: Models and Terminology," Research Triangle Park, NC: ISA.
2. ANSI/ISA-88.00.02-2001, "Batch Control Part 2: Data Structures and Guidelines for Languages," Research Triangle Park, NC: ISA.
3. ANSI/ISA-88.00.03-2003, "Batch Control Part 3: General and Site Recipe Models and Representation," Research Triangle Park, NC: ISA.
4. IEC 61131-3, "International Standard, Programmable Controllers — Programming Languages," IEC.
5. DeMarco, T., *Structured Analysis and System Design,* New York: Yourdon, 1978.
6. Yourdon, E., and Constantine, L., *Structured Design,* Englewood Cliffs, NJ: Prentice Hall, 1979.
7. Weinberg, V., *Structured Analysis,* New York: Yourdon Press, 1980.
8. Ghosh, A., "Object Lessons," *Chemical Engineering,* June 1991.
9. Ghosh, A., "The Object Is Control," *Chemical Engineering,* June 1991.
10. Ghosh, A., "Project Application Specification — What Is It and Why Do We Need It," Canadian Conference on Industrial Computer Systems, Montreal, May 1986.
11. Wensley, J. H., "Reliability in Batch Control Processes," ISA/82 Advances in Instrumentation, Part 3, October 1982.

Bibliography

ARC Report, "Batch Process Automation Strategies," ARC Advisory Group, October 1999.

Bristol, E. H., "A Design Tool Kit for Batch Process Control: Terminology and a Structural Model," *InTech,* October 1985.

Coad, P., and Yourdon, E., *Object-Oriented Analysis,* Englewood Cliffs, NJ: Prentice Hall, 1990.

Cox, B. J., *Object-Oriented Programming: An Evolutionary Approach,* Reading, MA: Addison-Wesley, 1986.

Fisher, T. J., *Batch Control Systems: Design, Application, and Implementation,* Research Triangle Park, NC: Instrument Society of America, 1990.

Ghosh, A., "Why Batch Process Control is not Chemical Engineering," AIChE Annual Meeting, Los Angeles, November 1991.

Ghosh, A., "Batch Control State of Art," in *Instrument Engineers' Handbook,* 3rd edition, Process Software and Digital Network, Sec 5.5, July 2002.

Glass, R. L., *Software Reliability Guidebook,* Englewood Cliffs, NJ: Prentice Hall, 1979.

Kopetz, H., *Software Reliability,* New York: Springer-Verlag, 1979.

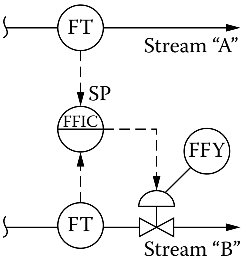
Rosenof, H. P., and Ghosh, A., *Batch Process Automation: Theory & Practice,* New York: Van Nostrand Reinhold, 1987.

Smith, D. N., *Concepts of Object-Oriented Programming,* New York: McGraw-Hill, 1991.

Sommerville, I., *Software Engineering*, Reading, MA: Addison-Wesley, 1984.

8.5 Blending and Ratio Controls

C. H. KIM, E. J. FARMER (1970, 1985) **B. G. LIPTÁK** (1995)
W. GARCÍA-GABÍN (2005)



Flow sheet symbol

Type of System:	A. Analog mechanical	
	B. Analog pneumatic	
	C. Analog electronic	
	D. Digital blending	
Features Available:	All designs are available with indicating, recording, or remote blend ratio adjustment features. In the case of digital systems, signal conversion is usually required	
Ratio Adjustment Ranges:	*Linear*	*Square Root*
	A. 0.1 to 3.0	0.5 to 1.7
	B. 0 to 3.0	0 to 1.7
	C. 0.3 to 3.0	0.6 to 1.7
	D. 0.001 to 1.999	
Precision (% Full Scale):	Inaccuracy	Repeatability
	A. ±2%	0.5%
	B. ±1%	0.25%
	C. ±0.5%	0.25%
	D. ±0.25% or better	0.1%
Controller Cost Ranges:	A. \$1500 to \$2500	
	B. \$2000 to \$3000	
	C. \$2000 to \$4000	
	D. \$5000 for stand-alone unit on "per flow stream" basis (can also be software in DCS system)	
Partial List of Suppliers:	Action Instruments (www.actionio.com)	
	Analogic Corp. (www.analogic.com)	
	Athena Controls Inc. (www.athenacontrols.com)	
	Barber-Colman Industrial Instruments (www.barber-colman.com)	
	Barton www.barton-instruments.com/index2.php	
	Bristol Babcock www.bristolbabcock.com	
	Dwyer Instruments Inc. www.dwyer-inst.com	
	Emerson Process Management www.emersonprocess.com	
	Fisher Controls International, Inc. (www.fisher.com)	
	Foxboro-Invensys (www.foxboro.com)	
	Honeywell Automation and Control (www.honeywell.com/acs/index.jsp)	
	ICS Triplex (www.icstriplex.com)	
	ISE, Inc. (www.instserv.com/prod01.htm)	
	Jumo Process Control. Inc. (www.jumousa.com)	
	Love Controls Corp. (www.love-controls.com)	
	Matricon Inc. (www.matricon.com)	
	Omega Engineering Inc. (www.omega.com)	
	Hanifin Corp. (www.pneutronics.com)	
	Partlow Process Instruments (www.partlow. com)	
	Powers Process Controls (http://www.powerscontrols.com/)	
	Robertshaw IDP, an Invensys Co. (www.robertshawindustrial.com)	
	Samson AG (www.samson.de/pdf_en/_ek16_ re.htm)	

Siemens Energy & Automation (www.sea.siemens.com)
Smar International (www.smar.com/products/technology.asp)
Spence Engineering Company, Inc. (www.spenceengineering.com/Handbook/index.htm)
Thermo Electric Co., Inc. (www.thermo-electric-direct.com)
Thermosystems (www.thermosystems.it)
Toshiba International Corp. (www.tic.toshiba.com)
Triplett Corp. (www.triplett.com)
United Electric Controls (www.ueonline.com)
Watlow (www.watlow.com)
Westinghouse Process Control (www.westinghousepc.com)
Wilkerson Instrument Co. (www.wici.com)
Yokogawa Corp. of America (www.yca.com)

INTRODUCTION

In addition to this section, blending and ratio control is also discussed in other sections of this handbook. Section 2.23 covers the theory of ratio control, while Section 2.12 discusses the interaction and decoupling in blending systems. Section 2.25 covers the relative gain aspects of ratio controls and also covers distribution control among a number of destinations.

Blending control is applied when the control objective is to mix two or more flows to achieve a specific composition. Ratio control keeps the proportion between two or more variables, often flows, at the desired set point. Blending controls are required in the production of solvents, paints, reactor feeds, foams, fertilizers, soaps, and liquid cleaners in the chemical industry. In the petroleum industry it is used in the production of gasoline, asphalt, lube and fuel oils, and distillates. In the food industry, in the making of wine, beer, candy, soap, ice cream mix, and cake mix. In the construction industry, in the making of cement, wire insulation, and asbestos products. Another example of ratio control is the control of the air/fuel ratio in a boiler or a furnace.

These applications provide the processor with economic advantages by controlling the consumption of materials (costly components and additives can be blended more precisely) and by reducing investment in floor space and batch tanks (costly blend tanks are eliminated). Through using continuous systems, the time lags resulting from batch methods are eliminated, productivity is increased, personnel needs are reduced, and inventory can be in the form of component base stocks rather than as partially blended or finished products.

Blending systems give technical advantages by accurately controlling the quality of the product and by providing the flexibility to blend a variety of finished products with minimum time required to change from one product to another.

This control approach achieves continuous control of the flow of each component with fixed ratios around components, so that when the streams are combined to form the finished blend at a fixed throughput rate, the composition of the finished product is within specifications.

In this section, blending systems will be described from the standpoint of control techniques. Analog and digital blending systems will be studied. A number of typical blending systems will be described to show the operating principles involved. A brief discussion of scaling procedures will be showed. Future trends on blending systems are illustrated, and some journals and papers with a practical point of view are available in the References. Finally, conclusions are presented.

BLENDING METHODS

Automatic, continuous, in-line blending systems provide control of gases, liquids, and solids in predetermined proportions at a desired total blend flow rate. Typical applications of blending and ratio control include maintaining a stoichiometric ratio of reactants to a reactor, keeping the air/fuel ratio to a boiler of furnace, and holding a reflux ratio for a distillation column at the optimum value. The blending systems consist of flow transmitters (to detect controlled variables), ratio relays (to set proportions), and controllers (to complete the closed-loop control).

A two-stream blending system is illustrated in Figure 8.5a. The component "A" flow controller is set by the total blend flow controller, and the component "B" flow is ratioed to "A."

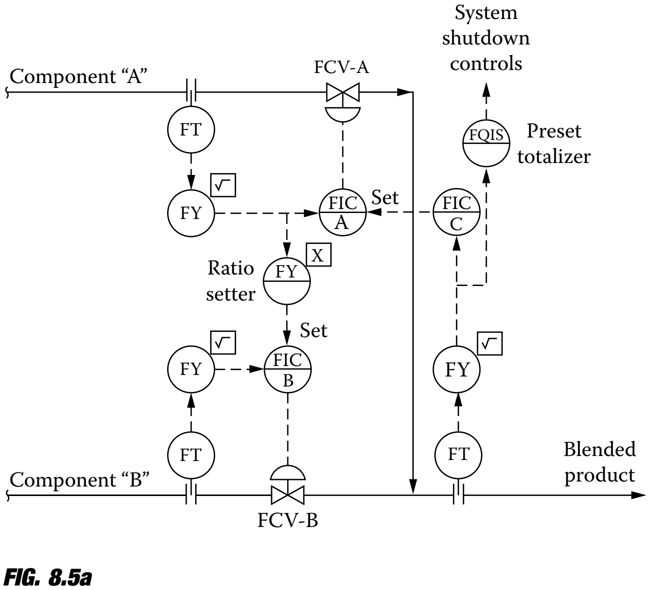


FIG. 8.5a
Analog rate blending of two components.

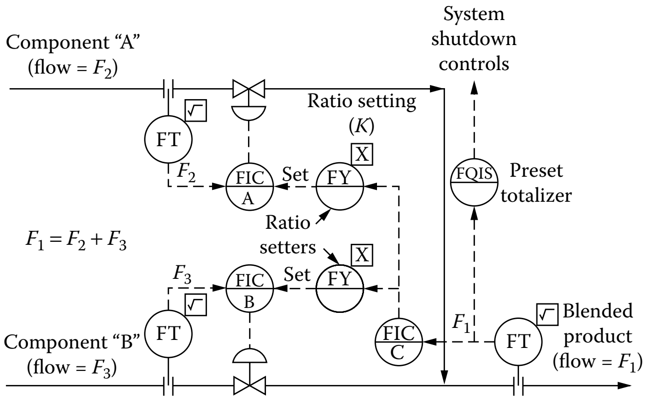


FIG. 8.5b
Both components can be directly ratioed to the total flow in an analog blending system.

It is also possible to have both blending components ratioed to the total blend flow, as shown in Figure 8.5b. In either case, the blending system maintains the blending ratio as well as the total flow rate. Incorporation of a preset totalizer with automatic system shutdown facilities provides batching capability, as well.

In both of these control configurations the total-flow controller is the cascade master, while the cascade slaves are the individual flow controllers. The only difference is that in Figure 8.5b the slaves are configured in parallel, while in Figure 8.5a they are in series. These control configurations can be further simplified by, for example, eliminating FIC-A in Figure 8.5a and allowing FIC-C to manipulate FCV-A directly. Because the time constants of the master and slave loops are similar in these cascade configurations, it is often necessary to detune the master (FIC-C), because otherwise it might change the set points of the slaves faster than they can respond to these changes.

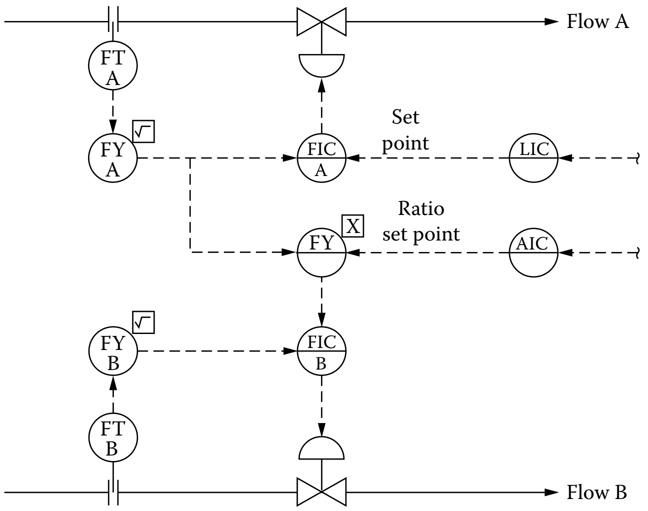


FIG. 8.5c
The level controller LIC manipulates the flow rate of "A," while the composition controller AIC manipulates the A/B flow ratio.

It is also possible to automatically control not only the total flow rate but also the composition of the blended stream. When composition control is also required (Figure 8.5c), it is recommended to let the total-flow master control the larger of the two streams and let composition set the set point of the smaller of the two streams. If the two blended streams are close to each other, such control systems will interact, and a change in one of the cascade loops will upset the other.

Such interactions can be decoupled by replacing the flow-based slave controllers (FIC-A and FIC-B) with slave controllers whose measurements have been modified as shown in Figure 8.5d. If total flow is the master of the slave that measures ($A + B$), while composition is the master of the slave that measures $A/(A + B)$, the interaction will be eliminated.

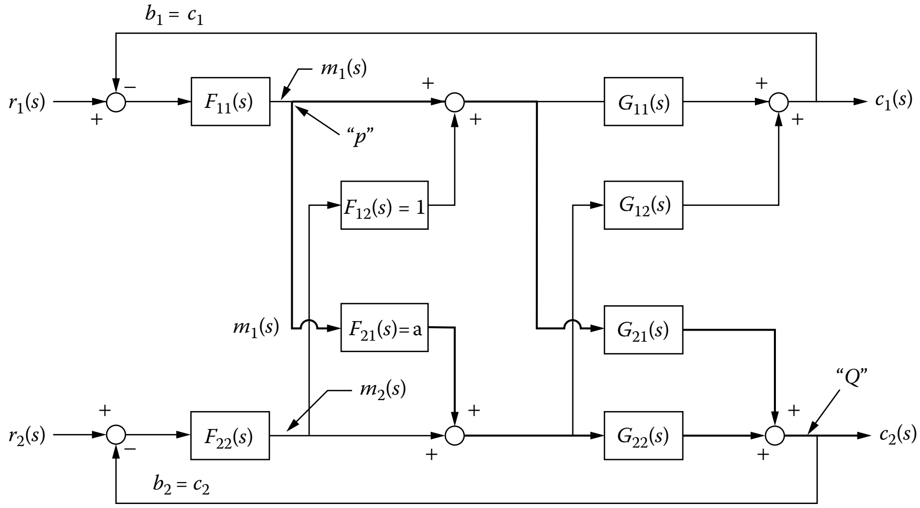


FIG. 8.5d
A generalization of a two-variable control system, which provides noninteracting control.

Many of the commercially available blending systems provide such accessories as flow alarm indications, system shutdown features, temperature compensation circuitries, scalers for conversion of transmitted signals to easily understood engineering units, pacing controls to slow down or shut down automatically, and manual or automatic adjustment of blend rate and ratio. All of these systems share one type or another of a ratioing mechanism. The methods of blending will be examined here. The working principles of components will be detailed in the latter part of this section.

Rate Blending

The blending system shown in Figure 8.5a is commonly found in the chemical industry for blending gas or liquid flows. A typical example is the manufacture of hydrochloric acid, which is done at fixed concentration and regulated flow rate by the absorption of anhydrous hydrogen chloride gas in water. The flow rate of hydrochloric acid from the absorption tower is measured to set the water flow, thereby maintaining the desired throughput, and the anhydrous hydrogen chloride gas is ratioed to water flow to give constant concentration.

A system in which all the blend components are ratioed to the total blend flow is illustrated in Figure 8.5b. A well-known example of this application is in the continuous or semicontinuous charging of a batch reactor, in which the recipe is given to set only the ratios of each ingredient and the total reactant charging rate. In the semicontinuous batch operation, a preset totalizer is used to terminate the charging operation when all ingredients have been charged. Numerous streams can be blended by incorporation of additional ratio devices and related controls.

Totalizing Blending

When totalized flows are ratioed, the integrated quantity of each component (over a period) is controlled in a direct ratio to the total quantity of the blended product. A schematic diagram of a totalizing blending system is shown in Figure 8.5e.

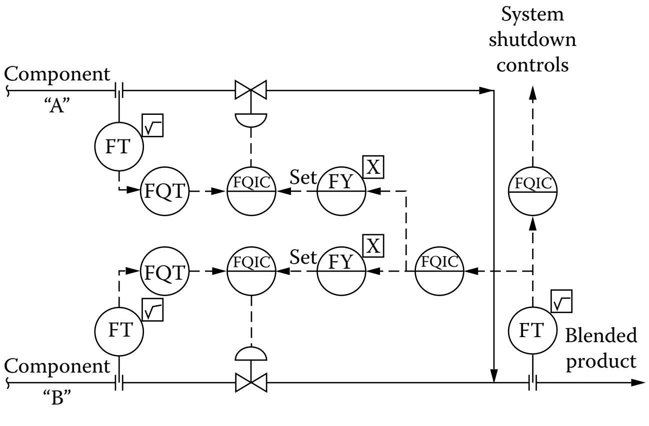


FIG. 8.5e
Analog totalizing blending system.

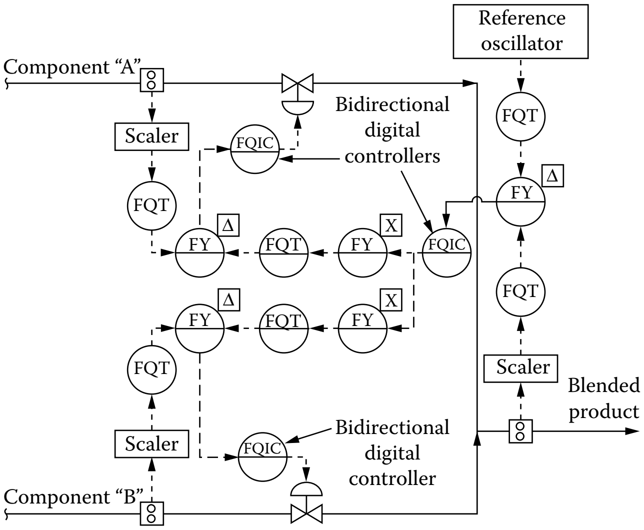


FIG. 8.5f
Digital totalizing blending system.

This system provides more precise control over the amount of each component than is possible with a rate-blending system. In the rate-blending system corrections are made to the ratio controller flow only after deviations have occurred (feedback control) and without correction for errors that have already occurred. In other words, the control system has no memory. Totalizing the flows and comparing the totals ensures the precise percentage of each component in the total blend.

Digital techniques have also been applied to digital blending systems, as shown in Figure 8.5f. Turbine meters or other pulse-generating devices can be used to generate digital flow signals. A bidirectional counter is used to integrate the flow measurement and demand signals and to compare them in order to generate a corrective control signal whenever the counts in the bidirectional counter memory are not zero.

The totalizing blending system (analog or digital) finds use in continuous in-line blending of petroleum products, such as gasoline and asphalts, in which long runs and batching operations require precise control of ingredients to ensure in-spec blending and uniform end products. The accuracy of this system allows for on-line blending of many petroleum, chemical, food, and cement products, with the product being sent directly to final shipping and storage containers. Multicomponent blending systems can be obtained by the addition of more flow ratio controllers.

Optimizing Blending

Almost the entire commercially available ratio relays and controllers are able to accept remote set points. Therefore, the blend ratios and total blend flow rates can be automatically adjusted by a process variable (output temperature, output pressure, composition, and so on). Thus, an optimizing blending system has the added capability of automatically manipulating ratio settings or the total rate, or both, based

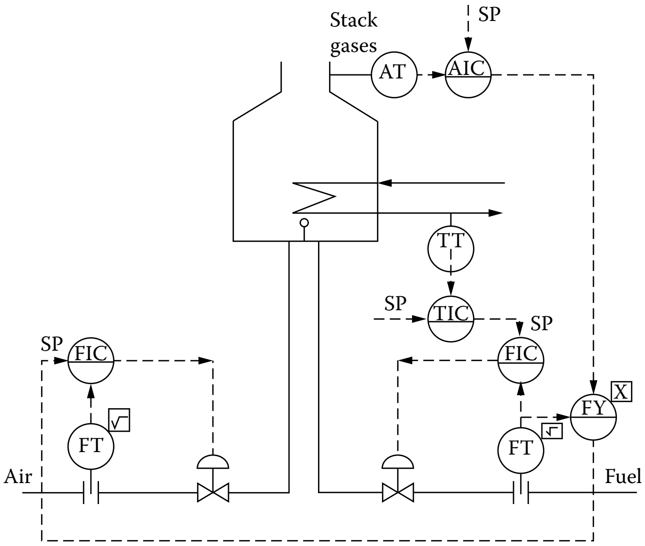


FIG. 8.5g
Control system where the air/fuel ratio is automatically adjusted to control the CO concentration in the stack gas.

on certain optimal criteria. Following this optimizing blending application, some approaches are illustrated.

The conventional air/fuel ratio control system is not versatile enough to maintain the air/fuel ratio in the range of high conversion efficiency at all operating conditions, especially in transient state or if the fuel characteristics change. Also, minimum pollution is desired in the stack gases. A ratio control based on an analysis of the stack gases in a furnace is shown in Figure 8.5g.

The control strategy adjusts the fuel/air ratio, optimizing CO percentage composition of the stack gases. Environmental and economical reasons require more complex configuration than classic control ratio, because an incomplete combustion results in the inefficient use of fuel and pollution. Optimizing ratio control employs a feedback signal based on the analyzer transmitter (AT), to regulate the fuel/air ratio. Thus, the temperature controller sets the reference value of the fuel control loop. The reference value of air loop control is set not only by fuel flow value (like classic ratio control) but also by CO analyzer information. This control approach optimizes the fuel/air ratio, avoiding atmosphere pollution and loss of energy.

A schematic of an optimizing blending system is shown in Figure 8.5c, in which the blend analyzer is used to measure composition and to adjust the ratio setting, while the level in some downstream tank is used to set the total flow rate.

A densitometer can be used to detect the solution concentration and to correct the ratio settings, avoiding set point deviations in the hydrochloric manufacturing process. A chromatographic analysis or Reid vapor pressure measurement of gasoline provides automatic adjustment of the blending ratio of components such as butane to give the proper octane number.

Another application is reactor feed rate control of jacketed exothermic reactors, based on the heat transfer coefficient and plant cooling capacity.

ANALOG BLENDING

The heart of the analog blending system is the mechanism for ratio control. This is often a separate component, although it may be housed with the controller. As shown in Figure 8.5a, a blending system can be constructed by ratio controlling the blend components with the total blend flow. Thus, the total blend rate is controlled together with the individual blend ratios. With incorporation of a preset totalizer, system shutdown can be initiated when batch blending is completed.

The ratio control relationship is derived with reference to the system shown in Figure 8.5b. It is assumed that C_1 and C_2 are the flow constants for the flow-measuring orifices. Then,

$$F_1 = C_1\sqrt{P_1} \quad \text{or} \quad P_1 = (F_1/C_1)^2 \qquad \textbf{8.5(1)}$$

and

$$F_2 = C_2\sqrt{P_2} \quad \text{or} \quad P_2 = (F_2/C_2)^2 \qquad \textbf{8.5(2)}$$

where

F_1 = total flow rate
F_2 = component "A" flow rate
P_1 = output signal of total flow transmitter
P_2 = output signal of component "A" flow transmitter

If K is the desired ratio setting,

$$P_2 = KP_1 \qquad \textbf{8.5(3)}$$

By substituting Equation 8.5(3) into Equation 8.5(2),

$$F_2 = C_2\sqrt{KP_1} \qquad \textbf{8.5(4)}$$

and by substituting Equation 8.5(1) into Equation 8.5(4) and simplifying,

$$F_2 = (C_1/C_2)(\sqrt{K})(F_1) \qquad \textbf{8.5(5)}$$

Using the same mathematical derivation, a ratioing system with linear input signals can be expressed as

$$F_2 = (C_1/C_2)(K)(F_1) \qquad \textbf{8.5(6)}$$

A graphical representation of Equation 8.5(6) is shown in Figure 8.5h, top, and the inverse-ratio controller characteristic is illustrated on the bottom. From Equation 8.5(5) it can be seen that the square root input signals from the

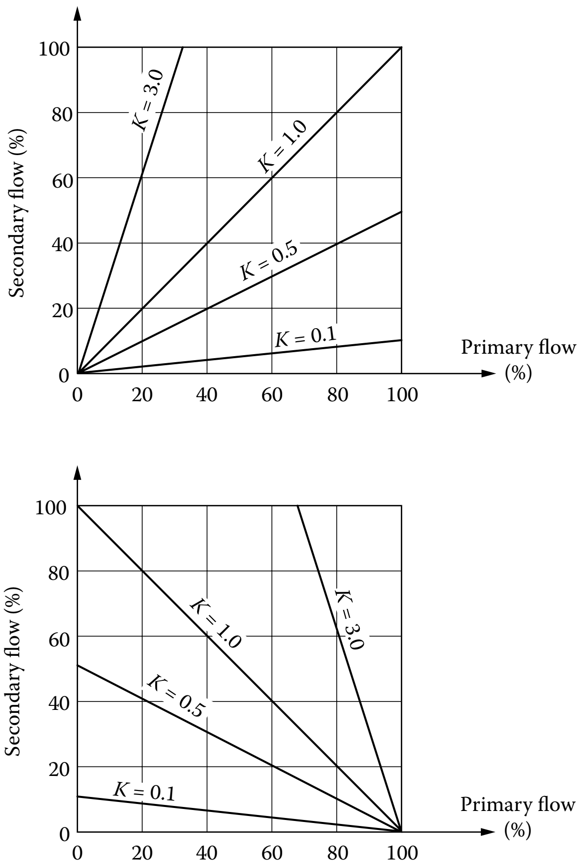


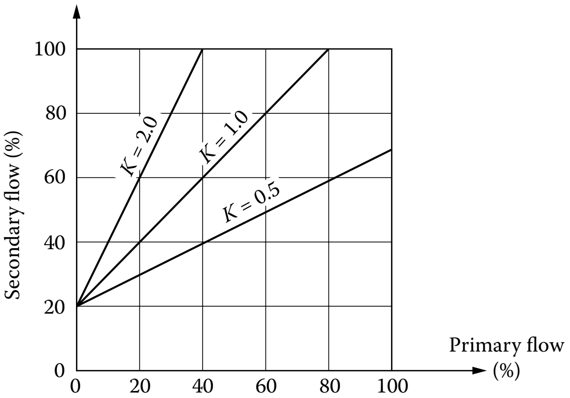
FIG. 8.5h
Direct ratio relationship (top) and inverse ratio relationship (bottom).

flowmeters can be used as in Figure 8.5f, with the ratio flow control circuit based on a linear relationship. The only modification is the square root calibration of the ratio setting dial and of the indicating dials.

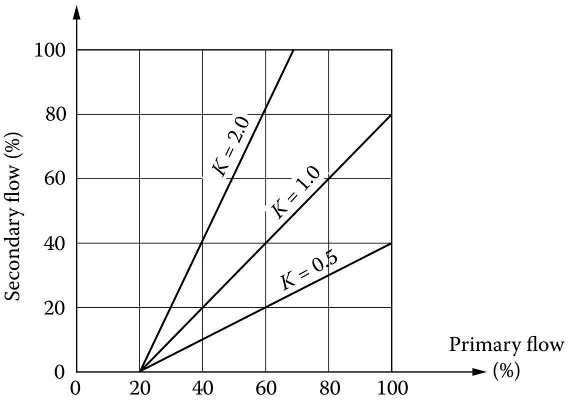
The ratio setting dial is graduated and calibrated as the square root of the linear ratio. Most ratio control mechanisms provide for bias adjustments to change the basic characteristics shown in Figure 8.5h, to meet various process requirements. Figure 8.5i shows some of these biased relationships in which the secondary flow has a preset minimum value or is kept at zero until the primary flow reaches some value.

Mechanical Ratio Control

The mechanical ratio control system consists of a proportioning mechanism, pneumatic or electronic flow signal receivers, and a case-mounted controller. The adjusting system resets the control point at a preset ratio by means of an adjustable mechanical linkage. The receivers are linked to a pen assembly, and output motion from the pen assembly operates the proportioning mechanism. Subsequent output motion from the proportioning mechanism positions the input lever of the pneumatic controller.



Secondary flow with minimum limit (positive bias)



Negative bias

FIG. 8.5i
Biased ratio relationships.

Figure 8.5j illustrates that each pen assembly spindle connects to one of the proportioning mechanism input levers through an adjustable link. Each input lever positions an

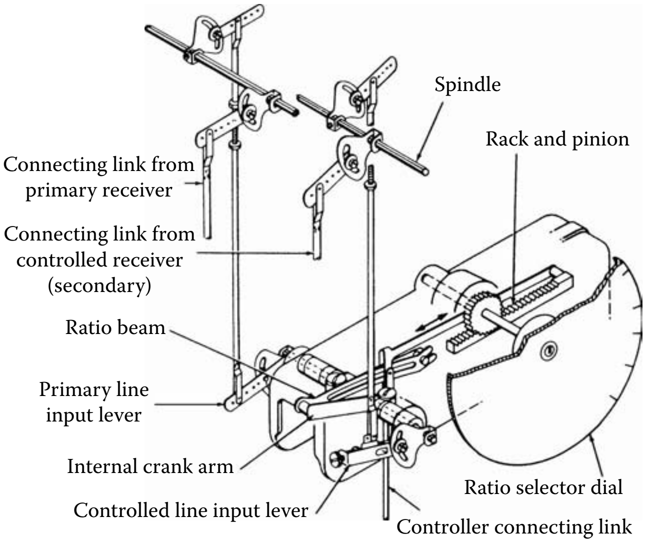


FIG. 8.5j
Ratio proportioning mechanism.

internal crank arm assembly and raises or lowers one end of the ratio beam. In turn, the link from the ratio beam connects to the input lever of the controller. The overall result is that controller output pressure changes whenever a receiver moves an input lever.

For remote adjustment of ratio set point, the manual set ratio mechanism can be replaced by a pneumatic receiver. In that case an external 3–15 PSIG (0.2–1.0 bar) set-point signal is used to position the receiver in proportion with the desired ratio. The ratio proportioning mechanism is precalibrated at the factory for the specific application, and all that is normally required is to check before use that the match marks are aligned and that the recorder pens and the transmitting meters are synchronized.

Lubrication is seldom required, but the mechanism should be periodically inspected, cleaned, and checked so that the proportioning mechanism operates frictionlessly.

An output inaccuracy of ±2% full scale can be obtained. Rangeability of this type of proportioning mechanism is approximately 40:1.

Pneumatic Ratio Control

The pneumatic ratio controllers contain no friction-producing mechanical links. The ratio relay modifies the input signal by means of the pneumatic circuit illustrated in Figure 8.5k.

The primary variable signal is tubed through a fixed restriction (FO) into an adjustable area restriction. If the variable restriction valve is closed, the signal is not modified. This condition represents 100% ratio, because the controlled variable signal (the set point of the secondary flow controller) must equal the primary variable signal for the control circuit to be satisfied.

If the adjustable area restriction is opened, the pressure between the two restrictions will drop until the flow through the fixed area restriction equals the flow through the adjustable area restriction. Thus, the pressure is modified as a function of the opening of the adjustable area restriction. By calibrating the adjustable restriction in terms of percentage ratio, one can set the relay for any desired ratio within its limits.

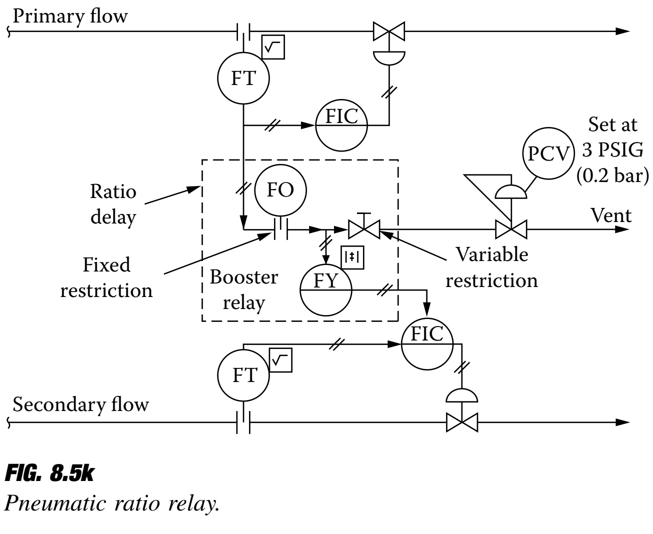


FIG. 8.5k
Pneumatic ratio relay.

A booster relay should be used to ensure rapid transmission of the modified signal from the ratio relay to the controller. For applications in which the secondary variable set point will always be less than 100% of the primary, a 1:1 booster relay is recommended. In other cases a 2:1 or higher booster may be used.

The pneumatic set ratio circuit is identical to the manual set ratio relay, except that the adjustable restriction opening is set by a pneumatic diaphragm motor. This allows for continuous automatic adjustment of the ratio in accordance with a pneumatic signal received from a quality controller or other optimizing device (Figure 8.5c). Incorporation of reset in the quality controller is recommended to eliminate the necessity for vernier adjustments to obtain the exact ratio and to compensate for the linearity limitations of the ratio unit.

The ratio relay should be calibrated at a specific ratio setting under actual operating conditions, even though the ratio setting is to be changed with operating conditions, to obtain maximum system accuracy. The accuracy (secondary flow set point) of ±1% of full scale can be expected. Signal rangeabilities of 50:1 can be obtained, but system rangeability is determined by the flowmeters used. Most pneumatic ratio control systems are not designed to operate below 20% of full-scale flow with square root signals or below 10% of full-scale flow with linear signals.

Electronic Ratio Control

Electronic ratio control systems operate on the Wheatstone bridge principle, shown in Figure 8.5l.

The bridge is said to be in a null, or balanced, condition when the ratios of resistance are such that $R_c/R_1 = R_f/R_2$, and no potential difference exists between points "A" and "B." If the ratio R_c/R_1 changes, then R_f/R_2 must also change by a like amount and in the same direction in order to maintain a null, or balanced, condition.

Figure 8.5m illustrates the operating principle of a Wheatstone bridge control system. Here, the fixed resistors, or bridge arms, are replaced by potentiometers so that the ratios previously mentioned are easily varied. Assuming an initial balance, an increase or decrease in the setting of the command potentiometer (primary) causes an error signal to

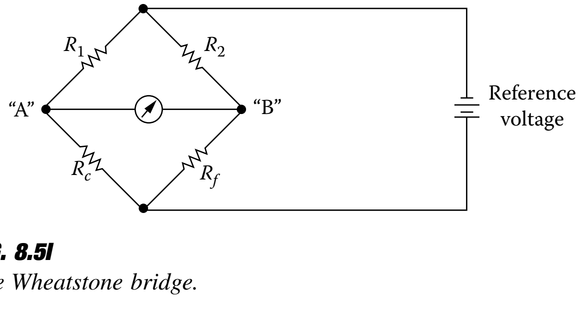


FIG. 8.5l
The Wheatstone bridge.

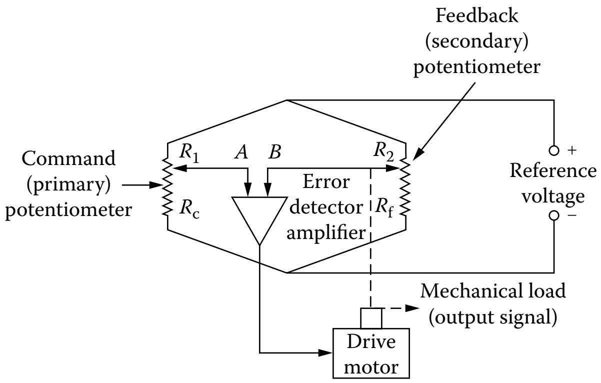


FIG. 8.5m
Wheatstone bridge operation.

appear at the input of a servo amplifier, which supplies power to the driven load.

The direction of movement is dependent upon the polarity (or phase) of the error signal. The brush arm of the feedback potentiometer is mechanically geared to the driven load (secondary) and thus rotates until $R_f/R_2 = R_c/R_1$. At this point a null again exists (error signal equals zero), and positioning ceases. The actual position of each pot wiper, expressed as a fraction of its total possible travel, may be written

$$\frac{R_c}{R_c + R_1} \quad \text{and} \quad \frac{R_f}{R_f + R_2} \tag{8.5(7)}$$

At null, $R_c/(R_2 + R_1) = R_f/(R_f + R_2)$. This is shown graphically in Figure 8.5n for a 0–100% movement of the command.

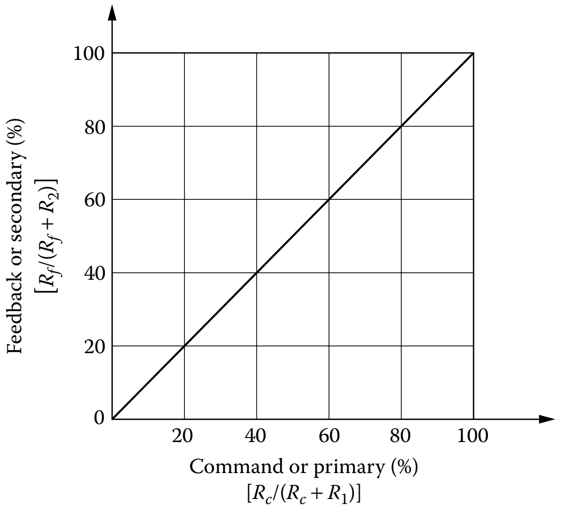


FIG. 8.5n
Command-to-feedback relationship.

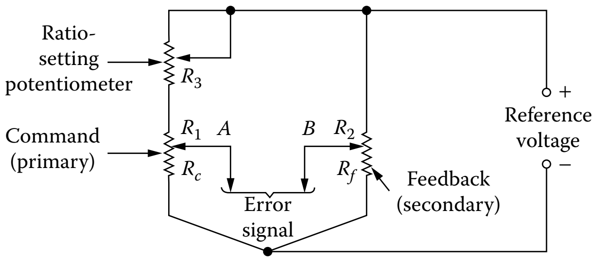


FIG. 8.5o
Electronic ratio circuit.

In this example the feedback signal is a measurement of actual movement or displacement, but when used in a flow ratio application, the feedback signal will be related to the secondary set point, and the command becomes the primary (wild) flow variable signal. Introduction of a fixed resistance in series with the command pot causes a change in the slope of the characteristic curve in Figure 8.5n.

By making this additional resistance a potentiometer, as shown in Figure 8.5o, one can limit the full-range travel of the feedback pot and of the driven load (secondary) to any desired degree for a 0–100% movement of the command.

In Figure 8.5o the feedback pot position at null will be

$$\frac{R_f}{R_f + R_2} = \frac{R_c}{R_c + R_1 + R_3} = K_1\left(\frac{R_c}{R_c + R_1}\right) \tag{8.5(8)}$$

where $K_1 = (R_c + R_1)/(R_c + R_1 + R_3)$. The slope of the characteristic curve in Figure 8.5p is K_1, and its value is dependent upon the setting of R_3, which is usually calibrated to represent K_1, the ratio setting.

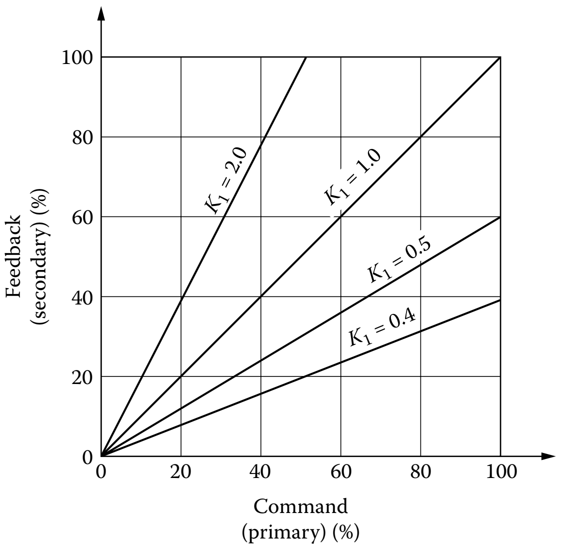


FIG. 8.5p
Command-to-feedback relationships.

Electronic ratio control provides fast, accurate, and adjustable ratios between input and output signals. Accuracy of ±0.5% of span is attainable.

Care should be exercised to provide a constant supply voltage and frequency. A change of 10% from the nominal voltage will cause a zero shift of as much as 0.5% of input value, and a change of 10 Hz over the range of 47–63 Hz will cause a zero shift of 0.25% of input value.

RATIO DIAL SETTING

The setting of the ratio relay is a function of the ranges of the transmitters. If the transmitters are measuring over the same range and in identical units, the graduations on the ratio dial represent the exact ratio between the primary and secondary flows. However, where maximum capacities and primary meter measurement units differ, the ratio selector dial-setting must be calculated for each ratio desired.

Commercially available ratio control units are graduated to handle signals of the same characteristics (either linear or square root) and of the same units. The following equation is used to calculate ratio dial settings:

$$\text{ratio dial setting} = \frac{(F_{\text{pm}})}{(F_{\text{sm}})}(R) \qquad \textbf{8.5(9)}$$

where

F_p = flow rate through primary
F_s = flow rate through secondary
F_{pm} = maximum capacity of primary flow transmitter
F_{sm} = maximum capacity of secondary flow transmitter
R = desired ratio of F_s/F_p

If F_{pm} is 50 GPM (189 1/m) and F_{sm} is 25 GPM (95 1/m), and it is desired to maintain the secondary at exactly 25% of the primary flow, then the ratio dial setting is (50/25)(0.25) = 50%. In selecting the ratio dial settings, one should keep in mind the rangeability limitations of the flow transmitters. (See Chapter 2, *Process Measurement and Analysis*.)

SCALING PROCEDURES

At this point a brief discussion of scaling in general may be helpful. Scaling is required not only in blending or ratio systems but in all control systems where calculations must be based on normalized (0–100%) transmitted signals, which represent different engineering quantities. The discussion on scaling here will consider the nature of traditional pneumatic relays (multiplying and dividing) in addition to analog electronic or DCS systems and will provide two examples.

The first example will be the scaling of a multiplier relay used in mass flow calculations, while the second example will describe the scaling of two relays (a multiplying one and a subtracting one), which are combined together to make a heat transfer calculation.

Multiplying and Dividing

Pneumatic multipliers typically have only two inputs, but several adjustable coefficients are available to facilitate scaling. The general formula for a multiplier is

$$A = a + f(B - b)[C(1 - c) + c] \qquad \textbf{8.5(10)}$$

Coefficients a, b, and c are bias or zero adjustments for the three signals, f is the gain of the device with both inputs at 100%, and $1 - c$ is the span of the C input. The formula for a divider is found by solving Equation 8.5(10) for signal B:

$$B = b + \frac{A - a}{f[C(1 - c) + c]} \qquad \textbf{8.5(11)}$$

Each instrument in a computing system, including indicators, recorders, and transmitters, should be calibrated against a common standard where possible. The most reliable standard to use is a mercury column, accurate to ±0.1% of the 3–15 PSIG (0.2–1.0 bar) range.

Calibrating multipliers and dividers is particularly painstaking because the adjustments must be made in a specific order. Referring to Equation 8.5(10) for a multiplier:

1. Zero adjustment “a” must be made with signals B and C at zero.
2. Zero adjustment “b” must be made with signal B at zero and C at 100%.
3. Zero adjustment “c” must be made with signal B at 100% and signal C at zero.
4. Span adjustment $1 - c$ must be made with B at some specified intermediate value and C at 100%.

A similar procedure must be followed with dividers.

After all components are calibrated individually, the system must be calibrated as a whole, to offset systematic (non-statistical) errors. Almost any of the adjustments can be used to calibrate the system at a single operating point, but the wrong choice may cause a greater error at some other point.

So, the accuracy of the system should be evaluated at several sets of conditions to determine which of the available adjustments would minimize the average error for all sets. Often more than one coefficient may require adjustment.

Scaling a Multiplier

Consider the example where a gas flowmeter requires compensation for absolute pressure:

$$W = k\sqrt{hP} \qquad \textbf{8.5(12)}$$

where W is the mass flow, h is the orifice differential pressure, and P is the absolute pressure. The orifice and differential range are selected so that 100% differential equals 100% flow at the normal operating pressure. With this in mind, scaling requires knowing only the normal pressure and the range of the pressure transmitter.

Let the normal pressure be 64.7 psia and the transmitter range be 0–75 PSIG with a base of 14.7 psia. Compensation requires that the multiplier exhibit a gain of 1.00 at the normal pressure. The absolute pressure range is 14.7–89.7 psia; therefore,

$$P = 14.7 + 75C \qquad \textbf{8.5(13)}$$

Here, signal C represents the pressure input to the multiplier. Then, the compensating factor determined at the normal pressure is

$$\frac{P}{64.7} = 0.227 + 1.160C \qquad \textbf{8.5(14)}$$

To conform with Equation 8.5(10), the maximum value of the compensating factor is extracted as $f = 89.7/64.7 = 1.387$:

$$\frac{P}{64.7} = 1.387(0.836C + 0.164) \qquad \textbf{8.5(15)}$$

Then, the scaled equation for the multiplier is

$$A = 1.38B(0.836C + 0.164) \qquad \textbf{8.5(16)}$$

where A is the multiplier output and B and C are the orifice differential and pressure inputs, respectively. A square-root extractor following the multiplier completes the calculation.

Scaling a Heat-Transfer Calculation

A simple yet highly effective procedure to be followed when scaling is:

1. Write the equation to be solved, including all conversion factors, with all signals given in the units in which they are to be measured or displayed.
2. Relate each input and output signal (having a range of 0–1.0) to the range of each variable, by a set of "normalizing" equations.
3. Substitute the normalizing equations into the original equation and solve for the output signal.

As an example, apply this procedure to the calculation of the rate of heat transfer, Q, to a liquid cooling medium. The liquid is flowing at rate F, with an inlet temperature T_1 and a higher outlet temperature T_2. The conditions are given in Table 8.5q.

The equation to be solved is

$$Q = Fk(T_2 - T_1) \qquad \textbf{8.5(17)}$$

TABLE 8.5q
Heat Exchanger Operating Conditions

Signal	Variable	Range	Normal Value
B	F	0–100 gpm (0–0.38 m^3/m)	60 gpm (0.23 m^3/m)
A	T_2	25–75°F (–3.9–23.9°C)	50°F (10°C)
C	T_1	0–50°F (–18–10°C)	30°F (–1.1°C)
E	Q	0–60,000 BTU/hr (0–17,400 W)	48,000 BTU/hr (13,920 W)

Coefficient k includes liquid density and specific heat, but it can be calculated from the normal operating conditions given in Table 8.5q.

$$k = \frac{Q}{F(T_1 - T_2)} = \frac{48,000}{60(50-30)} = 40 \qquad \textbf{8.5(18)}$$

Next, the normalizing equations are written, relating signals A, B, C, and E to the variables in Equation 8.5(17):

$$Q = 60,000E \qquad \textbf{8.5(19)}$$
$$F = 100B \qquad \textbf{8.5(20)}$$
$$T_2 = 25 + 50A \qquad \textbf{8.5(21)}$$
$$T_1 = 50C \qquad \textbf{8.5(22)}$$

Note that when signal A is 0, T_2 is actually 25°F, and when signal A is 100% (1.0), T_2 is 75°F.

Substituting the normalized equation into Equation 8.5(17) yields

$$60,000E = (100B)40(25 + 50A - 50C) \qquad \textbf{8.5(23)}$$

Solving for E,

$$E = 3.33B(0.5 + A - C) \qquad \textbf{8.5(24)}$$

Equation 8.5(24) must be solved by a subtractor and multiplier in combination, as shown in Figure 8.5r.

Because it is solved in two operations, Equation 8.5(24) must be separated into two pieces. Let the intermediate variable be identified as D:

$$D = 0.55 + A - C \qquad \textbf{8.5(25)}$$
$$E = 3.33BD \qquad \textbf{8.5(26)}$$

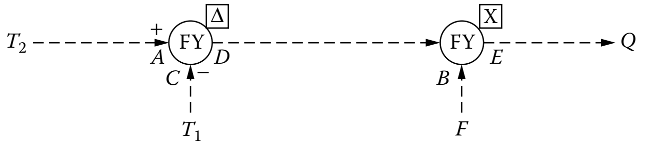


FIG. 8.5r
Scaling a heat-transfer calculation.

The entire factor 3.33 in the equation is shown applied to the multiplier. This is not altogether necessary; for example, the subtractor could have a gain of 2.0 and the multiplier a gain of 1.67. In that case:

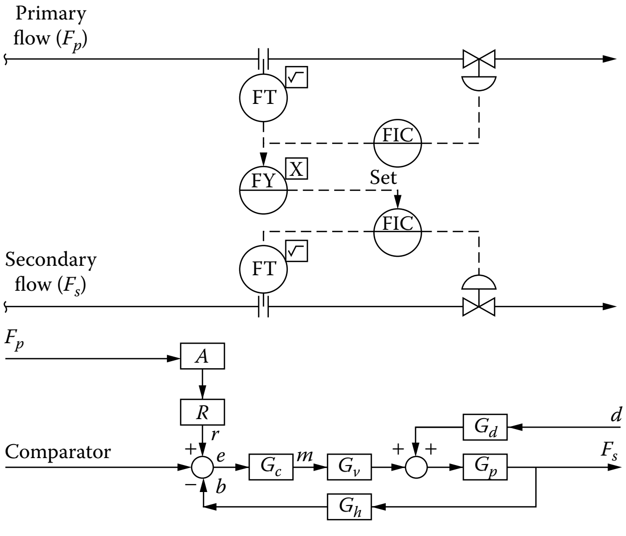
$$D = 1.0 + 2.0(A - C) \qquad \textbf{8.5(27)}$$

$$E = 1.67BD \qquad \textbf{8.5(28)}$$

This tends to improve the accuracy of the calculation but increases the danger of saturating the subtractor. Whenever more than one device is used in this way, each operation should be tested for saturation with reasonable combinations of inputs. In this example, a combination of 75°F for T_2 and 0°F for T_1 would not be reasonable.

RATIO CONTROLLER TUNING

A block diagram of a simple ratio control system is shown in Figure 8.5s. In the ratio control system, the set point of the secondary controller is directly related to the output of the primary flow transmitter. As the primary flow changes, the secondary controller assumes a new set point to maintain the desired ratio.



A = Input element.
b = Feedback variable.
d = Disturbance or load variable.
e = Error (deviation) signal.
F_p = Primary flow.
F_s = Secondary flow.
G_c = Controller transfer function.
G_d = Disturbance or load transfer function.
G_h = Feedback sensor transfer function.
G_p = Process transfer function.
G_v = Control valve transfer function.
m = Manipulated variable.
r = Reference (set-point) input.
R = Desired ratio.

FIG. 8.5s
Ratio control loop.

If simple characteristics are assumed for the transfer functions in the block diagram, the overall system transfer function for set-point disturbances can be expressed as:

$$F_s = RF_p \left(\frac{T_s(s+1)}{T_p(s+1)} \right) \times \left(\frac{T_i(s+1)}{(T_iT_s/K_c)s^2 + T_i(1/K_c+1)(s+1)} \right) \qquad \textbf{8.5(29)}$$

where

T_p = lag of primary flow measuring element
T_s = lag of secondary flow measuring element
T_i = integral time of controller
K_c = controller gain

As can be seen from Equation 8.5(29), the best regulation of the controlled secondary variable (with primary variable changes) can be obtained when the lags T_p, T_s, and T_i are minimum and the controller gain K_c is maximum, without creating instability. This statement is true not only for ratio loops but for feedback loops of all types.

DIGITAL BLENDING SYSTEM

The application of digital techniques to ratioing and blending may result in the total elimination of control system errors. This system continuously compares the total accumulated flows from each additive line with the total accumulated signal from a master oscillator. If there is a difference between these two values the corresponding control valve is repositioned to correct the deviation.

An overall digital blending system is illustrated in Figure 8.5t. The flow of each component is digitized by a turbine or displacement-type flowmeter or by an analog-to-pulse generator, producing a pulse train whose frequency is proportional to flow rate. A standardizer is used to scale the transmitter output frequency to a common reference basis, such as 1000 pulses per gallon (265 pulses per liter). This frequency is compared with a reference frequency produced by a numerically controlled frequency generator, which is commonly referred to as a *binary multiplier.*

The inputs to the multiplier consist of a numerical quantity and a pulse frequency, and the output is a new pulse train whose frequency is the product of the two inputs. The multiplier produces two reference frequencies proportional to the manually set numerical ratio settings of K and $(1 - K)$.

Each digitized flow rate is compared with its corresponding demand signal generated by the ratio set module (binary multiplier). This comparison is performed by a *bidirectional binary counter.* The bidirectional counter counts in the positive direction on pulses from one input and in the negative direction on pulses from the other input.

The set-point pulses produce "add" pulses, and the measurement pulses produce "subtract" pulses. Hence, if the flow-generated pulses equal the demand pulses, the algebraic sum is zero and no change will occur in the binary memory, and no corrective action is taken.

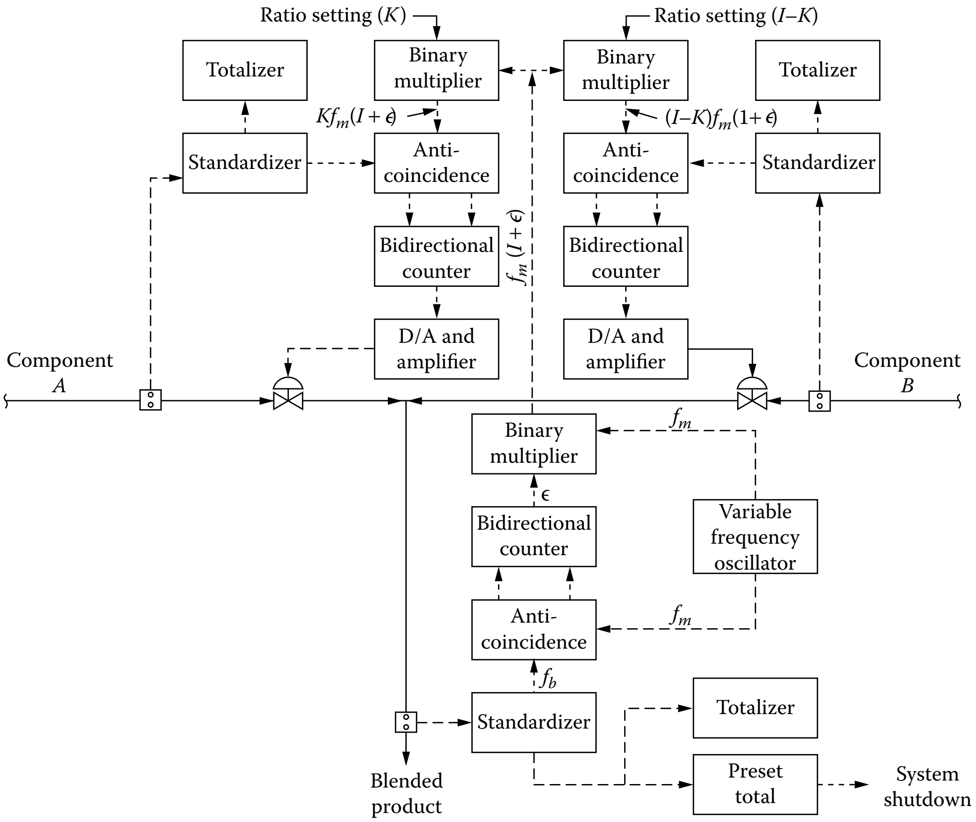


FIG. 8.5t
Schematic diagram of a two-element digital blending system.

Should the rate from one input exceed that from the other, an error count will accumulate in the memory, causing the valve-control logic to generate a proportional correction. This correction signal, after conversion and amplification, positions the control valve. Thus, the quantitatively controlled flow rates of the blend components are maintained at the prescribed ratio.

For applications requiring precise control of total flow rate as well as of the blend ratios, a further digital control loop is provided, as shown in Figure 8.5t. Here a variable frequency oscillator is manually set so that its frequency is proportional to the desired total blend flow rate. This reference signal (f_m), together with the signal generated by the actual total blended flow rate (f_b), is synchronized by anti-coincidence logic and accumulated in a bidirectional binary counter. The instantaneous counts (accumulation) of this counter are a measure of the difference between the total number of pulses generated by the reference oscillator and by the flowmeter, respectively. Thus,

$$\varepsilon = \Sigma f_m - \Sigma f_b \qquad \textbf{8.5(30)}$$

where

ε = instantaneous error accumulation in bidirectional counter
f_m = reference oscillator frequency
f_b = total blend flowmeter-generated frequency

This instantaneous error (ε) serves as the numerical input into a binary multiplier whose input frequency is f_m. Therefore, the multiplier output frequency is εf_m. This output is then mixed with f_m in such a manner as to avoid time coincidence, and it thereby yields a pulse train having the average frequency of $f_m(1 + \varepsilon)$. Thus, if the blend operation produces a flow rate that is less than the sum of the constituent flow rates, or if the blend output flow rate must be controlled while the blend ratios are kept constant, then the error term (ε) provides the necessary augmentation to the total flow rate reference frequency. The frequency input to the ratio setting binary multipliers is $f_m(1 + \varepsilon)$, and the resulting ratio demand outputs are

$$Kf_m(1 + \varepsilon) \quad \text{and} \quad (1 - K)f_m(1 + \varepsilon) \qquad \textbf{8.5(31)}$$

for the two-component blending system.

This principle can also be used automatically to slow down the total blend flow rate, by substitution of the master demand frequency (f_m) with a component flow frequency as an input to the master bidirectional counter. This feature is useful when one component may fall behind at start-up, when a strainer is plugged, or when a pump cannot meet the flow requirements.

When this occurs, the component controller takes over the pacing from the master demand unit if a predetermined error has been accumulated and adjusts the total flow rate to a value that the component can maintain. An alarm and automatic shutdown logic circuitry can also be incorporated to signal alarm conditions or automatically shut down the system if any of the components fall below their preset minimum rates.

The total blended product requirement may be preset on a totalizer to initiate batch shutdown. Analyzers or optimizers can be added to adjust automatically the blend ratios or the total blend flow rate as required.

The inaccuracy of the overall control system can exceed ±0.25%, with repeatability of better than 0.1%. The blend ratio setting can cover a range from 0.001 to 1.999, using four-digit thumbwheels. In a digital blending system, the dynamic response is limited only by the control valve stroke speeds, because the control system itself has practically no dead time.

TRENDS IN BLENDING SYSTEMS

High accuracy and optimal quality with easy recipe changes are the capabilities of microprocessor-based blending controls (Figure 8.5u). Nowadays, advanced control strategies are being applied to accurately compute and to track the target blending percentages and to obtain the quality of the products and flexibility to blend a variety of final products.

The following sections in this volume cover some of the more advanced control strategies that can be considered in optimizing blending systems: expert systems (Section 2.8), neural networks for process modeling (Section 2.18), hierarchical control (Section 2.11), genetic algorithms (Section 2.10), fuzzy logic control (Section 2.31) have been used in pilot plants and in some industrial processes to achieve these objectives.

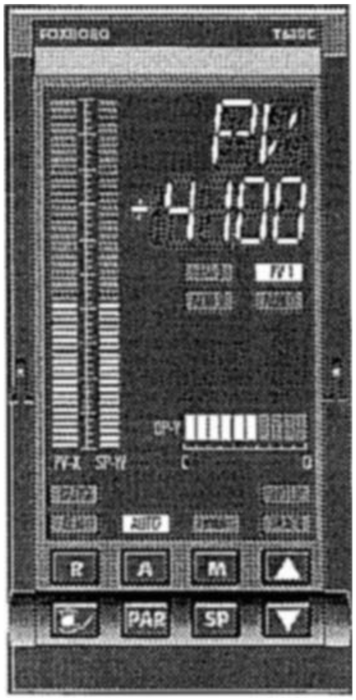


FIG. 8.5u
Microprocessor-based blending or ratio controller. (Courtesy of Foxboro Co.)

Application Examples

It is difficult to maintain constant product quality in coal preparation plants, because the properties of the raw coal supplies are highly variable. A fuzzy logic ash monitor can be used in a coal blending control system to stabilize the quality of the blend. Ash monitor has been used to achieve optimal blending of the coal supply. This approach allows coal loading for different clients having diverse requirements regarding the blending quality.[1]

An integrated approach to planning and coordinating the short-term scheduling of multiproduct blending with nonlinear recipe optimization can be achieved by hierarchical control. The resulting blending recipes and production volumes are provided as goals for scheduling level. The planning and scheduling approach is capable of switching between different recipes to obtain a high efficiency in solving industrial blending problems.[2]

When ratio control is done in software, it usually is resident in the "user layer" or Layer 8.[5] Ratio is a relatively simple function block to provide the "multiply" function. It is used in blending and other processes to establish one operator or recipe set value as a master, and to slave many different set point values based on the master.

In blending, a master flow is set as a constant with other flows proportional to the master flow. A ratio block would be configured in front of each set point of the flow control loops for each blended ingredient. As the master flow is changed, each ingredient flow will then be changed in constant ratio to the master flow.

An expert control strategy using neural networks, mathematical models, and rule models is applied to the coal blending process in an iron and steel plant. The complex model is constructed by using statistical data and empirical knowledge of the process and by forward-chaining and model-based reasoning. The predictions of the coal blend and coke quality are used to accurately compute the blending percentages. Target percentages are achieved by a distributed control technique employing PI control algorithms.[3]

Using traditional blending control, the desired ratio may be kept during a steady-state operation; however, during transients this ratio is hard to maintain. This is a serious problem, because blending control is applied to the process where the flows and blending ratios change frequently.

A blend control station that improves the ratio control performance during transients is applied in a paper mill process. The purpose of the controlled process is to add hydrosulfite to the pulp flow to bleach it. The goal is to keep the ratio between the pulp flow and the hydrosulfite flow constant. The control structure utilizes adaptive techniques to adjust on-line parameters based on the actual values of the pulp and hydrosulfite flows. In the adaptive blend station, no extra parameter tuning is required.[4]

CONCLUSIONS

In this section blending and ratio methods have been explained. In addition, methods of applying scaling procedures to blending and ratio systems been described. The industrial control configurations for rate blending, totalizing blending, and optimized blending have been shown.

Whether the blending system is designed by a user or is a package purchased from a manufacturer, the measuring and transmitting devices should be matched against the selected blending system in accuracy, rangeability, and flexibility. It is inconsistent to install an accurate digital blending system with low-accuracy sensors.

Also, the time lags in the flow measuring elements have a decisive role in the blending systems, because a good dynamic performance requires that the time constant of the transmitter be small. In-line blending systems have special importance when there is no downstream tank to accumulate and mix the blended product, because in such cases the blend composition must be maintained all the time. Advanced blending systems can economically improve the quality of the product and can provide flexibility to blend a variety of products.

References

1. Cierpisz, S., and Heyduk, S., "A Simulation Study of Coal Blending Control Using a Fuzzy Logic Ash Monitor," *Control Engineering Practice,* 10, 2002, 449–456.
2. Glismann, K., and Gruhn, G., "Short-Term Scheduling and Recipe Optimization of Blending Processes," *Computers and Chemical Engineering,* 35, 2001, 627–634.
3. Wu, M., Nakano, M., and She, J., "A Model-Based Expert Control Strategy Using Neural Networks for the Coal Blending Process in an Iron and Steel Plant," *Expert Systems with Applications,* 16, 1999, 271–281.
4. Hägglund, T., "The Blend Station: A New Ratio Control Structure," *Control Engineering Practice,* 9, 2001, 1215–1220.
5. SP50.8 Committee, "ISA TR50.02 Process Control User Layer," Research Triangle Park, NC: ISA, 2000.

Bibliography

Balchen, J. G., and Mummè, K. I., *Process Control Structures and Applications,* New York: Van Nostrand Reinhold, 1988.

Ball, S., *Analog Interfacing to Embedded Microprocessors: Real World Design,* Woburn Newnes, 2001.

Beard, M. J., "Analog Controllers Develop New Wrinkles," *Instruments and Control Systems,* November 1977.

Buckley, P. S., "Dynamic Design of Pneumatic Control Loops: Parts I and II," *InTech,* April and June 1975.

Buckley, P. S., *Techniques of Process Control,* New York: John Wiley & Sons, 1964.

Connel, B., *Process Instrumentation Applications Manual,* New York: McGraw-Hill, 1996.

Eckman, D. P., *Automatic Process Control,* New York: John Wiley & Sons, 1958.

Erickson, K. T., and Hedrick, J. L., *Plant-Wide Process Control,* New York: John Wiley & Sons, 1999.

IEC 614999-4, "Function Blocks for Industrial-Process Measurement and Control Systems, Part 4: Rules for Compliance Profiles," IEC, 2002.

Galovic, Z., "Advantages of Using Embedded Controllers on the Factory Floor," ETCON 2001.

Harriott, P., *Process Control*, New York: McGraw-Hill, 1964.

IEC 61506, "Industrial-Process Measurement and Control, Documentation of Application Software," IEC, 1997.

Jones, B. E., *Instrumentation, Measurement, and Feedback,* New York: McGraw-Hill, 1977.

Korn, G., *Interactive System Simulation,* New York: McGraw-Hill, 1989.

Luyben, W. L., and Luyben, M., *Essentials of Process Control,* New York: McGraw-Hill, 1997.

Luyben, W., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers,* 2nd edition, New York: McGraw-Hill, 1990.

Marlin, T., *Process Control: Designing Processes & Control Systems for Dynamic Performance,* New York: McGraw-Hill, 2000.

McAvoy, T. J., *Interaction Analysis,* Research Triangle Park, NC: Instrument Society of America (ISA), 1983.

Merritt, R., "Electronic Controller Survey," *Instrumentation Technology,* June 1977.

Murrill, P., *Application Concepts of Process Control,* Research Triangle Park, NC: ISA, 1988.

Murrill, P., *Fundamentals of Process Control,* 2nd edition, Research Triangle Park, NC: ISA, 1991.

Nachtigal, C. N., *Instrumentation and Control,* New York: Wiley, 1990.

Pal, J. K., and Tzafestas, S. G., *Real-Time Microcomputer Control of Industrial Processes,* Boston: Kluwer Academic Publishers, 2002.

Phillips, C. L., and Harbor, R. D., *Feedback Control Systems,* Englewood Cliffs, NJ: Prentice Hall, 1991, p. 136.

Predko, M. M., *Handbook of Microcontrollers,* New York, McGraw-Hill, 1999.

Roffel, B., and Chin, P., *Computer Control in the Process Industries,* Chelsea, MI: Lewis Publishers, 1989.

Rowan, D., "On-line Expert Systems in Process Industries," *AI/Expert,* August 1989.

Sadhukhan, D., and Dever, T. M., "The Application of Dedicated Controllers to Parallel/Serial Exhauster Trains," ISA EXPO 1998.

Seborg, D., and Mellichamp, D. A., *Process Dynamics and Control,* 2nd edition, New York: McGraw-Hill, 2004.

Shinskey, F. G., "Control Topics," Publ. no. 413-1 to 413-8, Foxboro, MA: Foxboro Company.

Shinskey, F. G., *Distillation Control,* 2nd edition, New York: McGraw-Hill, 1984.

Shinskey, F. G., *Process Control Systems,* 4th edition, New York: McGraw-Hill, 1996.

Smith, C. A., *Automated Continuous Process Control,* New York: John Wiley & Sons, 2002.

Smith, C., and Corripio, A., *Principles and Practice of Automatic Control Process Control,* New York. John Wiley and Sons, 1997.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice*, Englewood Cliffs, NJ: Prentice Hall, 1984.

Tocci, R. J., and Ambrosio, F. J., *Microprocessors and Microcomputers: Hardware and Software*, Upper Saddle River, NJ: Prentice Hall, 2003.

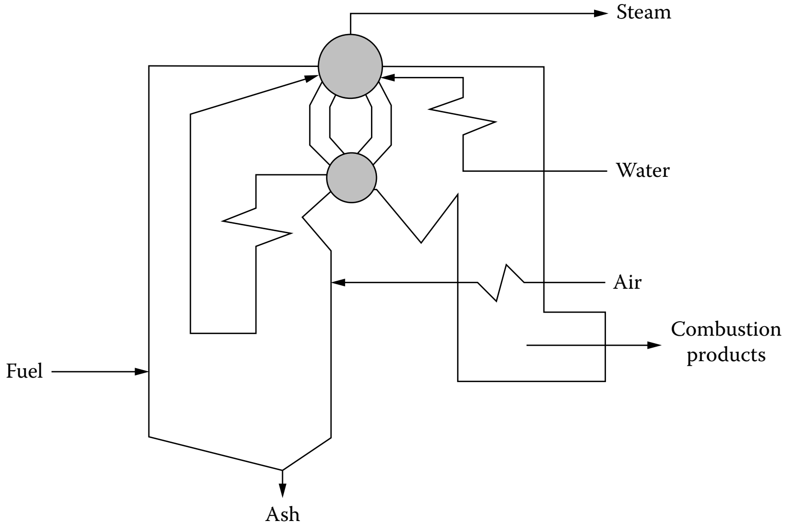
White, M. H., "Surge Control for Centrifugal Compressors," *Chemical Engineering,* December 25, 1972.

8.6 Boiler Control and Optimization

S. G. DUKELOW (1970) **B. G. LIPTÁK** (1985, 1995)

X. CHENG, R. H. MEEKER, JR. (2005)

Inputs by G. Liu (2005)



Flow sheet symbol

INTRODUCTION

This section is subdivided into four parts. The first part describes the characteristics of the boiler and its associated equipment (fans, dampers, and so on). The second part is devoted to a description of the conventional boiler controls, including steam pressure and temperature, air/fuel ratio, draft pressure, and feedwater controls. The third part describes the pollution control systems. And the fourth part discusses optimization and describes the methods of steam pressure floating, air/fuel ratio optimization, soot blower optimization, and blowdown controls.

Boilers are available in two basic designs: fire tube and water tube. Fire-tube boilers are generally limited in size to approximately 25,000 lb/hr (11,340 kg/hr) and 250 PSIG (1.7 MPa) saturated steam. Although they are noted for their ability to respond to changing demands, their size and pressure limitations preclude their use in large industrial facilities. Because of thermodynamic considerations, boilers should produce steam at high pressure and temperature to realize a maximum work efficiency. These conditions are achievable only with water-tube boilers—hence, they will be given prime consideration in this section.

Steam boilers are used by the electric utility industry to produce steam for power generation and by manufacturing and process industries (nonelectric utility) to produce steam for both power generation and process heating and energy conversion. Electric utility boilers tend to be larger and operate at higher pressures: A typical coal-fired utility boiler might produce 3 million lb/hr (1.36 million kg/hr) of superheated steam at 2400 PSIG (16.5 MPa). A typical nonelectric utility industrial boiler might produce 400,000 lb/hr (181,000 kg/hr) of steam at 900 PSIG (6.2 MPa). Industrial boilers are commonly referred to as co-generation or combined heat and power (CHP) applications.

In this section, the controls of both the electric utility and the process industry boilers are described. In general, the loads on the electric utility boilers are more stable, and when they change, they change slower than the often drastically varying loads that process industry boilers have to handle. As a consequence of this difference, the utility industry boiler control system shown in Figure 8.6l does not include the type of lead/lag compensation that is described in Figure 8.6s. Therefore, the overall controls shown in Figure 8.6l are for reference only, and the reader should study the discussion of the individual loops and select the appropriate ones for the type of load dynamics at hand.

The basic components of a water-tube steam boiler are the furnace, where air and fuel are combined and burned to produce combustion gases, and a water-tube system, the contents of which are heated by the combustion process. The tubes are connected to the steam drum, where liquid and vapor are separated and the generated water vapor withdrawn. If superheated steam is to be generated, the steam from the drum is

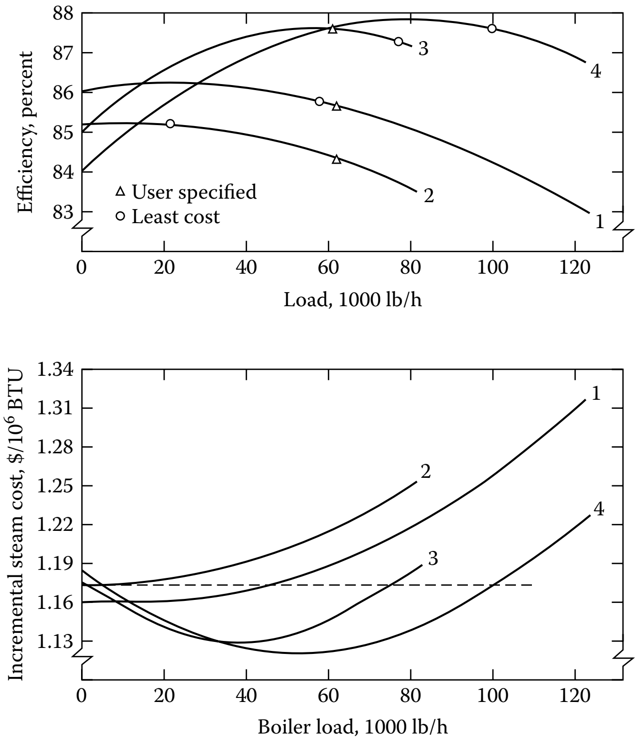


FIG. 8.6a

Boiler efficiencies and steam costs vary with both the design of the boiler and its loading.[25]

passed through the superheater tubes, which are exposed to the combustion gases. Supercritical or "once-through" boilers operate above the critical point of water where there is not a distinction between liquid and vapor; these boilers are not equipped with steam drums.

THE BOILER

Efficiency

The thermal efficiency of a steam generator is defined as the ratio of the heat transferred to the water (steam) to the heat input with the fuel. One of the goals associated with the operation, maintenance, and control of a boiler is to maximize its thermal efficiency.

The boiler efficiency is influenced by many factors. A fully loaded large boiler that is clean and properly tuned (with blowdown losses and pump and fan operating costs disregarded) is expected to have the following efficiencies:[1]

On coal: 88%, with 4% excess oxygen; 89%, with 3% excess oxygen
On oil: 87%, with 3% excess oxygen; 87.5%, with 2% excess oxygen
On gas: 82%, with 1.5% excess oxygen; 82.5%, with 1% excess oxygen

Boiler efficiencies seldom exceed 90% or drop below 60%. Efficiencies will tend to vary with individual design and with loading, as shown in Figure 8.6a. Efficiencies will also vary as a function of excess air, flue-gas temperature, and boiler maintenance. A 1% loss in efficiency on a 100,000 lb/hr (45,360 kg/hr) boiler will increase its yearly operating cost by about $20,000. A 1% efficiency loss can result from a 2% increase in excess oxygen[1] or from about a 50°F (28°C) increase in exit flue-gas temperature.[2]

Efficiency can be computed by the *direct* or the *indirect* method.[3] The direct method uses the ratio of the rate of heat transferred to the water (outlet steam specific enthalpy × steam mass flow−feedwater specific enthalpy × feedwater mass flow) to the rate of heat input by the fuel (higher heating value × fuel mass feed rate).

The indirect method uses fuel, ash, and stack gas analysis to do a per-unit-basis accounting of all heat losses, subtracting all losses from the higher heating value of the fuel and dividing the result by the higher heating value. The indirect method is more accurate, because it does not rely on the relatively inaccurate steam and fuel flow measurements. The major losses considered by the boiler indirect efficiency calculation equations are:[4]

Dry gas loss: sensible heat carried out of the stack with the combustion air and combustion products
Moisture loss: loss due to vaporizing the moisture in the fuel and the moisture produced from combustion of the hydrogen in the fuel
Incomplete combustion loss: loss due to combustion of carbon that results in carbon monoxide (CO), instead of the complete combustion product, carbon dioxide (CO_2)
Unburned carbon loss: loss due to carbon that does not get combusted and ends up in the refuse (ash)
Moisture in the combustion air loss: loss due to heating up water vapor contained in the combustion air
Radiation loss: heat lost from the external furnace walls to the surrounding air and other surfaces

If variations in only the dry gas loss are of primary interest (normally the largest source of energy loss[5]), then the efficiency can be approximated using Equation 8.6(1),[6] which assumes nominal fixed values for most of the above losses, based upon the type of fuel.

$$E = 100\left[1 - 10^{-3}\left(0.22 + \frac{K''y}{1 - y/0.21}\right)(T_s - T_a) - \frac{\Delta H_c}{H_c}\right] \qquad \textbf{8.6(1)}$$

where y is the mole fraction of oxygen in the flue gas, and K'' is a coefficient assigned to each fuel: 1.01 for coal, 1.03 for oil, and 1.07 for natural gas. The term $\Delta H_c/H_c$ is about 0.02 for coal, 0.05 for oil, and 0.09 for gas; the terms T_s and T_a are the stack and ambient temperatures (°F).

Where accuracy of the calculated efficiency is important, such as validation against performance guarantees, it is best to refer to the full indirect method calculations; a generally accepted standard for steam generating unit efficiency calculations is the ANSI/ASME Power Test Code (PTC) 4.1.[3]

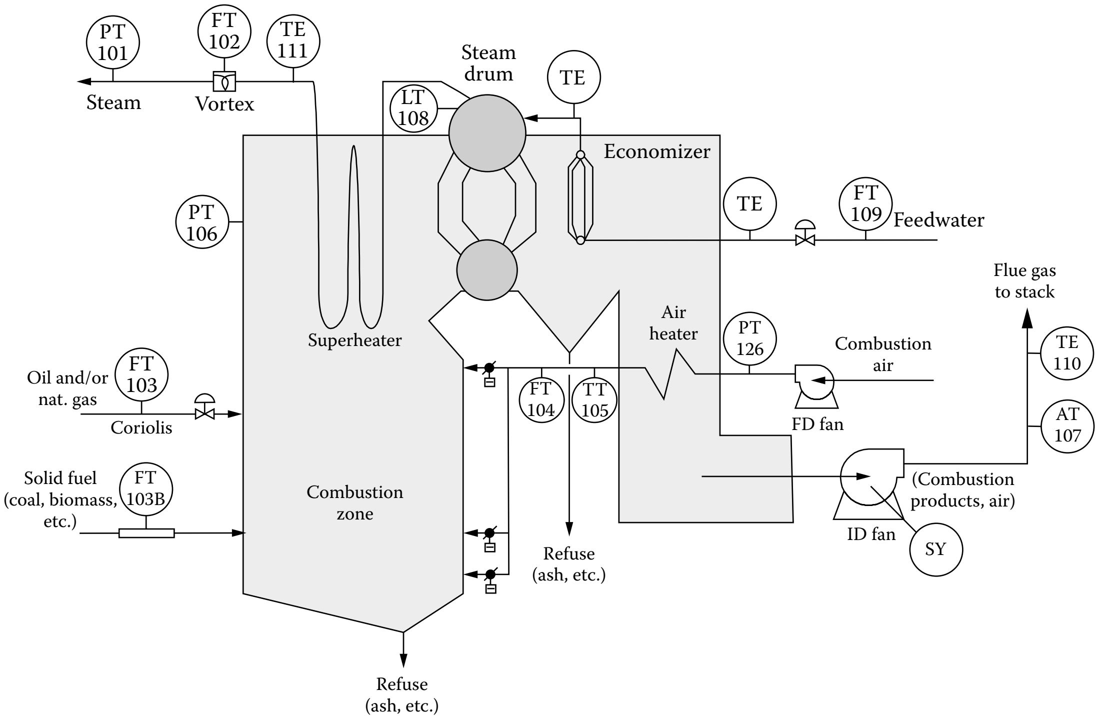


FIG. 8.6b

The main in-line instruments are shown here for a drum-type boiler.

Equipment

Steam boilers referred to in this section are drum-type boilers. Very large, supercritical pressure boilers are the "once-through" type and are found only in the largest electric generating plants.

In electric utility applications, a boiler is typically part of a *generating unit:* one boiler dedicated to one steam turbine. In industrial applications, often two or more steam boilers are connected to a common header supplying process steam users and, commonly, one or more turbine generators. The "load" on a steam boiler refers to the amount of steam demanded by the steam users (including turbines).

The boiler steaming rate must follow the steam demands arising from process heat or power generation requirements. The equipment and the control system often must be capable of satisfying rapid changes in load. Load changes can be a result of rapidly changing process requirements, power demand changes, or cycling control equipment. Whereas load may be constant and steady over prolonged periods, the boiler must have sufficient "turndown" to stay in operation at reduced capacities as portions of the plant may be shut down. This consideration usually leads to a greater "turndown" requirement for the boilers than for any other portion of the plant. At the same time, it is desirable to maximize boiler efficiency at all loads.

Boiler designs, in terms of air and gas flow configuration, are generally either forced-draft (FD) or balanced-draft boilers. Forced-draft boilers operate at positive pressure with air supplied to the boiler by a forced-draft fan. Balanced-draft boilers usually operate at slightly negative pressure with air supplied by an FD fan and flue gas withdrawn by an induced-draft (ID) fan (larger than the FD, due to the combustion products). Most boiler air and flue-gas fans are centrifugal (typically with backward-curved blades). Axial flow fans are used less often due to the nature of the axial flow static-pressure vs. capacity curve and the possibility of stall conditions.[7]

The Role of Sensors

Figure 8.6b shows a typical boiler arrangement for gas, oil, or solid fuel. It also shows the in-line instruments used on a boiler, together with some advice on the type of sensor to be used.

The normal "on-line" requirements for steam boilers serve to control steam pressure within ±1% of the desired pressure; air/fuel ratio within ±2% of excess air (±0.4% of excess oxygen), based on a desired "load" vs. "excess air" curve; steam drum water level within ±1 in. of desired level; and steam temperature (where provision is made for its control) within ±10°F (5.6°C) of desired temperature. In addition, the efficiency of the boiler should be monitored within ±1%.

In order to reach these performance goals, it is necessary to install accurate sensors and to make sure that the load does not change more than 10–35% of full scale per minute, depending on the size, fuel type, boiler design, and that there are no boiler design problems limiting this ability. The various loops tend to interact, so that integration into an overall system is necessary both during design and when the loops are being "field-tuned."

Flow Detectors Important and often disregarded are the flow detectors, which provide the basis for both material and heat balance controls. Outlet steam flow measurement, particularly if it is used as part of the boiler firing rate control strategy, for environmental permit compliance, or for on-line efficiency or energy use calculations, should be pressure- and temperature-corrected to a true mass flow. Most steam flow sensors in use (orifice, flow nozzle, vortex-shedding meters, and so on) measure velocity, which translates directly to volumetric flow.

Multivariable transmitters with integrated pressure and temperature inputs and mass flow computation in the meter are preferred. If an uncompensated velocity or volumetric flow signal is all that is available, then it should be compensated to a true mass flow measurement in the control system using Equation 8.6(2):

$$F_C = F_A\left(\frac{P+P_0}{P_R}\right)\left(\frac{T_R}{T+T_0}\right)\left(\frac{X}{X_R}\right)\left(\frac{Q_R}{Q}\right) \qquad \textbf{8.6(2)}$$

where

F_C = compensated flow (mass flow)
F_A = uncompensated flow (velocity or volumetric measurement)
P = actual measured steam pressure
P_0 = conversion to absolute pressure (14.7 psi for imperial units and P in PSIG)
P_R = reference pressure (pressure at which primary flow element was specified and sized, converted to same units as $P + P_0$)
T = actual measured steam temperature
T_0 = conversion to Rankine or Kelvin scale (459.69 for °F to °R, 273.15 for °C to °K)
T_R = reference temperature (temperature at which primary flow element was specified and sized, converted to same units as $T + T_0$)
X = measured actual steam compressibility
X_R = reference steam compressibility (at conditions for which primary flow element was specified and sized)
Q = measured actual steam quality
Q_R = reference steam quality (at conditions for which primary flow element was specified and sized)

In practice, the compressibility and quality terms are often dropped, lacking a good measure of the actual values for these. In most steam flow applications, the pressure compensation term is the most important one. Note also that the above equation assumes that, for a differential pressure-type flow measurement, the process variable, F_A, has already had square root extraction applied to convert it to velocity.

For successful control of the air/fuel ratio, combustion air flow measurement is important. In the past it was impossible to obtain ideal flow detection conditions. Therefore, the practice was to provide some device in the flow path of combustion air or combustion gases and to field-calibrate it by running combustion tests on the boiler.

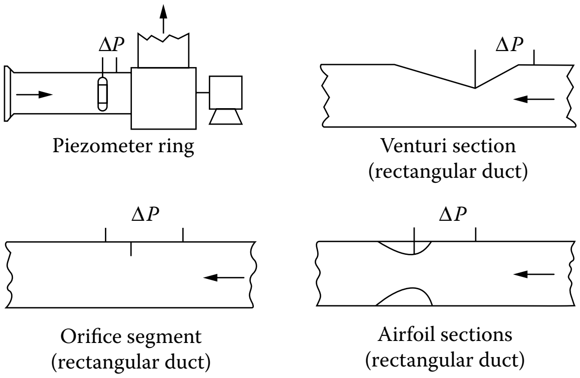


FIG. 8.6c
The various methods of air flow measurement.

These field tests, carried out at various boiler loads, used fuel flow measurement (direct or inferred from steam flow) and measurements of percentage of excess air by gas analysis; they also used the combustion equations to determine air flow. Because what is desired is a relative measurement with respect to fuel flow, the air flow measurement under these circumstances has historically been calibrated and presented on a relative basis.

Flow vs. differential pressure characteristics, compensations for normal variations in temperature, and variations in desired excess air as a function of load are all included in the calibration. With this traditional approach, using relative measurements, the desired result is to have the air flow signal match the steam or fuel flow signals when combustion conditions are as desired.

The following sources of pressure differential are normally considered:

- Burner differential (windbox pressure minus furnace pressure)
- Boiler differential (differential across baffle in combustion gas stream)
- Air heater differential (gas side differential)
- Air heater differential (air side differential)
- Venturi section or flow tube (installed in stack)
- Piezometer ring (at forced-draft fan inlet)
- Venturi section (section of forced-draft duct)
- Orifice segments (section of forced-draft duct)
- Airfoil segments (section of forced-draft duct)

Of these, the most desirable are the last four, because they use a primary element designed for the purpose of flow detection and measure flow on the clean-air side. Some of these typical traditional installations are shown in Figure 8.6c; none of these sensors meet the dual requirement of high accuracy and rangeability. In fact, they are of little value at 30% flow or less.

Flow Sensor Accuracy Table 8.6d lists some better flow sensors, such as the multipoint thermal flow probe or the area-averaging pitot stations provided with “hexcel”-type straightening vanes and with membrane-type pressure balancing *d/p* cells. Area-averaging pitot stations are also available in two-dimensional arrays, in a circular or rectangular

TABLE 8.6d
Flow Sensor Errors on Boilers

Flow Streams Measured	*Type of Flowmeter*	*Inaccuracy (% of flow)*			*Rangeability*	*Limitation*
		At 10%	*33%*	*100%*		
Fuel (oil)	Coriolis mass flow	0.5	0.5	0.5	20:1	
Fuel (natural gas)	Coriolis	0.5	0.5	0.5	20:1	
	Thermal	5.75	2.27	1.25	100:1	
	Turbine	0.5	0.5	0.5	10:1 to 100:1	
	Ultrasonic	0.1–1.0	0.1–1.0	0.1–1.0	30:1 to 50:1	
Fuel (solid: coal, wood, etc.)	Gravimetric feeders or belt scales	0.25–0.5	0.25–0.5	0.25–0.5	3:1	
Fuel (pulverized coal)	Coriolis			0.5	3:1	Max. rate ~29 tons/hr
	Microwave			5.0	5:1	
Steam and water	Vortex shedding					
	Steam	1–1.5	1–1.5	1–1.5	10:1	Min. R_e = 20,000
	Water	0.5–1	0.5–1	0.5–1		Max. temp = 750°F (400°C)
	Orifice	NG*	2–5	0.5	3:1	Rangeability can be increased if using two conventional *d/p* cells or a "smart" *d/p* cell
Air	Area averaging pitot traverse station	NG	2–10	0.5–2	3:1	
	Multipoint thermal	5–20	2–5	1–2	10:1	Dual-range unit required
	Piezometer ring, orifice segment Venturi section airfoil section	NG	3–20	2–3	3:1	Cannot be used below 25% of max. flow

*NG = Not Good.

mounting, that can average the entire flow-field. Thermal (hot-wire anemometer) sensor arrays can also be fairly accurate, offer good turndown, and produce a signal in direct proportion to mass flow. These represent major advances in combustion air flow detection.

This table shows the measurement errors that can be anticipated. Unfortunately, as the flow is reduced, the error—in percentage of actual measurement—increases in all cases except the first two. With linear flowmeters, the error increases linearly with turndown of 10:1. In case of nonlinear flowmeters, the error increases exponentially with turndown. Therefore, at a turndown of 10:1, the orifice or pitot error increases 100-fold and causes these devices to become useless. This situation can be alleviated somewhat by the use of two *d/p* cells on the same element or by the use of "smart" units.

Based on the data in Table 8.6d, if the boiler efficiency is to be monitored on the basis of time-averaged fuel and steam flows, the lowest error that can be hoped for is around 1%.

Similarly, the air/fuel ratio cannot be measured to a greater accuracy than the air flow. At high turndown ratios, this error can be very high. Considering that a 2% reduction in excess oxygen will increase the boiler efficiency by 1%, both the accurate measurement and the precise control of air flows are essential in boiler optimization. If combustion air temperature and pressure vary significantly, then air flow measurement can also be pressure- and temperature-compensated to a true mass flow, again, either with multivariable transmitters, or with Equation 8.6(2), dropping the compressibility and quality terms.

The impact of sensor inaccuracy on performance optimization elsewhere is not as critical. Standard instrumentation allows for the control of steam pressure within ±1%, furnace pressure within ±0.1 in. H_2O (25 Pa), water level within ±1 in. of desired level, and steam temperature to within ±10°F (5.6°C).

Inferential Measurements Inferential measurements, or soft sensors, are finding application in boiler measurement and control due to the difficulty or expense in directly measuring many of the important process variables related to operation of a boiler. Variables such as NO_x emissions, steaming rate, and turbine shaft temperature have been successfully measured inferentially. Two techniques that have successfully been employed are *neural networks* and *principal component analysis*. See Section 2.18 in Chapter 2 of this volume for discussion of neural networks. Regulatory agencies have accepted neural network-based emissions measurement and reporting in certain parts of the United States.

Safety Interlocks

Many of the interlocks related to the start-up, shutdown, and operation of a boiler are implemented for the purposes of protecting personnel and equipment. Most of the interlock and safety features directly related to the boiler can be classified as either *burner management* or *combustion control.* This delineation is made because boiler safety standards define very specific functions for burner management and require it to be implemented in a dedicated system, separate and apart from other control functions.[8]

A burner management system (BMS) is primarily concerned with the interlock, sequence, and timing functions required to safely put burners into service and to stop fuel and trip the boiler on detection of potentially unsafe conditions (*master fuel trip*). Other combustion control interlocks and protection functions, not necessarily a part of BMS, include furnace draft (implosion protection) control, fuel/air cross-limiting, and "runbacks."

An overview of some of the most common boiler safety interlocks is as follows:

PURGE INTERLOCK Prevents fuel from being admitted to an unfired furnace until the furnace has been thoroughly air purged.

LOW AIR FLOW INTERLOCK OR FAN INTERLOCK Fuel is shut off upon loss of air flow or combustion air fan or blower.

LOW FUEL SUPPLY INTERLOCK Fuel is shut off upon loss of fuel supply that would otherwise result in unstable flame conditions.

LOSS FLAME INTERLOCK All fuel is shut off upon loss of flame in the furnace, or fuel to an individual burner is shut off upon loss of flame to that burner.

FAN INTERLOCK Stops forced draft upon loss of induced-draft fan.

LOW WATER INTERLOCK (OPTIONAL) Shuts off fuel on low water level in boiler drum.

HIGH COMBUSTIBLES INTERLOCK (OPTIONAL) Shuts off fuel on highly combustible content in the flue gases.

Where fans are operated in parallel, an additional interlock is required to close the shut-off dampers of either fan when it is not in operation. This is necessary to prevent air recirculation around the operating fan.

Burner Management Systems BMS interlocks must be implemented with dedicated systems. They can be accomplished by hard-wired relay logic, solid-state logic, or programmable logic controllers (PLCs). The BMS is considered a safety instrumented system (SIS). Therefore, if PLC technology is used, it is often based on 1oo2 (one-out-of-two) or 2oo3d (two-out-of-three with diagnostics) logic.

The first one has two channels (two independent CPUs); the second has three channels (three independent CPUs, as in *triple modular redundant* systems). The criterion for selecting a certain type of SIS equipment is based on a safety integrity level (SIL) assessment that determines the degree of integrity required of the SIS based on probability and impact severity of risks. The Instrumentation, Systems, and Automation Society (ISA) has published detailed standards on SIL assessment.

The minimum interlocks required by the National Fire Protection Association (NFPA) for basic furnace protection for a multiple-burner boiler[8] are illustrated in Figure 8.6e. An important part of the burner logic is the purge system, illustrated at the bottom of Figure 8.6e. *Master fuel trip* cannot be reset; i.e. the burner light-off sequence cannot be started unless a proper purge sequence has been executed. The purge helps prevent accumulation of unburned fuel in the boiler after trips or failed start sequences. Furnace explosions are often related to accumulation of unburned fuel.

Automatic start-up sequencing for lighting the burners and for sequencing them in and out of operation is common. Timing for a portion of the typical light-off sequence for a 350 MW dual-fired (oil and coal) burner with gas igniter is illustrated in Figure 8.6f.

Combustion Control Safety Features Additional safety features required by code that are considered the responsibility of the combustion control system (as opposed to the BMS) are primarily aimed at:

1. Maintaining proper combustion zone conditions (air and fuel, air/fuel ratio) to support complete and safe combustion
2. Preventing furnace implosion

Fuel demand should never exceed the capability of the combustion air system to supply necessary air for complete combustion. If an ID or FD fan trips in single-fan systems, then a master fuel trip occurs. If there are fans in parallel, as in large electric utility units, then *runbacks* are employed in the control logic. This technique applies to any of the critical boiler equipment that can be operated in parallel: FD and ID fans, feedwater pumps, and so on.

In these cases, interlocks are provided that "run back" the boiler firing rate to an operating point that can safely be supported by only one fan or pump when one of a pair fails. For example, if one of a pair of FD fans fails, the boiler firing rate must, in a rapid controlled fashion, be brought down to a point that the air supplied by the one fan is more than adequate for complete combustion.

Another feature to ensure sufficient air/fuel ratio for complete combustion is *air/fuel cross-limiting.* This is a safety feature guaranteeing that no change in the firing rate (up or down) can result in a "fuel-rich" mixture. The consequence of this feature is that, on increasing firing rate, air increases before fuel, and, on decreasing firing rate, fuel decreases before air (greater than or equal to the sum of the stoichiometric plus minimum excess air requirement for the fuel being burned).

If the FD and ID fan capacities, combined with the total dynamic head characteristics of the entire air and gas path,

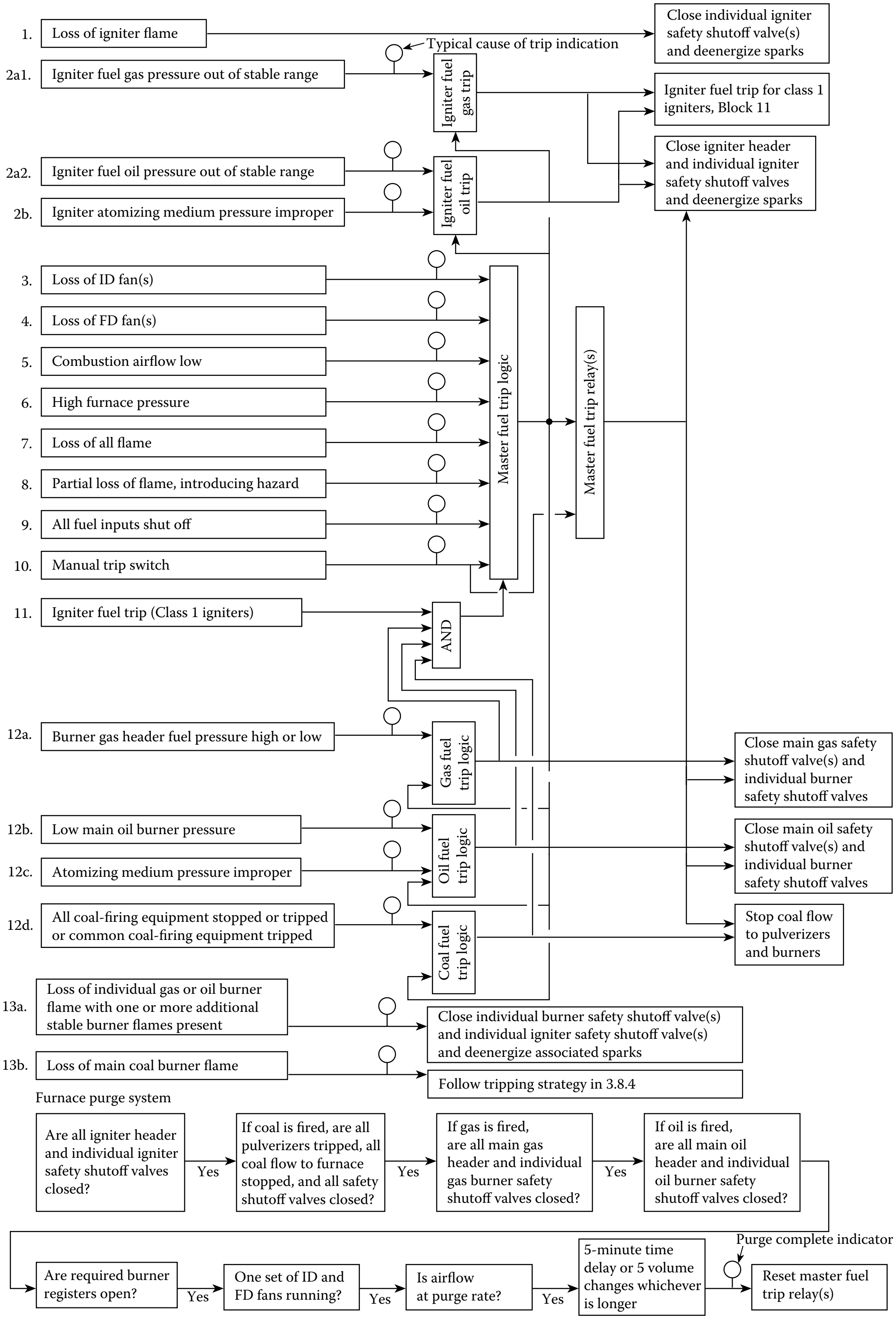


FIG. 8.6e

Interlock system for multiple burner boiler. (Reprinted with permission from NFPA 85-2004, "Boiler and Combustion Systems Hazards Code," copyright ©2004, National Fire Protection Association, Quincy, MA 02269. This reprinted material is not the complete and official position of the NFPA on the referenced subject, which is represented only by the standard in its entirety.)

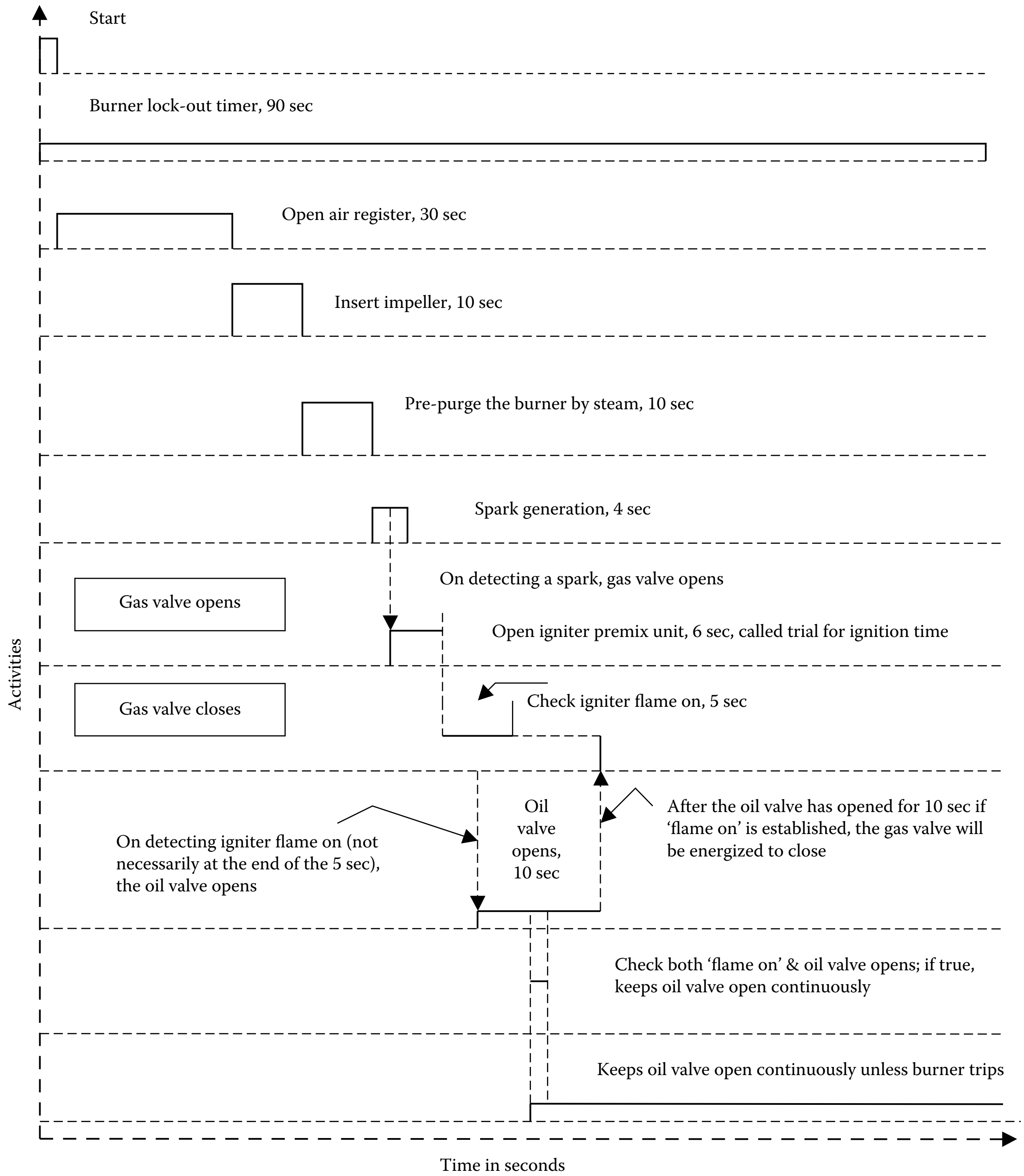


FIG. 8.6f

Timing diagram for typical multifuel burner light-off sequence.

can produce a draft or pressure that will exceed the design pressure ratings of the boiler or its associated ductwork, then the furnace pressure control system becomes more critical. In these cases, the furnace pressure control strategy should include features such as redundant furnace pressure transmitters, feedforward action from master fuel trip, and override action or directional blocking on large furnace draft error (Figure 8.6g).

Design and specification of the various safety interlocks are largely guided and governed by insurance company regulations, standards bodies such as the NFPA, and state regulations. NFPA 85 specifically addresses boilers. Insurance company standards

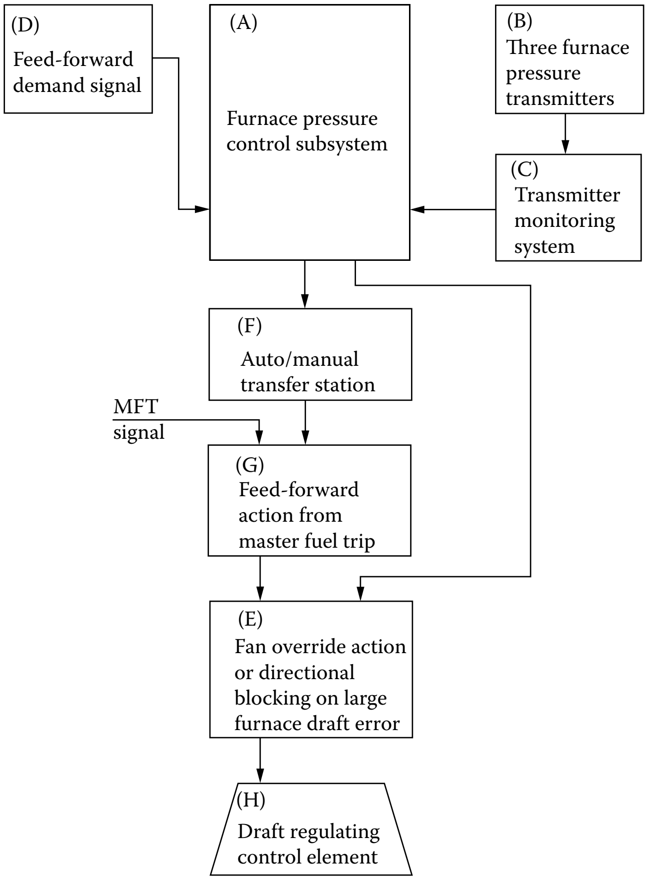


FIG. 8.6g
System requirements for furnace pressure protection and control when such pressure or vacuum can be applied that exceeds boiler or duct pressure ratings. (Reprinted with permission from NFPA 85-2004, "Boiler and Combustion Systems Hazards Code," copyright ©2004, National Fire Protection Association, Quincy, MA 02269. This reprinted material is not the complete and official position of the NFPA on the referenced subject, which is represented only by the standard in its entirety.)

may be more stringent than industry trade and professional organization (such as NFPA) standards.

Soot Blowers

Soot blowers are standard equipment on nearly all types of large water-tube boilers. Soot blowing equipment is used to periodically blow off deposits (fouling) that accumulate on the tubes on the inside of the boiler as a result of the combustion process. A clean heat-transfer surface plays a key role in achieving high thermal efficiency.

Steam, compressed air, or water have all been used as a blowing medium, with steam being the most common. Soot blowers installed in the furnace wall area are short fixed-length blowers (so-called IR blowers). Water lances are sometimes used for the furnace wall region, as well. The blowers located in the convection area, economizer area, and air heater area are long and retractable blowers (so-called IK blowers).

Recently, the water cannon has been introduced to improve cleaning on the furnace wall section.[9] Compared to a regular water lance, the water cannon has the following advantages: less maintenance work, higher cleaning efficiency, reduced NO_x, and longer tube life. Water is never used as a blowing medium, however, in black liquor recovery boilers (used in the pulp and paper industry) due to the potential for explosion from a smelt-water reaction.

Further, proper operation of soot blowers in recovery boilers becomes that much more critical due to the possibility of soot blower steam, directed at a localized area of tubes for too long, causing erosion-induced tube failure and, again, the catastrophic consequences of a smelt-water reaction from the failed water tubes.

In operation, soot blowers are often grouped into sequences based on their physical locations in the boiler. Each sequence consists of multiple steps. Each step may involve several blowers that can run simultaneously. Depending on the blowing medium limitations, normally three or four wall blowers can run at the same time, and one to two retractable blowers can run at the same time.

The operator can run sequences in a fixed schedule, or selectively pick up any individual sequence to run. An individual blow can always be selected whenever needed. Before any soot blower can start to run, the control system logic will first check if all permissives are met. The permissive conditions include whether the requested blower is in service, whether the required blowing medium is available, and whether the medium pressure is adequate.

If all starting permissives are met, the requested blower will start to run. In the retractable blower case, the blower travels forward to the end and reverses back. During the run time, the blower head pressure and the flow of the blowing medium have to be monitored throughout. A significant pressure or flow drop should cause the blower to stop and retract.

The control logic has to be programmed such that whenever abnormal conditions occur, an alarm signal is sent out and the troubled blower stops blowing and, in the IK blower case, fully retracts immediately. The most common failures are the following:

Fail to start: A blower has been commanded to start, but there is no indication that it actually started within an expected time interval.
Blow fail: Blowing medium pressure has dropped below an acceptable level for a specified time interval.
Motor overload or stall: The motor current of a retractable blower has exceeded its normal level by a set amount for a specified time period.
Elapsed time: A blower remained away from its rest position beyond a specified time.

Boiler Dynamics

Boiler response to load changes is usually limited by both equipment design and dead time considerations. Usually the maximum rate of load change that can be handled is from 20 to 100% per minute. This limitation is due to the maximum rates of change in burner flame propagation and to the "shrink/swell" effects on the water level. The period of oscillation of a typical boiler is between 2 and 5 min.

This is the result of a dead time of 30–60 sec and the integrating effect from the storage of energy (similar to tank level). The transportation delay in the boiler is partially due to the displacement volume of the furnace. For example, if the air/fuel ratio is changed, the furnace volume will have to be displaced before the flue-gas composition can reflect that change. The lower the air flow (the lower the load), the longer it will take to displace this fixed volume. Therefore, dead time increases as load is lowered on a boiler.

The transportation delay described above is only one component of the total dead time. The oxygen analyzer also contributes to the total delay, because of its location and its fly ash filter. In addition to the dead time contribution of instruments, control dead time is also created by the fuel/air cross-limiting.

Another way to reduce dead time and thereby increase boiler response is by using feedforward loops. The firing rate demand signal can be made more responsive by feedforward of steam flow, which responds to load changes faster than does steam pressure. Similarly, the induced-draft loop can be made more responsive by adding feedforward off the forced-draft position (damper, inlet vanes, or fan speed). In this system, as soon as the air flow into the furnace changes, the outflow is also modified in the same direction, so that furnace draft pressure is relatively unaffected.

Each of these loops will be discussed in some detail in the following paragraphs.

Air/Fuel Ratio Controls

Performance of air controls on traditional boilers has been limited by inaccurate and unreliable air flow sensors, particularly when air flow rates were less than 25% of maximum rates. This was unfortunate, because it is precisely at low loads that the boiler tends to be the least efficient to start with. Yet, for reasons of equipment inadequacy, some manufacturers will turn off the oxygen trim of the air/fuel ratio at low loads. Contributing to these problems were inaccurate sensors at low flows; leaking, nonlinear dampers with hysteresis and dead band; the use of constant-speed fans; and flame instability at low loads for certain boiler designs and fuels.

Legacy boiler control systems often have workarounds for such historical combustion air measurement and control challenges. In many cases, the firing rate signal itself has been characterized to set the excess air (Figure 8.6h); it has also been used as the oxygen set point on air/fuel ratio trim controls. Frequently, the excess air requirement curve was directly calibrated into the combustion system, and feedback trim based on excess oxygen was not used at all.

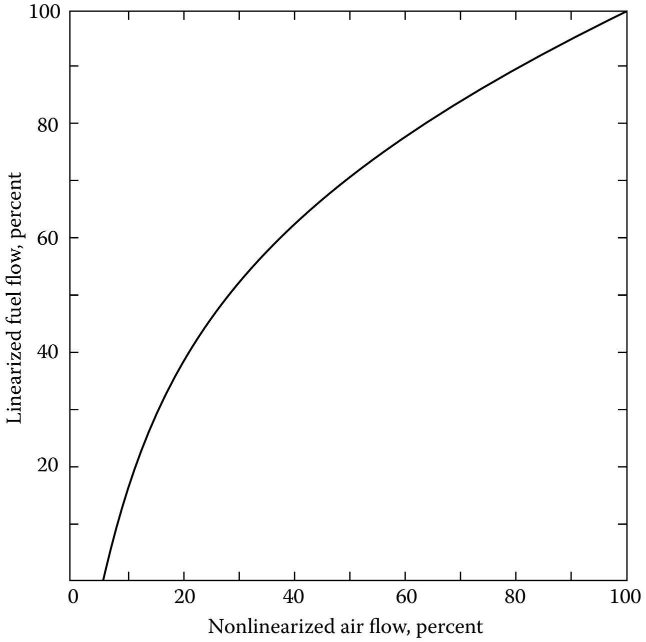


FIG. 8.6h
To apply air flow characterization, the boiler's fuel/air relationship must be determined empirically for the various boiler load levels.[14]

The advocates of this open-loop control strategy argue that the predetermined excess oxygen curve is rather permanent, and if an unmeasured effect necessitates a change in it, that change will be the same throughout the full firing range. This view of the process neglects nonlinearities, hysteresis, dead time, the play in linkages, sticking dampers, and other effects that are now understood.

In many traditional systems, the nonlinearity of dampers (Figure 8.6i) was taken into account by characterizing the fuel valve, the linkage, or the signal to the fan damper actuator. This is no easy task, because louver-type dampers are not only nonlinear, they also lack repeatability. In these systems, the air/fuel ratio trim was disabled below 25% load, because the conventional dampers could not be controlled.

It was argued that closed dampers leak as much as 10% of their full capacity and thus need to be opened only to 2% to deliver 25% flow. Therefore, according to this argument, if the oxygen trim signal were not disabled and it did request a 1% change in air/fuel ratio, this would mean a 1% change in the 2% opening of the damper, or a 0.02% change in its stroke or rotation; this the damper positioner could not handle. Naturally, the argument is correct in all respects except in its assumption that such dampers are a necessary limitation.

Other observations[10] include that PID-type controls cannot handle frequent load changes, because the speed of response of the fuel flow control loop is much faster than that

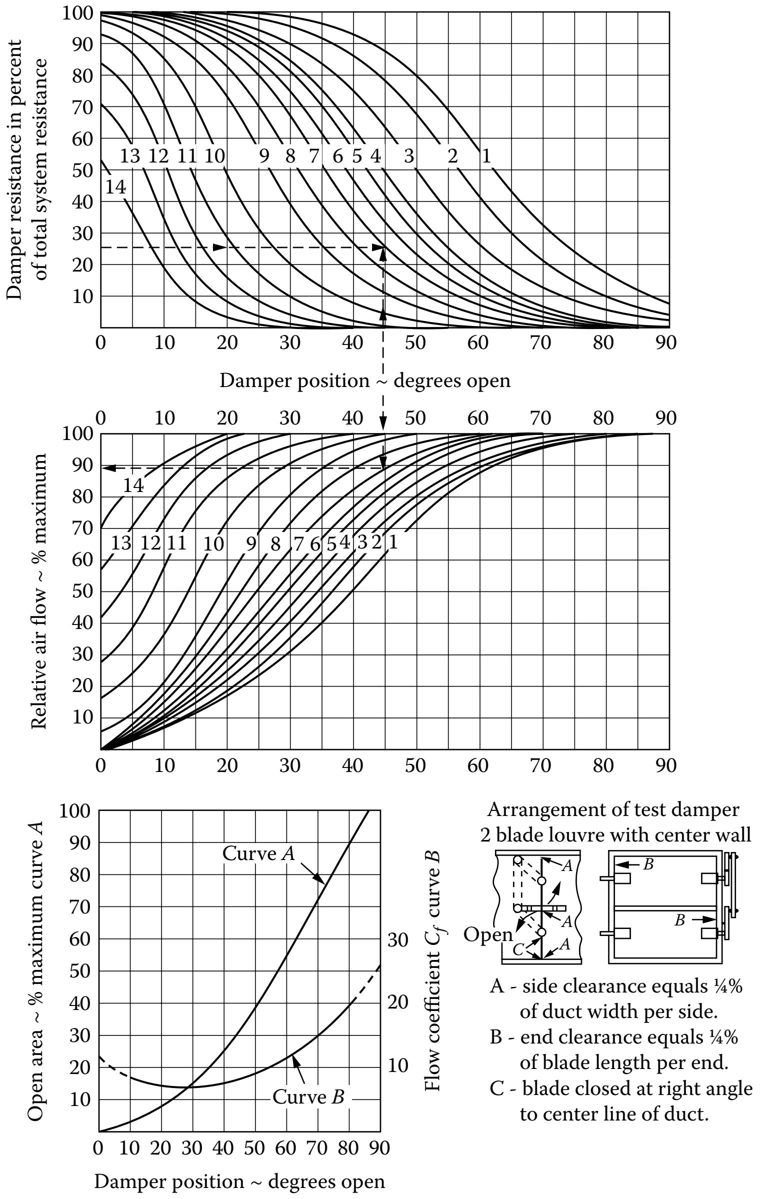


FIG. 8.6i
Multiple-leaf louvers with dividing partitions were used in many traditional systems.[11]

of the air flow control loop. Suppliers attempted to correct this by making the air actuator twice as fast as the fuel actuator.

In general, the nonlinear nature of the loops once created difficulties for the traditional boiler control designs. The need to change the firing rate in exact proportion to load changes was difficult to meet, because it required characterization in the field of the fuel valve and air damper actuators.

As will be discussed on the following pages, distributed microprocessor-based control systems make it much easier to linearize the nonlinear systems through characterizers and greater use of closed-loop control. It is also possible to memorize actuator dead band and speed of response. Still, the capability of state-of-the-art control equipment should not be used as a justification for installing inferior-quality dampers.

Dampers The recommendations for the use of accurate and high-rangeability sensors have already been made in connection with Table 8.6d. Now the desirable damper features will be discussed from a control performance viewpoint.

The reasons why dampers are undesirable as final control elements include their hysteresis, nonlinearity, and leakage. If the flow is accurately measured, the consequences of these undesirable features are easier to handle. For example, assuming that even under proper maintenance the damper displays some hysteresis or dead band, with a reliable flow sensor, the opening and closing characteristics of the damper can be determined.

In most digital control systems, this characterization curve can readily be accommodated, usually input as a series of paired points with linear interpolation point-to-point. The damper characteristic curve can also be automatically updated and corrected, as it might change because of wear, dirt build-up, or other causes.

Figure 8.6i shows the nonlinear characteristics of a multiple-blade damper.[11] Curve *A* on the bottom gives the relationship between damper rotation and open area, while curve *B* relates damper rotation and flow coefficient. The upper portion of Figure 8.6i describes the *installed* performance of the same damper, considering its pressure drop relative to the total system pressure drop.

For example, if the damper resistance is 25% of the total, when the damper rotation is 45%, curve #7 will give the damper characteristics, and actual air flow will be about 88% of maximum. From such curves as these, both the damper leakages and the nonlinearities can be estimated before actual installation.

It can be seen from Figure 8.6i that an increase of 1° in damper rotation that occurs at a 10° opening causes a much greater increase in the actual air flow than a 1° increase that occurs at a 60 or 70° opening. This is undesirable. Stable controls would require constant gain in the loop. Ideally, the change in air flow per increment change in damper opening should be uniform from the tightly closed to the wide-open position. Such inherently linear dampers are hard to manufacture, although the design illustrated in Figure 8.6j does hold such potentials. A more often used solution is the use of compensators and characterizer positioners, so that as the damper gain drops off—as it opens—these compensators introduce more gain into the loop, keeping its total gain nearly constant.

While it is possible to compensate for nonlinearity and hysteresis, leakage must be eliminated by selecting the correct design. Figures 8.2h and 8.6j illustrate some of the tight shut-off designs. If these are used, it is no longer necessary to turn off the excess oxygen-based air/fuel ratio trim when the load is below 25%. With such compensated, low-leakage dampers, trim need not be turned off, and the resulting increase in efficiency need not be abandoned, until load drops to 10%.

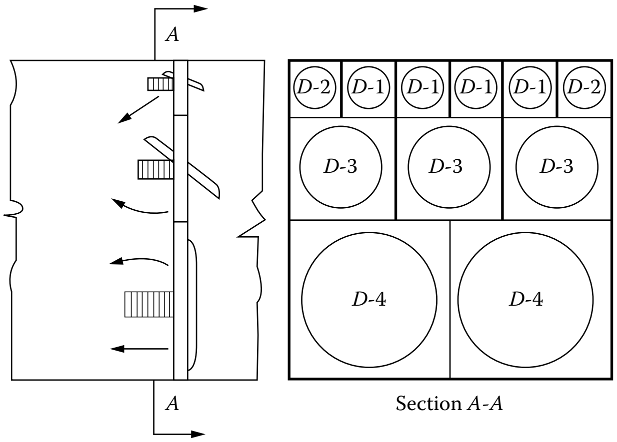


FIG. 8.6j
Improved dampers can be designed to provide low leakage (0.001% of maximum flow). (Courtesy of Mitco Corp.)

Nonlinearity and hysteresis, in the form of stiction and backlash, are often directly the result of the actuator itself. *Electric damper drives* have proven superior to pneumatic and hydraulic cylinder actuators and are becoming the norm for damper modulation.

Fans Even if the best damper is used, damper control will always have a disadvantage, namely that it burns up fan energy in order to control. Therefore, the best method of controlling air flow is at the fan: either by a variable-speed fan or by adjustable inlet vanes. By varying fan speed and eliminating a damper (or running it wide open), it is possible to avoid unnecessary energy loss in the form of damper pressure drop.

Variation of fan speed is accomplished with hydraulic or magnetic couplings, as well as variable-speed AC and DC electric drives. Adjustable inlet vanes, because they are integrated right into the suction area of the fan, are also much more efficient than dampers, though still not as efficient as varying the fan speed. Fan characteristic curves and a comparison of damper vs. inlet-vane vs. variable-speed control are illustrated in Figure 8.6k.

In addition to reducing the overall cost of boiler operation by conserving air transportation energy, variable-speed and variable inlet-vane fans offer better linearity, hysteresis, dead band, and leakage characteristics. For these reasons, the overall control diagram in Figure 8.6l shows fan speed or inlet-vane controls instead of damper controls.

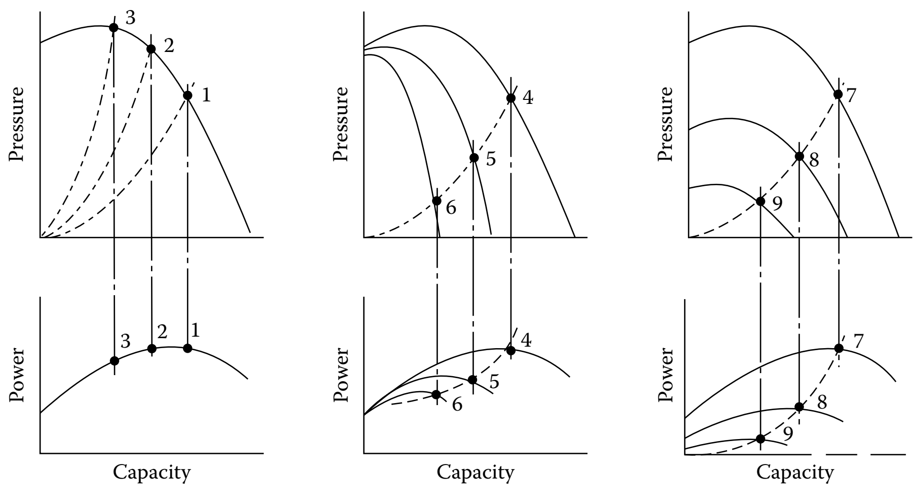
BASIC BOILER CONTROLS

Figure 8.6l shows a possible configuration of the basic boiler control loops and the tie-in points for optimization. Although this is a well-designed control system, it is just one of many possible configurations. Boiler size, steam pressure, and number and type of fuel(s) all can vary, necessitating variations in this scheme. The major loops shown in Figure 8.6l are numbered.

Boiler-Pressure and Firing Rate Controls

Realizing that the boiler is part of a larger plant system, consideration must be given as to how the boiler output will match the load demand placed upon it by the system. From a control perspective, it is useful to distinguish between electric utility and non-electric utility (i.e., industrial) boiler applications.

The firing rate of industrial boilers is most often manipulated to control to a constant header pressure. The primary controller that accomplishes this, whose output is firing rate, is usually termed the *boiler master*. In an industrial application,



(a) Discharge damper control

(b) Suction damper control

(c) Variable speed control

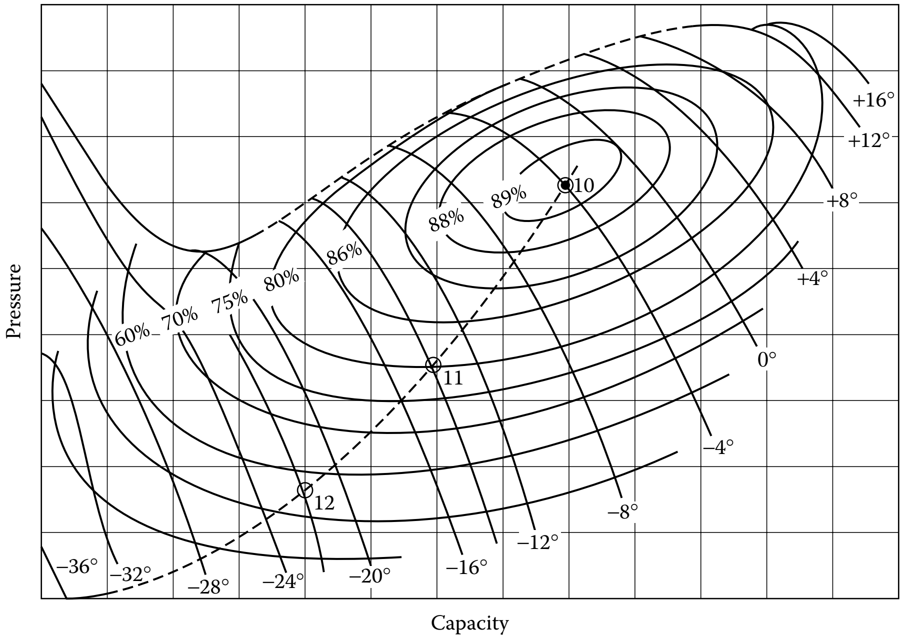


FIG. 8.6k
Axial-flow fans with variable-pitch control help eliminate the need to burn up unneeded air transportation energy in the form of damper pressure drops. Fan characteristics are shown with (a) damper control, (b) variable-inlet vane control, and (c) variable-speed control.[32]

multiple boilers may connect to one steam header, from which multiple turbine generators may be supplied as well as plant heating loads, through extraction turbines or pressure reducing valves (PRVs). In contrast, an electric utility boiler normally operates paired off as a single generation unit with a single turbine generator. The electrical output desired from the unit can be controlled in the following ways:

- *Boiler-following mode:* load demand is controlled first at the turbine generator, with the boiler supplying (following) whatever the turbine requires (header pressure control on boiler firing rate).
- *Turbine-following mode:* load demand is controlled first by boiler firing rate, with the turbine responding to (following) the boiler (throttle pressure control with turbine valves).
- *Boiler-turbine coordinated control mode:* turbine valves and boiler firing rate are manipulated in concert in response to load demand, while header pressure is maintained.
- *Sliding (variable-pressure) and free-pressure control modes:* control strategies that, by design, allow header pressure variation with load, to improve efficiency, speed of response to load changes, and turbine reliability.

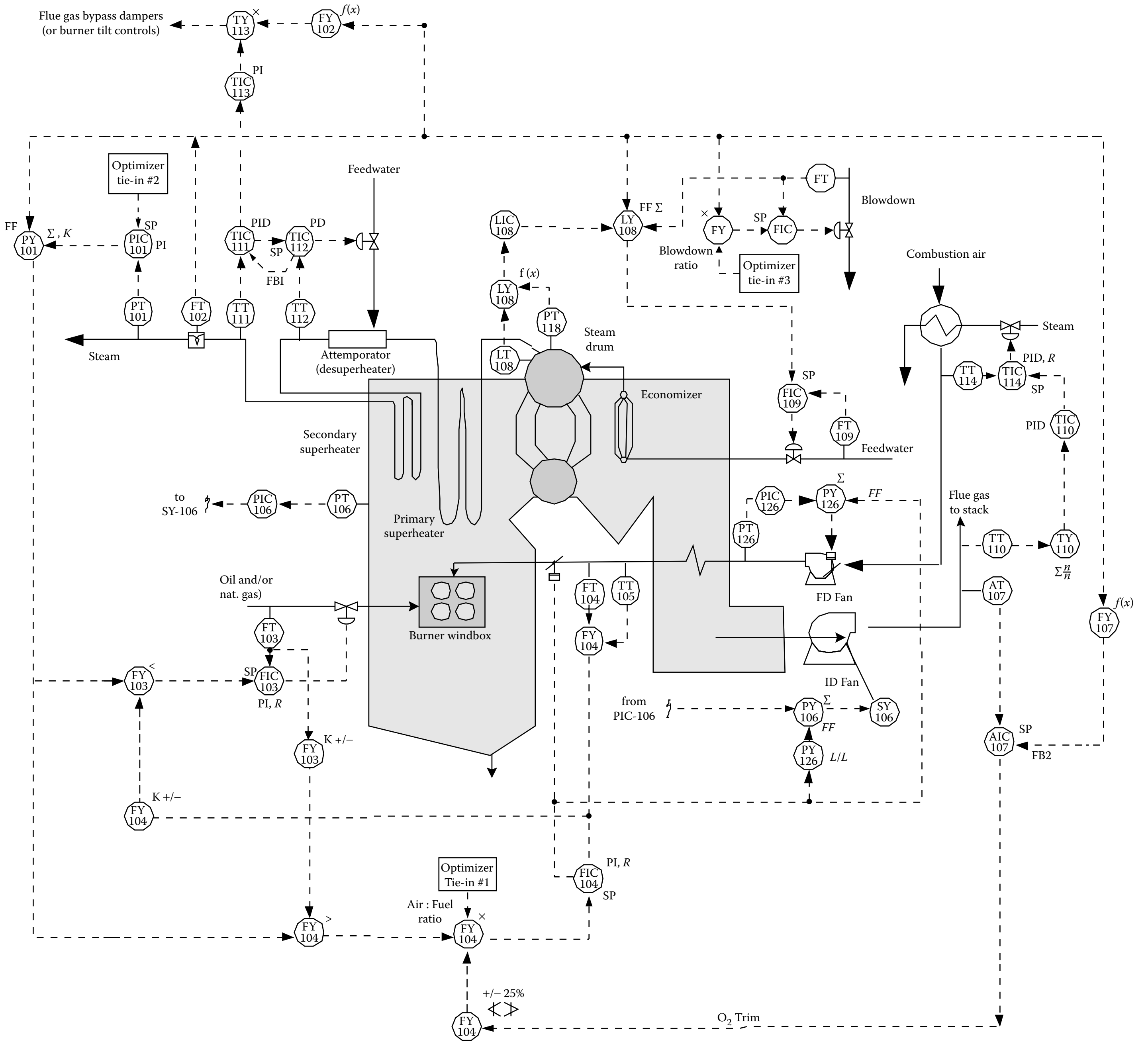


FIG. 8.6I
This diagram shows good boiler controls without optimization.

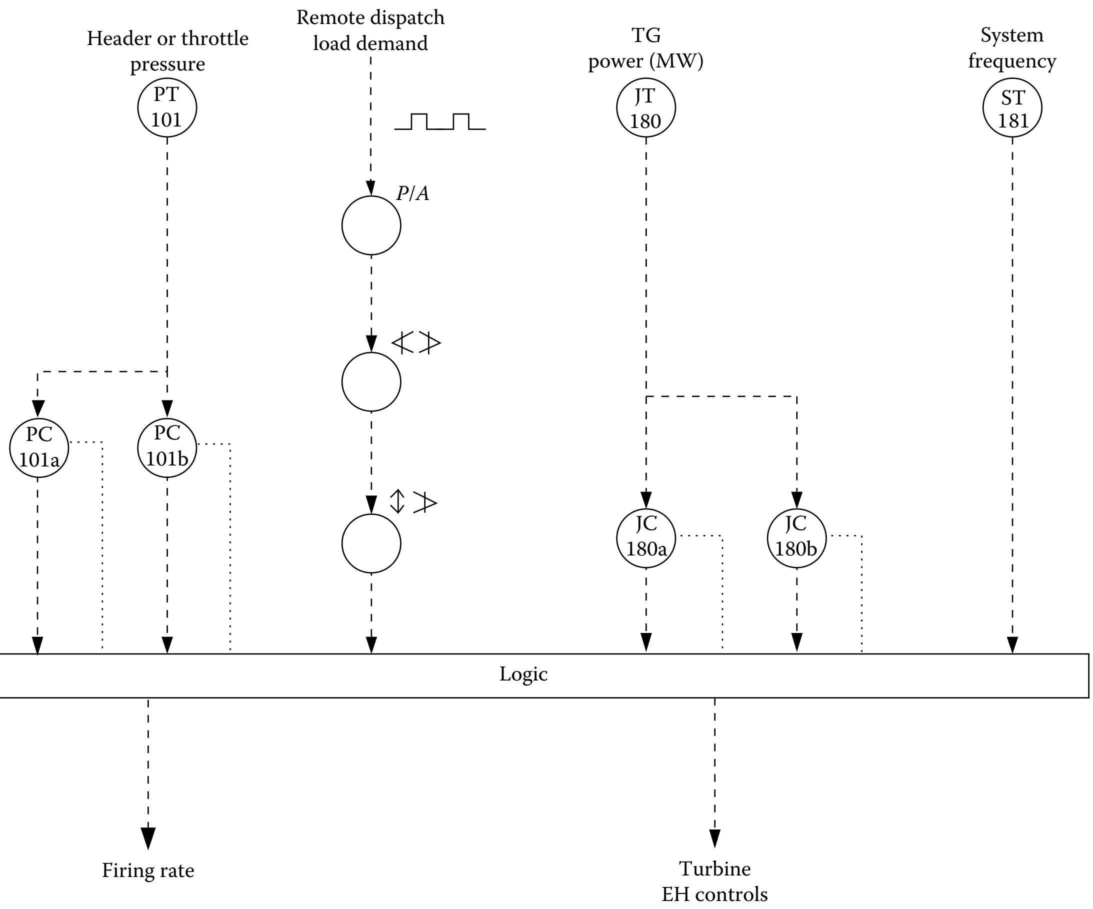


FIG. 8.6m
The load demand control architecture of an electric utility.

Turbine following is the most stable, but slowest mode; boiler following is faster, but limited to a maximum of about 2.5%/min for a controlled load change.[13] The coordinated control mode should be capable of 5%/min or better. Most control systems on large electric utility units are configured to allow selection from among the first two or three of these control modes, and, on some units, all four (some specific variant of the fourth scheme). Additionally, electric utility units usually have provision for the load demand to be generated remotely, referred to as remote dispatch.

A general, simplified structure for load demand control on a typical utility unit is illustrated in Figure 8.6m. The frequency input shown is sometimes necessary as an additional correction to prevent feedback control on turbine generator power (MW) or load demand (e.g., first-stage pressure) from counteracting the turbine generator's own power regulation response to frequency variations when connected to the electric power grid.

Firing Rate The firing rate signal becomes the set point for the fuel and air flow controls that make up the combustion control system. Additionally, *air/fuel ratio controls* provide for the correct stoichiometric amount of air to combust the fuel plus some percentage of excess air to account for nonideal mixing and imperfect combustion conditions. For safety purposes, fuel addition should be limited by the amount of available combustion air, and combustion air may need minimum limiting for flame stability.

Boiler outlet steam pressure is an indication of the balance between the inflow and the outflow of heat between energy supply and load. Therefore, by controlling the steam pressure, one can establish a balance between the demand for steam (process load) and the supply of steam (firing rate). A change in steam pressure will result from a change in firing rate only after a delay of a few seconds to a minute, depending on the boiler and the load level. Therefore, as will be examined in more detail, feedforward control, from load demand, is frequently employed to improve pressure control by adjusting firing rate (fuel and air) as soon as a load change is detected, instead of waiting for pressure to change first.

When more than one boiler is operated from the same master controller, the ability to individually bias and take control of each boiler should be provided, in addition to the ability to bias the master controller signal up or down when in automatic (Figure 8.6n).

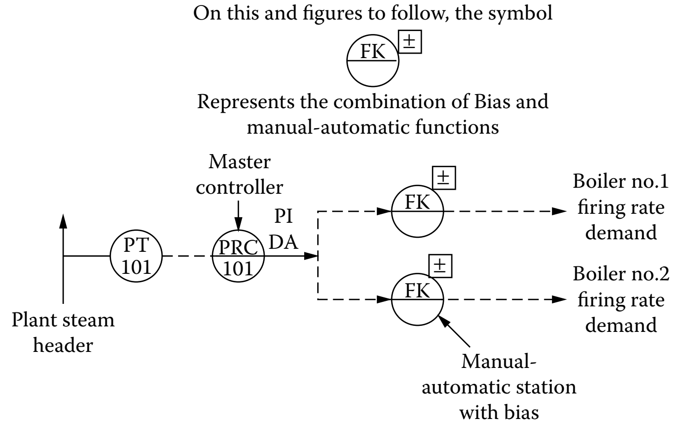


FIG. 8.6n
Load sharing controls among several boilers.

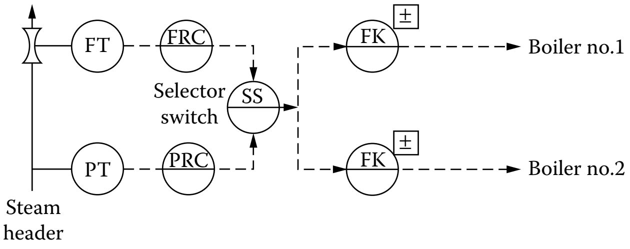


FIG. 8.6o
Boiler control with alternative pressure or flow master.

When steam pressure is controlled by other means, steam flow can be the master controller. If variations in fuel heating value are minor, the master flow controller shown might be eliminated and the master load signal generated by a manual loading station.

Situations may arise when it is desirable to have either flow or pressure control. In these cases, a master control arrangement, as shown in Figure 8.6o, can be used. Although it may appear simpler to switch transmitters, it is desirable to transfer the controller outputs so that the controller does not have to be returned each time the measurement is switched, and to make provision for initialization and bumpless transfer.

Typical proportional gain tuning setting for a pressure-controlled "master" is 6.25 (16% proportional band) and the typical integral setting is about 4.0 min (0.25 repeats per min). For a flow control "master" the comparable settings might be 1 for controller gain (100% proportional band) and 0.33 min for integral time (3.0 repeats per min). (See Sections 2.35 through 2.38 in Chapter 2 for details on controller tuning.)

A flow control master can be used for firing rate control on each of multiple boilers connected to a common header, with the flow controllers cascaded to the header pressure controller (Figure 8.6p).

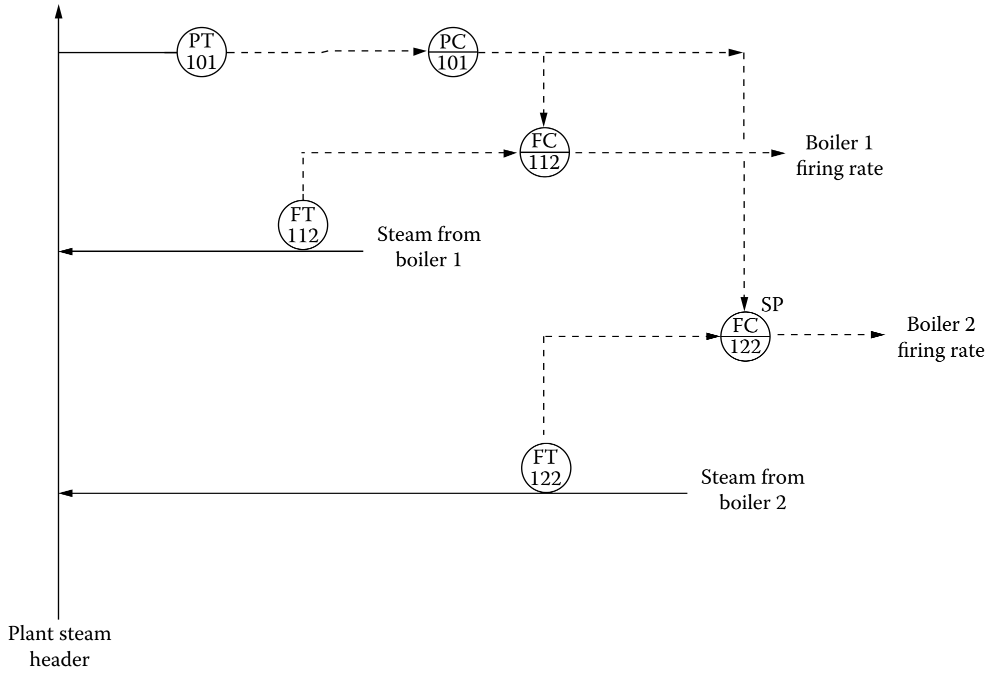


FIG. 8.6p
On multiple boilers connected to a common steam header, the header pressure controller can be the cascade master of a number of flow controllers, measuring the steam flows from the individual boilers.

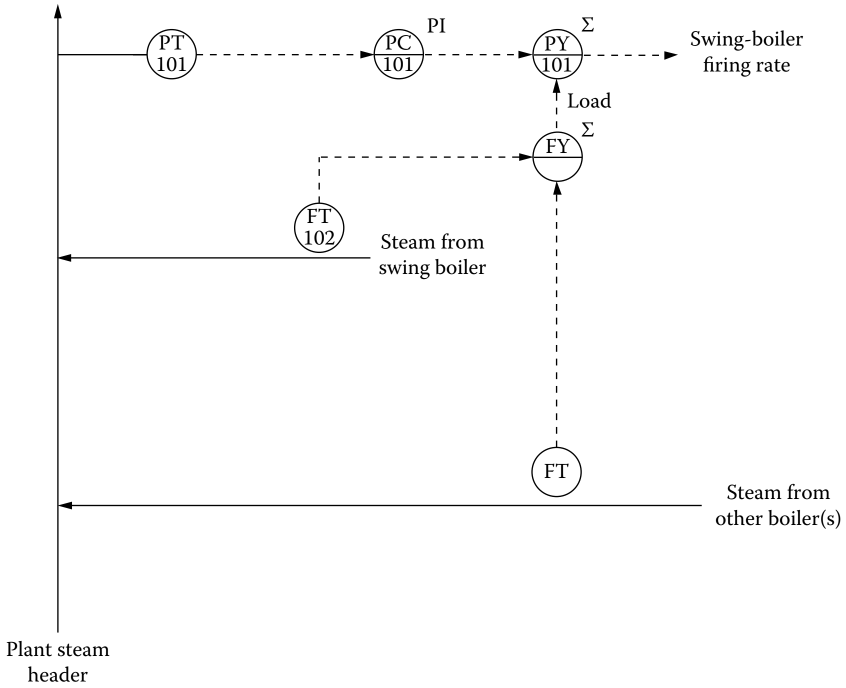


FIG. 8.6q
If the swing boiler is on header pressure control, the control dynamics can be improved by feedforwarding the measurement of the pressure master, by basing it on the sum of the steam flows from all the boilers.

Feedforward Control and Load Demand To improve the speed of response to load changes, a signal that represents load can be fed forward directly to firing rate. In the case of multiple boilers supplying steam to a header, load can be the total steam flow measurement from all boilers (Figure 8.6q) or the output of a high selector, selecting the highest of the steam flows. Steam flow measurements used in feedforward control should use mass flow detectors, either from a multivariable transmitter that computes mass flow or from a mass flow sensor; see Equation 8.6(2) and the paragraph titled The Role of Sensors.

If a boiler's own outlet flow is used as the load to feed forward to firing rate, the result will be positive feedback due to the impact of firing rate on flow. Used in this scenario, steam flow is called a *regenerative* load signal, and it will tend to cause instability and poor control. Even when combined with other steam flow measures (by average or high select) there will remain some amount of positive feedback.

On large electric utility boilers, due to the high temperatures and pressures, and large line sizes, steam flow measurement is often not provided at all. In that case, a good surrogate measure of load is the steam turbine first-stage shell pressure. The flow through the turbine varies almost linearly with first-stage shell pressure (Figure 8.6r).

In some cases (not to include sliding mode or free pressure control), it is also possible to use turbine inlet governor valve positions as an indication of load. However, first-stage shell pressure is generally preferred. These surrogate load parameters will behave more like *nonregenerative* signals, less coupled to boiler firing rate than boiler outlet steam flow. Therefore, even if steam flow is available it will be preferable in single-boiler arrangements to incorporate a surrogate load measurement for the purpose of feedforward firing rate control.

Feedforward Control and Stability The actual steam flow at any point is not necessarily a true indication of demand. For example, an increase in steam flow caused by increased firing should not be interpreted as a load increase—this would create a positive feedback loop, capable of destabilizing the boiler. According to Shinskey, the true load on a boiler can be approximated by $\sqrt{h/p}$, where h is the differential developed by a flow element and p is the steam pressure. Figure 8.6s describes a boiler-pressure control system using this type of feedforward model. Fuel flow is set proportionally to the estimate of load $\sqrt{h/p}$. Dynamic compensation is applied in the form of a lead-lag function to help overcome the heat capacity in the boiler.

The pressure controller adjusts the ratio of firing rate to estimated load, to correct for inaccuracy of the model and variations in heat of combustion of the fuel. The multiplier also changes the gain of the pressure feedback loop proportional to load. This feature is valuable in that boilers seem more difficult to control at lower loads because of lower velocities.

In the firing rate control system shown in Figure 8.6s, the loops 101 and 102 measure the pressure and flow of the generated steam, respectively. The flow transmitter (DPT-102)

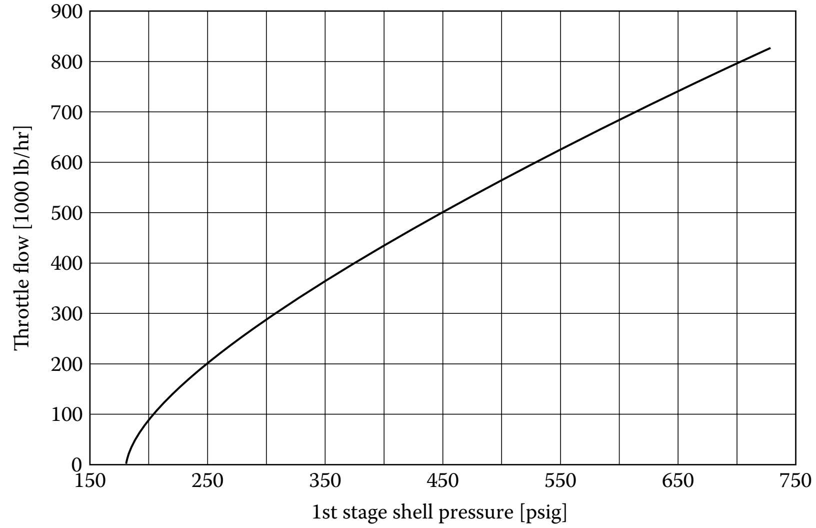


FIG. 8.6r
The relationship between the first-stage shell pressure of a steam turbine and the load (throttle flow); data taken from a typical 45 MW extraction (back-pressure) steam turbine.

is a high rangeability and linear device, capable of accurate measurements even at low loads. The firing rate demand signal is generated in a feedback manner by the pressure controller PIC-101.

In order to speed up the response of this loop to load changes, a feedforward trim is added. This trim is based on steam flow, because this flow is the first to respond to load changes. Therefore, as soon as the demand for steam changes, FY-102 will trim the firing demand signal, without waiting for the steam pressure to change. The dynamics of FY-102 are adjusted to reflect the time constants of the boiler, recognizing the time displacement between a change in firing rate and the resulting change in the rate of steam generation some time later.

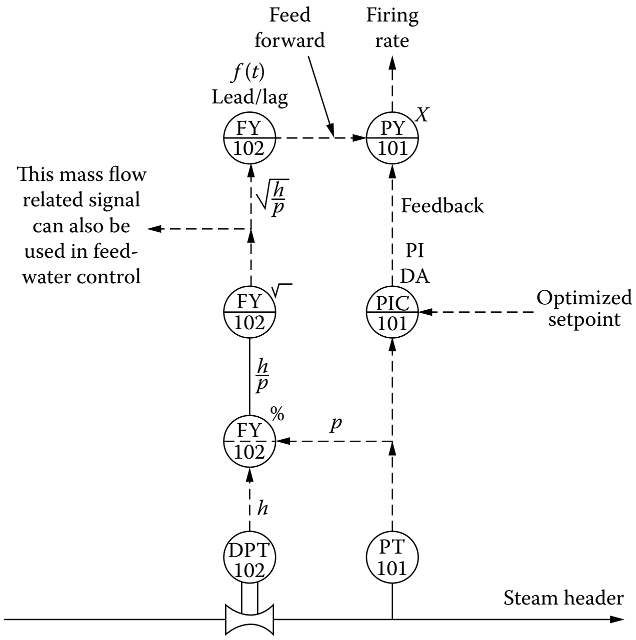


FIG. 8.6s
Firing rate determination using feedforward loop with feedback trim.

Fuel Controls

Measurable Fuels The primary boiler fuels are coal, oil, and gas, but there are a large variety of auxiliary fuels, such as waste gases, waste sludges, and waste wood products (bark, sawdust, hogged fuel, and coffee grounds). In many cases, these auxiliary fuels are dumped to the boiler plant on an uncontrolled basis for immediate burning. There are myriads of these combinations, and only the more common fuel control problems will be covered in the discussion to follow.

The control of gas and oil fuels tends to be straightforward, because they are easily measured and can be regulated with a control valve in the fuel line. If closed-loop control of flow is not available, then a valve positioner capable of providing a linear relationship between flow and control signal is desirable (Figure 8.6t).

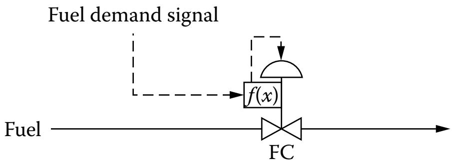


FIG. 8.6t
Positioner used to maintain linear relationship between demand and flow.

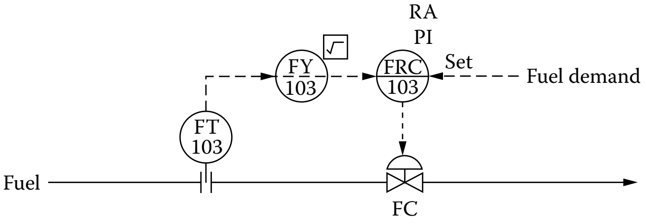


FIG. 8.6u
Fuel flow controller is used to keep demand and flow in linear relationship.

A flow control loop (preferred) is the more usual means of control, providing more precise linearization and immediate response to flow disturbances (Figure 8.6u). In cases in which better than 3:1 turndown and high measurement accuracy is desired, the Coriolis mass flow sensors should be used instead of orifices.

When it is desired to fire the fuels in a predetermined ratio to each other regardless of load, a manually adjustable signal splitter can be used, as shown in Figure 8.6v. The most precise and complex method of ratioing fuel (not shown) is to split the demand signal and send that to individual flow control loops. If the fuel is a gas at variable pressure, a pressure control valve is frequently installed upstream to the flow sensor, as shown in Figure 8.6v.

Both valves affect both variables (pressure and flow), and therefore they will interact. In order to eliminate the resulting oscillations, one should either leave the pressure unregulated and pressure-compensate the flow sensor or assign less pressure drop to the pressure control valve than to the flow control valve (thereby using a larger valve for pressure control than for flow control). Because the burner back-pressure will increase as flow increases, the available pressure differential for the flow control valve will decrease as the flow rises. In order to obtain an approximately linear relationship between fuel flow and valve position, an equal-percentage valve is needed.

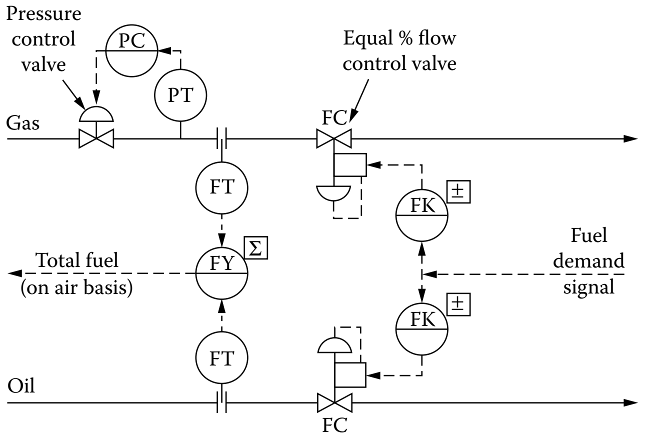


FIG. 8.6v
In this configuration, the fuel demand is manually split between the two fuels on an open-loop basis.

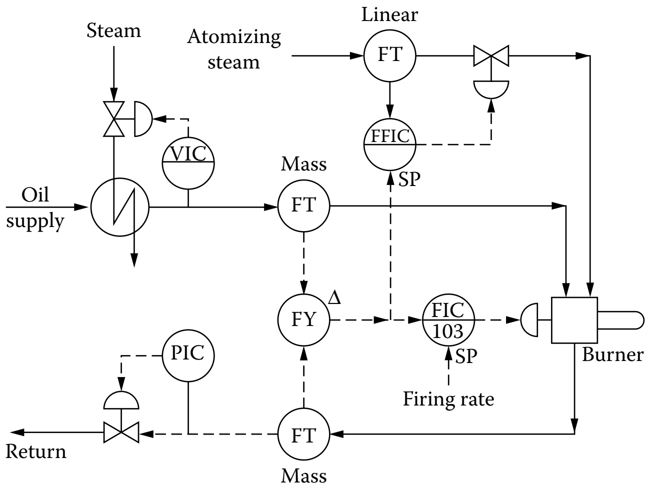


FIG. 8.6w
Oil flow control used for a recirculating burner that is provided with steam atomization.

In the case of oil fuels, proper atomization at the burner, and therefore complete combustion, will be achieved only if the oil is kept at constant pressure and viscosity. When heavy residual oils (e.g., no. 2 and no. 6 fuel oil) are burned, they must be continuously circulated past the burner and back (Figure 8.6w).

The difference between the readings of two Coriolis mass flowmeters indicates the net flow to the burner. The burner back-pressure is controlled by the control valve in the recirculating line, whereas the flow controller set point is adjusted by the firing rate demand signal. The firing rate is controlled by an alteration in the opening of the burner orifice. Atomizing steam is ratioed to the firing rate, and the heating steam is modulated to keep the fuel viscosity constant.

Figure 8.6x illustrates the controls required when the fuel demand is split between two fuels on a closed-loop (automatic)

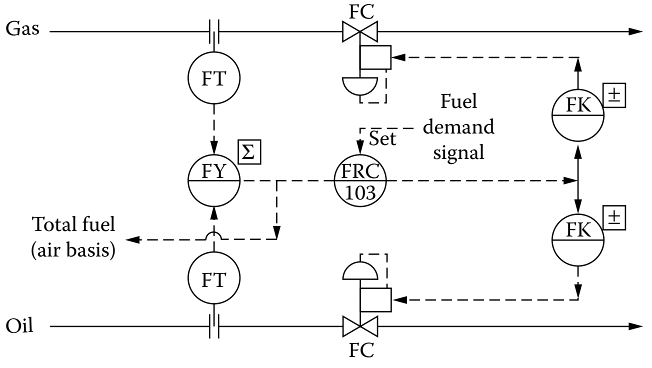


FIG. 8.6x
Fuel demand split between fuels on closed-loop basis.

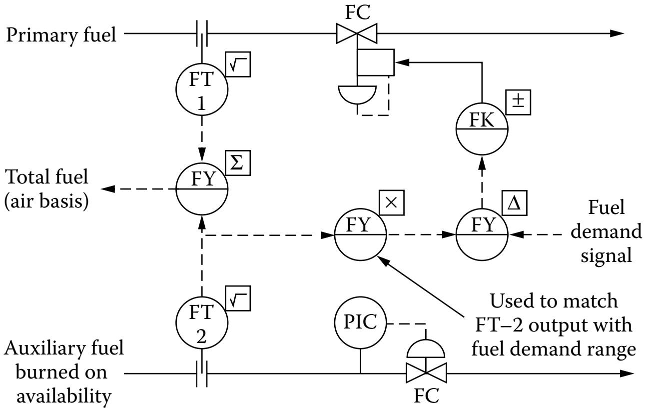


FIG. 8.6y
In this open-loop control configuration, the auxiliary fuel is burned on an uncontrolled-availability basis.

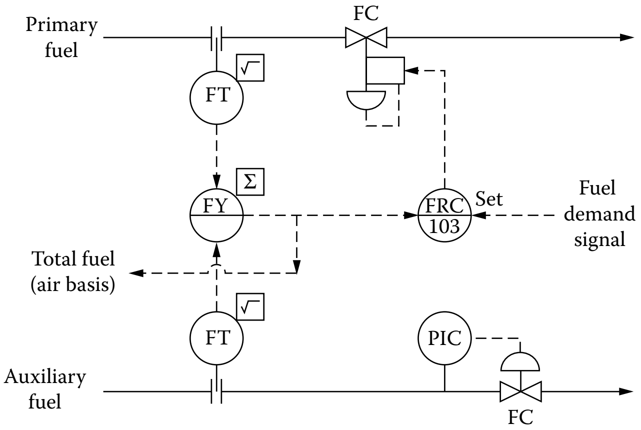


FIG. 8.6z
In this closed-loop control configuration the auxiliary fuel is burned on an uncontrolled-availability basis, while the flow controller throttles the primary fuel flow to meet the total demand.

basis. The instruments shown ("biasing stations") provide the means of manual control plus the ability of automatic control, with bias of one fuel with respect to the other.

Because one of the requirements ultimately is to have fuel ratioed to combustion air, any totalization of fuel for control purposes should be on an "air required for combustion" basis. If totalization is needed on any other basis, such as BTU for other purposes, a separate totalizer should be used.

Waste or Auxiliary Fuel Controls When auxiliary fuel is burned on an uncontrolled-availability basis, the fuel and air control system needs to be able to accommodate sudden changes in auxiliary flow without upsetting the master controller. The master controller should be designed and used to respond to total load demands only and not to correct for fuel upsets. A typical fuel control system for accommodating variations in auxiliary fuel without upsetting the master is shown in Figure 8.6y.

In the basis system without auxiliary fuel, the signal is relayed directly to the control valve. Addition of auxiliary fuel shifts the primary fuel control valve opening to prevent fuel variations from affecting overall boiler performance. A more precise system is shown in Figure 8.6z. Here, the flow controller adjusts the primary fuel control valve to satisfy total fuel demand and prevents auxiliary fuel variations from upsetting the master controller.

Figure 8.6aa describes a slightly more advanced system, in which the allowable maximum percentage of waste fuel that can be burned is set on the ratio relay FY-1. This ratio must be set under 100% if the heating value of the waste is so low that it could cause flameout if not enriched by supplemental fuel. For proper operation, the subtractor FY-2 must be scaled with the flowmeter ranges taken into consideration, and further scaling is required if the heat flow range of the total heat demand does not match that of the waste fuel flowmeter. When waste fuel gas availability becomes limited, waste fuel gas pressure will drop and PY-3 will select it for control, thereby overriding the waste flow controller. FY-2 will respond to this by increasing the supplemental fuel flow.

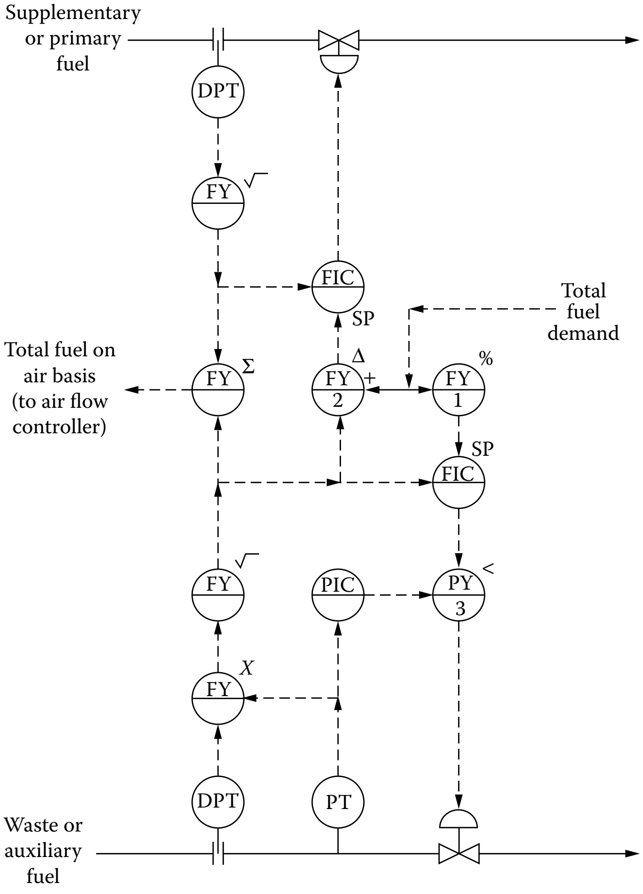


FIG. 8.6aa
Automatic control system for burning limited-availability waste fuel up to preset maximum percentage in the total mixture.

Closed-loop control will give greater precision and better linearization, but the performance can be limited by poor hardware selection or poor installation practices. An example of the first case is the problem of transmitter rangeability if orifices are used instead of turbine or mass flowmeters. An example of poor installation practices is the case of long transmission lines being allowed to introduce dead time into the loop because pneumatic leads were installed without boosters.

Whatever type of fuel control is used, the maximum flexibility in design will be present if all flow signals are linear and all control valve characteristics are also linear. In this manner, the various flows and signals can be combined, subtracted, multiplied, or divided to produce the desired control. One optimal condition is to have the total fuel demand signal linear, with fuel totalized on a basis of required combustion air. The other desired end condition is to have the total fuel control capacity maximum at a value approximately 10% greater than that required for maximum boiler-capacity. This excess is necessary for control flexibility at maximum boiler load. Additional excess capacity should not be considered, because it reduces turndown capability.

Unmeasured Fuels Coal can be an unmeasured fuel. In such cases coal control systems are open loop, wherein a control signal positions a coal-feeding device directly. This is the case with a spreader stoker or cyclone furnace or indirectly with pulverized coal.

A spreader stoker consists of a coal hopper on the boiler front with air jets or rotating paddles that flip the coal into the furnace, where a portion burns in suspension and the rest drops to a grate. Combustion air is admitted under the grate. There is no way to control fuel to a spreader stoker except in an open-loop manner by positioning a feeder lever that regulates coal to the paddles.

In pulverized coal-fired boilers, the coal is ground to a fine powder and is carried into the furnace by an air stream. There are normally two or more pulverizers (in parallel) per boiler. Pulverized coal flow is regulated at the pulverizer, and each manufacturer has a different design requiring different controls. One control arrangement is shown in Figure 8.6bb. Here, the primary air comes from a pressure fan that blows through the pulverizer, picking up the coal and transporting it to the furnace.

In addition to the controls shown, an air temperature control is required. In this loop, cold and hot combustion air is mixed ahead of the primary air fan to control the temperature of the coal air mixture in the pulverizer. This control is necessary to maintain a maximum safe operating temperature in the pulverizer. This is a simple feedback loop, usually involving proportional control only.

A control arrangement for a bowl-type pulverizer is shown in Figure 8.6cc. In this type of pulverizer, the air fan sucks air through the pulverizers with the fan (called an exhauster fan) located between the pulverizer and the burners. The coal-air temperature loop is similar to the one described in Figure 8.6bb.

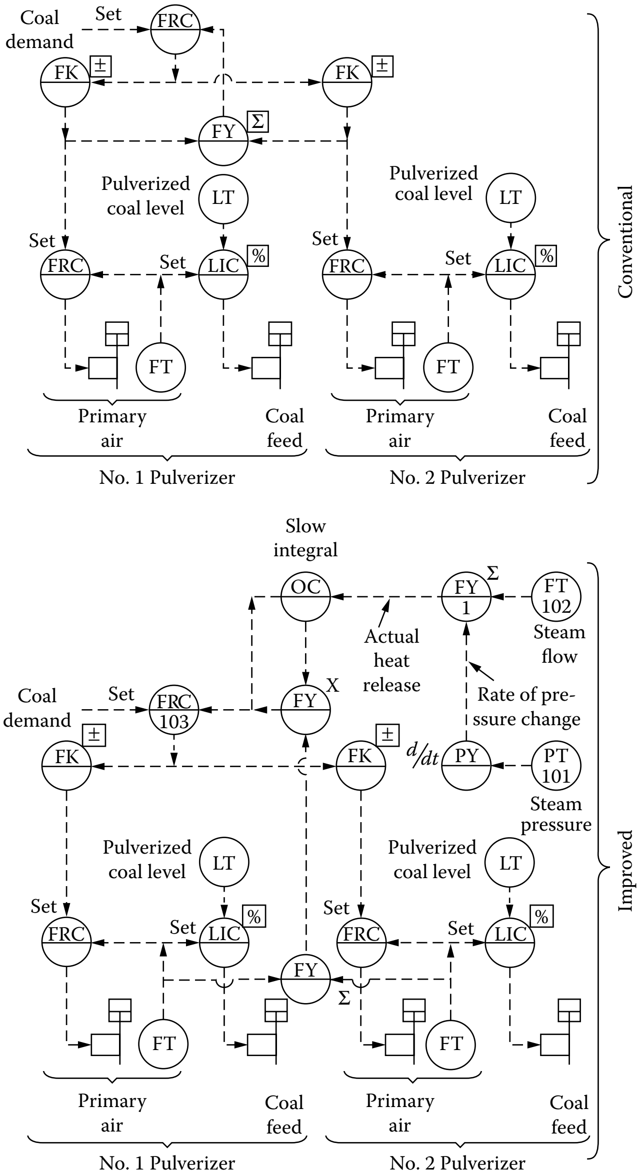


FIG. 8.6bb
Conventional (top) and advanced (bottom) control of ball-type pulverizers.

The control of a ball mill-type pulverizer is again different from a control standpoint. This is shown in Figure 8.6dd, including the application of manual compensation for the number of pulverizers in service.

The ball mill level measurement shown in Figure 8.6dd is normally based on the differential pressure value taken from the dip-tube. This measurement is inaccurate during start-up and during load swings due to large coal particle size. A more delicate control approach maintains mill motor

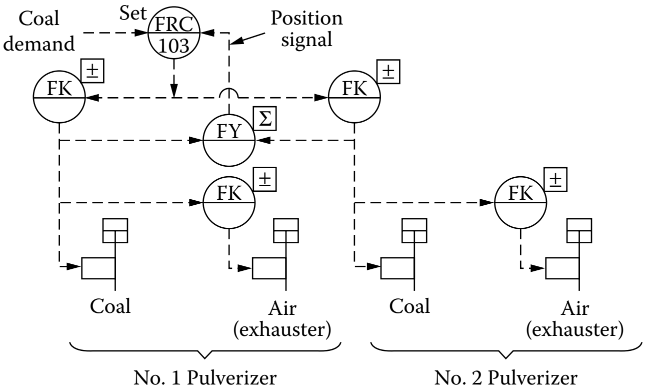


FIG. 8.6cc
Coal fuel control using bowl-type dual pulverizers.

power levels around operator-selected set points. A sonic signal taken from a microphone is often required to assist the control. The underlying mechanism is that the amount of coal charge in the ball mill is functionally related to the mill motor power level and the mill noise level. The weight of the coal charge can be controlled directly through modulation of the associated mill feeder(s).

A simplified control flow diagram is illustrated in Figure 8.6ee. The system utilizes the mill power level (kilowatts), the mill sound level (decibels), and the operator-selected set point (kilowatts) as the principal inputs to the mill process controller (MIC-150). Boiler fuel demand (or steam flow) can be used by the control system as a feedforward signal to anticipate major changes in mill load demand, and to preposition the mill feeder(s) at the proper steady-state speed.

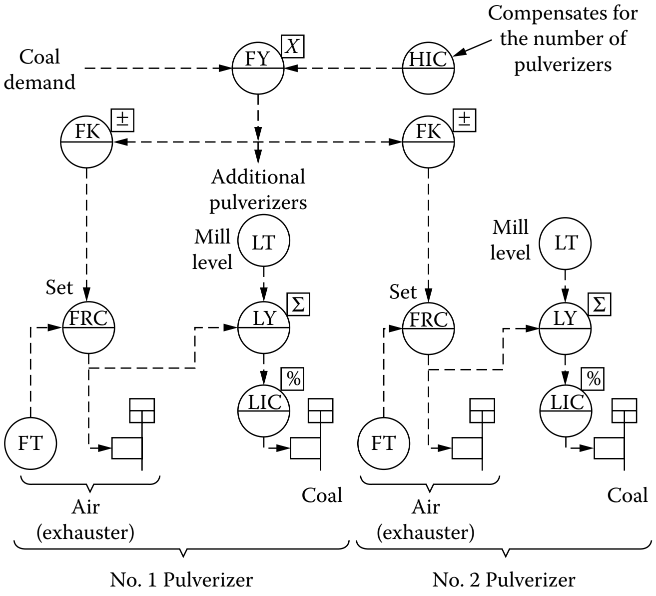


FIG. 8.6dd
Ball mill-type pulverizer controls.

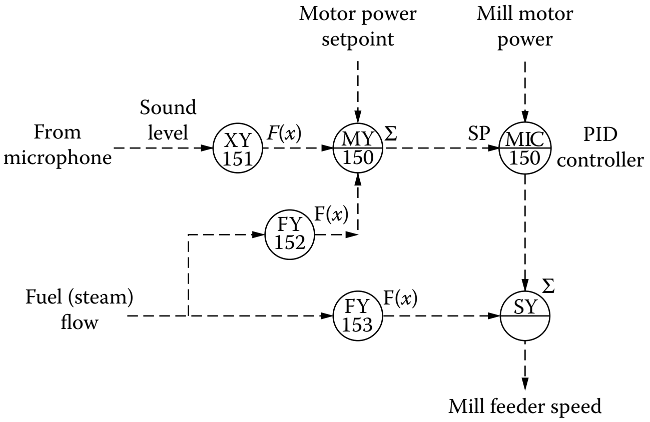


FIG. 8.6ee
Alternative, improved ball mill pulverizer control that utilizes both mill power level and sound level.

Metered Coal Controls When the flow of coal is controlled, it is done at the inlet to the mill, as illustrated in Figure 8.6ff. The capacity of the pulverizer contributes a delay of a few seconds to several minutes, depending on the design of the mill. Hammer mill delays are the shortest, whereas ball and roller mill delays can reach several minutes. If the delay is less than 1 min, a corresponding delay can be inserted into the coal flow transmitter output (FY-1), which will enable the loop to overcome this problem.

If the mill delay exceeds 1 min, the flow of the primary air that conveys the pulverized coal must be manipulated as a function of coal demand. Because the coal loading of the air is not uniform, Shinskey recommends the improvements noted in Figure 8.6bb. These include the determination of the

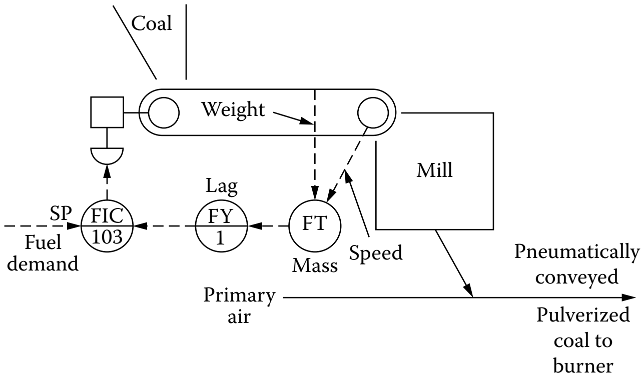


FIG. 8.6ff
Compensation for mill capacity.

actual heat release based on steam reading (FY-1 in Figure 8.6bb) and the slow integral correction of the total flow of primary air until the estimated and actual heat release are matched.

Highly Variable Fuels Solid fuels, such as coal, biomass, and municipal waste, tend to vary significantly in heating calorific value. This variation can be detected and compensated for by comparing the apparent *heat release* of the boiler process to the apparent heat input and adjusting the fuel flow accordingly to achieve the desired heat input. One approach to quantifying heat release is to sum the load indication (e.g., outlet steam flow or turbine first-stage pressure) with the rate of change of pressure in the drum.

Figure 8.6bb shows the use of steam flow (FT-102) for load. PT-101 and PY in Figure 8.6bb are the steam drum pressure and its rate of change. In effect, this is summing the rate of energy out with the rate of change of energy storage (FY-1, Figure 8.6bb). This heat-release measure is then compared to the current compensated fuel flow and the difference, considered to be the error between the two due to heating value variation in the fuel, is integrated slowly and multiplied by the fuel flow, producing an adjusted process variable for the fuel controller.

Air flow Measurement and Control

Combustion air for steam boilers may be supplied by induced draft (suction fan at boiler outlet or stack draft), forced draft (pressure fan at inlet), or a combination of forced and induced draft known as balanced-draft fans. With balanced-draft boilers, a slight negative pressure is maintained in the furnace.

In the control of combustion air (if there are both forced- and induced-draft fans), one fan should be selected for basic control of air flow and the other assigned to maintaining the draft pressure in the furnace. The following discussion is based on a single air flow source (fan) per boiler. Balanced draft and its effects on air flow control will be covered later in this section.

For successful control of the air/fuel ratio, combustion air flow measurement is important; refer back to Boiler Equipment, The Role of Sensors for detailed discussion on air flow measurement issues and techniques (Figure 8.6b).

Primary flow elements in the ducts, as well as the furnace pressure transmitter, will frequently produce noisy signals because of pulsation from the pumping action of the fans or from the combustion process. Provision should be made for dampening the flow signals to facilitate more optimum tuning of the controllers. Normal differential pressure ranges for these measurements are between 1 in. (25.4 mm) H_2O and 6 in. (152.4 mm) H_2O for conventional sensors and as low as 0.1 in. (2.54 mm) H_2O for area-averaging pitot stations.

Damper and Fan Controls Control devices for boiler air flow control on pneumatic installations are double-acting pistons, but in some cases electric motors are also used. In either case, a linear relationship is required between control signal and combustion air flow. Characteristics of most constant-speed fans or dampers approximate those given in Figure 7.1a in Section 7.1 in Chapter 7, and the desired relationship is linear.

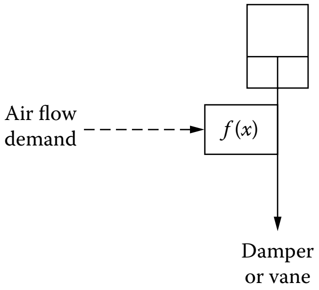


FIG. 8.6gg
Open-loop air control with single damper.

The relationship between open- and closed-loop control that was noted in connection with fuel control also applies to air flow control. Closed-loop air control is more precise and is self-linearizing, if the integrated system does not compare air flow with fuel flow (or as inferred from steam flow) directly in a ratio or difference controller.

Open-loop air control variations that may be used depending on the arrangement of fans are as shown in Figures 8.6gg through 8.6kk. Figure 8.6gg illustrates a single-damper-controlled open loop.

Figure 8.6hh shows a combination of damper and speed control. A system of this sort is often necessary to increase turndown (rangeability) where fan speed is variable. Good response of air flow based on fan speed adjustment alone is normally not attainable below approximately 1/3 maximum speed. Depending on fan design, this may correspond to 50% of boiler capacity. Use of a damper in combination with speed adjustment allows further turndown, because fan speed is blocked at approximately 1/3 speed. Split ranging, as shown in Figure 8.6hh, conserves steam or fan power.

As a result of inlet damper leakage that is normally present, it may be necessary for wide-range low-load or start-up control to parallel inlet and outlet dampers. To save fan

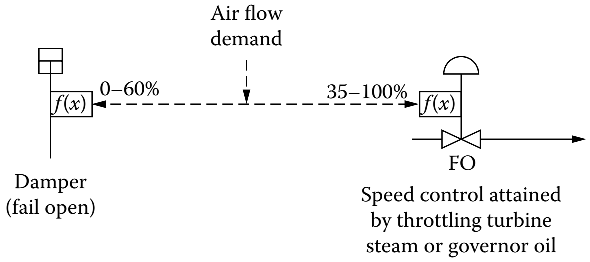


FIG. 8.6hh
Combination damper and speed control to increase rangeability.

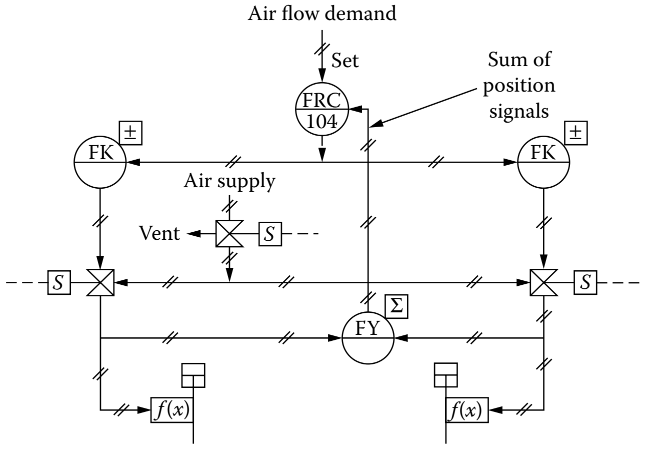


FIG. 8.6ii
Parallel fans with automatic air flow-balancing controls.

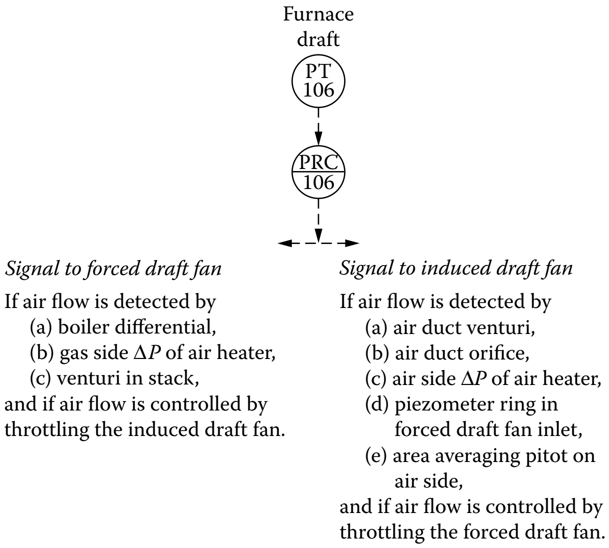


FIG. 8.6kk
A listing of the considerations for deciding whether to use the forced (left) or the induced (right) draft fan to control the furnace draft pressure.

power, the inlet damper may be operated over the full 3–15 PSIG (0.2–1.0 bar) range, whereas the discharge damper can be fully open at 3 PSIG (0.2 bar) and closed at 9 PSIG (0.6 bar).

As shown in Figure 8.6ii, when two fans are operated in parallel or on single-fan operation, the idle fan should have its damper closed to prevent recirculation from the operating fan.

When one fan of a two-fan system is switched on or off, considerable manual operation is required to prevent serious air flow upsets. The system shown in Figure 8.6ii eliminates this problem by automatically compensating the operating fan damper position as the parallel fan is started up or shut down. Separate discharge dampers may be used for shut-off purposes, supplementing the interlocks shown.

Closed-loop versions of the loops illustrated in Figures 8.6gg, 8.6hh, and 8.6ii would consist of flow controllers with air flow feedback superimposed on the components shown. For example, Figure 8.6jj shows the closed-loop version of Figure 8.6ii.

Furnace Draft Control

Whenever both forced draft and induced draft are used together, at some point in the system the pressure will be the same as that of the atmosphere. Balanced-draft boilers are not normally designed for positive furnace pressure. Therefore, the furnace pressure must be negative to prevent hot gas leakage. Excessive vacuum in the furnace, however, produces heat losses through air infiltration. The most desirable condition is thus one in which there is a very slight (approximately 0.1 in. H_2O, or 0.025 kPa) negative pressure at the top of the furnace.

Pressure taps for measuring furnace pressure may be located some distance below the top. Because of the chimney effect of the hot furnace gases, pressures measured below the top of the furnace will be lower by approximately 0.01 in. H_2O per foot (0.008 kPa/m) of elevation. Thus, if the pressure tap is 20 ft (6 m) below the top of the furnace, the desirable pressure to maintain is approximately 0.3 in. H_2O (0.075 kPa) vacuum.

In the case of a balanced-draft boiler, the maintenance of constant furnace pressure or draft keeps the forced and induced draft in balance. The purpose of this balance is to share properly the duty of providing combustion air and to protect furnaces not designed for positive-pressure operation.

The measurement of furnace draft produces a noisy signal, limiting the loop gain to relatively low values. In order to provide control without undue noise effects, it is desirable to use a full span of approximately 4–5 in. H_2O. This is normally a compound range, such as +1 to –4, or +2 to –3, in. H_2O (+0.25 to −1 kPa, or +0.5 to –0.75 kPa). Even with this span and with a set point of –0.1 to –0.3 in. H_2O (0.025–0.075 kPa), the controller gain can still not exceed 1.0. In some cases, it may be necessary to use integral control only to stabilize the loop.

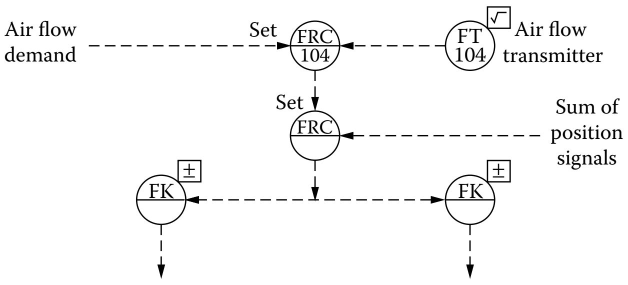


FIG. 8.6jj
Closed-loop control of parallel fans with balancing controls.

Air Flow and Furnace Pressure Interaction Additionally, stability problems and interactions may occur in the overall system because of measurement lags. It is recommended that the pressure transmitter connections to the boiler furnace be made with pipe at least 1 in. (25 mm) in diameter because of the very low pressures involved. If the distance is less than 25 ft (7.5 m), ¾ in. (18.75 mm) pipe may also be used.

Either the forced- or the induced-draft fan can be used to control the furnace draft, with the other fan performing the basic air flow control function. Interaction cannot be completely eliminated between these two loops, but it can be minimized by system designs such as those shown in Figures 8.6kk and 8.6ll.

The common rule is that air flow should be measured and controlled on the same side (air or combustion gas) of the furnace to minimize interaction between the flow and pressure loops.

If the combustion air is preheated, then its temperature will vary substantially and compensation is needed. The mass flow of air is related to $\sqrt{h / T}$, where h is a differential across a restriction and T is absolute temperature. This loop is shown in Figure 8.6ll, together with an excess oxygen trim on the air flow controller. Pressure and temperature compensation in engineering units can be carried out in the control system with Equation 8.6(2) (drop all but the temperature term if only temperature compensation is required).

The air flow and furnace pressures interact similarly to the phenomenon in Figure 8.6v. Because in the case of air flow controls the dampers are the same size and therefore their pressure drops are similar, decoupling needs to be applied. One might reduce interaction by connecting the two dampers (or fans) in parallel and using the furnace pressure as a trimming signal. (This can also be used to overcome problems resulting from noisy furnace pressure signals and slow response caused by the series relationship between flow and pressure loops.) This control system is illustrated in Figure 8.6ll.

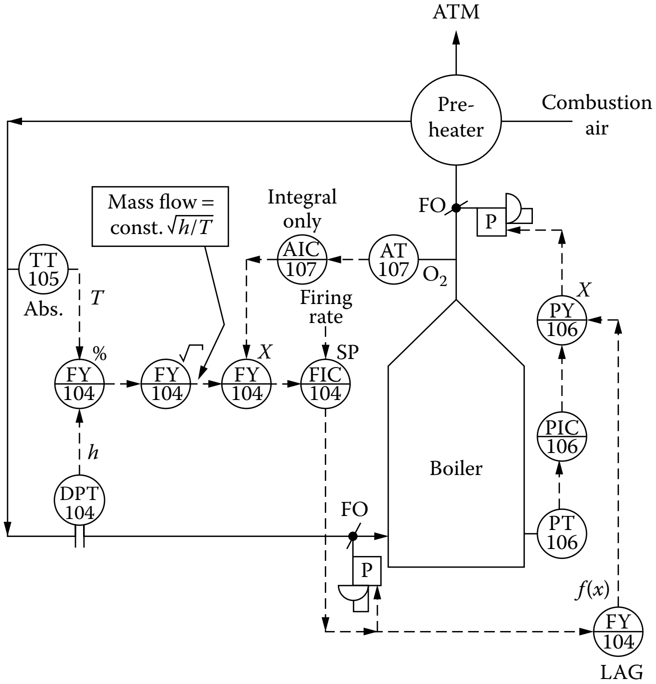


FIG. 8.6ll
Parallel control of inlet and outlet dampers reduces interactions.

In this arrangement the air flow controls move both dampers equally, and the furnace pressure corrects for any mismatch. The furnace pressure might respond faster to a change in the downstream damper opening, and therefore a dynamic lag (FY-104) is provided. In the combined control system shown in Figure 8.6l, basically the same control concept is implemented as shown in Figure 8.6ll. The main difference is that in Figure 8.6l, the furnace draft is throttled not by a discharge damper, but by an induced-draft fan.

The air flow is also detected by a high rangeability flowmeter (FT-104) that is compensated for temperature variations to approximate mass flow.[14] The air/fuel ratio is adjusted by applying a gain to the air flow (FY-104). The continuous optimization of this ratio is one of the major tools of boiler optimization.

The air flow controller throttles the speed or inlet vanes of the forced-draft fan and the signal to the induced-draft fan in a feedforward manner. Dynamic compensation is provided by the lag module (FY-104). Thus, as soon as the air inflow to the furnace is changed, the outflow will also start changing. This improves the control of the furnace draft, as the feedback pressure controller (PIC-106) will need only to trim the feedforward signal at PY-106 to account for measurement and other errors.

Air/Fuel Ratio

When the controls for air/fuel ratio are considered, one point is very important. Because of the combustion gas velocity through the boiler, for safety reasons the fuel/air ratio should be maintained on an instant-by-instant rather than a time-averaged basis.

As a general rule (except for the case of very slow-changing boiler loads), fuel and air should be controlled in parallel rather than in series. This is necessary because a lag of only 1 or 2 sec in measurement or transmission will seriously upset combustion conditions in a series system. This can result in alternating periods of excess and deficient combustion air. Consequently, the discussion here will be limited to parallel air and fuel control systems.

The simplest control of air/fuel ratio is with a system calibrated in parallel, with provision for the operator to make manual corrections (Figure 8.6mm). In this system, the operator uses the bias provision of the panel station (FK) to compensate for variations in fuel pressure, temperature, or heating value or for air temperature, humidity, or other factors. A system of this sort should be commissioned with detailed testing at various loads for characterizing and matching fuel

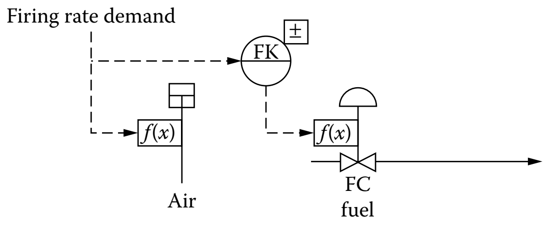


FIG. 8.6mm
Simple parallel air/fuel ratio system.

and air control devices. In addition, simple systems of this type should be adjusted for higher excess air, because they have no means of automatically compensating for the fuel and air variations.

The next higher degree of sophistication is a system with simple proportional compensation. This can be done by balancing fuel burner pressure to the differential produced between windbox and furnace pressures.

In the system shown in Figure 8.6nn, the burner fuel apertures are used to measure fuel flow, and the burner air throat is used as a primary element to detect air flow. There is no square-root extraction (such extraction would not show true flows because of the nature of the primary elements), and therefore the actual controller loop gain changes with load (capacity). Rangeability is limited unless there are multiple burners that can be put into or taken out of service.

The arrangement in this control system can also be reversed, with firing rate demand directly adjusting the fuel and the correction control being on air flow. This choice will be considered later in this section.

As boilers become larger, the need for precise control becomes greater, together with the potential for savings. The following series of diagrams represents further degrees of system sophistication.

The system in Figure 8.6oo is quite similar to that shown in Figure 8.6nn, except that here flows are measured as accurately as possible and are used in a flow controller to readjust the primary loop through the combining relay (FY).

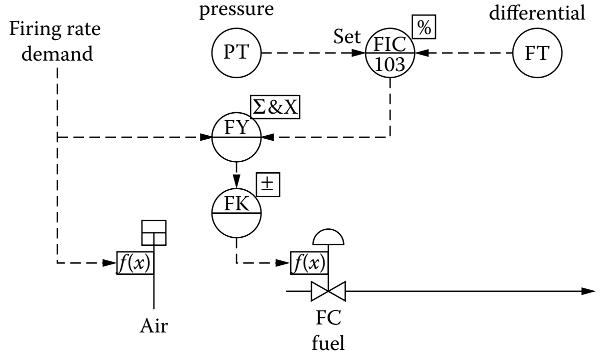


FIG. 8.6nn
Proportional compensation in air/fuel ratio control.

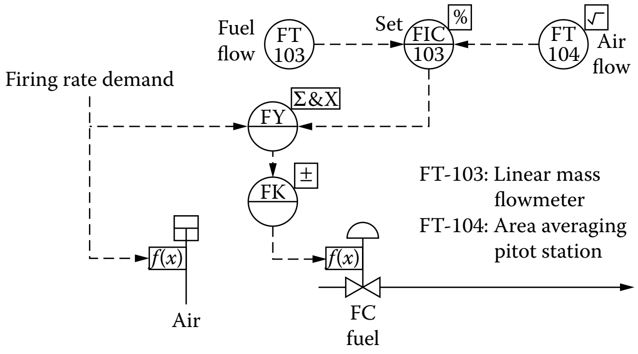


FIG. 8.6oo
Proportional compensation with accurately measured flows.

In this system fuel and air flow are open-loop controlled, and only secondary use is made of their measurements. The undesirable consequence is that fuel disturbances need master controller action for correction, and the fuel and air loops can interact with each other. The effect of interaction and disturbances in the fuel and air control loops can be minimized by the use of closed-loop fuel and air control. The system shown in Figure 8.6pp (which uses essentially the same equipment) is more desirable from the noninteracting, self-linearizing standpoint but must be provided with high turndown flow sensors and signal transmission that does not contribute a time lag.

In the systems shown in Figures 8.6oo and 8.6pp, fuel flow and air flow signals for proper fuel/air ratio are matched. This is done by matching air to fuel in the field combustion test calibration of the air flow measurement. Field testing is less stringent with the system shown in Figure 8.6pp, because the self-linearizing feature of the closed-loop system reduces the work to characterize the fuel and air control devices.

The systems shown here are for measurable fuels. In burning coal, wood, and refuse, fuel flow is often inferred from steam flow, and the steam-flow/air flow relationship is used as a control index.

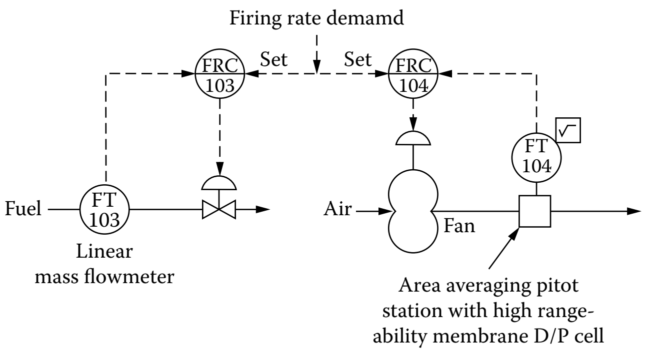


FIG. 8.6pp
Closed-loop air/fuel ratio control.

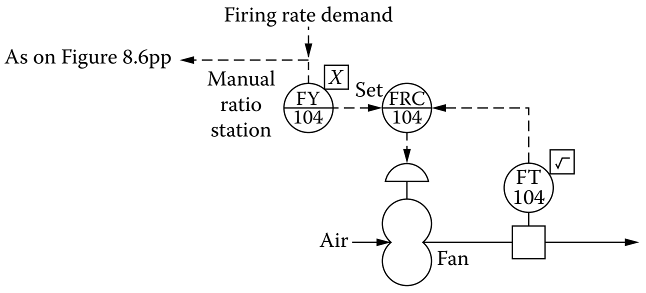


FIG. 8.6qq
Closed-loop air/fuel ratio control with manually adjusted ratio station.

In the event that the heat of combustion or flow rate of the fuel varies unpredictably, as frequently happens during feeding of coal, wood, or refuse, air flow should be set in ratio to the heat released by the fuel. For a boiler, heat release is indicated by steam flow combined with the rate of rise of steam pressure. This is the same calculation that was made in Figure 8.6bb to recalibrate the coal flow measurement. In the steady state, the heat release, coal flow, and total heat demand signal are identical, so any of the three may be used to set air flow.

The fastest response to a change in demand will be achieved if total heat demand sets air flow. The best response to an unmeasured change in coal flow or heating value would be obtained if air flow is set by calculated heat release. If the higher of the two signals is selected to set air flow, then excess air will be provided in either event.

A manual adjustment of the fuel/air ratio can be provided with either of the systems illustrated in Figures 8.6oo and 8.6pp. This adjustment can be inserted as shown in Figure 8.6qq to keep from changing the gain of the loop as adjustments are made. If ratio adjustment is manual, flue-gas analyzers, such as oxygen, carbon monoxide, or combustibles analyzers, should be provided for the operator's guidance or as feedback control of the ratio desired. A measurement of percentage of oxygen in the combustion gases is most necessary for operator guidance when multiple fuels are being burned or when fuel properties vary.

Fuel and Air Limiting Maintaining the correct fuel/air ratio also contributes to limiting the fuel rate to available air and limiting air minimum and maximum to fuel flow, air leading fuel on load increase, and air lagging fuel on load decrease. This arrangement also protects against a fan failure or a sticking fuel valve.

Though not normally furnished on smaller boilers, the following limiting actions are desirable for safety purposes:

1. Limiting fuel to available air flow
2. Minimum limiting of air flow to match minimum fuel flow or to other safe minimum limit
3. Limiting minimum fuel flow to maintain stable flame

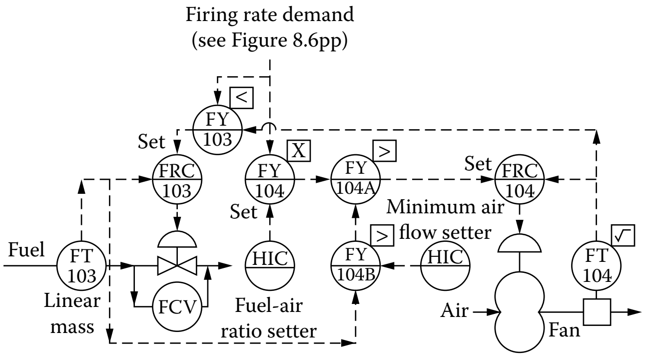


FIG. 8.6rr
Parallel, closed-loop control with air and fuel limiting.

These limiting features are simple to apply with the basic noninteracting, self-linearizing system in Figure 8.6pp. Figure 8.6rr shows the necessary modifications that can provide these features without upsetting the set point of the fuel/air ratio.

The following is accomplished by the illustrated system:

1. If actual air flow decreases below firing rate demand, then the actual air flow signal is selected to become the fuel demand by low selector (FY-103).
2. If fuel flow is at minimum and firing rate demand further decreases, actual fuel flow becomes the air flow demand, because FY-104A will select the fuel signal if it is greater than firing rate demand signal. A manual air flow minimum is also available to come into use through FY-104B, such that if fuel flow signal drops below the HIC setting, this manual setting will become the air flow set point.
3. Fuel is minimum limited by separate direct-acting pressure of flow regulator (FCV).

The combined control system in Figure 8.6l also shows air and fuel limiting on its air/fuel ratio system, but it is configured differently from the scheme shown in Figure 8.6rr. Figure 8.6l does not show a minimum air setter (HIC), the air/fuel ratio relay (FY-104) is located on the measurement signal to the air flow controller (FIC-104), and its ratio setting is not adjusted manually (HIC) but is optimized by "tie-in #1." Otherwise, the two loops are quite similar.

In the combined control system of Figure 8.6l, the fuel flow is detected by a high-rangeability, accurate mass flowmeter (FT-103). The set point for the fuel flow controller (FIC-103) is the smaller of either the biased air flow or the firing rate (FY-103). The set point of the air flow controller (FIC-104) is selected as the higher of either the biased fuel flow or the firing rate (FY-104).

These high and low selectors provide the so-called cross-limited parallel metering system.[15] The coefficients in the system are adjusted so that the pairs of signals provided to the high and low selectors are equal under steady-state conditions. When the firing rate demand increases, the high selector provides it as an air flow set point while the low selector transmits the air flow process variable as the fuel flow set point. Air flow, therefore, starts to increase immediately; fuel flow rises only after the air flow responds.

When the firing rate demand signal decreases, the low selector provides it as the fuel flow set point while the high selector transmits the fuel flow process variable as the air flow set point. Fuel flow starts decreasing immediately; air flow drops only after fuel flow responds. Likewise, if a disturbance causes air flow to drop, the low selector transmits the air flow signal to the fuel flow controller. Fuel flow then decreases regardless of the steam demand, preventing a fuel-rich condition.

The disadvantage of the system is that overall response is constrained by the slower of air flow or fuel flow response to changing demand signals. For example, if the firing rate demand rises slightly, the system first positions the damper to increase air flow; then, as air flow rises, the system opens the fuel valve.

The bias and gain modules FY-103 and 104 were added in Figure 8.6l to improve the system response to *small* load changes. These reduce the effective fuel flow signal presented to the high selector while raising the air flow variable provided to the low selector.

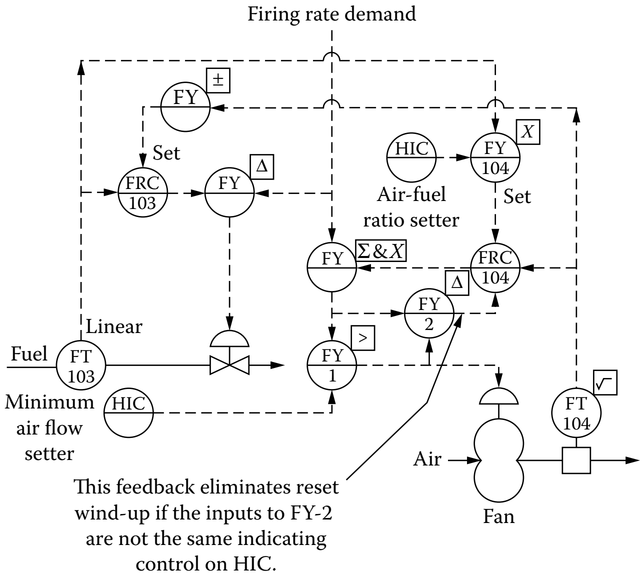
Under steady-state conditions, the firing rate demand is presented as the set point to both controllers. Likewise, if the firing rate demand changes only slightly, it will still be transmitted by both signal selectors and will cause the fuel and air flows to increase or decrease accordingly. If the changes in firing rate demand exceed the offset introduced by the bias and gain modules, the system will operate like a cross-limited system. As a result, fuel flow can respond to small increases in firing rate demand without raising air flow. Similarly, air flow may be adjusted slightly downward if firing rate demand falls, without having to decrease fuel flow.

In open-loop systems, the fuel and air limits are more difficult to apply. The application of these limits to an open-loop system is shown in Figure 8.6ss. Here, the fuel set point is determined (limited) by actual air flow, and a "fuel cutback" necessitated by reduced firing rate demand is accomplished at the expense of temporary fuel/air ratio offset.

Limiting combustion air flow to a minimum or to the rate at which fuel is being burned creates special problems because when the limit is in force, provision must be made to block the integral action in the air flow controller.

It may seem that a better way to limit fuel would be to have the firing rate demand directly set the air flow, with fuel being controlled through the combining relay (Figure 8.6tt). In this arrangement final fuel/air ratio correction occurs through the integral mode correction of the fuel/air ratio controller. This system is only partially effective, however, because on a sudden decrease of firing rate demand, the resulting reduction in fuel flow will occur only after the air flow has already been reduced.

A further consideration in setting fuel rather than air directly by the firing rate demand is that the parallel boilers can be more easily kept in balance, because balancing fuel directly balances the heat input without consideration of excess air between boilers.



This feedback eliminates reset wind-up if the inputs to FY-2 are not the same indicating control on HIC.

FIG. 8.6ss
Parallel, open-loop control with fuel and air limiting applied.

Feedwater and Drum-Level Control

Feedwater control is the regulation of water to the boiler drum. This water is admitted to the steam drum and, after absorbing the heat from the furnace, generates the steam produced by the boiler. On most boilers, makeup water for the feedwater system is filtered, deionized, treated, and deaerated prior to entering the boiler. It is usually preheated through one or more feedwater heaters and an economizer boiler tube section. The final control elements for feedwater are control valves, pump speed, or some combination. Pumps may be driven by electric motor or steam turbine.

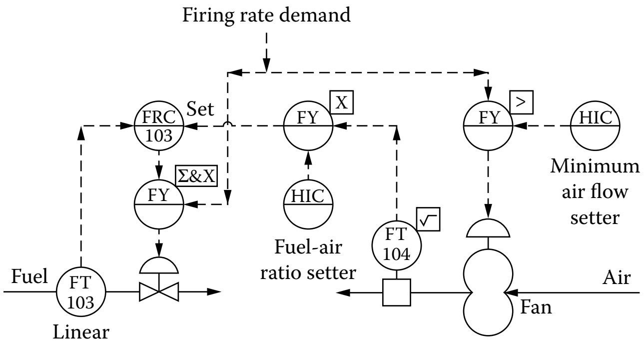


FIG. 8.6tt
Firing demand determines air flow, while fuel set point is adjusted by air flow.

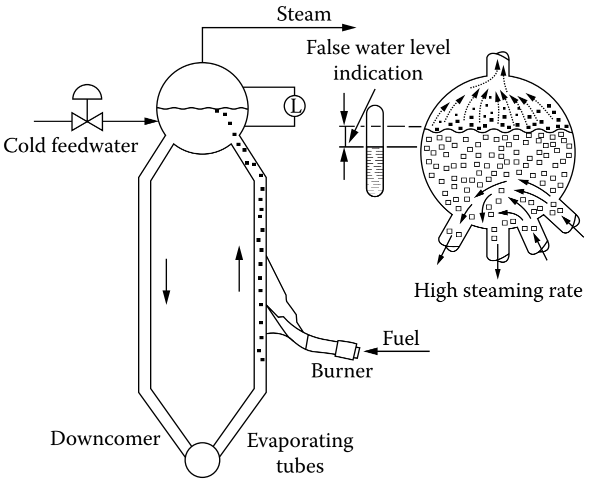


FIG. 8.6uu
Partial vaporization in the evaporating tubes causes drum level to shrink when feedwater flow increases and when pressure rises. On the other hand, an increase in the demand for steam causes the level to "swell."

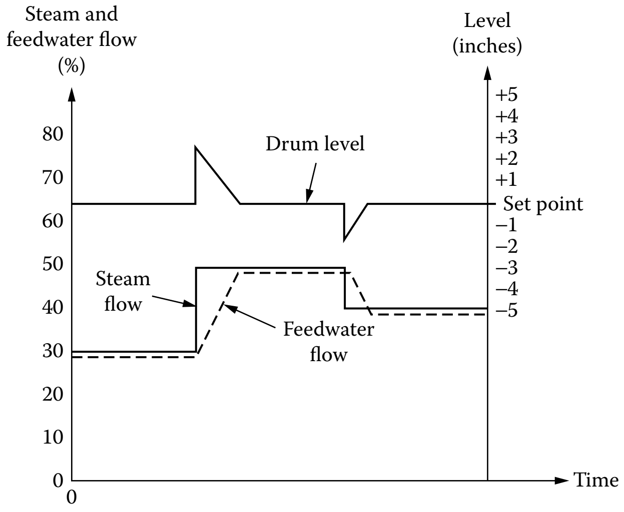


FIG. 8.6vv
Response relationship among steam flow (load), feedwater flow (manipulated variable), and level (controlled variable) in a properly designed system.

Proper boiler operation requires that the level of water in the steam drum be maintained within a certain band. A decrease in this level may uncover boiler tubes, allowing them to become overheated. An increase in this level may interfere with the operation of the internal devices in the drum that separate the moisture from the steam and may cause liquid carryover that can damage the steam turbine.

The water level in the steam drum is related to, but is not a direct indicator of, the quantity of water in the drum. At each boiler load, there is a different volume in the water that is occupied by steam bubbles. Thus, as load is increased there are more steam bubbles, and this causes the water to "swell," or rise, rather than fall, because of the added water usage (Figure 8.6uu). Therefore, if the drum volume is kept constant, the corresponding mass of water is minimum at high boiler loads and maximum at low boiler loads. The control of feedwater, therefore, needs to respond to load changes and to maintain water by constantly adjusting the mass of water stored in the system.

Feedwater is always colder than the saturated water in the drum. Some steam is then necessarily condensed when contacted by the feedwater. As a consequence, a sudden increase in feedwater flow tends to collapse some bubbles in the drum and temporarily reduce their formation in the evaporating tubes. Then, although the mass of liquid in the system has increased, the apparent liquid level in the drum falls. Equilibrium is restored within seconds, and the level will begin to rise.

Nonetheless, the *initial* reaction to a change in feedwater flow tends to be in the wrong direction. This behavior, called "inverse response" or "nonminimum phase response," causes an effective delay in control action, making control more difficult. Liquid level in a vessel lacking these thermal characteristics can typically be controlled with a controller gain of 10 (proportional band of 10%) or less. By contrast, the drum-level controller needs a controller gain closer to 1 (proportional band of 100%) to maintain stability. Integral action is then necessary, whereas it can usually be avoided when very narrow proportional band settings can be used.

Control of feedwater addition based on total drum level alone tends to be self-defeating, because on a load increase it tends to decrease water feed when it should be increasing. Figure 8.6vv shows the response relationship among steam flow, water flow, and drum level that should be present in a properly designed system if constant level under variable load is desired. For special reasons, one may wish to increase level with load. Boilers are designed for constant level operation, however.

Single- and Two-Element Feedwater Systems For small boilers having relatively high storage volumes and slow-changing loads, a simple proportional control may suffice, imprecise as it is. Integral action should not be used, because of resulting instability that is a result of integration of the swell on load changes that must later be removed. Control of this type, therefore, involves the addition of feedwater on straight proportional level control.

For larger boilers and particularly when there is a consistent relationship between valve position and flow, a two-element system (Figure 8.6ww) can do an adequate job under most operating conditions. Two-element control involves adding the steam flow as a feedforward signal to the feedwater valve (or boiler feed-pump speed). Two-element control is primarily used on intermediate-size boilers, in which volumes and capacities of the steam and water system would make the simple "total" level control inadequate because of

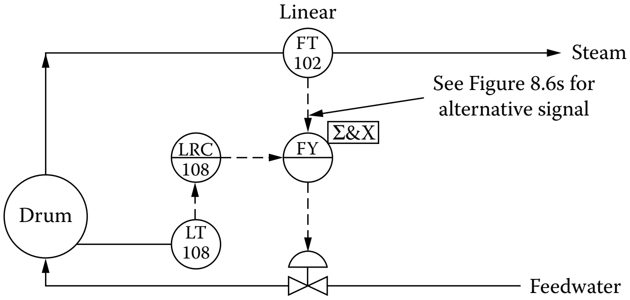


FIG. 8.6ww
Two-element feedwater control.

"swell." Total level control is undesirable when it is detected by sensors that are insensitive to density variations, such as the conductivity type. Displacement and d/p cell-type sensors are preferred from this perspective because they respond to hydrostatic head. Smaller boilers, in which load changes may be rapid, frequent, or of large magnitude, will also require the two-element system.

Field testing, characterization, and adjustment of the control valve are required so that the relationship of control signal to feedwater valve flow matches that of the steam flow to the flow transmitter output.

Any deviations in this matching will cause a permanent level offset at the particular capacity and less than optimal control (Figure 8.6xx). The level controller gain should be such that, on a load change, the level controller output step will match the change in the steam flow transmitter signal.

Three-Element Feedwater Systems As boilers become greater in capacity, economic considerations make it highly desirable to reduce drum sizes and increase velocities in the water and steam systems. Under these conditions, the boiler is less able to act as an integrator to absorb the results of incorrect or insufficient control. A three-element system is used on such large boilers to arrest disturbances and react to load changes more rapidly, as they occur.

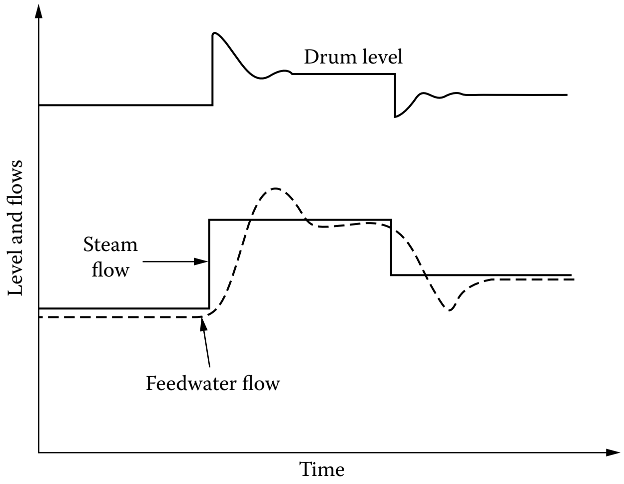


FIG. 8.6xx
The effect of mismatching between steam flow transmitter and valve flow characteristics.

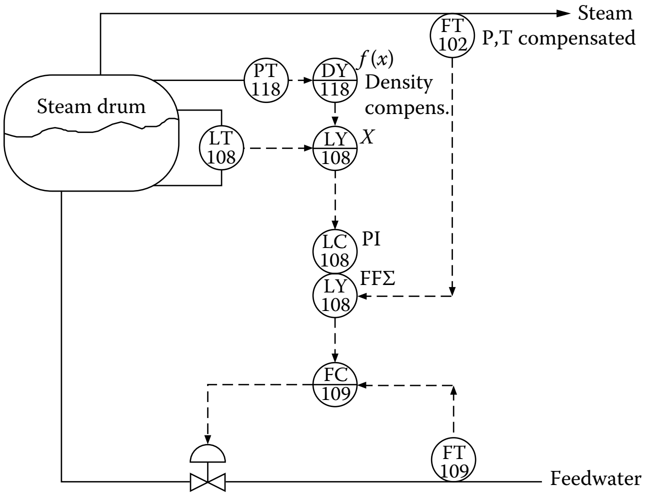


FIG. 8.6yy
Three-element feedwater system, plus density compensation with drum pressure.

Three-element control is similar to the two-element system, except that the water flow loop is closed rather than open. In this way, pressure disturbances that would affect feedwater flow are handled immediately by the fast response of the feedwater flow loop. There are several ways of connecting a three-element feedwater system, each of which can produce the results shown in Figure 8.6vv. Figure 8.6yy illustrates the most common way of connecting this system.

In addition to the three primary control variables (three elements)—drum level, steam flow, and feedwater flow—drum vapor-space pressure can be utilized to compensate for density changes.[16] The pressure is passed through a calculator, DY-118 in Figure 8.6xx, that calculates a multiplier to apply to the raw level signal. The multiplier is based on the density change vs. pressure for saturated steam, as taken from the steam tables.

In making gain adjustments on a three-element feedwater system, the first step is to determine the relative gains between level and flow loops. By observing a change in boiler load one can note the particular boiler "swell" characteristics. Maximum system stability results when the negative effect of swell equals the positive effect of flow. For example, if a 20% of maximum flow change produces a 2.4 psi (0.16 bar) change in flow transmitter output and this flow change also produces a 3 in. (75 mm) swell on a 30 in. (750 mm) range transmitter or a 1.2 psi (0.08 bar) transmitter output change, then the gain of the level loop should be double the gain on the flow loop.

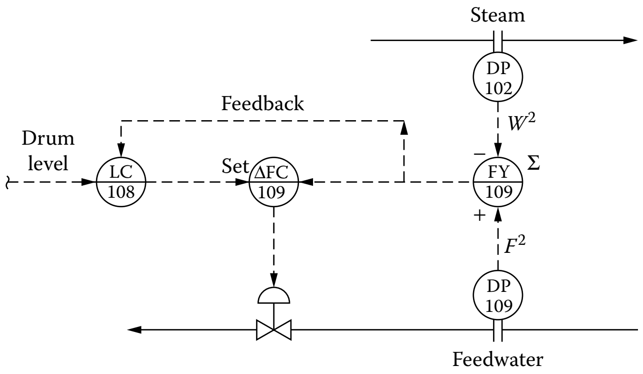


FIG. 8.6zz
Feedforward control of drum level.

Feedforward Control A feedforward variation is recommended by Shinskey to maintain a steam-water balance, reducing the influence of shrink-swell and inverse-response phenomena. The system shown in Figure 8.6zz causes feedwater flow to match steam flow in absence of action by the level controller. The two flowmeters have identical ranges, and their signals are subtracted. If the two flow rates are identical, the subtractor sends a 50% signal to the flow-difference controller. An increase in steam flow will call for an equal increase in feedwater flow to return the difference signal to 50%.

Errors in the flowmeters and the withdrawal of perhaps 2.5% water as “blowdown” (which is not converted to steam) will prevent the two flow signals from being identical. Any error in the steam–water balance will cause a falling or rising level. Therefore, the level controller must readjust the set point of the flow-difference controller to strike a steady-state balance.

The system assumes orifice-type flow sensors and does not use square-root extractors, because the period of oscillation and dynamic gain of a two-capacity level process varies directly with flow. The gain of the feedwater control loop without square-root extraction seems to compensate correctly for the process gain change.

Figure 8.6zz also shows external feedback from the flow-difference measurement applied to the level controller. This will precondition the level controller during start-up or at other times when feedwater is controlled manually or otherwise limited. Otherwise, an increase in steam or blowdown flow will increase the feedwater flow immediately, without depending on the level controller. This means that the feedback portion of the loop (LC-108) will need only to trim the ΔFC-109 set point to correct for flowmeter errors.

Because the role of the feedback portion is reduced from manipulating feedwater flow across its entire range to adjusting only for flowmeter errors, deviations in level from the set point will be minimized. Controller mode settings are not as critical in this situation, and incorrect actions caused by shrink, swell, and inverse-response are reduced.[6]

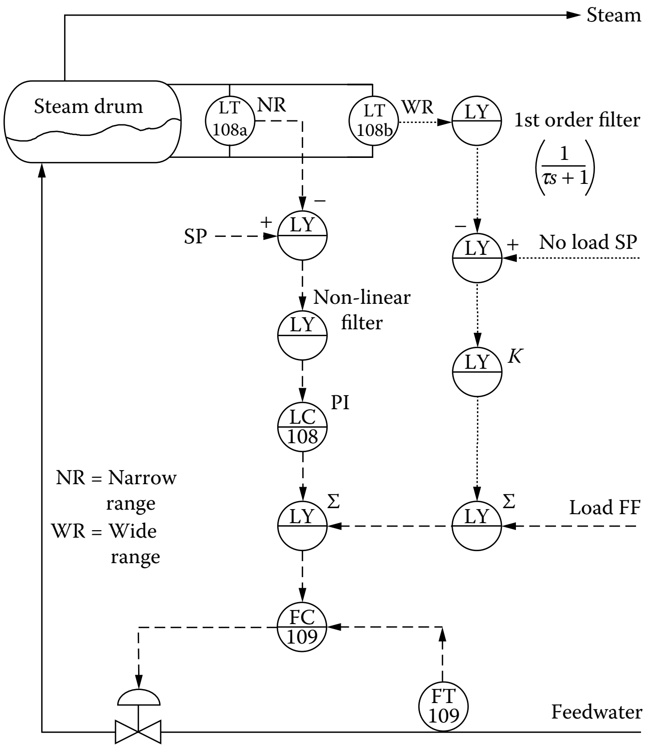


FIG. 8.6aaa
Wide-ranging proportional feedback drum level control, to help reduce performance degradation due to shrink–swell effect.

Shrink–Swell Compensation Another variation of level control proposed by Shinskey to overcome shrink–swell effects involves the use of proportional feedback from a wide-ranging level measurement.[17] On large boilers, primary drum level control is often accomplished with a narrow-range transmitter for more accurate control, with a wide-range transmitter present to show large excursions and handle alarms and trips outside the narrow control range.

The wide-range transmitter can be utilized to provide proportional feedback to sum with the output of the narrow-range (NR) level control (Figure 8.6aaa). When rising vapor volume in the water, resulting from an increasing load, causes the water to swell, the wide-range (WR) measurement, sensing much more of the total water inventory, will tend to properly measure lower, while the narrow-range instrument may indicate a rising level. Used with the proper filtering and tuning, the wide-ranging signal can be used to offset the narrow-range control loop’s tendency to initially respond in the wrong direction.

Feedwater Valves During start-up, when the boiler tubes and drum are being filled, the feedwater control valve must absorb the large pressure drop from the full feedwater pump discharge or feedwater header to the unpressurized drum (less dynamic head and elevation losses). Initial flows can be relatively small. At operating pressure, the pressure drop across the feedwater control valve is much smaller and the flow much larger. Due to this drastic variation in service conditions, it is common on large units to split feedwater control into a parallel-piped system with two control valves, one for start-up and one for normal operation.

On small to medium-sized units, it is possible to handle the full range of service conditions with one valve, but sizing, trim design, and material selection are critical to the success of control with a single valve. The start-up valve in a parallel two-valve system, or the regulating/start-up valve in a single-valve system, requires an anticavitation trim design (Figure 6.1y in Section 6.1).[18]

The duty required of a feedwater control valve is quite heavy, not only because of the large energy dissipation as the water passes through the valve, but also because the high-purity water normally used in boilers tends to be "metal-hungry." This, combined with high water velocity, produces corrosion, erosion, and cavitation effects that call for a chrome-moly or 400 series stainless steel valve body and trim (or other suitably resistant materials).

A general criteria has been proposed for an acceptable trim exit kinetic energy at all flows, for single-phase fluid applications.[19] The guideline set forth states that valve trim exit kinetic energy should be limited to 70 psi (480 kPa), equivalent to about 100 ft/s (30 m/s), and, for cavitation-prone applications, limited to 40 psi (275 kPa), equivalent to about 77 ft/s (23 m/s). At trim exit velocities of around 150 ft/s (50 m/s), boiler feedwater valves are reported to consistently fail within 3 years. Care in selecting the right valve design can result in much longer service life.

To keep the feedwater flow controller gain independent of load variations, the feedwater control valve should have a linear installed flow characteristic. An equal-percentage or modified equal-percentage inherent characteristic is appropriate for electric utility boilers with dedicated feedwater pumps for each boiler. On systems where multiple feedwater pumps supply a pressure-controlled header, which, in turn, supplies two or more boilers, a linear inherent characteristic may be appropriate. A characterizing positioner can be used to further ensure the linearity of the feedwater control valve. If there is noise in the loop, then dampening may be required, as well.

Valve Sizing For control valve sizing, a system "head" curve showing the relationship between system pressures and capacities should be developed. A typical head curve is shown in Figure 8.6bbb. The head curve demonstrates a basic problem in selecting the flow rate and pressure drop when the feedwater control valve is sized. Capacity C_2 and differential X are the most desirable from a control standpoint. Capacity C_3 and differential Y or Z are often used in an attempt to furnish sufficient water to the drum with safety relief valves blowing. It is not necessary to provide all this capacity in the primary feedwater regulating control valve; this will result in an oversized valve and will degrade control performance. It is not uncommon to see a valve that was designed for more than required capacity and for a 30 psi (207 kPa) differential, operating at a 500 psi (3.45 MPa) differential and at a fraction of its design capacity. Capacity for relief valve service can be provided with bypass valves.

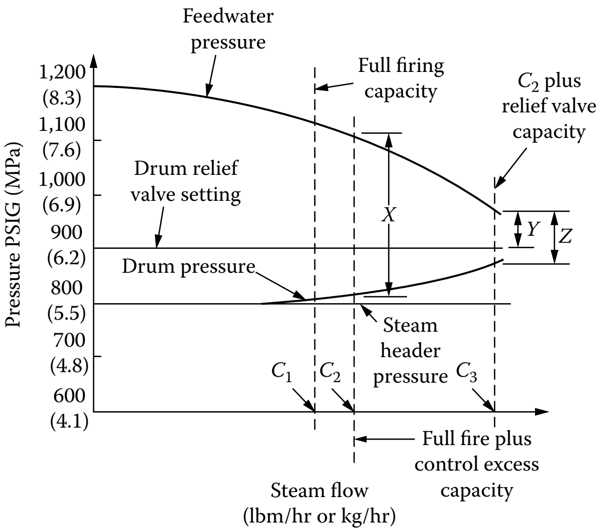


FIG. 8.6bbb
Head curve of the feedwater pump (top), the setting of the relief valve (horizontal) and the drum pressure curve (bottom).

In addition to the feedwater control valve(s) regulating water to the boiler, there is a feedwater recirculation valve controlling flow back to the deaerator (or sometimes the turbine condenser). This valve is a very severe service application, with the valve having to take the extreme pressure drop from the feed pump discharge back to the deaerator. In some plants, this may be from 5500 PSIG (38 MPa) down to 150 PSIG (1 MPa).[18] Cavitation abatement trim is required (Figure 6.1y in Section 6.1). The most efficient control strategy for this valve is to modulate it to maintain the net positive suction head (NPSH) available on the boiler feed pump greater than the NPSH required to prevent cavitation in the pump.

Pump Speed Control Control of pump speed to regulate feedwater flow can be accomplished if the pump is driven by a steam turbine or is furnished with variable-speed electric drive. This can be used in place of a control valve to save pump power on single-boiler systems. In systems in which several boilers are operating in parallel, the speed control can be used to save pump power by controlling the discharge pressure at a constant differential pressure above boiler pressure.

When pump speed is being used in place of a feedwater valve on a single boiler system, a large part of the speed control range is used in developing pump head at low flow. Characterization of the signal is, thus, necessary for good operation and constant gain throughout the operating range. On large units, there is still usually a start-up feedwater control valve for filling and low load, with pump speed taking over as load rises past some threshold.

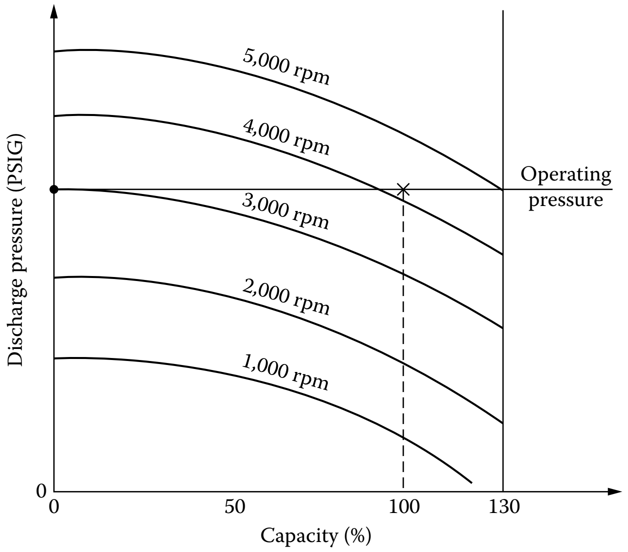


FIG. 8.6ccc
Feedwater pump on speed control. Speed range at 0 capacity to reach operating pressure is 0–3000 rpm. Speed range for 0–100% capacity is 3000–4200 rpm.

Figure 8.6ccc shows typical pump characteristics. Considerable pump speed is necessary just to build pressure in the system, before any significant load is being supplied. (Further discussion of variable-speed pumping is provided in Section 8.34 of this volume and in Section 2.14 of the *Process Measurement and Analysis* volume.)

Steam Temperature Control

The purpose of steam temperature control is to improve the thermal efficiency of steam turbines. Its most common application is for steam turbine electric power generation. The factors affecting steam temperature in a convection-type superheater are superheater area, flue-gas flow pattern across the superheater, flue-gas mass flow, temperature of flue gases leaving the furnace, and steam flow through the superheater. Additionally, furnace temperature affects radiant superheaters. Some superheaters may be designed for a flat curve, combining radiant and convection surface, but most superheaters are the convection type.

Control-station steam temperatures are limited to approximately 1050°F (566°C), whereas those in industrial units may be considerably less. If these temperatures can be controlled with extreme precision, they can be pushed closer to the allowable limits. Temperature can be controlled by adjustment of the amount of recirculation of the flue gas or by "attemperation," which is an energy-wasting method of superheat removal through feedwater spraying.

The required positions of burners or recirculating and bypass dampers as a function of load are well-established for any given boiler design. Therefore, it is common practice to program their positions directly from load. Readjustment to correct for inaccuracies in the program and changes in the characteristics of the boiler is accomplished by feedback control of temperature, using proportional and integral action (Figure 8.5ddd). Being applied through a multiplier, the feedback loop gain varies directly with load, and according to Shinskey this tends to cancel the inversely varying process gain.

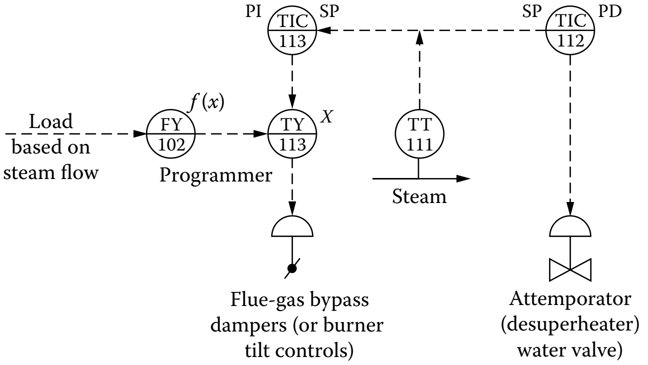


FIG. 8.6ddd
Steam temperature control.

Desuperheater Spray Controls Temperature control by attemperation is more responsive and can be used to supplement flue-gas manipulation. To minimize water usage, however, and to avoid conflict with flue-gas manipulation, proportional-plus-derivative control should be used for attemperation. The controller may be biased to deliver a nominal amount of feedwater at zero temperature deviation. The control system is described in Figure 8.6ddd.

To use a desuperheater spray for steam temperature control, the boiler would normally be provided with added superheater area. Figure 8.6eee demonstrates the effect of the water spray (which is usually between a primary and a secondary superheater section) for temperature control.

Provision must be made to prevent reset windup when in the uncontrolled load range. Control of this system is shown in Figure 8.5fff.

Large boilers may have burners that tilt up and down approximately 30° for steam temperature control. This effectively changes the furnace heat-transfer area, resulting in temperature changes of flue gases leaving the boiler. Spray is frequently used with these systems as an override control. These systems are used on large power plant boilers and normally require the type of controls illustrated in Figure 8.6ddd.

Work efficiencies are maximized by operating at the highest steam temperature at which the metals are capable of operating. In central stations, this limit is 1050°F (566°C); in industrial applications it is lower.[6] If improved control can elevate the TIC set point from 1000°F (538°C) to 1040°F (560°C), this will increase the available work by 17 BTU/lbm (9.4 cal/kg) in the steam.[6]

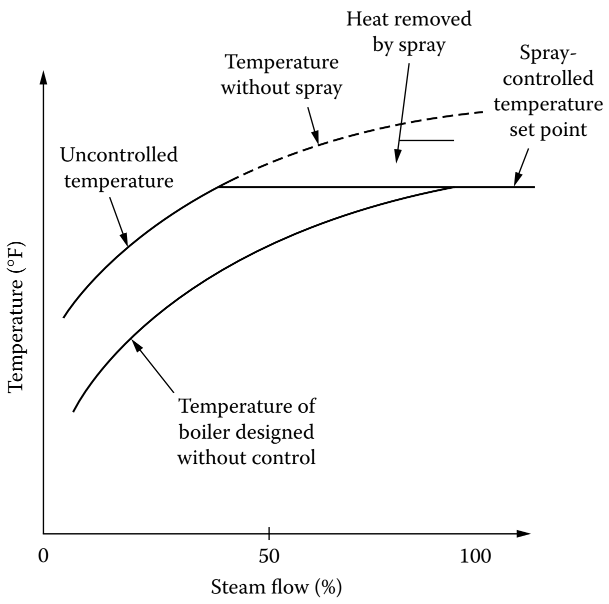


FIG. 8.6eee
Desuperheater characteristics.

Flame temperature does not vary much with load, but the hottest gases do tend to propagate further at higher loads. To increase the steam temperatures at low loads, recirculation blowers or tilting burners are used, which will direct the heat at the superheater sections. At high loads, the rise in the steam temperature is prevented by opening up some of the flue-gas bypass dampers or by desuperheating the steam through attemporation with water. For each boiler, the relationship between load and the required damper, recirculation fan, or burner tilt positions is well-established and, therefore, can be preprogrammed.

Figure 8.6l shows the input signal to the programmer as coming from FT-102, the steam flow transmitter. (Vortex shedding meters are limited to about 750°F, or 400°C.) In other designs, the input to the programmer is taken from the combustion air flow signal (FY-104). In either case, the preprogrammed relationship does need to be adjusted to overcome inaccuracies and changes in boiler characteristics.

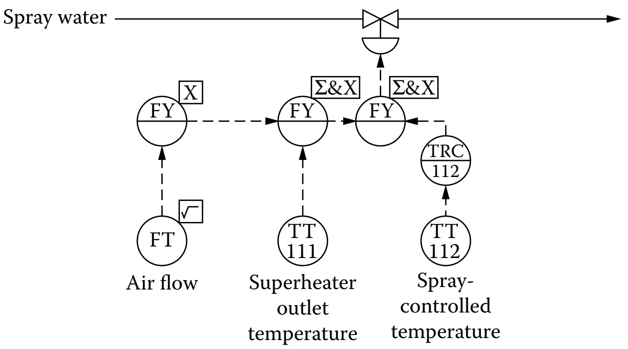


FIG. 8.6fff
Desuperheater spray control system.

TIC-111 throttles the set points of both slave controllers in cascade. Reset windup in TIC-111 is prevented by the external feedback (FB). Without it, windup would occur at low loads, when high temperatures cannot be maintained.[20] The task of temperature control is split between two slaves. The slower of these slaves—the PI controller—modulates burner or damper positions for long-term control. The desuperheater controller is faster; actually, it is faster than its cascade master. In order to use that speed, to minimize water usage (an irreversible waste of available work), and to avoid conflict with the PI controller, only P&D control modes are used.[6]

The desuperheater control loop can also be configured as a cascade loop, as illustrated in Figure 8.6ggg. Here, the

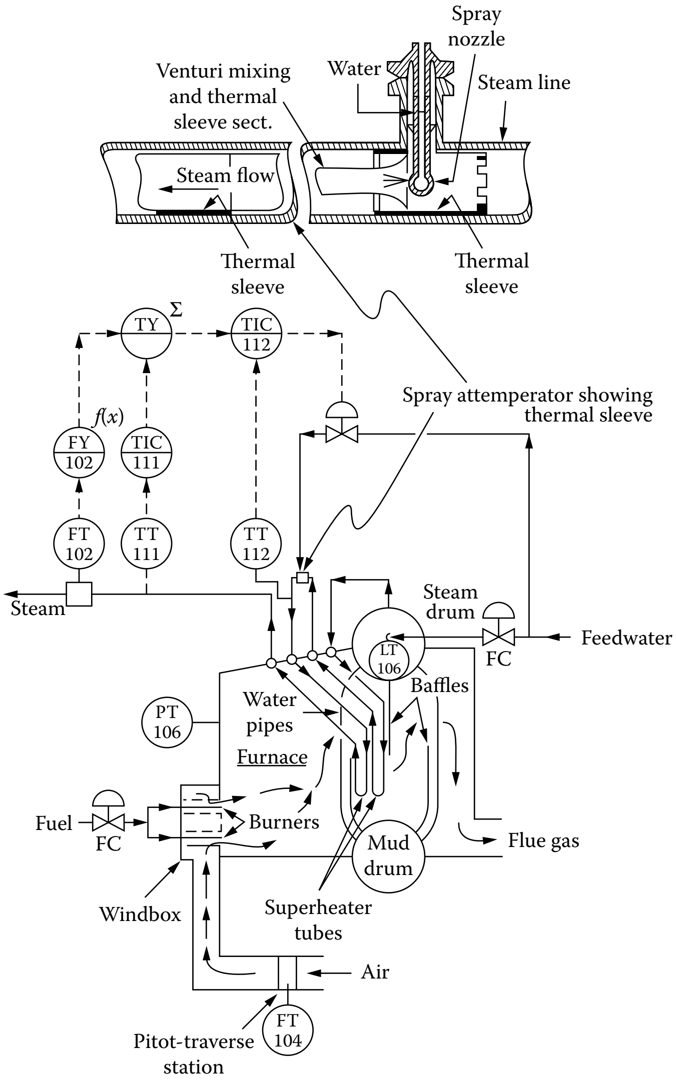


FIG. 8.6ggg
Cascade configuration for controlling the desuperheater control valve.

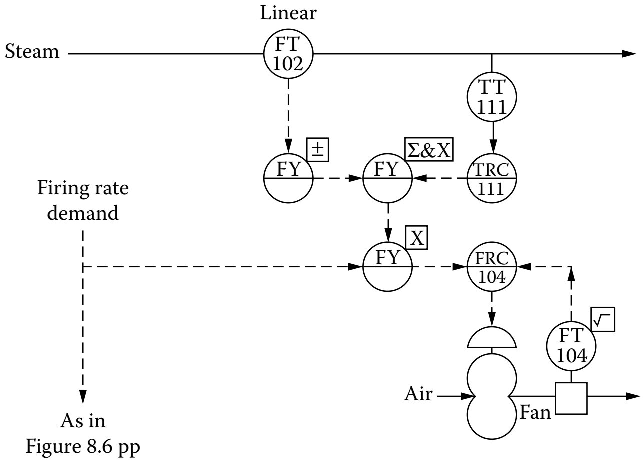


FIG. 8.6hhh
Steam temperature control by adjustment of excess air.

saturated steam from the steam drum is returned back into the furnace, where it is superheated. If the amount of superheat is excessive, then water is sprayed into the steam, and it is returned once more to the furnace to make sure that all water is vaporized.

The slave temperature controller (TIC-112) is placed right after the spray attemperator, while the cascade master (TIC-111) is located at the steam outlet. In Figure 8.6ggg, the feedforward correction is based on steam flow (FT-102), and the relationship between load and uncontrolled temperature (Figure 8.6eee) is predicted by the FY-102 function relay.

Cooling by Variable Excess Air Because mass flow of flue gas affects steam temperature, variation in excess combustion air can be used to regulate steam temperature. This method may be used on a boiler that was not specifically designed for temperature control. Although increasing excess air flow increases boiler stack heat losses, turbine thermal efficiency will also increase, and maintaining steam temperature can provide the greater economic benefit. The control arrangement in Figure 8.6hhh implements this method.

Flue-Gas Temperature

Flue-gas temperature is important for two reasons: first, as an indicator of boiler efficiency, and second, because if it drops below the dew point, the condensate formed will dissolve the oxides of sulfur in the flue gas, and the resulting acid will cause corrosion.

Figure 8.6l shows the temperature control loop TIC-110. The purpose of TIC-110 is to keep the flue gas dry and above its dew point. The flue-gas temperature at the cold end of the stack is usually arrived at as the average of several sensor outputs, measuring the temperature at various points in the same plane. When this averaged temperature drops near the dew point, TIC-110 will start increasing the set point of the air preheaters (TIC-114). Through the use of steam or glycol coils in the combustion air, the inlet air temperature to the boiler is increased, which in turn raises the flue-gas temperature.

Integration of Loops

Loops and subsystems can be combined to create an integrated control system (Figure 8.6l). Other combinations of the subsystems can similarly be put together to form a coordinated system. When designing, it is advisable to break the overall system down into these subsystems and examine them individually. Only then should the subsystems be put together into the total system.

Major design and operation problems in complex systems are created by inadvertently creating additional interaction, positive feedback, or tracking and initialization problems. In a complex system, as a result of adding, subtracting, multiplying, dividing, and comparing control signals or transmitter signals, it is difficult to get the system on “automatic” control easily and quickly.

The chief points to remember include

1. The systems often interact, e.g., air flow affects steam temperature, feedwater flow affects steam pressure, and fuel flow affects drum level and furnace draft.
2. For flexibility and rangeability, linear flow signals are necessary. Control valves and piston operators need linearizing positioners.
3. Fuels should be totalized on an air-required basis.
4. Tie-back arrangements, which simplify the task of getting quickly on automatic control, are very important in complex systems.
5. The flows of fuel and air should be controlled such that the flow rates reaching the burner always represent a safe combination.

POLLUTION CONTROL

Today, most electric utility-related environmental regulation is directed at reducing emissions of sulfur dioxide (SO_2) and nitrogen oxides (NO_x) as well as carbon dioxide (CO_2), mercury, volatile organic compound (VOC), particulate, stack opacity, wastewater, and other greenhouse gas pollutants. NO_x, CO_2, particulate, and opacity are considerably influenced by the boiler design, operation, and control. However, reduction or removal of some of these pollutants from the flue gas often must be accomplished between the boiler and the stack.

Electrostatic precipitators are commonly used for removal of particulate. Scrubbers are commonly used for removal of particulate and SO_2. From a process control viewpoint, two commonly regulated pollutants, NO_x and SO_2, are of particular interest. Their formation and control will be discussed in more detail.

NO_x Control[21,22]

Reduction of NO_x emissions is a major goal of the Clean Air Act amendments because of its known role in the formation of ground-level ozone and acid rain. A number of NO_x control technologies have been successfully applied to utility and industrial boilers. (For a discussion of the various nitrogen oxide sensors and analyzers, refer to Section 8.37 in Chapter 8 of the first volume of this handbook.)

NO_x Formation NO_x is formed by two primary mechanisms, resulting in thermal NO_x and fuel-bound NO_x. A third mechanism, “prompt NO_x,” also accounts for a minor share of NO_x formation. Thermal NO_x formation occurs only at high flame temperatures when dissociated nitrogen from combustion air combines with oxygen atoms to produce oxides of nitrogen, mainly NO and NO_2. Formation of thermal NO_x increases with combustion temperature and the presence of oxygen. However, the reactions are reversible.

Thermal NO_x usually comprises 25% of emission for a coal-fired plant, but the majority of emissions for gas-fired combustion. Fuel-bound NO_x formation is not limited to high temperature, but is dependent upon the nitrogen content of the fuel. In addition to the high flame temperature, quantity of excess air, nitrogen content in the fuel, the characteristics of the combustion process and the residence time at high temperature also play important roles in NO_x formation.

NO_x Reduction Strategies The best way to minimize NO_x formation is to reduce flame temperature, reduce excess oxygen, or burn low-nitrogen-containing fuels. NO_x reduction strategies and technologies for combustion sources can be classified by three major categories: precombustion processing (fuel switching), combustion modifications, and post-combustion processing. Combustion modifications include, but are not limited to, derating, burner system modification, low NO_x burners, and diluent injection. Major post-combustion processing techniques include selective catalytic reduction (SCR) and selective noncatalytic reduction (SNCR).

Low NO_x Burners These burners effectively reduce the formation of fuel and thermal NO_x. These burners generate air-fuel mixing patterns that lower peak flame temperature and oxygen concentrations. Reducing local oxygen concentration can be achieved by introducing flue-gas recirculation zones, air or fuel staging combustion zones, or flameless oxidation burners. Installation of these types of burners can reduce NO_x emissions by up to 50%.

Injection of water, steam, or flue gas (*diluent injection*) also helps to lower the NO_x level. By injecting a small amount of water or steam into the immediate vicinity of the flame, the flame will be cooled and the local oxygen concentration reduced. This would result in decreased formation of thermal and fuel-bound NO_x. However, this process generally lowers the combustion efficiency of the unit.

Flue gas can also be injected with the influent gas ahead of the burner to reduce the prompt NO_x formation. Advanced overfire air, in which air is injected in the combustor, can usually reduce NO_x emissions by 10–25%.

Excess air optimization also benefits NO_x reduction (see the forthcoming paragraphs on boiler optimization).

Selective Catalytic Reduction Claimed to be the most effective technology currently available for NO_x removal, selective catalytic reduction is a post-formation NO_x control technology that uses a catalyst to facilitate a chemical reaction between NO_x and ammonia to produce nitrogen and water. This process is implemented by injecting ammonia/air or ammonia/steam mixture into the exhaust gas, which then passes through a catalyst. The major NO_x reduction reaction is often presented as follows:

$$4NO + 4NH_3 + O_2 \rightarrow 4N_2 + 6H_2O \qquad \textbf{8.6(3)}$$

To optimize the reaction, the temperature of the exhaust gas must be in a certain range when it passes through the catalyst bed. Depending on the catalyst type (usually vanadium/titanium) and the location (usually economizer outlet), the typical flue-gas temperature should be in the range of 600–750°F (315–400°C). Removal efficiencies above 80% can be usually achieved, regardless of the combustion process or fuel type used.

Among its disadvantages, SCR requires additional space for the catalyst and reactor vessel, as well as an ammonia storage, distribution, and injection system. In the United States, the SCR control system is designed for operation during the ozone season (May–October), and for bypass operation in the non-ozone season.

The SCR typically consists of an NH_3 injection flow control system, an ammonia injection grid, and inlet and outlet NO_x monitoring equipment. Precise control of ammonia injection is critical. An insufficient amount of ammonia can result in unacceptable high NO_x emission rates, while excess ammonia can lead to ammonia “slip,” or the venting of undesirable ammonia to the atmosphere. NO_x reduction efficiency is directly proportional to the NH_3:NO_x ratio up to about 90%.

Specifically, the stoichiometry of the reaction is such that 1 mole of NH_3 reacts with 1 mole of NO_x, producing nitrogen (N_2) and water (H_2O). Adjustments to the molar ratio must be made to account for ammonia slip, i.e., the portion of the injected NH_3 that passes through the SCR unreacted.

In practice, as illustrated in Figure 8.6iii, the NH_3 flow control system anticipates the ammonia demand for NO_x emissions based on the boiler load (or the reactor inlet NO_x). This ammonia demand (FY-203) can be used to derive the ammonia feed rate and anticipates demand changes due to load swings. This ammonia demand signal is then trimmed using a feedback controller (XIC-202) that compares the

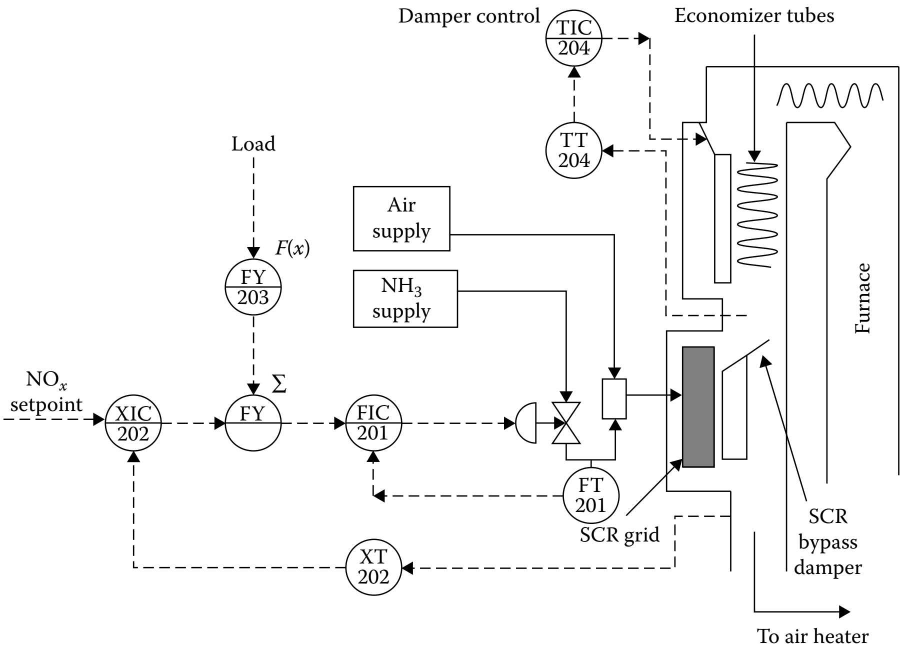


FIG. 8.6iii
NO_x control by using the selective catalytic reduction (SCR) process.

measured SCR outlet NO_x to the required outlet NO_x. Finally, the resulting ammonia demand signal is compared to the measured ammonia flow rate. The difference is conditioned by another slave controller (FIC-201), and the resulting control signal is used to modulate the ammonia flow control valve.

In order to maintain the flue-gas temperature above the minimum required, a set of dampers and a divider are often installed in the economizer, concurrently with the SCR project. The dampers will block only a portion of the economizer, restricting the flue-gas flow that passes through the economizer tubes. As the dampers close, they will force more flow through the open section, effectively reducing the area of economizer tubes that are exposed to the flue-gas flow. This control can be automatic (TIC-204), which compares the economizer outlet temperature (TT-204) to the set point and adjusts the position of the dampers accordingly.

Interlock functions need to be set up to prevent the ammonia system from starting or from continuing operation in abnormal situations. Important failures include, but are not limited to, NH_3 vaporizer outlet temperature too low, flue-gas flow too low, SCR flue-gas inlet temperature too low, dilution air flow too low, or vaporizer ambient NH_3 level too high.

Selective Noncatalytic Reduction Utility applications of SNCR processes involve the injection of a nitrogen-based reagent into the upper furnace or convective sections, where the injected chemical reacts with NO_x to form molecular nitrogen and water vapor. The most common types of reagent used commercially on large utility boilers are urea and ammonia. The optimum injection temperature when using ammonia is 1850°F (1010°C), at which 60% NO_x removal can be approached. The optimum temperature range is wider when using urea. Below the optimum temperature range, ammonia is formed, and above, NO_x emission actually increases.

The success of NO_x removal depends not only on the injection temperature, but also on the ability of the agent to mix sufficiently with flue gas. Compared to SCR, this technology is relatively capital inexpensive and used for smaller boilers. Typical NO_x reduction efficiency is 30–60%. In principle, the general control philosophy of an SNCR system is similar to the SCR control.

SO_2 Control

The formation of sulfur dioxide (SO_2) is a major concern with coal-fired boilers, due to the relatively high sulfur content in the fuel. For these units, post-combustion removal of SO_2 is realized through the flue-gas desulfurization (FGD) system. FGD systems can be classified into two basic categories: wet scrubbers and dry scrubbers.

Wet SO_2 scrubbers are the most widely used FGD technology for SO_2 control. Calcium-, sodium-, and ammonia-based sorbents are normally used in a slurry mixture, which is injected into a specially designed vessel to react with SO_2 in the flue gas. The most popular sorbent in operating wet scrubber is limestone, followed by lime. They are favored because of their availability and relative low cost. The overall

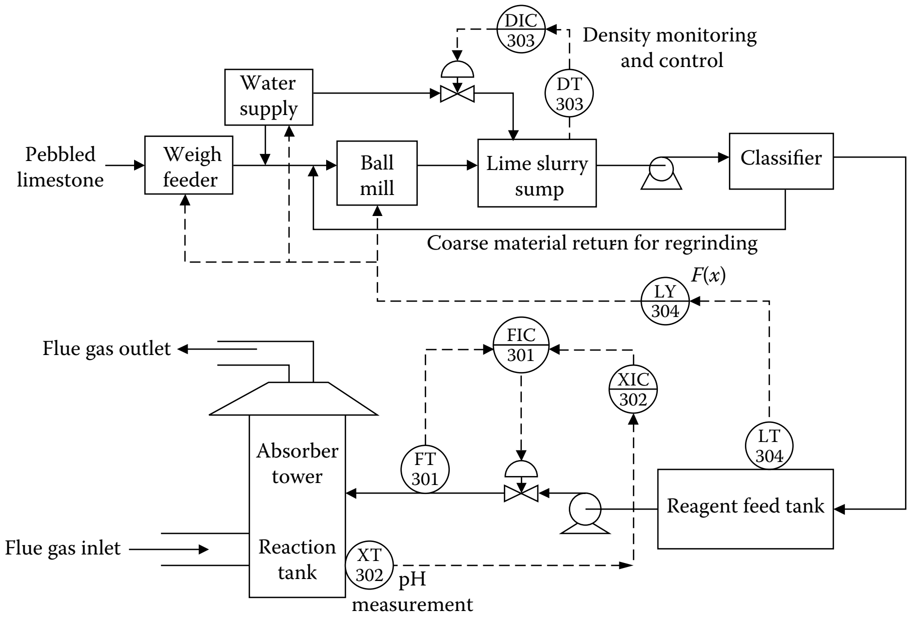


FIG. 8.6jjj
The control of a flue-gas desulferization (FGD) system.

chemical reaction that occurs with a limestone or lime sorbent can be expressed in a simple form as:

$$SO_2 + CaCO_3 = CaSO_3 + CO_2 \qquad \textbf{8.6(4)}$$

A typical FGD system should consist of several major subsystems, i.e., the limestone grinding and supply system, reagent feed system, absorber/reaction tank system, forced oxidation system, process water distribution system, dewatering system, filtrate system, and possibly, flue-gas reheat system. A simplified FGD system is illustrated in Figure 8.6jjj.

The limestone grinding loop is designed to supply limestone slurry containing 40% solids to the reagent feed loop. Usually, pebbled limestone from the limestone storage silo is conveyed by a variable-speed conveyor to the ball mill. Limestone is mixed with water and ground in the ball mill. The resultant slurry flows from the ball mill into the mill slurry sump, where it is pumped to the riser in the classifier. The slurry then flows from the riser into the cyclones that separate the coarse and fine slurry. The overflow (at about 40% solids) goes to the reagent feed tank, and the underflow goes back to the ball mill for regrinding. The reagent feed tank, which stores the slurry, allows for extended periods of operation with the ball mill out of service.

The SO_2/reagent reaction occurs in the absorber/reaction tank. The major equipment associated with the absorber loop typically are absorber feed tank (reaction tank), absorber spray pumps, wetted film contactor and its associated pump, and demister. The goal of the absorber loop is to absorb the SO_2 in the flue gas in the spray tower and wetted film contactor sections and precipitate calcium sulfite in the reaction tank. This is accomplished by controlling the reaction tank at proper pH level and density.

Reaction tank pH is the most important process variable in terms of total tower SO_2 removal efficiency and scale control. Operating at a lower pH would decrease the efficiency and increase the tendency of scale formation. Higher pH operation would cause high limestone consumption overall and decrease by-product quality. The pH value can be controlled by adjusting the reagent feed valve. The master control (XIC-302) compares the difference between the desired pH set point and the measured pH in the reaction tank. The output becomes the set point to the inner-loop slave controller (FIC-301), which compares this set point to the actual measured lime slurry flow rate and sends a command signal to adjust the valve opening.

Operating at lower density would remove valuable seed crystals from circulation, making control of the crystal types and the formation location difficult. Operation at a much higher density would increase pump horsepower requirements. Operation within a range of 5–15% solids is tolerable. A density element (DT-303) can be located in the mill slurry sump to measure the density of the mill slurry. Dilution water can be fed into the mill slurry sump to control the density via DIC-303.

Because the lime slurry flow from the reagent feed tank to the reaction tank results in a rising and falling level in the reagent feed tank, the reagent feed tank level control is also important. As shown in Figure 8.8jjj, the level in the reagent feed tank can be controlled by the signals (LT-304 and LY-304) transmitted to the ball mill, limestone conveyors, and ball mill water supply to adjust the limestone feeding and grinding. The reaction tank level is often monitored. However, it can also be controlled by adjusting the reagent slurry flow and the tank makeup water flow, subject to tank pH and density constraints.

Other important control loops (not shown in the figure) may involve the oxidation air blower control and the flue-gas reheat control. The oxidation air system normally consists of several air compressors and is designed to oxidize the solids in the reaction tank. The purpose of oxidizing the solids is to convert calcium sulfite ($CaSO_3$) to calcium sulfate ($CaSO_4$), which dewaters much more easily to achieve a drier waste product. The automatic control can be placed either on the oxidation air blower discharge pressure or the absorber mass flow.

Sometimes, a reheat system is designed to increase the temperature of the gas in the absorber outlet duct. To prevent acid condensation and to allow the gas to have enough buoyancy, the outlet temperature must be maintained at a minimum of around 180°F (82°C). Reheating of the gas is accomplished by passing the gas exiting the tower through reheater coils that are internally heated with steam. Whenever the flue-gas temperature exiting the reheater drops below the desired set point, the appropriate steam flow valve starts to open and effectively raises the temperature of the flue gas as it exits the absorber tower.

In practice, gypsum ($CaSO_4$), the by-product resulting from this process, can be reused in other applications. Modern post-combustion SO_2 technologies, however, consume as much as 1% of the energy produced in coal-fired power plants.

OPTIMIZATION OF BOILERS

The purpose of optimization is to continuously maximize the boiler efficiency, as variations occur in the load, fuel, ambient, and boiler conditions. When the yearly boiler fuel cost is in the millions, even a few percentage points of improved efficiency can justify the costs of added instruments and controls. In the following paragraphs, a number of optimization techniques will be described. The various goals of optimization include:

1. Minimize excess air and flue-gas temperature
2. Minimize steam pressure
   a. Turbines thereby open up turbine governors
   b. Reduce feed pump discharge pressures
   c. Reduce heat loss through pipe walls
3. Minimize blowdown
4. Measure efficiency
   a. Use the most efficient boilers
   b. Know when to perform maintenance
5. Provide accountability
   a. Monitor losses
   b. Recover condensate heat

The first three methods of optimization are achieved by closed-loop process control and can be superimposed upon the overall boiler control system shown in Figure 8.6l. The tie-in points for these optimization strategies are also shown in that figure. The benefits of the last two methods (efficiency and accountability) are not obtained in the form of closed-loop control signals, but they do contribute to better maintenance and better understanding of heat losses and equipment potentials.

Excess Air Optimization

If a boiler is operating on a particular fuel at a specific load, it is possible to plot the various boiler losses as a function of air excess or efficiency, as shown in Figure 8.6kkk. The sum total of all the losses is a curve with a minimum point. Any process that has an operating curve of this type is an ideal candidate for instrumental optimization. Such process control systems operate by continuously determining the

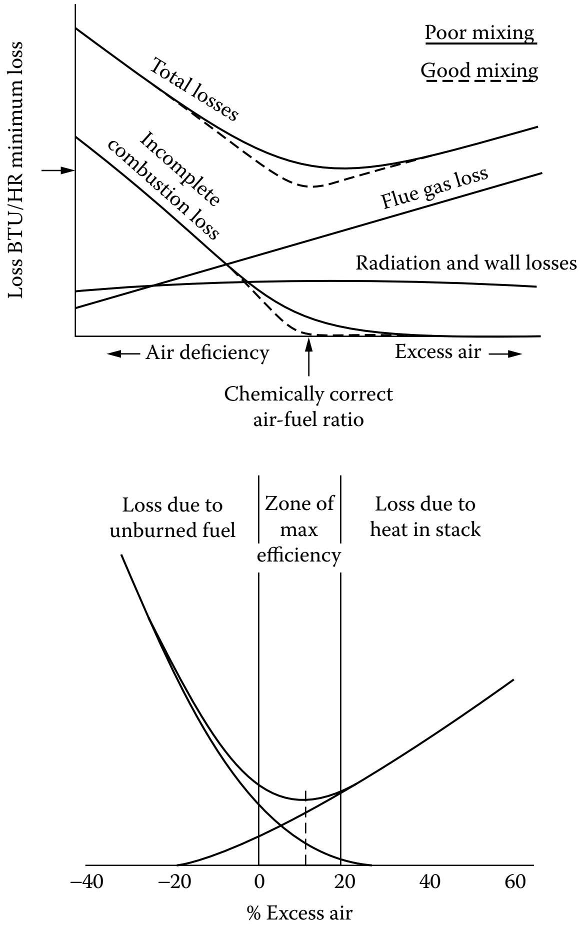


FIG. 8.6kkk

Boiler losses can be plotted as a function of excess air (top). The minimum of the total loss curve of a boiler is the point where optimized operation is maintained (bottom). Most efficient operation of a boiler occurs when the amount of excess air in the stack balances the losses in unburned fuel. (Adapted from Reference 2.)

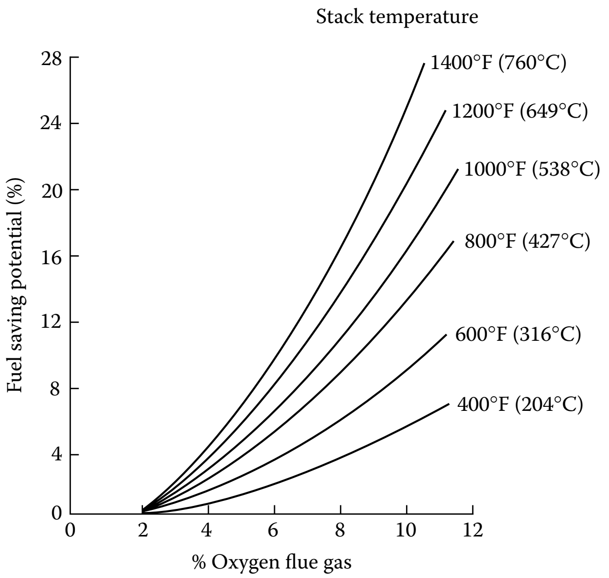


FIG. 8.6lll
The percentages of fuel-saving potentials in a boiler that operates most efficiently at 400°F stack gas temperature and 2% excess oxygen.

minimum loss point of the system at that particular load and then shifting the operating conditions until that point is reached.

As shown in Figure 8.6kkk, the radiation and wall losses are relatively constant. Most heat losses in a boiler occur through the stack. Under air-deficient operations, unburned fuel leaves, and when there is an air excess, heat is lost as the unused oxygen and its accompanying nitrogen are heated up and then discharged into the atmosphere. The goal of optimization is to keep the total losses at a minimum. This is accomplished by minimizing excess air and by minimizing the stack temperatures (Figure 8.6lll).

The minimum loss point in Figure 8.6kkk is not where excess oxygen is zero. This is because no burner is capable of providing perfect mixing. Therefore, if only as much oxygen would be admitted into the furnace as is required to convert each carbon molecule into CO_2, some of the fuel would leave unburned, as not all O_2 molecules would find their corresponding carbon molecules. This is why the theoretical minimum loss point shown by the dotted line in Figure 8.6kkk is to the left of the actual one. This actual minimum loss or maximum efficiency point is found by lowering the excess oxygen as far as possible, until opacity or CO readings indicate that the minimum has been reached. At this minimum loss point the flue-gas losses balance the unburned fuel losses.

Assume that for a particular boiler design using a particular fuel at normal loading, the optimum flue-gas temperature is 400°F (204°C) with 2% oxygen. The potential fuel savings through optimization can be estimated by determining the fuel loss using Figure 8.6lll, where the present, unoptimized stack gas conditions are entered.

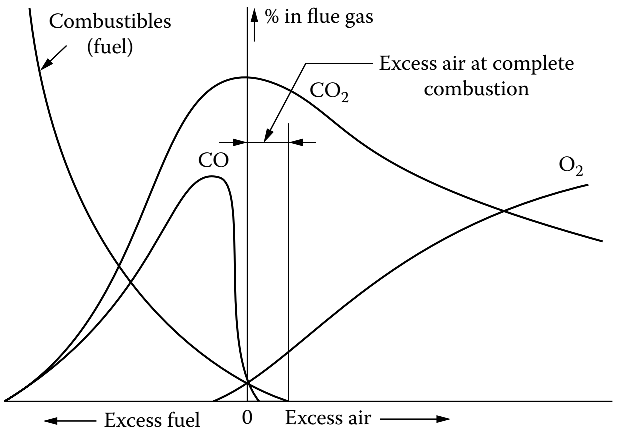


FIG. 8.6mmm
The major components of the flue gas are oxygen, carbon dioxide, carbon monoxide, and unburned hydrocarbons. (Adapted from Reference 2.)

Flue-Gas Composition Figure 8.6mmm shows the composition of the flue gas as a function of the amount of air present. The combustion process is usually operated so that enough air is provided to convert all the fuel into CO_2, but not much more. This percentage of excess oxygen is *not* a constant. It varies with boiler design, burner characteristics, fuel type, air infiltration rates, ambient conditions, and load.

The top portion of Figure 8.6nnn shows that the percentage of excess air must be increased as the load drops off. This is because at low loads the burner velocities drop off and the air flow is reduced, while the furnace volume remains constant. This reduces turbulence and lowers the efficiency of mixing between the fuel and the air. This loss of mixing efficiency is compensated for by the higher percentage of excess oxygen admitted at low loads.

The upper curve shown in Figure 8.6nnn theoretically illustrates the relationship between the excess O_2 requirement and load, and the lower plot provides actual test data for a specific boiler. Because each boiler has its own unique personality, this relationship must be experimentally determined. Once established, it can be used with a fair degree of confidence, although small shifts are still likely to occur as the equipment ages.

A simple demonstration of this law is the union of 1 lb of carbon with oxygen to produce a specific amount of heat (about 14,100 BTU, or 3,553 kcal). The gaseous product of combustion, CO_2 can be formed in one or two steps. If CO is formed first, it produces a lesser amount of heat (about 4,000 BTU, or 1,008 kcal), which, when it is converted to form CO_2, releases an additional 10,100 BTU (2,545 kcal). The sum of the heats released in each of these two steps equals the 14,100 BTU (3,553 kcal) that evolve when carbon is burned in one step to form CO_2 as the final product.

The Effect of the Fuel Used In a boiler furnace (where no mechanical work is done), the heat energy evolved from the

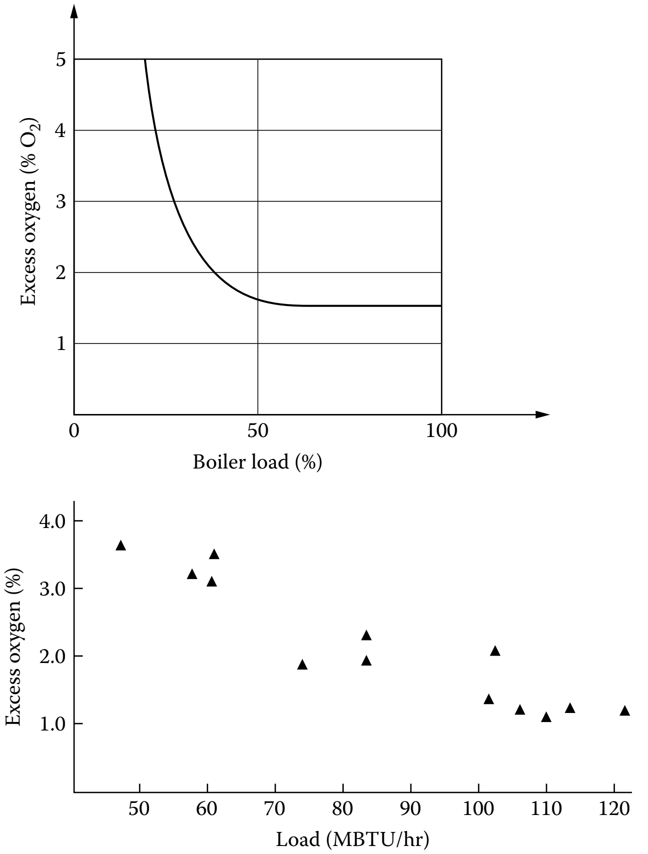


FIG. 8.6nnn

The top portion of this figure shows the theoretical relationship between load and excess oxygen. The bottom portion of the figure shows the test-based relationships between load and O_2. (Adapted from Reference 33.)

combustion of the fuel with oxygen depends on the heating value (calorific value) of the fuel, the degree to which the combustion reaction goes to completion.

Because of the weight ratios of oxygen and nitrogen in air (0.2315 and 0.7685, respectively), to supply 1 lb (0.45 kg) of oxygen for combustion it is necessary to supply 1/.2315 = 4.32 lb (1.96 kg) of air. In this amount of air, there will be 4.32 × .7685 = 3.32 lb (1.5 kg) nitrogen, which does not enter directly into the combustion process but which nevertheless remains present.

When burning carbon to carbon dioxide, 12 parts by weight of carbon (the approximate molecular weight of C) combine with 32 parts by weight of oxygen (the molecular weight of O_2) to form 44 parts by weight of carbon dioxide (the approximate molecular weight of CO_2). By simple division, 1 lb (0.45 kg) of carbon plus 2.66 lb (1.2 kg) of oxygen will yield 3.66 lb (1.66 kg) of carbon dioxide.

The theoretical amount of air required for the combustion of a unit weight of fuel can be calculated as follows:[4]

$$A{:}F = \frac{2.66\text{C} + 7.94\text{H}_2 + 0.998\text{S} - \text{O}_2}{0.232} \qquad \textbf{8.6(5)}$$

where $A{:}F$ is the air/fuel ratio, and C, H_2, S, and O_2 are the mass fractions (as-burned basis) of carbon, hydrogen, sulfur, and oxygen, respectively, in the fuel.

If theoretical or total air is defined as the amount required on the basis of the above equation, then excess air is the percentage over that quantity. Table 8.6ooo lists the excess air ranges required to burn various fuels. It can be seen that excess air requirement increases with the difficulty to atomize the fuel for maximum mixing.

Figure 8.6ppp also illustrates that gases require the lowest and solid fuels the highest percentage of excess oxygen for complete combustion. The ranges in Table 8.6ooo and the curves in Figure 8.6ppp also illustrate that as the load drops off, the percentage of excess oxygen needs to be increased.

Figure 8.6qqq illustrates the relationship between excess air and excess oxygen for a particular fuel. The optimum

TABLE 8.6ooo

Usual Amount of Excess Air Supplied to Fuel-Burning Equipment[*]

Fuel	*Type of Furnace or Burners*	*Excess Air %*
Pulverized coal	Completely water-cooled furnace—wet or dry-ash-removal	15–20
	Partially water-cooled furnace	15–40
Crushed coal	Cyclone furnace—pressure or suction	13–20
	Fluidized bed	15–20
Coal	Spreader, vibrating, and chain grate stokers	25–35
	Underfeed stokers	25–40
Fuel oil	Register-type burners	3–15
	Multifuel burners and flat flame	10–20
Acid sludge	Cone and flat flame-type burners, steam-atomized	10–15
Natural, coke oven, and refinery gas	Register-type burners	3–15
	Multifuel burners	7–12
Blast-furnace gas	Register-type burners	15–30
	Intertube nozzle-type burners	15–18
Wood/bark	Traveling grate, water-cooled vibrating grate	20–25
	Fluidized-bed	5–15
Refuse-derived fuels	Completely water-cooled furnace—traveling grate	40–60
Municipal solid waste	Water-cooled refractory covered furnace reciprocating grate	80–100
	Rotary kiln	60–100
Bagasse	All furnaces	25–35
Black liquor	Recovery furnaces for kraft pulping processes	15–20

[*] From Babcock & Wilcox Co., Reference 7.

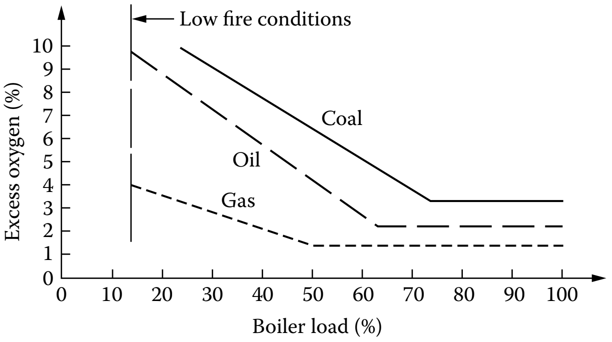


FIG. 8.6ppp
The ideal amount of excess oxygen provided to a boiler depends on the load as well as on the fuel properties.

excess oxygen percentages for gas, oil, and coal are around 1, 2, and 3%, respectively.

Detectors of Flue-Gas Composition The analyzers available for the detection of carbon dioxide, carbon monoxide, and excess oxygen are discussed in Sections 8.9, 8.10, and 8.42, respectively, in Chapter 8 in Volume 1 of this handbook. As shown in Figure 8.6mmm, excess air can be correlated to O_2, CO, CO_2, or combustibles present in the flue gas. Combustibles are usually detected either as unburned hydrocarbons or in the form of opacity. These measurements are not well suited as the basis for optimization, because the goal is not to maintain some optimum concentration, but to eliminate combustibles from the flue gas. Therefore, such measurements are usually applied as limit overrides.

The measurement of CO_2 is not a good basis for optimization either, because, as shown in Figure 8.6rrr, its relationship to excess O_2 is very much a function of the type of fuel burned. The CO_2 concentration of the flue gas also varies slightly with the CO_2 content of the ambient air.

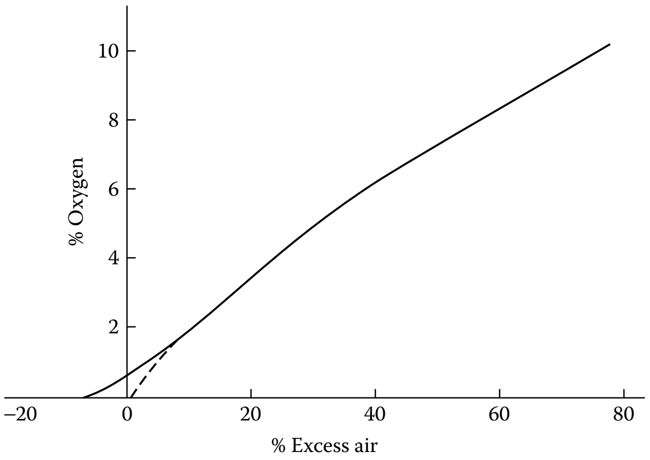


FIG. 8.6qqq
The amounts of oxygen and excess air in the flue gas can be correlated as shown.[2]

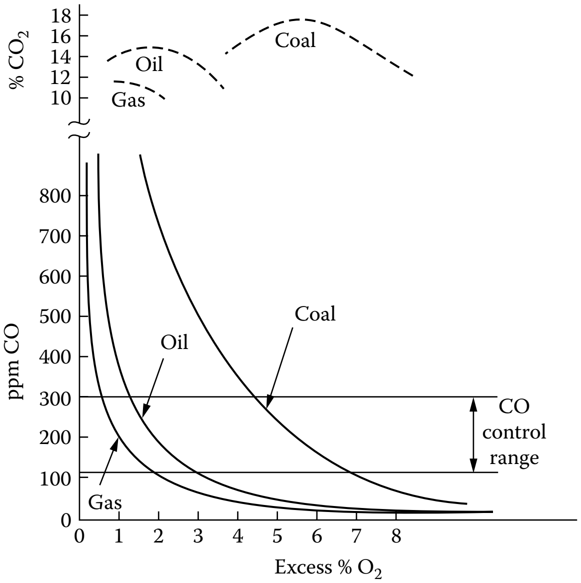


FIG. 8.6rrr
The relationship between excess O_2 and CO or CO_2 in the flue gas of a boiler operated at a constant load is a function of the type of fuel burned.[24]

It can also be noted from Figure 8.6mmm that CO_2 is not a very sensitive measurement. Its rate of change is rather small at the point of optimum excess air. In fact, the CO_2 curve is at its maximum point when the combustion process is optimized.

Excess O_2 as the basis of boiler optimization is also a relatively insensitive measurement, but it is popular. It uses zirconium oxide probes. In order to minimize duct leakage effects, the probe should be installed close to the combustion zone (Figure 8.6sss) but still at a point where the gas temperature is below that of the electrically heated zirconium oxide detector.

The flow should be turbulent at the sensor location, if possible, to ensure that the sample will be well mixed and representative of flue-gas composition. The output signal of these zirconium oxide probes is logarithmic. According to Shinskey[6] this is desirable. The correct location of the probe will reduce but will not eliminate the bias error caused by air infiltration. Ambient tramp air enters the exhaust ductwork (which is under vacuum) not only through leakage but also to cool unused burners and registers. The O_2 probe cannot distinguish the oxygen that entered through leakage from excess oxygen left over after combustion.

Another limitation of the zirconium-oxide fuel cell sensor is that it measures *net* oxygen. In other words, if there are combustibles in the flue gas, they will be oxidized on the hot surface of the probe and the instrument will register only that oxygen that remains *after* this reaction. This error is not

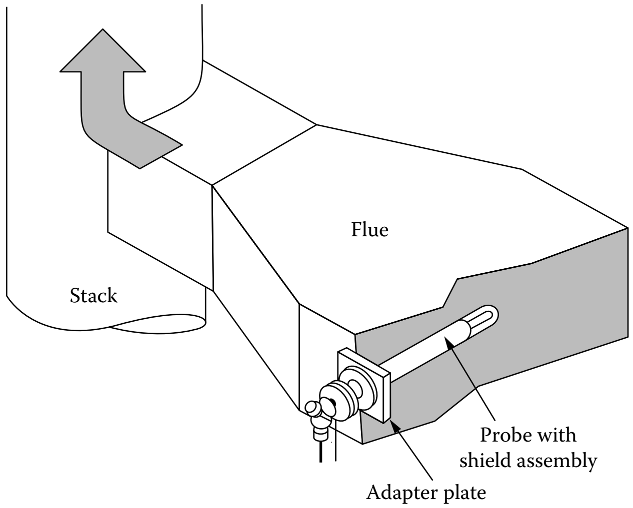


FIG. 8.6sss
The probe-type oxygen analyzer should be installed close to the combustion zone but at a point where the temperature is below the limit for the zirconium oxide detector.

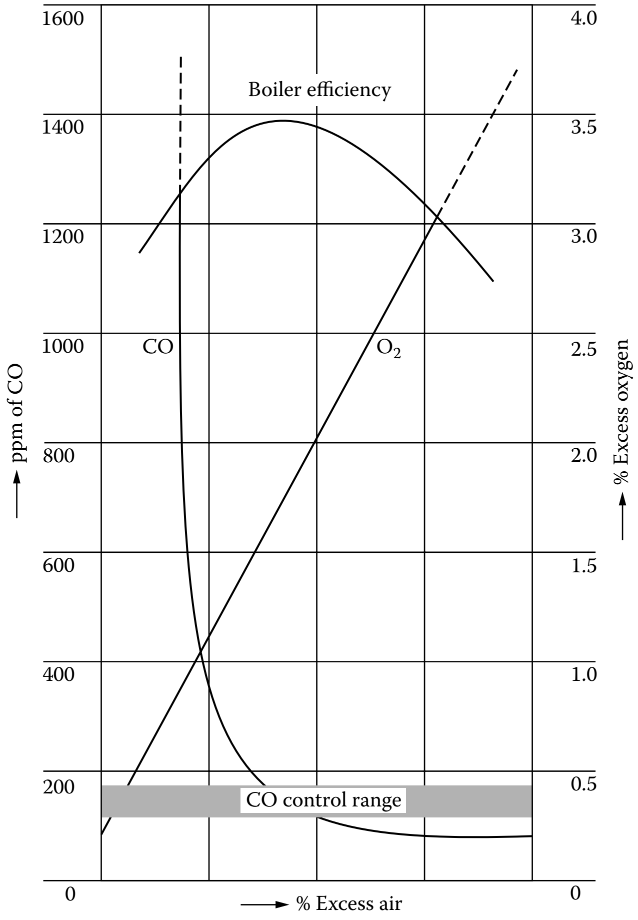


FIG. 8.6ttt
Gas burning boiler efficiency is maximum when CO is within the control range shown. (Courtesy of Econics Corp.)

substantial when the total excess oxygen is around 5%, but in optimized boilers in which excess oxygen is only 1%, this difference between total and net O_2 can cause a significant error. As infiltration tends to cause an error toward the high side, while the fuel-cell effect results in a low reading, the amount of uncertainty is too high to rely on O_2 sensors alone when maximum efficiency is desired.

Other limitations of optimization based on excess oxygen include the fact that local problems at the burners can result in incomplete combustion, even when the excess oxygen in the flue gas is normal. Another limitation is the precision and accuracy of such excess oxygen curves, as shown in Figure 8.6nnn. This precision is a function not only of the resolution at which the curve was prepared but also of changes in fuel composition and boiler conditions.

CO Measurement As shown in Figure 8.6mmm, the most sensitive measurement of flue-gas composition is the detection of carbon monoxide. As can be seen from Figures 8.6rrr and 8.6ttt, optimum boiler efficiency can be obtained when the losses due to incomplete combustion *equal* the effects of excess air heat loss. These conditions prevail at the "knee" of each curve. While the excess O_2 corresponding to these "knee" points varies with the fuel, the corresponding CO concentration is relatively constant.

Theoretically, CO should be zero whenever there is oxygen in the flue gas. In actual practice, maximum boiler efficiency can usually be maintained when the CO is between 100 and 400 ppm. CO is a very sensitive indicator of improperly adjusted burners; if its concentration rises to 1000 ppm, that is reliable indication of unsafe conditions. Because CO is a direct measure of the completeness of combustion and nothing else, it is also unaffected by air infiltration, other than the dilution effect.

For these reasons, control systems utilizing the measurement of both excess O_2 and carbon monoxide can optimize boiler efficiency, even if load, ambient conditions, or fuel characteristics vary. Also, when these systems detect a shift in the characteristic curve of the boiler, that shift can be used to signal a need for maintenance of the burners, heat-transfer units, or air and fuel handling equipment.

Nondispersive infrared (IR) analyzers can be used for simultaneous in situ measurement of CO and other gases or vapors such as that of water. This might signal incipient tube leakage. Most IR sensors use a wavelength of 4.7 μ for CO detection, because the absorption of CO peaks at this wavelength, whereas that of CO_2 and H_2O does not. CO_2 is also measured and is used to determine the dilution compensation factor for CO.

The CO analyzers cannot operate at high temperatures and therefore are usually located downstream of the last heat exchanger or economizer. At these points, the flue-gas dilution due to infiltration is frequently high enough to require compensation. The measurement of CO_2 is used to calculate this compensation factor.

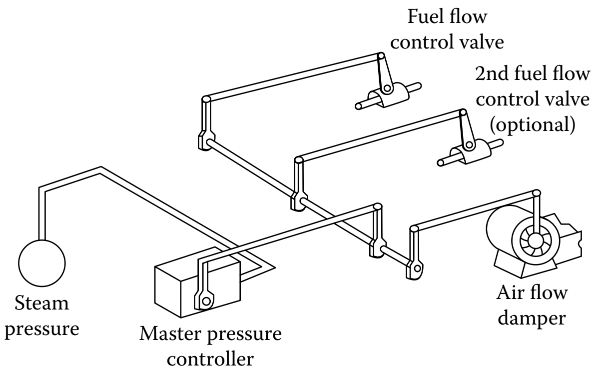


FIG. 8.6uuu
Early small boilers often used a direct positioning jackshaft modulated combustion control system with air/fuel ratio established through fixed mechanical linkages.[1]

Setting the Air/Fuel Ratio In Figure 8.6l, the set point of the air/fuel ratio relay is designated as the #1 tie-in point for the optimizer controls. This was done to emphasize the importance of this setting and to show that the method used to determine the correct air/fuel ratio will determine if the boiler is optimized or not.

In the early designs of small boilers, the air/fuel ratio was set by mechanical linkages between valve, damper, and the common jackshaft, as illustrated in Figure 8.6uuu. Even at fixed loads, these controls were only as good as the setting of the linkages that had to be readjusted manually as conditions changed.

When the importance of feedback control based on flue-gas analysis was better understood, such mechanically linked boiler controls were retrofitted as shown in Figure 8.6vvv. In this system the excess oxygen content of the flue gas is used to provide a feedback trim on the present relationship between firing rate (ZT) and damper opening (AZ). The influence of this trimming signal is bounded by the high/low limiter (AY-3) as a safety precaution to prevent the formation of fuel-rich mixtures as a result of analyzer or controller failure.

Air/Fuel Ratio with Excess Oxygen Trim Figure 8.6www shows an example of automatic fuel/air ratio correction based on load and excess air indicated by percentage of oxygen. In this control system, FY-102 provides the relationship between the load (steam flow) and the corresponding excess oxygen set point for optimum performance. Therefore, FY-102 memorizes the characteristic curve of the boiler for the particular fuel being used (see Figures 8.6nnn and 8.6ppp) and generates the excess oxygen set point based on that curve.

To obtain some of the advantages of the closed-loop fuel system, noninteracting oxygen analysis may be used to calibrate continuously the inherently poor fuel flow signal, if it could not otherwise be used with accuracy. An example of how a satisfactory coal flow signal can be obtained by continuously calibrating a summation signal of pulverizer feeder speeds is shown in Figure 8.6xxx.

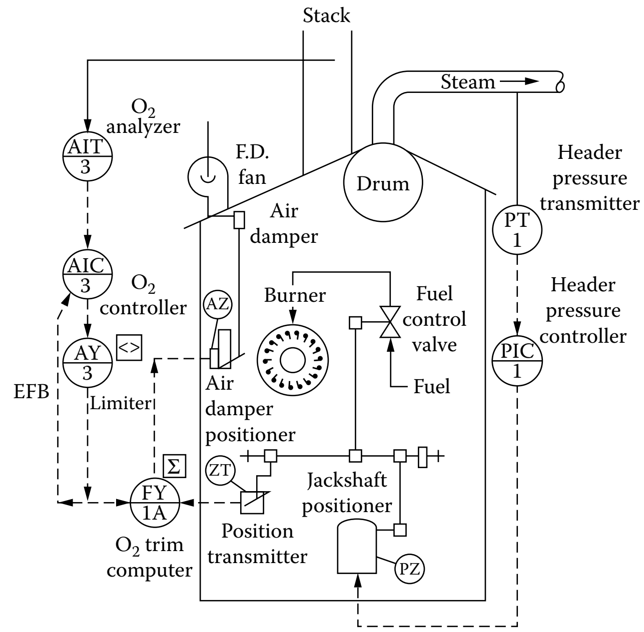
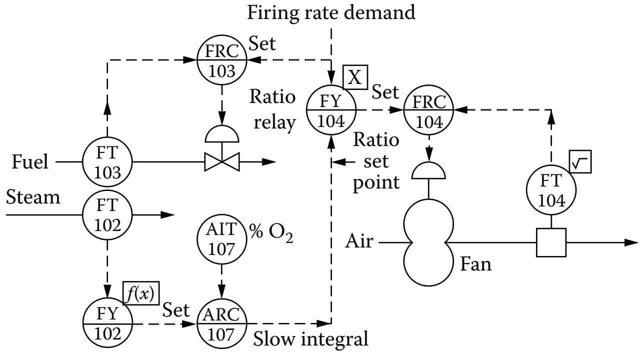


FIG. 8.6vvv
Parallel positioning combustion control systems can be retrofitted with excess oxygen trim.[34]

In both Figure 8.6www and Figure 8.6xxx, the set point of the excess oxygen controller is based on the steam flow.

Figure 8.6yyy illustrates a closed-loop control system corrected by oxygen analysis and provided with safety limits to protect against air deficiency. For the dual-selector system to function, air and fuel flows must be scaled on the same heat-equivalent basis.

The dual-selector system forces air flow to lead fuel on an increasing load and to lag on a decreasing load. Then, flue-gas oxygen content tends to deviate above the set point on *all* load changes. If the oxygen controller were allowed to react proportionally to these deviations, it would tend to defeat the security provided by the selectors. Consequently,



FT–103: Linear mass flowmeter
FT–102: Linear vortex shedding transmitter
FT–104: Area averaging pitot station

FIG. 8.6www
Air/fuel ratio control, with load vs. excess air curve (Figure 8.6nnn) considered.

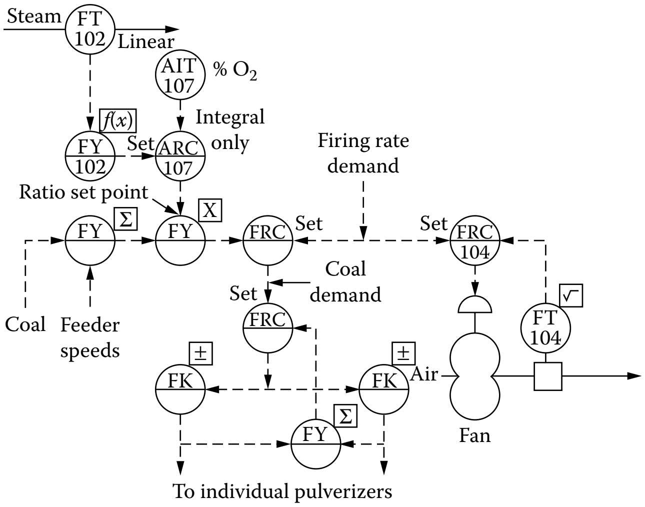


FIG. 8.6xxx
Load vs. excess air correction applied to the air/fuel ratio of a coal-burning boiler.

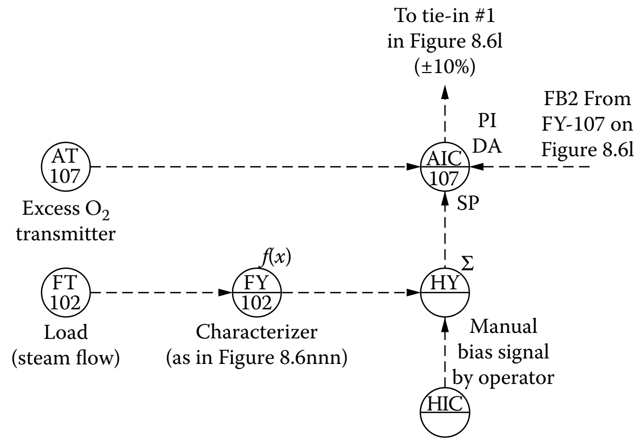


FIG. 8.6zzz
In this control system configuration, the excess oxygen trim has been characterized for load variations.

the integral control mode alone is used on the oxygen signal, so that reaction to rapid fluctuations is minimized. The principal function of the controller is to correct for long-term deviations caused by flowmeter errors and variations in fuel quality.

A variation of the previously described control system is shown in Figure 8.6zzz, in which FY-102 represents the relationship between load and excess oxygen. The input to FY-102 is steam flow (in other systems, firing rate is used as the input), and the output is the set point of the excess oxygen controller, AIC-107. The summer (HY) provides a bias so the operator can shift the characterizer curve up or down to compensate for changes in air infiltration rates or in boiler equipment performance.

The oxygen controller compares the measured flue-gas oxygen concentration to the load-programmed set point and applies PI action to correct the offset. Antireset windup and adjustable output limiting are usually also provided. The oxygen controller is direct-acting; the air/fuel ratio adjustment factor, therefore, increases if oxygen concentration in the stack rises because of effects such as reduced fuel heating value at constant flow. Increasing the air/fuel ratio adjustment factor raises the compensated process variable transmitted to the air flow controller, FIC-104 in Figure 8.6l.

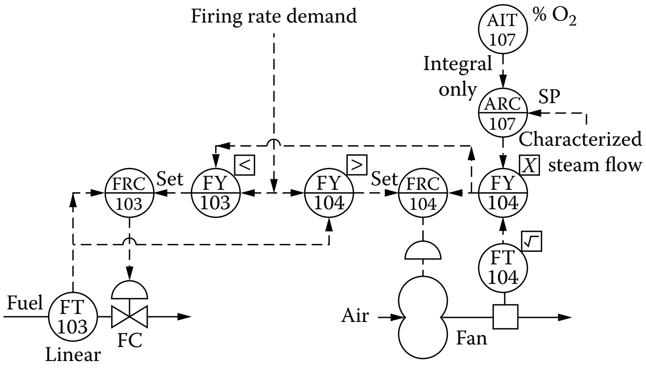


FIG. 8.6yyy
Feedforward control system that automatically maintains excess air during upsets.

The control system in Figure 8.6zzz can be further improved by increasing its speed of response. The transportation lag in a boiler can be as much as a minute. This lag is the time interval that has to pass after a change in firing rate before its effect can be detected in the composition of the flue gas. This dead time varies with flue-gas velocity and usually increases as the load drops. In boilers in which loads are frequently shed or added at irregular intervals, this dead time can cause control problems.

The feedforward strategy[25] shown in Figure 8.6aaaa can substantially lower this dead time. Here, the air flow damper position is controlled in a closed loop based on a set point developed through a characterization curve from the firing rate command. This loop acts to adjust the excess oxygen in a feedforward manner.

Feedback trim is provided by an oxygen measurement that modifies the set point to the damper position controller. This system anticipates the need for excess oxygen changes by responding to load swings, then correcting oxygen concentration to correspond with the excess air curve. A further refinement can be implemented using the corrections provided by the oxygen controller to adapt the damper characterization curve for a particular fuel to the current position in a learning mode.

Multivariable Envelope Control Multiple measurements of flue-gas composition can be used to eliminate the manual bias in Figure 8.6zzz and to obtain more accurate and faster control than what is possible with excess O_2 control alone.

Figure 8.6bbbb shows a control system in which the manual bias is replaced by the output signal of a CO controller,

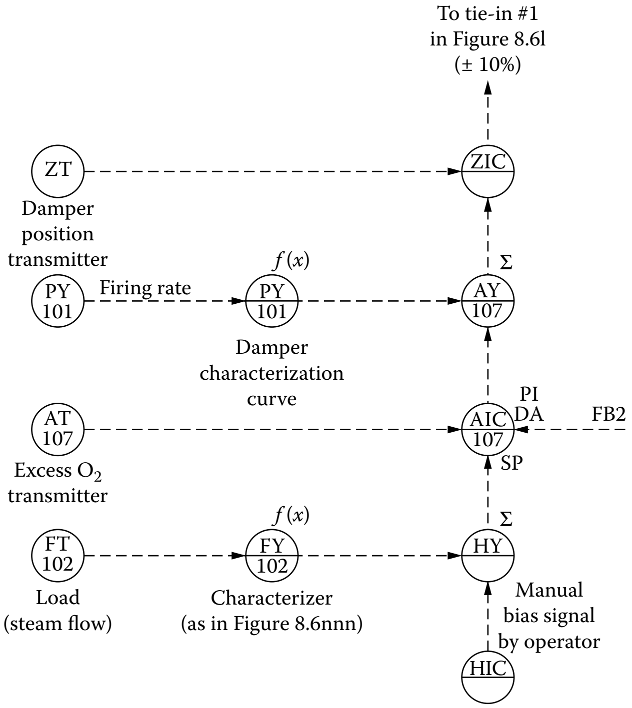


FIG. 8.6aaaa
The feedforward correction that is based on the characterized damper position increases the sensitivity of the control system. (Adapted from Reference 25.)

trimming the characterized set point of AIC-107. This trimming corrects the characterized excess O_2 controller set point for changes in the characteristic curve. These changes can be caused by local problems at the burners, resulting in incomplete combustion, or by changes in fuel characteristics, equipment, or ambient conditions.

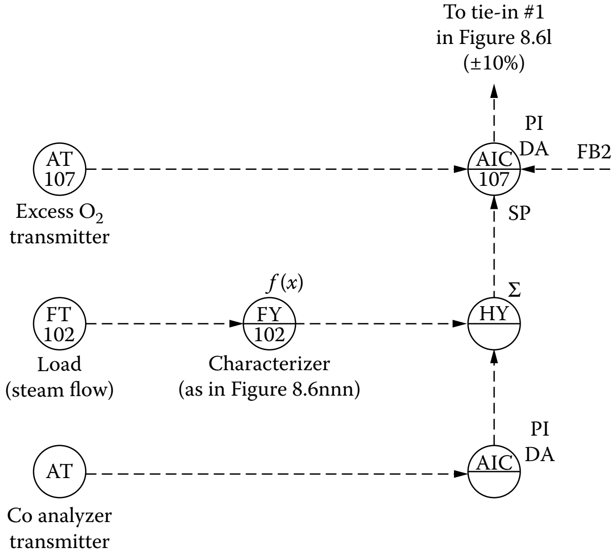


FIG. 8.6bbbb
In this configuration, the need for manual biasing is eliminated by the addition of fine trimming based on CO. (Adapted from Reference 25.)

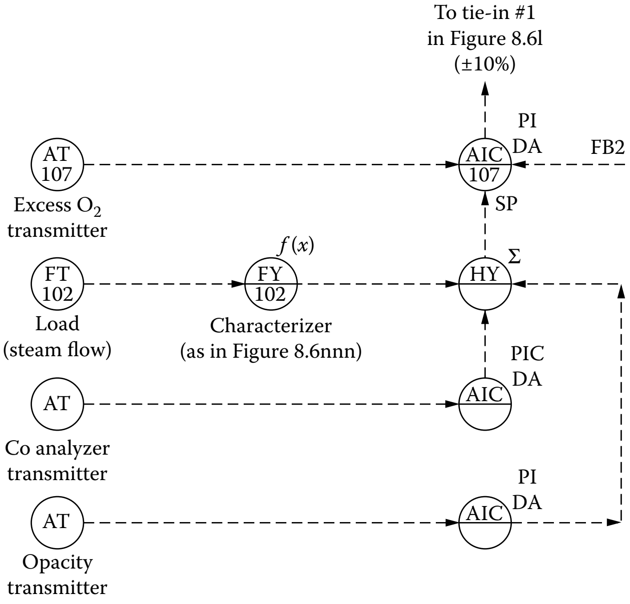


FIG. 8.6cccc
An opacity override has been added to the characterized CO to O_2 cascade system of Figure 8.6bbbb to meet environmental regulations. (Adapted from Reference 42.)

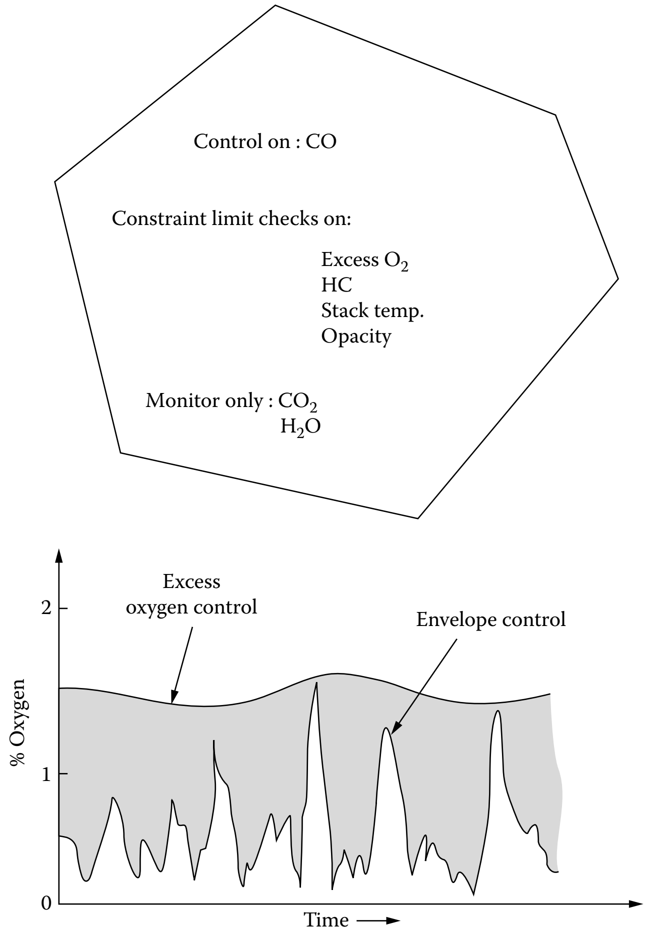
The control systems shown in Figure 8.6bbbb could be further improved if the CO measurement signal was corrected for dilution effects due to air infiltration, or if the CO controller set point was also characterized as a function of load. As was shown in Figure 8.6rrr, such characteristics can be determined for each fuel and firing rate. The control range for CO tends to remain relatively constant. As CO gives an indication of the completeness of combustion, it is *not* a feasible basis for control if the boiler is in poor mechanical condition or if the fuel does not combust cleanly.

Figure 8.6cccc provides the added feature of an opacity override to meet environmental regulations. In this system, under normal conditions, the cascade master is CO, just as it was in Figure 8.6aaaa. Similarly, the cascade slave is excess O_2, but when the set point of the opacity controller is reached, it will start biasing the O_2 set point upward until opacity returns to normal.

With microprocessor-based systems, it is possible to configure a control envelope, such as that shown in Figure 8.6dddd. With these control envelopes, several control variables are simultaneously monitored, and control is switched from one to the other, depending on which limit of the envelope is reached.

For example, assuming that the boiler is on CO control, the microprocessor will drive the CO set point toward the maximum efficiency (“knee” point in Figure 8.6rrr), but if in so doing the opacity limit is reached, that will override the CO controller and will prevent the opacity limit from being violated.

Similarly, if the microprocessor-based envelope is configured for excess oxygen control, it will keep increasing



Measured variable	Method of detection
CO	Infrared
Excess O_2	Zirconium oxide
HC (unburned hydrocarbons)	Infrared
Opacity	Visible light
CO_2	Infrared
H_2O	Infrared

FIG. 8.6dddd
Multivariable envelope-based constraint control can lower overall excess oxygen. This can be achieved by monitoring both carbon dioxide and water and by performing constraint limit checks on excess oxygen, hydrocarbons, stack temperature, and opacity.

boiler efficiency by lowering excess O_2 until one of the envelope limits is reached. When that happens, control is transferred to that constraint parameter (CO, HC, opacity, and so on); through this transfer, the boiler is "herded" to stay within the envelope defined by these constraints. These limits are usually set to keep CO under 400 ppm, opacity below #2 Ringlemann, and HC and NO_x below regulations.

Microprocessor-based envelope control systems usually also include subroutines for correcting the CO readings for dilution effects or for responding to ambient humidity and temperature variations. As a result, these control systems tend to be both more accurate and faster in response than if control was based on a single variable. The performance levels of a gas-burning boiler under both excess O_2 and envelope control are shown on the lower part of Figure 8.6dddd.

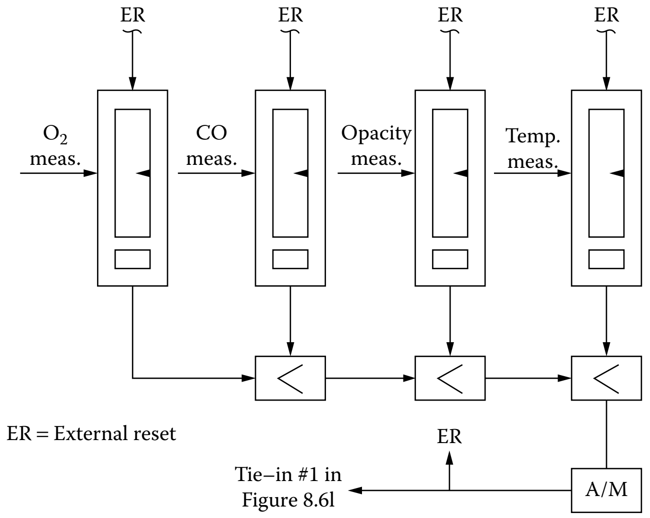


FIG. 8.6eeee
Analog controllers configured into a multiple selective loop can also be used to implement envelope control.

Envelope control can also be implemented by analog controllers that are configured in a selective manner. This is illustrated in Figure 8.6eeee. Each controller measures a different variable and is set to keep that variable under (or over) some limit. The lowest of all the output signals is selected for controlling the air/fuel ratio in Figure 8.6l, which ensures that the controller that is most in need of help is selected for control. Through this herding technique, the boiler process is kept within its control envelope. As shown in Figure 8.6eeee, reset windup in the idle controllers is prevented by the use of external reset, which also provides bumpless transfer from one controller to the next.

Operator access is shown to be provided by a single auto/manual (A/M) station. A better solution is to provide each controller with an A/M station. Then, if a measurement is lost, only the defective loop needs to be switched to manual, not the whole system.

Flue-Gas Temperature

The amount of energy wasted through the stack is a function of both the amount of excess air and the temperature at which the flue gases leave. The flue-gas temperature is a consequence of load, air infiltration, and the condition of the heat-transfer surfaces. Like any other heat exchanger, the boiler will also give its most efficient performance when clean and well maintained.

In optimized boiler controls, a plot of load vs. stack temperature is made when the boiler is in its prime condition, and this plot is used as a reference baseline in evaluating boiler performance. If the stack temperature rises above this reference baseline, this indicates a loss of efficiency. Each 50°F (23°C) increase will lower the boiler efficiency by about 1%.

On a 100,000 lb/hr (45,450 kg/hr) boiler, this is a yearly loss of about $20,000.

The reason for rising flue-gas temperatures can be fouling of heat-transfer surfaces in the air preheater, scale build-up on the inside of the boiler tubes, soot build-up on the outside of the boiler tubes, or deteriorated baffles that allow the hot gases to bypass the tubes. Some microprocessor-based systems will respond to a rise in flue temperature by taking corrective action, such as automatically blowing the soot away, or by giving specific maintenance instructions (see the paragraph on Soot Blowing Optimization).

If the stack temperature drops below the reference baseline, this does *not* necessarily signal an increase in boiler efficiency. More likely, it can signal the loss of heat due to leakage. Cold air or cold water can leak into the stack gases if the economizer or the regenerator are damaged. The consequence of this is loss of efficiency and the danger of corrosive condensation (sulfuric acid formation), if the temperature drops down to the dew point. Minimum flue-gas temperatures are approximately 250°F (121°C) for natural gas and 300–325°F (149–163°C) for heavy oil and wood.[5] A drop in flue-gas temperature will also lower the stack effect, thus increasing the load on the induced-draft fan.

For the above reasons, the advanced envelope control systems (Figure 8.6dddd) include both high- and low-limit constraints on stack temperature, using the above-described baseline as a reference.

Fuel Savings through Optimization

The overall boiler efficiency is the combined result of its *heat-transfer efficiency* and its *combustion efficiency.* Heat-transfer efficiency is reflected by stack temperature, which at the hot end should not exceed steam temperature by more than 150°F (65°C) when excess air is near optimum.[23] Combustion efficiency is tied to excess oxygen, which is brought as low as possible without exceeding the limits on CO (usually 400 ppm), opacity (#2 Ringlemann), unburned carbon, and NO_x (although reduction in excess O_2 is usually accompanied by reduction in NO_x).

When optimization reduces the flue-gas losses, the resulting savings can be estimated from the amount of reduction in these losses. In the case of a 100,000 lb/hr (45,450 kg/hr) steam boiler, a 1% reduction in fuel consumption (a 1% increase in efficiency) will lower the yearly operating costs by about $20,000 if the fuel cost is estimated at $2 per million BTU.

With the current average efficiency and the projected or actual efficiency after optimization, fuel savings can be calculated with Equation 8.6(6):

$$S = \left(1 - \frac{\eta_b}{\eta_i}\right)(F) \qquad \textbf{8.6(6)}$$

where S is the fuel savings, η_b is the base efficiency, η_i is the improved efficiency, and F is the current or base fuel usage.

The fuel savings resulting from the lowering of excess O_2 can be estimated from graphs, such as those shown in Figures 8.6ffff and 8.6gggg. On these graphs the temperature is the difference between the stack and ambient temperatures. If, in a boiler operating at a 500°F (260°C) stack temperature difference, optimization lowers the excess O_2 from 5 to 2%,

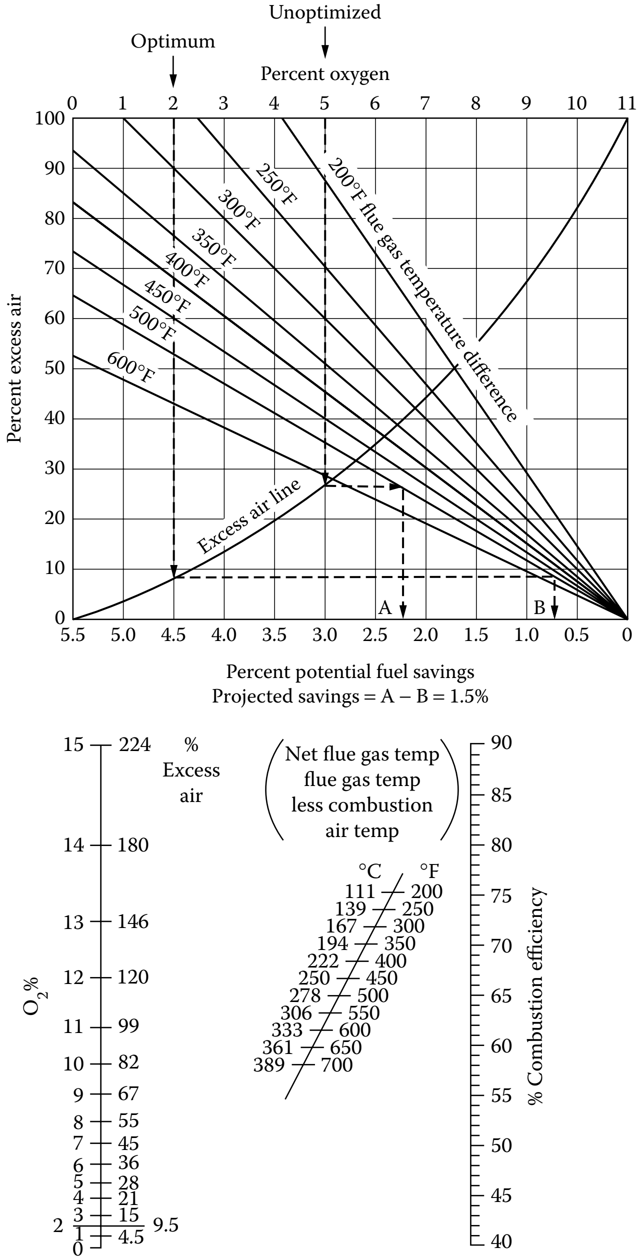


FIG. 8.6ffff

In the upper portion of this figure, one can determine the achievable fuel savings resulting from the reduction in excess oxygen. From the lower portion, one can obtain the combustion efficiency by drawing a straight line through the points corresponding to the applicable excess oxygen and flue-gas temperature. (Adapted from Reference 36.)

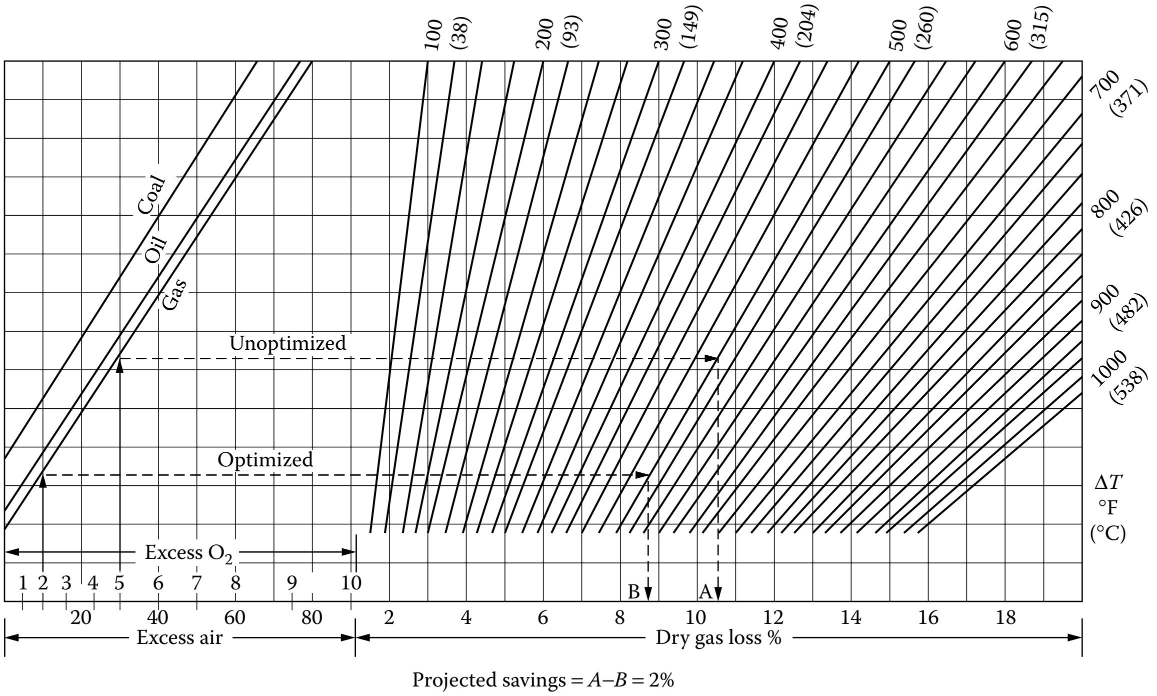


FIG. 8.6gggg
The fuel-savings potential can also be computed as shown here. (Courtesy of Dynatron, Inc.)

this will result in a fuel savings of 1.5% (2.25% − 0.75%), according to Figure 8.6ffff.

For the same conditions, Figure 8.6gggg gives a savings of 2% (10.5% − 8.5%). Such differences are acceptable, as the size and designs of the equipment do influence the results. Although optimization has lowered the excess O_2 in some boilers to around 0.5%, the value of these last increments of savings tends to diminish in proportion to the cost of accomplishing them. Lowering excess O_2 from 1 to 0.5% will increase boiler efficiency by only 0.25%, a savings of about \$5,000/yr in a 100,000 lb/hr (45,000 kg/hr) boiler, and the controls needed to sustain such operation need to be as sophisticated as the ones described in Figure 8.6dddd. For these reasons, an optimization target of 1% excess O_2 on gas fuel is reasonable.

The fuel savings resulting from improved thermal efficiency can be estimated from Figure 8.6gggg or Figure 8.6hhhh. Lowering the stack temperature difference from 500°F (278°C) to 400°F (222°C) at an excess O_2 of 2% will result in a savings of 1.7% (8.5% − 6.8%), according to Figure 8.6gggg. Assuming an ambient temperature of 60°F (16°C), these same conditions are marked in Figure 8.6hhhh at 560°F (293°C) and 460°F (238°C), resulting in a savings of about 2%. Therefore, as a crude approximation, fuel savings of 1% can be estimated for each 50°F (28°C) reduction in stack gas temperature.

As was already discussed, stack gas temperature reductions must be limited to achieving a temperature above dew point that is high enough to provide the required stack effect.

Consequently, the total savings from improved thermal and combustion efficiencies can be estimated on the basis of both stack temperature and excess oxygen being lowered to their limits. The resulting total fuel savings can be around 5%.

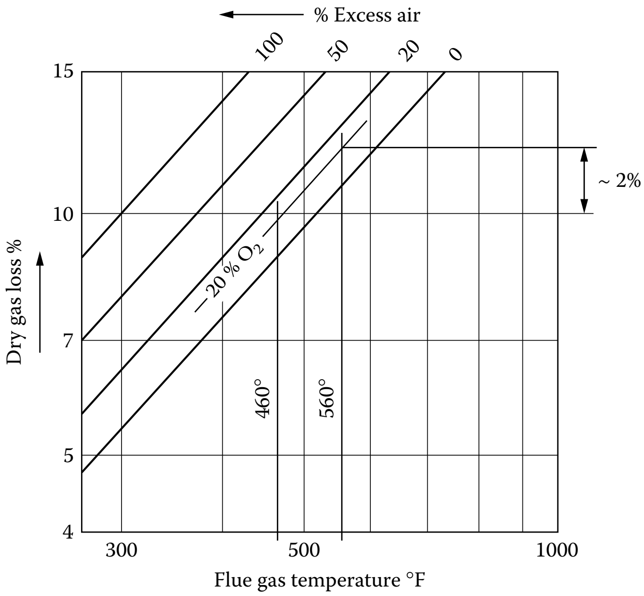


FIG. 8.6hhhh
Temperature of flue gas leaving a medium-size boiler has a direct effect on combustion efficiency. (Adapted from Reference 2.)

Steam Pressure Optimization

In Figure 8.6l, the second tie-in point for the optimizer is the set point of the pressure controller PIC-101. In basic boiler operation, the steam pressure is maintained at a constant value. Optimization is possible by allowing this pressure to vary.

In co-generating plants, in which the boiler steam is used to generate electricity and the turbine exhaust steam is used as a heat source, optimization is obtained by maximizing the boiler pressure and minimizing the turbine exhaust pressure, so as to maximize the amount of electricity generated (Figure 8.6iiii).

In plants that do not generate their own electricity, optimization is achieved by minimizing boiler operating pressure. This reduces the pressure drops in turbine governors by opening them up further, lowers the cost of operating feedwater pumps because their discharge pressure is reduced, and generally lowers radiation and wall losses in the boiler and piping.

In electric utility plants, optimization is achieved through sliding-pressure or free-pressure control strategies that allow the pressure to vary with load, within certain constraints. This reduces temperature fluctuations in the steam turbine, thereby reducing fatigue and improving turbine life, and it improves heat rate at lower loads. In basic sliding pressure control, the turbine governor valves are run wide-open (full-arc admission), and pressure reducing valves are used upstream to control pressure, or the boiler itself is cycled to change header pressure with load.

Pump power is particularly worth saving in high-pressure boilers, because as much as 3% of the gross work produced by a 2400 PSIG (16.56 MPa) boiler is used to pump feedwater. Figure 8.6jjjj illustrates the method of finding the optimum minimum steam pressure, which then becomes the set point for the master controller PIC-101 in Figure 8.6l.

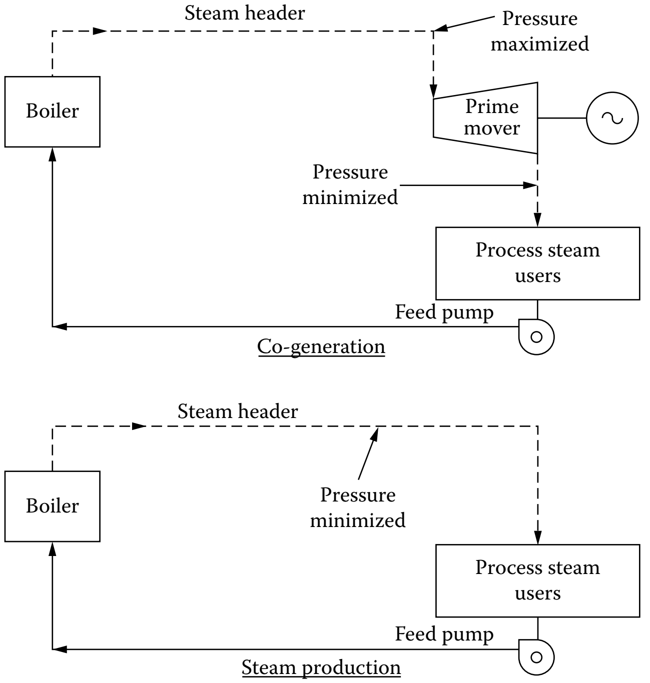


FIG. 8.6iiii
The optimum steam pressure at the boiler depends on the purpose that the steam serves.

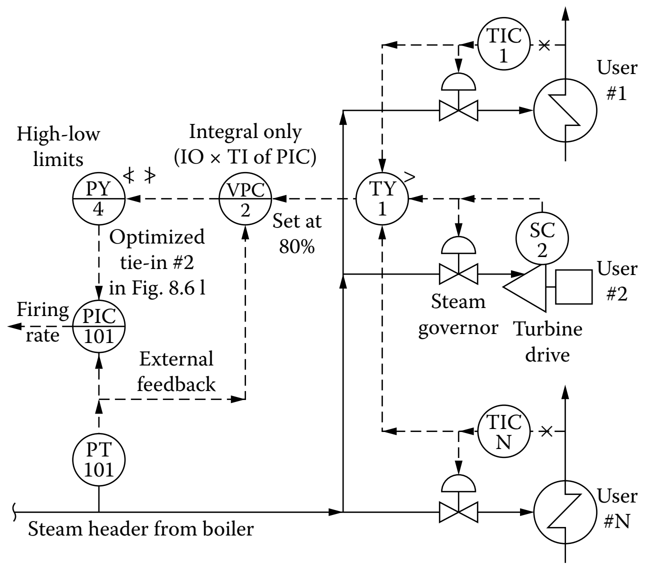


FIG. 8.6jjjj
The set point of the steam supply header pressure controller (PIC-101) can be floated, so as to keep the most-open user valve nearly full open.

As long as all steam user valves (including all turbine throttle valves) are less than fully open, a lowering in the steam pressure will not restrict steam availability, because the user valves can open further. The high-signal selector (TY-1) selects the most-open valve, and the valve position controller (VPC-2) compares that signal with its set point of, for example, 80%. If even the most-open valve in the plant is less than 80% open, the pressure controller set point is slowly lowered. VPC-2 is an "integral only" controller; its reset time is at least ten times that of PIC-101.

This slow integral action guarantees that only very slow "sliding" of the steam pressure will occur and that noisy valve signals will not upset the system, because VPC-2 responds only to the integrated area under the error curve. The output signal from VPC-2 is limited by PY-4, so that the steam pressure set point cannot be move outside some preset limits. This necessitates the external feedback to VPC-2, so that when its output is overridden by a limit, its reset will not wind up.

This kind of optimization, in which steam pressure follows the load, not only increases boiler efficiency but also does the following:

1. Prevents any steam valve in the plant from fully opening and thereby losing control
2. Opens all steam valves in the plant, thereby moving them away from the unstable (near-closed) zone of operation

3. Reduces valve maintenance and increases valve life by lowering pressure drop
4. Increases turbine drive efficiencies by opening up all steam governors

The total savings in yearly operating costs resulting from optimizing the steam pressure to follow the load is a small percentage of the total cost.

Steam Temperature Optimization

Dynamic Feedforward Although the feedforward function FY-102 in Figure 8.6ggg is commonly used in superheater steam temperature control, the transient response due to major disturbances (e.g., load swing) is not properly compensated under this design. The reason is that the process dynamics are not taken into account. By incorporating dynamic models in the feedforward control design, the transient error, such as the temperature overshoot, can be effectively eliminated. The dynamic feedforward compensator can be simply chosen in the form of a causal transfer function that is equivalent to a lead-lag type compensator readily available in most control systems' algorithm library.

Refer to Figure 8.6ggg and let the dynamics from the set point of TIC-112 to TT-111 be modeled by the transfer function $G(s)$, and the dynamics from FT-102 to TT-111 be represented by the transfer function $G_d(s)$. This assumes no control action from TIC-111. Let us also assume the steam flow disturbance can be transformed as $d(s)$, and the static feedforward compensator FY-102 is replaced by a dynamic controller $G_{ff}(s)$. To completely eliminate the transient error, the following equation should hold if there is no mismatch between the model and the actual process:

$$G_{ff}(s) \cdot G(s) \cdot d(s) + G_d(s) \cdot d(s) = 0 \qquad \textbf{8.6(7)}$$

This immediately yields

$$G_{ff}(s) = -\frac{G_d(s)}{G(s)} \qquad \textbf{8.6(8)}$$

If the feedforward controller $G_{ff}(s)$ is designed as such, the disturbance effect can be completely eliminated before it creates any transient error. However, partly due to the complexity of data collection and modeling task, this approach has not been widely used in current industrial practice. In fact, for steam temperature dynamics modeling purposes, much of the existing plant data stored for the conventional control design can be utilized.

For example, as part of the unit calibration and tuning process, the relationship between the steam flow rate and the spray water control valve input signal needs to be determined by testing at the design steam temperature while operating at different boiler loads. Much of the test data can be directly used for dynamic modeling purposes, although these data are mainly used to characterize the steady-state relationship in a conventional scheme.

In order to make the dynamic feedforward controller practically useful, the process model $G(s)$ has to be a minimum-phase transfer function. This means that the process model cannot have inverse dynamics at the beginning of the transient. Details about the dynamic feedforward design can be found in Section 2.9 in Chapter 2. Due to process nonlinearity, multiple linear models should be identified and used for control design at different loads.

It is suggested that the flue-gas temperature at the superheater inlet is a better signal to be used for feedforward compensation. Although the firing rate master signal and its associated variables (e.g., air/steam flow) are often indicators for steam temperature change, the relationship that can be derived is really a coarse estimate. It is the flue-gas temperature and mass flow that directly influence the convection area and the resulting steam temperature. A dynamic feedforward control scheme based on radiation pyrometry is briefly introduced in Reference 35.

Model-Based Multivariable Control The steam temperature control strategies discussed so far are mainly for superheater sections. In reality, many large utility boilers have at least one reheater section. Although the reheat steam temperature control is similar to the superheat steam temperature control in principle, the subtle differences are often overlooked. First, in the current practice, the flue-gas bypass damper and burner tilting, whenever available, are normally used as the first choice for reheat steam temperature control.

Spray water will be engaged in control action only when other methods are ineffective (possibly due to control output saturation). This measure can significantly reduce spray water usage and help to improve unit heat rate. Second, the reheat outlet steam temperature control is usually more challenging due to the fact that more flue-gas variation is expected in the reheat section.

Given the fact mentioned above, it is obvious that the superheat and reheat steam temperature control is a highly coupled process. Manipulation of flue-gas bypass damper or burner tilts will inevitably affect both superheat and reheat steam temperature, either in the same direction or in the opposite directions. For boilers with split furnace, sometimes the burner tilting and bypass damper movement can have different impact on steam temperatures of different sides. Moreover, as the steam temperature control fights against the interactions, load changes might be called for. The changing firing rate may introduce a number of other disturbances that do not necessarily come in the same fashion.

These include, but are not limited to, air flow, furnace-to-windbox differential pressure, steam flow, and steam pressure. More interactions are identified in Reference 12. A coordinated multivariable control design becomes a natural choice to achieve better performance for the steam temperature regulation. The manipulated variables may include the burner tilting angles, flue-gas damper positions, and spray water flows. The controlled variables will be superheater and reheater outlet steam temperatures. The details of model-based multivariable control system design can be found in Section 2.13 in Chapter 2.

Water Side Optimization

The optimization of the water side of a steam generator includes the optimized operation of the feedwater pump at the condensate return system and of the boiler blowdown. As pumping system optimization is the subject of a separate chapter in this book, only the boiler blowdown will be discussed here.

Blowdown Optimization In Figure 8.6l, the third tie-in point for the optimizer is the set point of the blowdown flow controller. The goal of optimization is to minimize blowdown as much as possible without causing excessive sludge or scale build-up on the inside surfaces of the boiler tubes. The benefits of such optimization include the reduction in the need for makeup water and treatment chemicals and the reduction in heat loss as hot water is discharged. About 90% of the blowdown should occur continuously, and 10% would result from the periodic blowing down of the mud drum and of the headers.

Blowdown can be optimized by automatically controlling the chloride and conductance of the boiler water. The neutralized conductivity set point is usually around 2500 micromhos. Automatic control maintains this set point within ±100 micromhos. The required rate of blowdown is a function of the hardness, silica, and total solids of the makeup water and also of the steaming rate and condensate return ratio of the boiler.

The amount of blowdown can be determined as follows:

$$\mathrm{BD} = \frac{S - R}{C - 1} \qquad \mathbf{8.6(9)}$$

where

BD = blowdown rate, lb per hour
R = rate of return condensate, lbs per hour
S = steam load, lbs per hour
C = cycles of concentration based on makeup

The value for cycles of concentration is generally determined on the basis of the chloride concentration of the boiler water divided by the chloride content of the makeup water. The value is also given by dividing the average blowdown rate into the average rate of makeup water, assuming no mineral contamination in any returned condensate.

Figure 8.6kkkk illustrates that the rate of blowdown accelerates as the boiler water conductivity set point is lowered. A reduction of about 20% can result from converting the blowdown controls from manual to automatic.[31] In the case of a 100,000 lb/hr (45,450 kg/hr) boiler, this can mean a reduction of 1,340 lb/hr (600 kg/hr) in the blowdown rate. If the blowdown heat is not recovered, this can lower the yearly operating cost by about $10,000 (depending on the unit cost of the fuel).

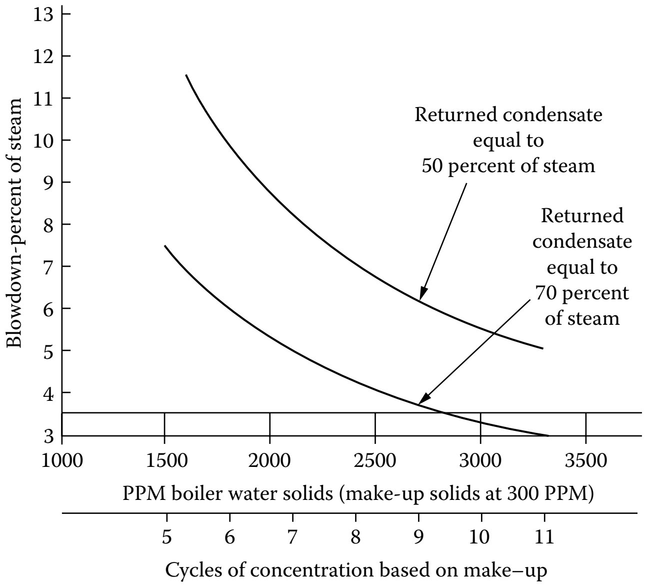


FIG. 8.6kkkk
The rate of blowdown increases as the boiler water conductivity set point is lowered.[31]

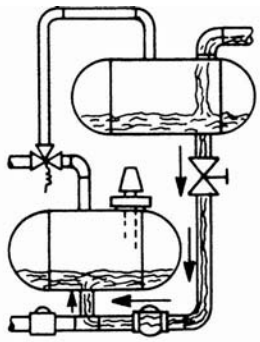
Overall boiler efficiency can also be increased if the heat content of the hot condensate is returned to the boiler. Pumping water at high temperatures is difficult; therefore, the best choice is to use pumpless condensate return systems. Figure 8.6llll illustrates the operation of such a system, which uses the steam pressure itself to push back the condensate into the deaeration tank. This approach eliminates not only the maintenance and operating cost of the pump but also the flash and heat losses, resulting in the return of more condensate at a higher temperature.

Alternatively, blowdown heat can be recovered by using a heat exchanger to preheat boiler makeup water. Consider a boiler producing 400,000 lb/hr (180,000 kg/hr) steam at a drum pressure of 900 PSIG (6.2 MPa), with a blowdown rate of 5% (percentage of feedwater); Assuming makeup water at 60°F (16°C), a heat exchanger “approach” ΔT of 2°F (1.1°C), and 90% heat recovery from blowdown, energy savings would be nearly 9 million BTU/hr (9.5 million kJ/hr).

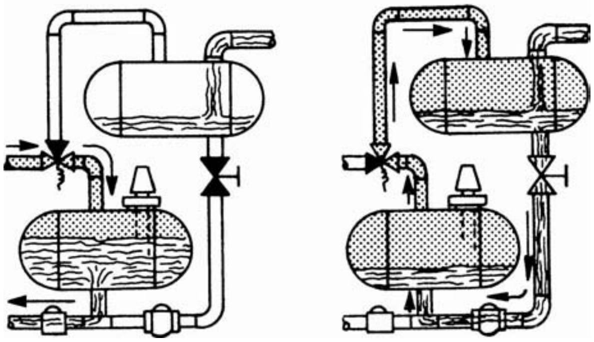
The performance of the steam and condensate piping system in the plant can also be improved if steam flows are metered. Such data is helpful not only in accountability calculations but also in locating problem areas, such as insufficient thermal insulation or leaking traps.

Load Allocation-Based Optimization

The purpose of load allocation between several boilers is to distribute the total plant demand in the most efficient and optimized manner. Such optimization will reduce the steam



Fill cycle



Equalization and discharge cycles

Vent cycle

FIG. 8.6llll
The "pumpless" condensate return system uses the steam pressure itself to push the condensate back into the deaeration tank. (Courtesy of Johnson Corp.)

production cost to a minimum. Such computer-based energy management systems can operate either in an advisory or in a closed-loop mode. The closed-loop systems automatically enforce the load allocation, without the need for operator involvement. The advisory system, on the other hand, provides instructions to the operator but leaves the implementation up to the operator's judgment.

The load allocation techniques discussed in this section, which are often referred as economic dispatch, only cover the scenario where all units in consideration are located in one plant or in a nearby neighborhood. Load allocation at grid level is more complicated, because the transmission loss, transmission flow constraints, reactive power constraints, fuel transportation, production scheduling, and many other factors need to be taken into account.

In simple load allocation systems, only the starting and stopping of the boilers is optimized. When the load is increasing, the most efficient idle boiler is started (Figure 8.6a); when the load is dropping, the least efficient one is stopped. In more sophisticated systems, the load distribution between operating boilers is also optimized. In such systems, a computer is used to calculate the real-time efficiency of each boiler. This information is used to calculate the incremental steam cost for the next load change for each boiler.

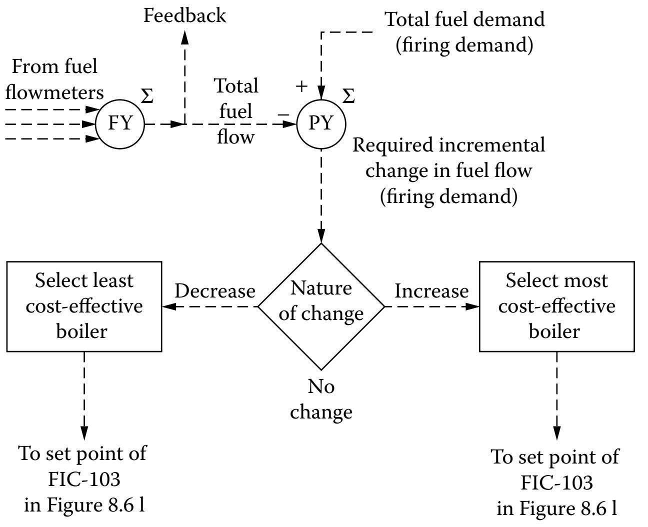


FIG. 8.6mmmm
Computer-based load allocation directs load increases to the most cost-effective boiler and sends load decreases to the least cost-effective boiler.

For example, if the load increases, the incremental increase is sent to the set point of the most cost-effective boiler. If the load decreases, the incremental decrease is sent to the least cost-effective boiler (Figure 8.6mmmm). The required software packages with proven capabilities for continuous load balancing through the predictions of costs and efficiencies are readily available.[32] With the strategy described in Figure 8.6mmmm, the most efficient boiler either will reach its maximum loading or will enter a region of decreasing efficiency and will no longer be the most efficient.

When the loading limit is reached on one boiler, or when a boiler is put on manual, the computer will select another as the most efficient unit for future load increases. On the other hand, the least efficient boiler will accept all decreasing load signals until its minimum limit is reached. Its load will not be increased unless all other boilers are at their maximum load or in manual. As shown in Figure 8.6a, some boilers can have high efficiency at normal load while being less efficient than the others at low load. Such units are usually not allowed to be shut down but are given a greater share of the load by a special subroutine.

If all boilers are identical, some will be driven to maximum capacity and others will be shut down by this strategy, and only one boiler will be placed at an intermediate load.[6] Boiler efficiency can be monitored indirectly (by measurement of flue-gas composition, temperature, combustion temperature, and burner firing rate) or directly (through time-averaged steam and fuel flow monitoring).

For the direct efficiency measurement, it is important to select flowmeters with acceptable accuracy and rangeability (Table 8.6d). In order to arrive at a reliable boiler efficiency reading, the error contribution of the flowmeters, based on actual reading, must not exceed $\pm 1/2$ to $\pm 3/4$%.

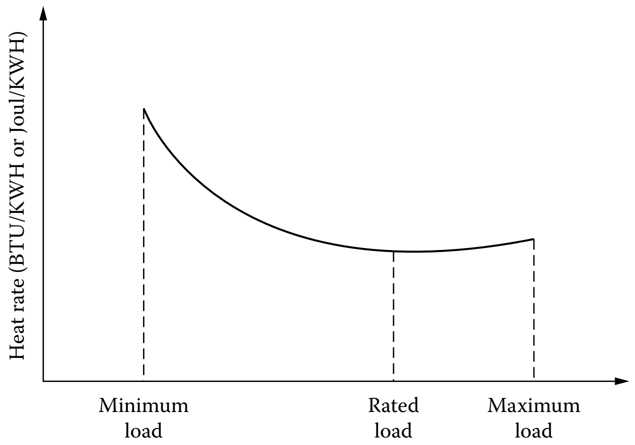


FIG. 8.6nnnn
Heat rate vs. load.

Boiler allocation can be based on actual measured efficiency, on projected efficiency based on past performance, or on some combination of the two. The continuous updating and storing of performance data for each boiler is also a valuable tool in operational diagnostics and maintenance.

The load allocation strategy described above is sometimes based upon heat rate curve (particularly for electric utility units). Most plants' heat rate characteristics are nonlinear, in that they have a high value at the low load and are flat out at the high end (Figure 8.6nnnn). Usually a unit's heat rate curve can be approximated by a polynomial function of the load. The incremental heat rate curve can be obtained by taking derivative of the heat rate curve with respect to the load.

Whether the load allocation is based on the incremental steam cost curve or the incremental heat rate curve, the simple method discussed so far suffers from the following deficiencies:

1. The incremental cost curves do not take operating constraints into account. Especially in the past decade, emission constraints are imposed on almost every power plant. Pollution control credits and penalties should all be taken into account when the optimal load allocation is considered.
2. Methods simply based on the change of incremental cost curve do not usually work well with nonsmooth and nonconvex cost functions, as is often the case for plants that have combined cycle units. The overall heat rate curve for multiple combustion turbogenerators is discontinuous.[33]

To overcome the disadvantage of incremental cost curve-based approach, a relatively complete, yet still simplified, economic load allocation method can be formulated as the following optimization (We assume there are total number of N units available for generation.):

$$\text{Minimize} \quad J = \sum_{i=1}^{N} (F_i + H_i - C_i) \qquad \mathbf{8.6(10)}$$

subject to constraints:

$$\sum_{i=1}^{N} L_i = L_{\text{total}} \qquad \text{(Total load constraint)}$$

$$L_{i,\min} \le L_i \le L_{i,\max} \qquad \text{(Single load constraint)}$$

$$E_i \le E_{i,\text{limit}} \qquad \text{(Single unit emission constraint)}$$

where

L_i = the ith unit load (the decision variable)
$L_{i,\min}$, $L_{i,\max}$ = the low and high limits for the ith unit
L_{total} = the total load demand
E_i = the ith unit emission ($E_{i,\text{limit}}$ is the corresponding limit)
F_i = the ith unit fuel cost
H_i = the ith unit emission control cost
C_i = the ith unit emission credit

ED: Superscript OK?

The fuel cost for each unit is a function of its load level, heat rate, and fuel price. The emission output is also a function of the load. The emission credit C_i can be the result from the emission credit trading market. A negative C_i would indicate penalty. This constrained optimization is nonlinear in general and can be directly solved by a state-of-the-art nonlinear programming algorithm. Alternatively, this problem can be tackled by a standard linear programming approach if nonlinear models are piecewise linearized first.

Note that many electric utility units have implemented automatic generation control (AGC). In AGC mode, raised or lowered pulses of varying lengths are transmitted to the unit from a central location. The control logic changes the unit's load set point up or down in proportion to the pulse length. If the optimal load allocation program is not integrated with the AGC program, then the unit under AGC mode has to be tuned out from the load allocation program.

Soot Blowing Optimization

The impact of soot deposit and soot blowing on boiler performance is complicated (Figure 8.6oooo). The complication is not only due to the obvious fact that the fouling reduces the heat transfer, but also because the fouling changed the heat distribution pattern along the flue-gas path. Besides the heat-transfer pattern, soot blowing can also affect steam temperature, thermal NO_x emission, and stack opacity. For example,[34] one study shows that soot-blowing impact on NO_x can be as high as 6%, change in steam temperature can be as much as 40°F (22°C), and change in the heat rate can be up to 110 BTU/kWh (116 kJ/kWh).

Efficient removal of fireside soot deposit has long been a challenging task. Frequent operation of soot blowers wastes

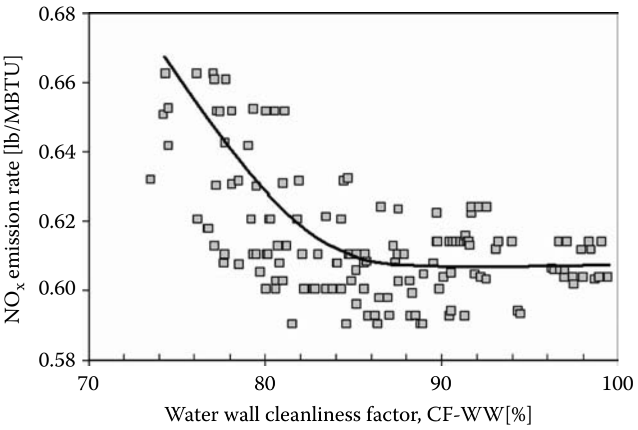


FIG. 8.6oooo
Effect of water wall cleanliness on NO_x *emissions in a coal-fired boiler. (Courtesy of Energy Research Center, Lehigh University.)*

steam, increases blower maintenance cost, and aggravates the tube erosion. Conversely, far less frequent blowing allows too much soot accumulation and, hence, decreases efficiency. It may also cause high stack opacity when a heavily fouling area gets blown. Therefore, intelligent adjustment of the cleaning schedule according to the actual cleaning need becomes the means of achieving our primary goal: efficiency improvement. This is realized through advanced control software involving a cleanliness factor calculation and a rule-based expert system.

Cleanliness Factor Calculation Boiler section fouling status can be quantified by the section cleanliness factor (CF). By usual definition, the heat-transfer effectiveness ε is the ratio of actual to design heat-transfer rate, i.e.,

$$\varepsilon = \frac{Q_{\text{actual}}}{Q_{\text{ideal}}} \qquad \mathbf{8.6(11)}$$

where Q_{actual} and Q_{ideal} are the actual and ideal section heat absorption rate (BTU/hr or kJ/hr), respectively.

Then, the cleanliness factor is defined as the ratio of the actual effectiveness vs. the baseline effectiveness.

$$\text{CF} = \frac{\varepsilon_{\text{actual}}}{\varepsilon_{\text{baseline}}} \qquad \mathbf{8.6(12)}$$

The baseline effectiveness is determined by design and can usually be calibrated in the field. Because the baseline heat-transfer effectiveness is most likely assumed to be a constant in practice, the cleanliness factor CF can be conveniently represented by the $\varepsilon_{\text{actual}}$.

A conventional method of calculating the heat absorption rate is the log-mean-temperature-difference approach, which requires steam (water) and flue-gas temperature measurements at each heat-transfer section inlet and outlet. The formula is

$$Q = \mu A T_{lm} \qquad \mathbf{8.6(13)}$$

and

$$T_{lm} = \frac{(T_{g_i} - T_{s_o}) - (T_{g_o} - T_{s_i})}{\log((T_{g_i} - T_{s_o}) / (T_{g_o} - T_{s_i}))} \qquad \mathbf{8.6(14)}$$

where

μ = surface heat-transfer coefficient (same as ε)
A = heat exchange section area
T_{g_i} = flue-gas temperature measured at section flue-gas inlet
T_{g_o} = flue-gas temperature measured at section flue-gas outlet
T_{s_i} = steam temperature at section steam inlet
T_{s_o} = steam temperature at section steam outlet

The steam/water temperature can be measured at many places along the boiler heat-transfer path. On the other hand, flue-gas temperature measurements are usually only available around air heater inlet and outlet. At all other locations, flue-gas temperatures have to be obtained by backward computations according to a system of energy balance equations.

Another method for calculating heat absorption and cleanliness factor is empirical model based.[35] Given the steam flow rate, temperature, and pressure at each section inlet and outlet, the actual heat absorption can always be calculated as

$$Q = F_s \cdot (H_o - H_i) \qquad \mathbf{8.6(15)}$$

where

F_s = steam flow rate (lb/hr)
H_i = steam enthalpy at section inlet (BTU/lb)
H_o = steam enthalpy at section outlet (BTU/lb)

Enthalpy H, as a function of steam temperature and pressure, can be determined from the standard ASME steam tables.

In order to model the ideal heat absorption for a clean section, fire-side influence (i.e., the flue-gas temperature) needs to be identified. The idea is that the steam temperature at section outlet is largely decided by section inlet flue-gas temperature and section fouling status. So, in the ideal clean section situation, if all required measurements are available, the calculation of ideal heat transfer can be modeled as

$$Q_{\text{ideal}} = f(F_s, T_{s_i}, G_i) \qquad \mathbf{8.6(16)}$$

where G_i generically represents all variables that have influence on the section inlet flue-gas temperature.

This empirical method relies on the system's ability of identifying all fire-side influencing variables and correctly interpreting the acquired data that represents a clean boiler section. A typical boiler will have cleanliness calculation for

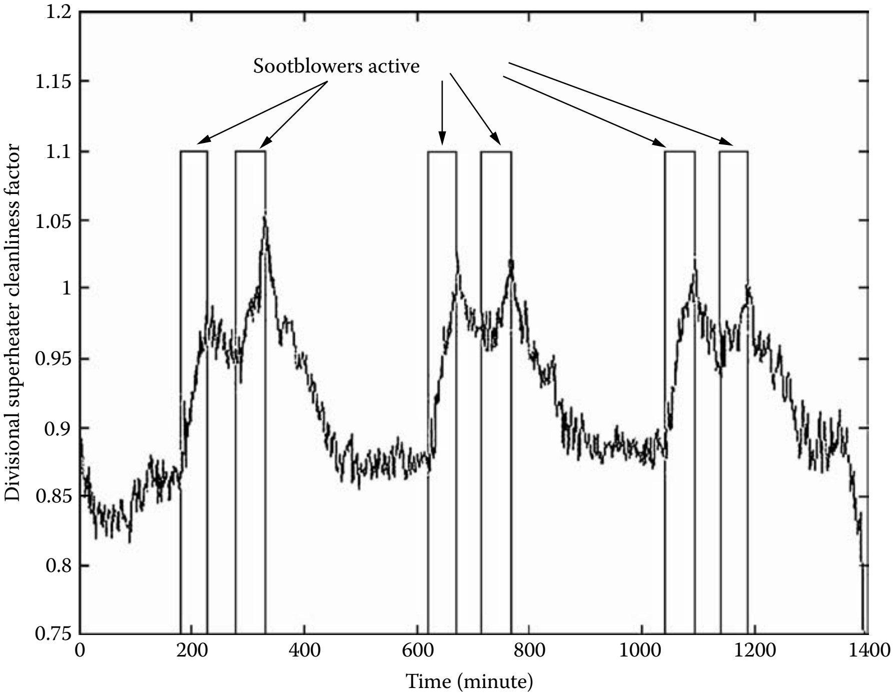


FIG. 8.6pppp
Trend recording of the cleanliness factor.

the furnace wall section, the economizer section, the air heater section, and each of the superheater and reheater sections.

Figure 8.6pppp shows an example of the cleanliness factor calculation results. Also worth mentioning is that, in order to directly measure heat absorption rate for the furnace wall, in situ heat flux sensors can be installed in the wall area. However, the cost is high for purchasing, installation, and maintenance.

Rule-Based Expert System In addition to cleanliness factor considerations, expert systems also play a key role. Because operators are most familiar with the daily operation, their experience in detecting and handling different fouling scenarios is important and, therefore, should be incorporated into the rule base.

Expert rules can be implemented from the following perspectives.

- Determine the desired cleanliness factor for each boiler section. The soot-blowing decision can be made based on the difference between the desired and actual cleanliness factors.
- At low loads, the fouling is built up slowly, and hence longer idle time between running sequences is expected. Therefore, allowing the furnace wall to have relatively low cleanliness factor will leave more heat for the following convection sections. This also increases the opportunity to blow convection sections more, and should raise steam temperatures without having to lift the firing rate.
- At high loads, the fouling is built up rapidly, and hence shorter idle time between running sequences is expected. In this situation, superheat temperature tends to run too hot, requiring attemperating spray water to prevent overheating. Therefore, convection sections should be allowed to have relatively low cleanliness factors and the furnace wall section should be cleaned more often.
- In order to limit stack opacity, the fuel/air ratio, operating status of precipitator, and fuel burner will be frequently checked, and the result will be taken into account by the rule base.
- Regardless of cleanliness factors, there should be a minimum idle time between runs for each blower sequence.
- Regardless of cleanliness factors, each blower sequence will have a maximum allowed off-time so that slag will not be heavily accumulated in one area.

Model-Based Boiler Optimization

Once surrounded by skepticism, model- and computer software-based boiler optimization schemes have now been applied and proven successful in many utility and industrial boiler applications. Optimization typically involves O_2 and CO, and is targeted at efficiency improvement and, often, NO_x reduction as well. NO_x reduction of 10–30%, and heat rate

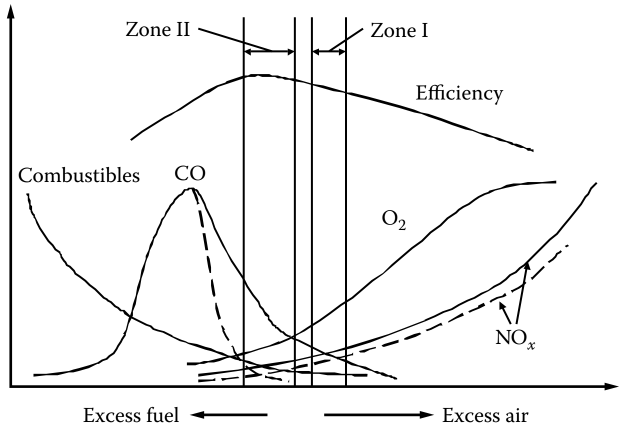


FIG. 8.6qqqq
Zones of optimum boiler performance.

improvement of 1–2% are possible.[36,37] Successful NO_x reduction through this kind of optimization can avoid or postpone large capital expenditures for low NO_x burners, over-fire air modifications, and SCR/SNCR.

Depending on the type and size, a fossil-fired generating unit may contain as many as hundreds of highly correlated, nonlinearly related, and time-varying variables. The Zone I shown in Figure 8.6qqqq is the optimal region for a traditional excess air-based efficiency optimization. However, the curve in the figure is usually obtained from design, or from simple experimental characterization at specific load levels for a specific control system setting.

The basis for boiler combustion optimization lies in identifying the relationship of important fuel/air parameters affecting the mixture and distribution of fuel and air within the firebox. By finding a better combination of all variables that affect fuel and air distribution, the NO_x curve in Figure 8.6qqqq can be shifted to the right and the CO curve can be shifted to the left. This would open up more room for improved optimization compared to strategies that rely only on excess air (O_2 trim) control. Advanced model-based control techniques can result in reduced O_2, CO, and NO_x, while keeping the boiler operated at a stable low oxygen zone (Zone II in Figure 8.6qqqq). Conditions leading to improved efficiency generally also result in better control and emissions performance.

Modeling The models are a set of equations that relate key performance measurements to the major influencing combustion control variables. Due to the complexity and uncertainty of the analytical models that are derived from physical principles, empirical models based entirely upon the plant data are typically used for practical boiler optimization control. Modeling starts with identifying potential controlled variables (CVs), manipulated variables (MVs), and disturbance variables (DVs).

The controlled variables are those performance variables that we would like to drive to the most cost-effective and regulatory compliant region. The manipulated variables are those that affect the CVs and can be directly adjusted by the control strategy. In the boiler optimization context, they are various supervisory set points (or their biases). The disturbance variables are the variables that affect one or more performance variables. They are measurable but cannot be adjusted by the control strategy directly. In the boiler case, it is usually the load (or steam flow) change. Depending on boiler size and configurations, the following controlled and manipulated variables are normally selected for the initial experiment.

Potential CVs are CO, NO_x, O_2, and boiler efficiency/heat rate.

Potential MVs are feeder speed bias, mill exit primary air temperature bias, O_2 trim bias, FD/ID fan bias, furnace pressure, auxiliary air damper bias, fuel flow bias, flue-gas damper bias, burner tilt bias, and secondary over fire air (SoFA) damper bias.

Step tests are normally carried out at different load levels, e.g., high, medium, and low load. Although it is desirable to exercise most parameters to values beyond those encountered in normal operation, due to operating constraints in most plants, the magnitude of the test signal is normally selected at 5–10% of the overall operating range. Correlation analysis can be performed to sort through the large number of variables involved and to identify the variables that have significant impact on the performance.

Modeled relationships can take the form of step response, impulse response, state-space representation, or a neural network (a direct nonlinear form). If a linear form is assumed, then the model is linearized about some operating point, or a series of linear models is produced; each represents a specific operating condition (usually load level). The obtained model can be used for solving a static optimization problem to find out the optimal operating point. The "optimal" criterion can be user selectable. For example, the selection can be minimizing NO_x, minimizing heat rate, or a combination of both. The model can also be used for carrying out closed-loop control, i.e., to use the identified MVs and drive the CVs to the region of optimal performance.

Closed-Loop Control and Optimization Closed-loop multivariable boiler control has to be planned and performed carefully, because plant operators are not traditionally willing to reduce air/fuel ratios due to concerns about CO and other symptoms associated with oxygen-deficient combustion. Model predictive control (MPC) is by far the most widely used technique for conducting multivariable boiler optimization and control. Forms of MPC that are inherently multivariable and that include real-time constrained optimization in the design are best suited for boiler application.

For example, when NO_x and CO are selected as CVs, in a constrained optimization they do not have to be controlled to any specific set point as long as they are all held below specified limits. NO_x limit can be specified either by regulation or by the plant operator. The CO limitation can be specified by the operator as a performance constraint.

Fuel quality, boiler loading, heat exchanger surface fouling, ambient condition, and aging of equipment will all cause process to drift and affect the model accuracy. Adaptive tuning computations can be built in to take care of known quantifiable relations (e.g., variation of dead time with load). Online training, incorporating an adaptive learning algorithm, can be used to automatically train models in real time, combining recent results with the initial and historical training data.

Certain results from the optimization calculation may be very intuitive. For example, it might call for reducing the top elevation mill coal flow whenever feasible, or removing the top elevation mill from service whenever the load can be sustained with the lower level mills.[29] This coincides with our intuition that reducing the fuel input for the upper level mill will result in a lower fire-ball position, and effectively lower the upper furnace flame temperature, which in turn reduces the thermal NO_x formation.

CONCLUSIONS

The various goals of boiler optimization include the following.

- To minimize excess air and flue-gas temperature
- To measure efficiency (use the most efficient boilers; know when to perform maintenance)
- To minimize steam pressure (open up turbine governors; reduce feed pump discharge pressures; and reduce heat loss through pipe walls)
- To minimize blowdown
- To provide accountability (monitor losses; recover condensate heat)
- To minimize transportation costs (use variable-speed fans; eliminate condensate pumps; and consider variable speed feedwater pumps)

If the potentials of all of the above optimization strategies are fully exploited, the unit costs of steam generation can usually be lowered by about 10%. In larger boiler houses, this can represent a savings that will pay for the optimization system in a year or less.[38]

ACKNOWLEDGMENTS

The co-authors wish to express their appreciation to Dave Fatkin, Plant Engineer, A.B. Hopkins Power Plant, City of Tallahassee Electric Utility, Tallahassee, FL, and John Hlavac, P.E., Plant Controls Engineer, Big Bend Station, Tampa Electric Co., Tampa, FL, for their input and review of the control strategies of electric utility fossil-fuel boilers.

References

1. Congdon, P., "Control Alternatives for Industrial Boilers," *InTech*, December 1981.
2. Walsh, T. J., "Controlling Boiler Efficiency," *Instruments and Control Systems,* January 1981.
3. ANSI/ASME, "PTC 4.1, Steam Generating Units Power Test Codes," American Society of Mechanical Engineers (ASME), 1965.
4. Culp, A. W., Jr., *Principles of Energy Conversion*, New York: McGraw-Hill, 1979, p. 102, pp. 204–207.
5. MacDonald, J. A., "Optimizing Power Boiler Efficiency Calls for Heat Loss Cuts, System Insulation, and Modifications," *Energy-Tech*, April 2004.
6. Shinskey, F. G., *Energy Conservation through Control,* Academic Press, 1978.
7. Babcock and Wilcox, *Steam: Its Generation and Use,* 40th ed., New York: Babcock and Wilcox, 1992.
8. NFPA, *NFPA-85—Boiler Combustion Safety*, National Fire Protection Association, 2004.
9. Garton, D., "Water Cannon for Water Wall Cleaning Applications," *EPRI Intelligent Soot Blowing Conference*, December 6–7, 1999.
10. McMahon, J. F., president, Cleveland Controls, verbal communications.
11. Dickey, P. S., "A Study of Damper Characteristics," Bailey Meter Co., Reprint No. A8.
12. Jorgensen, R., "Fans," in *Marks Standard Handbook for Mechanical Engineers*, 8th Ed., McGraw-Hill, New York, 1978, p. 14.53.
13. Dukelow, S. G., *The Control of Boilers,* 2nd ed., Research Triangle Park, NC: Instrumentation, Systems, and Automation Society, 1991.
14. Hurlbert, A. W., "Air Flow Characterization Improves Boiler Efficiency," *InTech,* March 1978.
15. DeLorenzi, O., Combustion Engineering, Inc., 1967.
16. ANSI/ISA-77.42.01, "Fossil Fuel Power Plant Feedwater Control System: Drum Type," Instrumentation, Systems, and Automation Society, 1999.
17. Shinskey, F. G., "Taming the Shrink–Swell Dragon," *Control*, March 2004.
18. Fisher Controls International, *Control Valve Sourcebook: Power and Severe Service*, Marshalltown, IA: Fisher Controls International, 1990.
19. Miller, H. L. and Stratton, L. R., "Fluid Kinetic Energy as a Selection Criteria for Control Valves," American Society of Mechanical Engineers, Fluids Engineering Division, summer meeting, June 22–26, 1997.
20. Shinskey, F. G., *Process Control Systems,* New York: McGraw-Hill, 1979.
21. Latta, C. A., "Methods for Reducing NO_x Emissions," *Plant Engineering*, September 1998.
22. McDonald, J. A., "Controlling the NO_x after the Burn," *Energy-Tech*, December 2002.
23. Schwartz, J. R., "Carbon Monoxide Monitoring," *InTech*, June 1983.
24. O'Meara, J. E., "Oxygen Trim for Combustion Control," *InTech*, March 1979.
25. McFadden, R. W., "Multiparameter Trim in Combustion Control," *InTech,* May 1984.
26. American Technical Services, "Boiler Audits," June 1982.
27. McMahon, J. F., President, Cleveland Controls, verbal communications.
28. Westinghouse Electric Corp., "Oxygen Trim Control," *AD-106-125,* Westinghouse Electric Corp., June 1979, January 1985.
29. Moelback, T., "Advanced Control Superheater Steam Temperatures: An Evaluation Based on Practical Applications," *Control Engineering Practice*, Vol. 7, 1999, pp. 1–10.
30. Dukelow, S. G., *The Control of Boilers,* 2nd ed., Research Triangle Park, NC: ISA, 1991.
31. Schieber, J. R., "The Case for Automated Boiler Blowdown," Universal Interlock, 1969.

32. Cho, C. H., "Optimum Boiler Allocation," *InTech,* October 1978.
33. Wood, A. J., and Wollenberg, B. F., *Power Generation, Operation, and Control*, 2nd ed., New York: Wiley Interscience, 1996.
34. Romero, C., Sarunac, N., and Levy, E., "Soot Blowing Optimization in Coal-Fired Boilers," Lehigh University Energy Research Center Energy Liaison Program Annual Meeting, 2001.
35. Cheng, X., Kephart, R. W., and Williams, J. J., "Intelligent Soot Blower Scheduling for Improved Boiler Operation," *Proceedings of ISA POWID/EPRI Instrumentation and Control Conference*, St. Petersburg, FL, June 1999.
36. Booth, R. C., and Roland, W. B., "Neural Network-Based Combustion Optimization Reduces NO_x Emissions while Improving Performance," *Proceedings of 1998 ASME International Joint Power Generation Conference*, Baltimore, MD, 1998.
37. Lefebvre, C., Lynch, M., and Roland, R., "Application of Process-Link Real-Time Optimization System to Cajun Electric Power Cooperative's Big Cajun II Generating Station," *Proceedings, 1999 International Power-Gen Conference*, New Orleans, LA, 1999.
38. Lipták, B. G., "Save Energy by Optimizing Your Boilers," *InTech,* March 1981.

Bibliography

ANSI/ISA-77.41, "Fossil Fuel Power Plant Boiler Combustion Controls," Instrumentation, Systems, and Automation Society, 1992.

ANSI/ISA-77.43.01, "Fossil Fuel Power Plant Unit/Plant Demand Development — Drum Type," Instrumentation, Systems, and Automation Society, 1994.

ANSI/ISA-77.44.01, "Fossil Fuel Power Plant Steam Temperature Control System — Drum Type," Instrumentation, Systems, and Automation Society, 2000.

ASME Research Committee on Industrial and Municipal Wastes, "Combustion Fundamentals for Waste Incineration," New York: American Society of Mechanical Engineers, 1974.

Babcock and Wilcox, *Steam: Its Generation and Use,* 40th ed., New York: Babcock and Wilcox, 1992.

Bell, A. W., and Breen, B. P., "Converting Gas Boilers to Oil and Coal," *Chemical Engineering,* April 26, 1976.

"Boiler Controls," *Power*, October 1972, pp. 21–28.

Buntzel, H., "Calibrating Boiler Control Systems," *InTech,* January 1974.

Champigny, R. A., "Retrofitting Turbine and Boiler Controls," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Cho, C. H., "Optimum Boiler Load Allocation," *InTech,* October 1978.

Claypoole, G. T., "The Conceptual Design of a Utility," Paper #76-560, ISA 1976 Annual Conference, Houston, 1976.

Council of Industrial Boiler Owners (CIBO), *Energy Efficiency Handbook*, Council of Industrial Boiler Owners, 1997.

Criswell, R. L., "Control Strategies for Fluidized Bed Combustion," *InTech,* January 1980.

Davis, T. P., "Control Strategies for Fuel Oil Heating," *InTech,* September 1977.

Dukelow, S. G., "Charting Improved Boiler Efficiency," *Factory,* April 1974, pp. 31–34.

Dukelow, S. G., "Combustion Controls Save Money," *Instruments and Control Systems,* November 1977.

Dukelow, S. G., "Control of Steam Boilers," in *Instrumentation in the Processing Industries*, 1973, Chapter 10.6, pp. 632–667.

Environmental Protection Agency (EPA), "Controlled Flare Cuts Smoke, Noise, Saves Steam," *Technology Transfer,* Washington, D.C.: U.S. EPA,

Environmental Protection Agency, Rep. No. 650/2-74-066, Washington, D.C.: U.S. EPA.

Flynn, D., *Thermal Power Plant Simulation and Control*, London: Institution of Electrical Engineers, 2003.

The Foxboro Company, "Differential Vapor Pressure Cell Transmitter, Model 13Va," Tech. Inform. Sheet 37–91a, April 15, 1965.

Gilbert, L. F., "Precise Combustion Control Saves Fuel and Power," *Chemical Engineering,* June 21, 1976.

"Guide to Selecting Combustion Stack Analyzers," *InTech,* April 1982.

Hockenbury, W. D., "Improve Steam Boiler Control," *Instruments and Control Systems,* July 1981.

Johnson, A. J., and Auth, G. H., *Fuels and Combustion Handbook,* New York: McGraw-Hill, 1951, p. 355.

Kaya, A., "Optimum Soot Blowing Can Improve Your Boiler Efficiency," *Instrumentation Technology,* January 1982.

Keller, R. T., "New Boiler Control System," *Instruments and Control Systems,* June 1976.

Kolmer. J. W., "Programmed O_2 Optimizes Combustion Control," *Instruments and Control Systems,* December 1977.

Krigman, A., "Combustion Control," *InTech,* April 1982.

Kuhlman, L. J., "Waste Firing Control in Industrial Boilers," *Instrumentation Technology,* February 1976.

Lawrence Berkeley National Laboratory, Resource Dynamics Corp., *Improving Steam System Performance — A Sourcebook for Industry*, U.S. Department of Energy Office of Industrial Technology.

Louis, J. R., "Control for a Once-Through Boiler," *Instruments and Control Systems,* January 1971.

Matthys, W. J., "Retrofitting Industrial Boilers to Meet the Fuels Market," paper presented at American Power Conference, Chicago, April 21–23, 1975.

May, D. L., "Cutting Boiler Fuel Costs with Combustion Controls," *Chemical Engineering,* December 22, 1975.

"Microprocessors that Run Boilers," *Business Week,* August 13, 1979.

Mockenbury, W. D., "Improve Steam Boiler Control Step by Step," *Instruments and Control Systems,* July 1981.

Nathan, P. and Wasserman, R., "Energy Control Cuts Electric Bills," *Instrumentation Technology,* December 1975.

Neuberger, E. D., "Programmed O_2 Set Point Optimizes Combustion Control," *Instruments and Control Systems,* December 1977.

Nordmann, B., "Compensate Combustion Controls for BTU Variations," *Instruments and Control Systems,* April 1978.

Pavinski, D., Romero, C., Levy, E., Cramer, D., Maines, P., Williams, S., and Petrill, E., "An Approach to Boiler Soot Blowing for Optimizing NO_x and Heat Rate," *Proceedings of the 6th Joint ISA/EPRI Controls and Instrumentation Conference*, Baltimore, 1996.

Payne, F. W., and Thompson, R. E., *Efficient Boiler Operations Sourcebook,* 4th ed., Fairmont Press, Inc., 1996.

Perry, D., "Upgrade Boiler Controls with DDC," *Power,* October 1975.

Pocock, R. E., Ross, D. F., and McGandy, E. L., "Oxygen Analyzers for the New Proposed EPA Pollution Stack Monitoring Regulations," 68th Annual Meeting of the Air Pollution Control Association, June 1975.

Pollard, E. V., and Drewry, K. A., "Estimating Performance of Automatic Extraction Turbines," West Lynn, MA, General Electric Co.

Robnett, J. D., "Energy Optimization," *InTech*, June 1983.

Rodman, C. W., "Monitoring Combustion Chamber Pulsations," *InTech*, April 1982.

Roth, J. W., "Development of a Multi-Element Boiler Combustion Control System," Southern California Meter Association, October 1961.

Scheib, T. J., "Instrumentation Cuts Boiler Fuel Costs," *Instruments and Control Systems,* November 1981.

Shinskey, F. G., "Adaptive Control? Consider the Alternatives," *Instruments and Control Systems,* August 1974.

Shinskey, F. G., *pH and pIon Control in Process and Waste Streams,* New York: Wiley (Interscience), 1973, pp. 88–92.

Shinskey, F. G., *Process Control Systems,* New York: McGraw-Hill, 1967, pp. 188–190.

Shinskey, F. G., "Process Control Systems with Variable Structure," *Control Engineering,* August 1974.

Shinskey, F. G., "Two-Capacity Variable Period Processes," *Instruments and Control Systems,* October 1972.

Singer, J. G., *Combustion Fossil Power*, Combustion Engineering, Inc., 1991.

Tomlinson, L. O., and Snyder, R. W., "Optimization of STAG Combined Cycle Plants," presented at the American Power Conference, Chicago, April 29–May 1, 1974.

Trimm, T. G., "Perform and Energy Audit on Your Boilers," *InTech,* July 1982.

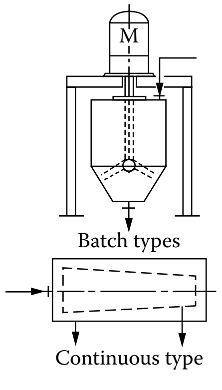
Wells, R. D., "Computer Cuts Steam Cost by $1.3 Million," *InTech,* January 1983.

Yarway Steam Conditioning Valve Manual, Blue Bell, PA: Yarway Corp.

Zimmer, H., "Chiller Control Using On-Line Allocation for Energy Conservation," Paper No. 76-522, Advances in Instrumentation, Part 1, ISA/76 International Conference, Houston, 1976.

8.7 Centrifuge Controls

C. J. SANTHANAM (1970, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

This section reviews some of the basic centrifuge control systems used on both sedimentation and filtering centrifuges. The sensors used in measuring the product quality are also discussed.

CENTRIFUGE TYPES

Centrifuges can be used for both liquid–liquid and liquid–solid separation. The centrifuge designs can be classified into two main groups: sedimentation and filtering centrifuges.

Sedimentation centrifuges: These designs have solid walls, and separation occurs by sedimentation. This process is illustrated in the top portion of Figure 8.7a, where the feed enters a solid-walled bowl, which is rotating about a vertical axis. Centrifugal force and gravity both act upon the solid and liquid phases. The dominant force is the centrifugal one, which moves the heavier liquid phase or the heavier solid particles to the perimeter and thereby the separation of solid and liquid phases (or heavy and light liquid phases) takes place as shown.

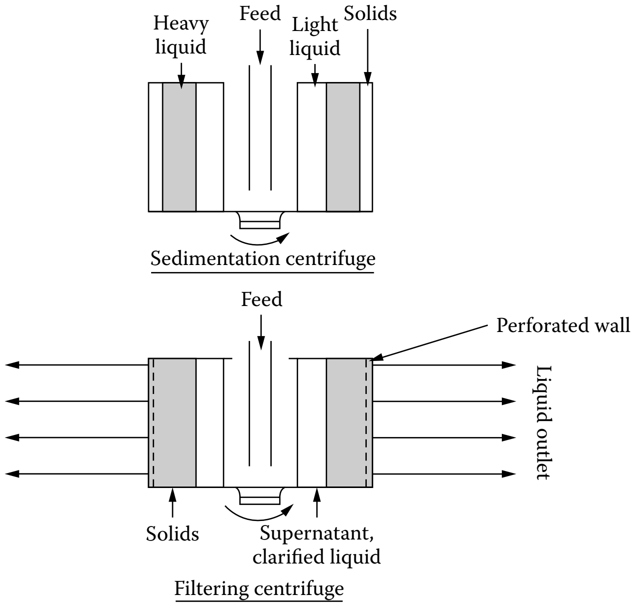


FIG. 8.7a
The operating principles of the sedimentation (top) and filtering (bottom) centrifuge designs.

Filtering centrifuges: These machines have perforated walls that retain the solids on a permeable surface through which the liquid can escape. This design is shown in the bottom portion of Figure 8.7a. The operation that takes place is similar to that of a filter but with a much higher "g force" than what can be obtained in gravity or pressure filtration. In these filtering centrifuges, nearly all the liquid is removed, leaving behind an almost dry cake.

The centrifugal force obtained in industrial machines is several times the force of gravity. If a particle is rotating with an angular velocity w and is located at a distance r (radius) from the axis of rotation, the centrifugal separating effect or g force is:

$$G = \frac{wr}{g} \qquad \textbf{8.7(1)}$$

Filtering centrifuges operate at a g force range of 400 to 1,800, whereas the g force in sedimentation units ranges from 3,000 to over 60,000 in laboratory machines (ultracentrifuges).

The critical speed phenomenon must be considered in the design and operation of centrifuges, as with any high-speed machine. At critical speed, the frequency of rotation matches natural frequency of the rotating member. At this speed, even the minute vibrations that can be induced by slight imbalances are drastically reinforced. Centrifuges pass through the critical speed during acceleration and deceleration, because their normal operating speeds much exceed the

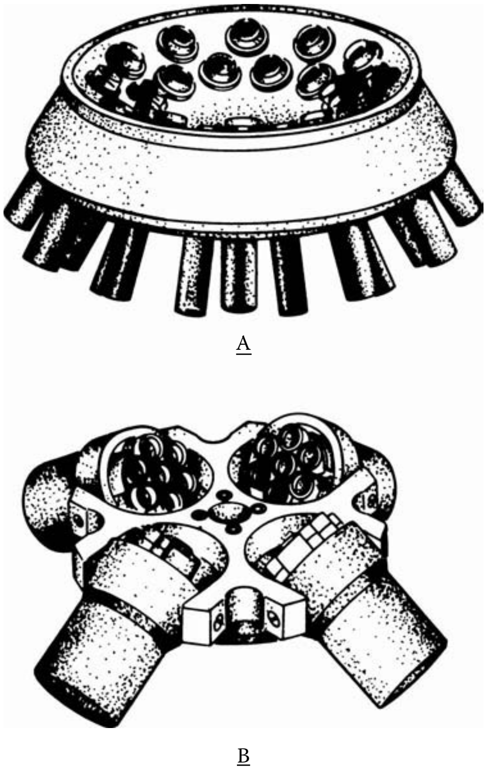


FIG. 8.7b
Tube holders in laboratory centrifuges can be of the fixed-angle (A) or the swing-out (B) design. (Courtesy of Cole-Parmer Instrument Co.)

critical. Therefore, except in case of major bowl imbalance, this does not cause problems.

Laboratory Centrifuges

Laboratory centrifuges are usually compact benchtop units. The samples that are to be centrifuged are usually placed in sample tubes, which are inserted into either fixed-angle or swing-out tube holders (Figure 8.7b). The sample sizes that the individual sample tubes can hold usually range from 1.5 to 15 ml, although special tubes are available for samples from 0.5 to 100 ml. The rotor assemblies can usually hold from 1 to 24 sample tubes. The speed of centrifuging can be fixed at around 3,500 rpm, or it can be variable, usually up to 14,000 rpm.

The maximum rotational centrifugal force (RCF) that can be obtained varies with the dimensions of the machine and with its rpm, from about 1,000 g to about 16,000 g. Some of the more sophisticated units can be microprocessor-controlled and provided with refrigeration and temperature controls. A simple fixed tube unit with 3,600 rpm and RCF equal to 1,400 g can be obtained for about $1,000; nonrefrigerated universal units cost about $2,500, while the refrigerated, microprocessor-controlled ones cost about $10,000.

INDUSTRIAL CENTRIFUGES

In selecting a centrifuge, several factors need to be considered, including (1) the ability of the machine to process the given feed slurry or emulsion to provide the desired degree of separation, (2) the reliability of the machine, (3) the operating and maintenance requirements, and (4) the investment.

Table 8.7c provides an overall orientation among the types of centrifuges, their sizes, *g* forces, capacities, and methods of solids and liquid discharging.

In the majority of cases, equipment manufacturers have standard machines that are adapted to the applications specified by the customer. Sedimentation machines are usually chosen on the basis of small-scale tests in laboratory centrifuges. Filtering centrifuges are chosen on the basis of tests in batch machines. Based on such tests, the manufacturers will offer specific machines and will outline the anticipated performance.

One important consideration in the selection among centrifuge designs is the size of the particles that are present in the feed slurry and are to be separated from it (Figure 8.7d).

Another important consideration in the centrifuge selection is the percentage of solids that are present in the feed slurry (Figure 8.7e).

The process control engineer should concentrate on two main control systems when working on a process centrifuge installations:

Feed slurry control: Regulation of the feed slurry at the correct continuous flow rate or in the right batch sizes. This is very important, because the machine cannot usually tolerate major variations in feed rate or feed composition. The accurate control of the wash liquor feed flow rate is equally important.

Sequencing operations: All batch machines are sequentially operated, and the related interlock design is one of the important steps in engineering a system. Figure 8.7f illustrates the sequencing of the program flow in a semiautomatic centrifuge operation.

In the semiautomatic system, the program can be run manually, one module at a time. The individual sequences can be started manually or by a higher level overall control system. This type of configuration is particularly suited for control systems where the sequencing program can be preset for making several products. Figure 8.7g illustrates the sequencing of the operation of an automatically operated centrifuge.

Sedimentation Centrifuges

Sedimentation units are used as clarifiers, desludgers, and liquid-liquid phase separators. Particle size in this processing is usually such that separation obeys Stokes' law. Sedimentation units are generally of small diameter and run at high speed. These can be classified into the following types: (1) tubular, (2) disk, and (3) solid-bowl. Overall data on present-day machines are presented in Table 8.7c.

TABLE 8.7c
Classification of Centrifuges

Rotor Type	*Range of Bowl Diameter in Inches (mm)*	*Maximum Centrifugal Force (g)*	*Method of Solids Discharge*	*Method of Liquid Discharge*	*Maximum Capacity*	
					gals/hr or ft^3/batch short tons/hr	*(liters/hr or m^3/batch) (short tons/hr)*
I. Sedimentation Centrifuges						
Tubular	2–6 (50–150)	60,000	Manual (batch)	Continuous	3,000 g/hr	(11,250 l/hr)
Disk	9–32 (230–800)	2,500–8,000	Batch or semi continuous	Continuous	12,000–24,000	(45 to 90,000 l/hr)
Solid bowl						
Constant-speed (horizontal)	14–36 (350–900)	1,000–3,000	Automatic batch	Continuous overflow	60 ft^3/batch	(1.7 m^3/batch)
Variable-speed (vertical)	12–84 (300–2 m)	Up to 3,200	Automatic batch	Continuous	15 ft^3/batch	(0.42 m^3/batch)
Continuous		Up to 3,200	Continuous	Continuous	Up to 65 tons/hr solids	(59 tons/hr)
II. Filtering Centrifuges						
Conical screen						
Wide-angle		Up to 1,400	Continuous	Continuous	15,000 g/hr	(56,250 l/hr)
Differential scroll		Up to 1,800	Continuous	Continuous	70 tons/hr solids	(63.5 tons/hr)
Vibrating screen		Up to 500	Continuous	Continuous	100 tons/hr solids	(90.7 tons/hr)
Pusher		1,800	Batch	Continuous	10 tons/hr solids	(9.1 tons/hr)
Cylindrical screen						
Pusher		1,500	Continuous	Continuous	40 tons/hr solids	(36.3 tons/hr)
Differential scroll		1,500	Continuous	Continuous	40 tons/hr solids	(36.3 tons/hr)
Horizontal		1,300	Batch	Intermittent	25 tons/hr solids	(22.7 tons/hr)
Vertical		900	Batch	Intermittent	10 tons/hr solids	(9.1 tons/hr)

The *g* concept is often used to compare the performance of solid-wall centrifuges. This equivalence converts the geometry, size, and speed of the bowl to the area of a settling tank theoretically capable of the same separation in a gravity field of unity.

The control strategies used for these centrifuges are different for batch, semicontinuous, and continuous applications.

Batch Centrifuge Control All laboratory centrifuges and also some of the smaller sedimentation machines in industry

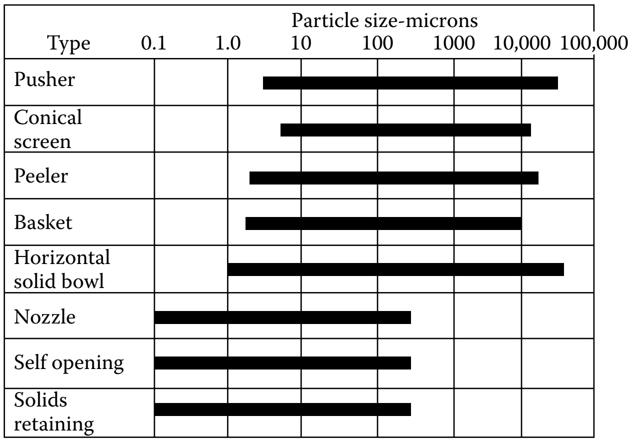


FIG. 8.7d
The relationship between the size of solid particles in the feed slurry and the type of centrifuge design recommended for use.

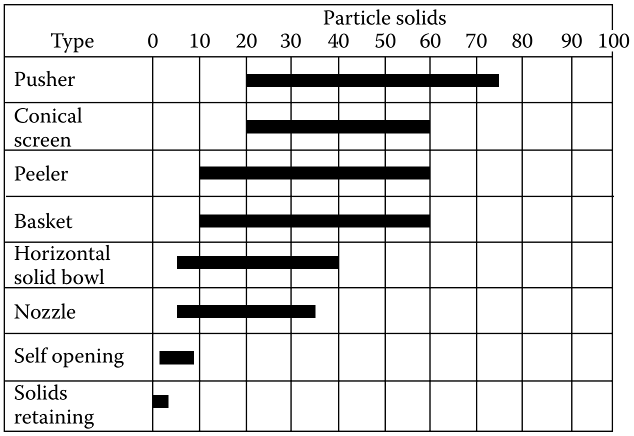


FIG. 8.7e
The relationship between the percentage of solid particles in the feed slurry and the type of centrifuge design recommended for use.

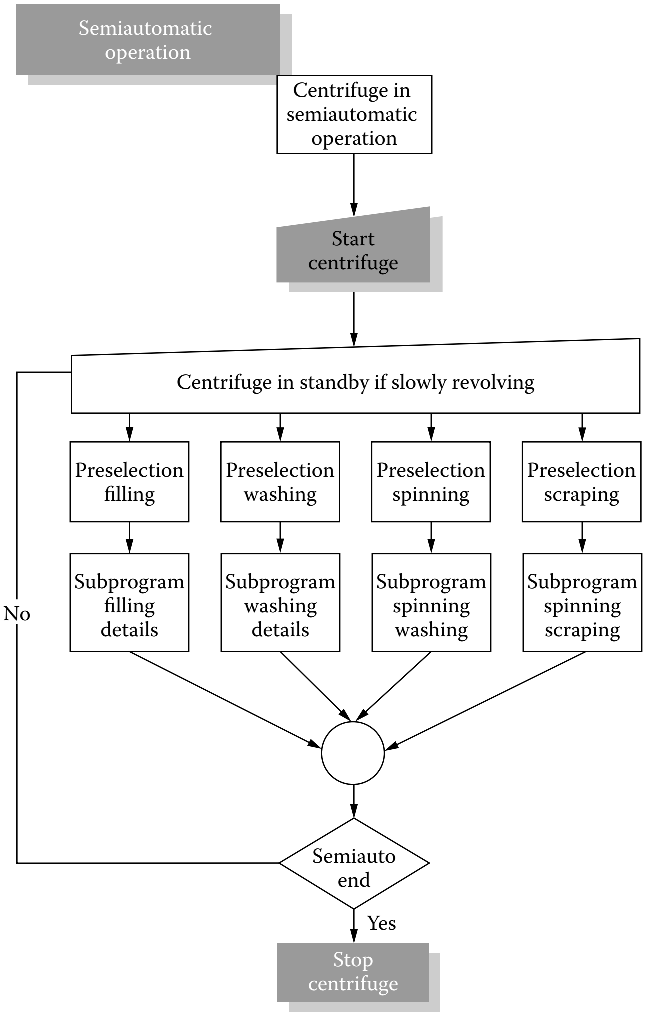


FIG. 8.7f

Sequencing of the operation of a semiautomatic centrifuge. (Courtesy of Ferrum Ltd.)

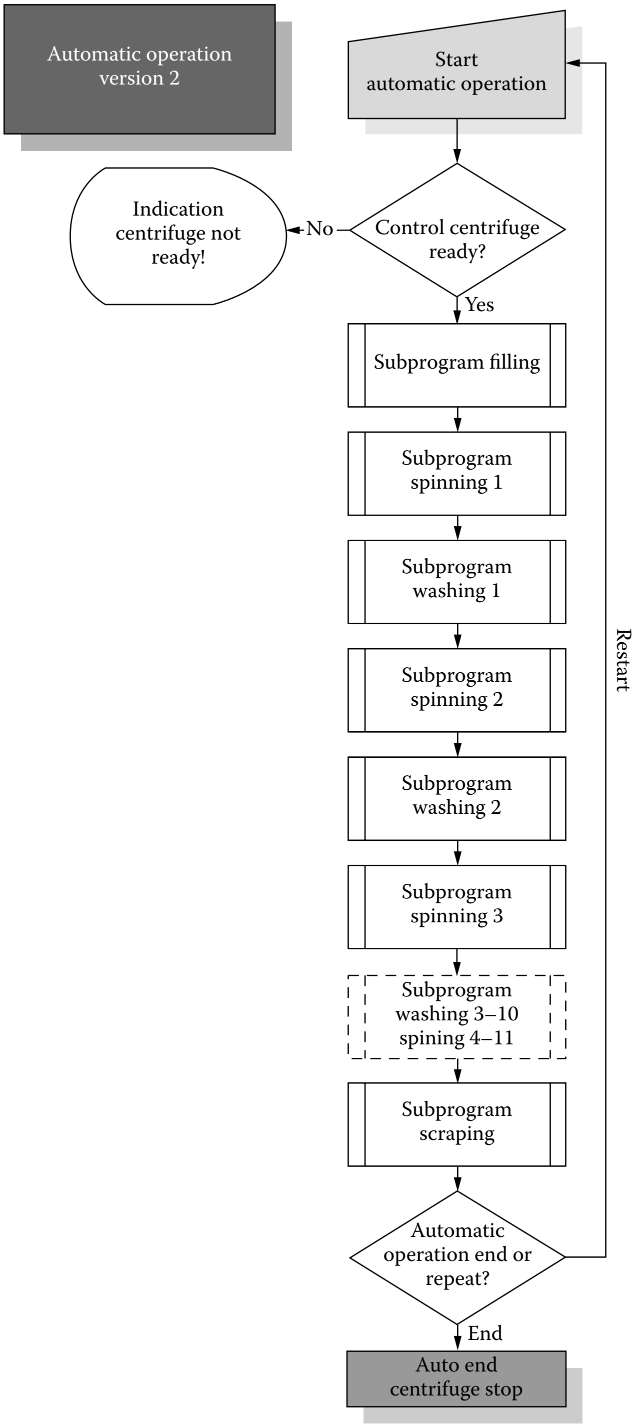


FIG. 8.7g

Sequencing of the operation of an automatic centrifuge. (Courtesy of Ferrum Ltd.)

are operated as batch units with manual control of the feed and of the discharge of solids. In such instances, little automatic control is required, because all sequencing is done manually.

In the category of automatic batch-type centrifuges, one finds the fall disk-type machines, such as the desludging centrifuges, and also some solid-bowl types. A simple control scheme for a desludging, or a clarifier-type, centrifuge is shown in Figure 8.7h. In this control system the controls are operated strictly on a time sequence, without any feedback controls of product quality.

The feed to the machine is introduced from a sludge or magma feed tank by gravity through valve #1. This valve is open throughout the cycle, except during flushing. Valve #2, the purge valve, is closed throughout the cycle, except during flushing, while valve #3, the sludge valve, is operated intermittently to discharge solids.

If a minimum head of 10 ft (~3m) is available, gravity flow is adequate for control. Feed tank location and tank and pipe size are determined to match the capacity of the specific application. The feed enters this unit at a predetermined rate and is continuously clarified. The clarified effluent is then discharged. The sludge accumulates in the system and is periodically ejected. The unit may or may not run at full speed while sludge is ejected. The interlock system or

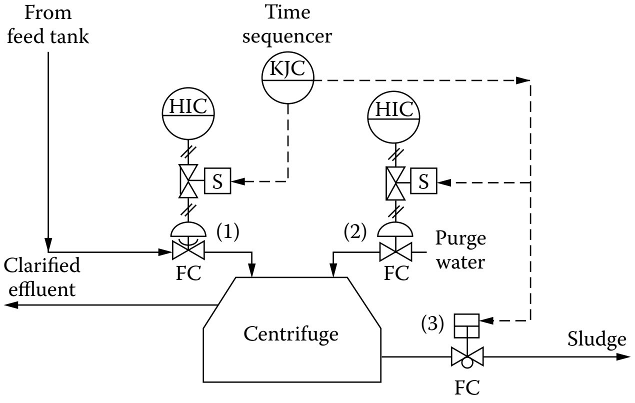


FIG. 8.7h
The most basic controls that can be provided for a desludging or clarifier-type centrifuge.

sequencing software requires a sequence timer and three working contacts to control the three control valves: 1) feed, 2) purge water, and 3) sludge. Usually added to this basic control system are partial desludging and other refinements, such as water sprays in specific parts of the unit.

In these units the choice of control valve is of major importance. For on/off service, properly chosen ball- or plug-type valves work satisfactorily. For throttling applications, refer to the slurry service valve designs shown in Figure 6.1ee and the associated discussion in Chapter 6.

In one modification of this machine, the sludge accumulates in an outer chamber. When this chamber is full, a sensing device initiates the desludging operation. Such centrifuges can tolerate significant variations in the sludge concentration in the feed liquor.

Continuous Centrifuge Control Among the most versatile centrifuge designs used in the petrochemical processing, polymerization, and waste treatment industries are the fully continuous solid-bowl type sedimentation centrifuges.

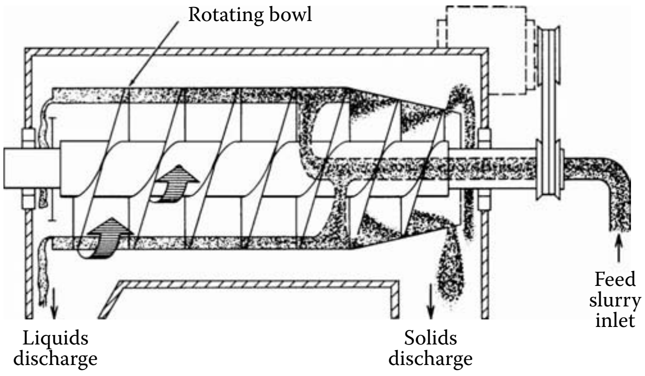


FIG. 8.7i
The operation of the rotating solid-bowl-type continuous centrifuge.

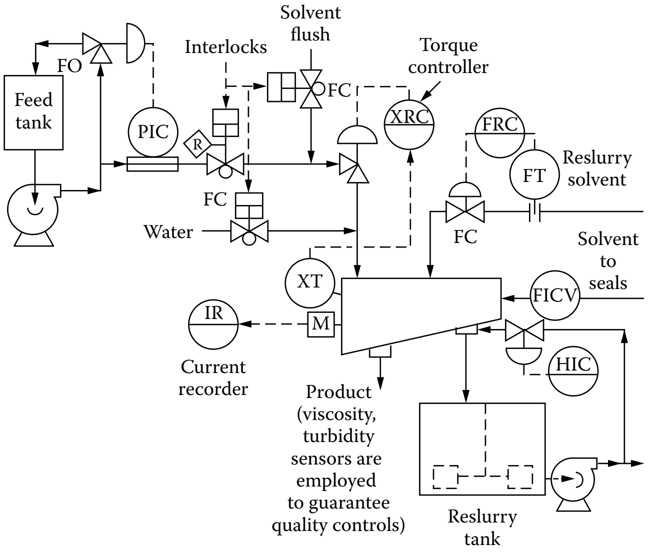


FIG. 8.7j
Solid-bowl centrifuge with feed pressure and machine torque controls.

One example of a solid-bowl centrifuge design is shown in Figure 8.7i. The slurry is introduced in the revolving bowl of the machine through a stationary feed tube at the center. It is acted upon by centrifugal force, and the solids (denser phase) are thrown against the wall. Inside the rotating bowl is a screw conveyor, which rotates in the same direction as the bowl, but with a slight speed differential with respect to the bowl rotation. This screw conveyor moves the solids up the beach and out of the liquid layer. Solids and clarified liquid are continuously discharged.

Control of these solid-bowl machines requires that the torque, which is developed by the unloading plow, be controlled. The control scheme shown in Figure 8.7j automatically adjusts the feed rate, in order to maintain the torque at a constant value.

The torque can be detected by rotating torque transducers with slip-rings and bearings, by inductively coupled torque transducers, or by noncontacting torque sensors, which are all discussed in Section 7.21 in the first volume of this handbook. Load cells are also frequently applied in such installations as torque detectors (Figure 8.7k).

The safety interlocks, not shown in Figure 8.7j, result in the shut-down of the centrifuge by closing the feed and opening the flush valve. Such shut-downs are usually initiated by the following conditions: (1) low coolant or flush flows to seals, (2) low flow or high temperature in the lubricating oil system, (3) high motor current, or (4) high torque. Other, less severe conditions, such as low reslurry solvent flow, might temporarily stop the feed and open the flush valve, but without shut-down.

It is important that a reasonably uniform slurry feed be supplied to the centrifuge. For this purpose, a circulating pump with a recycle line is usually installed to keep the slurry in motion and thus prevent the settling out of crystals in the tank or in the pipelines. In case the slurry does not settle

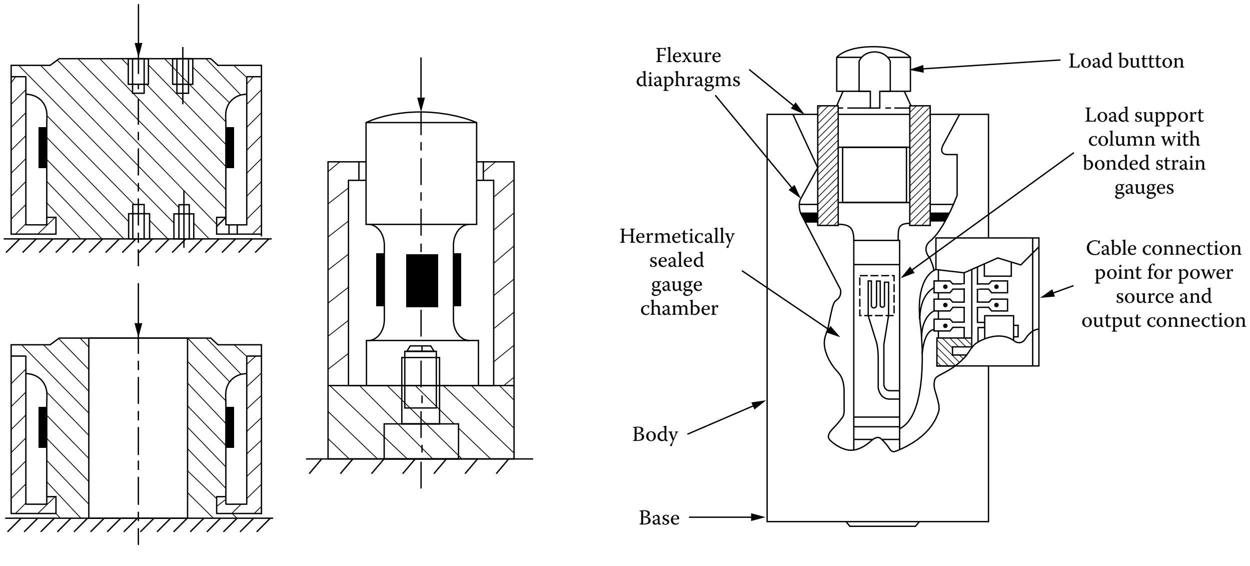


FIG. 8.7k
The force caused by torque can also be detected by canister-type load cells.

quickly, magma extraction can be performed by siphon. In Figure 8.7j, the recycle flow is throttled so as to keep the slurry feed flow to the centrifuge at constant pressure. On slurry service, it is important to protect the pressure sensor from plugging, and Figure 8.7l shows one of the chemical seal designs that can serve that purpose.

If the process slurry feed to the machine must be on flow control, the torque controller can throttle the relative speed of the conveyor and the bowl to balance the unloading requirements. The adjustment of the differential on the scroll (conveyor) is a manipulated variable that is also available to control other variables besides the torque on the machine. Product quality indicators, such as turbidity or viscosity, can also be controlled by feedback control of the speed differential in some applications.

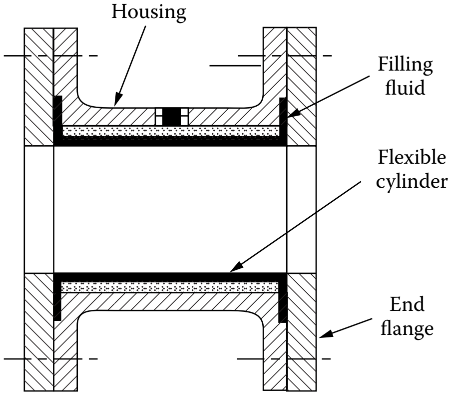


FIG. 8.7l
A full-stream spool-type seal can be used to protect the pressure sensor from plugging.

Filtration Centrifuges

Filtration centrifuges can also be batch, automatic-batch, and fully continuous. In all of them a cake is deposited on a filter medium held in a rotating basket, which is then washed and spun dry. The method of solid discharge is the main distinction among the various designs. Table 8.7c and Figures 8.7d and 8.7e provide performance data for some of the more common designs.

Automatic Batch Units Figure 8.7m illustrates two versions of intermittent feed controls for a horizontal cylindrical basket machine (also called a “peeler”) on automatic-batch sequence control. In the upper portion of the figure, a circulating pump feeds one or more centrifuges against a head tank, and the overflow from this tank is returned to the crystallizer. In the

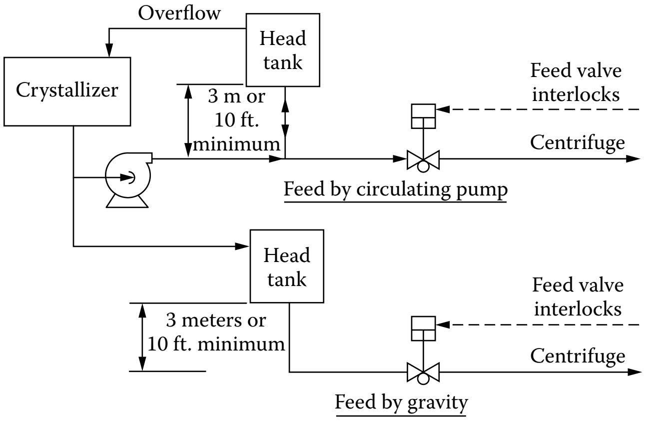


FIG. 8.7m
The intermittent feed controls of automatic batch centrifuges can utilize pumped (top) or gravity flow systems (bottom).

configuration shown at the bottom, the feed is charged by gravity alone.

A typical constant speed, automatic-batch machine will operate on the following cycle sequence:

1. *Screen rinse:* Residual layer of crystals is rinsed prior to loading.
2. *Loading:* Feed is admitted through the feed valves.
3. *Cake rinse:* After the crystal layer is established, feed is stopped and rinsing is started.
4. *Drying:* The cake (after washing) is spun dry.
5. *Unloading:* The unloading knife is cut into discharge solids.

In these machines the feed slurry and wash liquor line pressure are assumed to be reasonably steady, and therefore feed flow controllers are often not used. A variable-speed machine has a similar sequence, but in that sequence, at the appropriate points, periods are provided for acceleration and deceleration.

Continuous Centrifuges In a fully continuous machine, which can be the reciprocating pusher type or the differential scroll type, the only controls required are flow control loops to maintain the slurry and wash-water feed rates constant.

When the magma is continuously extracted, the total quantity of solids withdrawn must be controlled. This requires the measurement of the volumetric flow rate and of the solid concentration in the centrifuge feed and the throttling of the laundering fluid flow to maintain the solid charge rate to the centrifuge constant. However, in some cases this approach is considered to be too expensive.

There are less expensive methods for the continuous withdrawal of crystal magma at controlled rates: One is the removal in small batches but at such high frequency as to make the flow virtually continuous. This can be achieved by the use of ball-check-type or pulsing valves with adjustable stroke and frequency.

Another method of magma extraction involves the use of an elutriation leg attached to a crystallizer and the throttling of the laundering fluid inflow to control the magma density. One version of this control configuration is shown in Figure 8.7n. The laundering fluid is on flow control with its set point adjusted to maintain density as a measure of solid concentration. Measurement of the differential pressure between two points in the crystallizer can be a method of density detection.

An important aspect of this control system is the measurement of density in the elutriation leg. The measurement of differential pressure in this leg or between some other two points in the crystallizer can be used as a method of density detection (Figure 8.7o).

If higher accuracy is desired, ionization chamber-type radiation detectors can be used (Figure 8.7p). If they are provided with stable amplification, simple, reliable, and accurate density measurements can be obtained from them.

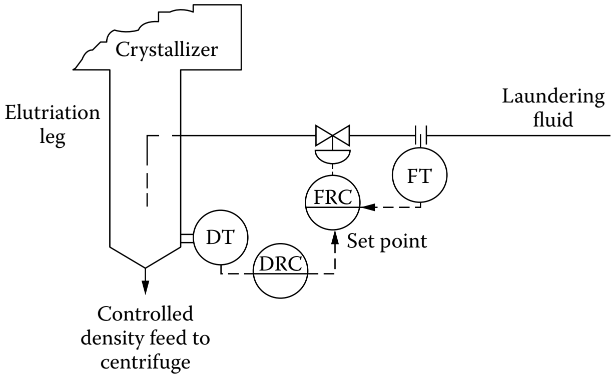


FIG. 8.7n
Controlling the density of the centrifuge feed can utilize an elutriation leg on the crystallizer.

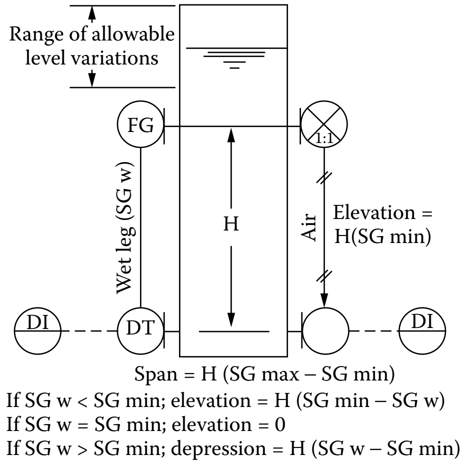


FIG. 8.7o
Density detection in the elutriation leg can be based on differential pressure measurement.

FIG. 8.7p
Ionization cell-type radiation detectors can be used to more accurately measure density.

In the crystallizer, occasionally, raw feed may be used as laundering fluid, but filtrate is more commonly used. Washing in these machines is also fully continuous, and wash fluid flow is on automatic control.

A special application is that of a scraped surface crystallizer, which is producing the slurry for separation in a continuous centrifuge. In such cases the liquor can be fed by a metering pump, and the centrifuge can be mounted in series with the crystallizer.

CONCLUSIONS

In addition to the main control aspects discussed, the actual centrifuge operation also involves a number of auxiliaries, such as wash lines for preventing the solid build-up in various parts of the unit, lube oil pumps and coolers, and discharge hoppers, which are provided with devices to prevent bridging.

Over the years, manufacturers have developed standard units that are adapted to specific applications. Operating characteristics of the various designs vary significantly, and the nature of the sludge and solids involved can necessitate extensive testing in order to minimize handling problems.

Bibliography

Bloch, H., *Process Plant Machinery,* Butterworths, 1989.

Burd, R. S., "A Study of Sludge Handling and Disposal," Washington, D.C.: Department of the Interior, 1968.

Couper, J., et. al., *Chemical Process Euipment: Selection and Design,* Gulf Professional Publishing, 2004.

Eckenfelder, W. W., and Santhanam, C. J., *Sludge Treatment,* New York: Marcel Dekker, 1982.

EPA, *Process Design Manual for Sludge Treatment and Disposal,* Washington, D.C.: Environmental Protection Agency, 1974.

EPA, *Sludge Treatment and Disposal,* Vols. I and II, Washington, D.C.: Environmental Protection Agency, 1978.

Jenkins, D., et. al., *Manual on Causes and Control of Activated Sludge Bulking and Other Solids Separation Problems*, 3rd ed., Boca Raton, FL: CRC Press, 2003.

Lipták, B., *Environmental Engineers' Handbook,* 2nd ed., Boca Raton, FL: CRC Press.

McIlvaine Co., *The Sedimentation and Centrifugation Manual,* Chicago: McIlvaine Co., 1981.

Perry, R. H., and Chilton, C. H., *Chemical Engineering Handbook,* 7th ed., New York: McGraw-Hill, 1997.

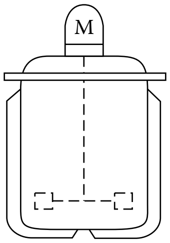
Schweitzer, P. A., *Handbook of Separation Techniques,* 3rd ed., New York: McGraw-Hill, 1997.

Walas, S., *Chemical Process Equipment: Selection and Design,* Butterworth, Heinemann, June 2002.

Woon-Fong Leung, W., *"Industrial Centrifugation Technology,* New York: McGraw-Hill, 1998.

8.8 Chemical Reactors: Batch Sequencing

D. C. KENDALL, W. F. SCHLEGEL (1970)   **H. I. HERTANU** (1985)
B. G. LIPTÁK, F. MOLNÁR (1995)   **R. A. DWIGGINS** (2005)



Flow sheet symbol

INTRODUCTION

This section concentrates on batch reactor automation and its component parts. Batch processes in general are covered in Sections 8.3 (Batch Control Description, Terminology, and Standard S88) and 8.4 (Batch Processes and their Automation). These sections should be studied when integrating a chemical reactor into a larger processing facility or into the whole plant.

Refer to Section 8.9 (Chemical Reactors: Control and Optimization) for the regulatory control and optimization strategies associated with chemical reactors, including the control of temperature and pressure and the determination of reaction endpoint. Section 8.10 (Chemical Reactors: Simulation and Modeling) describes the modeling and simulation of chemical reactors. The various aspects of DCS-based batch control systems were discussed in Chapter 4.

Control issues associated with external heat transfer equipment and transfer pumps are further described in Sections 8.11 (Chiller Control), 8.28 (Heat Exchanger, Condenser, and Evaporator Controls), and 8.34 (Pump Controls and Optimization). Many reactions require control of dissolved oxygen, redox potential, pH, or inerting systems, which are respectively covered in Sections 8.1 (Aeration and DO Controls), 8.31 (ORP Controls), 8.32 (pH Control), and 8.29 (Inerting Systems).

A batch process produces discrete quantities of material in a discontinuous fashion, such that the equipment states, process conditions, and batch composition vary with time. The starting of a pump, the opening of a valve, the completion of charging of a particular material into the reactor, and the reaching of the endpoint of a reaction are all events in time, requiring different operator responses and automated control actions. Designers and users of batch control systems must deal with time-varying process conditions and transition phenomena.

The control system will frequently serve the creation of multiple products and accommodate idle times between batches. Start-up and shutdown procedures are integral to the operation of each batch and are often automated to improve cycle time (for higher throughput) and repeatability (for quality and regulatory compliance). Preparatory procedures that must execute periodically or before every batch, such as cleaning, sterilizing, and pressure testing, may be automated as well.

Batch chemical processing has increased in importance in those industries that manufacture low-volume and high-value-added chemicals. The main advantage of batch processes is their great flexibility and rapid response to changing market conditions, but their unsteady-state operation makes tight quality control difficult. In these markets, increased customer demand for rapid commercialization of differentiated high-quality products has fueled the construction of highly flexible and highly automated batch facilities.

BATCH CONTROL CHARACTERISTICS

Sequential control encompasses all control functions (discrete and analog) triggered by events (time-based or process-based), whether normal or abnormal. In its simplest form, this comprises basic control functions analogous to those seen in continuous processes and relies on the operator to coordinate all control actions and status reporting required before, during, and after each process step. This can be overwhelming, causing overall performance to deteriorate. In a fully automated batch plant, these coordination functions are executed by the control system for multiple simultaneous batches queued from a production schedule, using preprogrammed product-specific material, equipment, and processing requirements for each batch.

Traditional batch automation is characterized by high project costs, lack of revision flexibility, and difficulties in maintaining the system. A desirable batch automation system, on the other hand, is easy to use and modify by the chemical and production engineer. It is also standardized and therefore can be designed by configuration and not by programming, plus it avoids costly maintenance and support associated with custom software. A desirable system should also be hierarchically built from a reusable library of preprogrammed and user-definable modules (application objects), so as to further simplify application development and eliminate re-engineering of commonly used components.

As shown in Table 8.8a, the automation of batch reactors can be structured in layers. Table 8.3b and Figure 8.4d have also defined such levels or layers. This section will discuss

TABLE 8.8a
Five Levels of Batch Automation

Process	*Control*
Plant	Plant management
Production	Batch scheduling, resource allocation, and reporting
Unit	Recipe-based unit control
Basic operations	Equipment control
Devices	Regulatory sequences, and discrete control safety interlock

the bottom three levels in Table 8.8a, which corresponds to Levels 2 and 3 in Table 8.3b. For detailed information about the automation of higher-level functions relating to production and plantwide automation, refer to Sections 8.3 (Batch Control Description, Terminology, and Standard S88) and 8.4 (Batch Processes and their Automation).

BATCH CONTROL STANDARDS

As covered in Sections 8.3 (Batch Control Description, Terminology, and Standard S88) and 8.4 (Batch Processes and their Automation), standards for batch and sequential automation have been prepared by the ISA SP88 Standardization Committee (1988, USA), based upon earlier work of the NAMUR Advisory Committee on Instrumentation and Control (1986, Europe).

The American National Standards Institute and the International Electrotechnical Commission have adopted the recommendations of the SP88 committee as U.S. Standards ANSI/ISA S-88.01 (1995), 88.00.02 (2001), and 88.00.03 (2003) and with minor revisions as European Standards IEC 61512-1 (1997) and 61512-2 (2001). These standards have enabled industrywide development of project methodologies and batch control system capabilities that facilitate organizing, communicating, and implementing batch automation requirements, and enable efficient skills transfer and database portability among disparate platforms. These developments should be taken into consideration when designing the automation of chemical batch reactors.

Standard Terminology

The following summarizes some of the core terminology identified in the international Purdue Workshop on Industrial Computer Systems[1] and the ISA S88.01/IEC 61512-1 standard[2] for precise definition, discussion, and solution of batch control problems:

BASIC CONTROL Continuously executed algorithms that drive the process or equipment to a specified state and keep it there, such as indicators, regulatory and device controls, and interlocks.

BATCH The product that is produced by one execution of a recipe.

CAMPAIGN The total production run of one product, for example, for a single order or season, consisting of one or more lots.

COMMON RESOURCE An equipment entity that services more than one unit, either simultaneously (shared-use resource) or one at a time (exclusive-use resource).

CONTROL MODULE Lowest-level equipment grouping that acts as a single entity from a control standpoint, but cannot execute procedural elements. May be an individual measurement (with suitable signal conditioning or state names) or a grouping of directly coupled actuators with their associated measurements, alarms, and control actions, including subordinate control modules as appropriate. Examples are an uncontrolled temperature, a flow control loop, an automatic block valve with limit switches, or a header containing interlocked (mutually exclusive) block valves.

COORDINATION CONTROL Control functions existing at multiple levels to schedule batches and manage recipes, procedural control execution, equipment entity allocation, and batch data.

EQUIPMENT CONTROL The procedural, basic, and coordination control capability that enables an equipment entity to perform its function. It is not part of a recipe, but may be directed by a recipe.

EQUIPMENT ENTITY A set of process and control equipment and associated control capability that has been grouped together to provide some specified process or equipment functionality.

EQUIPMENT MODULE An equipment entity incorporating necessary devices, control modules, and application-specific states and modes to execute some basic process-oriented task or group of tasks. May include equipment procedural elements (typically phase logic) or subordinate equipment modules, but may not overlap other equipment modules. Examples are a common (exclusive-use) weigh tank, a recirculation/transfer pump within a unit, and a shared-use ingredient dosing system.

FORMULA The list of process inputs, outputs, and data (operating set points, reported values, timing) required to execute the batch procedure.

LOT A collection of batches prepared using the same recipe. Typically, all batches of a lot are prepared from the same homogeneous source of raw material.

OPERATION A major programmed processing action or set of related actions, normally consisting of one or more phases.

PHASE The lowest level of procedural control to accomplish a process-oriented task. Phases may be further subdivided into equipment-oriented steps and transitions for executing a defined task, as described in

European Standard IEC 60848 (1988) for specification of sequential function charts. Normally, the phase boundaries represent points of process transition, hold, or activity. The boundaries define major milestones and possible points of safe intervention. Phases may exist either as part of a recipe procedure (recipe phase) or independently for equipment control (equipment phase); however, any constituent steps are always part of the equipment phase.

PROCEDURAL CONTROL Control that sequentially directs subordinate procedural elements or basic controls to execute the steps required by its defined process-oriented task.

PROCEDURE A user-defined set of instructions that define the strategy for making a single batch of a particular type or grade of product.

PROCESS CELL A grouping of equipment that comprises one or more complete trains and defines the immediate local domain for production scheduling (analogous to a work cell in discrete manufacturing).

RECIPE The complete set of data and operations that define the control requirements for a particular type or grade of final or intermediate product. Specifically, each recipe comprises a header, formula, procedure, and equipment requirements.

TRAIN A grouping within one process cell of units and associated lower-level equipment that is capable of making a batch of material. A train may define a single equipment path for a batch or multiple possibilities, of which one will be selected based on availability during execution of the control recipe. Multiple batches can be processed simultaneously in the same train (but different units).

UNIT An equipment entity that contains and performs some major processing activity or activities (e.g., react, crystallize, make solution) on one batch at a time. A unit normally comprises a major piece of equipment and directly associated control modules or equipment modules that are not shared with other units.

UNIT PROCEDURE A major programmed processing action or set of related actions, normally consisting of one or more operations. Unit procedures are naturally related to a distinct regime of production: for example, all processing carried out in one batch unit of a multiunit production line.

BATCH REACTOR CONTROL

Regardless of whether the ISA S88/IEC 61512 standards will be fully applied in the automation of a particular batch reactor, their basic concepts should be understood to provide a common basis for analysis, modularization, and communication of the control requirements. The following summary identifies key points of Sections 8.3 (Batch Control, Terminology) and 8.4 (Batch Processes and their Automation) that are required to discuss batch reactor control in this context.

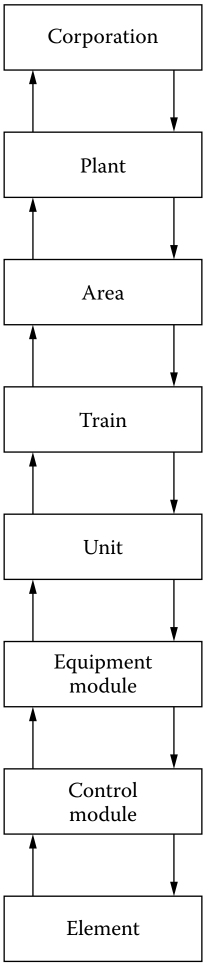


FIG. 8.8b
The equipment model of hierarchically grouped physical assets.

The physical assets of a manufacturing enterprise are hierarchically grouped within a physical model into the following levels: enterprise, site, area, process cell, unit, equipment module, and control module (Figure 8.8b).

The ISA S88/IEC 61512 documents identify a standard structure for batch control at the bottom four levels of this hierarchy (see Definitions above). Suitable decomposition of process cells into units, equipment modules, and control modules underpins successful application of the standards, as does the long-term development of well-constructed specification and component libraries for recurring module and procedural element functionalities. Table 8.8c illustrates how the various conceptual models defined in the standards and summarized below align with this physical model.

Control Concepts

Batch control spans multiple levels of the physical model and comprises three distinct types of control:

TABLE 8.8c
Conceptual Model Relationships in ISA S88/IEC 61512

Physical Model Structure	*Batch Control Concepts*	*Process Model Structure*	*Procedural Model Structure*	*Recipe Types*	*Basis of Recipe Procedure Definition*
Enterprise				General recipe	Process model (general processing requirements)
Site				Site recipe	
Area					
Process cell	Coordination control			Master recipe	Procedural model (procedure to deploy process model using available equipment)
		Process	Procedure	Control recipe	
Unit*	Procedural control	Process stage	Unit procedure		
		Process operation	Operation		
		Process action	Phase		
Equipment module*					
	Basic control				
Control module*					

* May comprise common resources that service multiple units, either simultaneously (shared-use resource) or one at a time (exclusive-use resource).

Basic control is responsible for establishing and maintaining associated process and equipment variables in defined states based on external commands, analogous to continuous process control. It comprises continuously executed functions of control modules and equipment modules such as process indication, regulatory (including PID, fuzzy, multivariable, and custom algorithm) and discrete control, interlocking, and cyclic operations (such as sump pump or bag house solenoid control). These control functions may be executed automatically, manually, or external to the control system.

Procedural control comprises both equipment-specific (equipment control) and product-specific (recipe control) functions that are hierarchically organized to determine the execution order of equipment phases and, through those phases, provide supervisory commands to basic control. Recipe control requires definition of a procedural control element at the recipe procedure level and supporting equipment phases at the unit or equipment module level, with peer-to-peer transference from recipe control to equipment control occurring at some level of the recipe and equipment procedure hierarchies.

Coordination control operates primarily at the process cell level to schedule batches and manage recipes, procedural control execution, entity allocation, and batch data. Status, availability, and control mode propagation (downward or upward in the control hierarchy), however, are parts of coordination control that are best managed by the equipment entities themselves (units, equipment modules, and control modules).

Recipe Types

A recipe typically defines the product-specific procedural control requirements for producing a single batch of a specific product. Additional recipes may be defined to separately schedule ancillary activities for production equipment, such as cleaning and sterilizing. Four standard recipe types defined at various levels of the physical model are used as follows:

General recipes are defined at the enterprise level and tightly controlled by R&D to standardize the process across multiple sites and support multisite production planning. These material-centric recipes define a complete process from raw materials or intermediate products to final product and are generally represented in material-centric Process Procedure Chart (PPC) notation defined in Section 7 of ANSI/ISA-88.00.03.

Site recipes are defined in PPC notation based on corresponding general recipes, with some added site-specific information. They standardize the process across the site, support long-term production scheduling, and reflect only that portion of each general recipe that can be executed locally.

Master and Control Recipes *Master recipes* unique to each product or intermediate product made in each process cell are constructed from corresponding site recipes either manually or automatically, reflecting only the portion of the production activity that can be completed in that cell. These standardize the actual production procedure to be used based on the type of equipment available in the process cell and support detailed production scheduling.

Master recipes are generally represented in equipment-centric Procedure Function Chart (PFC) notation defined in Section 6 of ANSI/ISA-88.00.02. Ancillary master recipes may also be defined to efficiently schedule and execute product-independent recipes for such activities as equipment cleaning and sterilization, which are excluded from general and site recipes because they are not process-related.

Control recipes are batch-specific and identify the specific equipment used to make the batch. They are normally created by a batch server from the corresponding master recipe upon scheduling and initialization of each batch and include all associated schedule details, parameter changes, and results.

Recipe Structure Each recipe is defined by four basic elements, with a fifth category for ancillary design and operation data. The content of each category varies with recipe type as follows:

Header contains administrative information (e.g., recipe name, product identification, and version control).

Procedure hierarchically defines the processing strategy, based on the process model (process, process stage, process operation, process action) for general and site recipes and based on the procedural model (procedure, unit procedure, operation, phase) for master and control recipes. Each level of the hierarchy is defined by a recipe-specific series and parallel combination of subordinates and transition conditions.

Equipment requirements are listed for the entire batch and for each procedural element. These comprise design data in general and site recipes (such as materials of construction and design temperatures) vs. selection criteria for master and control recipes (such as first available sterilized "type 2" reactor).

Formula comprises all parameters to properly execute the procedure. All recipe types require process operating conditions, but general and site recipes may include acceptable parameter ranges and detailed heat and mass balances vs. master and control recipes with settings related to the local equipment type and capabilities.

General information, optionally provided, can range from special safety considerations and conceptual process flow diagrams in general and site recipes to links to material safety data sheets, standard operating practices, and P&IDs in master and control recipes.

Programming Concepts

Three underlying concepts that also must be understood for successful implementation of these models relate to object-oriented programming, object commands and states, and exception handling. These concepts apply to the models described above as follows:

Object-Oriented Programming The power of the S88 models lies in the separate layering, modularity, and reusability of equipment and procedural elements. Object-oriented programming is an enabling technology for efficient implementation (and specification) of these structures, whereby the functionality (including commands and state definitions), data links, and parameters associated with each control element are encapsulated in a single object.

Similar objects may be generated through simple replication with new data links and parameter values, but this creates multiple copies of the program that must be individually corrected and retested for validation. If the same objects are instead *instantiated* from an object class with the same properties, then a single copy of the source code provides the designed functionality. In this case, full validation is required for only one instance and data links and parameter values checked for all instances.

Combining object classes with minor functional differences into a single class can yield even greater efficiency. Depending upon the target system capabilities, this may be accommodated through either additional parameters or subclasses. Following validation, the specifications and code for each object class and subclass should be stored in a library for reuse on subsequent projects.

Object Commands and States *State* is defined in S88.01 as "the condition of an equipment entity or of a procedural element at a given time." Equipment states are represented by user-defined enumerations, which normally relate to availability for units (e.g., reactor with states of dirty, clean, sterile, and out of service) and combinations of input/output (I/O) states for control modules (e.g., block valve with states of open and close) and equipment modules (e.g., pumping station with states of stop, recirculate, and transfer).

Unit objects do not normally receive commands, as in principle they are just containers for lower-level entities and procedural elements. Control module and equipment module state commands comprise requests to set the module's output(s) to the indicated state, based on an internal lookup or hard-coded correlation (note that the outputs are not commanded directly, and the receiving module ultimately decides if the requested transition is permitted).

Phase objects have a standard set of states and commands, which usually corresponds to the example given in S88.01. Under this model, the phase executes normal logic and monitoring functions in the running state and optional exception logic in the pausing, holding, stopping, or aborting state before entering the quiescent paused, held, stopped, or aborted state, respectively.

These state transitions may be triggered by the operator (if permitted by system security) or automatically (via its internal exception monitoring functions or cascaded from higher-level procedural elements) and may be required to cascade upward through the procedural hierarchy. Higher level procedural elements have the same state enumerations and propagation capabilities, but generally handle exception conditions through normal transition logic.

Exception Handling *Exception handling* is defined in S88.01 as "those functions that deal with plant or process contingencies and other events which occur outside the normal or desired behavior of batch control." Due to transient behavior of batch processes, they are inherently more problematic to operate than continuous units and must respond appropriately to a large number of possible failures in the context of what is happening at the time. The problems encountered may relate to unavailability of material or equipment, process problems, or equipment malfunctions. This occupies a large portion of the control definition and is usually underestimated by engineers lacking batch control experience.

Pushing exception handling as far down the control hierarchy as possible and maximally utilizing low-level object classes simplifies the task by minimizing the amount of detail required in supervisory logic. For example, process alarms and device failures (e.g., failure to reach the commanded state) should be monitored and alarmed by the responsible control modules and responded to by operational interlocks.

Exception handling by associated phase logic need only enable and adjust any process-dependent alarm settings, confirm that commanded device states are permitted, and trigger its own exception logic when warranted. If recovery requires aborting one phase and executing another (such as reheating to reaction temperature following a feed loss and subsequent cool-down in a semibatch operation), then it must be dealt with manually or through appropriate transition logic above the phase level. Either way, appropriate interlocks must be present at the control module level (e.g., to prevent feed resuming below minimum temperature).

SEQUENCING LOGIC CONTROLS

A batch reactor control system is a combination of analog and discrete control functions supervised by procedural control logic. The various analog loops used in reactor control are discussed in Section 8.9 (Chemical Reactors — Control and Optimization). The discrete controls provide the capability to turn things on and off, and the procedural controls serve to initiate and step through the various process phases in logical sequence, as shown in Figure 8.8d. The equipment phases in a unit or common resource execute distinct minor processing activities in the batch cycle, such as charge, mix, heat, and transfer.

Equipment phase logic details can be reduced to a time-ordered combination of a few basic state transitions, resulting

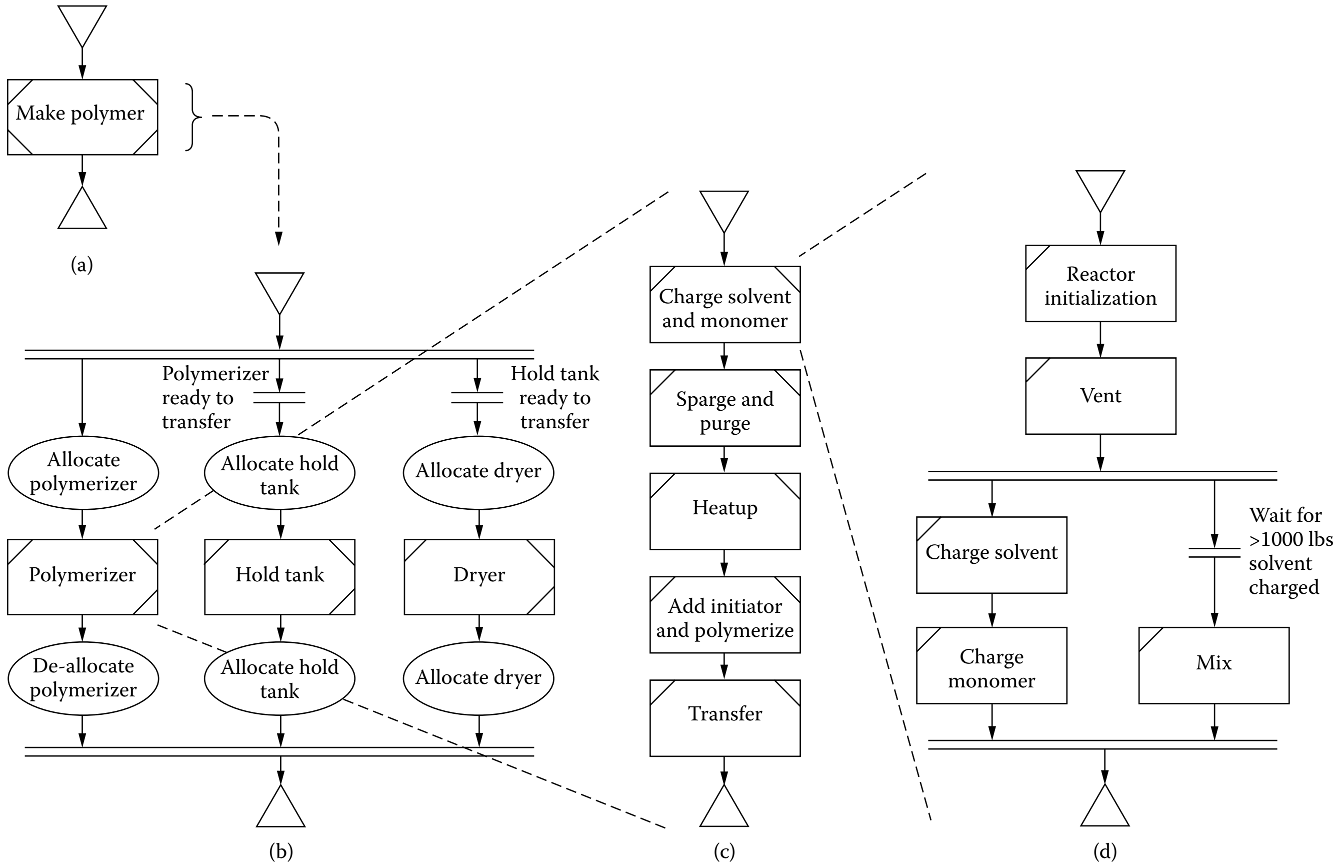


FIG. 8.8d

The sequence of steps required to produce a polymer product are shown in a hierarchy of increasingly detailed Procedure Function Charts (PFCs). These PFCs are comprised of recipe procedural elements down to the phase level for one operation of one unit procedure in a recipe: Procedure for recipe "Polymer X," (b) PFC for procedure "Make polymer," (c) PFC for unit procedure "Polymerizer," (d) PFC for operation "Charge solvent and monomer."

from the procedural steps described in Figure 8.8d. These actions are listed below:

1. Operate on/off devices (pumps, valves, and so on)
2. Activate or deactivate control loops
3. Open and close cascade loops
4. Supply control loop parameters, such as set point, ramp rate, alarm limits, tuning parameters, and flow integrals
5. Integrate flows
6. Initiate times or delays between processing steps
7. Perform calculations
8. Compare values, measurements, and test indicators and branch according to the results of the comparisons
9. Initiate alarm status or operator messages
10. Release control if nothing more can be done until the next sequencing interval

It is important to note that the data acquisition and loop control functions are not included as part of the phase logic. Rather, the equipment phase logic plays a supervisory role, communicating changes to the control modules and equipment modules as required in the same way that an operator would adjust set points and alarm limits during the course of a batch. In turn, the execution of equipment phases within each unit is coordinated and parameterized by the currently operating control recipes (or in some circumstances by the operator). Additionally, when purely manual actions are required, the phase will trigger an operator message to perform the activity and confirm after it is complete.

Figure 8.8e gives a specific example of the above generalized description. The upper portion of the illustration shows the physical equipment involved, including the nine controlled devices whose states are being logically sequenced. The lower portion of the illustration identifies the sequenced process states. During process state A the required quantity of raw material A is being charged. The prerequisites for the initiation of state A are that 1 and 4 be on and 6 be off. This means that the agitator must be on, the vent valve must be open, and the drain valve must be closed.

When these prerequisites are satisfied, state A is initiated by turning 2 on. This means that the charge valve of reactant A is opened and a control loop, such as that shown in Figure 8.8f, is activated. The required quantity of this ingredient is set by the recipe as Q_A. When this target is reached, process state A is terminated and process state B is initiated, which then performs its own tasks defined by the sequencing logic. In a more sophisticated reactor control system, the number of controlled devices and the number of logical sequence steps is much greater, but the basic concept is the same.

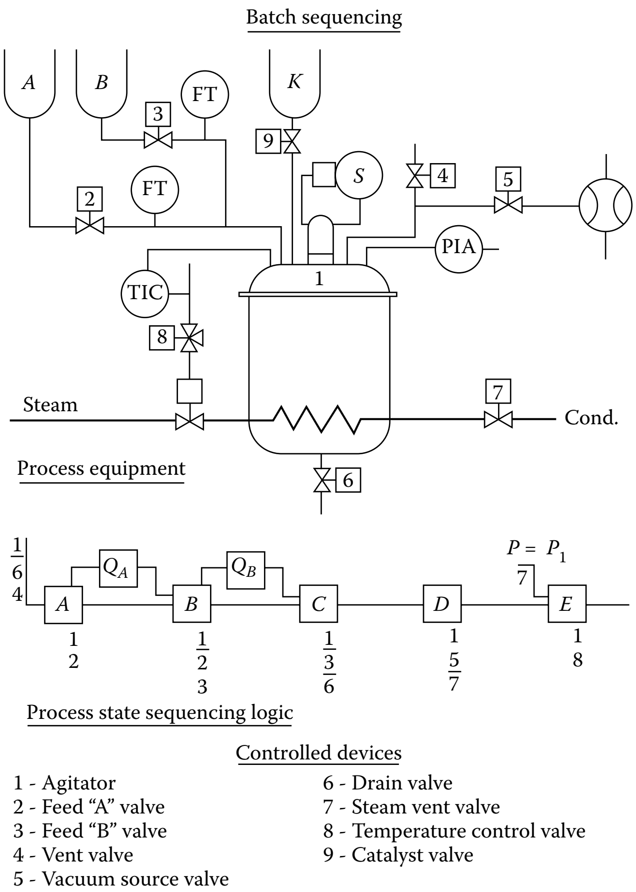


FIG. 8.8e

Example of batch reactor sequencing logic. On top is the physical equipment used; on the bottom is a flowsheet of the sequenced process states.

Recipe Procedures

The standard graphical notation for representing complex procedural control is the Procedure Function Chart for master and control recipes and the Process Procedure Chart for general and site recipes, which are respectively defined in the S88.00.02 and S88.00.03 standards. S88.00.03 also defines a more convenient tabular method that may be used in lieu of the PFC or PPC to specify procedural logic that is not highly branched. This method simply lists the operations and phases in sequence, using icons and path numbers to identify each parallel series of steps.

A simple example of a PFC is illustrated in Figure 8.8d. The triangles at the top and bottom of each PFC represent

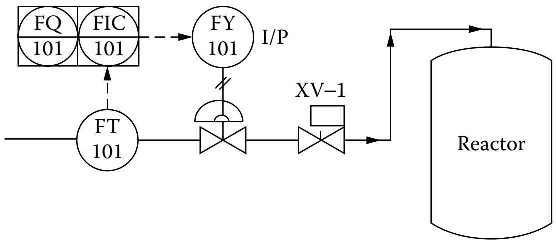


FIG. 8.8f

Flow control loop for charging reactant A.

its start and endpoints and rectangles indicate the procedural elements, with corner markers distinguishing the symbols for a procedure, unit procedure, operation, and phase. Ovals are used to indicate equipment allocation steps, and connecting arrows represent the sequence in which procedural elements are initiated at the next lower level, upon the prior element completing its execution.

Transition markers, indicated by a short double line, may be inserted at any point to provide additional conditions for initiation of the subsequent element. (Note that a procedural element may be notified when the transition below it goes TRUE, which provides a simple mechanism for requesting early termination of that element.) Parallel paths with a connecting double line above and below, as illustrated in Figure 8.8d(d), are initiated simultaneously and executed independently, with control passing beyond the endpoint after all paths have completed their execution.

Branching (not illustrated) among two or more alternative paths is possible simply by splitting the connecting line and placing a transition condition at the start of each path. This feature can be used in a reaction operation, for example, to repeat periodic catalyst booster charge and product sampling phases until a satisfactory lab result is obtained (i.e., operation-level exception handling).

Procedure Function Chart Execution

Execution of the illustrated procedure will be as follows. Procedure “Make Polymer” in Figure 8.8d(a) commences operation when a batch has been scheduled to execute on some train (which may specify a single product path or multiple options), initialized, and started. Figure 8.8d(b) specifies parallel execution along three paths for different unit procedures. The first branch causes an immediate attempt to allocate a polymerizer for the batch, selected polymerizers in the designated train.

If no polymers are immediately available (i.e., they are in use for other batches) or none are in the correct state (e.g., they have been cleaned by a separate equipment recipe that manages equipment and cleaning fluids), execution is suspended pending allocation of a polymerizer for the batch. The next step initiates the polymerizer unit procedure and waits for it to complete before deallocating the unit. Meanwhile, the parallel path for the hold tank waits for the transition condition “Polymerizer ready to transfer,” which becomes TRUE when the polymerizer reaches its transfer phase.

(Because this phase must work in tandem with the hold tank’s charge phase to move material between the units, it is necessary that both units be allocated to the batch during the transfer and that the unit procedures appear in parallel on the PFC. The transition condition prevents premature allocation of the hold tank, which might rob the unit of valuable operating time on another batch.)

After the transition goes TRUE, the hold tank unit level sequencing is analogous to the polymerizer (except that connecting pathways are also validated before allocating a hold tank); likewise for the dryer after “Hold tank ready to transfer” is satisfied. The “Make polymer” procedure is complete when all three paths have fully executed.

Continuing with this example, execution of one unit procedure, “Polymerizer,” triggers the PFC sequence illustrated in Figure 8.8d(c), in which each operation simply begins and executes to completion after its predecessor is complete. The “Polymerizer” unit procedure is complete when the entire series of operations is finished. The details of the “Charge solvent and monomer” operation, illustrated in Figure 8.8d(d), reveal how the recipe ultimately sequences the phases.

The progression of series and parallel execution for this typical operation is analogous to that described previously for the unit procedure. The mechanism for control propagation, however, differs significantly in that recipe phases merely contain parameters, links to equipment phase parameters, and equipment phase states and commands, with all of the procedural details located in the equipment phase. Moreover, these links are established only at run time within the control recipe, based on the currently allocated units for the associated batch. This is because each recipe phase (product-specific) must be able to interface with any corresponding equipment phase (equipment-specific) that might be selected to run the product.

Likewise, each equipment phase can be commanded at different times by a number of recipe phase instances, which may appear in the same or different recipes. Although the recipe phases and equipment phases belong to different object classes, this interface requires that all phases that are defined in the recipe unit class be present in the equipment unit. The corresponding phases in different equipment units need not be of the same object class, provided they expose the same parameter interface for the recipe phase. (It should be noted that this transference from recipe to equipment control may be done at any procedural level, but it usually occurs at the phase level and is discussed here only in that context.)

Equipment Control Sequence Logic

The sequential control functions of equipment phases, equipment modules, and control modules represent the heart of batch control functionality and the vast bulk of control software that must be defined and implemented. Various tools and methods have been used to describe the intended functions, including ladder and Boolean diagrams, state diagrams, matrix diagrams, time-sequence diagrams, Gantt charts, and sequential function charts, usually supplemented by a written description of the control objectives for each control object. Time-sequence diagrams are used here to illustrate detailed sequences of events.

Time-Sequence Diagrams

The time-sequence diagram shows the sequential changes in control actions and their time relationships. The sequence of events can be followed along dotted lines for vertical time coincidence and along the solid lines in a horizontal direction for sequence control.

The diamond symbol is used to indicate a trigger event, and a vertical dotted line indicates the time coincidence trigger events. When two or more diamonds occur in the same relative time line, an "AND" logic condition is assumed.

The format of the time-sequence diagram proceeds from left to right in discrete time steps, with relative time being the horizontal coordinate.

Figure 8.8g illustrates a time-sequence diagram for a flow loop, such as the one shown in Figure 8.8f. First, the valve XV-1 is opened and the controller FIC-101 is placed in automatic (t_0). When XV-1 is open and FIC-101 is in "auto," the continuous task of flow integration is started (t_1). When the desired total is reached (t_2), XV-1 is closed. When XV-1 is closed (t_3), the integration of flow continues to check for leakage. If, at the end of a preset time interval, the total flow is still zero, FIC-101 is switched back to manual (t_4).

Figure 8.8h describes a control system in both ladder and time-sequence terminology. Here, when the start button is closed and XV-1 is open (its contact is closed), motor coil *M* is energized. When the coil is energized, its auxiliary "M-contact" closes. Once the M-contact is closed, it keeps the coil energized even if the start button contact reopens. If at any time the XV-1 valve would fail closed, its limit switch contact will open to de-energize the M coil. Once the coil is de-energized, the auxiliary M-contact reopens.

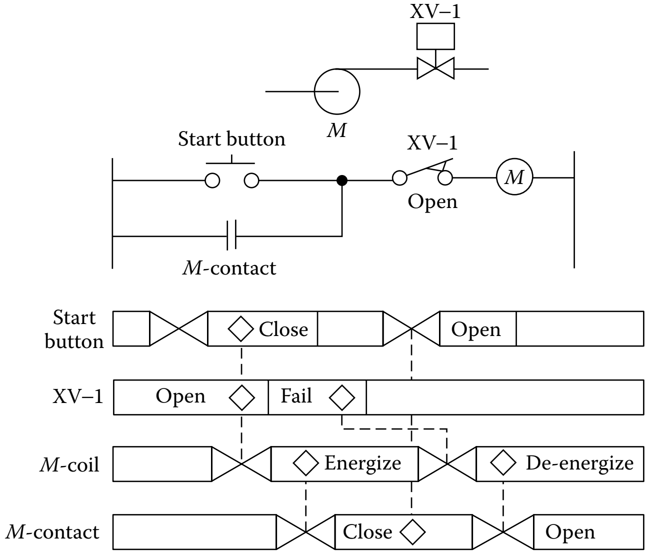


FIG. 8.8h
The time-sequence diagram of the flow control loop in Figure 8.8f is described in ladder diagram format at the top and by time-sequence diagram in the lower part of this figure.

In batch process units, the batch sequence is subdivided into *process states*. Each state is given a unique name, such as CHARGE, REACT, HEAT, COOL, HOLD, DISCHARGE, WASH, and EMPTY. Within each state, discrete

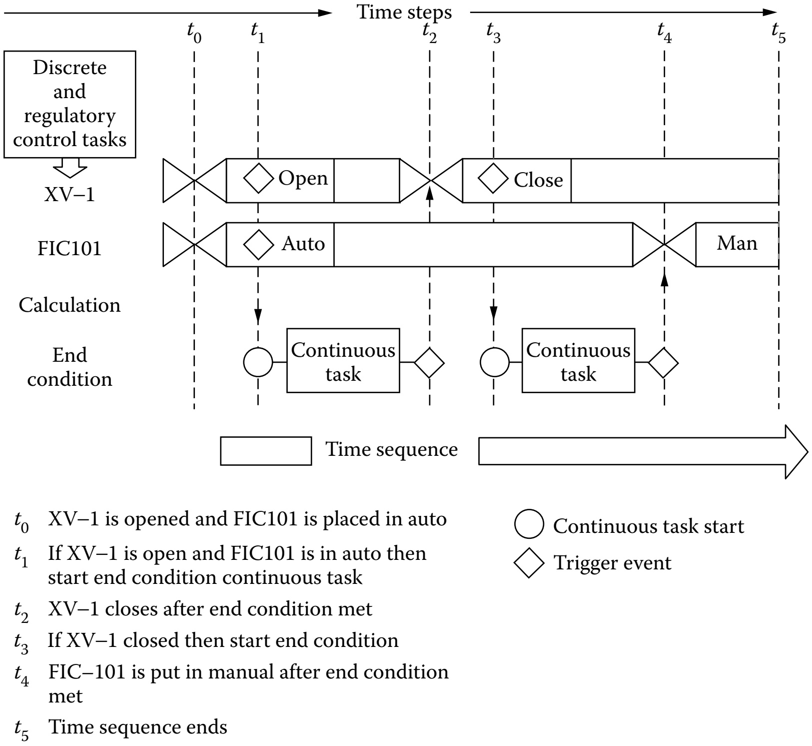


FIG. 8.8g
The time-sequence diagram for the flow control loop in Figure 8.8f. The sequence proceeds in discrete steps from left to right.

control functions, continuous-regulatory control functions, and safety and permissive interlocks are performed.

ENGINEERING A BATCH CONTROL STRATEGY

When microprocessor-based controls are used, the control system design follows the step-by-step approach described in Figure 8.8i to specify the functional requirements. The iterative nature indicated by the decision at the bottom is due to refinement of P&IDs, process definition, and control logic interactions over the duration of the project. Of the eight steps shown, steps 1-2 provide an overall foundation for the work to follow. Steps 3-5 specify equipment-centric requirements, steps 6-7 define the process sequencing and supervision, and steps 8-9 identify exception handling and reporting requirements.

Designing a suitable degree of control modularity at the equipment module level is the most important and challenging aspect of establishing overall design modularity and maximizing reuse of prior engineering efforts. Typical sequence control tasks and events to consider are shown in Table 8.8j.

Preparatory Steps

In the first step of Figure 8.8i (step 1), the process control engineer should review the P&ID flow sheets to make sure that the measurement and control devices and control strategies are accurately represented and that they are capable of meeting the requirements of the applicable general/site recipes or process description, standard automation practices, and expected cycle times. Ancillary automation requirements that are not covered by general/site recipes, such as clean in place (CIP), should also be reviewed. Any subsequently identified P&ID deficiencies must be noted as soon as possible to avoid costly rework later in the project.

In the next step (step 2), the engineer must identify and tabulate the process cells, units, trains, and common resources. As defined by ISA S88/IEC 61512 standards, process cells identify the physical limits for any single recipe/batch. Units and their associated controls may be allocated to only one batch at a time, so vessels that sequentially hold the same batch within a process cell line must be identified as separate units if they are to simultaneously process successive batches.

Equipment and controls that service more than one unit (e.g., a shared transfer pump) must be separately identified as common resources. Similar assignments must be made for continuous parts of the process, as units also define the common basis for alarm groupings and database structure.

Equipment-Related Requirements

In step 3, the process control engineer should identify on the P&IDs the discrete and analog measurement and control devices and tag them in terms of input and output signals. (If the instrument index is available at this time, the I/O list items may easily be extracted from it. Once the I/O list is established, it should be checked against the P&ID for recent changes. Note that assigning I/O to units and common resources may not be necessary, but will be helpful for sorting purposes during detailed design.)

In step 4, the engineer must define all basic control requirements (control modules and equipment modules, except phase logic) associated with each unit and common resource identified in step 2. In general, control modules are at least structurally well-defined by the P&ID, but equipment modules entail coordinated control (usually command- or event-driven) of multiple devices that is more subjectively defined based on P&ID symbols and notes, coupled with the engineer's analysis of operational requirements.

A control specification comprising a written description (or cause-and-effect matrix for simple interlocks) or control diagrams (based on ISA symbology, Boolean or ladder logic, sequential function chart, or time-sequence notation) is required to define each unique module class. If such classes have been well designed on previous projects, most new requirements will be drawn from a previously developed specification library, supplemented with a project-specific tabulation (sorted by unit or common resource) of tags, connections, and parameters that are unique to each required instance.

In order to maximize future reuse, specifications for each class of new modules should be as general as practical, consolidating similar types by using internal parameters to select features. Operating modes, human and supervisory control requirements, indicators (e.g., states, fault, module performance statistics), and relevant process calculations (e.g., material and heat balances and recipe corrections based on quality control data) should also be defined.

In step 5, the equipment procedural control is defined for each equipment module and unit, comprising phases and higher-level elements as appropriate. Supervisory control and parameter interfaces for each phase and suitable transition conditions and actions for each phase state must be identified.

Phase logic is the workhorse of procedural control, administering the real-time control functions primarily by sequentially manipulating and monitoring states of the various equipment modules and control modules via the previously defined basic controls. Equipment procedural elements, like basic control entities, can be reused across multiple products and projects by creating libraries of flexible and modular object classes for previously specified functionalities.

To be usable, however, such classes must be based on indirect addressing schemes that reference tag names generically within the unit or equipment module. Procedural control object classes, like those for basic control, can thus be documented in detail just for a single instance, with an appended tabulation of suitable linkages and parameter values for each instance.

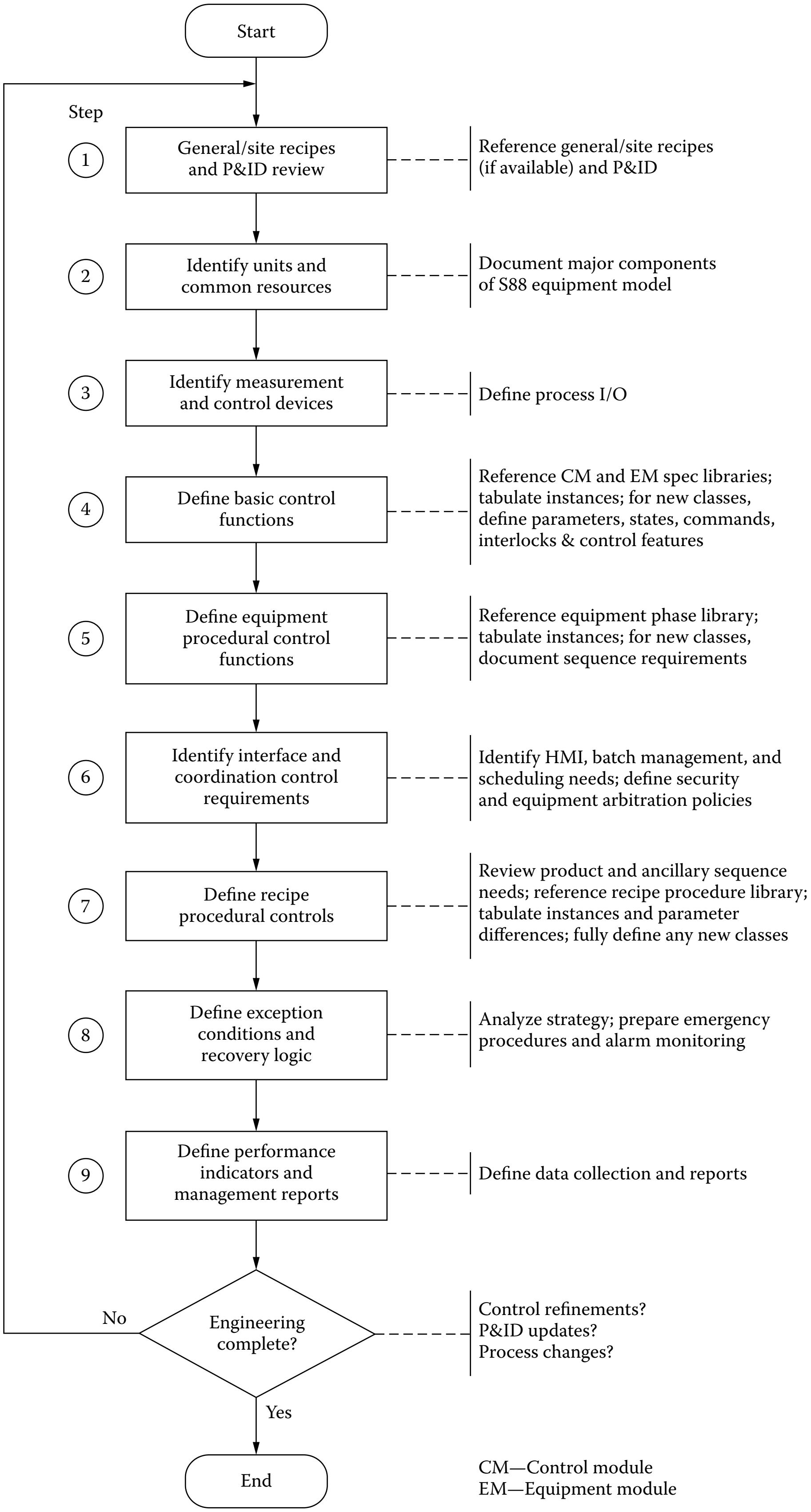


FIG. 8.8i

Step-by-step approach to specifying a microprocessor-based batch control system.

TABLE 8.8j
Typical Sequence Control Tasks and Events

1. Sequence start prerequisites and permissives (e.g., unit available, previous batch complete, interlocks satisfied)
2. Discrete element and devices (e.g., outlet valve, inlet value, manifold, agitator)
3. Regulatory control loop setpoint changes, mode changes, output changes, profiles, etc. (A/M-PID algorithm, SP profile, output)
4. End conditions
5. Timers
6. Calculations (e.g., heat or mass balance)
7. Operator action request/response (e.g., clean, take laboratory analysis, check batch end)
8. Process failure conditions (e.g., level too high, pressure too low)
9. Failure actions (branch to alternative sequence)

Process Sequencing and Supervision

In step 6, the engineer must identify the interface (human and data) and coordination control requirements, including how recipe management, production planning and scheduling, and production information management will work. If a commercial batch server will be used to implement the upper-level procedural control requirements, it will also perform most coordination control functions; otherwise, custom programming must be specified for this purpose (not usually cost-effective unless the required capability is very limited).

Guidelines must be developed for equipment arbitration, batch historization (including product genealogy tracking and electronic signatures, as needed), and application security. A custom operator interface may be required (along with extra phase logic for setting display variables such as status, step descriptors, prompts, fault messages, and parameters) to monitor and control phase execution details. Custom interlock displays and faceplates (and module-based display variables) may be required to adequately monitor and control nonstandard basic control functions.

In step 7, the master recipe is defined for each product, based on the process information identified in step 1, and for each ancillary equipment operation (e.g., CIP) that will be routinely scheduled. The procedure, unit procedure, and operation levels of the control strategy must be defined, as well as the equipment (unit and common resource) and formula requirements. Depending upon the equipment controls specified in step 5, the product recipes may be either unit-specific or not. Definition of master recipes can also utilize component libraries for rapid construction, particularly in circumstances where tabulation of formula differences is sufficient to accommodate most products.

Exception Handling and Reporting

In step 8, an overall failure analysis must be done, with the plant considered in its entirety. The consequences of potential device and equipment malfunctions, unexpected process delays and events, and abnormal process and quality measurements must be evaluated relative to the need for alarms or messages to alert the operator, for automatically holding or aborting the batch, and for any special response or recovery procedures. Various adjustments of the previously designed strategies may be required.

In step 9, all the historical data trending and reporting necessary for adequate production and performance management are designed and formatted. Batch processes have unique data storage and retrieval requirements, as information must be identified with individual batches as they move through the plant, as well as their ultimate material origins and disposition. Adjustments of the previously designed strategies may be required to cover unanticipated data requirements.

Implementation

Various microprocessor- and computer-based systems are specifically designed to perform the wide variety of control tasks involved in batch processes.

In the area of continuous (regulatory) control, a fill-in-the-blank-type language supported by CRT-based operator consoles has become an accepted standard. The fill-in-the-blank language is used to configure the control system database and to create many of the necessary control algorithms.

The discrete functions (direct on/off control, interlock control, sequence control) require *programming.*

Many programming languages available for generating control software are of the “high-level” type, meaning that the instructions are expressed in English. Various structures and procedures contained in the programming languages offered for control applications include:

- Processing discrete inputs and outputs
- Timing and counting routines
- Special routines applied to start up electric motors sequentially, in order to limit the inrush current imposed on the power supply
- Integration of analog variables
- Message triggering
- Transfer routines for synchronization of dual processors (whenever 100% redundancy is required).
- Dynamic links of discrete variables to P&ID-type graphic displays for monitoring batch processes
- Interface routines for linking recipe data files with the main control program

Process management functions are programmed using high-level languages that are well suited for data manipulations.

The application software development for a certain defined and specified project is considered finished when all the programs are debugged and are running within the constraints of the operating system and the hardware configuration. To avoid unexpected control interactions during start-up,

TABLE 8.8k
Performance Capabilities of Batch Charging Instruments

Detector	Type	Min. Error	Rangeability
Load cell weighing	Mass	±0.1% FS	NL
Turbine-type flowmeters	Volumetric	±0.25% AF	10:1
Positive displacement flowmeters	Volumetric	±0.25% AF	10:1
Metering pumps	Volumetric	±0.5% FS	20:1
Mass flow meters	Mass	±0.2% AF	20:1
Vortex shedding flow meters	Volumetric	±0.75% AF	10:1
Level	Volumetric	±0.25% in.	NL

AF—Actual flow.
FS—Full scale.
NL—No limitation.

a full simulation of the sequential control functions (run using the as-designed operator interface) is an essential debugging aid for complex batch processes.

Batch Charging

Reactants and catalysts can be charged into the reactors, either by weight or by volume, using flow meters, pumps, feeders, and weighing systems, as shown in Table 8.8k. The main considerations in making the selection are accuracy and reliability of the data obtained. Because material balances are always based on the weight (not the volume) of the reactants, all things being equal, the mass- or weight-type sensors would be preferable; using such sensors would eliminate the need for density or temperature compensation. If they cannot be considered for some reason, the solid-state volumetric sensors (vortex) would be preferred over the ones with moving parts (turbines, pumps, positive displacement meters), because sensors with moving parts tend to be high-maintenance devices, with frequent need for recalibration.

Therefore, in a new installation with no restrictions on hardware selection, it might be advisable to place the reactor on load cells and charge all ingredients that weigh more than 10% of the total batch by weight. Reactants and additives that are needed in lesser quantities can be added under the control of Coriolis mass flowmeter-based measurements with the load cells providing only backup. Catalysts or ingredients needed in extremely small quantities, representing less than 1% of the total batch, can be added by metering pumps or by specialized feeders.

UNIT OPERATIONS CONTROLLERS

In the last decade, manufacturers of shared, distributed controllers (which started out as microprocessor-based multiple PID controllers) have added more and more logic capability to their products. At the same time, manufacturers of PLCs (which started out as microprocessor-based programmable logic controllers) added more and more PID capability to their products. The result today is the merging of these products into devices that are capable of handling both analog and logic sequencing control tasks. Although the control industry does not yet speak in these terms, a new era of process control has arrived: the era of the “unit operations controller.”

The unit controller is capable of controlling a unit operation in the plant. That unit operation can be a reactor, a distillation tower, a compressor, or any other subsystem. This represents a major step forward because it makes it necessary to stop thinking in terms of controlling single loops—such as pressures, flows, or temperatures—and to start thinking in multivariable terms, controlling the overall unit operation.

This is a prudent and logical direction to take because plants do not produce and sell pressures, flows, or temperatures: they sell a product. Therefore, the truly relevant control variable is maximum productivity. To move from the age of single to multivariable mentality, the unit controller is required. The microprocessor is needed to memorize the complex nature of the unit process in order for control to take place on this new and higher level.

The chemical reactor is a natural candidate for control by unit controller. The number and types of input and output signals required and the types of control actions to be performed on them are within the capabilities of many microprocessor-based multiloop products on today’s market. Therefore, the user can implement right now all that is discussed in this section and in the previous section.

Today’s process control industry is somewhat software-limited. A large variety of well-designed “empty boxes” are available, which are provided with a library of low-level software subroutines for PID, logic, and other functions. Unfortunately, with few exceptions, it is up to the user to educate this box to become a reactor controller or some other unit operation controller. Depending on the manufacturer, these empty boxes are more or less “friendly” to the user.

High-level software packages will be developed in the future to give “personality” to the previously empty boxes. One plug-in cassette or floppy disk might educate an empty box to become a distillation tower controller, and other disks could give it the personality of a dryer, evaporator, reactor, or other controller. It goes without saying that these high-level programs would also require adjustment to fit the universal unit controller to the particular process, but this “custom fit” would occur at a much higher level.

All user inputs could be provided by process engineers. These inputs could deal with recipes, temperature profiles, equipment or piping configuration, and so on. While such high-level, all-purpose packages are not yet available, flexible and reconfigurable reactor control packages (available with or without the reactor equipment) have been developed. One such reactor unit controller is described in the following paragraphs.

REACTOR UNIT CONTROLLER

Reactor control systems can be distinguished based on many criteria. One of these is flexibility. If a reactor control system is specifically designed for a particular product to be made in a particular unit, the control system has no flexibility at all. Flexibility can be built into the controls, so that they might be used even if operating condition, raw material, or product specifications change ("static flexibility"). A higher level of flexibility is one where the control system can be used to make different products and to maintain different operating conditions ("dynamic flexibility").

If the equipment configuration and the associated piping connections can also be changed, this is called structural flexibility in a control system. In plants where products change often, such as in pharmaceutical plants, flexibility in the reactor control is an important consideration, because it allows for the implementation of new production strategies without loss of time or need for additional engineering design.

Figure 8.8l shows a reactor control system that is provided with some structural flexibility. The control system is basically a unit controller, which is initially "empty" in the sense that it has a certain I/O, display, and memory capability. That capability is not utilized by the supplier to configure a single reactor control system (nonflexible system) but rather is used to configure a variety of control systems, which can be selected by the operator as a function of the desired equipment and operations.

Figure 8.8l illustrates how a flexible reactor control system might be reconfigured into different control structures. In the "nonflexible" (dedicated) reactor control systems, the operator would only adjust the controller set points, tuning parameters, and the quantities of ingredients in the recipes.

With a flexible reactor unit controller, the same reactor might be used for reaction, stripping or distillation, or crystallization, and for each operation a different control configuration and different tuning constants are needed. Because of the many operations taking place within the same reactor, the sensors must have wide rangeability and the controllers must be adapted for variable gain operation, because the time constants (dynamics) of the process changes with time.

The equipment accompanying this reactor unit controller is shown in Figure 8.8m.[3] The most important piece of equipment is the reactor itself. In addition, the following equipment modules exist and will be described in the following paragraphs:

1. Jacket system, with pump, on/off and control valves, and steam ejector
2. Distillation system, consisting of reflux cooler, condenser, water separator, and receiving tanks
3. Charging system, with weighing equipment and other instruments
4. Feeding system, consisting of a feeding tank, a dosing pump, and control devices

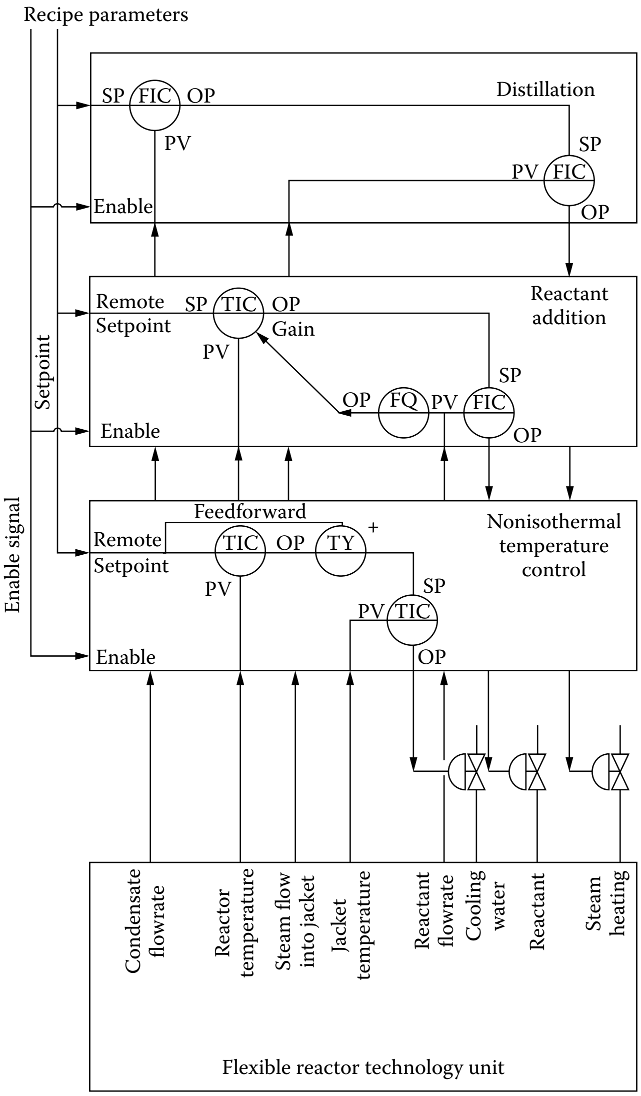


FIG. 8.8l
Illustration of the flexibility that a unit controller can provide.

5. Auxiliary systems, with equipment to perform operations such as clean in place, material transfer, inertization, and so on

Each equipment module is provided with the measurements necessary for operations and control.

Jacket Equipment Module

In addition to the jacket itself, the jacket equipment module consists of a pump, a steam ejector, two control valves (TCV-1 and -2), and ten on/off valves (V1 to V10). The various options on the operation of the jacket controls are described in Section 8.9. The jacket system can be in several states, as shown by the state diagram in Figure 8.8n.

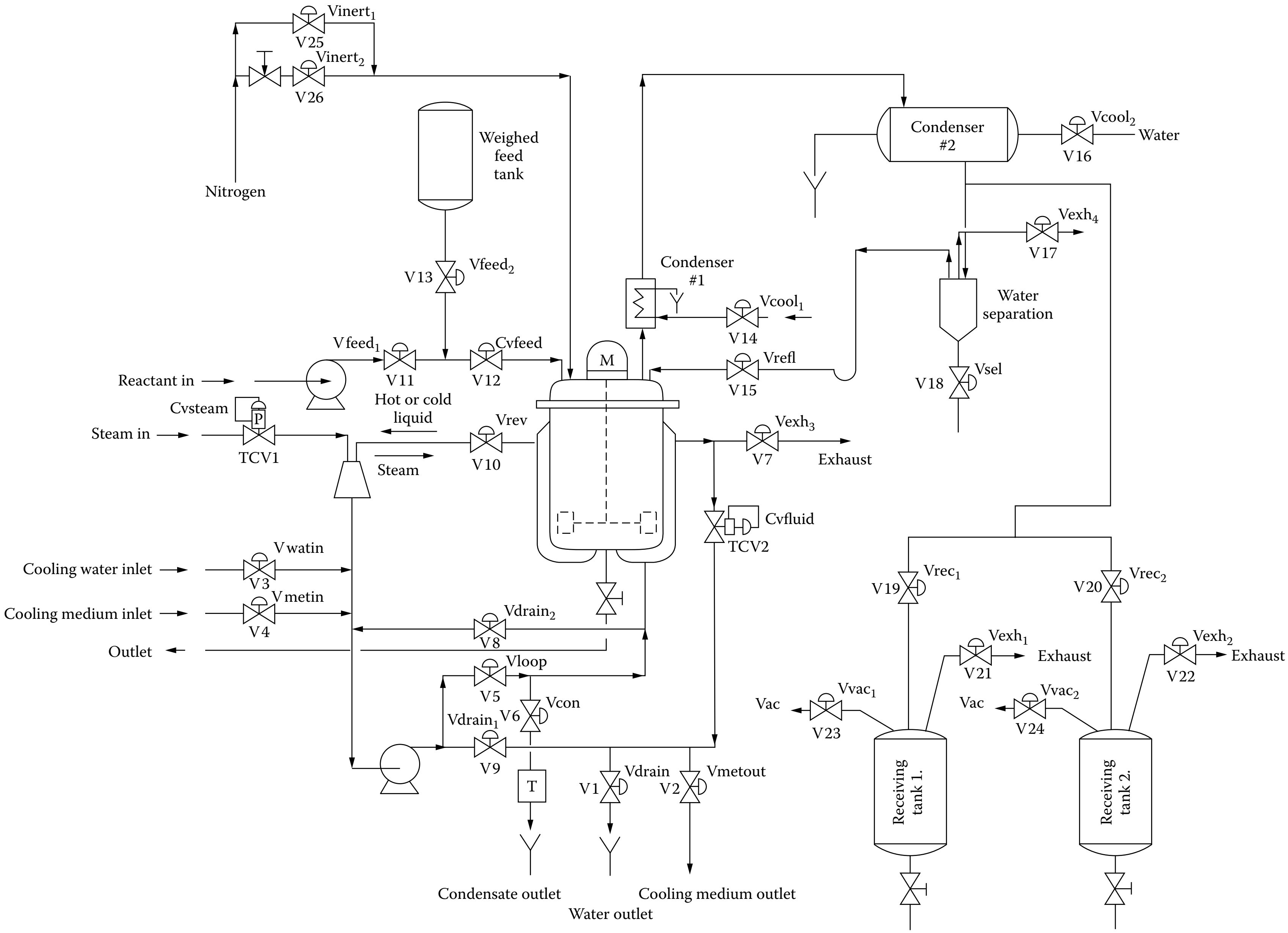


FIG. 8.8m
Prepackaged and automated reactor control system.[4]

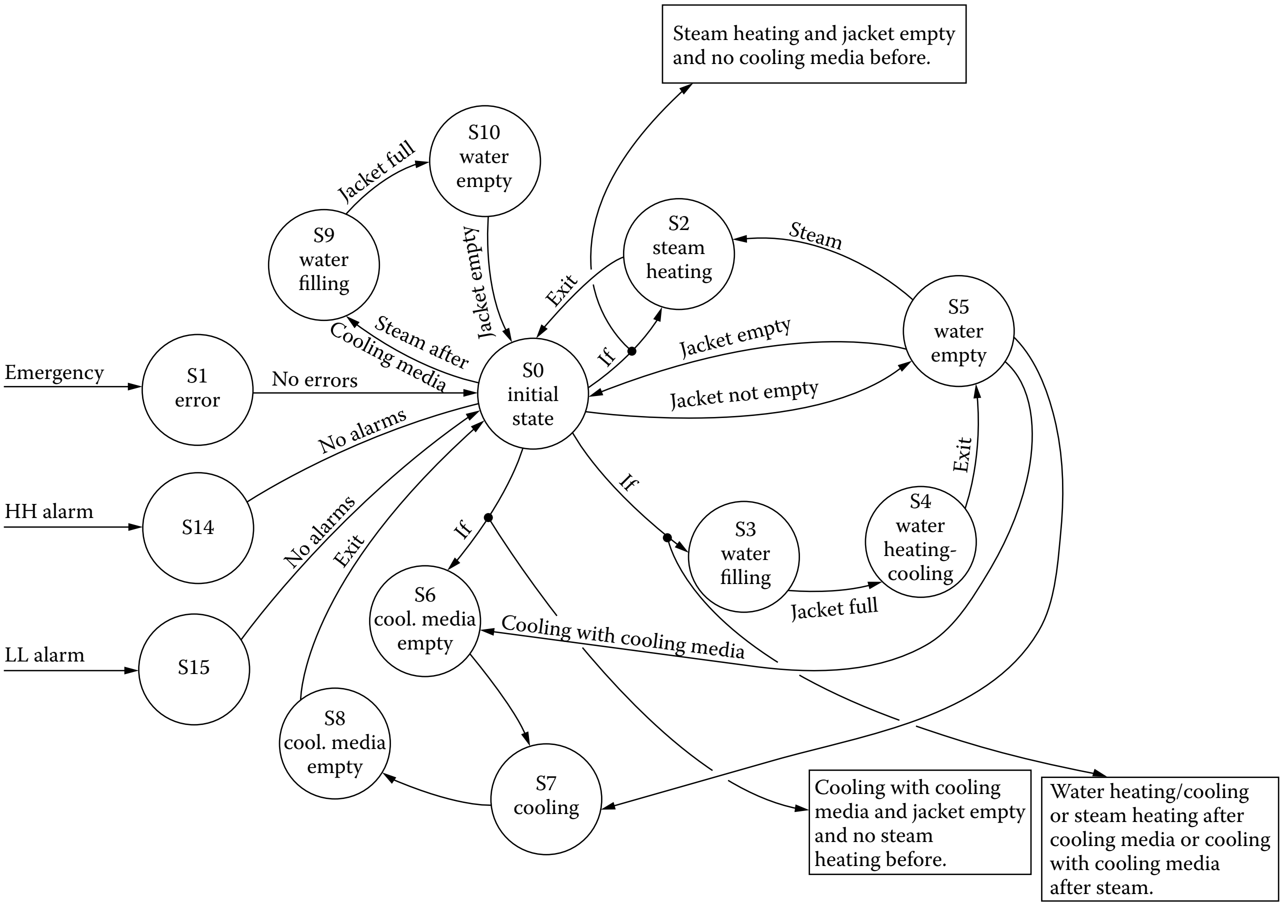


FIG. 8.8n
State diagram of a reactor jacket system.[4]

In state S0 all the valves are closed except the vent valve V7, which is open. If any of the safety interlocks are activated, hardware errors or device malfunctions are detected, and the system is returned to state S1, which is the "error state." In this state the control is returned to the operator. If the controlled process variables are not on set point and high or low alarms are registered, these bring the system into states S14 or S15, where those corrective responses are initiated that have been preprogrammed for such conditions.

When either the higher level of software or the operator is ready to initiate the steam heating state (S2), valves V10 and V6 are opened and TCV-1 is assigned to throttle the steam. In order to make sure that the steam condensate will not be contaminated, if prior to the steam heating state the jacket was filled with a cooling media (such as methanol or brine), the jackets will be first flushed with water (V3, V5, TCV-2, V1, and pump on), and then drained (V7, V8, V9, V1, and pump on).

Because of the multiple heat-transfer media used, automatic controls are provided not only to protect against cross-contamination but also to protect the glass lining of the reactor against damage by excessive temperature differences.

Therefore, if the particular batch sequence calls for consecutive stages that could thermally shock the glass lining (direct steam heating followed by cooling, or the other way around), the jacket interlocks automatically guarantee that intermediate temperature operations will be inserted, to make the transition gradual and thereby protect the glass.[4] Therefore, when direct steam heat is followed by cooling, states S9 and S10 are interposed to bring the jacket temperature to an intermediate value to avoid shocking the glass.

Prior to starting to heat or cool with water, the jacket is filled in state S3 (V3, V5, V7 opened, and pump on), and only when the jacket is full does the heating (V10, TCV-1 open, while V7 closes) or cooling (TCV-2 and V1 open) state (*S*4) start. When this is over, the water is drained from the jacket (V1, V8, V9, V19 open, pump on), and when the jacket is empty, it might return to the initial state (S0).

When a cooling media other than water is used (methanol or brine), the same sequence of filling (S6), cooling (S7), and draining (S8) states is initiated as was the case with the water-based operation. With these various control states, it becomes possible to match the desired batch temperature profile for a particular reaction with a profile of jacket temperatures that is reasonably close to it.[4]

Recipe Charging Module

The reactants, catalysts, solvents, modifiers, and other chemicals are charged into the reactor either before or during the reaction. The operation that takes place prior to starting the reaction is called charging, while the continuous addition during the reaction is often called feeding. The chemicals can be added on the basis of their weight or their volume, and the devices used in charging them include mass or volumetric flowmeters, metering pumps, feeders, or weighing systems.[5]

When weighing is applied, it can be done by weighing the reactor itself or by providing a separate weighed charge tank, as is the case in Figure 8.8m. The accuracy, reliability, and rangeability of the charging devices are the main criteria for selection, as was discussed in connection with Table 8.8k.

The ingredients that are fed while the reaction is in progress can be charged all at once (dumped), or they can be charged as a function of time, temperature, pressure, pH, or other variables. They are usually charged from a separate feeding tank under level or flow control or added by dosing pumps. In Figure 8.8l the on/off valves V11, V12, and V13 serve to control feeding. For an example of an installation in which the reactant is added under temperature control, refer to Figure 8.8o.

Stripping or Distillation Module

The addition of the distillation equipment module to the reactor provides the capability to perform the type of batch distillation that is illustrated in Figure 8.8p. The resulting quality of separation is naturally lower than if a distillation column is used. During this stripping phase the reactor is either in the hot water or in the direct steam heating mode, which guarantees a state of slow boiling in the reactor. During the stripping phase, valves V15, V16, and V17 are open. This state of slow boiling can be terminated on the basis of time or on the basis of the rise in vapor pressure in the reactor. If desired, water can be separated before the reflux is returned to the reactor.

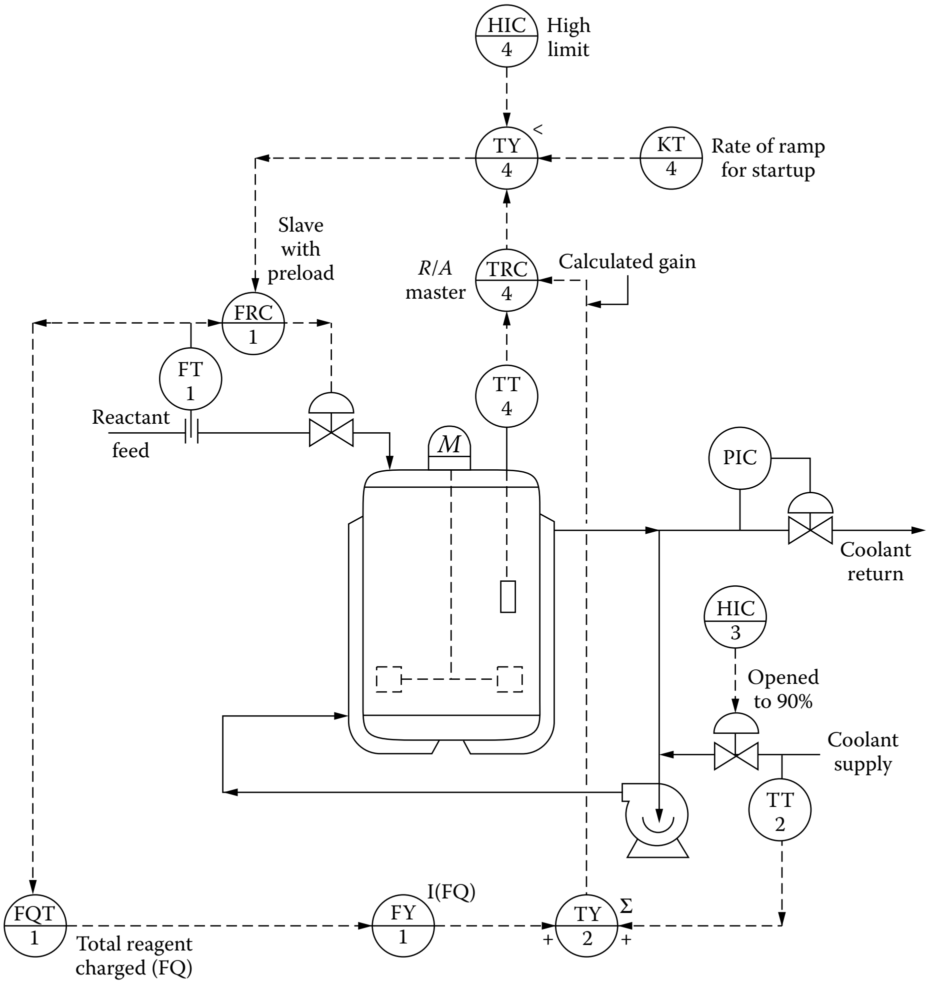


FIG. 8.8o
With constant coolant flow rate, the reactant flow can be the cascaded slave; because the process gain drops as the amount of reactants converted to products increases, the gain of TRC-4 is automatically increased as a function of total reactant charged.

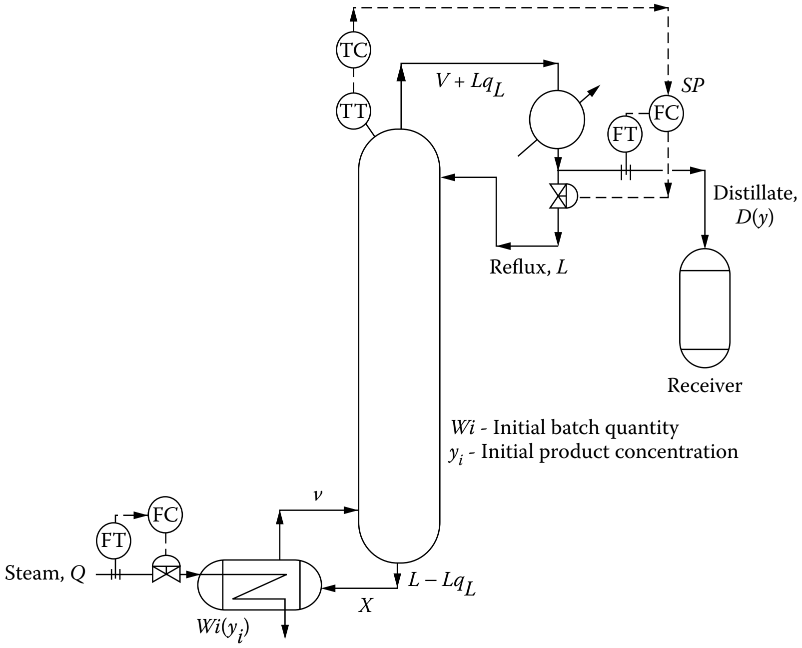


FIG. 8.8p
Batch distillation.

When stripping is to take place at atmospheric pressure, the valves V16, V19, V20, V21, and V22 are opened. If reflux is needed, V14 is also opened. The heat input into the reactor is limited by the cooling capacity of the condensers and by the foaming characteristics of the batch. If the stripping is to take place at a temperature lower than the boiling point of the solvent, the process is operated under vacuum, by closing V21 and V22 and by opening V24. The states of the other valves remain unaltered.

When more than one chemical is to be stripped off and separated, the distillation configuration is similar except that during the early phase of stripping V19 is held open (when the first chemical has been removed, V19 is closed and V20 opens). The switching of the valves is based on endpoint analyzer control, which is discussed in Section 8.9 (Chemical Reactors: Control and Optimization).

Cleaning and Interting Modules

In the food, pharmaceutical, and fine chemicals industries, it is important to clean the reactor equipment before or after each batch. Cleaning in place is usually done using the appropriate solvent, which is applied through jets that reach all parts and surfaces that require cleaning. While CIP is effective, it cannot meet all the requirements of all industries.

Figure 8.8m also shows the components needed to create an inert atmosphere in the reactor. The inertization sequence can be started by purging the vapor space with nitrogen (V25 open) until the oxygen content of the exhaust drops below the required limit. The purge phase can be terminated on the basis of time or by the use of on-line oxygen analyzers (see Section 8.42 in the first volume of this handbook).

During the reaction the nitrogen blanket is by V26 (V25 closed), which admits the continuous flow of purge nitrogen that sweeps the vapor space; if only the reactor is to be intertized, the nitrogen leaves through V17.1f other components also need to be intertized, the nitrogen might leave through V21 or V22.[4]

Recipe Controls

In the reactor process described in Figure 8.8m, the recipe formula and procedure contain all of the instructions needed to produce a particular product. Consistent with collapsibility of the procedural control model defined in S88.01, the procedure level has been eliminated in this example because it is not needed in a single-unit recipe, and the operation level (which is an optional convenience used for bookmarking purposes) was also eliminated. The recipe procedural control for a particular product thus consists of a single unit procedure and its sequentially executed phases, which provide parameter values (formula) for the reactor's equipment phases (Figure 8.8q).

Each unit procedure can be broken down into a number of recipe phases, which are building blocks linking the equipment phases to such parameters as time, temperature, or other quantities or variables. Each recipe phase corresponds to a particular equipment phase presented in a form that is suitable for automatic execution by the system. The actual building of a recipe from a sequence of phases is illustrated in Figure 8.8r. When an equipment phase is linked to a particular phase of a recipe, it is also provided with the associated parameters and endpoints, which serves the completion of the phase and the need to transfer to the next.

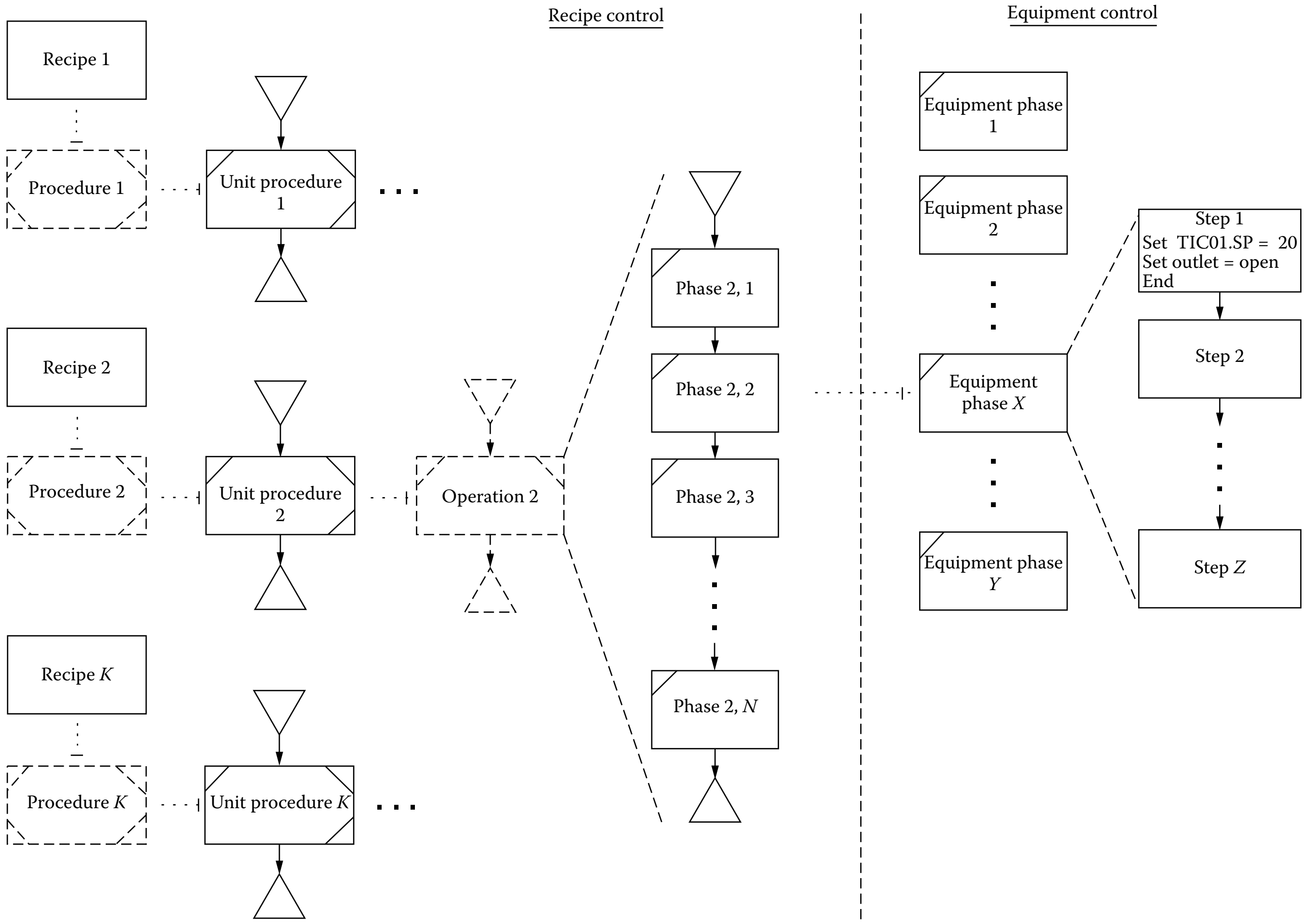


FIG. 8.8q
Building blocks of the unit controller recipe (dashed borders indicate implicit elements of S88 model).

Some of the more complex phases are made up of several steps, while the individual steps, in this case, are made up from structured text statements. The execution of each recipe is controlled by the coordination control software, which serves to organize the execution of the recipe. This function is illustrated in Figure 8.8s.

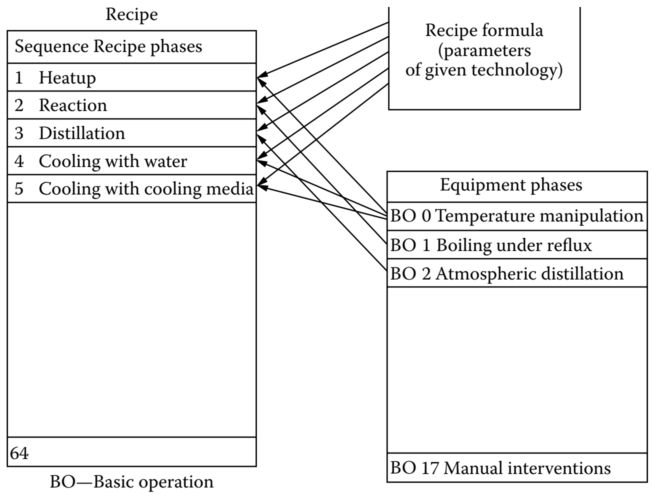


FIG. 8.8r
Recipe building.

Equipment Phases

This unit controller utilizes the fact that most processes consist of the same basic operations—only the set point and the parameters vary. Therefore, the equipment operations can be separated into the algorithm, which does not change from recipe to recipe, and the variable part, which contains the parameters and settings that are adjusted by the recipe. The larger the variable portion of a phase, the more flexible its operation but also the more engineering work will be needed for the application. The use of phases provides for the reuse of canned knowledge and for the efficient building of recipes.

Some of the equipment phases configured for the chemical batch reactor unit controller are listed in Table 8.8t. The parameters that need to be set prior to or during the phase include the set points at the beginning and during the phase, the controller mode (auto, manual, cascade) and the tuning settings, the operation times, alarm settings (LL, L, H, HH),

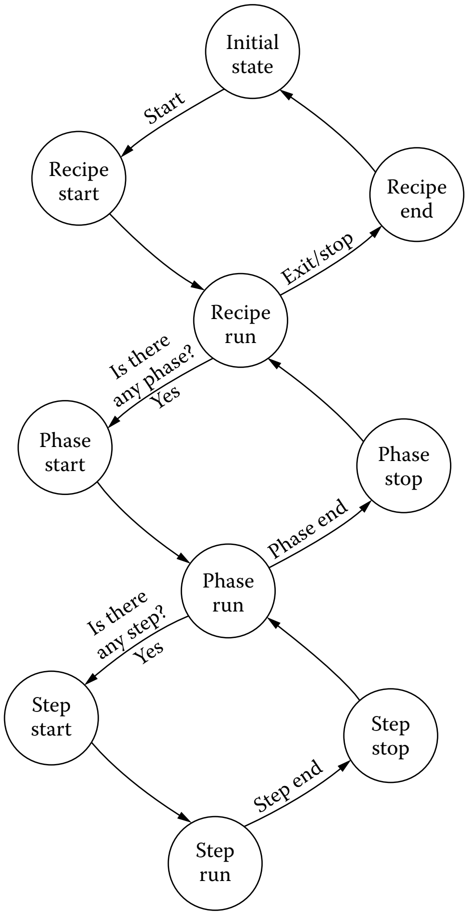


FIG. 8.8s
One function of the custom coordination control software in the example "unit controller" is to sequentially execute both the phases comprising each unit procedure and the steps comprising each phase.

the desired actions initiated by the alarms, and the conditions that must prevail before the process can be transferred to the next phase.

One of the critical equipment phases involves temperature control. A temperature-related equipment phase must be parameterized to define the associated jacket mode (steam or water heating, water or methanol cooling), the required controller configuration (slave or cascade), maximum allowable temperature difference between batch and jacket temperatures, and so on.

In some processes the temperature controls are reconfigured when the reaction starts, and therefore two equipment phases might be used. When the phase is for stripping or distillation, additional parameters need to be set for the jacket operation (water or steam heat), for temperature limits, and for receiver tank selection and switching.

The phases that describe charging are provided with parameters to state the required quantities, rates, and sequences. Similarly, the phases that define feeding must state such parameters as the jacket mode (steam or water heating, water or methanol cooling), the safety limits for the jacket medium, preload settings (flow rates at the start of the phase), maximum limits on flow, total quantity of reactants to be fed, and controller configuration.

TABLE 8.8t
Equipment Phases (EP) Available in a Flexible Read Package[4]

Designation	*Name of Equipment Phase*
EP 0	Temperature manipulations
EP 1	Boiling under reflux
EP 2	Distillation at atmospheric pressure
EP 3	Vacuum distillation
EP 4	Reflux with water separation
EP 5	Evaporation at atmospheric pressure
EP 6	Evaporation under vacuum
EP 7	Fractional distillation
EP 8	Steam distillation
EP 9	Feeding as a function of time
EP 10	Feeding as a function of temperature
EP 11	Feeding as a function of pH
EP 12	Cleaning in place
EP 13	Forwarding under suction
EP 14	Forwarding under pressure
EP 15	Inertization
EP 16	Initial charging
EP 17	Manual interventions (dummy operation)

The basic operation of reactor inertization operates parallel with other operations and also has parameters such as jacket modes, number of rinsing cycles, start points, and endpoints.[4] To accommodate this while maintaining a simple linear recipe flow, the inertization sequencing can be programmed as part of the equipment module's basic functionality, which is merely set up and initiated by the equipment phase execution.

In general there are two methods of setting up equipment phases. They can be simple (valve opens/closes, motor starts/stops) or complex, where the process is divided into larger, more complicated operations such as feeding, distillation, and crystallization. The above example illustrates the second, more complex approach, where the manufacturer of the reactor package has already designed and checked out the details, and therefore the uncertainties associated with a development project are eliminated. Such predesigned packages also reduce installation time and engineering design costs, while they also allow for fast changeover between products and fast implementation of new developments.

Component or Device Control

The lowest level of reactor automation is the control of individual devices. This includes continuous regulatory control functions, discrete control functions, safety and permissive interlocks, time-sequencing controls, and the like.

The continuous regulatory control functions of batch processes are similar to those of continuous processes; however, batch processes have additional requirements: for handling transients (e.g., heat-up without overshoot), for adaptation to varying process dynamics, and for operating against constraints.

In DCS systems the fill-in-the-blank-type language supported by CRT-based operator consoles became an accepted standard. The fill-in-the-blank language is used to configure the control system database and to create many of the necessary control algorithms (Table 8.8u).

In industrial controllers continuous recording charts are also frequently used. One possible cascade configuration for batch application is shown in Figure 8.8v. Other manufacturers approach the task of loop configuration by a statement list using function blocks as subroutines (Table 8.8u).

The design of discrete control functions (direct on/off control, interlock control, sequence control) first requires the definition of the logic using ladder diagrams, flowcharts, decision tables, function charts, function plans, time-sequence diagrams, or logic diagrams.

Once defined, the logic can be entered into the controller package by programming or by configuring. Methods describing discrete control functions are theoretically equivalent, but they each have different application areas. For example, ladder diagrams are preferred by maintenance engineers, and function charts are preferred by control engineers.

TABLE 8.8u
*Programming an Auto-Manual PID Controller by Answering a Series of Questions**

	Step		
No.	*Description*	*Parameter*	*Value*
1.	Engineering Units 0% Value	EU 0%	–99999 to 99999 (E.U.)
2.	Engineering Units 100% Value	EU 100%	–99999 to 99999 (E.U.)
3.	Gain	GAIN	0 to 128
5.	Reset	RESET	0 to 293 (repeats/minute)
6.	Rate	RATE	0 to 895 (minutes)
8.	Process Variable Filter Time Constant	PV FTIM	0 to 1.8 (minutes)
9.	Reverse Control Action	REV ACT?	YES or NO
10.	Increase Closes	INC CLO?	YES or NO
11.	Alarm A Trip Point	ALM A TR	0 to 136% of E.U. Span (E.U.)
12.	Alarm B Trip Point	ALM B TR	–16 to 136% of E.U. Span (E.U.)
13.	Alarm C Trip Point	ALM C TR	–16 to 136% of E.U. Span (E.U.)
14.	Alarm Dead Band	ALM DBND	0 to 136% of E.U. Span (E.U.)
16.	Setpoint Low Limit	SP LO LM	–13.97 to 113.97% of E.U. Span (E.U.)
17.	Setpoint High Limit	SP HI LM	–13.97 to 113.97% of E.U. Span (E.U.)
18.	Valve Output Low Limit	VO LO LM	–13.97 to 113.97(%)
19.	Valve Output High Limit	VO HI LM	–13.97 to 113.97(%)
20.	Anti-Reset Windup Low Limit	ARW LO LM	–13.97 to 113.97(%)
21.	Anti-Reset Windup High Limit	ARW HI LM	–13.97 to 113.97(%)
22.	Feedforward Gain	FF GAIN	0 to 1
23.	Feedforward Reverse Action	FF REV?	YES or NO
24.	Feedforward Filter Time Constant	FF FTIM	0 to 112 (minutes)
25.	Track Filter Time Constant	TK FTIM	0 to 112 (minutes)
28.	Restart Valve Output	RST MD	1 to 5
29.	Restart Valve Output	RST VO	–13.97 to 113.97(%)
30.	Restart Set Point	RST SP	–13.97 to 113.97% of E.U. Span (E.U.)

*Courtesy of Honeywell Inc.

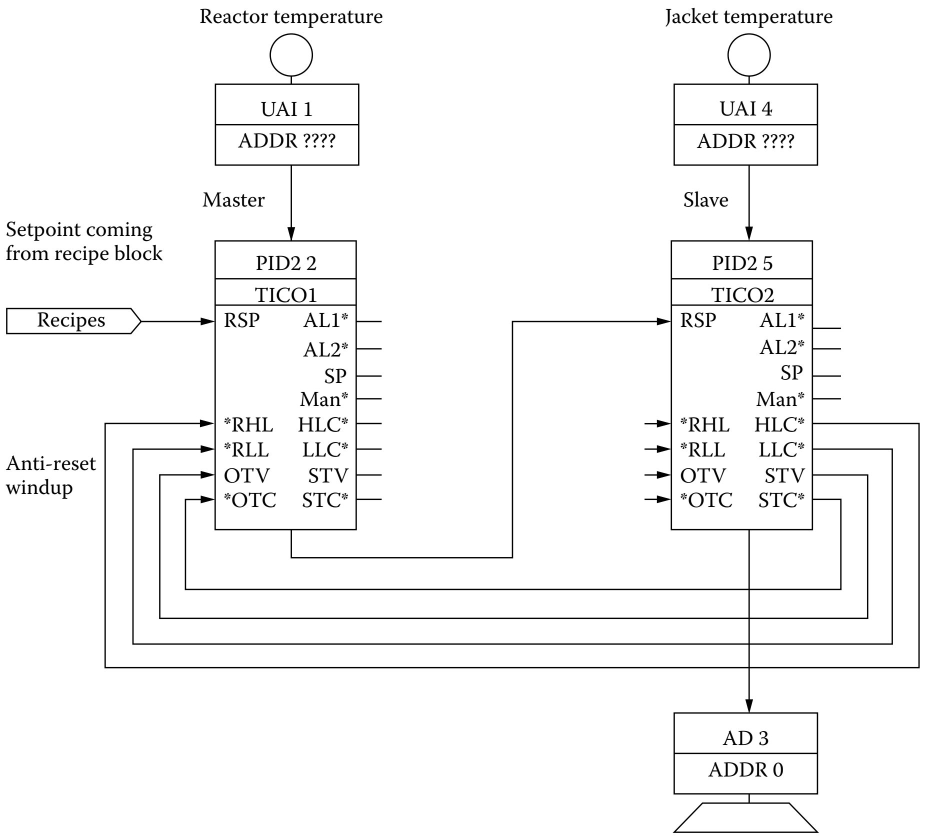


FIG. 8.8v

Building continuous regulatory functions on the continuous control chart.

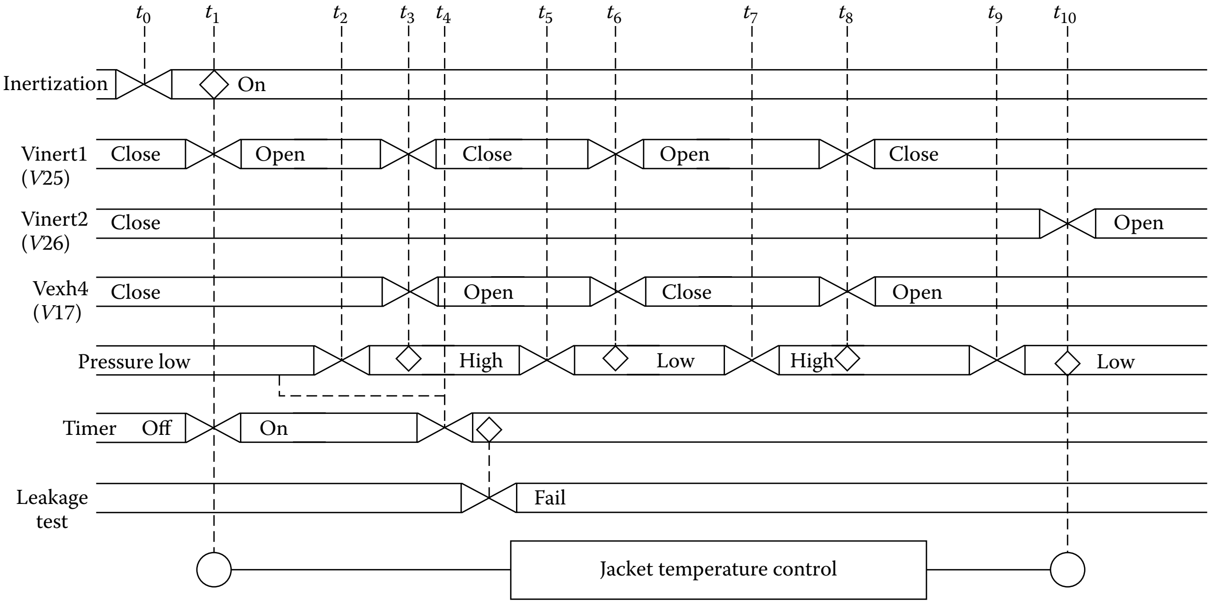


FIG. 8.8w

Time-sequence diagram for reactor inertization.

In high-performance process control, high-level control languages are now standard.

As was discussed earlier, for batch processes, time-sequence diagrams have an important role. The time-sequence diagram shows the sequential changes in control actions and their time relationships. The sequence of events can be followed along dotted lines for vertical time coincidence and along the solid lines in the horizontal direction for sequence control. The diamond symbol is used to indicate a trigger event, and a vertical dotted line indicates the time coincidence of trigger events. When two or more diamonds occur in the same relative time line, an "AND" logic condition results.

As an example, the time-sequence diagram for inertization in Figure 8.8w shows the basic operation of inertization. When the operation is started, valve V25 is opened and the timer is started. If the pressure in the reactor does not reach the given value (e.g., 2 bar) by the preset time, a leakage is assumed and the system goes into error state. In the normal sequence the system has two cycles of rinsing (t_1 to t_6 and t_6 to t_{10}), and after completion of these cycles all valves except V26 and V17 will close and inertization continues at a predetermined constant flow of inert gas.

Safety interlocks are designed to ensure the safety of operating personnel and to protect plant equipment and environment. These types of interlocks differ from ordinary process interlocks, in that they detect equipment malfunction or shutdown. These interlock functions are frequently implemented on a separate set of control equipment.

CONCLUSIONS

Many batch reactors must consistently execute complex product-dependent sequences of discrete events to produce a variety of products through multiple unit operations over a broad range of operating conditions and compositions, often in regulated industries that demand a detailed accounting of each step. This requires a very flexible and intricate control system.

The tasks associated with designing modular sequential control strategies for such an application and several reactor sequence control examples have been discussed in this section, based on the batch control standards developed by ISA and approved by ANSI and IEC. Batch server packages based on these standards greatly facilitate and enhance high-end coordination control capabilities, while design efficiency is maximized by good modularization practices and reuse of flexible previously engineered control. Because the role of the operator is especially important in batch plants, the project's overall business success often depends upon such things as adequately addressing operator interface and exception handling needs and on adequately defining and enforcing the operator's responsibilities.

References

1. Bristol, E. H., "A Design Tool Kit for Batch Process Control," *InTech,* October 1985.
2. ISA SP88, "Batch Control Part 1: Models and Terminology," ANSI/ISA-S88.01, Research Triangle Park, NC: ISA, October 1995 (adopted as IEC 61512-1, 1997).
3. Molnar, F., Szuk, G., and Zeller, I., "Modular Batch Automation System for Pharmaceutical Applications," (in Hungarian), Magyar Kemikusok Lapja, 1991.
4. "Chemiflex Reactor User's Manual," (in Hungarian), Chinoin Pharmaceutical and Chemical Works, Budapest, 1991.
5. Lipták, B. G., *Optimization of Unit Operations,* Radnor, PA: Chilton Book Company, 1987.

Bibliography

Alsup, M., "Recipe Configuration for Flexible Batch Systems," ISA/93 Technical Conference, Chicago, IL, September 19–24, 1993.

"Batch Control Systems," ISA Transactions, Vol. 28, No. 3, 1989.

Brandl, D., "General to Master," *InTech,* Vol. 51, No. 4, 2004.

Crowl, T. E., "S88.01 Concepts Streamline Control Software Application for Biotech Plant," *Proceedings of ISA TECH Conference,* Anaheim, CA, October 1997.

Dove, A. P., and Deitz, D., "Applying Object-Oriented Design Patterns to Control Strategies," *Proceedings of ISA TECH Conference,* Anaheim, CA, October 1997.

Faccenda, J. F., and Baker, T. E., "An Integrated AI/OR Approach to Blocked Operations Scheduling," *Proceedings of the NPRA Computer Conference,* Seattle, WA, October 1990.

Feldmeyer, E., Alsup, M., and Burns, H., "Recipe Interface Requirements for Flexible Batch Control Systems," *Proceedings of the Industrial Computing Conference,* Vol. 1, Instrument Society of America, Anaheim, CA, October 1991.

Fisher, T. G., *Batch Control Systems: Design, Application, and Implementation,* Research Triangle Park, NC: Instrument Society of America, 1990.

Fleming, D. W., and Pillai, V., *S88 Implementation Guide*, New York: McGraw-Hill, 1999.

Ghosh, A., "An Approach to a Flexible and Easy-to-Use Batch Control System," *Proceedings of the Industrial Computing Conference,* Vol. 1, Instrument Society of America, Anaheim, CA, October 1991.

Hedrick, J. L., and Severns, G., "Planning Control Methods for Batch Processes," *Chemical Engineering,* April 18, 1983, pp. 69–78.

Hogenson, D. C., "Automatic Scheduling of Batch Operations," *In Tech,* April 1990, pp. 30–32.

Hornbeck, D., "A New Approach to Batch Process Control," *Proceedings of the Industrial Computing Conference,* Vol. 1. Instrument Society of America, Anaheim, CA, October 1991.

ISA SP88, "Batch Control Part 2: Data Structures and Guidelines for Languages," ANSI/ISA-S88.00.02, Research Triangle Park, NC: ISA, February 2001 (adopted as IEC 61512-2, 2001).

ISA SP88, "Batch Control Part 3: General and Site Recipe Models and Representation," ANSI/ISA-S88.00.03, Research Triangle Park, NC: ISA, March 2003.

ISA SP88, "Batch Control Part 4: Production Records," Editor's Draft 4, ISA-ds88.00.04, Research Triangle Park, NC: ISA, March 2004.

ISA SP88, "Batch Control Part 5: Recipe Equipment Interface – Procedural Elements," Editor's Draft 1, ISA-ds88.00.05, Research Triangle Park, NC: ISA, March 2004.

Juba, M. R., and Hamer, J. W., "Progress and Challenges in Batch Process Control," Proceedings of the CPC III, CACHE, Amsterdam, 1986.

Kompass, E. J., Witlock, S. K., and Williams, T. J., "Generic for Batch Manufacturing," *Proceedings of the Sixteenth Annual Advanced Control Conference,* West Lafayette, IN, September 1990.

Larson, K., "Flexibility Takes Center Stage in Batch Process Automation," *Control,* November 1991, pp. 24–33.

Lipták, B. G., *Optimization of Unit Operations,* Radnor, PA: Chilton Book Company, 1987.

Massey, L. E., "The Development of Modular Batch Automation Techniques," *Proceedings of the Industrial Computing Conference,* Vol.1, Instrument Society of America, Anaheim, CA, October 1991.

Molnár, F., and Nagy, T., "Flexible Autoclave Process Control System, Based on Micro TDC 3000," Honeywell Application Notes, Brussels, 1991.

Molnár, F., and Nagy, T., "Flexible Reactor Controller in the Pharmaceutical Industry," *Meres es Automatika*, Vol. 46, 1991.

Morari, M., "Flexibility and Resiliency of Process Systems," *Comp Chem. Eng*., Vol. 7, 1983, pp. 423–437.

Muller-Heinzerling, T., "Standard Software for Batch Processes with TELEPERM M-BATCH TM." Seminar on INTERKAMA, Dusseldorf, Germany, 1989.

Nowicki, P. L., "OOP & Expert Systems Transform Batch Recipes," *InTech,* September 1992.

Nowicki, P. L., "Transforming Recipes is the Key to Flexible Batch Processing," *Proceedings of the ISA/91,* Vol. 46, Part 2, Instrument Society of America, Anaheim, CA, October 1991.

Parshall, J., and Lamb, L., *Applying S88: Batch Control from a User's Perspective,* Research Triangle Park, NC: ISA, 2000.

Pillai, V., "Good Practices for S88 Modularization," *Control,* January 1999.

Pillai, V., and Korkmaz, B., "How to Implement Unit Supervision," *Control,* August 1998.

Pillai, V., "S88 Phase Fundamentals," *Control,* February 1999.

Rosenof, H. P., and Ghosh, A., *Batch Process Automation,* New York: Van Nostrand Reinhold, 1987.

Szeifert, F., Chovan, T., and Nagy, L., "Dynamic Simulation and Control of Flexible Chemical Technologies," Automation '92, Budapest, 1992.

"Taylor Mod 300: The System of Choice for Flexible Batch," *Combustion Engineering*, Rochester, NY, 1989.

Tom, T. H., "DCS Selection for Flexible Batch Automation," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

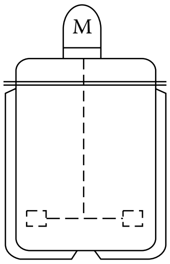
Vass, J., and Szeifert, F., "Adaptive Temperature Control of Chemical Batch Reactors," *Mereses Automatika,* Vol. 36, 1986.

WBF-WG Flow Analysis, "ISA-88 Modeling Using Flow Analysis," Editor's Draft 2, Draft Technical Report, World Batch Forum (WBF), February 2002.

Woolley, K., and Christie, D., "Achieving Increased Batch Plant Productivity by the Application of a Multilevel DCS Solution," *Proceedings of ISA EXPO Conference,* Houston, TX, October 1998.

8.9 Chemical Reactors: Basic Control Strategies

D. C. KENDALL, W. F. SCHLEGEL (1970) **H. I. HERTANU** (1985)
F. MOLNÁR (1995) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

INTRODUCTION

It should be noted that the coverage of this section and that of the next one (8.10) are similar. If you are interested in reading a detailed and complete treatment of the subject of chemical reactor control, read this section. On the other hand, if you have little time and are an experienced process control engineer, familiar with the basics of chemical reactor control, and want only to refresh your memory about the most important aspects of their control and optimization, read Section 8.10.

This section is started with a description of reactor characteristics, reaction rates, and time constants. This is followed by a discussion of the various methods of reactor temperature control, initial heat-up control, end-point detection, pressure, and safety controls. Other aspects of reactor control are covered in Sections 8.8, 8.10, and 8.11, covering the topics of batch sequencing, optimization, and modeling of chemical reactor controls.

Chemical reactor designs include the continuous stirred tank reactors (CSTRs), the batch stirred tank reactors (BSTRs), the tubular reactors, and the packed bed reactors. The optimization of batch and continuous chemical reactors has many potential benefits, including increase in productivity and improvement in safety, product quality, and batch-to-batch uniformity. The combined impact of these factors on plant productivity can approach a 25% improvement.[1]

Such overall results are the consequences of many individual control loops and control strategies. These loops will program temperature and pressure and maintain concentration and safety, while providing sequencing and record-keeping functions. All elements of the overall chemical reactor control system are discussed in this chapter.

REACTOR DESIGNS AND CHARACTERISTICS

In a batch cycle, there is no steady state and therefore no "normal" condition at which controllers could be tuned. The dynamics of the batch process vary with time; thus, the process variables, the process gains, and time constants also vary during the batch cycle. In addition, there are the problems of runaway reactions and batch-to-batch product uniformity.

Runaway reactions occur in exothermic reactions, in which an increase in temperature speeds up the reaction, which in turn releases more heat and raises the temperature further. In order to counter this positive feedback cycle, highly self-regulating cooling systems are required. One of the most self-regulating cooling systems is a bath of boiling water, because it needs no rise in temperature to increase its rate of heat transfer. Endothermic reactions are inherently self-regulating.

Batch-to-batch uniformity is a function of many factors, from the purity of reactants, catalysts, and additives to the repeatability of controllers serving to maintain heat and material balance. Before addressing such complex topics, it is necessary to review the basic batch process.

Most batch cycles are started by charging reactants into the reactor and then mixing and heating them until the reaction temperature is reached. The reaction itself is frequently started by the addition of a catalyst. *Exothermic reactions* produce heat, and *endothermic reactions* consume heat. The reactor itself can be *isothermal*, meaning that it is operated at constant temperature, or *adiabatic*, meaning that heat is neither added nor removed within the reactor; the reaction is controlled by other means, such as the manipulation of pressure, catalyst, and reactants.

Chemical reactions can follow quite complex paths and sequences, but for engineering purposes such as equipment design and control system analysis, most reactions can be considered as one of four types: irreversible, reversible, consecutive, or simultaneous. Most reactions are reversible — that is, there is a ratio in product-to-reactant concentration that brings about equilibrium. Under equilibrium conditions the production rate is zero, because for each molecule of product formed there is one that converts back to its reactant molecules.

The equilibrium constant (K) describes this state as the ratio of forward- to reverse-rate coefficients. The value of K is also a function of the reaction temperature and the type of catalyst used. K naturally places a limit on the conversion that can be achieved within a particular reactor, but conversion can usually be increased, if at least one of the following changes can be made:

- Reactant concentration can be increased
- Product concentration can be decreased through separation or withdrawal
- Temperature can be lowered by increased heat removal in reversible exothermic reactions
- A change in operating pressure can be affected (this increases conversion only in certain reactions)

The catalyst does not take part in the reaction, but it does affect the reaction rate (k). Some catalysts are solids and are packed in a bed; others are fluidized, dissolved, or suspended. Metal catalysts are frequently formed as flow-through screens. Whatever their shape, the effectiveness of the catalyst is a function of its active surface, because all reactions take place on that surface. When it is fouled, the catalyst must be reactivated or replaced.

The time profiles of heat release, operating temperature, and chemical concentrations are illustrated in Figure 8.9a for a consecutive reaction,[2] in which first ingredient A is converted into intermediate product B, and then intermediate product B reacts to form final product C. The reaction temperature is controlled so as to maximize the production of C while minimizing the cycle period.

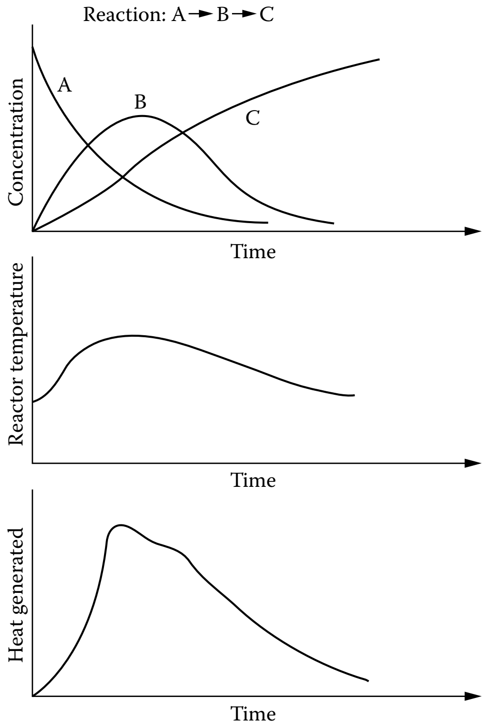


FIG. 8.9a

Concentration, temperature, and heat variables are shown as a function of time for a consecutive reaction, where A is first converted to B and then B into C. In this reaction the heat generated by the reaction is greater during the conversion of A to B, but the reactor temperature is controlled at an optimum setting to ensure maximum conversion in a minimum amount of time.[2]

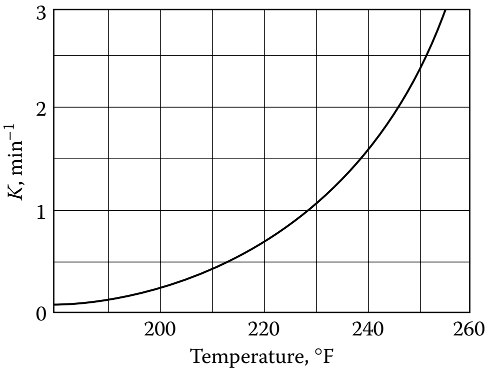


FIG. 8.9b

The influence of temperature on reaction rate coefficient is substantial.[3]

Reaction Rates and Kinetics

The reaction rate coefficient exponentially increases with temperature. The activation energy (E) determines its degree of temperature dependence[3] according to the Arrhenius equation:

$$k = \alpha e^{-(E/RT)} \qquad \textbf{8.9(1)}$$

where

k = specific reaction rate coefficient (min^{-1})
α = pre-exponential factor (min^{-1})
E = activation energy of reaction (BTU/mole)
R = perfect gas constant (1.99 BTU/mole·°R)
T = absolute temperature (°R)

Figure 8.9b illustrates the strong dependency of the reaction rate coefficient (k) on reaction temperature[3] for the values of $\alpha = e^{29}$ and $E/R = 20{,}000$.

Figure 8.9c illustrates the three basic reactor types: 1) plug flow, 2) continuous stirred tank, and 3) batch; it mathematically defines their fractional conversion of the reactant(s) into product (y).

Because the continuous plug flow-type reactor is dominated by dead time, its temperature control is difficult. On the other hand (as shown in Figure 8.9c), the plug flow reactor gives higher conversion than a back-mixed reactor operating under the same conditions. If the reaction rate is low, a long tubular reactor or a larger back-mixed reactor is required to achieve reasonable conversions.

In a batch reactor, after the initial charge there is no inflow or outflow. Therefore, an isothermal batch reactor is similar in its conversion characteristics to a plug-flow tubular reactor. If the residence times are similar, both reactors will

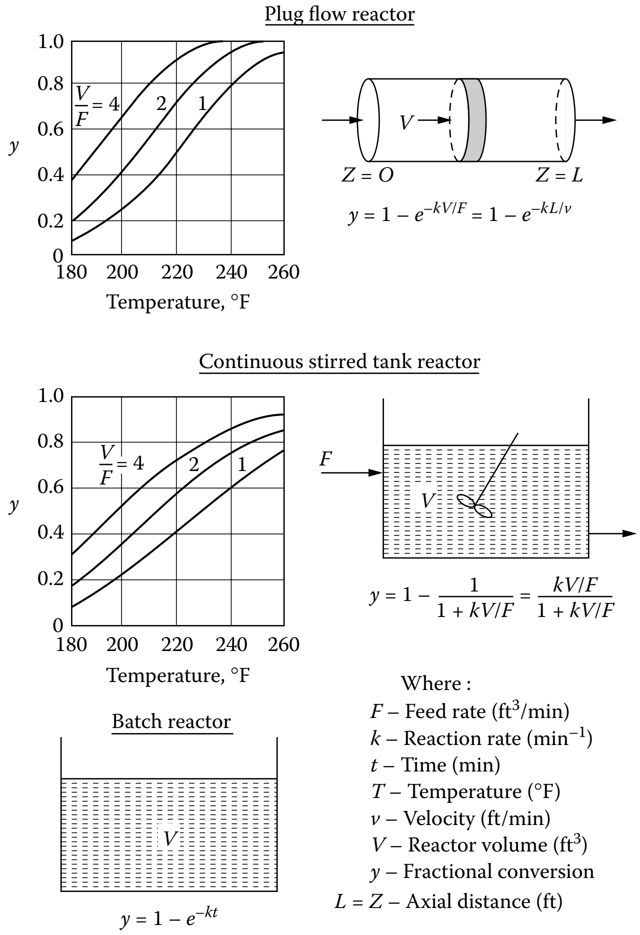


FIG. 8.9c

Conversion equations and conversion vs. temperature characteristics vary with reactor design.[2,3]

provide the same conversions. Batch reactors are usually selected when the reaction rates are low, when there are many steps in the process, when isolation is required for reasons of sterility or safety, when the materials involved are hard to handle, and when production rates are not high.

As shown in Figure 8.9d, the batch (or tubular) reactor is kinetically superior to the continuous stirred tank reactor.[2] The batch reactor has a smaller reaction time and can produce the same amount of product faster than the back-mixed one.

Reactor Time Constants

The amount of heat generated by an exothermic reactor increases as the reaction temperature rises. If the reactor is operated without a temperature controller (in an open loop), an increase in the reaction temperature will also increase heat removal, because of the increase in ΔT between process and coolant temperatures. If an increase in reaction temperature results in a greater increase in heat generation than in heat removal, the process is said to display positive feedback; as such, it is considered to be "unstable in the open loop."

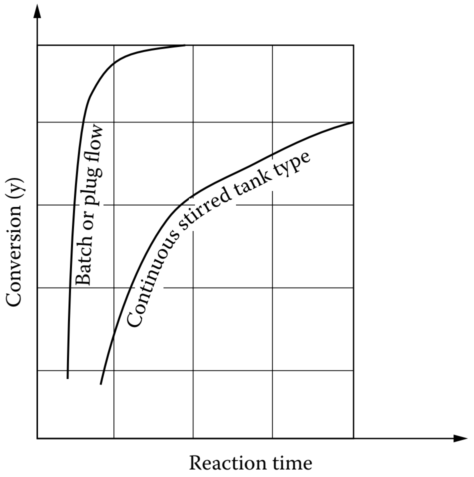
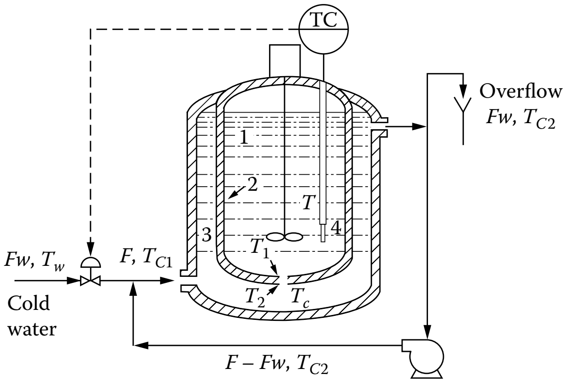


FIG. 8.9d

Batch reactors have better conversion efficiencies than back-mixed reactors.[2]

The positive feedback of the open-loop process can be compensated for by the negative feedback of a reactor temperature controller, which will increase the heat removal rate as the temperature rises. The addition of such a feedback controller can stabilize an open-loop unstable process only if the control loop is fast and does not contain too much dead time. Cascade control can increase speed, and maximized coolant flow can reduce dead time. Shinskey[3] suggests that if the dead time can be kept under 35% of the thermal time constant of the reactor, the process can be stabilized, whereas if it approaches 100% the reactor will not be controllable.

A real reactor has several lags and delays, including those of measurement and heat removal, as illustrated in Figure 8.9e.



The process gain: $K_P = \dfrac{T_{C2} - T_w}{Fw}$

FIG. 8.9e

There are four interacting time lags in a chemical reactor process. By minimizing the process gain, the controller gain can be maximized, and the narrower the proportional band the more sensitive the control loop will be.[3]

The equations for calculating the thermal, reactor wall, coolant, and thermal bulb time delays are listed below.[3] Typical values of these time constants are:

τ_1 = thermal time constant = 30–60 min
τ_2 = reactor wall time constant = 0.5–1.0 min
τ_3 = coolant time constant = 2–5.0 min
τ_4 = thermal bulb time constant = 0.1–0.5 min (can be minimized by the use of bare bulbs)

$$\text{Thermal time constant: } \tau_1 = \frac{W_1C_1}{k_1A} = \frac{W_1C_1}{Q}(T - T_1) \quad 8.9(2)$$

$$\text{Reactor wall time constant: } \tau_2 = \frac{W_2C_2l}{k_2A} = \frac{W_2C_2}{Q}(T_1 - T_2) \quad 8.9(3)$$

$$\text{Coolant time constant: } \tau_3 = \frac{W_3C}{k_3A} = \frac{W_3C}{Q}(T_2 - T_c) \quad 8.9(4)$$

$$\text{Thermal bulb time constant: } \tau_4 = \frac{W_4C_4}{k_1A_4} \quad 8.9(5)$$

where

A = heat-transfer area, ft^2
A_4 = surface area of bulb, ft^2
C = specific heat of coolant, BTU/(lb)(°F)
C_1 = specific heat of reactants, BTU/(lb)(°F)
C_2 = specific heat of wall, BTU/(lb)(°F)
C_4 = specific heat of bulb, BTU/(lb)(°F)
k_1 = heat-transfer coefficient, BTU/(h)(ft^2)(°F)
k_2 = thermal conductivity, BTU/(h)(ft^2)(°F/in)
k_3 = heat-transfer coefficient, BTU/(h)(ft^2)(°F)
l = wall thickness, in.
Q = rate of heat evolution, BTU/h
T = reactor temperature, °F
T_1 = wall temperature, °F
T_2 = outside wall temperature, °F
T_c = average coolant temperature, °F
W_1 = weight of reactants, lb
W_2 = weight of wall, lb
W_3 = weight of jacket contents, lb
W_4 = weight of bulb, lb

The total dead time in the loop is the sum of jacket transport lag, the dead time due to imperfect mixing, and miscellaneous smaller contributing factors. The dead time due to jacket displacement can be reduced by increasing the pumping rate. This should be kept under 2 min in a well-designed reactor.[3] The dead time caused by imperfect mixing can be reduced by increasing the agitator pumping capacity. In a well-designed reactor it should be held to less than 10% of the thermal time constant τ_1.

In the case of a typical reactor, the period of oscillation might be around 30 min. This period approximately equals four dead times; therefore, the total dead time of such a loop is around 7.5 min. In the case of interacting controllers, the correct setting for such a controller would be 7.5 min for both integral and derivative times.

Two types of reactors are used in chemical plants: continuous reactors and batch reactors. Continuous reactors are designed to operate with constant feed rate, withdrawal of product, and removal or supply of heat. If properly controlled, the composition and temperature can be constant with respect to time and space. In batch reactors, measured quantities of reactants are charged in discrete quantities and allowed to react for a given time, under predetermined controlled conditions. In this case, composition is the function of time.

TEMPERATURE CONTROL

The control loop features required during heat-up are substantially different from those needed during an exothermic reaction or those required during stripping or refluxing. Each will be discussed in the following paragraphs, starting with the controls of exothermic reactors.

Reaction temperature is frequently selected as the controlled variable in reactor control. It may be necessary to control reaction rate, side reactions, distribution of side products, or polymer molecular weight and molecular weight distribution. All of these are sensitive to temperature. It is frequently necessary to control reaction temperature to within 0.5°F (0.28°C). Many reactions are exothermic. In order to control reaction temperature, the released heat must be removed from the system as it is liberated by the reactants.

A simple temperature control scheme is depicted in Figure 8.9f. The reaction temperature is sensed, and the flow of heat-transfer medium to the reactor jacket is manipulated. For many installations this scheme is considered to be unsatisfactory because of the reactor nonlinearity and dynamic features. This "once-through" method of cooling is undesirable

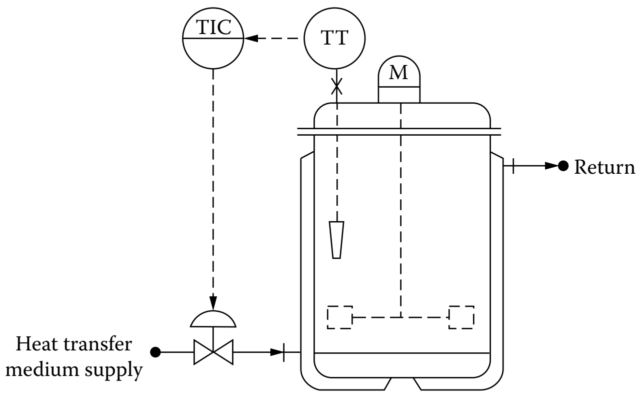


FIG. 8.9f
In chemical reactors with once-through cooling, the coolant temperature is not uniform and the process dead time varies with load.

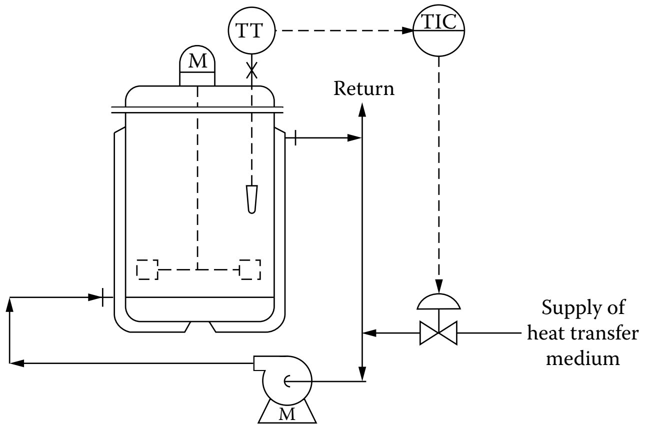


FIG. 8.9g
If the cooling water is recirculated around the jackets of chemical reactors, the water temperature will be more uniform, the process dead time will be minimized, and the heat transfer will be maximized.

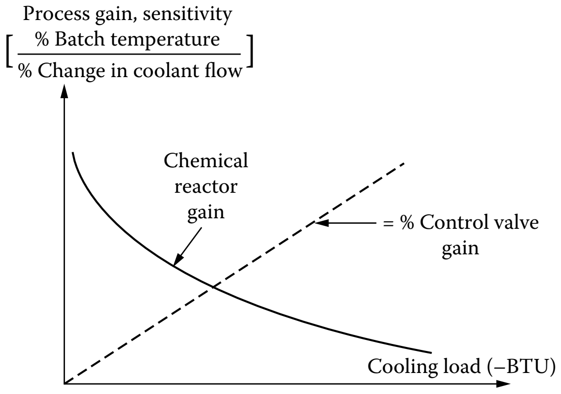


FIG. 8.9h
The sensitivity (gain) of a jacketed reactor drops in a nonlinear manner as the cooling load rises because the increasing amount of heat must be transferred through a fixed heat-transfer area. Because the gain of an equal-percentage coolant control valve linearly rises with load, using such a valve will partially compensate for the nonlinearity of the gain of a jacketed reactor.[6]

because the coolant temperature is not uniform. This can cause cold spots near the inlet and hot spots near the outlet.

Another disadvantage of this configuration is the variable residence time of the cooling water within the jacket as the flow rate changes. This causes the dead time of the jacket to vary, which in turn necessitates the modification of the control loop tuning constants as the load varies. In addition, when the water flow is low, the Reynolds number will drop off, and with it, the heat-transfer efficiency will also diminish. Low water velocity can also result in fouling of the heat-transfer surfaces.

For all the above reasons, the recirculated cooling water configuration shown in Figure 8.9g is more desirable, because it guarantees a constant and high rate of water circulation. This keeps the jacket dead time constant, the heat-transfer coefficient high, and the jacket temperature uniform, thereby eliminating cold and hot spots.

The fluid velocity in the reactor jacket is maintained high enough to produce satisfactory film coefficients for heat transfer. The fluid velocity can be further increased by additional jets. In addition, the liquid is circulated at a high enough rate to keep the temperature gradient in the heat-transfer medium, as it passes through the jacket, at a high enough level to maintain the jacket wall temperatures throughout the reactor. This keeps the jacket dead time constant and eliminates fouling of the heat-transfer surfaces.

Because the jacket provides a constant heat-transfer area, when the cooling load is low, the process is sensitive and the process gain is large. As shown in Figure 8.9h, as the load rises, the process gain drops in a nonlinear manner. The variable process gain can be partially compensated for by using a variable gain control value (equal-percentage valve); thereby, when the process gain drops, the valve gain rises and the total loop gain remains relatively constant.

Figure 8.9i illustrates the temperature response of any uncontrolled chemical reactor to a step change in load, assuming that the coolant is applied in a once-through manner (Figure 8.9f), without recirculation. The solid line depicts the temperature response at low loads, and the dotted line depicts temperature response at high loads.

Both dead time and the process gain increase as the load drops. In other words, at low loads it takes longer for the process to start responding, but once it has, the full response develops quickly. At high loads the opposite is the case.

Cascade Control

A superior method of reactor temperature control, a cascade loop, is depicted in Figure 8.9j. Here the controlled process

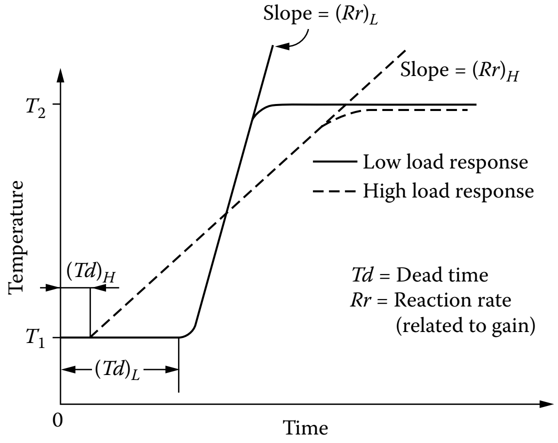


FIG. 8.9i
In a chemical reactor with a once-through jacket, without a water circulating pump, both the dead time and the gain of the process drop as the cooling load rises.

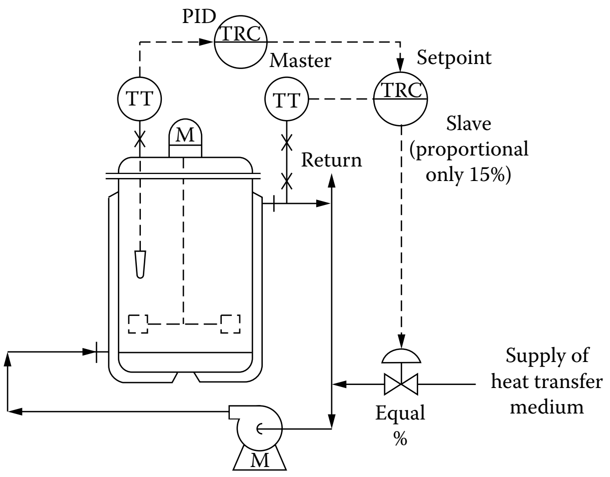


FIG. 8.9j
The addition of a cascade control loop to a chemical reactor that is provided with coolant recirculation reduces the period of oscillation of the master temperature controller.

variable (reactor batch temperature, hereafter called reactor temperature), whose response is slow to changes in the heat-transfer medium flow (manipulated variable), is allowed to adjust the set point of a secondary loop, whose response to coolant flow changes is rapid. In this case, the reactor temperature controller varies the set point of the jacket temperature control loop.

The purpose of the slave loop is to correct for all outside disturbances, without allowing them to affect the reaction temperature. For example, if the control valve is sticking or if the temperature or pressure of the heat-transfer media changes, this would eventually upset the reaction temperature, if the control system was configured as in Figures 8.9f or 8.9g but not in Figure 8.9j.

This is because in Figure 8.9j the slave would notice the resulting upset at the jacket outlet and would correct for it before it had a chance to upset the master. As pointed out in the detailed discussion of cascade systems in Section 2.6, the process lags should be distributed between master and slave loops in such a way that the time constant of the slave is one tenth that of the master. Cascade loops will not function properly if the master is faster than the slave.

It is preferred that the slave controller be used to maintain the jacket outlet (and not inlet) temperature, because this way the jacket and its dynamic response is included in the slave loop. Another advantage of this configuration is that it removes the principal nonlinearity of the system from the master loop, because reaction temperature is linear with jacket-outlet temperature.

The nonlinear relationship between jacket-outlet temperature and heat-transfer-medium flow is now within the slave loop, where it can be compensated for by an equal-percentage valve, whose gain increases as the process gain drops off (Figure 8.9h). In most instances the slave will operate properly with proportional-plus-derivative or proportional only control, which is set for a proportional band of 10–20%.

The period of oscillation of the master loop is usually cut in half as direct control is replaced by cascade. This might mean a reduction from 40 to 20 min in the period and a corresponding reduction of perhaps 30 to 15% in the proportional band.[3] The derivative and integral settings of an interacting controller would also be reduced from about 10 min to about 5 min. This represents a fourfold overall loop performance improvement.

Using jacket-inlet temperature as an override or measurement for the slave may be useful in cases in which the jacket temperature must be limited, e.g., when the reactor is used as a crystallization unit, or in safety systems that serve to protect glass-lined reactors from thermal shock. Such protection might be needed when hot water is generated by direct injection of steam (see Figure 8.9k).

If it is desired to reduce the heat-up time by applying direct steam heating to the reactor jacket and to use both water and methanol as cooling media in the same system, the configuration in Figure 8.9k can be considered. In order to guarantee that water and methanol will not intermix, even accidentally, in addition to the coolant control valve (TCV-2), tight shut-off on/off valves are provided on both coolants (V3 and V4). Such a control system can be operated in a variety of modes.

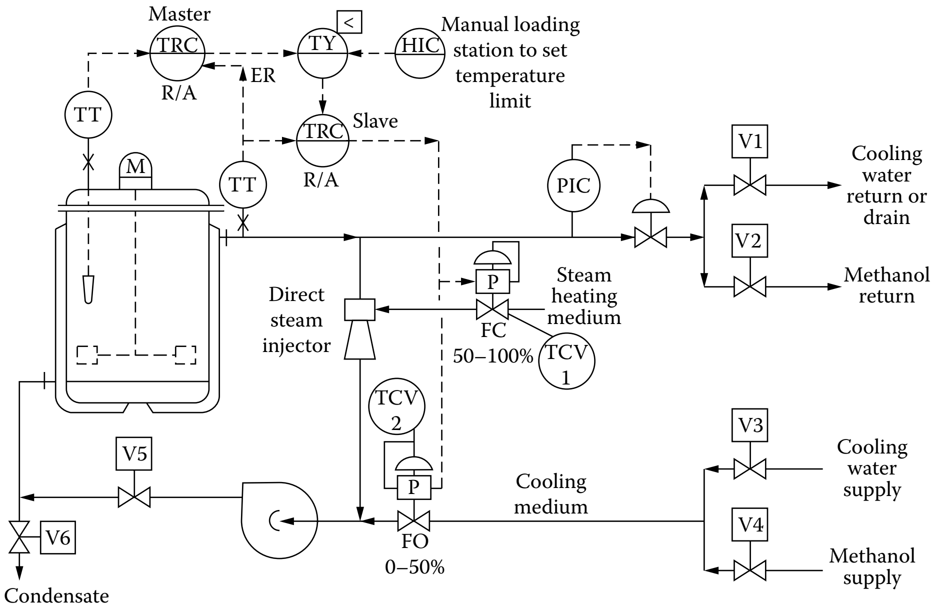
In Figure 8.9k five operating modes are listed. These can be implemented with positive interlocks. The return flow path is selected to match the type of coolant supply and is provided with a back-pressure regulator to prevent draining of the jacket.

If heat needs to be added in some phases of the reaction while in other phases it must be removed, the controls must be configured in a two-directional manner. Figure 8.9l depicts a cascade temperature control system with provisions for batch heat-up. The heating- and cooling-medium control valves are split-range controlled, such that the heating-medium control valve operates between 50 and 100% control output signal and the cooling-medium control valve operates between 0 and 50%.

It is important to match the characteristics of the valves (zero point) and to avoid nonlinearity at the transition, which can result in cycling. It is equally undesirable to keep both valves open simultaneously, because it results in energy waste. The control system shown in Figure 8.9l is a fail-safe arrangement, because in case of air failure the heating valve is closed and the coolant valve is opened.

Figure 8.9l also shows an arrangement whereby an upper temperature limit is set on the recirculating heat-transfer-medium stream. This is an important consideration if the product is temperature-sensitive or if the reaction is adversely affected by high reactor wall temperature. In this particular case, the set point to the slave controller is prevented from exceeding a present high-temperature limit.

Another feature shown is a back-pressure control loop in the heat-transfer-medium return line. This may be needed to impose an artificial back-pressure, so that during the heat-up



Heat transfer mode	Valve status (O = open)							
	TCV1	TCV2	V1	V2	V3	V4	V5	V6
Direct steam heat	0							0
Heat by water	0		0		0		0	
Cooling by water		0	0		0		0	
Heat or cool by water	0	0	0		0		0	
Cool by methanol		0		0		0	0	

FIG. 8.9k
Multimedia reactor jacket temperature control system that allows heating by direct steam injection and cooling by either water or methanol.

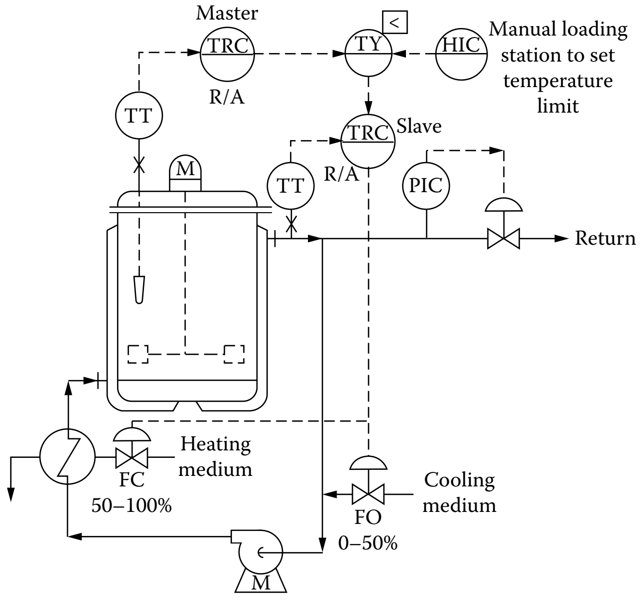


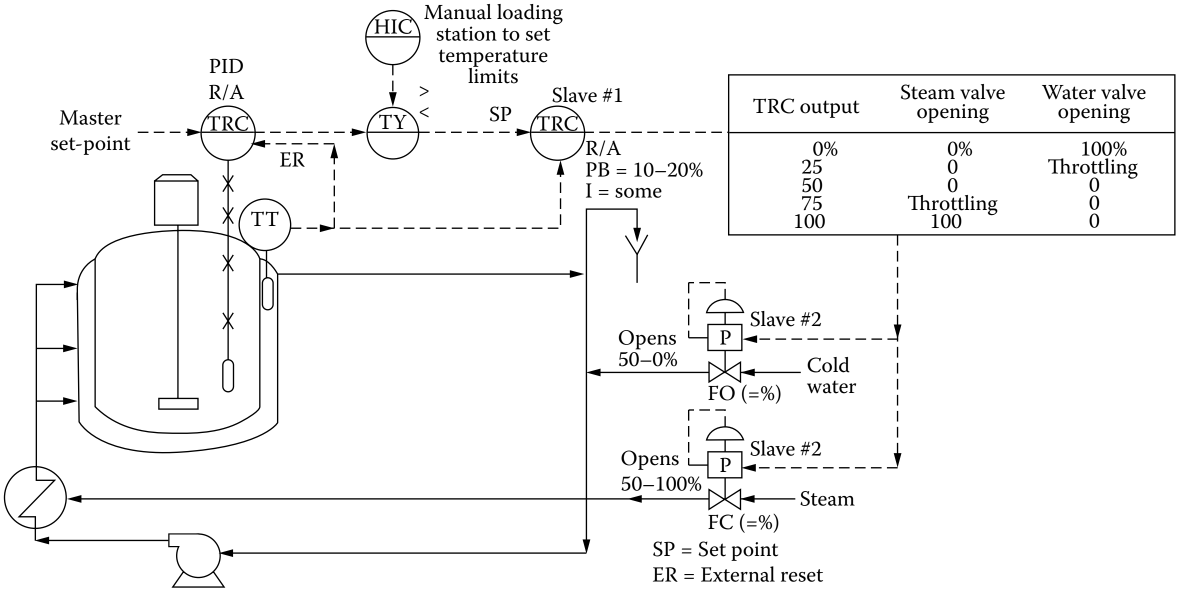
FIG. 8.9l
Two-directional cascade loop with a maximum limit on jacket temperature allows heat to be added in some phases of the operation and removed in others.

cycle no water leaves the recirculation loop and therefore the pump does not experience cavitation problems.

Limitations of Cascade Control Figure 8.9m illustrates a cascade loop consisting of three controllers in series. The master is the reaction temperature controller, the primary slave is the jacket temperature controller, and the secondary slave is the valve position controller (see Sections 6.2 and 6.12).

A positioner is a position controller that detects the opening of the valve and corrects for any deviations between measurement (the mechanically detected position) and set point. In such a hierarchical arrangement, only the master can control its variable independently; the slave controller set points must be freely adjustable to satisfy the requirements of the master.

Reset Windup Whenever the master is prevented from modifying the set point of the slave — because of a limiter, such as in Figure 8.9m, or because the slave has been manually switched from remote to local set point — reset windup can occur. Reset windup is the integration of an error that the controller is prevented from eliminating. Consequently, the



TRC output	Steam valve opening	Water valve opening
0%	0%	100%
25	0	Throttling
50	0	0
75	Throttling	0
100	100	0

FIG. 8.9m

If the control valves are provided with positioners, there will be three controllers in series in a cascade loop. It is recommended that all cascade masters be provided with an external reset from the measurement of their slave controller.

controller output is saturated at an extreme value. Once saturated, the controller is ineffective when control is returned until an equal and opposite area of error unsaturates it.

This problem is eliminated by the external reset (ER) shown in Figure 8.9m. The external reset signal converts the contribution of the integral mode to just a bias (Equation 2.28[7]) and thereby stops the integral action whenever the slave is not on set point. This feature eliminates the need for switching the master to manual and thereby also eliminates the need for the auto/manual station. In addition, it eliminates reset windup upset due to start-ups, shutdowns, or emergency overrides. Whenever external reset is used the slave must have some integral to eliminate the offset; otherwise, the slave offset would cause an offset in the master.

Another limitation is that the cascade loop will be stable only if each slave is faster than its master. Otherwise, the slave cannot respond in time to the variations in the master output signal, and a cascade configuration will in fact degrade the overall quality of control. A rule of thumb is that the period of oscillation of the slave should not exceed 30% of the period of oscillation of the master loop. This requirement is not always satisfied.

For example, in Figure 8.9m it is important to select valve positioners that are faster than the slave temperature controller on the jacket. Similarly, the jacket control loop should contain less dead time than its master, which would usually not be possible if a once-through piping configuration (Figure 8.9f) is used.

One possible method of reducing the dead time of the cascade slave loop is to move the measurement from the jacket outlet (Figure 8.9j) to the jacket inlet. This usually is not recommended, because when this is done, the slave will do much less work because the nonlinear dynamics of the jacket (Figure 8.9h) have been transferred into the master loop.

Multiple Heat-Transfer Media The use of a single coolant and single heating media (shown in Figure 8.9m) is often insufficient or uneconomical. If one type of coolant (or heating media) is less expensive than another—for example, the cold water used in the system in Figure 8.9n might be less expensive than the chilled water—it is desirable to fully utilize the first before starting to use the second.

For best performance, the fact that there are three valves should be transparent to the temperature controllers. Their gain should be the same, and their combined range should appear as the straight line in Figure 8.9h. This is not easy to achieve, particularly when the valves are nearly closed (see Figure 6.7e in Chapter 6 for a discussion of valve gains), which happens to be the case when the controller output is 11 PSIG in Figure 8.9n.

Therefore, some users prefer to provide some overlap so that the water valve might start opening at 11 PSIG, while the steam valve does not close fully until the signal drops to 10.5 PSIG. Overlapping at the transition points improves control quality but at the price of energy efficiency.

Figure 8.9n also shows that the destination of the returning water should not be selected on the basis of the origin of that water but rather should be based on temperature. This will reduce the upset caused in the plant utilities when a reactor switches from heating to cooling.

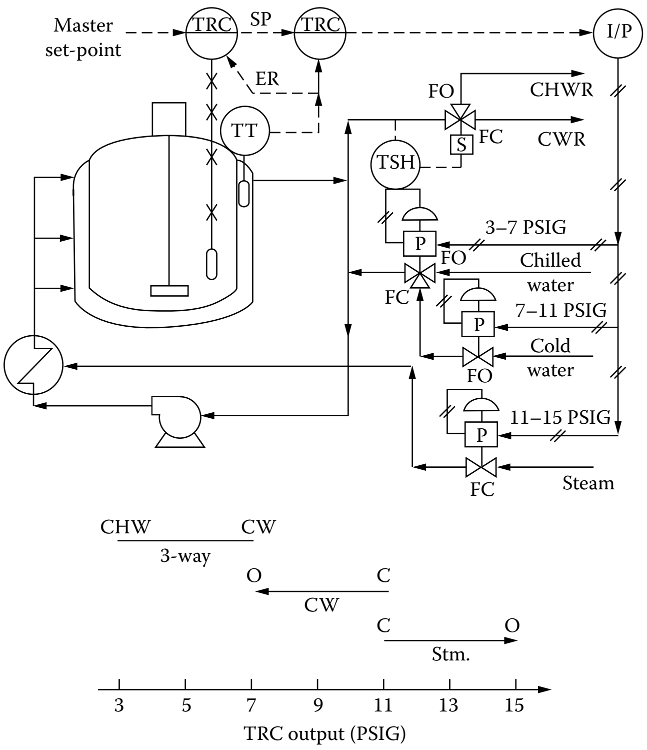


FIG. 8.9n
When multiple coolants are available, the total cost of cooling can be minimized by split-range sequencing of the valves.[4]

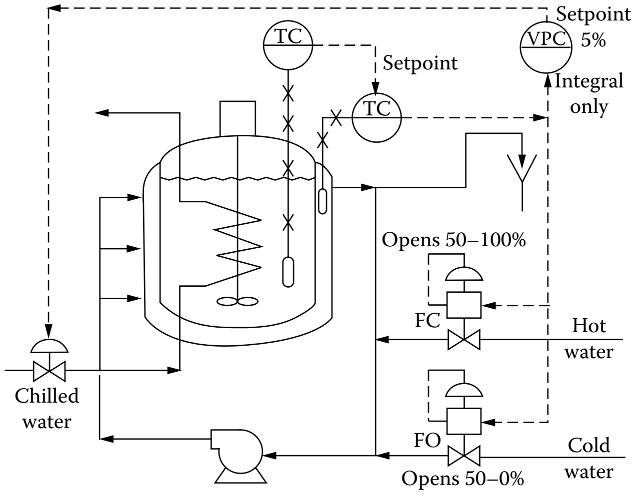


FIG. 8.9o
In this configuration, the higher-cost chilled water is used only when the availability of the lower-cost cold water is insufficient to meet the total demand for cooling.[3] *In all cases, when a separate heat-transfer surface is used for auxiliary cooling, the valve position controller (VPC) is needed to manipulate the flow of the auxiliary coolant.*

Figure 8.9o describes a reactor with a separate chilled water coil. This coil is inoperative until the cold water valve approaches full opening. When the valve position controller (VPC) detects that condition, it starts opening the chilled water valve. The resulting increased heat removal will cause the temperature cascade loop to close the cold water valve until it drops to the setting of the VPC.

Figure 8.9p serves the same purpose as the system depicted in Figure 8.9k, except that the designer placed the throttling valve for the coolant into the return flow path so that a separate back-pressure regulator would not be needed.

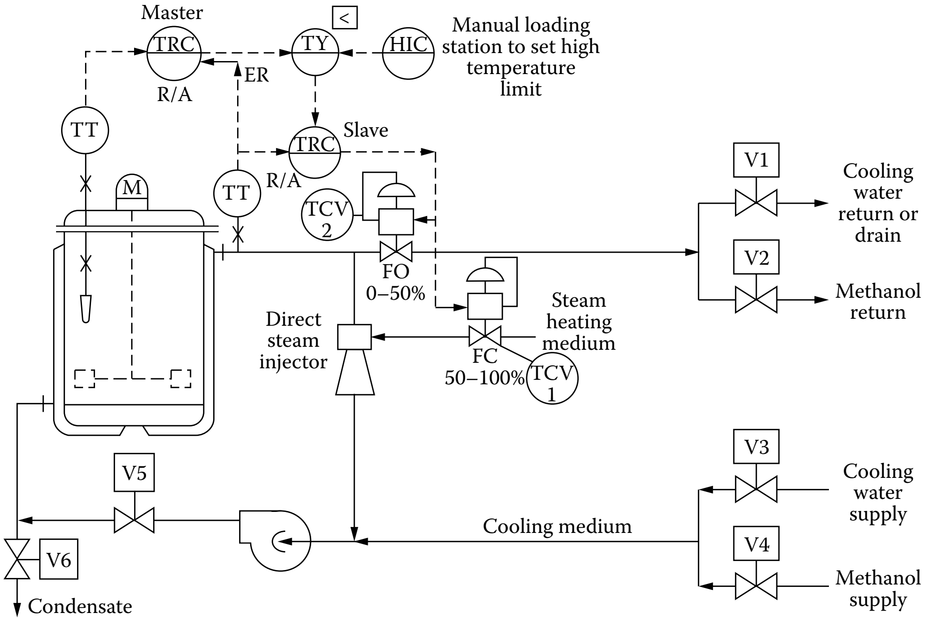


FIG. 8.9p
Multimedia temperature control system with coolant control valve in the return line.

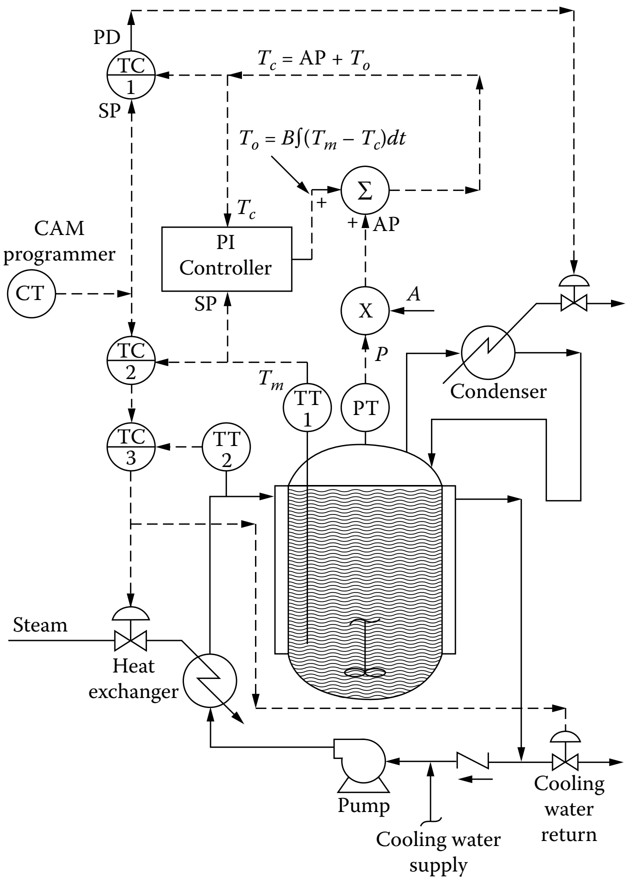


FIG. 8.9q
Pressure-compensated temperature control can be used to improve the control system's speed of response. (Adapted from Byrd Hopkins, U.S. Patent No. 3,708,658.)

Pressure-Compensated Temperature Control

In a process where the reactor pressure is a function of batch temperature (e.g., the reactor pressure is essentially the vapor pressure of one of the major components in the reaction), this pressure may be sensed and used to speed up temperature control. In large polymerization reactors having low heat-transfer coefficients and large changes in heat evolution, the conventional temperature cascade loop is not fast enough. On the other hand, pressure measurement gives an almost instantaneous indication of changes in temperature.

Figure 8.9q illustrates the application of a pressure-compensated temperature control system to a reactor with both jacket and overhead condenser cooling (U.S. Patent No. 3,708,658, January 2, 1973, assigned to Byrd Hopkins of the Monsanto Company). This same approach can also be applied to reactors with jacket cooling only. Under steady-state conditions the reactor temperature (T_m) is on setpoint, set by the cam programmer, and therefore the calculated temperature $T_c = T_m$.

When an upset occurs, the pressure transmitter (PT) will detect it first, causing the calculated temperature (T_c) to change as the AP part of the expression is changed. This will make the measurement of TC-1 much faster than it otherwise would have been; it also allows the overhead condenser to start removing the excess heat even before the temperature transmitter (TT-1) is able to detect it. After each dynamic upset, the PI controller slowly returns the calculated temperature (T_c) to equal the measured temperature (T_m). This then automatically reestablishes the correct pressure-temperature relationship as the composition in the reactor changes.

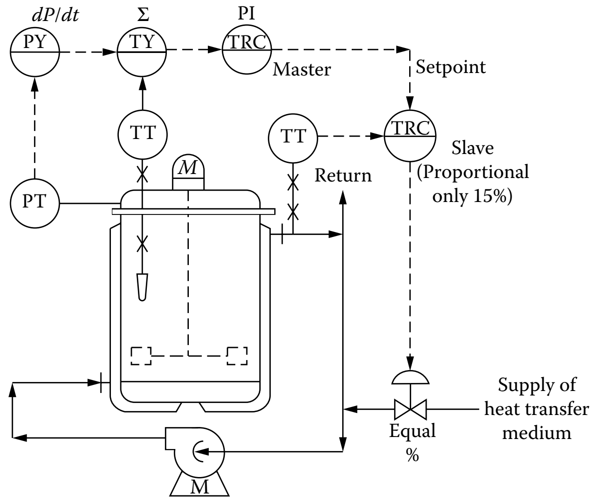


FIG. 8.9r
The derivative mode can be applied to the quickly responding process variable (pressure), and the integral mode can be applied to the more slowly responding process variable (temperature).

The net result is the ability to operate the reactor at a much higher reaction rate, thereby obtaining higher productivity than was possible with temperature control alone.

TC-1 in Figure 8.9q would normally be provided with proportional and derivative control modes only. When TC-1 is on set point, the output signal returns to a value set by an adjustable internal bias. If the preferred means of cooling is through the jacket, then this bias will be set to a low value, but not to zero.

If there is only a single manipulated variable, the control system would be configured as in Figure 8.9r. The control modes are distributed between the measured variables so that integral action will act on the slowly responding temperature, while derivative action is applied to the more sensitive pressure.[5]

The derivative time setting will be much shorter than the integral, because pressure responds faster than temperature. Integral action cannot be applied to the pressure measurement, because the pressure can vary even under constant temperature conditions (as a result of variation in feed composition or catalyst activity), and the intent here is to use the pressure loop only in the unsteady state.

Naturally, integral action is applied to the temperature measurement signal, because it is steady-state temperature

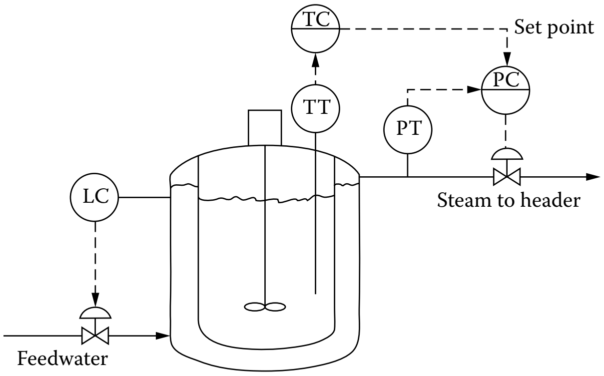


FIG. 8.9s
Exothermic reactors can be stabilized by the internal heat sink, which is provided by the boiling coolant in the jacket.[5]

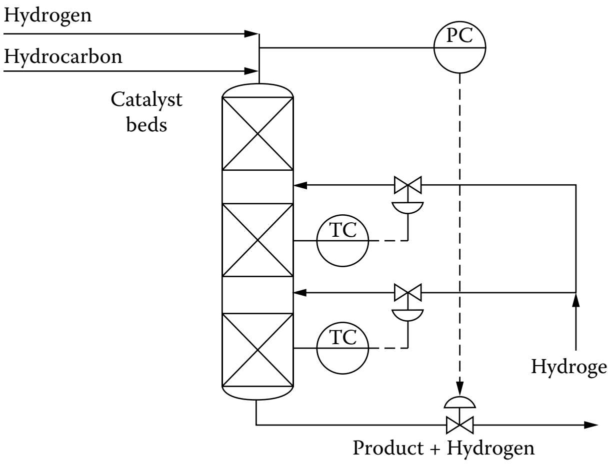


FIG. 8.9t
In some processes, the reaction rate (and reaction temperature) can be reduced by lowering the concentration of the reactant. This is usually done by introducing more diluent (in this case, hydrogen).[5]

that determines product quality, and integral action will ensure its return to set point.

Temperature Controlled by Boiling Coolant Gas phase reactions (such as ammonia synthesis) are usually conducted at relatively high temperatures and pressures. Because of the competing reactions taking place in parallel with the main reactions, productivity and partial pressures are interrelated.

For example, in an ammonia synthesis process, a pressure increase is required to increase production when the ammonia concentration is low. On the other hand, a pressure decrease is required to increase production when the ammonia concentration is high.[5] If inerts are also present, partial pressures are usually determined by detecting both the total pressure and the gas composition.

In such processes, which are both temperature- and pressure-sensitive, a very stable heat removal system is desired. If the reactor jacket is filled with boiling water, the rate of heat removal can vary without causing a change in the jacket temperature. Figure 8.9s illustrates such a cascade system. In order to increase the sensitivity of the loop, pressure is selected as the controlled variable for the slave controllers.

Temperature Controlled by Feed Rate In a hydrocracking reactor (Figure 8.9t), product quality, catalyst life, and productivity are each a function of accurate temperature control. In this process, the reaction rate is fast, residence time is a few seconds, and reactant concentration is low. This combination would allow the reaction temperature to be controlled by manipulating the feed rate. If it is desirable to set production rate, and therefore feed rate independently of cooling capacity, the temperature can be controlled by throttling a diluent.

In Figure 8.9t the diluent is hydrogen, which is admitted under separate temperature controls into each zone. The introduction of hydrogen diluent lowers the reaction rate by reducing the reactant concentration and also by cooling.

The temperature of the batch at any one time is a function of the balance between the exothermic heat that is generated and the quantity of heat that is being removed by the coolant. If the two are not in balance, the batch temperature changes.

In the previously described control schemes, the exothermic heat was uncontrolled, and the cascade loop controlled the batch temperature by adjusting the coolant flow to match that load. In some of the fast and more stable processes, this configuration can be reversed, as shown in Figure 8.9u. Here the rate of cooling is fixed (at a relatively high rate) and the temperature controller (TRC-4) is adjusting the reagent flow (exothermic heat generation) to maintain a balance.

At the beginning of the batch the concentrations of precharged other reactants are high, the concentration of the product in the batch is low, and therefore the process gain is very high. As the reaction progresses, product concentration rises, unreacted reactant concentrations drop, and, as a consequence, the gain of the process also drops. By the end of the reaction, the process gain (*Gp*) can be reduced 10- or 100-fold.

As was shown in Figure 2.1x in Chapter 2, good control requires that the gain product of the loop be constant. Therefore, as *Gp* drops it is necessary to increase the controller gain (*Gc*) in proportion. This is done by FY-1 in Figure 8.9u, which measures the total amount of reactant charged and increases the gain of TRC-4 as that total rises. The function of the reaction rate (FY-1) is nonlinear reaction-specific. TT-2 provides a feedforward signal so that if the coolant temperature would rise and thereby the cooling rate would drop off, this would also decrease the gain of TRC-4, thereby making it cause a smaller change in slave set point.

The set point of the slave (FRC-1) is limited by the low-signal selector TY-4. It compares two limits to the set point of FRC-1, generated by TRC-4 and picks the lowest of the three signals. During start-up, safety is served by KT-4, which

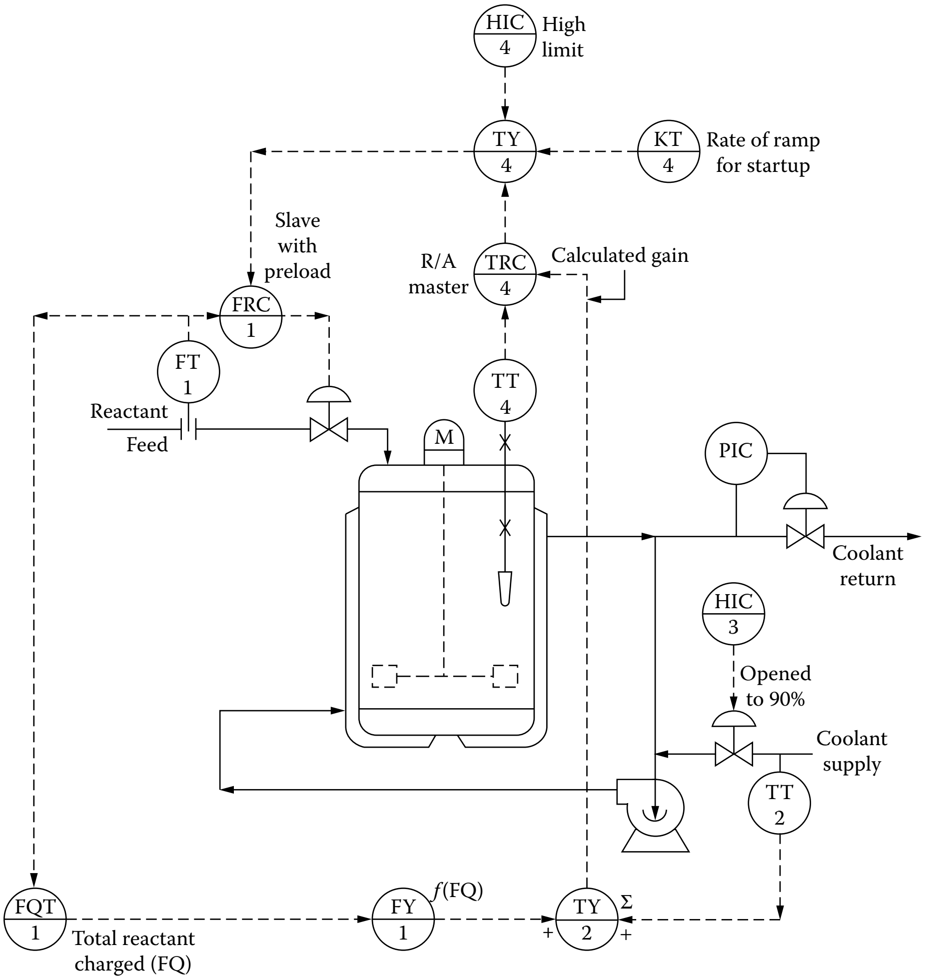


FIG. 8.9u

If the amount of cooling is constant, the cascade slave of the reactor temperature controller can be the reactant flow controller. Because the process gain drops off as the production rate rises, the gain of the cascade master controller (TRC-4) can be automatically increased as a function of total reactant charged.

ramps the slave set point up at a safe rate to protect against overshooting. The slave (FRC-1) is preloaded to minimize start-up overshoot or transients.

The amount of preload is based on experience and serves to stabilize the loop more quickly than if the initial value of the FRC-1 output was zero. The main advantage of such control systems is that they minimize the batch reaction time and thereby maximize plant productivity.

Model-Based Temperature Control

When a reaction is highly exothermic and if runaway reactions can occur, the use of a model-based control system is justified. In order to develop the total model, three component models need to be developed:

1. *Equipment model*, describing the effect of heat-transfer media on reactor temperature (T_r)
2. *Kinetic model*, describing the effect of T_r on the chemical reaction
3. *Calorimetric model,* describing the effect of reaction rate on T_r[7]

When an exothermic reaction is taking place within the reactor, a feedback cascade loop such as that shown in Figure 8.9m is needed to provide stable temperature control by matching the rate of heat removal to the rate of heat generation. If the TRC in Figure 8.9m was tuned at a time when there was no reaction taking place within the reactor, it will not perform properly when exothermic heat is being generated.

If the gain of the master TRC is fixed, the loop might become unstable if the reaction is autocatalytic (does not require a catalyst). Model-based controls can improve on such performance. One such model-based approach is to use a heat-release rate estimator. The rate of heat release (Q in Figure 8.9v) can be estimated on the basis of reactor

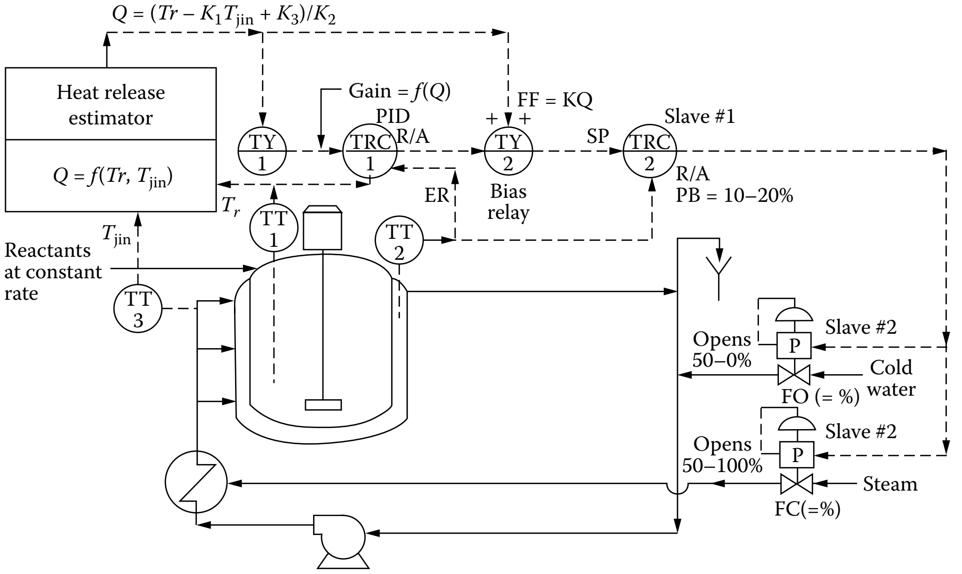


FIG. 8.9v

Heat-release estimator increases the gain of the master temperature controller (TRC-1) when the estimated total exothermic heat release rises and, as a consequence, the process gain in a batch reactor drops.

temperature (T_r) and jacket-inlet temperature (T_{jin}) in accordance with:

$$Q = (T_r - K_1\, T_{jin} + K_3)/K_2 \qquad \textbf{8.9(6)}$$

In exothermic reactions, Q is the major load variation that tends to upset the stability of the temperature controls. Once this disturbance load is estimated, it is possible to feedforward that estimate to the slave set point as a bias. This bias relay is TY-2 in Figure 8.9v, and the feedforward bias (FF) it adds to the slave set point is a function of Q. The feedback master (TRC-1) naturally corrects the total effect on the jacket temperature set point and thereby overcomes the errors in the feedforward model.

The gain of the feedback master (TRC-1) is varied as a function of Q [Gain = $f(Q)$]. As Q increases, the process gain drops, and therefore the gain of this nonlinear adaptive-gain feedback controller is increased in order to keep the gain product for the loop constant.

INITIAL HEAT-UP

In most chemical reactions, a certain temperature must be reached in order to initiate the reaction. An ideal reactor temperature controller will permit rapid automatic heat-up to reaction temperature without overshoot and then accurately maintain that reaction temperature for several hours. This is a difficult goal to accomplish, because the dynamics of the controlled process will go through a substantial change as the heat load first drops to zero and then the cooling load gradually evolves when the reaction is started.

The master temperature controller is usually a three-mode one, tuned for the exothermic phase of the reaction cycle. It might have a 30% proportional band and a 5 minute setting for both integral and derivative. If such a controller were kept on automatic during heat-up, a substantial temperature overshoot would result (Figure 8.9w) because of windup. Therefore, the conventional PID controller must be supplemented with added features to provide it with the required start-up characteristics.

The added feature is either the “batch unit” or the “dual mode unit,” depending on the proportional band. If it is less than 50%, the batch unit will give good results, whereas if a

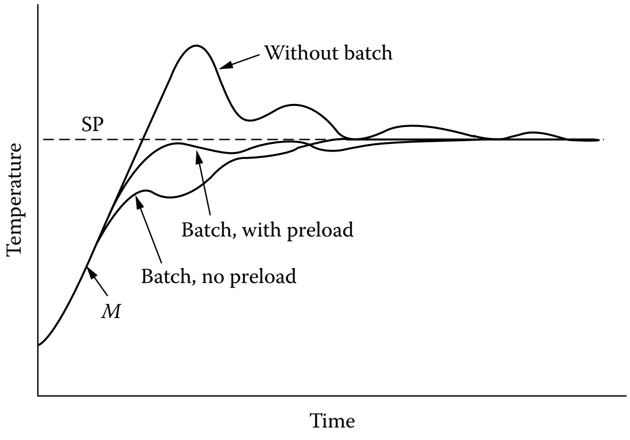


FIG. 8.9w

Temperature overshoot can be eliminated by properly adjusting a “batch unit.”[3]

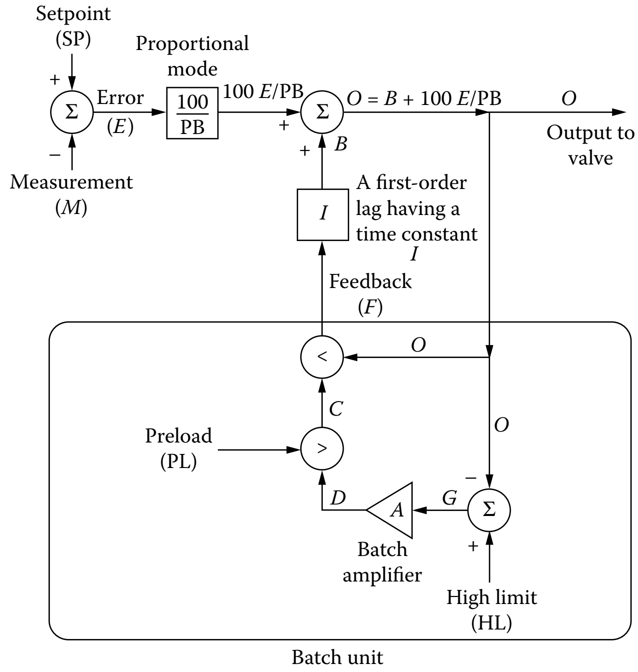


FIG. 8.9x

The batch unit disables the integral control mode until the error is nearly zero.[3] *Thereby, the batch unit maintains the controller output at its limit.*

proportional band wider than 50% is required, the dual-mode unit will be more effective.

The Batch Unit

Without the batch unit, the PI controller illustrated in Figure 8.9x would receive a feedback signal (*F*) equaling the output (*O*). Therefore, whenever there is an error (*E*), the output signal is driven continuously by the positive feedback through *I* (a first-order lag, having a time constant *I*) until *B* reaches the saturation limit. Once in this saturated state (the reset is wound up), the output signal "*O*" will equal "*B*" even if the error has returned to zero. This is the reason that in Figure 8.9w the temperature keeps rising even after it has reached set point (SP = *M*, *E* = *O*).

Without the batch unit, therefore, control action cannot begin until an equal and opposite area of error is experienced. This is why reset windup always results in overshoot and why this windup must be eliminated by the addition of the batch unit shown in Figure 8.9x.

Under normal operation, the output to the valve is below the high limit. Therefore, *G* is positive and the amplifier drives *D* upward, which causes *O* to be less than *C*. In this state, the low selector selects *O* as the feedback signal, and the controller behaves as a conventional PI controller.

When *O* exceeds HL, the amplifier drives down *D*, *C*, *F*, and *B* and thereby limits *O* from exceeding the HL setting.

It is also necessary to provide a low limit (called preload) to the feedback signal; otherwise, the opposite of an overshoot would be experienced—an excessively sluggish approach to set point, as shown in Figure 8.9w by the "no preload" curve. If the PL setting did not prevent the feedback (*F*) from dropping too low at times of high error (such as at the beginning of heat-up), *B* could saturate at the low limit, keeping output (*O*) below zero even when the measurement has returned to set point. With preload, the controller output "*O*" will equal PL when the error is zero.

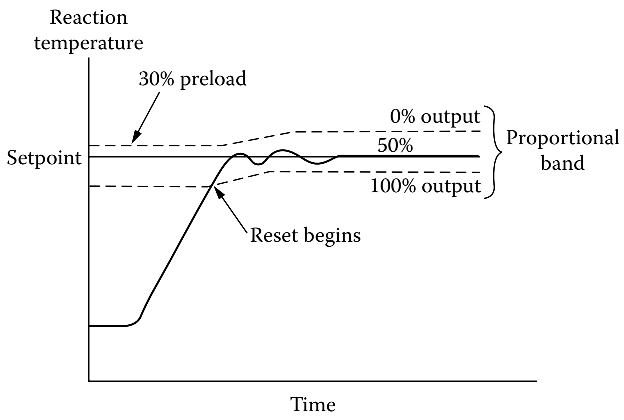
It can be seen from the above that a PID-type batch controller requires a total of five adjustments, because HL and PL must be set and the three control modes must be tuned. HL should be set at the maximum allowable jacket water temperature, which would then eliminate the need for a separate limit, such as the HIC in Figure 8.9l.

The correct setting for PL is the master controller output at that time when reaction has started and a steady state has been reached between the generation and the removal of the heat of reaction. If, for example, the jacket water temperature during steady state is 90°F, this value could be selected as the preload setting, which will be the output of the master (and the set point of the slave) when the reaction temperature has been reached.

Actually, the PL setting should be a few degrees lower than this value, say 87 or 88°F, to allow for the contribution of the integral action of the controller from the time the proportional band is entered to the time where the set point is reached. This is illustrated in Figure 8.9y.

Dual-Mode Controller

The effectiveness of the batch unit, described earlier, is lost when the reactor requires a wide proportional band, say in excess



100% output means full heating
0% output means full cooling
30% output might correspond to the normal level of cooling at steady state

FIG. 8.9y

If preload is correctly set, overshoot will be eliminated and heat-up time will still be held at a minimum.[3]

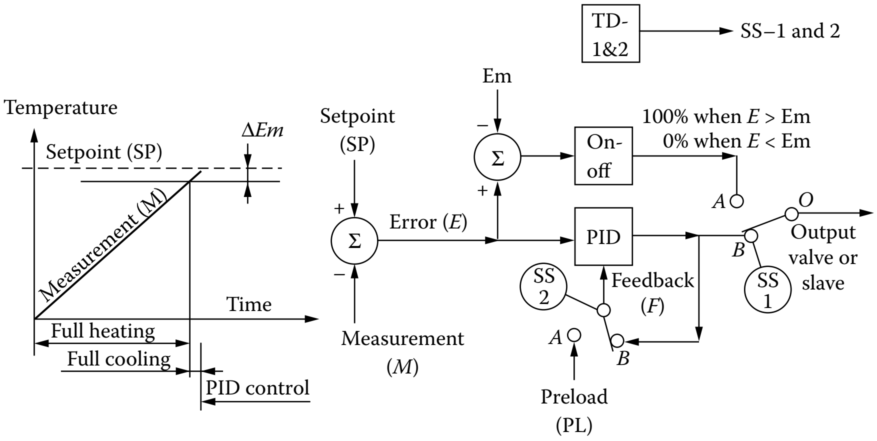


FIG. 8.9z

The dual-mode controller switches from full heating to full cooling, which is then followed by switching to PID control with preload. It is used for optimum start-up of potentially unstable batch reactors. The value of ΔE_m *is the minimum error setting, which corresponds to a state when the measurement has approached the set point to within 1 or 2%.*

of 50%. With a wide band (as shown in Figure 8.9y) reset action would begin much earlier, which would lengthen the heat-up time. In such a case, the dual-mode unit (Figure 8.9z) is the proper selection.

In the dual-mode unit, the preload is estimated as in the case of the batch unit, but it is not reduced for integral correction. It is not lowered from 90°F to 87 or 88°F in our example, because in this case reset does not begin until the error is zero.

The sequence of operation is as follows:

1. Full heating is applied until the reactor is within 1 or 2% of its set-point temperature. This margin is set by the minimum error setting (Em). During this state SS-1 and SS-2 are in position "A."
2. When E drops to Em, time delays TD-1 and -2 are started, and full cooling is applied to the reactor for a minute or so to remove the thermal inertia of the heat-up phase. When TD-1 times out the period required for full cooling, SS-1 switches to position "B" and the PID controller output is sent to the slave as set point. This output is fixed at the preload (PL) setting, which corresponds to the steady-state jacket temperature (estimated in our example as 90°F).
3. When the error and its rate of change are both zero, estimated by TD-2, this time delay will switch SS-2 to position "*B*." This switching also transfers the PID loop from manual to automatic, with its external feedback loop closed.

If properly tuned, the dual-mode unit is the best possible controller, because by definition, optimal switching is unmatched in the unsteady state by any other technique.[3] On the other hand, this loop requires seven settings. Three of these—*P*, *I*, and *D*—pertain only to the steady state of the process; the other four—PL, Em, TD-1, and TD-2—will determine start-up performance. The effect of these adjustments is self-evident:

Em should be increased in case of overshoot and lowered if undershoot is experienced.
PL has the same effect as in Figure 8.9w.
TD-1, if set too long, will bring the temperature down after the set point is reached.
TD-2 is not very critical.

Figure 8.9aa illustrates that from the start-up performance of the reactor, it can be determined which setting needs adjustment.

Rate of Temperature Rise Constraint

In highly unstable, accident-prone reactors that have a history of runaway reactions, an added level of protection can be provided, based on the permissible rate of temperature rise

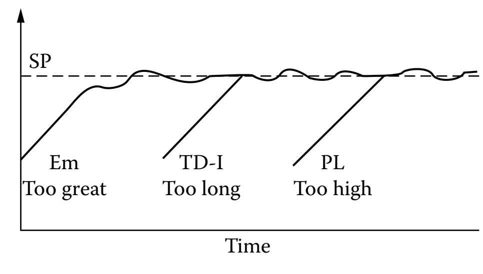


FIG. 8.9aa

It is possible to diagnose the tuning constants used in dual-mode loops on the basis of the reactor's start-up performance.[3]

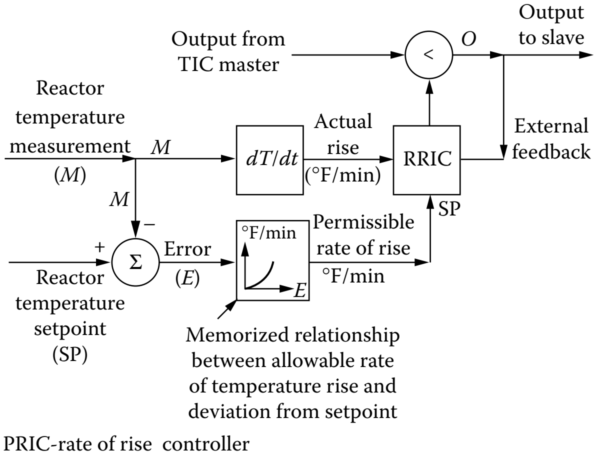


FIG. 8.9bb
The rate of temperature rise constraint is usually incorporated as added protection.

during heat-up. This is usually superimposed on the previously discussed control systems on a selective basis as a safety backup. The protection is provided by first calculating the actual rate of temperature rise in units of °F/min and then sending it as the measurement signal to the RRIC controller in Figure 8.9bb.

The set point of this controller is programmed as a function of the heat-up state. Toward the end of the heat-up period, when the reaction temperature has been almost reached (the error is near zero), the value of the permissible rate of rise is set to be much lower than at the beginning of heat-up. Doing so prevents the process from building up a thermal inertia as it approaches the region of potential instability.

Model-Based Heat-Up Control

Model-based heat-up algorithms can be configured in two blocks, the "predictor block" and the "corrector block." The predictor block is a process model of the reactor and serves to determine the amount of heat input required (*Qh*) at any point in the batch heat-up cycle, so as to minimize the heat-up period while eliminating overshoot.

The corrector block serves to refine the model by correcting the constants and calculation parameters in the process model by comparing the model to the actual process. Figure 8.9cc illustrates the operation of such a model-based heat-up control system.

The variable manipulated by the "predictor" is the heat input rate (*Qh*), which initially is set to maximum, similar to the action of the dual-mode control algorithm in Figure 8.9z. The difference with the dual-mode strategy is that here (in Figure 8.9cc), as the reactor temperature approaches its target set point, the manipulated variable is not switched to cooling but to a minimum rate of heating (*Qh*)min. The switching from (*Qh*)max to (*Qh*)min occurs when the steady-state reactor model indicates that with the amount of heat already introduced, once the reactor reaches its new steady state, its temperature will match the set point.

Adaptation The nature of the batch reactor process is such that it never reaches a steady state. The process is in a continuous state of transformation and change; therefore, the control system parameters should also be in a continuous state of change. Some of these changes are rather drastic; for example, when the system switches from heating to cooling. When such changes occur, the controller parameters must be changed immediately and equally drastically, which is usually done by "programmed adaptation," meaning that the PID algorithm settings for the various stages of the reaction cycle are stored in a table and are recalled as needed.

Adaptation of tuning constants can also occur slowly using various adaptive algorithms (Chapter 2) as a function of various measurements, such as, for example, the rising of the level (mass) in the reactor or the reduction in the exothermic heat release as the reaction rate drops off. Adaptation can also be based on various rules of thumb, such as utilizing the rate of temperature rise during heating up the reactor to predict the reactor time constants, which in turn can be used to tune the temperature controller.

Initialization, Preload, and Feedforward When the transition from heat-up to the reaction phase takes place, it is important that the temperature controller be properly "initialized" (placed into the correct initial state). That means that the controller must be provided with the proper tuning constants and bias (output signal at zero error) that will make the transition smooth. This initialization of the controller is often called preloading. As was shown in Figure 8.9w, the use of the right amount of preload is essential to obtain maximum heat-up rate without overshoot.

In a batch process, where the same product is made in consecutive batches, preload is often determined empirically. This means that if in batch #19 the transition to reaction phase was stable and the controller output at the beginning of that phase was 52%, it is a fairly safe assumption that in batch #20 the same 52% preload will also provide stable transition. In some reactor heat-up applications the preload can also be predicted on the basis of the rate of temperature rise during heat-up.

Feedforward features can also be incorporated into the overall control system. For example, in Figure 8.9n one could feedforward a change in the set point in the master directly into a change in the set point of the slave, if a bias unit was inserted to detect the master set point and to generate the correct amount of change based on that measurement. By such feedforward action, the integral-mode offset error, caused by set point change, can be reduced.

This is a good feature, because even if the control algorithm is so configured that the proportional and derivative modes act only on the measurement (Section 2.3 in Chapter 2), the integral mode usually acts on the total error, and therefore it is responsive to set-point changes.

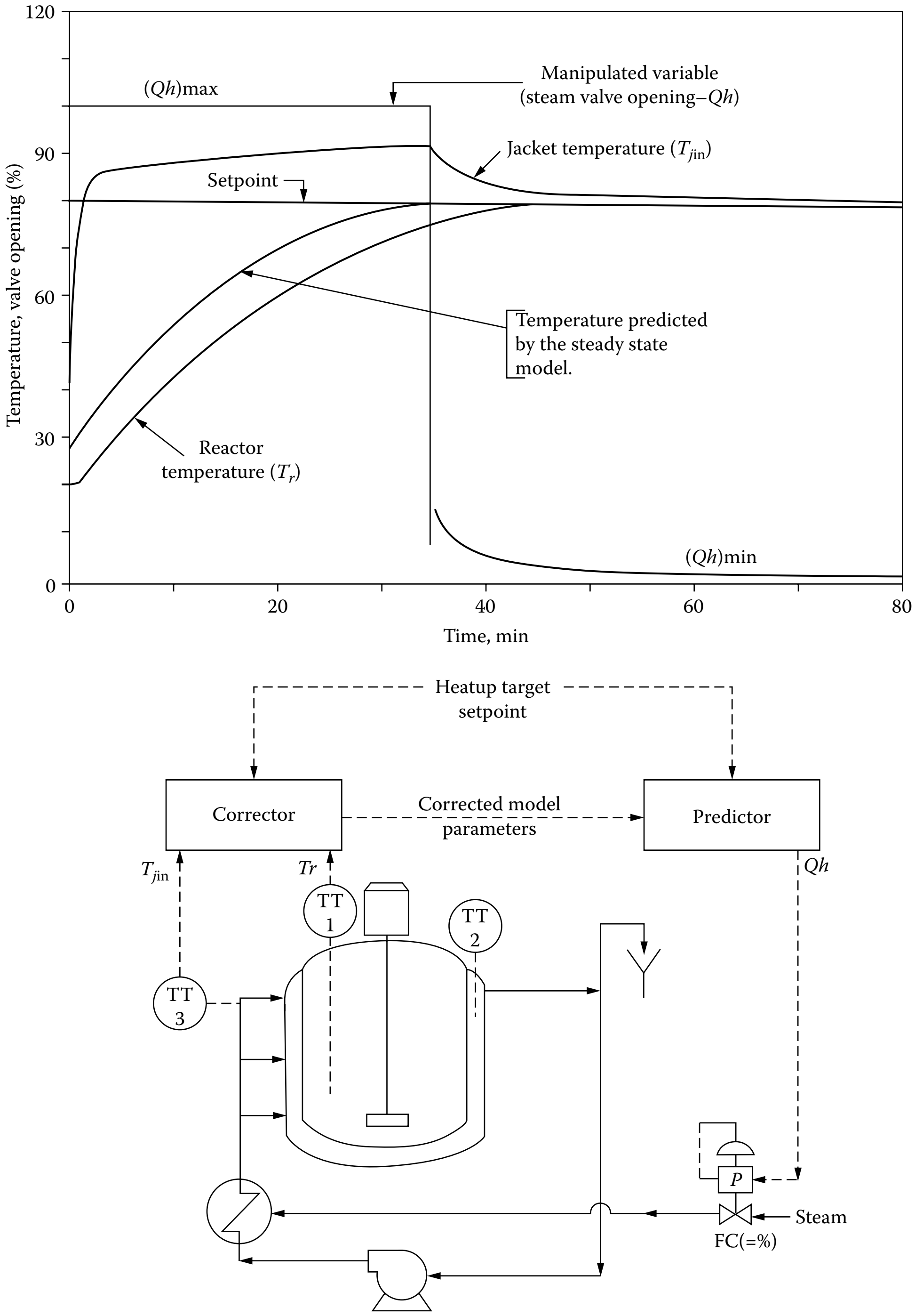


FIG. 8.9cc
The model-based heat-up strategy controls the point of switching from maximum to minimum heating. This switching occurs when the model predicts that enough heat has been introduced for the reactor temperature to reach its set point (top), once the new steady state is reached. The lower half of the figure shows the roles of the corrector and predictor segments of the system.

SPECIAL AND OPTIMIZING STRATEGIES

Maximized Production by Constraint Optimization

In exothermic reactions, one of the critical safety constraints is coolant availability. That limitation is automatically configured into the control loop of a continuous reactor, illustrated in Figure 8.9dd. Here, the optimizing controller (OIC) detects the opening of the coolant valve, and if it is less than 90% open, it admits more feed by increasing the set point of the FRC. Thus, production rate is always maximized, but only within the safe availability of the coolant. When the

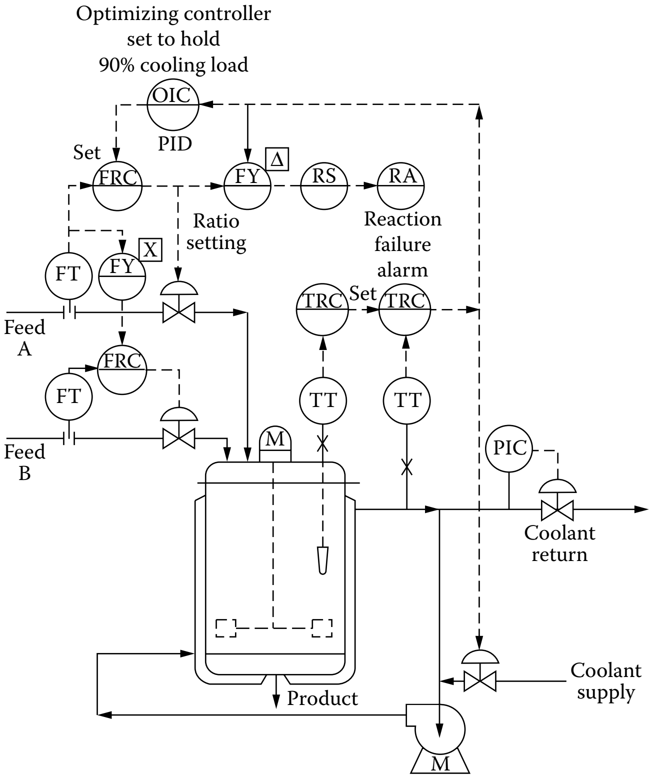


FIG. 8.9dd
The reaction rate in this continuous reactor is continuously matched to the full available capacity of the cooling system.

coolant valve opens beyond 90%, production rate is lowered so that the reactor will never be allowed to run out of coolant.

Another safety feature shown in Figure 8.9dd is the reaction failure alarm loop (RA). This loop compares the reactant charging rate to the reactor with the heat removal rate from the reactor and actuates an alarm if the charging rate is in substantial excess.

In Figure 8.9dd, the valve position controller (OIC) maintains the feed rate as high as the cooling system can handle. This results in a variable production rate, because changes in coolant temperature or variations in the amount of fouling of the heat-transfer areas will change production.

The dynamic response of the reactor temperature to changes in reactant flow is not favorable. A change in feed flow must change the reactant concentration before a change in reaction rate (and therefore in heat evolution) can change the reaction temperature.

Reactant concentration adds a large secondary lag to the valve position control loop. This necessitates the use of a three-mode controller,[5] which in turn limits the application to stable loops only. The slower the OIC loop (the longer its lag), the lower should be the set point of OIC in Figure 8.9dd, to provide the required margin for stability.

Maximizing Batch Reactor Production In a batch reactor, production rate increase with temperature, and therefore production can be maximized by maximizing temperature. This can be accomplished by a valve position controller that raises the batch temperature set point whenever the coolant valve is less than 90% open.

Unfortunately, the dynamic characteristics of this loop are also undesirable because of the inverse response of the loop.[5] When the coolant valve opens to more than 90%, the VPC will lower the set point of the temperature controller. This will temporarily *increase* the demand for coolant, although once the excess sensitive heat is removed, it will lower it. The longer the dead time introduced by this inverse response, the large the margin needed for safety and stability, and hence the lower the set point of the VPC.

When the concentration time constant is equal to or larger than the thermal time constant, temperature control through feed flow manipulation is no longer practical.[5] In such reactors, only the manipulation of the coolant will give stable operation. The strategy of valve position control is still useful in such applications, but in these cases it must manipulate supplementary cooling.

Heat Release Control

By multiplying the jacket circulation rate by the difference in temperature between inlet and outlet, it is possible to determine the amount of heat released or taken up by the reaction. The reaction is in an endothermic state when the value of Q in Figure 8.9ee is negative, and it is in an exothermic phase when Q is positive.

Under steady-state conditions, at high circulation rates the difference between T_o and T_i is about 5°F or less. The

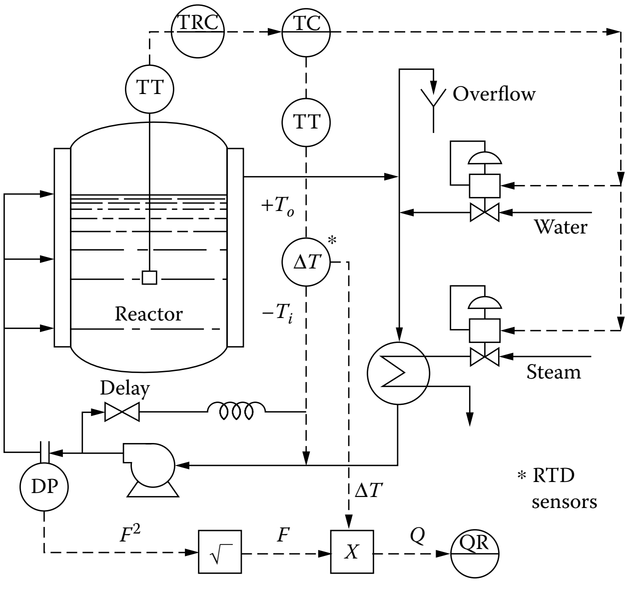


FIG. 8.9ee
To correctly determine the instantaneous amount of heat release, the jacket inlet temperature reading must be delayed by the residence time of the jacket, so that the ΔT transmitter will be dynamically compensated.[3]

method of reading this ΔT is critical, because during load changes, dynamic compensation is needed. As more cooling is requested, the jacket inlet temperature will drop, and the jacket outlet temperature will stay unaffected until the jacket contents are displaced. If Q is to be used for control purposes, such dynamic errors must be removed by compensation.

This is accomplished by delaying the jacket inlet temperature into the ΔT transmitter by a time equal to the delay through the jacket.[3] In Figure 8.9ee, this is accomplished by simulating the jacket using a length of tubing whose dead time is adjustable by changing the flow rate through it. A similar result can be obtained by placing an inverse derivative relay, set for a jacket displacement time of, say, 30 secs, into the transmitter output signal; the desired goal can also be reached by electronically simulating the transportation lag in digital systems.

The correct selection of the ΔT transmitter is very important, because in order to be able to detect Q within a ±1% error, the ΔT must be detected within ±0.2°F (±0.01°C). This can be accomplished only by using the best quality RTD-type transmitters that can provide spans, which are as narrow as –5 to 0 to +5°F. When the absolute value of the jacket temperature can vary from, say, 50 to 250°F throughout the reaction cycle, it will also be necessary to correct the calibration of the ΔT transmitter as the absolute temperature changes.

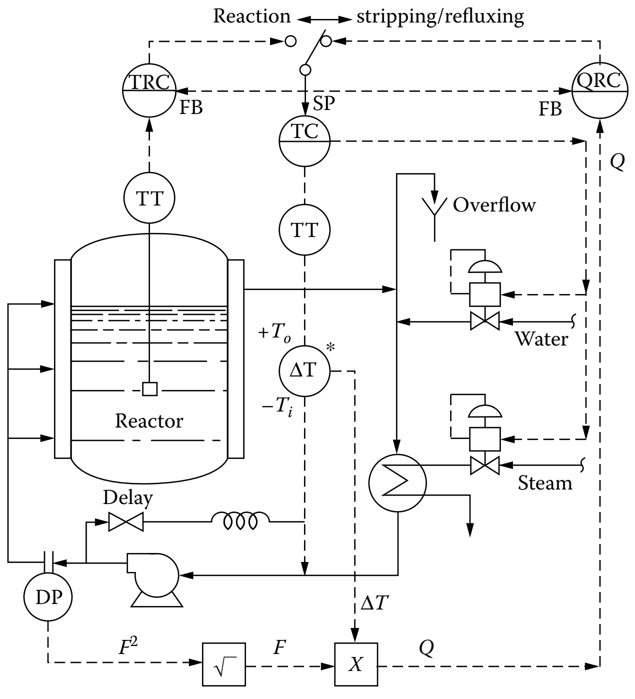
Once the value of Q is accurately determined, it can be used for many purposes. The instantaneous value of Q signals the rate of heat removal or addition, and the time integral of Q gives the total heat that has been added or removed. A change in the value of ΔT under standard conditions can signal fouling and the need for cleaning the reactor heat-transfer surfaces.

Stripping and Refluxing Controls

During stripping or refluxing phases, the reactor might be controlled on the basis of heat input (Q), because refluxing tends to be done at constant temperature, and the increase in temperature during stripping is usually too small for control purposes. Therefore, the system would be switched automatically from temperature control (Figure 8.9ee) to heat input control (Figure 8.9ff) whenever the reactor enters a stripping or refluxing phase.

The heat input during refluxing is usually set to be sufficient to maintain a state of slow boiling. During stripping, the heat input is usually set to complete the stripping in some empirically established time period; for example, an hour or two.

The time integral of Q represents the total reaction heat, which is an indicator of product concentration or percentage conversion. It can be used to introduce additives at predetermined conversions and to determine reaction endpoint. Through these automated steps, the need for taking grab samples can be eliminated. This in turn will result in reduced overall cycle time and, therefore, in increased production rate. The control of vacuum strippers will be discussed in the Pressure Control subsection.



* with RTD sensors

FIG. 8.9ff
Control can be switched from temperature control to heat input control as the reaction cycle enters a stripping or refluxing phase.

Constant Reaction Rate

Optimization of a chemical batch reactor usually means that its throughput is maximized while the operating costs are minimized. These goals can be met by simple, traditional techniques such as by optimizing the utility distribution in such a way that no reactor is held up in its reaction sequence because it has to wait for the availability of a particular resource.

Other obvious optimization techniques include the increasing of the reactant charging rates, the maximization of reactor heat-up or cool-down rates, and the maximized utilization of coolant availability during the exothermic reaction itself (Figure 8.9dd). When attempting to maximize the heat removal capability of a reactor, one can also precool the reactants prior to charging them to the reactor.[7]

Figures 8.9ee and 8.9ff show how the reaction rate can be measured in a reactor by detecting the exothermic heat release. Figure 8.9v shows a method to estimate the rate of heat release. Whether an accurate measure or just an estimate of the heat release rate (Q) has been made, that information can be used for control.

In Figure 8.9v, the heat release rate was used only as the basis for adjusting the gain of the TRC master and as a feedforward anticipator. However, it can do more, because by converting the nonlinear reaction rate of the batch into a linear rate, the batch reaction time is much reduced.

As shown in Figure 8.9gg, the batch reaction is completed much faster when the reaction rate is constant (linear) than when a nonlinear (first- or second-order) curve describes

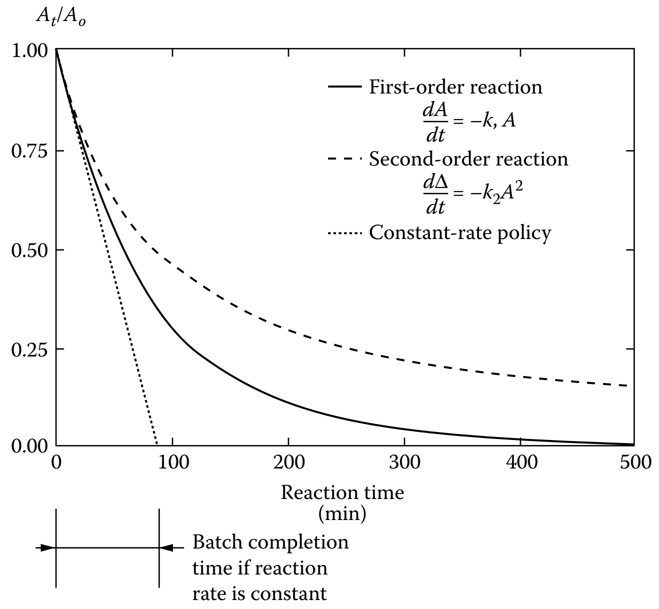


FIG. 8.9gg
The batch reaction time is a function of the reaction rate within the reactor. If constant rate is maintained, the batch reaction time can be drastically reduced.[7]

its time characteristics.[7] In order to keep the reaction rate from dropping off, the reaction temperature has to be increased (it is a nonisothermal batch reaction).

Changing the operation of a batch reactor from a constant reaction temperature to a variable reaction temperature mode (Figure 8.9b) is not an easy proposition to carry out.[11] One of the disadvantages of a nonisothermal operation is that the formed products might change as the reaction temperature changes. Yet, in some processes, the economic advantages outweigh the required complexities in modeling and control.

One possible configuration is to use the reactor shown in Figure 8.9v, in which the reactant is charged at constant rate, and use the heat-release estimator in that figure to modify the set point of TRC-1 in such a way that the reaction rate will be held constant. For some reactions the temperature settings required to provide constant reaction rates can be calculated, while for more complete reactions it is necessary to determine them experimentally.

Figure 8.9hh shows a cascade configuration in which the reaction rate controller (XRC-4) is the master, and it adjusts the batch temperature (TRC-1) set point so as to keep the reaction rate constant. The measurement to the reaction rate controller can come from a heat-release estimator (Figure 8.9hh) or from the actual measurement of the exothermic heat release (Q in Figures 8.9ee and 8.9ff).

ENDPOINT DETECTION

The reliable determination of batch cycle endpoint is important for two reasons:

1. One way to increase plant productivity is to reduce batch cycle time.

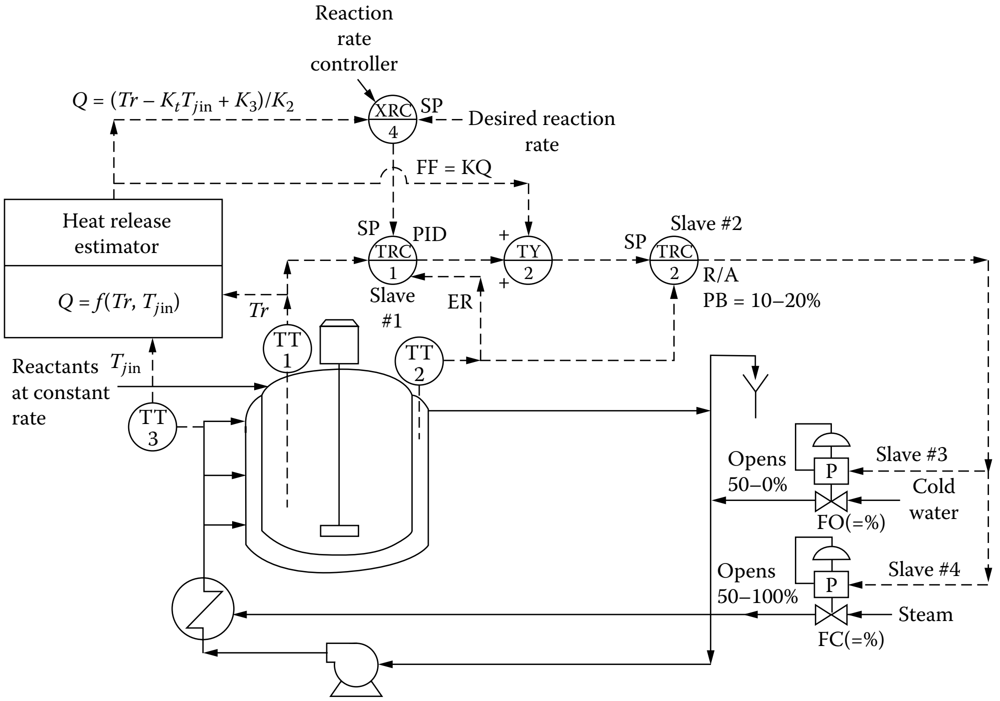


FIG. 8.9hh
Nonisothermal (variable reaction temperature) batch reactor control can shorten the batch cycle time by keeping the reaction rate constant.

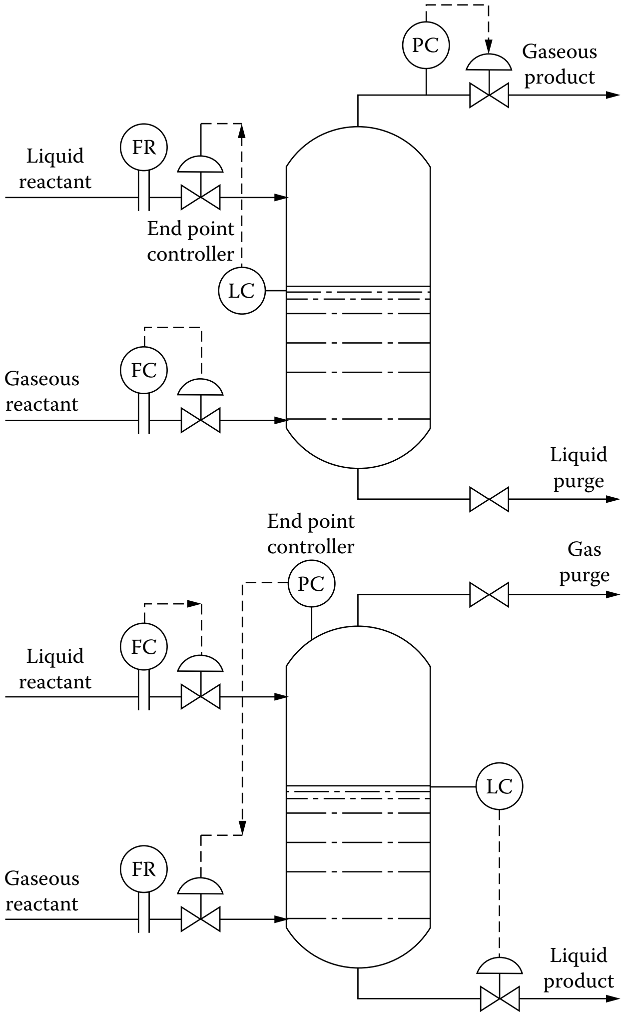


FIG. 8.9ii
One of the reactants can be added at the same rate as it is consumed (inventory control), if the phase of that reactant is different from the phase of the other streams entering the reactor.[3]

2. If the reaction continues beyond the optimal endpoint, the process yield can be reduced due to side reactions.

In batch reactors, if the completion of the reaction cannot be reliably detected, it is common practice to terminate reactant flow on the basis of total charge. This can be based on flow or weight measurements. In other reactions, the endpoint is accompanied by the fall or rise of the reaction pressure. This change in pressure can be used to have a pressure switch or pressure controller to shut off the reactant flow (Figure 8.9ii). In still other reactions, the total reaction heat signals the endpoint, as was illustrated in Figures 8.9ee and 8.9ff.

pH Endpoint If some analytical property, for example, pH, is used in detecting the endpoint, it is important to realize that during most of the reaction the measurement will be away from set point. Therefore, integral action must not be used, because it would saturate and overshoot would result. This is one of the few processes in which the proper selection of control modes is a proportional plus derivative with zero bias, modulating an equal-percentage reagent valve.

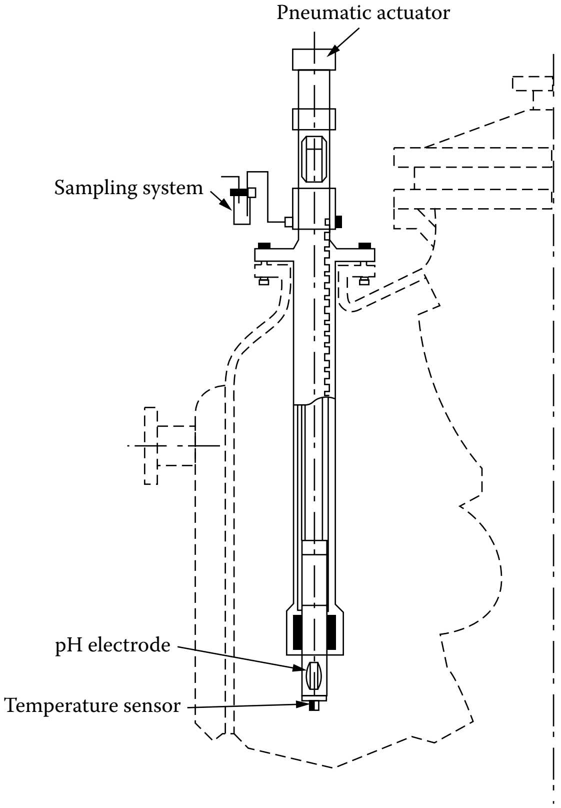


FIG. 8.9jj
The immersion-type pH probe can be used in chemical reactors.

If the derivative time is set correctly, the valve will shut quickly as the desired endpoint pH is reached. With excessively long derivative time setting, the valve will close prematurely, but this is not harmful because it will reopen as long as the pH is away from the set point.

For the measurement of pH, conductivity, and resistivity, on-line analyzers are available. The pH probe shown in Figure 8.9jj can be used for on-line measurement in a glass-lined reactor. When maintenance or recalibration is needed or when the reactor is empty, the probe can be lifted by a pneumatic actuator and immersed into buffer a solution to protect it from getting dry.

In some cases it might be preferred to keep the pH sensor outside the reactor and bring a sample of the reactor contents to it. Figure 8.9kk illustrates such a design configuration.

In other installations, such as in fermenters, the preference is for the use of side entry, retractable pH probes. These are illustrated in Figure 8.9ll. Other pH probe designs are described in Section 8.48 in Chapter 8 of the first volume of this handbook.

If one reactant is charged to the reactor in substantial excess to the others, this will usually guarantee the complete

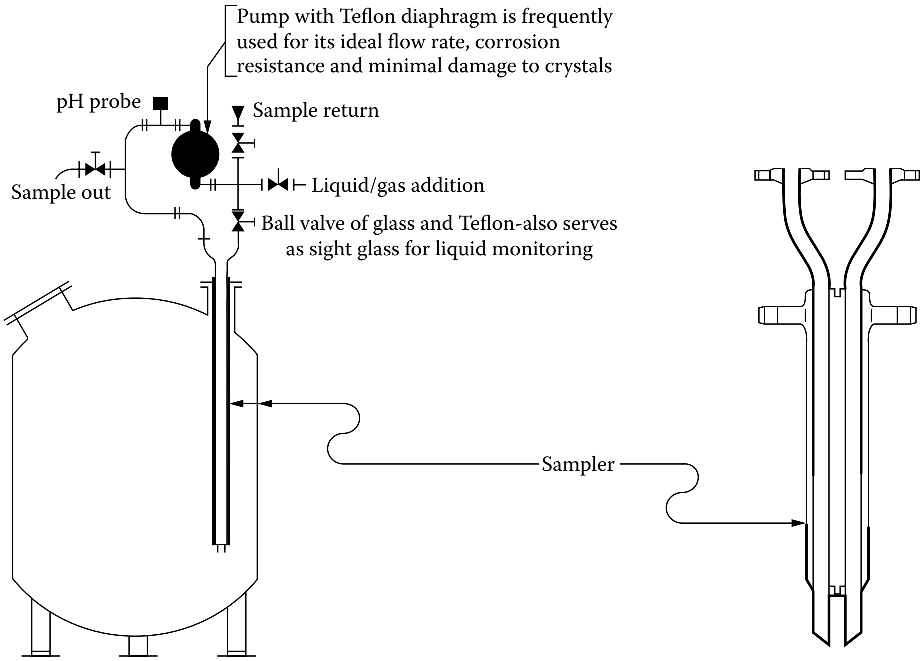


FIG. 8.9kk
The design components of a chemical reactor sampling system.[12]

conversion of the other reactants, and endpoint control is not required, unless the product is also used as the recycle stream. If the reactants are fed to the reactor in the same proportion in which they react, endpoint control is required. If the reactants and product are all in the same phase (all liquids or all gases), the endpoint is usually determined through the use of an analyzer, such as the pH unit.

On the other hand, when the reactants are in different phases, the endpoint can be controlled on the basis of level or pressure. As was shown in Figure 8.9ii, if the product is

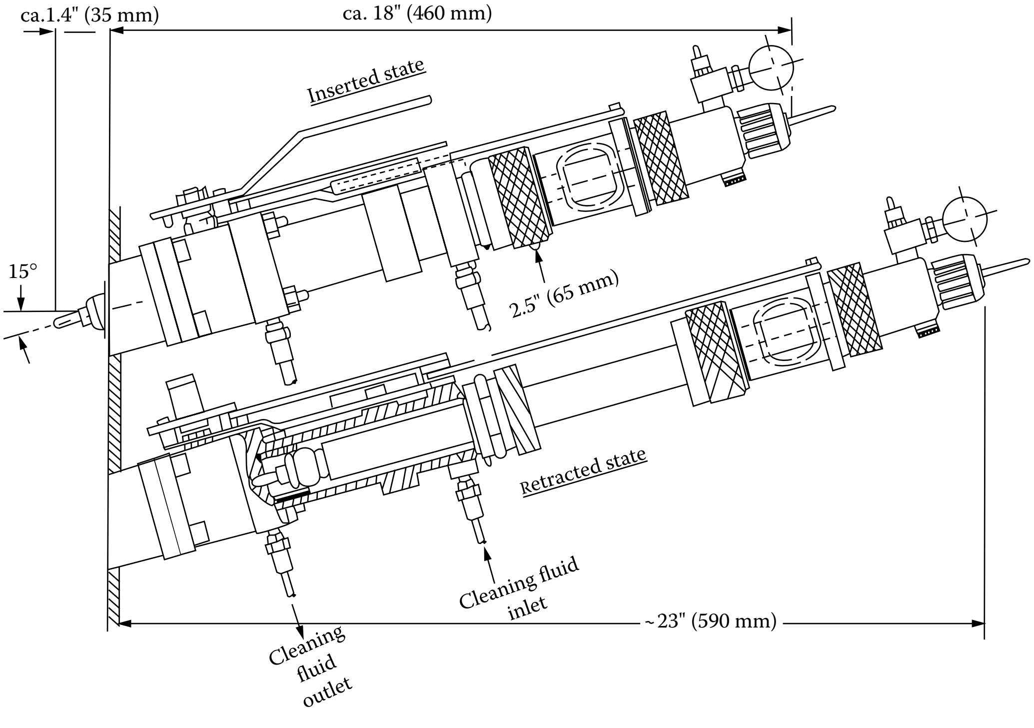


FIG. 8.9ll
Retractable, side-entry probe designs.

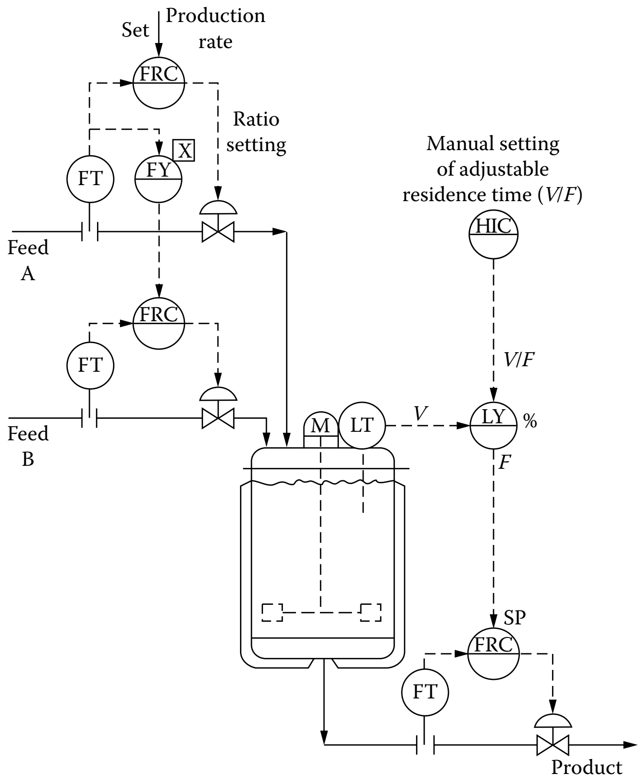


FIG. 8.9mm
Residence time control is accomplished by maintaining a constant ratio between reactor volume and reactor outflow.

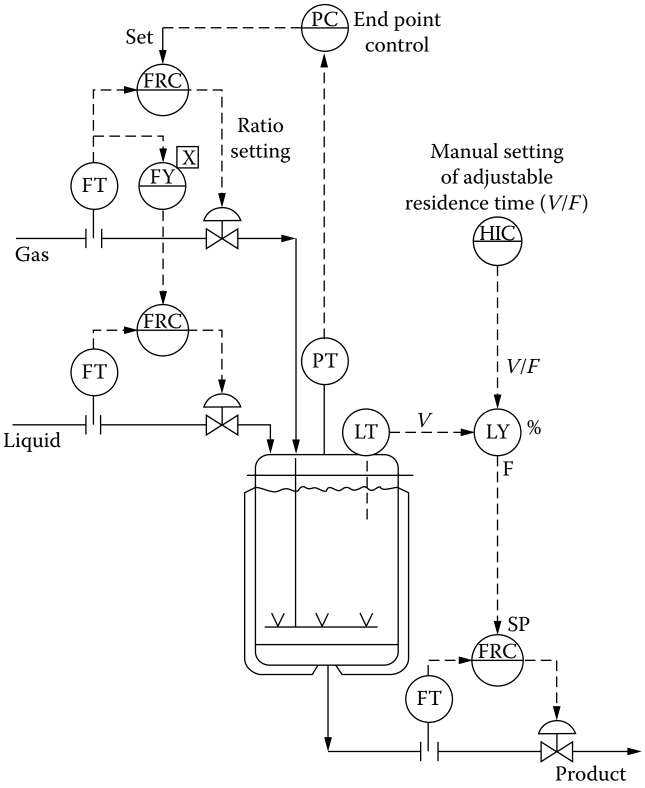


FIG. 8.9nn
In this control system, where the product is in the liquid phase, the endpoint control is based on pressure, while the residence time control is based on level.

gas, the endpoint is maintained by the level controller, whereas if the product is liquid, the endpoint control is based on pressure. When this pressure rises, that is an indication of excess gaseous reactant accumulation in the reactor. In response, the pressure controller (PC) will lower the gas flow; thereby, it will increase the liquid reactant concentration by increasing its relative flow. This in turn will increase gas consumption, and the balance will be reestablished and the correct endpoint maintained.

Residence Time Control

In continuous reactors it is usually desirable to provide a constant residence time for the reaction. If the production rate (inflow to the reactor) varies, the most convenient way of keeping the residence time constant is to maintain a constant ratio between reactor volume (V) and reactor outflow (F). The ratio V/F is the residence time, where the volume (V) can be approximated by a level measurement, and F can be measured as outflow. Figure 8.9mm illustrates a residence time control system using this approach.

All reactions with a liquid product can benefit from residence time control. This requires the combination of the features in the control system shown in Figures 8.9ii and 8.9mm, resulting in Figure 8.9nn.

Analyzers for Endpoint Detection

Probe-type analyzers do not require sampling systems, because the analyzer itself is inserted directly into the process. In the case of chemical reactors, analyzers can be inserted into either the reactor vessel or the discharge piping. Regardless of their locations, probes will give accurate readings only when clean. Therefore, automatic cleaning of such on-stream analyzer probes is a good idea.

Figure 8.9oo illustrates a flow-through, brush-type probe cleaner assembly that is provided with a sight glass, allowing visual observation of the cleanliness of the probe. Such units are also available with spray nozzle attachments for water or chemical cleaning of the electrode.

One common method of endpoint detection is the measurement of viscosity either by monitoring the torque on the agitator or by using viscosity sensors. Such a viscometer can be inserted in pipelines or directly into the reactor. It is essential that they be located in a representative area and be temperature-compensated.

In reactions in which the endpoint can be correlated to density, some of the more sensitive and less maintenance-prone densitometers can also be considered (Chapter 6 in Volume 1). The radiation-type design tends to meet these requirements if pipe line installation is acceptable. These

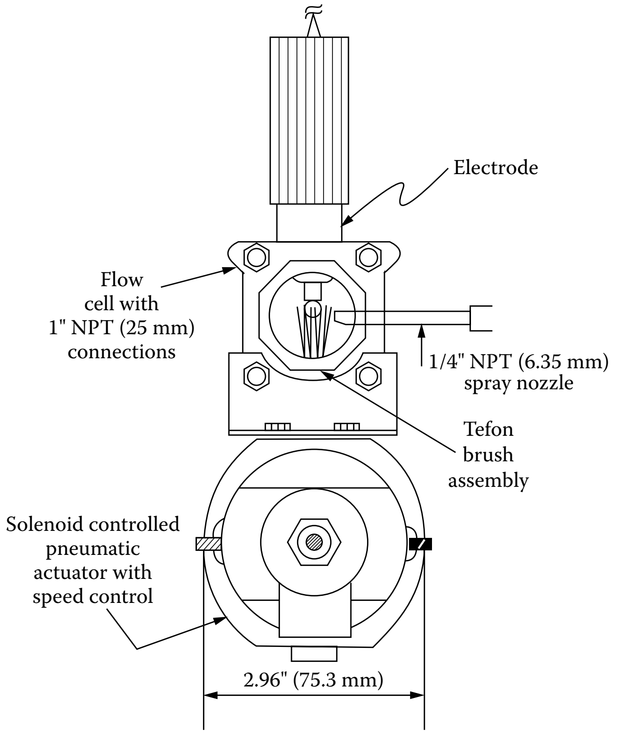


FIG. 8.9oo
Brush cleaner with optional water jet.

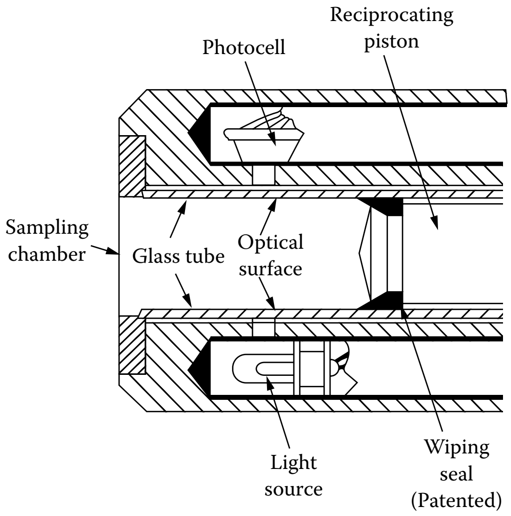


FIG. 8.9pp
Probe-type, self-cleaning suspended solids detector. (Courtesy of Monitek Technologies Inc., a Metrisa Company, http://www.monitek.com)

units can operate with a minimum span of 0.05 specific gravity unit and an error of ±1% of that span, or an absolute error of ±0.0005 specific gravity.

In a number of reactions, refractometers (Section 8.52 in Chapter 8 in Volume 1) have been found to be acceptable solutions to endpoint detection. In many installations, they are used as laboratory devices: A grab sample is brought to the bench analyzer. In other installations, they are used on-stream. When the refractive index (RI) analyzer is on-stream, human error tends to be reduced and analysis time is shortened, with a corresponding increase in productivity.

On the other hand, an on-stream installation requires more maintenance attention, because the RI measurement must be made to high precision, such as ±0.0005 RI units. This requires accurate temperature compensation, plus protection of the heat-sensitive photocells through cooling, which are time-consuming and expensive to replace when ruined by high temperature.

If the endpoint of a reaction is detectable by such measurements as a change in color or opacity, or the concentration of suspended solids, the self-cleaning probe shown in Figure 8.9pp can be considered. Here, the reciprocating piston in the inner cavity of the probe not only serves to clean the optical surfaces with its wiping seals but also guarantees the replacement of the sample in the sampling chamber by fresh material upon each stroking of the piston. The temperature considerations and limitations mentioned in connection with the RI analyzer also apply here.

If the endpoint of the reaction is detectable by infrared (IR) beam attenuation, such analyzers can also be put on-stream or can be inserted in pipelines and will detect the attenuated total reflectance of the IR beam. This type of analyzer is particularly suitable for the detection of water in hydrocarbons.

In addition to the analyzers discussed above a wide variety of computer-supported fiber-optic probes (FOPs) are also available for on-line analysis. Other analyzers, such as the gas chromatograph (GC), the high-pressure liquid chromatograph (HPLC), and the mass spectrometer (MS), are also powerful analytical tools and are available both as laboratory or on-line devices. For an in-depth discussion of all the analyzers available, refer to Chapter 8 in the *Process Measurement and Analysis* volume of this handbook.

Product Quality Control

Measuring the composition of the contents of a chemical reactor is a method that can help maintain product quality in continuous reactors and can detect endpoints in batch reactors. The analyzers are limited by sampling difficulties, by the intermittent nature of some models, and by incomplete mixing within the reactor. Because sampling and sample preparation result in dead time, sampling time should be minimized or eliminated altogether; this can be done by placing the sensor directly into the reactor, which results in tight control and fast response.

Because analyzers in general are low-reliability and high-maintenance devices, most users are reluctant to close an automatic loop around them. The main concern is that the failure of the analyzer might cause a hazardous condition by driving the reactor into an unsafe state. As shown in Figure 8.9qq, this concern can be alleviated by the use of

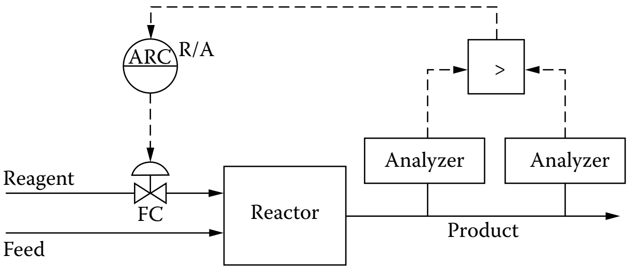


FIG. 8.9qq
If an unsafe condition can only arise in one direction (high or low concentration), a high or low signal selector can protect the reactor from the consequences of analyzer failure.[3]

redundant sensors through a high signal selector.[3] In case of a "downscale" failure of one of the analyzers, the backup unit takes over automatically. An "upscale" failure in this arrangement would shut down the reactor, providing safety at the cost of production.

If neither an accident nor an interruption of production can be tolerated, three analyzers are required, arranged in a "voting" configuration. In such an arrangement, if one of the sensors disagrees with the others, it is automatically disregarded and the "majority view" is used for closed-loop control. Maintenance and recalibration of the defective sensor are then initiated. Another solution is to use a median selector in combination with the three analyzers.

PRESSURE CONTROL

In gas phase reactions, in oxidation and hydrogenation reactions, or in high-pressure polymerization, the reaction rate is also a function of pressure. If, in a batch reaction, the process gas is completely absorbed, the controls in Figure 8.9rr would apply. Here, the concentration of process gas in the reactants is related to the partial pressure of the process gas in the vapor space. Therefore, pressure control results in the control of reaction rate. This loop is fast and easily controlled.

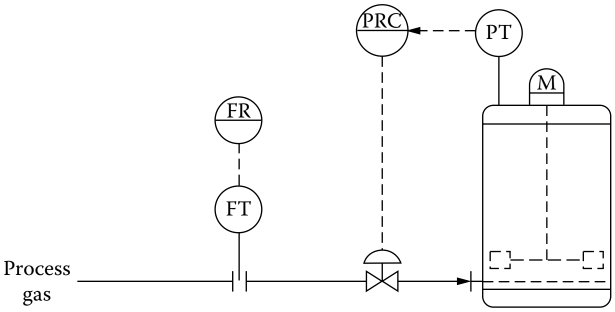


FIG. 8.9rr
Reactor pressure controlled by gas makeup.

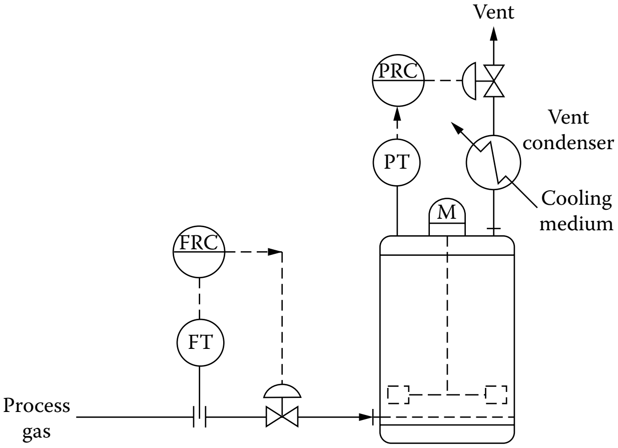


FIG. 8.9ss
Reactor pressure controlled by vent throttling.

Certain reactions not only absorb the process gas feed but also generate by-product gases. Such a process might involve the formation of carbon dioxide in an oxidation reaction. Figure 8.9ss illustrates the corresponding pressure control system. Here, the process gas feed to the reactor is on flow control and reactor pressure is maintained by throttling a gas vent line. This particular illustration also shows a condenser, which is used to minimize the loss of products through the vent.

In the case of continuous reactors, a system such as shown in Figure 8.9tt is often employed. Here the reactor is liquid full, and both the reactor liquid and any unreacted or resultant by-product gases are relieved through the same

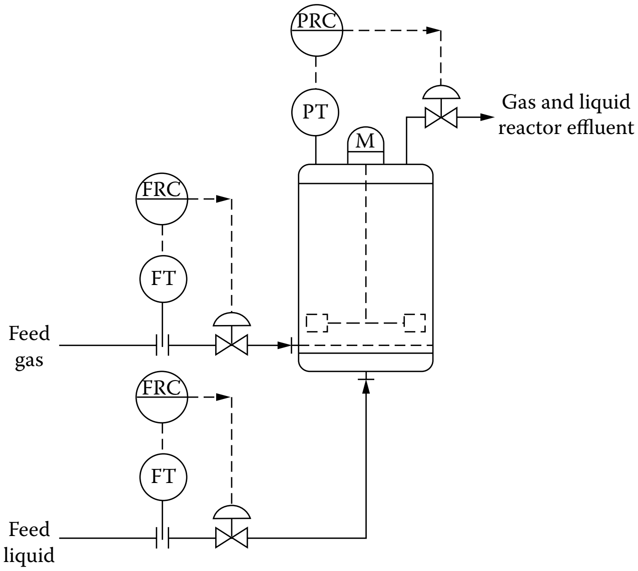


FIG. 8.9tt
Arrangements such as this can be used to control the pressure in continuous reactors.

outlet line. Reactor pressure is sensed, and the overflow from the reactor is throttled to maintain the desired operating pressure. Process gas feed and process liquid feed streams are on flow control.

All these illustrations are simplified. For instance, it may be desirable to place one of the flow controllers on ratio in order to maintain a constant relationship between feed streams. If the reaction is hazardous, and there is the possibility of an explosion in the reactor (for example, oxidation of hydrocarbons), it may be desirable to add safety devices, such as a high-pressure switch, to stop the feed to the reactor automatically.

If one of the reactants differs in phase from both the other reactants and the product, inventory control can be applied (Figure 8.9ii). In the case of gaseous products, the reactor level is controlled by modulating the liquid reactant; with liquid products, the gaseous reactant is modulated to keep reactor pressure constant. A purge is needed in both cases to rid the reactors of inerts either in the liquid or in the gas phase.

Vacuum Control

Some reactions must be conducted under vacuum. The vacuum source is frequently a steam jet-type ejector. Such units are essentially venturi nozzles with very little turndown. This constant-capacity vacuum source is frequently matched to the variable capacity reactor by wasting the excess capacity of the ejector. This is done by creating an artificial load through the admission of ambient air.

When the vapor generation rate in the reactor is low, the steam jet is still operating at full capacity, sucking in and ejecting ambient air. The corresponding waste of steam can be substantially reduced if the system shown in Figure 8.9uu is applied. Here, a low-load condition is detected by the fact that the air valve is more than 50% open, and this automatically transfers the system to the "small" jet. A later increase in load is detected as the air valve closes (at around 5%), at which point the system is switched back to the high-capacity jet.

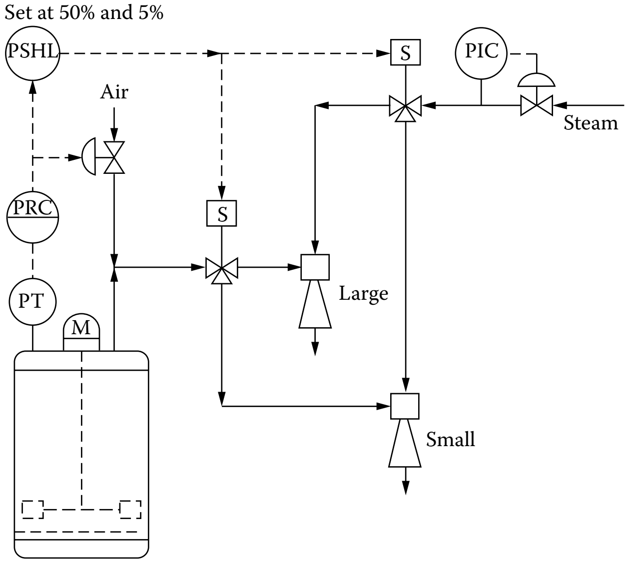


FIG. 8.9uu
Cost-effective operation of vacuum pressure controls can substantially reduce the waste of steam.

Vacuum Stripping

When stripping is done by direct steam injection, the controls shown in Figure 8.9ff are not applicable. The controls of a vacuum stripper serve to remove the solvent and the unreacted monomers from the batch in minimum time but without causing foaming.

The steam addition to the stripper (FIC in Figure 8.9vv) is usually programmed as a function of time (FY). As the batch gets more concentrated with polymer products and leaner in monomers or solvent, the more likely it is to foam, and therefore the rate of addition is reduced. This is usually controlled by hardware or a software program (FY) that relates steam flow to time. This relationship is adjustable, and if foaming is experienced during a stripping cycle (LSH in Figure 8.9vv), the steam rate curve programmed in FY can be lowered to avoid the repetition of foaming in the next batch.

The temperature of the batch is a function of (1) the balance between the amount of heat introduced by the steam and (2) the amount of cooling caused by the vaporization of the monomers. The vacuum level in the stripper is therefore controlled by temperature. This guarantees that the vacuum is low enough so that all the solvents and monomers are removed, but that it does not drop too low, to the point where the stripping steam would condense and would not be removed by the vacuum but would cause dilution of the batch.

REACTOR SAFETY

For a more detailed discussion of plant and control system safety, the reader is referred to Sections 4.11 and 5.8 of this volume. Here, only some of the "reactor-specific" safety concerns will be considered.

Safety problems can arise in chemical reactors as a result of many causes, including equipment failure, human error, loss of utilities, or instrument failures. Depending on the nature of the problem, the proper response can be to "hold" the reaction sequence until the problem has been cleared or to initiate an orderly emergency shutdown sequence. Such actions can be taken manually or automatically.

After a "hold" or "emergency" condition has been cleared, an orderly sequence of transitional logic is required to return the reactor to normal operation. It makes no difference whether the emergency was caused by a pump or valve failure or whether the reactor was put on hold to allow for

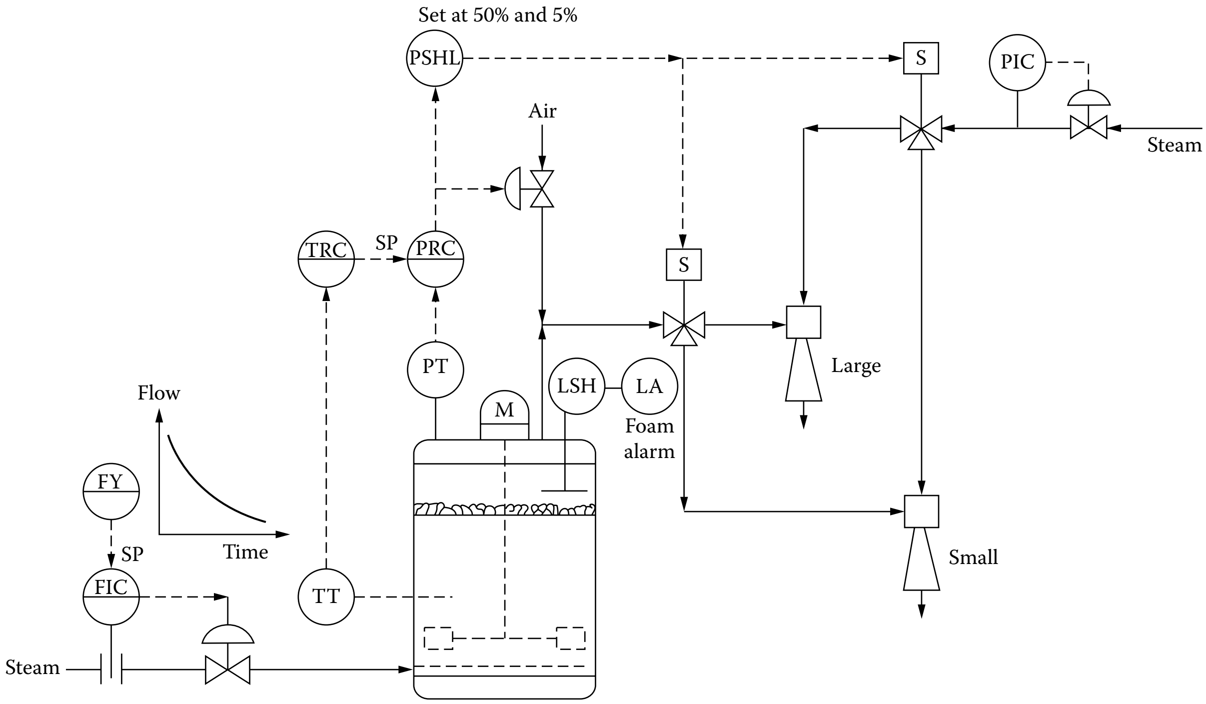


FIG. 8.9vv
The controls of a vacuum stripper serve to remove the solvent and the unreacted monomers from the batch in minimum time but without causing foaming.

manual sampling and laboratory analysis of the product: A return sequence is still required. The reentry logic determines the process state when the interruption occurred and then decides whether to return to that process state or the previous one in order to reestablish the conditions that existed at the time of the interruption.

A reactor control system should provide the following features:[9]

- The ability to maximize production
- The ability to minimize shutdowns
- The ability to maximize on-line availability of the reactor
- The ability to minimize the variations in utility and raw material demand
- The ability to provide smooth operation in terms of constant conversion, yield, and product distribution
- Easy start-up and shutdown

However, the overriding, primary design objective is that the reactor must be safe for both the operating personnel and the environment.

Safety is guaranteed by monitoring both the potential causes of emergencies and their consequences. If the cause is detected and responded to in time, the symptoms of an emergency will never develop. It is for this reason that the monitoring of potential causes of safety hazards is preferred. Some of the potential causes of emergencies and the methods used to sense them are listed in Table 8.9ww.

Each of the above failures can actuate an annunciator point to warn the operator and can also initiate a hold or shutdown sequence, if the condition cannot be corrected automatically. Automatic correction will require different responses in each case.

TABLE 8.9ww
Methods of Detection and Potential Causes of Reactor Emergencies

Causes	*Methods of Detection*
Electric power failure	Volt or watt meters
Instrument air failure	Pressure switch
Coolant failure	Pressure or flow switches
Reactant flow failure	Pressure or flow switches
Catalyst flow failure	Pressure or flow switches
Loss of vacuum	Vacuum switch
Agitator failure	Torque or RPM detector
Pump failure	Pressure or flow switch
Valve failure or mispositioning	Position-sensing limit switches
Measuring device failures	Validity checks

TABLE 8.9xx
Unsafe Reactor Conditions and the Sensors That Can Detect Them

Symptom	*Detectors*
High reactor temperature	High temperature switch
High jacket temperature	High temperature switch
Low jacket temperature	Low temperature switch
High reaction rate	High heat release detector
Reaction failure	Low heat release detector
High rate of temperature rise	Rate of rise detector
High pressure	High-pressure switch and relief valve
Abnormal level	Level and weight switches
Abnormal composition	On-stream analyzers

For example, in case of electric power failure, a diesel generator or steam turbine might be started. If instrument air failure is detected, an alternate compressor might be started, or an alternate gas, such as dry nitrogen, might be used. In addition, the safe failure positions for all control valves must be predetermined, as in Figure 8.9m. In case of loss of coolant, an alternate source of coolant might be used, or the type of automatic constraint control that was shown in Figure 8.9dd might be applied.

It is also desirable to provide intelligent human–machine interfaces that will protect against unsafe operator actions. The simplest example of such limits is to prevent the operator from moving set points outside safe limits.

If reactant or catalyst flow fails, automatic ratio loops (Figure 8.9dd) can cut back the flow rates or related streams; alternatively, if there are alternate storage tanks from which the reactant can be drawn, these tanks might be accessed. A loss of vacuum can also be corrected by switching to an alternate vacuum source if such redundancy is economically justifiable.

Critical pumps and valves can be provided with spare backup units, and agitators can be furnished with multiple drives. Failed sensors and defective instruments can be automatically replaced by redundant and voting systems.

If the cause of an emergency goes undetected, eventually it will affect the operating conditions of the reactor. These symptoms might involve temperature, pressure, composition, or other conditions listed in Table 8.9xx.

Runaway Reactions

The control system will respond to the symptom of an abnormal process variable condition much as it did to the previously discussed failure causes, but here the probability of an emergency shutdown is higher. For example, if, in a potentially unstable reactor, the temperature or pressure has exceeded its high limit, it is possible to make sure that the reactor is on full cooling and that the reactant feeds are shut off, but after that little else can be done. Once such conditions have evolved, only drastic action can prevent a runaway reaction, such as depressurization or even the transferring of the reactor contents into a blow-down tank.

The consequences of high jacket temperature are less serious, and the limit shown in Figure 8.9l is usually sufficient to prevent the formation of hot spots. In some processes, the critical consideration is low jacket temperature,[9] because that can cause "frosting" or "freezing" the reactants onto the inner wall of the reactor. As the frozen layer thickens, heat transfer is blocked and the reactor must be shut down. If used on such a process, the control system in Figure 8.9l would also be provided with a low limit on the slave set point.

When the reaction rate suddenly rises or when it approaches the capacity of the cooling system, it is necessary either to increase the heat removal capacity quickly (Figures 8.9o and 8.9bb) or to cut back on the reactant flows (Figure 8.9i). It is also necessary to detect if the reaction has failed by comparing the reactant feed rate with the heat removal rate (Figure 8.9dd) or by monitoring heat release (Figure 8.9ee). In yet other processes, it is desirable to limit the rate of temperature rise (Figure 8.9bb).

Glass Lining Protection

Some of the chemical reactors are glass-lined to resist corrosion or to protect against contamination or discoloring of the product. When a reactor is glass-lined, the integrity of the lining can be damaged, because the thermal expansion characteristics of glass and metal are different. Therefore, it is necessary to provide control systems that serve to protect the lining against cracking (Figure 8.9yy) and that are able to detect if such damage has occurred (Figure 8.9zz).

The integrity of various glass-lining materials are guaranteed for different temperature differentials between the metal and the glass. The lining protection controls serve to keep the difference between the temperature of the lining (batch temperature detected by TT-1 in Figure 8.9yy) and the temperature of the metal (jacket temperature detected by TT-2) under the guaranteed limit.

Figure 8.9yy illustrates the protection loop for a situation in which the maximum differential is 100°. The goal of the controls is to keep the set point of the slave (TRC-2) within 100° of the batch temperature (TT-1). The purpose of TY-3 and TY-4 is to make sure that the slave set point does not drop too low. Therefore, TY-3 subtracts 100° from the TT-1 signal, while TY-4 selects the higher between that signal and the master (TRC-1) output. Similarly, the purpose of TY-5 and TY-6 is to prevent the slave set point from rising too high. TY-5 adds 100° to the TT-1 signal, while

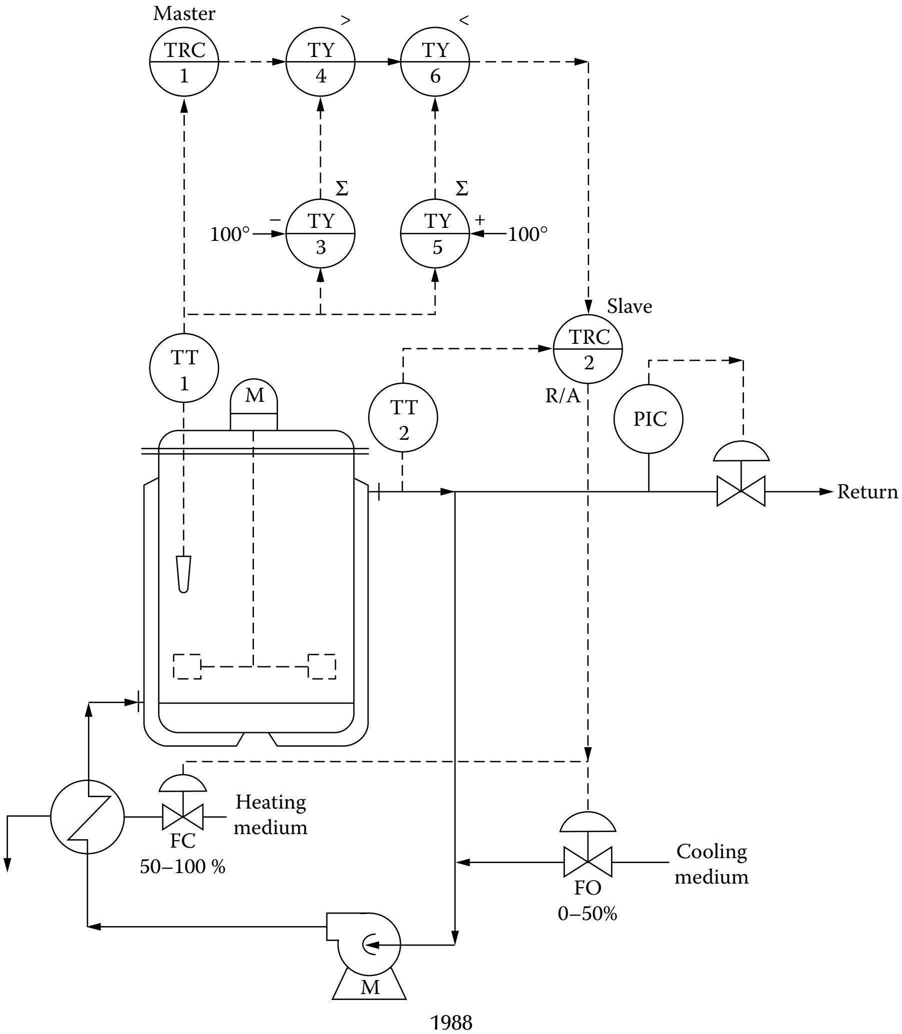


FIG. 8.9yy

The illustrated glass-lining protection strategy guarantees that the slave set point (jacket temperature) will never be more than 100° away from the batch temperature.

TY-6 selects the lower between that signal and the TRC-1 output.

Figure 8.9zz shows the control system used to monitor the integrity of the glass lining, when the batch is electrically conductive. The fault test is made by inserting a probe into the batch and checking the resistance between that and the metallic tank wall. If the glass lining is damaged, the resistance drops as the metal is exposed to the conductive batch. The fault monitor shown in Figure 8.9zz is designed for the scanning of ten glass-lined reactors.

Multiple Sensors

In connection with the topic of increasing the reliability of composition analyzers, we have already seen some options on using redundant measurements in our control systems. Figure 8.9qq described one possible option, where a pair of redundant analyzers in combination with a high signal selector can increase reliability.

Multiple temperature sensors can also be used for safety or maintenance reasons. In a fixed-bed reactor, for example, in which the location of the maximum temperature might shift as a function of flow rate or catalyst age, multiple sensors would be installed and the highest reading selected for control, as shown in Figure 8.9aaa.

On the other hand, if the reason for the use of multiple sensors is to increase the reliability of the measurement, the use of median selectors is the proper choice (Figure 8.9bbb). A median selector rejects both the highest and the lowest signals and transmits the third. This type of redundancy protects against the consequences of sensor misoperation, while

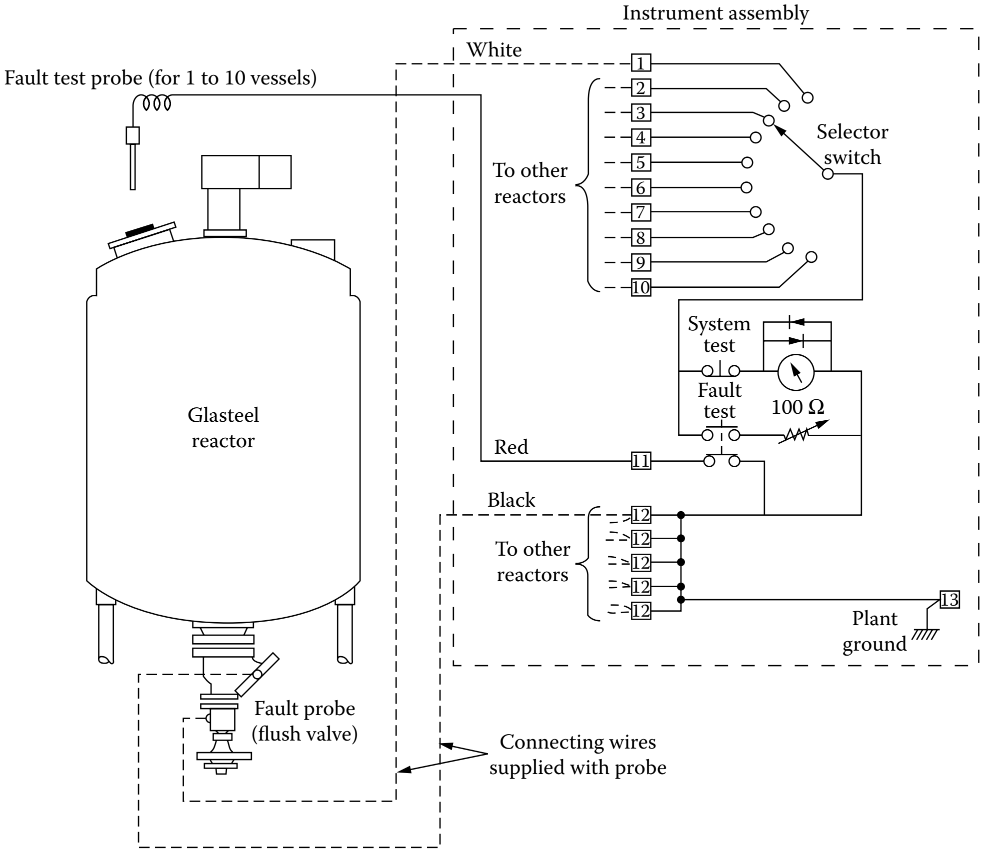


FIG. 8.9zz

The illustrated fault finder signals if the glass lining of a reactor is damaged. (Courtesy of Factory Mechanical Systems.)

also filtering out noise and transients that are not common to two of the signals.[5]

Another method of increasing sensor reliability is the use of voting systems. These also consist of at least three sensors. Reliability is gained, as the voting system disregards any measurement that disagrees with the majority view.

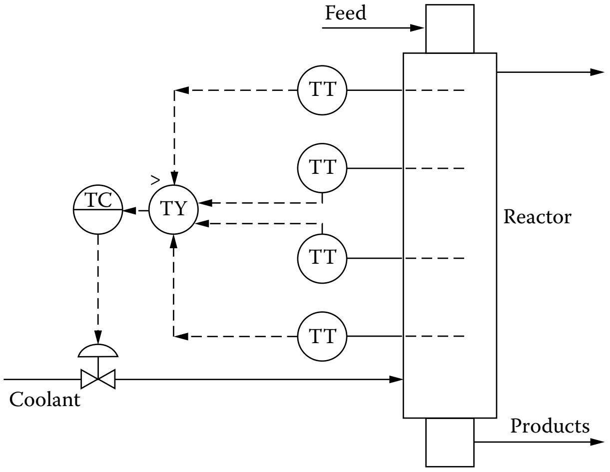
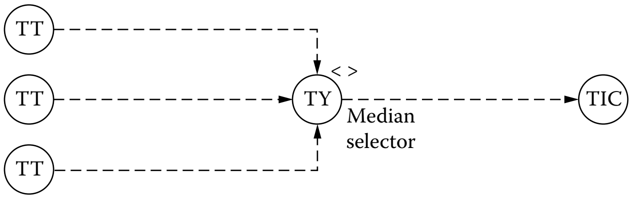


FIG. 8.9aaa

This method can be used to control the highest temperature in a fixed-bed reactor.[3]

Instrument Reliability

A risk analysis requires data on the failure rate for the components of the control system. Such data has been collected from user reports; the findings of one such survey are shown in Table 8.9ccc. When the mean time between failure (MTBF) of the individual instrument components has been



*

FIG. 8.9bbb

Median selector rejects both the highest and the lowest signals and can be the logical choice in some applications.

TABLE 8.9ccc

*Failure Rate for Control System Components**

Variable	*Instrument*	*Mean Time between Failures*
Level	Bubbler	1–2 yrs.
	d/p transmitter	1–5 yrs.
	Float and cable	0.2–2 yrs.
	Optical	0.1–5 yrs.
Flow	Flume and weir	0.5–5 yrs.
	Venturi, etc.	2 mo.–5 yr.
	Propellers	1 mo.–1 yr.
	Positive displacement	1 mo.–1 yr.
	Magnetic	0.5–10 yrs.
Density	Nuclear	1–3 yrs.
	Mechanical	1–6 mos.
Analysis	pH and ORP	1–4 mos.
	Dissolved O_2	1–9 mos.
	Turbidity	1–6 mos.
	Conductivity	1–4 mos.
	Chlorine gas	0.5–1 yr.
	Explosive gas	0.2–1 yr.
	TOC	0.1–1 mo.
Miscellaneous	Temperature	0.5–2 yrs.
	Pressure	0.1–5 yrs.
	Speed	0.6–5 yrs.
	Weight	0.6–2 yrs.
	Position	0.1–1 yr.
	Sampling	0.1–1 yr.

*Adapted from Reference 10.

estimated, the MTBF of the various control loops can be established.[9]

An example of predicting the MTBF of a system, based on the MTBF of its components, is illustrated in Table 8.9ddd. Once the MTBF of each loop has been established, the reliability of the whole reactor control system can be increased by increasing the MTBF of the loops that pose the highest risks. This usually is done through the use of self-diagnostics, preventive maintenance, backup devices, and, in critical instances, the technique of "voting," which was described in connection with Figure 8.9bbb.

While risk analysis is complex and time-consuming (see Sections 4.11 and 5.8), a plant can be designed for a particular level of safety, just as it can be designed for a particular level of production. The reliability of the result is as good as the data used in the analysis. For this reason, data collected by users, testing laboratories, or insurance companies should be used, instead of manufacturers' estimates. For instrument reliability and performance data, good sources are the International Associations of Instrument Users, SIREP and WIB.

TABLE 8.9ddd

The Mean Time between Failure (MTBF) of a Control Loop Can be Calculated by Considering the MTBF of its Components[9]

	Flow Transmitter	*Flow Totalizer*	*Solenoid Valve*	*On/Off Valve*
Failures/10^6	10	10	30	50
Operations/Process Cycle	100	100	2	2
Mean Cycles between Failure	1,000	1,000	16,000	10,000
Mean Time between Failure, Days	250	250	4,000	2,500

CONCLUSIONS

As was discussed in the previous paragraphs, the goals of good chemical reactor control and optimization include the following:

- Precise recipe charging
- Minimum heat-up time without temperature overshoot
- Maximized reaction rate within limits of coolant availability
- Accurate and fast endpoint determination

These goals can be met by various levels of automation. As the requirements for safety and optimized production are increased, higher and higher levels of automation become necessary. The trend is towards model-based anticipation and towards the integration of the individual control loops into integrated multivariable reactor controller systems.[13]

References

1. *Control Roundup,* Vol. 4, No. 1, January 1984, The Foxboro Company.
2. Luyben, W. L., "Batch Reactor Control," *Instrumentation Technology,* August 1975.
3. Shinskey, F. G., *Process Control Systems,* 2nd edition, New York: McGraw-Hill, 1979.
4. Lipták, B. G., *Optimization of Unit Operations,* Radnor, PA: Chilton Book Company, 1987.

5. Shinskey, F. G., *Controlling Multivariable Processes,* Research Triangle Park, NC: Instrument Society of America, 1981.
6. Lipták, B. G., "Process Control and Optimization," CFPA Course, Amsterdam, March 1993.
7. Juba, M. R., and Hamer, J. W., "Progress and Challenges in Batch Process Control," *Chemical Process Control-CPC III,* Amsterdam, 1986.
8. Szeifert, F., Chovan, T., and Nagy, L., "Dynamic Simulation and Control of Flexible Chemical Technologies," Automation '92, Budapest, 1992.
9. Buckley, P. S., "Protective Controls for a Chemical Reactor," *Chemical Engineering,* April 20, 1970 and April 4, 1975.
10. Molvar, A. C., "Instrumentation and Automation Experiences," U.S. Environmental Protection Agency publication no. EPA-600/2-76-198, October 1976.
11. Kladko, M., "The Case of a Real Engineering Design Problem," *Chem-tech,* 1971.
12. "System Samples Material without Opening Manway," *Chemical Processing,* November 1986.
13. Lipták, B. G., "Process Control Trends—Parts 1 & 2," *Control,* March and May, 2004.

Bibliography

Alsup, M., "Recipe Configuration for Flexible Batch Systems," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

ANSI/ISA-88.00.03-2003, "Batch Control Part 3: General and Site Recipe Models and Representation," Research Triangle Park, NC: Instrumentation, Systems, and Automation Society (ISA), 2003.

ARC Report: "Batch Process Automation Strategies," ARC Advisory Group, October 1999.

Armstrong, W. S., and Coe, B. F., "Multiproduct Batch Reactor Control," *Chemical Engineering Progress,* January 1983.

Benett, C. E., "Evaluating Controls for Batch Processing," *Instrumentation Technology,* July 1979.

Bonvin, D., and Srinivasan, B., "Optimal Operation of Batch Processes via the Tracking of Active Constraints, a Case Study,: *ISA Transactions,* Research Triangle Park, NC: ISA, 2003.

Branch, T., and Ray, T., "Batch Automation Project Increases Production," *Proceedings of the World Batch Forum,* Orlando, FL, April 2001.

Bristol, E. H., "A Design Tool Kit for Batch Process Control," *InTech,* October 1985.

Buckley, P. S., "Designing Override and Feedforward Controls," *Control Engineering,* Part I, August 1971; Part II, October 1971.

Carlo-Stella, G., "Distributed Control for Batch Systems," *InTech,* March 1982.

Christie, D., "A Methodology for Batch Control Implementation–Real World Lessons," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Clark, D. W., and Gawthrop, P. J., "Implementation and Application of Microprocessor-Based Self-Tuners," *Automatica,* 17(1), 1981.

Corripio, A. B., *Design and Application of Process Control Systems*, Research Triangle Park, NC: ISA, 1998.

Crowl, T., and Wolin, D., "The Paperless Batch Plant," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Emerson, D., "What Does a Procedure Look Like? The ISA S88.02 Recipe Representation Format," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

Fihn, S. L., "Specifying Batch Process Control Strategies," *InTech,* October 1982.

Fleming, D. W., and Schreiber, P. E., "Batch Processing Design Example," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Ghosh, A., "Batch Control State of Art," *Instrument Engineers' Handbook,* 3rd edition, Process Software and Digital Network, Sec. 5.5, July 2002.

Ghosh, A., "Maximizing the Potential of Batch Process Control," *Proceedings of the World Batch Forum,* Brussels, Belgium, October 2000.

Hancock, J., and Hopkinson, P., "A Case History of the Implementation of an S88-Aware Batch Control System," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Hodson, A. J. F., and Clarke, D. W., "Self-Tuning Applied to Batch Reactors," *Proceedings of the IEE Vacation School on Industrial Digital Control Systems,* London: IEE, September 1984.

Hopkins, B., and Alford, G. H., "Temperature Control of Polymerization Reactors," *Instrumentation Technology,* May 1973, pp. 39–43.

Horak, J., and Jiracek, F., "Adaptive Nonisothermal Computer Control of a Batch Reactor," *Collection of Czechoslovakia Chemical Communicable,* 1983.

Huzmezan, M., et. al., "Advanced Control of Batch Reactor Temperature," *Proceedings of the International Society for Measurement and Control (ISA) Expo,* 2000.

John, B., "Batch Control Systems Current and Future," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2002.

Jutan, A., and Uppal, A., "Combined Feedforward-Feedback Servo Control Scheme for an Exothermic Batch Reactor," *Industrial Engineering Chemical Process Design and Development, 23,* 1984.

Kawano, K., "Applications of Batch Progress Prediction by Rolling Scheduling to Operation Support System," *Proceedings of the World Batch Forum,* Baltimore, MD, April 1998.

Krigman, A., "Guide to Selecting Weighing, Batching, and Blending Systems," *InTech,* October 1982.

Longwell, E. J., "Endo/Exo Batch Reaction Requires Complex Temperature Control," *InTech,* August 1986.

Lorenzo, D., and Pérez, N., "Applying Model Predictive Control to a Batch Process," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Luyben, W. L., "Batch Reactor Control," *Instrumentation Technology,* August 1975.

Luyben, W. L., Tyreus, B. D., and Luyben, M. L., *Plantwide Process Control*, New York: McGraw-Hill, 1998.

Lynch, E. P., *Applied Symbolic Logic,* New York: John Wiley & Sons, 1980.

Marroquin, G., and Luyben, W. L., "Experimental Evaluation Cascade Controllers for Batch Reactors," *Industrial Engineering Chemical Fundamental,* 1972.

Mehta, G. A., "The Benefits of Batch Process Control," *Chemical Engineering Progress,* October 1983.

Parshall, J. H., and Lamb, L. B., "Applying S88: Batch Control from a User's Perspective," Research Triangle Park, NC: ISA, 1999.

Procyk, L. M., "Batch Process Automation," *Chemical Engineering,* May 1991.

Schnelle, P. D., and Richards, J. R., "A Review of Industrial Reactor Control: Difficult Problems and Workable Solutions," *Chemical Process Control-CPC III,* Amsterdam, 1986.

Shaw, W. T., "Computer Control of Batch Processes," *EMC Controls,* 1982.

Shaw, W. T., "Structured Design Produces Good Batch Programs," *Control Engineering,* 30(12), 1983.

Shinskey, F. G., "Controlling Temperature in Batch Reactors," *In Tech,* April 1992.

Shinskey, F. G., "Exothermic Reactors: The Stable, the Unstable, the Uncontrollable," *Chemical Engineering,* March 2004.

Shinskey, F. G., *Process-Control System,* New York: McGraw-Hill, 1979.

Shinskey, F. G., *Process Control Systems*, 4th edition, New York: McGraw-Hill, 1996, p. 317.

Smith, F. A., "Development of Remote Auto-Manual Control System for Batch Organic Processes," *Chemical Industry,* 1984.

Staus, R., "Improved Temperature Control in Batch Production Systems," *Proceedings of the International Society for Measurement and Control* (ISA), 2002.

"Temperature Control for Batch Processes," *Chemical Engineering,* April 1984.

Thome, R. J., Cline, M. W., and Grillo, J. A., "Batch Process Automation," *Chemical Engineering Progress,* 75(5), 1979.

Tom, T. H., "DCS Selection for Flexible Batch Automation," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Ward, J. C., "Digital Batch Process Control," *InTech,* October 1982.

Wensley, J. H., "Fault-Tolerant Control for Batch Processes," *In Tech,* October 1982.

Wooley, K., and Christie, D., "Achieving Increased Batch Plant Productivity by Application of a Multilevel DCS Solution," *Proceedings of the ISA,* Houston, October 1998.

8.10 Chemical Reactors: Control and Optimization

F. G. SHINSKEY (2005)

INTRODUCTION

It should be noted that the coverage of this section and of the previous one (8.9) are similar. If the reader is interested in a detailed and complete treatment of the subject of chemical reactor control, it is recommended to also read that section. On the other hand, if you have little time and you are an experienced process control engineer who is familiar with the basics of chemical reactor control and want only to quickly refresh your memory about the most important aspects of this area of process control and optimization, you should read this section.

OBJECTIVES

The optimization of batch and continuous chemical reactors has many potential benefits, including increase in productivity and improvement in safety, product quality, and batch-to-batch uniformity. Such results are the consequences of many individual control loops and strategies. These loops will maintain material balances, hold optimum temperature and pressure, and control concentrations, while providing safe operating conditions. All elements of the overall chemical reactor control system are discussed in this chapter.

The reactor is the heart of most chemical production facilities. Other operations such as heat transfer, separation, recycle, purification, and drying all are dependent on the reactor's production and demands for services. Most reactors are "base-loaded," i.e., their feed rate is fixed, as determined by market demand and constraints. All other operations are then dependent on the reactor's production rate.

Therefore, it is essential that the reactor operation be stable. If the reactor is run efficiently, with steady product of uniform quality, all other operations may proceed smoothly and the plant should be profitable. But if temperature control is erratic, for example, product quality will suffer and all downstream plant operations will be upset.

Safety

Several objectives must be met for successful reactor operation. The first is always *safety.* Most chemical reactions in production facilities are *exothermic*, which means that they produce heat, much like the combustion of fuel. That heat must be removed as it is produced, to maintain a constant temperature. If it is not, the temperature will rise, which increases the rate of reaction, creating the possibility of a thermal runaway—in the extreme case resulting in a fire or explosion. For the protection of life, equipment, and the environment, such an outcome must be prevented. But less extreme conditions can damage product, deactivate catalyst, and foul heat-transfer surfaces—these also must be avoided.

Maximizing Yield

Operating a reactor safely is not enough. It must also operate profitably. This means maximizing yield and minimizing operating costs. *Yield* is the fractional or percentage salable product that can be made from a given feedstock.

Ideally, 100% of the ethylene fed to an oxidation reactor should be converted to ethylene oxide, but some is invariably converted to carbon dioxide and other by-products of the reaction. If the reaction that produced some of the by-products is reversible, then they may be recycled to reduce their further production and do not represent a loss in yield. But some, like carbon dioxide, will have to be purged from the system and therefore lost.

To maximize yield, it may be necessary to limit *conversion*, which represents the fraction of the feed that is converted to product on one pass through the reactor. The unconverted feed must be separated from the product and recycled. Operating costs increase with recycling, as it increases the loading on pumps, compressors, and separation equipment. So, there must be a balance struck between yield and conversion, reflecting the cost of lost yield and the operation of the equipment carrying the recycle, along with their constraints.

The Catalyst

Most reactions are catalyzed, as well. A *catalyst* is a substance that does not change during the reaction, but increases the reaction rate, allowing a higher yield, or higher conversion, or production to take place under less severe conditions of pressure and temperature. Reactions take place on the catalyst surface, so that a large, active surface is essential. Most catalysts are porous, and pore size can affect their selectivity for particular reactions. Many are inert ceramics coated with active metals like platinum.

The catalyst may be fixed, such as shaped material packed or spaced in a bed. Some catalysts are very fine and fluidized by the upward flow of the reaction mass, as in a fluid cat-cracker, which breaks heavy oil into smaller gasoline molecules. Some catalysts travel with the reaction mass as a slurry, or may even be dissolved. Polymerization reactions usually require an initiator, which is used to create an active site on a molecule, attracting other molecules into a chain reaction.

Catalysts are costly, many having active surfaces of noble metals. While they are not changed by the reaction, they can be degraded, and even poisoned. Degradation consists of clogging the pores or coating the active surface with coke or tar, for example. The fluid cat-cracker contains an auxiliary unit for burning the coke continuously off the catalyst, thereby regenerating its surface and recirculating it to the reactor.

A fixed catalyst bed must be regenerated off-line, which takes time and interferes with production. Some catalysts are poisoned by heavy metals and sulfur, which may require their eventual replacement. Close control of operating conditions can minimize the cost of catalyst loss and regeneration.

CHEMICAL REACTION KINETICS

The conduct of chemical reactions follows an exact science called *chemical kinetics*, which describes the relationships that produce reactions. The first relationship is one of *equilibrium*. Most chemical reactions are reversible and terminate in a condition of equilibrium, where forward and reverse rates are equal. As an example, consider the partial oxidation of ethylene to ethylene oxide:

$$C_2H_4 + \tfrac{1}{2}O_2 \Leftrightarrow C_2H_4O + \text{heat}$$

This equilibrium is defined by an equilibrium constant K, defined as

$$K = \frac{[C_2H_4O]}{[C_2H_4][O_2]^{1/2}} \qquad \textbf{8.10(1)}$$

where the bracketed terms are molar concentrations of the components. To shift the equilibrium to the right, i.e., to make more product, the concentration of either ethylene or oxygen may be increased, or the concentration of ethylene oxide reduced. The reactant concentrations can be increased by increasing the operating pressure, as they are gases, and the product concentration can be reduced by condensing it to a liquid.

This forward reaction is of 1.5 order, in that it depends on the first power of ethylene concentration and 0.5 power of oxygen concentration; the reverse reaction is first-order. Many reactions are second-order, but most ultimately can be treated as first-order. The reason for this is that favorable conditions for most reactions have one reactant in dominant concentration and the other in a much smaller concentration, the smaller thereby controlling the reaction. At equilibrium, forward and reverse reaction rates are equal, the equilibrium constant being their ratio.

Batch Reactions

In a first-order reaction, the rate of conversion of reactant of concentration x varies linearly with that concentration and rate coefficient k, expressed in units of inverse time:

$$-\frac{dx}{dt} = kx \qquad \textbf{8.10(2)}$$

In a batch reactor, that concentration will change with time t, as the rate equation is integrated between initial and final conditions:

$$\int_{x0}^{x} \frac{dx}{x} = \int_{0}^{t} -k\,dt$$

where x_0 is the initial concentration and x is its value at time t.

The solution of the above is

$$x = x_0 e^{-kt}$$

where e is 2.718, the base of natural logarithms. The fractional conversion of feed to product in a batch reactor at time t is

$$y = \frac{x_0 - x}{x_0} = 1 - e^{-kt} \qquad \textbf{8.10(3)}$$

Continuous Plug-Flow Reactors

Continuous reactors operate with a uniform flow of feed in and product out. The average time spent by a molecule in a continuous reactor is called its *residence time* and is simply reactor volume V divided by total volumetric flow F, i.e., V/F. In an ideal plug-flow reactor, ingredients flow from inlet to outlet without any longitudinal mixing. In this way, the ideal plug-flow reactor resembles a batch reactor, in that the fractional distance a molecule travels from inlet to outlet is also its fractional residence time in the reactor. As a result, fractional conversion at the exit of a plug-flow reactor is

$$y = 1 - e^{-kV/F} \qquad \textbf{8.10(4)}$$

At $kV/F = 1$, fractional conversion is 0.632. Actual plug-flow reactors may have some longitudinal mixing, but the intent of the design is to avoid it, as *back-mixing* reduces conversion.

Continuous Back-Mixed Reactors

If a vessel is thoroughly mixed, reactant concentration is the same everywhere, including in the product stream withdrawn. The rate of reaction is therefore determined by the outlet concentration, which is lower than that at any point within the plug-flow reactor. Production rate in a perfectly back-mixed reactor is its volume times the rate of loss of reactant concentration, which is also its feed rate multiplied by the loss in reactant concentration from inlet to outlet.

$$-\frac{V\,dx}{dt} = kVx = F(x_0 - x)$$

Solving for exit concentration gives

$$x = \frac{x_0}{1 + kV/F}$$

And fractional conversion is

$$y = 1 - \frac{1}{1 + kV/F} \qquad \textbf{8.10(5)}$$

For the same residence time and rate coefficient, conversion will always be less in the back-mixed than the plug-flow reactor, as shown in Figure 8.10a. At *kV/F* = 1, fractional conversion is 0.5. However, a back-mixed reactor is lag-dominant and is, therefore, easier to control than the plug-flow reactor, which is dominated by dead time.

Effect of Temperature

Temperature has a profound effect on reaction rate, through an exponential relationship to the rate coefficient:

$$k = ae^{-E/RT_r} \qquad \textbf{8.10(6)}$$

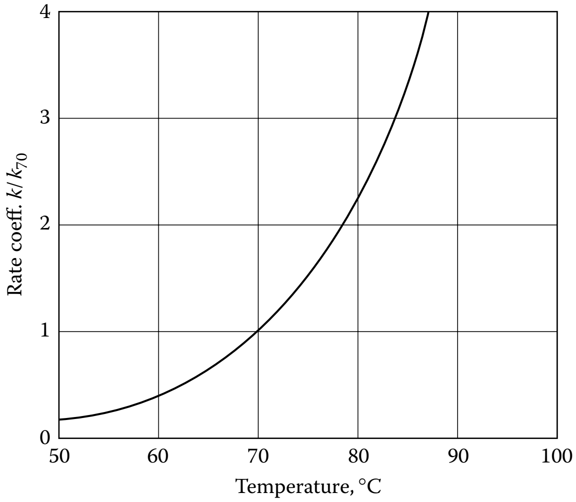


FIG. 8.10b
Reaction increases exponentially with temperature.

where *a* and *E* are constants particular to the reaction, *R* is the universal gas constant, and T_r is the absolute temperature of the reaction mass. The nature of the relationship is that reaction rates tend to double with each increase in temperature of about 10°C (18°F). The rate coefficient for a typical reaction normalized to what it would be at 70°C is plotted vs. temperature in Figure 8.10b. This effect is so dominant that precise control of temperature is the most important aspect of operating reactors of any type.

For all reactors, then, conversion is a function of both temperature and time. Because conversion in plug-flow and back-mixed reactors differs as a function of residence time, they can also be expected to differ as a function of temperature.

This is shown in Figure 8.10c, where both reactors are operated at *kV/F* = 1 at 70°C. Other values of *V/F* produce

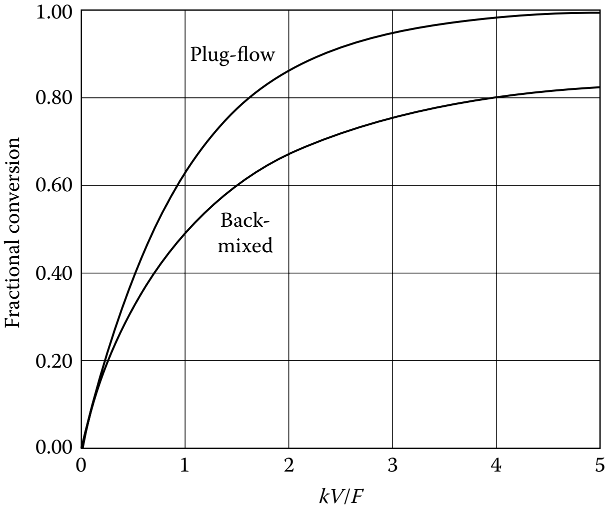


FIG. 8.10a
Plug-flow reactors are used for their higher conversion.

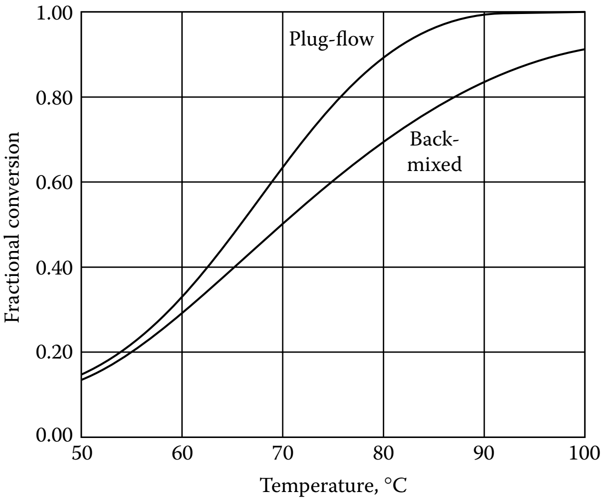


FIG. 8.10c
Conversion in a plug-flow reactor is more temperature-sensitive.

parallel curves. Because Figure 8.10b shows a doubling of k for about 9°C temperature rise, parallel curves for doubling or halving of V/F will be spaced 9°C apart. For example, the plug-flow reactor will produce 63.2% conversion at 61°C and half the feed rate, and also at 79°C with twice the feed rate.

EXOTHERMIC REACTOR STABILITY

Endothermic reactions are inherently self-regulating, because heat must be continuously applied for them to proceed—remove the heat and they die. Ordinary heat transfer is also self-regulating. Increasing the temperature of a hot fluid entering a heat exchanger will increase the rate of heat transferred to the coolant without any change in its flow. Exothermic reactors, by contrast, can be steady-state unstable—depending on their heat-transfer capability—and even uncontrollable. Moreover, some have been observed to change between stable and uncontrollable states.

Steady-State Stability

Consider an exothermic reactor operating at steady-state—constant temperature—in the open loop. A small upset causes its temperature to rise, which increases the rate of reaction, thereby releasing more heat. The increase in heat release then raises the temperature further—positive feedback. At the same time, however, the rising temperature increases the rate of heat transfer to the coolant—negative feedback. Which has the stronger influence will determine the steady-state stability of the reactor. Figure 8.10d compares the heat-release and heat-transfer capabilities of a stable exothermic reactor.

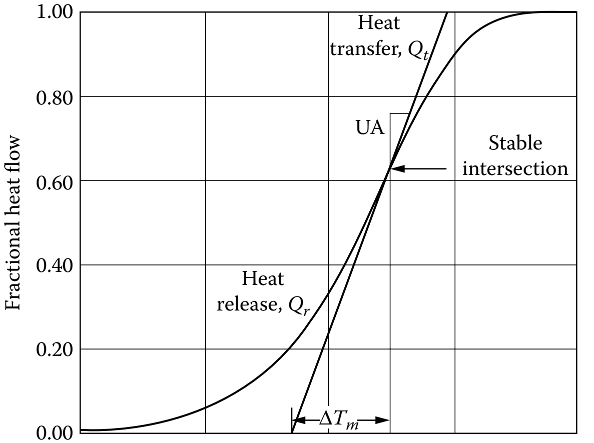


FIG. 8.10d
Steady-state stability is provided by sufficient heat-transfer surface.

In Figure 8.10d, the conversion vs. temperature plot of the plug-flow reactor is equated to heat release by multiplying by the reactant feed rate F and the heat of reaction H_r:

$$Q_r = yx_0FH_r \quad \textbf{8.10(7)}$$

Here, it is normalized to 100% conversion at nominal feed conditions. On the same plot is a line representing heat transfer from the reactor to the coolant:

$$Q_t = UA\Delta T_m \quad \textbf{8.10(8)}$$

where U and A are the heat-transfer coefficient and area respectively, and ΔT_m is the mean temperature difference between the reactor and the coolant. (If the reactor were a well-mixed vessel and the coolant rapidly circulated past its heat-transfer surface, ΔT_m could be a simple difference between reactor and coolant temperatures. In practice, this is not likely to be the case, but a mean temperature difference formula can be fitted to whatever heat-transfer configuration exists.) Heat flow in and out of the reactor is set equal at 70°C, the control point.

If the slope of the heat-transfer line exceeds that of the heat-release curve at the operating point, as is the case in Figure 8.10d, the reactor is stable. A rise in temperature transfers more heat than is released, resulting in a stabilizing of temperature. The key to achieving this stability lies therefore in designing the heat-transfer system to have sufficient area for the expected heat load. Area cannot be traded for colder coolant—they are not equivalent as a guarantor of reactor stability. Note that ΔT_m for the stable reactor in this example is only 16°C.

The slope of the heat-release curve can be determined by differentiating the conversion curve. For the plug-flow reactor, the slope is

$$\frac{dy}{dT} = -\frac{E}{RT^2}(1-y)\ln(1-y) \quad \textbf{8.10(9)}$$

And for the back-mixed reactor, it is

$$\frac{dy}{dT} = \frac{E}{RT^2}(1-y)y \quad \textbf{8.10(10)}$$

The maximum slopes of both curves occur where $kV/F = 1$. At that point, the slope of the plug-flow curve is exactly twice that of the back-mixed curve. The plug-flow reactor is therefore twice as demanding of the heat-transfer surface to remain stable.

For pilot-scale reactors, jacket cooling may be sufficient for stability, especially with stirred tanks. However, there is a scale-up problem: Production increases with volume, which increases with the cube of diameter, whereas jacket area increases only with its square. At some scale, jacket cooling is insufficient, and additional heat-transfer surface is required. This can be provided by internal coils, by a reflux condenser, or by an external exchanger through which reactor contents are circulated. Most full-scale plug-flow reactors have external exchangers.

Unstable but Controllable

The heat-transfer surface in Figure 8.10e has been reduced by half from that in Figure 8.10d—increasing ΔT_m to 32°C—and now the reactor is steady-state unstable. Any rise in temperature above the control point will release more heat than is transferred, augmenting the rise until the upper stable intersection of the line and the curve is reached. Conversely, any drop in temperature will cause heat release to fall more than heat transfer until the lower stable intersection is reached. In the open loop, the temperature will seek one of the two stable intersections, running away from the desired control point.

A simple model of the unstable reactor is the inverted pendulum—a stick balanced vertically in the hand, for example. Its balanced position is an unstable steady state, in that the slightest disturbance will cause it to accelerate away from that state. However, it is possible for a person, or even a trained seal, to balance a stick for a time. To be successful in doing so reveals the control effort required. As soon as the stick begins to deviate from a true vertical position, the hand must move *farther* in the same direction to restore balance. In other words, the gain of the (human) controller must exceed unity to succeed in controlling a steady-state-unstable process.

That is, the process is controllable if the gain of the controller can be set high enough. Reference 1 identifies the limit of controllability of the steady-state-unstable process as when the dead time in the loop approaches its time constant. In other words, a lag-dominant unstable reactor is controllable, whereas a dead time-dominant one is not.

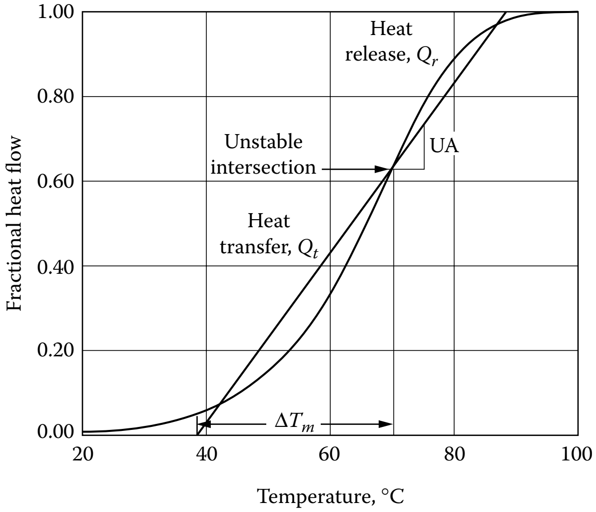


FIG. 8.10e
Insufficient heat-transfer surface makes a reactor unstable.

A stirred tank is more likely to be stable than a plug-flow reactor, because the maximum slope of its heat-release curve is only half as great. But even if unstable, it is controllable because it is dominated by the large lag of its well-mixed contents. To be sure, its control loops need to be properly structured, a subject to be covered. But the proportional gain of a typical temperature controller on a back-mixed reactor is in the order of 3:5, which is entirely capable of controlling it even if it is unstable.

Uncontrollable Reactors

Plug-flow reactors are dominated by dead time and, if unstable, are therefore also uncontrollable. During start-up, as temperature is raised, they may behave well because reaction rate is low at the lower temperatures. But as the desired operating temperature is approached, there is a tendency for the reactor to speed right past it. Even a carefully structured control system will not be able to stop it in time. As the temperature rises further, the slope of the heat-release curve moderates, and the manipulated cooling begins to bring the temperature down. But it then goes past the control point in a downward direction, until the control system arrests the fall.

The result is a limit-cycling of the temperature—at a given set of conditions, the cycle will be of uniform amplitude and period and nonsinusoidal, possibly sawtooth in shape. Its amplitude will depend on the difference between the slopes of the heat-release curve and the heat-transfer line at the control point.

One of the characteristics of a limit cycle is its resistance to change. Variations in the tuning of the temperature controller(s) tend to have little effect—its period and amplitude may change slightly, but the cycle persists. The only way to stop the temperature from cycling is to lower the production rate, which reduces the slope of the heat-release curve, or to improve the heat-transfer condition. A reactor that was once controllable and is no longer, may be suffering from a fouled heat-transfer surface—cleaning has been observed to restore stability.[2]

TEMPERATURE CONTROL

Unless reactor temperature can be controlled precisely, product quality will vary. This is especially true of polymerization reactions, where molecular weight, density, and viscosity are all functions of reactor temperature. Variations as small as 1°C may produce unsatisfactory product.

Measuring Temperature

Resistance-bulb sensors are universally used to measure reactor temperature, as they are more accurate than thermocouples. In a well-stirred tank, the measurement may be made anywhere within the mixed zone, velocity being necessary for dynamic response and to avoid fouling. The sensor must be protected against corrosion and erosion, requiring installation in a well in most cases—then close thermal contact of the sensor with the well is necessary to maximize dynamic response.

The temperature will vary along the length of a plug-flow reactor, due to the variable rate of reaction within. In an adiabatic reactor (no heat removal), temperature will rise throughout. For one that is cooled by a jacket, a peak temperature will usually be reached along the first third of travel from the inlet.

In a packed bed, that hot spot may move downstream gradually as catalyst decays. There may also be a radial gradient. Therefore, most plug-flow reactors have multiple sensors, one of which may be manually selected as the control point. Selection may even be automatic, as the highest of several measurements. An average of two or more temperatures may also be used for control, or the median of three or five measurements, as being more representative of the reaction mass. However, if the combined sensors are spaced too far apart, a difference in their dynamic responses may require the controller settings to be compromised.

Reactor pressure, or rather its change, has been used to predict reactor temperature change. The rationale is that pressure changes directly with temperature, but faster, because it lacks the lags in the bulb and thermowell. However, the steady-state value of pressure is misleading, because it is too easily influenced by variations in composition, particularly the presence of noncondensible gas.

Therefore, the predicted temperature is best calculated as the measured temperature plus a gain factor times the rate of change of pressure. This method was tried by the author on an uncontrollable reactor, but unsuccessfully. The predicted temperature led the measured temperature when falling, but not when rising, and so was ineffective in stabilizing the loop. It was only a half-predictor.

Manipulating Coolant Flow

It is possible to control the temperature of a continuous reactor by manipulating feed rate at a constant coolant flow—but it is also very dangerous! If the reaction mass is cold, reaction rate and heat release will be low, and adding more feed will not raise the temperature. Under these conditions, it is possible to overfeed a reactor, so that when reaction temperature is reached, there will not be enough cooling available to remove the heat released, resulting in a thermal runaway. But even beginning at a steady state, manipulating feed rate is poor practice.

A change in feed rate must first change reactant concentration through the concentration lag, and then change temperature through the thermal lag. These are two large lags in series. The concentration time constant[3] is

$$\tau_x = \frac{1}{k + F/V} \qquad \textbf{8.10(11)}$$

and the thermal time constant is

$$\tau_T = \frac{MC}{UA} \qquad \textbf{8.10(12)}$$

where M and C are the reaction mass and specific heat, respectively. There is also the problem of inverse response: Increasing the flow of cold feed will cause the temperature of the reaction mass to fall before the effect of the increased concentration of the reactant causes it to rise. This reversal could last for several minutes, resulting in very unfavorable dynamic response. This scheme is not recommended to control an exothermic reactor.

Manipulating the flow of coolant is necessary for successful temperature control. If it can be injected directly into the reaction mass, then the lags associated with heat transfer are avoided. If not, then a boiling coolant is the best heat sink, as it is isothermal, but such a fluid may not be available or safe or economical in the temperature range desired. Boiling part of the reaction mass and condensing it in a reflux condenser is another self-regulating means of heat removal. However, the most common coolant used for exothermic reactors is water flowing past a heat-transfer surface.

Tempering Loops

Manipulating coolant flow in a single pass by the heat-transfer surface is not good practice. Consider, for example, the jacketed batch reactor shown in the previous section in Figure 8.9f, whose cooling load varies with time. When little coolant is required, its low flow through the jacket creates a long dead time—jacket volume divided by the low flow. At higher demands, this dead time decreases. The temperature controller cannot be optimally tuned for all of the possible flowing

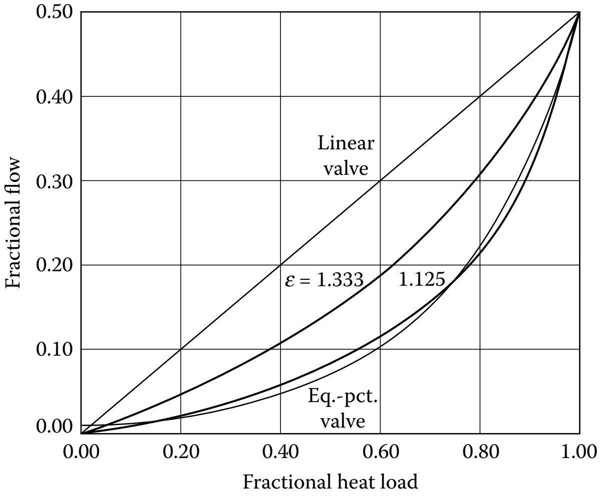


FIG. 8.10f
The nonlinear relationship of coolant flow to heat load is compensated by an equal-percentage valve.

conditions, and must favor the worst case of low flow, giving less-effective control during the major portion of the reaction.

This dead time varying with flow also creates a limit-cycling condition in the temperature loop. A rising temperature causes the controller to increase coolant flow, which decreases jacket dead time and thereby drops temperature quickly. But a falling temperature causes the controller to decrease coolant flow, which increases dead time and delays temperature response.

A temperature cycle is created that then has sharp peaks and flat valleys, and tuning may not be able to remove it. As the cooling load increases, higher flow rates tend to decrease the period and amplitude of the cycle, and it may even disappear at the highest rates.

This behavior can be avoided entirely by the use of a circulating pump on the coolant, as shown in the previous section in Figure 8.9g. Adding coolant lowers the temperature of the jacket contents without affecting internal velocities. The coolant enters the jacket at constant flow and variable temperature, instead of constant temperature and variable flow.

Valve Selection

The rate of heat transfer is not linear with coolant flow. Consider, for example, a batch reactor controlled at 70°C by cooling water supplied at 20°C. If there is no cooling load, then the coolant in the circulating loop will be at the same temperature as the reactor, with no fresh coolant added. Opening the coolant valve will add water at the supply temperature of 20°C while displacing water at 70°C—each increment of water carries away heat proportional to that 50°C of temperature rise. However, at full heat load, the coolant exit temperature will be much lower, possibly as low as 30°C. Now the same change in water added removes heat proportional to a temperature rise of only 10°C.

The gain in converting coolant flow to heat transfer varies by 5:1 between zero and full load.

When a logarithmic-mean temperature difference is used to relate heat transfer to coolant temperatures in a jacketed stirred tank, the ratio of fresh coolant mass flow W_C to circulating flow W is derived as a hyperbolic function of heat load:

$$\frac{W_C}{W} = \frac{1}{WC_C(T_r - T_C)/Q_t - 1/(\varepsilon - 1)} \qquad \mathbf{8.10(13)}$$

where C_C is the specific heat of the coolant, and

$$\varepsilon = e^{UA/WC_C} = \frac{T_r - T_{C1}}{T_r - T_{C2}}.$$

Temperatures T_{C1} and T_{C2} are those of the circulating coolant entering and leaving the jacket, respectively, at any given steady state; if U, A, and W are constant, ε is also constant.

This relationship is plotted for the reactor described above as the numbered curves in Figure 8.10f, for a range of coolant flow up to half the circulated flow, and two values of ε. $\varepsilon = 1.125$ corresponds to jacket inlet and outlet temperatures of 25 and 30°C, and 1.333 corresponds to jacket inlet and outlet temperatures of 30 and 40°C at full heat load. The latter case represents a value of UA/WC_C that is 23% higher. Linear and equal-percentage valve characteristics are plotted for comparison, with the rangeability of the equal-percentage valve being 50:1. The latter matches the required characteristic for $\varepsilon = 1.125$ almost perfectly. Matching the process curve by the valve means that heat flow will be linear with valve stroke, and therefore so should temperature.

However, the curve for $\varepsilon = 1.333$ shows that such a close match of valve characteristic to the heat-transfer process is not always attained. In fact, Reference 2 describes a gas-phase plug-flow reactor cooled by recirculation of gas through a counterflow cooler, where the process curve is closer to that of $\varepsilon = 1.333$. The equal-percentage valve is still the preferred choice, because its variable gain compensates the loop for the effect of changes in coolant supply temperature and reactor temperature set point.

Equation 8.10(13) shows that the relationship between coolant flow W_C and heat transfer Q_t is affected by the temperature difference $(T_r - T_C)$ as a gain factor. A lower temperature difference decreases the process gain dQ_t/dW_C and requires a higher coolant flow for the same level of heat transfer. But the gain of an equal-percentage valve increases directly with flow delivered, thereby compensating for the process-gain change.

Cascade Control

Cascade control is recommended for critical variables that are difficult to control and subject to several disturbances, and the temperature of an exothermic reactor certainly fits this case. Cascade control is possible whenever the process can be divided into separate dynamic pieces by the measurement of an appropriate secondary variable. In the case of an exothermic reactor, the two pieces are the reactor itself and its cooling system, and the secondary variable is the temperature of the coolant as it enters or exits the heat-transfer surface.

As shown in Figure 8.9l in the previous section, the secondary or slave loop is closed by a temperature controller whose set point is adjusted by the primary or master temperature controller. The primary controller sees the secondary closed loop and the reactor in series. It is imperative that the secondary loop be faster than the primary, and yet it should contain significant dynamics and disturbance variables to be effective.

The periods of the two loops should differ by a factor of 4 or more to avoid resonance. For the stirred tank, controlling the jacket outlet temperature closes the loop around the jacket, thus speeding the response of the primary loop and preventing cooling-system disturbances from upsetting the reactor. But plug-flow reactors are faster, and controlling coolant inlet temperature as the secondary variable gives the two loops better dynamic separation.[2]

Batch exothermic reactors require heating as well as cooling, both for startup purposes and to cure the product and drive off solvent at the end of the reaction. The secondary controller therefore opens cooling and heating valves in split range. The split-ranging is usually accomplished with valve positioners as shown in the previous section in Figure 8.9m, calibrated to open the heating valve from 50 to 100% of the controller output, and the cooling valve from 50 to 0% (the cooling valve fails open).

Because of the difficulty of controlling exothermic reactors and of the critical role played by valves, smart digital positioners are recommended.[2] They eliminate stiction and dead band, and maximize response speed. Each valve positioner closes a loop around its valve and therefore constitutes a second level of cascade control, as was shown in Figure 8.9m.

With the provision of heating, additional precautions are advised. A high limit should be set on coolant temperature as shown, and for glass-lined reactors, the temperature difference between the reactor and jacket should be limited to avoid excessive thermal stress. Some polymerization reactors are also fitted with interlocks that prevent heating after the initiator has been introduced into the reactor.

In some polyvinyl-chloride (PVC) processes, the reaction mass is heated to about 70°C without any initiator, and allowed to stabilize at that temperature without any reaction taking place. Then initiator is injected, and the heating valve locked out for safety. A rising temperature then signals the beginning of the reaction. The temperature controllers respond by opening the cooling valve, and after a peak deviation of 2–3°C is reached, the temperature returns toward set point, and in fact, overshoots.

At this point, the secondary controller output rises above 50%, but the heating valve is locked out, so that the secondary loop is open. The temperature slowly begins to rise again, but during the open-loop condition, the primary controller has been winding up—it has raised the set point of the secondary controller, which was unable to respond because of the interlock. As a result, the secondary set point is too high, causing the primary temperature to rise past its set point again. The scenario then repeats, settling into a limit cycle of ±1°C or thereabouts, which does not respond to tuning changes and adversely affects product quality.

The limit cycling is avoided through the use of *external reset* (ER) from the secondary temperature measurement to the integral-feedback port of the primary controller, as was shown in Figure 8.9m. As long as the secondary temperature is at its set point, the ER signal and the primary controller output are identical, allowing the primary controller to integrate normally. But when the secondary is prevented from controlling (by the locked-out valve, or any other obstacle such as a set-point limit), primary integration stops. The secondary controller must have integral action for this strategy to be successful, because any secondary-loop offset will cause a proportional primary-loop offset.

This feature successfully eliminated limit-cycling in the PVC reactors with the locked-out heating valve. It was also used on batch reactors for latex paint and found to be very robust. In trials, the primary time constant of the reactor was varied over a range of 4:1 by changing batch size and heat-transfer area, each by a factor of 2:1—yet the control system did not require any tuning adjustments. The rate of temperature change was obviously affected by these process conditions, but the dynamic response of the secondary temperature altered the rate of integration of the primary controller sufficiently to accommodate them.

The secondary controller in a cascade system must be tuned before the primary, as it is a part of the primary loop. It should be tuned for optimum *load* response—not set-point response—regardless of having to respond to its set point. The set-point changes that it sees are gradual, not stepwise, and therefore affect the loop dynamically much the same as load changes do. For the same reason, filtering should never be applied to the secondary set point, and the set point should receive full proportional action (some controllers have the option of eliminating proportional action to the set point). Otherwise, the secondary control action will effectively lag the primary output, negating the dynamic benefit of cascade control.

Batch Reactor Start-Up

The problem of integral windup is particularly troublesome during start-up of a batch reactor. The goal is to raise reactor temperature from its initial value (possibly ambient) to an elevated set point (where the reaction proceeds) as quickly as possible, but without overshoot. In this situation, a conventional PID controller will invariably overshoot set point.

Following a large set-point change, its output will saturate, and during the rise time, the integral mode will also saturate. The result is that the output remains at its high limit until the set point is crossed. If the new steady state requires no heating or cooling in absence of a reaction, the new steady-state value of the controller output should be 50%; if an exothermic reaction has begun, the new output should be even lower. As a result, there will be an unacceptable temperature overshoot, such as that shown in the top curve in Figure 8.10g.

Integral windup in the start-up of a batch process can be overcome by the circuit, which was shown in the previous section in Figure 8.9x. The output of the controller is compared with a high-limit setting, and their difference sent to a high-gain amplifier, whose function is to control the controller output at that limit by manipulating the integral feedback term.

If the controller output is below the high limit (say 100%), that output is selected by the low-signal selector (<)

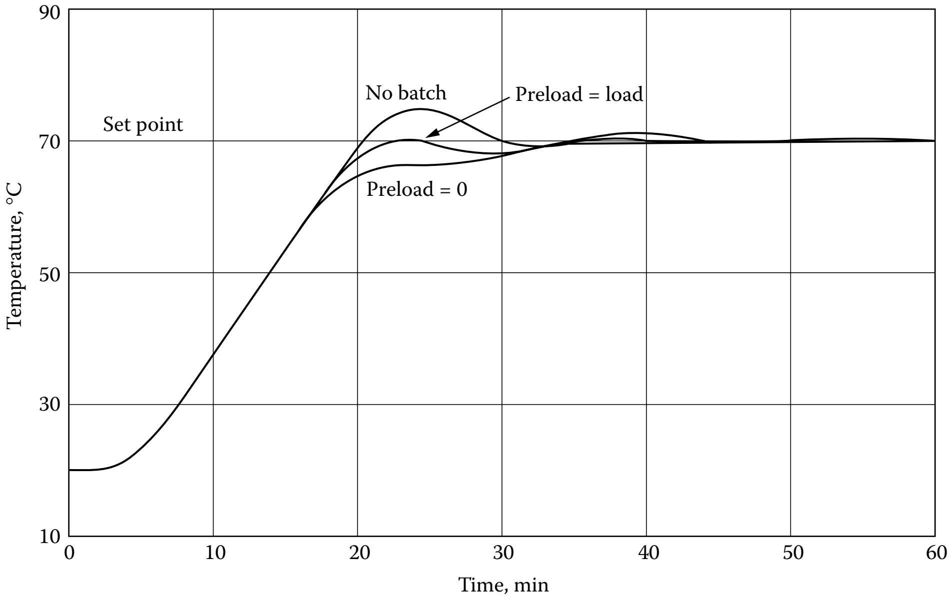


FIG. 8.10g
A preloaded batch PID controller can avoid overshoot.

as integral feedback, allowing the controller to integrate normally. If the output moves above the limit setting, the amplifier will manipulate the integral feedback term as needed to hold it at the limit. However, the feedback term can only be reduced to the point where the preload setting is selected by the high-signal selector (>); preload is adjustable to match the expected load when the new steady state is reached.

Overshoot is caused by the integral term staying at the high output limit until set point is reached. Reducing that term to match the expected load—the controller output required to hold temperature at set point in the new steady state—should eliminate the overshoot, and it does.

Without preloading the controller, however, *undershoot* will result, as shown in the lowest curve in Figure 8.9g. These curves were produced by an interacting PID controller having derivative action on the measurement (derivative was not shown in Figure 8.9x), which also assists in limiting overshoot. If derivative is not used, then the preload setting will have to be reduced somewhat below the expected load, because integration begins as soon as the controller output leaves its limit, which is well before the set point is crossed.

Another method of avoiding overshoot while minimizing start-up time has been used successfully on polystyrene reactors—the *dual-mode* control system. It consists of an on/off controller for start-up and a PID controller to take over at the new set point, as was schematically depicted in the previous section in Figure 8.9z.

The on/off controller applies full heat until the temperature has almost reached set point—it then switches to full cooling. After a set time delay, control is transferred to a preloaded PID. Switching from full heat to full cooling takes place within 2–3°C of set point, and the time delay is long enough to replace the hot water in the jacket with cold.

The sequence of operation is as follows:

1. Initial conditions have both switches in the A position. Full heat (100% output) is applied by the on/off controller until the deviation falls to the set point of Em.
2. The on/off controller output then drops to zero, producing full cooling. It also starts two time-delay relays: TD-1 and TD-2.
3. When TD-1 times out, the period of full cooling terminates, and the system output is switched by SS-1 to the preloaded PID controller, without integration.
4. When TD-2 times out, the temperature should be settled at set point. It then transfers switch SS-2 to the B position, enabling integration.

While this system is capable of minimizing start-up time, it is not foolproof. It has three settings more than the batch controller: Em, TD-1, and TD-2. Furthermore, these settings need to be changed with process conditions such as batch size and reactivity, temperature set point, and heat-transfer coefficient, the last typically decreasing with the number of batches produced, as surfaces foul.

Efforts to adapt these parameters to changing conditions, or to program them with each recipe, have met with limited success. The preloaded batch PID controller with external reset feedback from secondary temperature is much simpler, more universally applicable, and more robust under changing conditions.

MATERIAL BALANCE CONTROL

Most chemical reactions require the precise matching of two reactants to produce a product. One of the challenges of process design is to ensure that all of each reactant is converted to product, regardless of the fractional conversion in the reactor.

A simple example is the neutralization of acid and base—their individual concentrations in solution must be typically matched within ±0.01% for the solution to fall within the pH range of 6–8. A pH sensor is sensitive enough to detect such a minute imbalance, but on-line analyzers are not always available and even when they are, end-point control can be difficult, both in a batch and in a continuous plant.

Even when there is but a single reactant, as a monomer being polymerized, end-point control is necessary to terminate the reaction at the desired chain length or molecular weight or density.

Stoichiometry

Stoichiometry is the term used to describe the balancing of a chemical reaction so that all of the reactants are converted to the desired product. A typical example is the oxidation of hydrogen sulfide to elemental sulfur, which is condensed:

$$H_2S + \tfrac{1}{2}O_2 \rightarrow H_2O + S\downarrow$$

A simple enough reaction, but if too much oxygen is fed, it will oxidize sulfur to sulfur dioxide in a consecutive reaction:

$$S + O_2 \rightarrow SO_2$$

Simply metering the gas and air is not accurate enough, and the H_2S content of the gas is also usually variable. The key to controlling this reaction is establishing an equilibrium between H_2S and SO_2 in the tail gas leaving the reactor:

$$H_2S + \tfrac{1}{2}SO_2 \Leftrightarrow 1\tfrac{1}{2}S + H_2O$$

At equilibrium, where the reaction stops, H_2S has twice the concentration as SO_2 in the gas phase, and Reference 4 demonstrates that this ratio results in the minimum sulfur loss.

So an analyzer on the tail gas reports both concentrations, and a ratio controller manipulates the air-to-gas flow ratio to hold the H_2S/SO_2 ratio in the tail gas at a value of 2.0. The system is shown in Figure 8.10h. There is no fixed or set value for either H_2S or SO_2; they float with production rate and other conditions, but are always at a minimum as long as their ratio is maintained at 2:1. (Note that control of the ratio is achieved without the use of a divider—placing a divider in a closed loop is inadvisable, because its gain varies with the value of the denominator.)

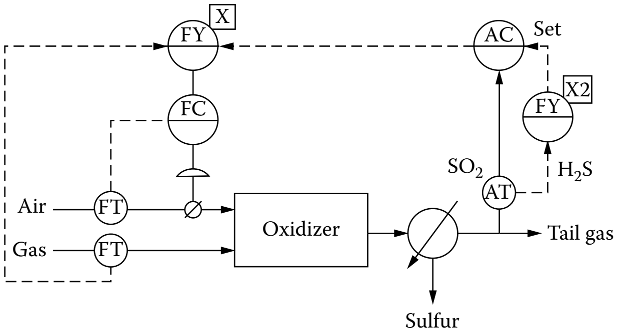


FIG. 8.10h
Minimum sulfur is lost when the ratio of H_2S to SO_2 in the tail gas is controlled precisely at 2:1.

Controlling Inventory

When one of the reactants differs in phase from the product(s) and other reactants, its inventory is measurable independently from the others, and it can be fed at the rate at which it is consumed. Two examples were shown in Figure 8.9ii.

A gaseous reactant forming a gaseous product in a liquid-phase reactor will consume the liquid reactant at the rate of reaction. Loss of liquid inventory is sensed by the level, which can be controlled to keep inventory constant. The manipulated flow of liquid feed then directly indicates reaction rate. Because no liquid product leaves the reactor, however, liquid or solid contaminants could accumulate at low points. There should be a liquid purge, either continuous or periodic—just like boiler blowdown—to eliminate the problem. Most plants have facilities for recovering feedstocks and by-products from purge streams—some are only operated periodically, as waste inventory builds.

Similarly, if the product(s) leave in the liquid phase, consumption of a gaseous reactant lowers pressure, indicating a loss in its inventory. A pressure controller can add gaseous reactant as it is consumed. In this case, a small gas purge is required to eliminate inerts such as nitrogen or carbon dioxide. Molecular sieves are often used to recover valuable gases from purge streams.

When feeds and products are in the same phase, inventory control is more difficult. For example, isobutane reacts with butylenes in a 1:1 ratio to produce high-octane gasoline called alkylate, but they are fed to the reactor in a 10:1 ratio to minimize formation of butylene polymers. The process therefore contains a large inventory of liquid isobutane, which must be controlled as it is consumed.

The reactor effluent, containing isobutane, alkylate, and inerts such as normal-butane, must be separated by distillation, with the isobutane being recycled. The recycle storage tank provides an outflow that is 10 times the rate of consumption and make-up of isobutane, and so its time constant is large compared to that of typical liquid-level loops. Proportional-only level control is recommended to allow the tank to absorb variations in production rate without the oscillations that integral action brings to slow level loops.

A similar refining process is the hydrocracking of heavy oil into gasoline fractions. Hydrogen and oil are heated and fed into a series of packed-bed reactors, where partial conversion takes place. Lighter products in the effluent are then removed by distillation, with the unreacted oil recycled under flow control from the column base to the reactors.

The oil is eventually all consumed, but at a rate that varies with temperatures and so on. Its inventory is sensed as the liquid level in the base of the distillation column, and that level controller is used to add more oil. It has also been used to set reactor temperature to convert a fixed flow of oil, but this is more difficult dynamically, as it places a temperature loop within a level loop.

Reactor Dynamics

Closing the composition loop on a reactor can be difficult, depending on the reactor dynamics. Ideally, a plug-flow reactor is pure dead time In fact, static mixer elements are often used in a plug-flow reactor to maximize radial mixing and minimize longitudinal mixing. Actually, however, the step response will appear more like a series of noninteracting lags, between 20 and 50 in number, as shown in Figure 8.10i. This would be equivalent to 20–50 ideal back-mixed reactors in series. The response does approach dead time, with 50% completion reached when the elapsed time is equal to the residence time, *V/F.* It also should be noted that the effective dead time is variable with flow, although large variations are not expected in production reactors.

Ideally, a back-mixed reactor is a pure first-order lag. But perfect mixing is unattainable, and therefore the actual step response reveals a small amount of dead time, dominated by a first-order lag. The sum of the dead time and time constant is the residence time, *V/F*—at that point, the response curve crosses 63.2% complete, just as does a single lag.

The back-mixed reactor curve in Figure 8.10i was simulated by 20 *interacting* lags, although increasing the number above 20 does not change the shape of the curve. It is essentially a distributed lag that is an accurate representation of a heat exchanger, where heat capacity and heat-transfer surfaces are distributed throughout the exchanger.

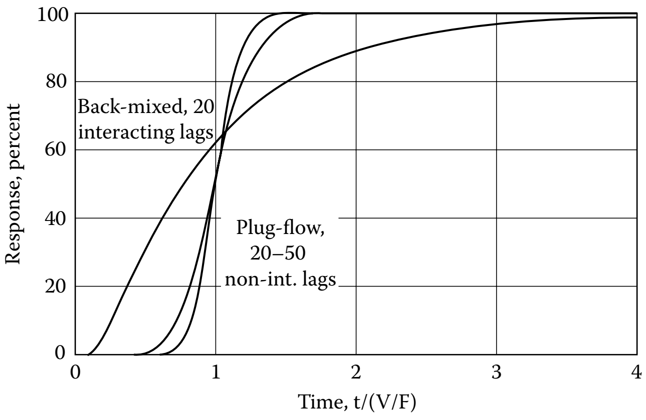


FIG. 8.10i
The plug-flow reactor is simulated by 20–50 noninteracting lags, the back-mixed-reactor by 20 interacting lags.

Reactor composition loops are ordinarily self-regulating, their step response approaching a final steady state as the figure shows. However, many reactors are enclosed in recycle loops that bring positive feedback into the process and eliminate their inherent self-regulation.

Consider, for example, an acid-base reaction carried out in an organic solvent. At the reactor exit, gaseous product is withdrawn, but any excess acid or base remains in the solvent and is ultimately recycled to the reactor. So if z concentration of excess base remains in the solvent, and the acid-base ratio is not adjusted to allow for its presence in the recycle stream, the concentration will increase to $2z$ the next time around, and continue to accumulate.

The integrating time of this recycle loop is the residence time of the solvent in the system—the time required to circulate the entire solvent inventory through the reactor. The significance of this loss of self-regulation cannot be understated. End-point control must be applied, or the composition of the reactor effluent will never reach a steady state. And integral-only control cannot be used to close the loop, or the composition will cycle endlessly—proportional action is essential for stability.

Batch End-Point Control

There are two basic types of batch reactors: pure-batch and fed-batch. A pure-batch reactor has all of its ingredients charged at once, and the reaction is allowed to proceed over time until complete. In a fed-batch reactor, one or more of the ingredients is left out of the charge and then fed at a controlled rate over time until the reaction is complete. In both cases, some determination must be made of the reaction's approach to completion, to avoid wasting both chemicals and time.

Polymerization converts light hydrocarbons into high-molecular-weight products having low vapor pressure. They are often conducted as pure-batch processes, with the initial charge under pressure. At reaction temperature, initially the pressure is close to the vapor pressure of the monomer. As the last of the monomer is consumed, the pressure starts falling, indicating completion of the reaction.

Another indication is the decline in heat evolution as the last of the reactant is consumed. Still another is the total quantity of heat transferred to the coolant, which must be determined by calculation and integration. Others measure or infer density or viscosity of the product to determine completion.

Fed-batch processes require termination of feed at the end point, to avoid loss of reactant. This can be done simply by stopping feed when a desired total is delivered, but this may not be accurate enough in some cases. The use of end-point analyzers is recommended, where available. A case in point is the control of pH in a batch of wastewater from a small processing plant—it must meet environmental specifications before discharge is allowed.

This is not a typical composition loop requiring a PID controller, because a batch process has no flow leaving it—it is essentially a no-load process. Integral action is needed to eliminate offset whenever a controller must move its steady-state output to match the current plant load. But when the load is zero and stays there, integral action is unnecessary and, in fact, causes irrecoverable overshoot.

Consider a tank filled with impounded industrial wastewater at pH 3—its pH must be raised to 6 prior to discharge, and so 6 is the set point of the pH controller. The large initial deviation of 3 pH will drive the caustic valve fully open, and it will remain open as long as necessary to neutralize the waste. During this time interval, if the integral mode is active it will wind up to the controller output limit, and the output will not fall below that limit until the pH crosses set point. That is clearly too late, and by the time the caustic valve closes, the pH will have risen well beyond set point—perhaps to 9 or 10.

In a continuous plant, overshoot is temporary, because its outflow will eventually purge away excess reagent. But not in a batch plant—the overdose is permanent. The batch will require re-treating with acid or more wastewater to lower the pH.

Because the load is zero, the controller output should also be zero when the controlled variable is at set point. This is achievable by elimination of integral action, replacing it with an output bias of zero—the controller must be proportional-only or proportional-plus-derivative.

This is not simply a matter of shutting off integration, as this leaves the constant of integration in place of an output bias, and it floats, changing every time the controller is transferred from manual to automatic. The bias must be a hard, permanent zero, and some controllers—of the incremental or velocity type—do not have this capability. If a pure proportional controller is unavailable, build one out of a calculation block, as follows:

$$m = K_c(r - c) \qquad \textbf{8.10(14)}$$

where m is the controller output, K_c its proportional gain, r the set point, and c the controlled variable. If derivative action is desired—usually helpful in a composition loop—a lead-lag function can be applied to the controlled variable before it enters the controller.

It is possible to feed two reactants simultaneously into a batch reactor with no outflow. The control problem is like trying to fill a bathtub in minimum time. To reach the right temperature at the desired volume, it really needs to be controlled *throughout* the filling process. This requires finding the flow ratio that produces approximately the right temperature at the start of the filling process, and adjusting it as necessary until the final volume is reached.

Initially, changing the ratio produces a large temperature change, but as the tub fills, this gain gradually decreases. In fact, the dynamic gain of the process varies inversely with the volume of the fill. To effectively control the composition of the variable-volume batch, the gain of the composition controller needs to be changed in direct proportion to the liquid level in the reactor.

REACTOR OPTIMIZATION

After satisfactorily controlling reactor temperature, material balance, and end-point, at the necessary production rate, optimization may be attempted. Optimization generally means maximizing economic return on the plant capital and operating costs, but there are many aspects of this, because there are many cost factors.

In an unlimited market, maximizing production maximizes return on both capital and labor. When the market is limited, meeting production goals with a minimum cost of utilities is essential. Maintenance is also costly, and operating a plant in a way that minimizes cleaning and wear has its savings.

Maximizing Production

Production rate from a reactor is determined by the rate coefficient and residence time. To increase flow rates through a reactor of fixed volume reduces residence time and conversion per-pass, and therefore does not result in a proportional increase in production. The production rate of a first-order reaction is

$$P = x_0 F y$$

Increasing production by raising the feed concentration x_0 of the controlling reactant is not often an option, in that it tends to reduce yield, through side reactions and consecutive reactions. In the case of isobutane alkylation, for example, increasing the ratio of butylenes to isobutane above 1:10 increases the formation of polymers having undesirable properties as motor fuel, and consumes more catalyst.

Increasing production by raising feed rate is viable, but inefficient. To determine the effect of feed rate on production it is necessary to differentiate Fy as a function of F. For the plug-flow reactor,

$$\frac{d(Fy)}{dF} = y + (1-y)\ln(1-y) \qquad \textbf{8.10(15)}$$

And for the back-mixed reactor,

$$\frac{d(Fy)}{dF} = y^2 \qquad \textbf{8.10(16)}$$

For the plug-flow reactor operating at $kV/F = 1$, $d(Fy)/dF = 0.264$, and for the back-mixed reactor it is 0.250. Therefore, increasing feed rate by 1% increases production only

0.25%, while increasing the load on the separation and recycle systems by 0.75%—not a favorable result.

To increase production without this penalty requires increasing reaction temperature along with feed rate, following the reaction-rate coefficient curve. Only small changes are required, given the rate coefficient doubling approximately every 10°C. In the region around 1.0 in Figure 8.10b, the normalized rate coefficient increases about 10% for each 1°C.

Maximizing production in a back-mixed exothermic reactor amounts to operating at the cooling constraints. Figure 8.9dd described how this is done when manipulating feed rate(s) to a continuous reactor. The position of the coolant valve is sent to an optimizing controller (OIC), which adjusts the feed rate(s) to drive the valve to about 90% open. In other applications, such a valve-position controller would have only the integral mode, but for an exothermic reactor, all three PID modes are necessary for stable operation.

In addition, if a large difference develops between the positions of the feed and coolant valves—or more accurately between feed and heat-transfer rates—this is an indication that the reaction is not proceeding normally and should be shut down before excessive amounts of reactant(s) have been fed. The OIC loop must be slower than the temperature loop that is inside it, and great care must be taken in closing it. The OIC loop arose out of operators' practice of observing the steady-state position of the coolant valve and gradually increasing the feed to a fed-batch reactor to drive the valve almost fully open.[2] The same operators respected the conditions under which this could be done safely, namely that the reactor had been properly charged and initiated, and was already at the controlled temperature.

There is no feed rate to manipulate in a pure-batch reactor. Furthermore, the reaction rate tends to vary as reactants are consumed. To minimize run time then requires the OIC to adjust the temperature set point to drive the coolant valve to 90% open. This is a difficult loop, however, exhibiting inverse response. To open the coolant valve further requires increasing temperature set point. However, when its set point is increased, the temperature controller responds by *closing* the coolant valve first. As the temperature then begins to rise, the controller will eventually open the coolant valve more than before, but this takes time. Again, the OIC requires full PID control, and must be much slower than the reactor temperature controller.

Maximizing production based on the position of the coolant valve is not always applicable on plug-flow reactors. If the coolant valve is obviously limiting production, as can happen during peak summer weather, then it may work. However, Reference 2 describes an uncontrollable reactor limit-cycling in winter, when the equal-percentage coolant valve was only 10–20% open.

Production rate was limited by the heat-transfer surface of the cooler and not by the coolant valve or supply temperature. The maximum controllable production rate of a plug-flow reactor is determined by the slope of the heat-transfer line in Figure 8.10d. At the limit of stability, the slopes of the heat-release and heat-transfer curves are the same. The slope of the heat-release curve is

$$\frac{dQ_r}{dT_r} = x_0 F H_r \frac{dy}{dT_r} \qquad \textbf{8.10(17)}$$

and that of the heat-transfer curve is *UA*. Therefore, the maximum controllable feed rate is

$$F_{\max} = \frac{UA}{x_0 H_r (dy/dT_r)} \qquad \textbf{8.10(18)}$$

To monitor *UA* requires the calculation of the logarithmic-mean temperature difference across the cooler:

$$UA = \frac{Q_t}{\Delta T_m} \qquad \textbf{8.10(19)}$$

where for counter-current heat transfer,

$$\Delta T_m = \frac{(T_{H1} - T_{H2}) - (T_{C1} - T_{C2})}{\ln \dfrac{T_{H1} - T_{C2}}{T_{H2} - T_{C1}}} \qquad \textbf{8.10(20)}$$

Q_t may be calculated either from fresh or circulated coolant flow:

$$Q_t = W_C C_C (T_{C2} - T_C) = W C_C (T_{C2} - T_{C1}) \qquad \textbf{8.10(21)}$$

T_{H1} and T_{H2} are temperatures of the reaction mass entering and leaving the heat-transfer surface, respectively.

For a plug-flow reactor prone to fouling, *UA* or ΔT_m should be monitored regularly and compared to production schedules. In this way, cleaning can be scheduled in advance, instead of waiting until a limit-cycling condition develops that forces a maintenance shut-down at an inconvenient time.

Supplemental Cooling

Some reactors have two sources of cooling, at different costs. For example, cooling water can be used up to the point where its valve is fully open, and then a supplemental medium such as chilled water may be used for additional cooling. If the two media can be mixed, then accommodating the additional source is as simple as sequencing the two valves. For example, the cooling-water valve could open from 50 to 25% of controller output, and the chilled-water valve from 25 to 0.

However, the media are likely to be separated from each other to avoid cross-contamination, with auxiliary cooling by chilled water passing through an internal tube bundle, and cooling water circulated through the jacket. This complicates a control system that features cascade control of jacket outlet temperature. The cascade system should be retained for its effectiveness, and the supplemental cooling manipulated to keep the jacket temperature operating in its control range.

This is done by the use of a valve-position controller (VPC) connected to the secondary output, as was shown in Figure 8.9o in the previous section. Using integral-only control, it manipulates the chilled-water flow to keep the cold-water valve about 90% open, which corresponds to a controller-output signal of 5% (its range being 50–0%).

As long as the heat load stays in the range of the cold-water valve, the chilled-water valve would remain closed. But when the VPC set point is passed, the chilled-water valve will be opened as necessary to match the load. This loop is not fast, but it need not be any faster than the expected load changes, because the jacket-temperature loop is always in control.

Whether the additional cost of the chilled water can be justified depends on the demand for product. In most cases where supplemental cooling is used, it is only needed during peak load, which only represents part of a typical batch operation.

References

1. Shinskey, F. G., *Process Control Systems*, 4th edition, New York: McGraw-Hill, 1996, p. 317.
2. Shinskey, F. G., “Exothermic Reactors: The Stable, the Unstable, the Uncontrollable,” *Chemical Engineering*, March 2004.
3. Shinskey, F. G., *Process Control Systems*, p. 312.
4. *Ibid*., pp. 324–325.

Bibliography

Corripio, A. B., *Design and Application of Process Control Systems*, Research Triangle Park, NC: ISA, 1998.

Juba, M. R., and Hamer, J. W., “Progress and Challenges in Batch Process Control,” *Chemical Process Control-CPC III*, Amsterdam: Elsevier, 1986.

Lipták, B. G., *Optimization of Unit Operations*, Radnor, PA: Chilton, 1987.

Luyben, W. L., “Batch Reactor Control,” *InTech*, August 1975.

Luyben, W. L., Tyreus, B.D., and Luyben, M. L., *Plantwide Process Control*, New York: McGraw-Hill, 1998.

Shinskey, F. G., *Controlling Multivariable Processes*, Research Triangle Park, NC: ISA, 1971.

8.11 Chemical Reactors: Simulation and Modeling

A. B. CORRIPIO, P. A. HOLST (1970) **A. B. CORRIPIO** (1985)
B. G. LIPTÁK (1995) **B. G. LAKATOS** (2005)

INTRODUCTION

In order to be able to control a process, the process control engineer must understand it and know how the process will respond to changes in loads or other inputs. Modeling and simulation are important tools that can be used in analysis of physical systems and in the understanding of the functioning of processes. Besides, the physical equipment to be controlled does not yet exist, and consequently, no experimental data exists for them. Even if the process equipment is available for experimentation, the procedure usually is very costly, and, consequently, it is preferred if one can use a model to predict the behavior of the processing systems.

The fundamental models can be derived from the basic principles and from thorough knowledge of the process. In characterizing a processing system, one needs (1) a set of fundamental variables describing the state of the system, and (2) a set of equations for these variables, which describe how the state of the given system changes with time. The dynamic characteristics of a process, the knowledge of which is fundamental to designing the control systems, are described by the dynamic models in their general time-dependent form from which the steady-state models can be obtained as limiting cases forcing the time derivatives to zero.

Model Building

In the case of a chemical reactor, several factors enter into the model-building process. The set of physical characteristics is described by the balance equations of three fundamental quantities: mass, energy, and momentum.

Reaction kinetics is the other main consideration in choosing conditions for carrying out a reaction. Quantities such as density, heat capacity, and other physical–chemical parameters also are part of the overall model. An important element in the overall model can be an economic submodel containing a cost or profit objective function, useful in evaluating the possible process variables and in optimizing the system. Statistical elements can also be incorporated into the model on the basis of which future behavior of the system can be predicted.

Once a model has been built it can be used for examining the process behavior, for operator training, and also for developing different control algorithms and configurations. Early predictions of the likely conditions that the reactor is approaching can also be predicted from the model, thereby giving time for the control system to correct such trends.

Steady-state simulations are often used in equipment design. Discrete dynamic simulations serve to optimize schedules, maximize production, and minimize equipment and inventory requirements. Continuous dynamic simulation is useful for improving control strategy, reducing start-up and heat-up times, and improving on-stream times. Real-time simulation is used both for operator training and for debugging DCS and PLC systems, thereby providing both improved performance and safety.

SIMULATION

All the control programs must be tested under simulated plant conditions before the control system is connected to the process. A step-by-step simulation and checkout must take place in order to ensure the validity of all control strategies designed for a particular application. Prior to the simulation, it is usually necessary to wire to the simulator a number of analog input signals (DC source and individual potentiometers), analog output signals (DC meters), discrete input signals (AC/DC source and individual switches), and discrete output signals (AC/DC lights).

These tools play an important role in control system analysis and design, and are still used today to some extent. However, these days most simulation work is carried out by means of digital computers, which allow engineers to analyze and design more sophisticated control strategies and systems. Digital computers offer a lot of function blocks by means of which practically all linear and nonlinear process and control elements can be modeled. Although simulation with digital computers is more effective, the methodologies of testing and checkout of control systems are similar.

As the first step in the testing, all continuous (regulatory) loops are tested under both open and closed conditions, while the discrete controls are deactivated. For inherently stable processes, usually open-loop testing (without process dynamics) is adequate. Whenever dynamic process conditions must be simulated, a program is required to generate these conditions. The process variables of the models are monitored to determine the responses and stability of algorithms to the set point and simulated process load changes, respectively.

The next step is the simulation and checkout of the discrete functions, namely direct on/off control, interlock control, and sequential control. Using the simulator or the simulation program, one can imitate normal or abnormal conditions and check the system responses. If a very large number of discrete I/O is required in order to check a very complicated sequential control strategy, a simulator program can be generated, and the test can be conducted in the same manner as a dynamic test for continuous (regulatory) control described earlier.

In addition to the previously described simulation, when the system is fully installed in the field, a test run is recommended prior to actual start-up.

Control System Simulation

The simulation of instrumentation and control systems includes the measuring or detecting devices, the controllers, and the actuating or manipulating of output elements. Most measuring devices may be simulated by one or more simple lags for dynamic representation, with added static nonlinearities for device characteristics, sensitivity, or operating point modeling.

For special functions, such as logarithmic input/output relationships (pH measurements), the general-purpose function-generating units may be used, or the functions may be generated by implicit techniques. When the measuring device puts out a discontinuous (pulse-type) signal, the signal-generating circuits may be applied.

A direct, three-mode (PID) controller has the transfer function:

$$M(s) = \frac{100}{\mathrm{PB}}\left(1+\frac{1}{T_I s}\right)\left(1+\frac{T_D s}{1+T_0 s}\right)E(s) \qquad \textbf{8.11(1)}$$

where (all variables are in Laplace representation)

$M(s)$ = controller output
$E(s)$ = actuating error, $E(s) = R(s) - Y(s)$
$Y(s)$ = measured (controlled) input variable
$R(s)$ = set point
PB = proportional band
T_I = integral (reset) time
T_D = derivative time
T_0 = filtering (stabilizing) time constant

The filtering time constant is usually made as small as possible and is often a fraction of the derivative time constant T_D (such as $T_0 = T_D/16$). Often, the proportional band is replaced by the proportional gain expressed as

$$K_c = \frac{100}{\mathrm{PB}} \qquad \textbf{8.11(2)}$$

The controller transfer function in Equation 8.11(1) is of the interacting type, which is the common case in most industrial applications. A noninteracting controller is obtained by means of the transfer function

$$M(s) = \left(\frac{100}{\mathrm{PB}}+\frac{1}{T_I s}+\frac{T_D s}{1+T_0 s}\right)E(s) \qquad \textbf{8.11(3)}$$

in which all three modes can be adjusted independently.

In general, the proportional band should have an overall adjusting effect, which leads to the practical noninteracting controller with the transfer function

$$M(s) = \frac{100}{\mathrm{PB}}\left(1+\frac{1}{T_I s}+\frac{T_D s}{1+T_0 s}\right)E(s) \qquad \textbf{8.11(4)}$$

Modern controllers often limit the derivative action to respond only to measured variable changes and not to rapid set-point changes, which could upset the process. This is stated in the modified transfer function

$$M(s) = \frac{100}{\mathrm{PB}}\left[\left(1+\frac{1}{T_I s}\right)E(s)-\frac{T_D s}{1+T_0 s}Y(s)\right] \qquad \textbf{8.11(5)}$$

which is of the noninteracting type.

Digital Control Simulation

Digital control may be simulated directly with a discrete controller, where the measured variable $y(t)$ and the set point $r(t)$ in the time domain are sampled at fixed intervals ΔT, called sampling period, and $t = nT$. Denoting now the actuating error by $e(nT) = y(nT) - r(nT)$, or $e_n = y_n - z_n$, the numerical approximation of the controller algorithm 8.11(5) is

$$\Delta m_n = \frac{100}{\mathrm{PB}}\left[e_n+\frac{T}{T_I}\sum_{k=1}^{n}e_k-\frac{T_D(y_n-y_{n-1})}{T_0 T}S_n\right] \qquad \textbf{8.11(6)}$$

where the term

$$S_n = \sum_{k=1}^{n} e^{-\frac{(n-k)T}{T_0}} \qquad \textbf{8.11(7)}$$

is due to the convolution of functions.

This is referred to as the position form of the PID control algorithm, because the actual control output is computed. An alternative form of this algorithm is the velocity form, in which the change of the controller output from the preceding period is computed rather than the actual controller output itself. In this case, the velocity form corresponding to 8.11(5) becomes

$$m_n - m_{n-1} = \frac{100}{\mathrm{PB}}\left[\left(1+\frac{T}{T_I}\right)e_n - e_{n-1} - \frac{T_D}{T_0 T}\right.$$
$$\left.\times[S_n y_n - (S_n + S_{n-1})y_{n-1} + S_{n-1}y_{n-2}]\right] \qquad \textbf{8.11(8)}$$

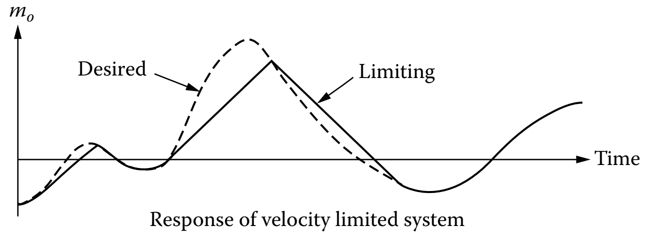


FIG. 8.11a
Velocity-limiting response.

The track-store operation is controlled by the pulse P occurring once during each time interval T to produce an output pulse $\Delta m_n = m_n - m_{n-1}$ of amplitude corresponding to the desired change in the control variable.

Control Valve Simulation

The most important actuating or manipulating output element is the control valve, which may have several distinct aspects in computer simulation. The flow characteristics can generally be linear, equal-percentage, hyperbolic, quick opening, or square root. The valve characteristics can be modeled by special computer circuits or by function-generating blocks for true representation. Additional factors, such as limiting or backlash in valve stem movements, must be included.

The dynamic behavior of a control valve may be represented by a time lag of first or second order, with a limited velocity in stem movement. The time lags are simulated by lag or lag-lead elements. Velocity limiting may be expressed by the differential equation

$$\frac{dx}{dt} = \min\left(v_L; \frac{dx_{\text{ideal}}}{dt}\right). \qquad \mathbf{8.11(9)}$$

where x is the stem position, v_L denotes the velocity limit, and x_{ideal} is the input position of an ideal, unconstrained valve.

The response of a velocity-limited device is shown in Figure 8.11a. This can respond perfectly to a driving function as long as the maximum velocity does not exceed the limit v_L. When the velocity exceeds the limiting value, the gain of the actuator decreases and its phase lag increases as shown in Figure 8.11a.

REACTOR MODELS

In any production system where a chemical change is taking place the chemical reactor is the heart of the process. The remaining operational units of the plant serve for preparing the reactants to make one or more products of greater value, and for the separation, purification, and formation of the products produced in the reactor or reactor system. Thus, the plant efficiency hinges to a great extent on the efficient operation of the reactor.

Chemical reactors may be divided into two main categories: homogeneous and heterogeneous. In homogeneous reactors only one phase, usually a gas or a liquid phase, is present. In heterogeneous reactors two, or possibly three, phases are present. Common examples are gas-liquid, gas-solid, liquid-solid, liquid-liquid, and gas-liquid-solid systems. In cases where one of the phases is a solid, it quite often serves as a catalyst.

The basis for the mathematical model of a chemical reactor are the fundamental physical and chemical laws such as the conservation of mass, energy, and momentum, as well as chemical equilibrium and chemical kinetics. The detailed mathematical model of a process reactor, however, often consists of a large number of nonlinear differential and algebraic equations, which usually involve more information that is relevant to the objectives of simulation and control. As a consequence, usually simplified, reduced models are used to model the flow conditions of process reactors.

The two basic types of flowing reactors are shown in Figure 8.11b. In the continuous stirred tank reactor (a) the agitator thoroughly disperses the reactants into the reaction mixture immediately as they enter the tank. In the tubular-flow reactor (b) there is as little mixing as possible between the reactants, which are flowing through the tube. These two reactors are described by the perfect mixing model and plug-flow model, respectively. These represent the two limiting cases of flow models.

Reaction Kinetics

The principle of chemical kinetics determines the rate at which a reaction proceeds toward the maximum extent determined by the chemical equilibrium. The reaction rate depends on the concentrations of the reactants and on temperature. For gas

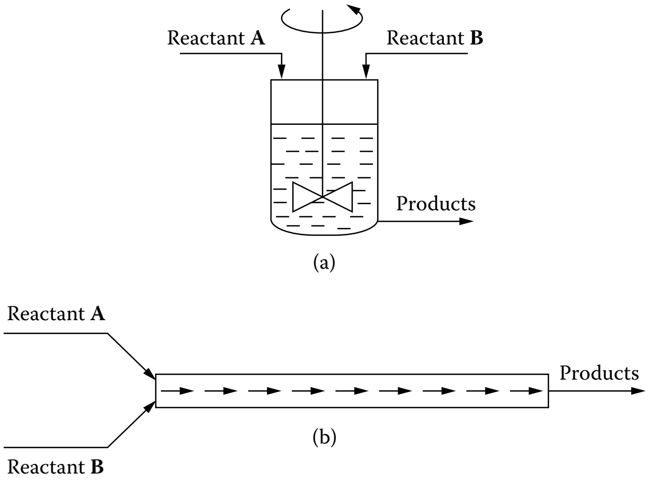


FIG. 8.11b
Two basic types of flowing chemical reactors are the continuous stirred tank (a) and the plug-flow variety (b).

phase reactions the concentrations are proportional to the pressure, and therefore the pressure also affects the rate of reactions.

In case of a homogeneous reaction:

$$a\mathbf{A} + a\mathbf{B} \rightarrow q\mathbf{Q} + s\mathbf{S} \qquad \mathbf{8.11(10)}$$

where: "*a*" moles of species **A** react with "*b*" moles of species **B** to produce "*q*" moles of species **Q** and "*s*" moles of species **S**. **A** and **B** are therefore the reactants, **Q** and **S** are the products, and *a, b, c,* and *d* stand for the corresponding coefficients in the stoichiometric Equation 8.11(10). Such a reaction is called irreversible, and it will stop when either reactant is completely consumed.

For homogeneous irreversible reactions the rate of reaction is measured as moles of one of the reactants transformed per unit volume of reacting mixture and unit time:

$$r_A = -\frac{1}{V}\frac{dn_A}{dt} \qquad \mathbf{8.11(11)}$$

where r_A is the rate of consumption of reactant **A**, and n_A denotes the total moles of **A** in volume V. The rate of reaction, however, can be expressed also in terms of other reactants and products. Because in this case the direction of change of species **Q** and **S** are reversed, these rates are related to rate r_A as

$$r_B = \frac{b}{a}r_A, \quad r_Q = -\frac{q}{a}r_A, \quad r_S = -\frac{s}{a}r_A \qquad \mathbf{8.11(12)}$$

The reaction that does not proceed to completion is called reversible, and it is necessary to include also the kinetics of the reverse reaction:

$$a\mathbf{A} + b\mathbf{B} \leftrightarrow q\mathbf{Q} + s\mathbf{S}. \qquad \mathbf{8.11(13)}$$

The arrow pointing to the left indicates that the reaction is reversible; i.e., both forward and reverse reactions take place simultaneously. Therefore, the state of chemical equilibrium is regarded as a dynamic balance between the forward and reverse reaction.

The equations that express the rate of reaction in terms of concentrations of the reactants and products are called the kinetic model of the reaction. A typical model for the reversible reaction represented by Equation 8.11(13) could be

$$r_A = -k_f\left(c_A^{\alpha}c_B^{\beta} - \frac{1}{K}c_Q^{\varsigma}c_S^{\sigma}\right) \qquad \mathbf{8.11(14)}$$

where c_A, c_B, c_Q, and c_S are the concentrations of the corresponding species in moles per unit volume. The exponents in this expression are often equal to *a, b, q,* and *s*, respectively, and then the reaction is said to be elementary.

If the exponents differ from the coefficients of Equation 8.11(13), the reaction is not elementary, meaning that the actual mechanism of the reaction consists of a series of intermediate elementary steps that are summarized by Equations 8.11(13) and 8.11(14). When the functional relationship has the form of Equation 8.11(14), the reaction is said to be of order α, β, ς and σ, with respect to the species, while the orders of the forward and reverse reactions overall are $\alpha + \beta$ and and $\varsigma + \sigma$, respectively.

In Equation 8.11(14), parameter k_f is the rate coefficient of the forward reaction, and K is the equilibrium constant for which the relationship

$$K = \frac{k_f}{k_r} \qquad \mathbf{8.11(15)}$$

holds where k_r is the rate coefficient of the reverse reaction. For irreversible reaction $k_r = 0$, so that $K \rightarrow \infty$ and the second term in parentheses in Equation 8.11(14) becomes 0.

Both the rate coefficient and the equilibrium coefficient are functions of temperature. Experimentally, the influence of temperature on the rate coefficient is well represented by the Arrhenius equation

$$k_f = k_{f0}e^{-\frac{E}{RT}} \qquad \mathbf{8.11(16)}$$

where E is called the activation energy of the reaction, R is the universal gas constant (1.98 BTU/(lb-mole °R) or in the SI system 8.314 J/(mol °K)), and T is the absolute temperature of the reacting mixture.

A kinetic model is essential for the simulation of a process reactor. The functional relationship and the parameters are identified with the aid of reaction rate data measured in the laboratory or pilot plant, fitting a rate equation of general form to a set of experimental data. However, the kinetic model does not have to be as accurate for control simulation as it would have to be for designing the reactor.

Perfectly Mixed Flow Reactor

The essential feature of the continuous stirred tank reactor is the assumption of complete uniformity of concentration and temperature throughout the reactor volume. Because with perfect mixing, the reactor contents are completely uniform, it is also called the perfectly mixed flow reactor in which the conversion takes place at unique concentration and temperature levels.

The mathematical model is developed using the principle of the conservation of energy and mass:

$$\begin{pmatrix}\text{rate of}\\ \text{quantity}\\ \text{accumulation}\end{pmatrix} = \begin{pmatrix}\text{rate of}\\ \text{quantity}\\ \text{in}\end{pmatrix} - \begin{pmatrix}\text{rate of}\\ \text{quantity}\\ \text{out}\end{pmatrix} + \begin{pmatrix}\text{rate of}\\ \text{quantity}\\ \text{produced}\end{pmatrix} - \begin{pmatrix}\text{rate of}\\ \text{quantity}\\ \text{consumed}\end{pmatrix} \qquad \mathbf{8.11(17)}$$

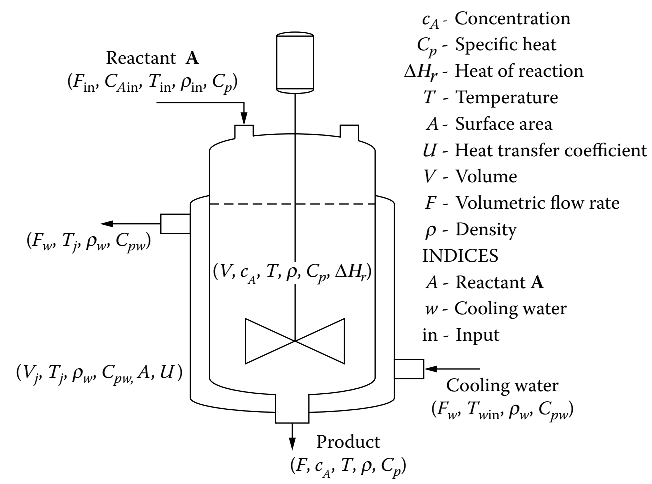


FIG. 8.11c

The parameters used in modeling a perfectly mixed flow reactor with a cooling jacket.

where the quantity may represent mass, mass of individual components, and energy. Equation 8.11(17) is called balance equation of the given quantity.

If an irreversible, exothermic reaction resulting in **A** → *Product* is carried out in the reactor, which is cooled by water flowing through a jacket, the total mass balance in the reactor (shown in Figure 8.11c) can be written as

$$\frac{d(V\rho)}{dt} = F_{\text{in}}\rho_{\text{in}} - F\rho \qquad \mathbf{8.11(18)}$$

where V is the volume and ρ is the mean density of the reacting mixture, and F denotes the volumetric rate of flow. The subscript in is referred to the inputs of the system.

If the density of the reacting mixture is not changed appreciably by the reaction ($\rho_{\text{in}} = \rho$), then Equation 8.11(18) can be simplified as

$$\frac{dV}{dt} = F_{\text{in}} - F \qquad \mathbf{8.11(19)}$$

describing only the change in volume of the reacting mixture as a function of time.

The mass balance equation on reactant **A** is given by

$$\frac{d(Vc_A)}{dt} = F_{\text{in}}c_{A\text{in}} - Fc_A + Vr_A \qquad \mathbf{8.11(20)}$$

where c_A denotes the concentration of reactant **A**, and r_A is the reaction rate expressed in terms of concentration of reactant **A**.

Taking into account that

$$\frac{d(Vc_A)}{dt} = V\frac{dc_A}{dt} + c_A\frac{dV}{dt} \qquad \mathbf{8.11(21)}$$

Equations 8.11(19) and 8.11(20) can be combined, which results in the equation

$$\frac{dc_A}{dt} = \frac{F_{\text{in}}}{V}(c_{A\text{in}} - c_A) + r_A. \qquad \mathbf{8.11(22)}$$

The ratio F_{in}/V is the reciprocal of the mean residence time of the reactor, which is, in principle, the time constant of the perfectly mixed reactor.

The Energy Balance The energy balance equation on the reactor is written as

$$\begin{aligned}\frac{d(V\rho C_pT)}{dt} &= F_{\text{in}}\rho_{\text{in}}C_{p\text{in}}T_{\text{in}} - F\rho C_pT \\ &\quad + V(-\Delta H_r)r_A - UA(T - T_j)\end{aligned} \qquad \mathbf{8.11(23)}$$

where T and C_p are the temperature and the mean specific heat of the reacting mixture, and, again, index in denotes the corresponding input variables. ($-\Delta H_r$) denotes the heat liberated in the reaction per unit mass of **A** reacted, T_j is the temperature of the cooling water in the jacket, A is the effective heat-transfer surface area between the reactor and the jacket, and U is the overall heat-transfer coefficient.

Assuming that the density and the specific heat are constant and using similar reasoning that resulted in Equation 8.11(22), Equation 8.11(23) can be reduced to the form

$$\frac{dT}{dt} = F_{\text{in}}(T_{\text{in}} - T) + \frac{(-\Delta H_r)}{\rho C_p}r_A - \frac{UA}{V\rho C_p}(T - T_j). \qquad \mathbf{8.11(24)}$$

Finally, the energy balance on the jacket gives the equation for the temperature of the jacket T_j taking the form

$$\frac{dT_j}{dt} = \frac{F_w}{V_j\rho_wC_{pw}}(T_w - T_j) + \frac{UA}{V_j\rho_wC_{pw}}(T - T_j) \qquad \mathbf{8.11(25)}$$

where F_w, C_{pw}, and T_w are the volumetric flow rate, the specific heat, and the inlet temperature of the cooling water, respectively.

Gas Phase Reactors In gas phase reactors, the reactor pressure significantly affects the performance of the reactor so that the description of the changes of pressure should be included into the model. In this case the density varies with pressure and composition, and the flow out of the reactor passes through a control valve. The density usually is computed using the perfect gas law

$$\rho = \frac{Mp}{RT} \qquad \mathbf{8.11(26)}$$

where M is the mean molecular weight of the reacting mixture and R is the universal gas constant. The concentration of reactant **A** is expressed as

$$c_A = \frac{p y_A}{RT} \qquad \textbf{8.11(27)}$$

where y_A is the mole fraction of species **A**.

In Equation 8.11(18), the outflow through the control valve can be expressed as

$$F = C_V \sqrt{\frac{p - p_{\text{out}}}{\rho}} \qquad \textbf{8.11(28)}$$

where C_V is the valve sizing coefficient, and p_{out} is the constant pressure controlling the out-flow. Equations 8.11(26)–8.11(28) can be substituted into Equations 8.11(18), 8.11(20), and 8.11(24) to express them in terms of ρ, p, y_A, and T to obtain the mathematical model of the gas phase reactor. Naturally, in this case the volume of the reactor V is constant.

Plug Flow Reactor

Plug flow is a simplified and idealized view of the motion of a fluid, whereby all the fluid elements move with a uniform velocity along parallel streamlines. It is also assumed that the concentrations and temperature are uniform across the cross-sectional area of flow. Tubular reactors, such as shown in Figure 8.11d, with very high velocities, tend to approach plug flow conditions.

Whereas composition and temperature are functions of time only in the perfectly mixed reactor, in the plug flow reactor they are functions of time and distance x from the entrance. Thus, the plug flow reactor is a distributed parameter system, having a dynamic model, which consists of partial differential equations. To solve a partial differential equation it must be approximated by a number of ordinary differential equations. However, a special case of a model containing N equal-size perfectly mixed tanks, called compartments in the present context, arranged in series forms, is a useful and frequently used approximation of the plug flow reactor. This method will be shown here.

The tubular reactor in Figure 8.11d, in which an irreversible, endothermic reaction **A** → *Product* is assumed to be carried out, is divided into a number of compartments, as is illustrated in Figure 8.11e. The compartments are assumed to be connected in cascade without any back-flow. The length of each compartment is Δx and its volume is $A\Delta x$, where A is the cross-sectional area of flow. As a consequence, if the volume of the reactor is V and N compartments are used in the modeling, then the length of a compartment is derived as $\Delta x = V/NA$. Further, if the linear velocity of flow is denoted by v then the volumetric velocity is Av.

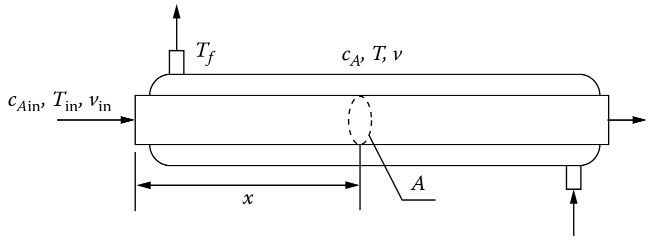


FIG. 8.11d
Plug flow reactor with heating jacket.

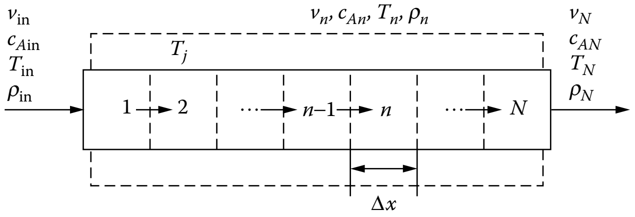


FIG. 8.11e
The model of a plug flow reactor consists of a number of compartments.

The mathematical model of the system now can be developed by sequentially applying the principle expressed in Equation 8.11(17) for each compartment. The mass balance around the nth compartment, shown in Figure 8.11f, can be written as

$$\frac{d(A\Delta x \rho_n)}{dt} = A v_{n-1}\rho_{n-1} - A v_n \rho_n, \quad n = 1,2\ldots N \qquad \textbf{8.11(29)}$$

where ρ_n is the mean density and v_n is the mean linear velocity of the reacting mixture in the nth compartment. For the first compartment $\rho_0 = \rho_{\text{in}}$ and $v_0 = v_{\text{in}}$, where the subscript in denotes the input variables to the reactor. In liquid phase reactors, the density is assumed to be constant. In gas phase reactors Equation 8.11(26) is used to compute the density from the pressure and temperature of the compartment.

Because the parameters x and A are constant, Equation 8.11(29) can be simplified to give

$$\frac{d\rho_n}{dt} = \frac{1}{\Delta x}(v_{n-1}\rho_{n-1} - v_n\rho_n), \quad n = 1,2\ldots N,$$
$$v_0 = v_{\text{in}}, \rho_0 = \rho_{\text{in}} \qquad \textbf{8.11(30)}$$

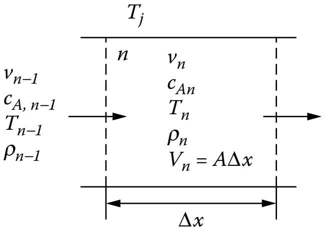


FIG. 8.11f
The compartment model of the nth compartment of a plug flow reactor.

The mass balance equation on reactant **A** becomes

$$\frac{d(A\Delta x c_{An})}{dt} = Av_{n-1}c_{A,n-1} - Av_n c_{An} + A\Delta x r_{An},$$
$$n = 1,2\ldots N, v_0 = v_{\text{in}}, \rho_0 = \rho_{\text{in}}, c_{A0} = c_{A\text{in}} \qquad \mathbf{8.11(31)}$$

where c_{An} is the concentration of reactant **A** in the nth compartment, and r_{An} is the reaction rate. Again, this equation can be simplified writing

$$\frac{dc_{An}}{dt} = \frac{1}{\Delta x}(v_{n-1}c_{A,n-1} - v_n c_{An}) + r_{An}, \quad n = 1,2\ldots N,$$
$$v_0 = v_{\text{in}}, \rho_0 = \rho_{\text{in}}, c_{A0} = c_{A\text{in}} \qquad \mathbf{8.11(32)}$$

The Energy Balance The energy balance equation around the nth compartment takes the form

$$\frac{d(A\Delta x \rho_n C_p T_n)}{dt} = Av_{n-1}\rho_{n-1}C_p T_{n-1} - Av_n\rho_n C_p T_n$$
$$+ A\Delta x(-\Delta H_r)r_{An} + Ua\Delta x(T_j - T_n)$$
$$n = 1,2\ldots N, v_0 = v_{\text{in}}, \rho_0 = \rho_{\text{in}}, T_0 = T_{\text{in}} \qquad \mathbf{8.11(33)}$$

where a is the perimeter of the tube, so that $a\Delta x$ is the heat-transfer area of one compartment, and T_j is the temperature of the jacket assumed to be uniform along the tube, as it takes place, for instance, in a firing box. This last equation can also be written in simplified form

$$\frac{dT_n}{dt} = \frac{v_{n-1}\rho_{n-1}}{\Delta x \rho_n}(T_{n-1} - T_n) + \frac{1}{\rho_n C_p}(-\Delta H_r)r_{An} + \frac{Ua}{A\rho_n C_p}(T_j - T_n)$$
$$n = 1,2\ldots N, v_0 = v_{\text{in}}, \rho_0 = \rho_{\text{in}}, T_0 = T_{\text{in}} \qquad \mathbf{8.11(34)}$$

Often, the pressure drop in the reactor is sufficiently large to be necessary to account for it, instead of using an average value. Then the velocity from each compartment to the next can be computed by using the usual Bernoulli equation

$$p_n - p_{n-1} = k_F\rho_n v_n^2 + \alpha\rho_n v_n(v_n - v_{n-1}),$$
$$n = 1,2\ldots N,\ v_0 = v_{\text{in}}, \rho_0 = \rho_{\text{in}} \qquad \mathbf{8.11(35)}$$

where the constant k_F is proportional to the Fanning friction factor, which can be computed using fluid dynamics correlation. α is a conversion factor in the Fanning pressure drop equation. In liquid phase reactors, the velocity can also be assumed constant, thus Equations 8.11(32), 8.11(34), and 8.11(35) by themselves form a closed model of the reactor.

Gas Phase Reactions In gas phase reactors, however, the mathematical model consists of Equations 8.11(30), 8.11(32), 8.11(34), and 8.11(35), and this set of equations can be closed by also applying Equation 8.11(26). If the temperature of the jacket T_j cannot be considered constant, then a heat balance equation, similar to Equation 8.11(33) without reaction term, should be added to the model.

The greater the number of compartments in this compartments model, the better the model approximates the distributed parameter plug flow model. However, in accordance with the engineering heuristics, five compartments usually are sufficient to describe the performance and behavior of plug flow reactors.

Nonideal Flow

The flow patterns of process reactors often deviate from the ideal flow conditions of perfect mixing and plug flow. In stirred tank reactors, this can be caused by such elements in real equipment as corners, baffles that can lead to stagnant regions, or nonuniform flow paths that can induce bypassing of fluid.

The way to model this type of imperfections is to divide the reactor into more compartments of different volumes. These compartments can be connected with each other using volumetric flows of different flow rates. The sum of the volumes of the compartments must equal the volume of the reactor. A simple model of a real stirred tank with bypassing and dead space is shown in Figure 8.11g.

The nonideal flow in tubular reactors can be modeled simply by adding back-flow from a compartment to the preceding one in the N compartments in series model. In this manner, one can allow for some mixing in the axial direction of the tube, as it is illustrated in Figure 8.11h.

Increasing the rate of back-flow, the model will approach that of the perfectly mixed reactor. In the case of fixed finite back-flow, when the number of compartments, N, approaches infinity, the model approaches the so-called plug flow with axial dispersion model.

Another type of nonideal flow occurs in a tubular reactor in which the velocity, temperature, and concentration varies across the radius of the tube. If axial symmetry is present, these phenomena may be modeled by assuming some velocity

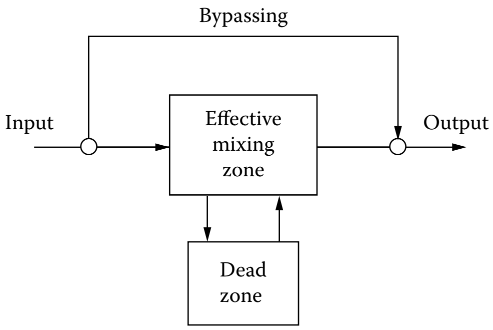


FIG. 8.11g
The model of a real stirred tank reactor includes both bypassing and dead space components.

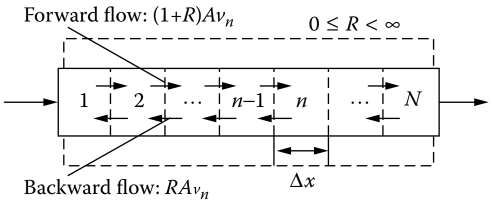


FIG. 8.11h
Plug flow reactor model that includes back-flow among the compartments.

profile along the radius that further can be combined with diffusion of chemical species and conduction of temperature in radial direction.

In such cases, the reactor can be divided into a number of "rings," and the balance equations can be written for each of these rings. For the transfer of heat and mass between the rings, appropriate flow models and interfacial heat and mass transfer equations must be used.

Heterogeneous Reactors

Modeling and simulation of multiphase heterogeneous reactors falls into a special category because an additional complexity, namely the transfer of mass and energy between phases, enters into the problem. In general, mass and energy balance equations must be written for each phase, and appropriate interfacial transfer equations must be used for describing the transfer of reactants and products between the phases.

The transport of mass and energy inside the phases are described by models discussed in the case of homogeneous reactors, while the interfacial mass and heat transfers are expressed in terms of appropriate driving forces as:

$$\Phi_A = k_L(c_{Ak} - c_{Ai}) \qquad \textbf{8.11(36)}$$

where Φ_A is the mass flux through the interface, k_L is the mass transfer coefficient, and $(c_{Ak} - c_{Ai})$ denotes the driving force expressed as a difference of concentrations of species **A** in the *k*th and *i*th phases. For this model, it is important to determine the interfacial surface area between the phases, which could be difficult. The two most significant physical properties of the system are the solubility and the phase equilibrium relationships.

For fast reactions and low diffusion or mass transfer rates, the rate of mass transfer has greater influence on the rate of chemical reactions than does the kinetics of the reaction. In such cases, the mass transfer is the step that determines the rate, and the reaction rate equations may be quite different compared with the previously discussed ones.

Catalytic Reactors

In catalytic reactors, the catalyst speeds up the reaction without being consumed or produced by it. The reaction rate coefficient is then a function of the temperature, catalyst concentration, and catalyst activity or age. Catalysts frequently lose a large portion of their activity while in operation. The age of the catalyst, being a measure of loss of the catalyst activity, is measured as the accumulated time that has passed since the catalyst was replaced or regenerated. These values usually are not readily available and must be determined experimentally in order to be incorporated into the model.

When the catalyst is a solid, the additional complications of heterogeneity arise in the modeling process. In fluid-solid catalytic reactors, the reaction itself takes place on the surface of the solid. For this reason, it is advantageous to have the particles of catalyst in a porous form making a large internal surface area available for the reaction. In this case, the rate of transport of gas to and from the surface within the pore structure will significantly affect the performance of the reactor and has to be taken into consideration. The effectiveness factors usually provide a useful tool to make the model of the reactor treatable. In general, this is defined as

$$\eta = \frac{\text{rate of reaction with pore diffusion resistance}}{\text{rate of reaction without pore diffusion resistance}} \qquad \textbf{8.11(37)}$$

and the actual reaction rate that would be observed is

$$(r_A)_{\text{observed}} = \eta (r_A)_{\text{ideal}} \qquad \textbf{8.11(38)}$$

where the reaction rate $(r_A)_{\text{ideal}}$ is derived with surface conditions. The effectiveness factor depends on the shape of catalyst particles.

Sometimes, especially in fixed-bed catalytic reactors, the equations can be considerably simplified by expressing the rate of reaction on the basis of unit mass of catalyst instead of on the basis of unit volume.

CONCLUSIONS

Presented in this section were methods for developing the equations that describe process reactors. Because the reactor is only one part of the total control loop, for a complete simulation of the loop, it is necessary to include the equations that represent the temperature, level, pressure, and concentration sensors, transmission lines, controllers, and control valves.

A reactor control simulation allows the control engineer to tune the controllers without any loss of production or risk of an accident. Simulation also provides the perfect tool to train plant operators for smoother and safer start-ups. In addition, it serves as a "live" model with which to try new control ideas that will result in safer and more efficient operation of the reactor; it gives an insight into the behavior of the reactor equivalent to several years of reactor operation. This last advantage derives from the ability to look at variables in the simulation that are impossible or impractical to measure in the process reactor.

When modeling on digital computers, one must recognize the following:

1. The digital computer is a discrete time device; therefore, all the computations required to implement must be expressed in discrete time forms.
2. Numerical integration involves approximating continuous differential equations with discrete finite difference equations. The accuracy and stability of these approximating equations must be constantly considered.
3. Interactive environments of digital computers provide convenient and effective tools and some kind of graphical capabilities for modeling, simulating, and analyzing dynamic, multidomain systems and control system design. It lets you accurately describe, simulate, evaluate, and refine a control system's behavior through standard and custom block libraries, also allowing you to develop and build your own special process and control elements.

Bibliography

Coulson, J. M., and Richardson, J. F., *Chemical Engineering,* Volume 3, (Richardson, J. F., and Peacock, D. G., Eds.), Oxford: Elsevier, 1994.

Douglas, J. M., *Process Dynamics and Control 1. Analysis of Dynamic Systems,* Englewood Cliffs, NJ: Prentice Hall, 1972.

Douglas, J. M., *Process Dynamics and Control 2. Control System Synthesis*. Englewood Cliffs, NJ: Prentice Hall, 1972.

Froment, G. F., and Bischoff, K. B., *Chemical Reactor Analysis and Design,* New York: John Wiley & Sons, 1979.

Gershenfeld, N., *The Nature of Mathematical Modeling,* Cambridge: Cambridge University Press, 1999.

Harriott, P., *Process Control,* New York: McGraw-Hill, 1964.

Levenspiel, O., *Chemical Reaction Engineering,* New York: Wiley, 1999.

Luyben, W. L., *Process Modeling, Simulation and Control for Chemical Engineers.* New York: McGraw-Hill, 1973.

Mayer, F. X., and Spencer, E. H., "Computer Simulation of Reactor Control," *ISA Journal,* July 1961, 58–64.

Morely, R., "Computer Control for the 1990s," *Automation,* June 1991, 28–31.

Morely, R., "Faster than Real Time," *Manufacturing Systems,* May 1992, 58–65.

Rao, M., and Qiu, H., *Process Control Engineering,* Chemin de la Sallaz: Gordon & Breach, 1993.

Shinskey, F. G., *Process Control Systems. Application, Design, and Tuning,* New York: McGraw-Hill, 1996.

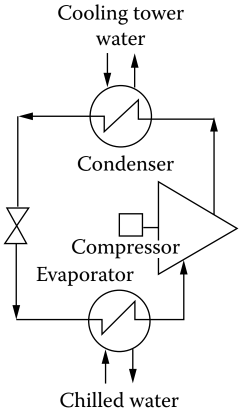
Smith, C. A., and Corripio, A. B., *Principles and Practice of Automatic Process Control,* New York: Wiley, 1985.

Stephanopoulos, G., *Chemical Process Control: An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

Wen, C. Y., and Fan, L. T., *Models for Flow Systems and Chemical Reactors,* New York: Marcel Dekker, 1975.

8.12 Chiller Control

B. G. LIPTÁK (1970, 1985, 1995, 2005)



Flow sheet symbol

INTRODUCTION

The chiller process and its basic controls are discussed in this section, while the next section covers various methods of chiller optimization. As the overall cooling process also includes cooling towers, compressors, fans, and pumps, the reader should also refer to the sections that discuss the control of these systems in this chapter.

The unit of refrigeration is the ton. The rate of cooling represented by the three widely used definitions of a ton of refrigeration are listed below:

Standard Ton	200 BTU/min (3520 W)
British Ton	237.6 BTU/min (4182 W)
European Ton (Frigorie)	50 BTU/min (880 W)

The performance and operating features of the various chiller designs vary over a wide range. Some of these are listed below:

Compression ratios (per stage): 3:1 to 8:1
Maximum turndown ratios: 8:1 for conventional, over 30:1 with hot gas bypass (HGBP)
Typical superheat at evaporator outlet: 9°F, or 5°C
Coefficients of performance: 2.5–3.5 for conventional, 5–7 for optimized chillers

When the operation of a chiller is optimized (subject of the next section) its yearly cost of operation is reduced by 1.5% for each °F reduction in chiller ΔT.

The Heat Pump

Heat pumps serve to transport heat from a lower to a higher temperature level. They do not make heat; they just transport it. This is similar to a pump transporting water from a lower to a higher elevation. The required energy input to a regular pump is a function not so much of the amount of water to be transported but of the elevation to which it is to be lifted. Similarly, the required energy input of a heat pump is a function not only of the amount of cooling it has to do but also of the temperature at which it must reject the heat, because it is this elevation against which it is pumping.

Each °F reduction in the temperature difference (ΔT), across which the heat pump is transporting the heat, will lower the yearly operating cost by 1.5% (each °C by 2.7%). As it will be discussed in the next section, there are not many unit operations whose efficiency can be doubled through optimization; chiller optimization is one of these few.

In this review of chiller controls, the thermodynamic aspects of the refrigeration process will be discussed first, after which conventional chiller controls will be discussed. The nomenclature and abbreviations used in connection with chiller controls and optimization are listed below.

Nomenclature

CHWP: Chilled water pump
CTWP: Cooling tower water pump
FAH, FAL: Flow alarm; H = high, L = low
FC: Fail closed
FO: Fail open
FSH, FSL, FSHL: Flow switch; H = high, L = low, HL = high-low
HLLS: High-low limit switch
HWP: Hot water pump
LCV: Level-control valve
$M_{1–5}$: Motors that drive equipment; 1 = cooling tower fan(s), 2 = cooling water pump(s), 3 = compressor(s), 4 = chilled water pump(s), 5 = hot water pump(s)

P-1: Pump
P, P_{1-3}: Pressures
PAH: Pressure alarm; H = high
PB: Pushbutton
PCV: Pressure-control valve
PDIC: Pressure-differential indicating controller
PI: Pressure indicator
PSH, PSL: Pressure switch; H = high, L = low
Q_H: Amount of heat delivered to cooling tower water
Q_L: Amount of heat removed from chilled water
RD: Rupture disk
S: Solenoid
SC: Speed controller
SIC: Speed-indicating controller
SP: Set point
SV: Solenoid valve
T_c: Temperature of refrigerant in condenser inlet
T_e: Temperature of refrigerant in evaporator inlet
T_H: Temperature of cooling water at condenser exit, absolute
T_L: Temperature of chilled water at evaporator exit, absolute
T_{chwr}: Temperature of chilled water return
T_{chws}: Temperature of chilled water supply
T_{ctwr}: Temperature of cooling tower water return
T_{ctws}: Temperature of cooling tower water supply
T_{hwr}: Temperature of hot water return
T_{hws}: Temperature of hot water supply
T_{wb}: Temperature of wet bulb
TAH, TAL: Temperature alarm; H = high, L = low
TCV: Temperature control valve
TIC: Temperature-indicating controller
TSH, TSL: Temperature switch; H = high, L = low
TT: Temperature transmitter
TY: Temperature relay
VPC: Valve position controller
W: Work
XLS: Limit switch
XSCV: Superheat control valve

THE COOLING PROCESS

Figure 8.12a illustrates how heat pumps can transport heat from a lower to a higher elevation and thereby provide cooling of a relatively low temperature process. The heat pump removes Q_L amount of heat from the chilled water and, at the cost of investing W amount of work, delivers Q_H quantity of heat to the warmer cooling tower (or condenser) water.

In the lower part of this illustration, the idealized temperature-entropy cycle is shown for the chiller. The cycle consists of two isothermal and two isentropic (adiabatic) processes:

1→2: Adiabatic process that occurs in the expansion valve
2→3: Isothermal process that takes place in the evaporator

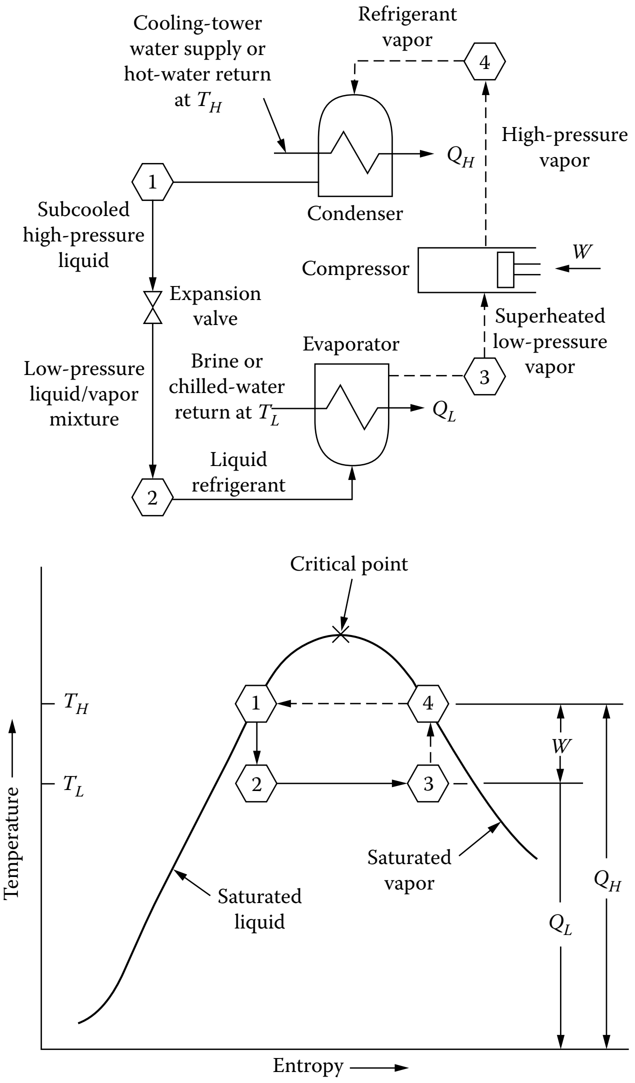


FIG. 8.12a
The top portion of this figure shows the main components of a refrigeration system. The bottom portion describes the refrigeration cycle, which consists of two isothermal and two isentropic processes.[1]

3→4: Adiabatic process that occurs in the compressor
4→1: Isothermal process that takes place in the condenser

The isothermal processes in this cycle are also isobaric (constant pressure). The efficiency of a heat pump is defined as the ratio between the heat removed from the process (Q_L) and the work required to accomplish this heat removal (W).

$$\beta = \frac{Q_L}{W} = \frac{Tl}{Th - Tl} \qquad \mathbf{8.12(1)}$$

Coefficient of Performance

Because the efficiency of a chiller can be much more than 100%, it is usually called the *coefficient of performance*

(COP). If a chiller requires 1.0 kWh (3,412 BTU/h) to provide a ton of refrigeration (12,000 BTU/h), its coefficient of performance is said to be 3.5. This means that each unit of energy introduced at the compressor will pump 3.5 units of heat energy into the cooling tower water.

Conventionally controlled chillers operate with COPs in the range of 2.5 to 3.5. Optimization can double the COP by increasing T_L, by decreasing T_H, and through other methods that will be described in the next section.

Figure 8.12a shows the four principal pieces of equipment that make up a refrigeration machine. In path 1→2, as it passes through the expansion valve, the high-pressure subcooled refrigerant liquid becomes a low-pressure mixture of liquid and vapor. In path 2→3, in passing through the evaporator, this mixture is vaporized into a superheated low-pressure vapor stream, whereas in the compressor (path 3→4) the pressure of the refrigerant vapor is increased. Finally, in path 4→1 this high-pressure vapor is condensed at constant pressure. The liquid leaving the condenser is usually subcooled, whereas the vapors leaving the evaporator are usually superheated by controlled amounts.

Refrigerants and Heat Transfer Fluids

The fluid that carries the heat from a low to a high temperature level is referred to as the refrigerant. Table 8.12b provides a summary of the more frequently used refrigerants.

The data in Table 8.12b assumes that the evaporator will operate at 5°F (−15°C) and that the temperature of the cooling water will keep the condenser at 86°F (30°C). Other evaporator

TABLE 8.12b
Refrigerant Characteristics

		Feature								
Refrigerant		*Applicable Compressor (R = Reciprocating, RO = Rotary, C = Centrifugal)*	*Boiling Point in °F* at Atmospheric Pressure*	*Evaporator Pressure in psia† if Operating Temperature is 5°F (−15°C)*	*Condenser Pressure in psia† if Operating Temperature is 86°F (30°C)*	*Latent heat in BTU/lbm‡ at 18°F (−7.8°C)*	*Toxic (T), Flammable (F), Irritating (I)*	*Mixes or Compatible with the Lubricating oil*	*Chemically Inert and Noncorrosive*	*Remarks*
Ethane	C_2H_6	R	−127	236	675	148	T&F	No	Yes	For low-temperature service
Carbon dioxide	CO_2	R	−108	334	1039	116	No	Yes	Yes	Low-efficiency refrigerant
Propane	C_3H_8	R	−48	42	155	132	T&F	No	Yes	
Freon-22	$CHClF_2$	R	−41	43	175	92	No	(1)	Yes	For low-temperature service
Ammonia	NH_3	R	−28	34	169	555	T&F	No	(2)	High-efficiency refrigerant
Freon-12	CCl_2F_2	R	−22	26	108	67	No	Yes	Yes	Most recommended
Methyl chloride	CH_3Cl	R	−11	21	95	178	(3)	Yes	(4)	Expansion valve may freeze if water is present
										Common to these refrigerants
Sulfur dioxide	SO_2	R	+14	12	66	166	T&I	No	(4)	
Freon-21	$CHCl_2F$	RO	+48	5	31	108	No	Yes	Yes	a. Evaporator under vacuum
Ethyl chloride	C_2H_5Cl	RO	+54	5	27	175	F&I	No	(5)	b. Low compressor discharge pressure
Freon-11	CCl_3F	C	+75	3	18	83	No	Yes	Yes	c. High volume-to-mass ratio across compressor
Dichloro methane	CH_2Cl_2	C	+105	1	10	155	No	Yes	Yes	

$$*°C = \frac{°F - 32}{1.8}$$

† psia = 6.9 kPa

‡ BTU/lbm = 232.6 J/kg

(1) Oil floats on it at low temperature.
(2) Corrosive to copper-bearing alloys.
(3) Anesthetic
(4) Corrosive in the presence of water.
(5) Attacks rubber compounds.

and condenser temperatures would have illustrated the relative characteristics of the various refrigerants equally well.

It is generally desirable to avoid operation under vacuum in the evaporator or any part of the cycle because of the resulting sealing problems. At the same time, very high condensing pressures are also undesirable because of the resulting structural strength requirements. From this point of view, the most desirable refrigerants are listed in Table 8.12b between propane and methyl chloride, because they display favorable characteristics.

An exception to this reasoning is when very low temperatures are required. For such service, ethane can be the proper selection in spite of the resulting high condenser design pressure.

Another consideration is the latent heat of the refrigerant. The higher it is, the more heat can be carried by the same amount of working fluid and, therefore, the smaller the corresponding equipment size will be. This feature has caused users in the past to tolerate the undesirable characteristics of ammonia.

One of the most important considerations in the selection of the refrigerant is safety. In industrial installations, the desirability of nontoxic, nonirritating, nonflammable refrigerants cannot be overemphasized. It is similarly important that the working fluid be compatible with the compressor's lubricating oil. Corrosive refrigerants are undesirable for the obvious reasons of higher initial cost and increased maintenance.

Most of the working fluids listed in Table 8.12b can be used with reciprocating compressors. Only the last four fluids in the table, which have high volume-to-mass ratios and low compressor discharge pressures, can justify the consideration of rotary or centrifugal machines.

Freon-12 is suited for the largest number of applications, but it does have the disadvantage of contributing to ozone depletion in the atmosphere.

In the majority of industrial installations, the refrigerant evaporator is not used directly to cool the process. More frequently, the evaporator cools the chilled water, which is then piped to cool the process.

For temperatures that are too low for water to be used as a coolant, brine is frequently selected. The limitation of weak brines is that they may freeze, while strong brines can plug the evaporator tubes, if they are not true solutions. For operation around 0°F (–18°C), the sodium brines (NaCl) are recommended, while for services down to –45°F (–43°C), the calcium brines (CaCl) are the best.

Care must be exercised in handling brines, because they are corrosive if they are not kept at a pH of 7 or if oxygen is present. In addition, brine will initiate galvanic corrosion between dissimilar metals.

REFRIGERATOR AND CHILLER DESIGNS

The more conventional heat pump controls are discussed below. For the more sophisticated and optimized control systems, refer to the next section in this chapter.

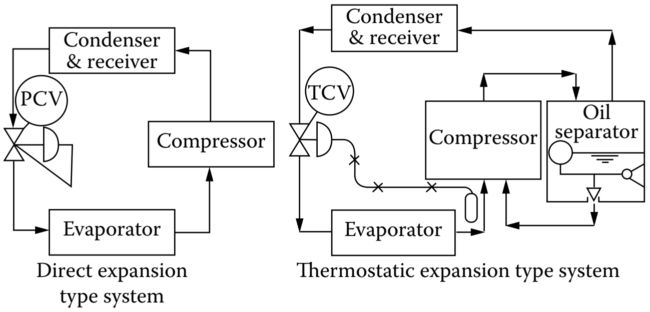


FIG. 8.12c
The small industrial refrigerators are usually provided with either direct expansion (left) or thermostatic expansion-based throttling control.

Small Industrial Refrigerators

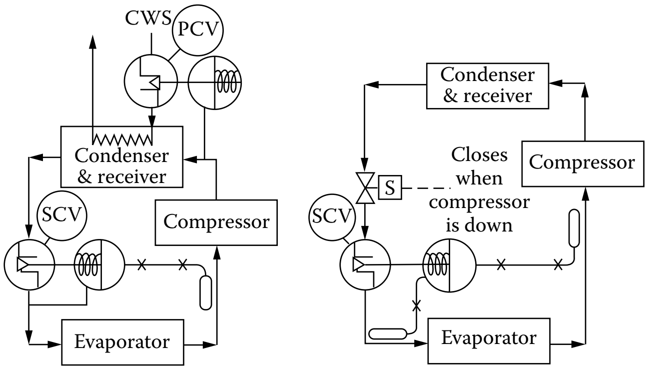
The direct expansion-type control of a refrigerator is shown on the left side of Figure 8.12c. Here a pressure-reducing valve keeps the evaporator pressure constant. The pressure setting is a function of load, and therefore these controls are recommended only for constant-load applications. The pressure setting is set manually by adjusting the pressure until the frost just stops at the end of the evaporator. This indicates that liquid refrigerant is present all the way to that point.

In the direct expansion-type system, shown on the left side of Figure 8.12c, when the load increases, all the refrigerant will vaporize before the end of the evaporator, causing low efficiency, because the unit is "starved." Starving can be eliminated by manually changing the pressure setting. When the evaporator is down, the pressure-control valve closes, isolating the high- and low-pressure sides of the system. This guarantees that the desirable high start-up torque on the compressor will be present.

The thermostatic expansion-type control is shown on the right side of Figure 8.12c. This system, instead of maintaining the evaporator pressure constant, controls the superheat of the evaporated vapors. This design is not limited to constant loads, because under all conditions it guarantees the presence of liquid refrigerant at the end of the evaporator.

Figure 8.12c also shows a typical oil separator.

Expansion Valves Figure 8.12d illustrates some expansion valve design variations. On the left side of Figure 8.12d, a fairly standard superheat control valve is shown. It detects the pressure at the inlet and the temperature at the outlet of the evaporator. If the evaporator pressure drop is low, these measurements (the saturation pressure and the temperature of the refrigerant) are an indication of superheat. The desired superheat is set by the spring in the valve operator, which together with the saturation pressure in the evaporator opposes the opening of the valve. The "superheat feeler bulb" pressure balances these forces when the unit is at equilibrium and is operating at the desired superheat (usually 9°F, or 5°C).



Expansion valve system for low evaporator drop with water regulating valve

Differential temperature expansion valve with overload protection

FIG. 8.12d
The control of the chiller's expansion valve depends on the type of operation. In installations with low evaporator drops, a superheat control valve is recommended (left), while at low-temperature applications, the differential temperature-type valve (right) gives more sensitive performance.

If the process load increases, it causes an increase in the evaporator outlet temperature. An increase in this temperature results in a rise in the "feeler bulb" pressure, which in turn further opens the superheat control valve. This greater flow of refrigerant liquid from the condenser to the evaporator increases the saturation pressure and temperature, and the increased saturation pressure balances against the increased feeler bulb pressure at a new (greater) valve opening in a new equilibrium. As the controls adjust to an increased load, the evaporator pressure increases, but the amount of superheat (set by the valve spring) is kept constant.

On the left side of Figure 8.12d, the operation of the cooling water regulating valve is also illustrated. This valve maintains the condenser pressure constant and at the same time conserves cooling water. At low condenser pressure, such as when the compressor is down, the water valve closes. It starts to open when the compressor is restarted and its discharge pressure reaches the setting of the PCV valve. The water valve opening follows the load by further opening at higher loads in order to maintain the condenser pressure constant. This feature too can be incorporated into any one of the refrigeration units.

When operating at very low temperatures, a small change in refrigerant vapor pressure results in a fairly large change in temperature. For example, a 5°F (2.8°C) temperature change when Freon-12 is at the temperature of –100°F (–73°C) will result in a 0.3 PSIG (2.1 kPa) change in saturation pressure. Therefore, a more sensitive measurement will result if a thermal bulb is used to indirectly detect the saturation pressure in the evaporator. On the right side of Figure 8.12d, a differential temperature-type expansion valve is shown, which takes advantage of the high temperature sensitivity at low operating temperatures.

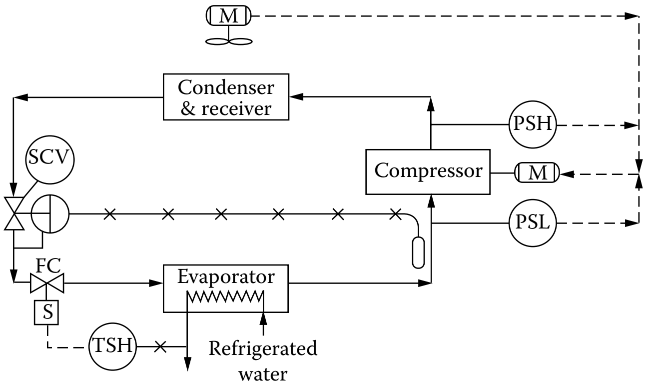


FIG. 8.12e
Small industrial refrigeration units can be controlled in an on/off manner. In this configuration the machine is started when the refrigerated water temperature reaches a preset high and stopped when it reaches a preset low limit.

On/Off Control of Small Units Figure 8.12e shows the controls for a fairly simple and small refrigeration unit. This system includes a conventional superheat control valve, a low-pressure-drop evaporator, a reciprocating on/off compressor, and an air-cooled condenser. The purpose of this refrigeration package is to maintain the temperature of a chilled water supply to the plant within some set limits.

The high-temperature switch (TSH) shown in Figure 8.12e is the main controller. Whenever the temperature of the chilled water drops below a preset value (usually 38°F, or 3.3°C), the refrigeration unit is turned off, and when the chilled water temperature rises to some other level (such as 42°F, or 5.6°C), it is restarted. This on/off cycling control is accomplished by the temperature switch, which is opening and closing a solenoid valve to keep the water temperature low enough. The closing of this valve causes the compressor suction pressure to drop. When it reaches the set point of the low-pressure switch (PSL), it stops the compressor.

While the unit is running, the expansion valve maintains the refrigerant superheat constant and the safety interlocks protect the equipment. These interlocks include such features as turning off the compressor if the fan motor stops or if the compressor discharge pressure becomes too high (PSH) for some other reason.

This unit cannot throttle its level of operation, it can operate only at full compressor capacity or not at all. This type of machine is referred to as a unit that has two-stage unloading. When continuous load following and therefore the varying of the cooling capacity is desired instead of turning it on and off, two possible control techniques are available. One of these approaches is the multistep unloading of reciprocating compressors. In a three-step system the available levels of loading are 100%, 50%, and 0%, whereas with five-step unloading, 100%, 75%, 50%, 25%, and 0% loads can be handled.

Multistage Refrigeration Units It is not practical to obtain a compression ratio outside the range of 3:1 to 8:1 with the compressors used in the process industry. This places a limitation on the minimum temperature that a single-stage refrigeration unit can achieve.

For example, if Freon-12 was the refrigerant, in order to maintain the evaporator at −80°F (−62°C) and the condenser at 86°F (30°C), which is compatible with standard supplies of cooling water, the required compression ratio could be calculated as follows:

$$\text{compression ratio} = \frac{\text{refrigerant pressure at } 86°\text{F}}{\text{refrigerant pressure at} - 80°\text{F}} = \frac{108}{2.9} = 37 \qquad \textbf{8.12(2)}$$

(In the next section, where chiller optimization is discussed, the control of multiple positive-displacement compressors, used in parallel and serving an unlimited number of evaporator coils, will be described.)

Industrial Chillers

The refrigeration unit shown in Figure 8.12f, although far from a "standard" system, does contain some of the features typical of conventional industrial units in the 500 ton (1760 kW) and larger sizes. These features include the capability for meeting continuous changing loads in a continuous manner, as contrasted with stepwise unloading. Other features included are the economizer expansion valve system and the hot gas bypass, which serve to increase rangeability.

The illustrated system provides refrigerated water at 40°F (4.4°C) through the circulating header system to the users of an industrial plant. If the flow rate is fairly constant, changes in the process load will be reflected by the temperature of the returning refrigerated water. Under normal load conditions, this return water temperature is controlled at 50°F (10.6°C) by TIC-2.

As process load decreases, the return water temperature drops correspondingly. With the reduced load on the evaporator, the refrigerated water supply temperature to the plant also drops and the direct acting TIC-1 gradually closes the suction damper or the prerotation vane of the compressor. By throttling the suction vane, a 10:1 turndown ratio can be accomplished. If the load drops below this ratio, the hot gas bypass system has to be activated.

Hot Gas Bypass The hot gas bypass is automatically controlled by TIC-2. Its purpose is to keep the constant-speed compressor out of surge. Therefore, when the load drops to levels sufficiently low to approach surge, this bypass valve is opened. If the chilled water flow rate is constant, the difference between chilled water supply and return temperatures is an indication of the load.

If full load corresponds to a 15°F (8.3°C) difference on the chilled water side of the evaporator and the chilled water supply temperature is controlled by TIC-1 at 40°F (4.4°C), then the return water temperature detected by TIC-2 is also an indication of the load.

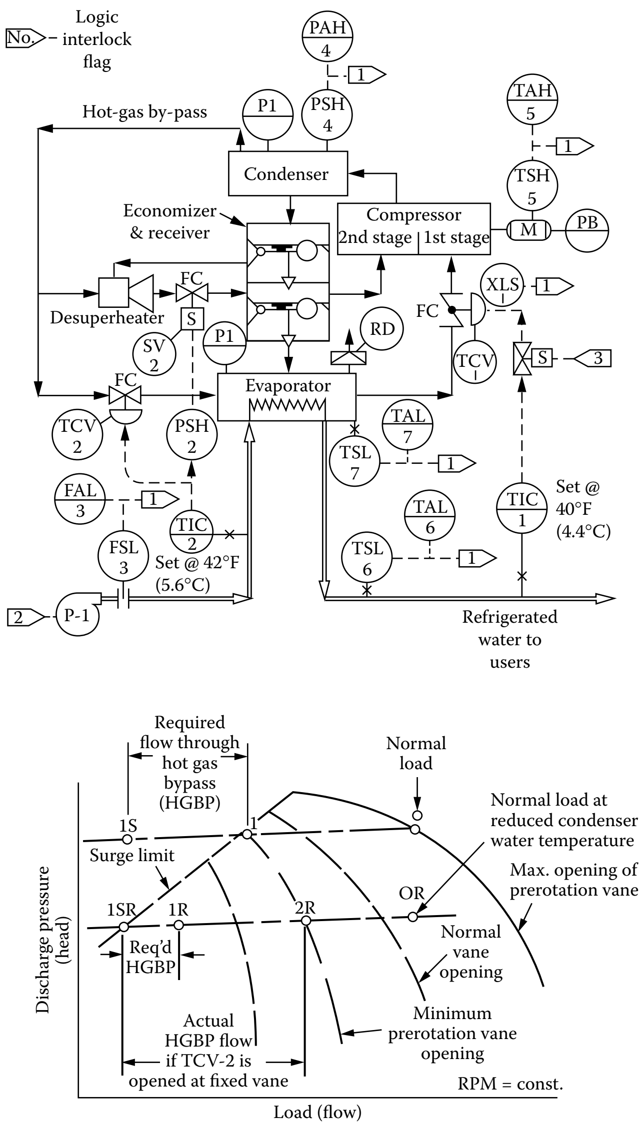


FIG. 8.12f

The rangeability of a conventionally controlled industrial refrigeration system is increased by the addition of a hot gas bypass. (Bottom part of this illustration adapted from Reference 2.)

If the compressor goes into surge at 10% load, this would correspond to a return water temperature of 41.5°F (5.3°C). In order to stay safely away from surge (for a detailed discussion of the phenomenon of surge, refer to the section on compressor control in this chapter), TIC-2 in Figure 8.12f is set at 42°F (5.6°C), corresponding to a load of approximately 13%.

When the temperature drops to 42°F (5.6°C), the bypass control valve TCV-2 starts to open, and its opening can be proportional to the load detected. This means that the valve is fully closed at 42°F (5.6°C), fully open at 40°F (4.4°C), and throttled in between. This throttling action is accomplished by a plain proportional controller (TIC-2), which has a 2°F throttling range. If the span is 0–100°F, this corresponds to a proportional band of 2% or a gain of 50.

The hot gas bypass makes it theoretically possible to achieve a very high turndown ratio by temporarily running the machine on close to zero process load. This operation can be visualized as heat pump, which is transferring heat energy from the refrigerant itself to the cooling water. During this process, some of the refrigerant vapors are condensed, resulting in an overall lowering of operating pressures on the refrigerant side.

The main advantage of a hot gas bypass, therefore, is that it allows the chiller to operate at low loads without going into surge. The price of this operational flexibility is an increase in operating costs, because the work introduced by the compressor is wasted in the form of friction through the hot gas bypass valve (TIC-2). As will be shown in the next section, optimized control systems eliminate this waste through the use of variable-speed compressors, which can respond to a reduction in load by lowering their speed instead of throwing away the unnecessarily introduced energy in the form of friction in TCV-1 and TCV-2.

Instead of controlling the hot gas bypass on the basis of return water temperature (as in the upper part of Figure 8.12f), other conventional packages control it on the basis of the opening of prerotation vanes. This is illustrated in the lower portion of Figure 8.12f. The problem with opening the hot gas bypass when the prerotation vane has closed to some fixed point is that this control technique disregards condensing temperature. This causes the hot gas bypass to open sooner and to a greater extent than needed.

This is also illustrated in the lower section of Figure 8.12f, where under normal loads the compressor operates at point "O." As the load drops off at constant discharge pressure (constant condenser temperature), the prerotation vane is gradually closed to its minimum allowable opening, until at point "1" the hot gas bypass is opened, to protect the compressor from going into surge. If the actual load drops to "1S," the HGBP will furnish the differential flow between "1S" and "1."

Now, if the cooling water temperature is reduced, normal operation falls to point "OR." As the load drops, the fixed minimum setting on the prerotation vane is reached at point "2R." This is much sooner than necessary. It is possible to obtain optimized controls that recognize the impact of cooling water temperature variations and in this case will open the HGBP only when load drops to point "1R."

Economizer Controls The economizer shown in Figure 8.12f can increase the efficiency of operation by 5–10%. This is achieved by the reduction of space requirements, savings on compressor power consumption, reduction of condenser and evaporator surfaces, and other effects. The economizer shown in Figure 8.12f is a two-stage expansion valve with two condensate chambers.

When the load is above 10%, the hot gas bypass system is inactive. Condensate is collected in the upper chamber of the economizer, and it is drained under float level control, driven by the condenser pressure. The pressure in the lower chamber floats off the second stage of the compressor, and it, too, is drained into the evaporator under float level control, driven by the pressure of the compressor second stage. Economy is achieved as a result of the vaporization in the lower chamber by precooling the liquid that enters the evaporator and at the same time desuperheating the vapors that are sent to the compressor second stage.

When the load is below 10%, the hot gas bypass is in operation, and the solenoid valve SV-2, which is actuated by the high-pressure switch PSH-2, opens. Therefore, some of the hot gas goes through the evaporator and is cooled by contact with the liquid refrigerant, and some of the hot gas flows through the open solenoid into the economizer. This second portion is desuperheated by the injection of liquid refrigerant upstream of the solenoid, which protects against overheating the compressor.

Safety Interlocks Operating safety in Figure 8.12f is guaranteed by a number of interlocks. The first interlock system prevents the compressor motor from being started when one or more of the below listed conditions exist, and it also stops the compressor if any except the first condition listed below occurs while the compressor is running:

- Suction vane is open, detected by limit switch XLS-1
- Refrigerated water temperature is dangerously low, approaching freezing, as sensed by TSL-6
- Refrigerated water flow is low, measured by FSL-3
- Evaporator temperature has dropped near the freezing point, as detected by TSL-7
- Compressor discharge pressure (and, therefore, pressure in the condenser) is high, indicated by PSH-4
- Temperature of motor bearing or winding is high, detected by TSH-5
- Lubricating oil pressure is low (not shown in Figure 8.12f)

The second interlock system guarantees that the following pieces of equipment are started or are already running upon starting of the compressor:

- Refrigerated water pump (P-1)
- Lubricating oil pump (not shown)
- Water to lubricating oil cooler, if such exists (not shown)

The third interlock usually ensures that the suction vane is completely closed when the compressor is stopped.

CHILLED WATER PIPING CONFIGURATIONS

The proper control and optimization of chilled water distribution systems can only be accomplished if the piping layout of the plant allows for it. Figure 8.12g illustrates both a desirable and an undesirable piping layout.

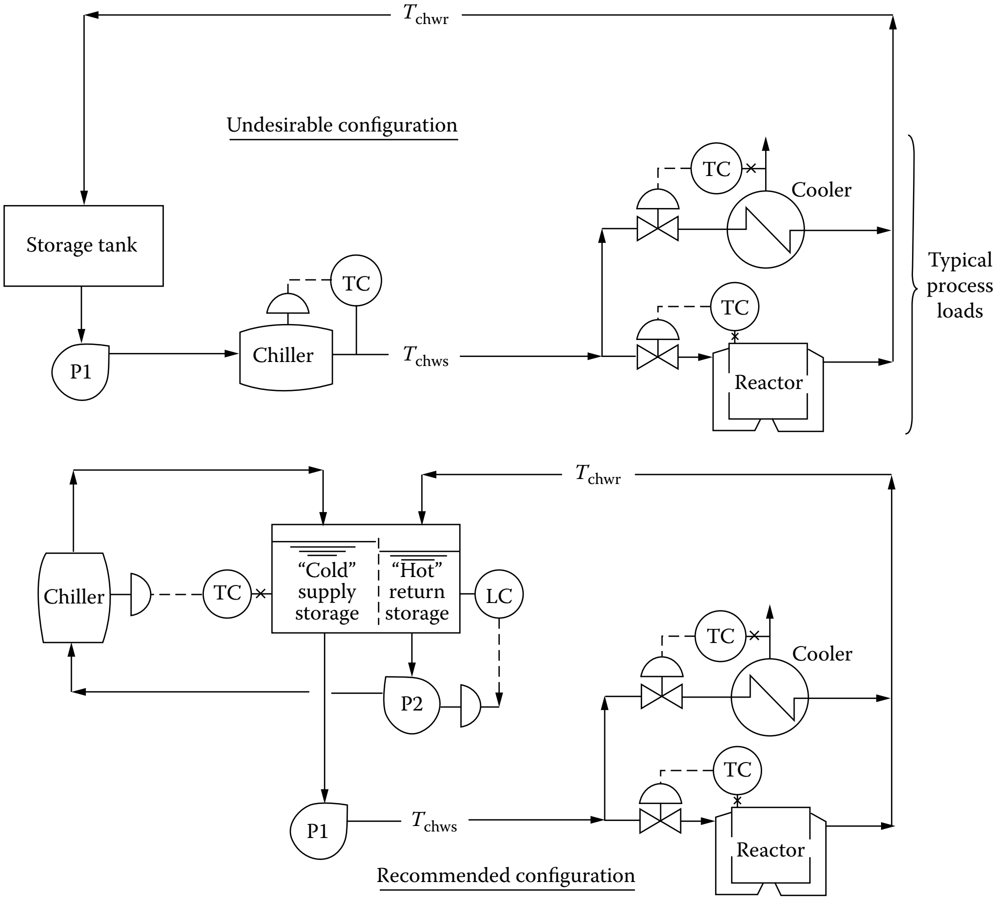


FIG. 8.12g

The proper configuration of the chilled water piping system is an essential element in minimizing upsets. The chiller should not be connected directly to the process loads (top). It is recommended that a storage tank be provided for both the supply of the cold chilled water and the water returning from the process (bottom).

There are two problems associated with the undesirable configuration shown at the top of the figure. One problem is that the layout does not provide chilled water storage, and therefore, if the chiller fails, the plant has to shut down. The other, even more serious problem is that if one of the loads (say the reactor) upsets the return water temperature (T_{chwr}), this in turn will upset the chilled water supply temperature (T_{chws}) to the whole plant.

The reason for this upset can be visualized by assuming that the reactor shown at the top of Figure 8.12g is a batch reactor that has just completed its heat-up phase, and the batch in the reactor is just starting to react. At this point exothermic heat is beginning to be generated, and therefore the jacket is switched from heating to cooling. If the operator is not careful, it is possible that as the chilled water enters the reactor jacket, it will displace the hot water still inside the jacket into the chilled water return line back into the storage tank.

As this slug of several hundred gallons of hot water travels through the storage tank, it will raise T_{chwr}, but the chiller will not know about that until this slug of warm water reaches it. At this point it is already too late, because the chiller can only start to increase its rate of cooling when the T_{chws} has already risen above the set point of the temperature controller (TC).

As a result, the chilled water supply temperature to the whole plant will be upset at each such occurrence. Actually, each of these occurrences will upset the plant's supply temperature twice: once when the slug of warm water arrives at the TC and once again when it had passed and therefore the T_{chws} temperature drops.

Adding Storage

The lower part of Figure 8.12g illustrates how one might eliminate both of these problems if the proper tanks and piping configuration are provided. Following the same sequence of events in this recommended piping layout, the same slug of warm water will not upset the whole plant, because it is received and blended in the "hot" return storage tank. Pump P2 takes the blended hot water from this tank (or tank section) and sends it through the chiller, which discharges it into the "cold" chilled water supply tank.

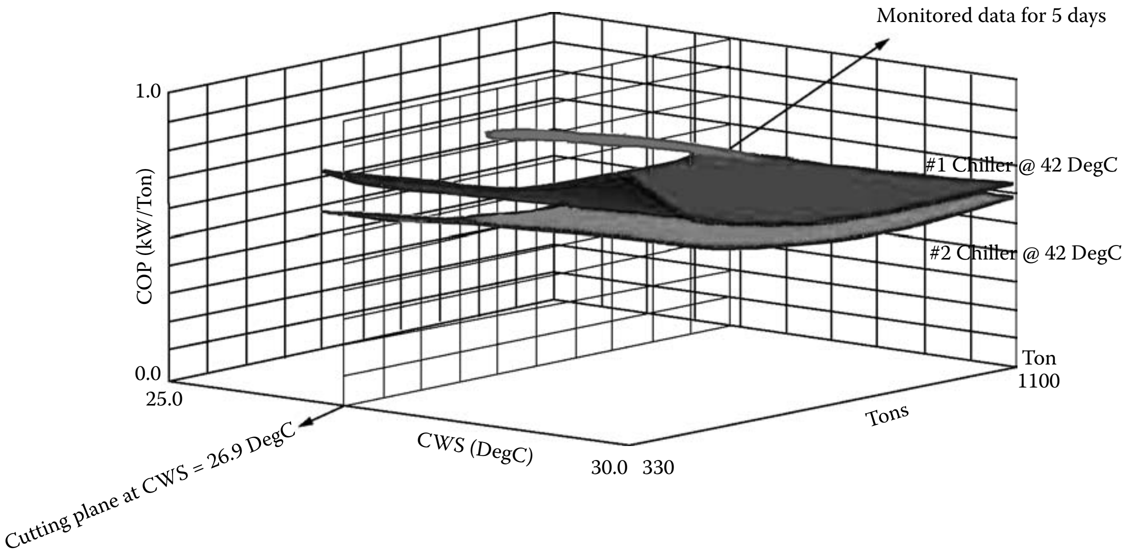


FIG. 8.12h

The differences between the performance of chillers and the changes in the coefficient of performance with load and temperature can be utilized to reduce operating costs.

In this configuration the TC is continuously maintaining the temperature of a fairly large amount of chilled water at the desired T_{chws} temperature. Therefore, the slug of warm water will not upset the supply to the plant at all. Even more important, the plant will be provided with a supply of ready-made chilled water. This safety storage of coolant can be very handy during emergencies, in power failures, or when the chiller is in need of maintenance.

CONCLUSIONS

The control of chillers is well understood. If the piping layout is properly configured and if both the cooling tower controls and the pumping controls are correctly designed, a well designed cooling system can be configured.

Figure 8.12h shows that the performance varies not only with the selection of the particular chiller, but also with the cooling load and with the temperature of the cooling water supply. These variations and other means can be exploited to continuously minimize the cost of meeting the cooling load of a plant. For a discussion of chiller optimization, refer to the next section.

References

1. Lipták, B. G., "Optimizing Controls for Chillers and Heat Pumps," *Chemical Engineering,* October 17, 1983.
2. Cooper, K. W., "Saving Energy with Refrigeration," *ASHRAE Journal,* December 1978.

Bibliography

American Society of Heating, Refrigerating, and Air-Conditioning Engineers, "Chilled and Dual-Temperature Water System," in *ASHRAE Handbook, Equipment Volume,* Atlanta, GA: ASHRAE, 1998.

Blevins, T. L., "Advanced Control Unleashed," *ISA,* 2003.

Carrier System Design Manual, Syracuse, NY: Carrier Corporation,

Cooper, K. W., and Erth, R. A., "Centrifugal Water Chilling Systems: Focus on Off-Design Performance," *Heating/Piping/Air Conditioning,* January 1978, p. 63.

Frayne, C., *Cooling Water Treatment*, New York: Chemical Publishing Co., January 2001.

Hartman, T., "Ultra-Efficient Cooling with Demand-Based Control," *HPAC Engineering,* December 2001.

Lipták, B. G., "Control of Refrigeration Units," *Instrumentation in the Processing Industries,* Section 10.5, Philadelphia: Chilton, 1972, p. 616.

Lipták, B. G., "Reducing the Operating Costs of Buildings," *ASHRAE Transactions,* Vol. 83, Part 1, 1977.

McMillan, G., "Good Tuning: A Pocket Guide," *ISA,* June 2000.

Null, H. R., "Heat Pumps in Distillation," *Chemical Engineering Progress,* July 1976.

Polimeros, G., "There Is Energy for Use in Refrigeration Waste Heat," *Specifying Engineer,* January 1980.

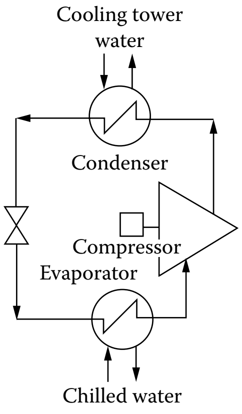
Shinskey, F. G., "Energy Conservation," *Instruments and Control Systems,* January 1977.

Stair, W. S., "Air Conditioning System," U.S. Patent Numbers 2,715,514 and 2,715,515 issued August 16, 1955.

Whitt, M. D., "Successful Instrumentation and Control Design," *ISA,* 2003.

8.13 Chiller Optimization

B. G. LIPTÁK (1970, 1985, 1995, 2005)



Flow sheet symbol

INTRODUCTION

There are not many unit operations whose efficiency can be doubled through optimization. The chillers are one of these few. One unit of energy introduced at the chiller can remove 5, 6, or even 7 units of heat energy from the process, if the chiller controls are optimized. The previous section discussed the conventional chiller controls, while this section is devoted to chiller optimization.

Listed below are some of the performance characteristics of optimized chillers as contrasted to conventional ones:

- Coefficients of performance: 2.5–3.5 for conventional, 5–7 for optimized chillers.
- Operating cost reduction: 1.5% of yearly chiller operating cost per °F reduction in chiller ΔT, 1% of total cooling system operating cost per °F decrease in the cooling tower's approach (see Figure 8.13b for the definition of "approach"). A 10°F reduction in cooling tower water temperature cuts chiller operating costs by 30%.
- Pumping cost savings: 50–60 cents/gpm for each psid.
- Operating cost distribution: 60% chiller compressor, 15% ea. cooling tower and chilled water pumps, 10% cooling tower fans.
- Maximum turndown ratios: 8:1 for conventional, over 30:1 with hot gas bypass.
- Typical superheat at evaporator outlet: 9°F (5°C).

In order to fully optimize a cooling system, in addition to the chiller, it is necessary to also optimize cooling tower controls, the water distribution, and the pumping systems, as will be discussed in the sections that follow.

THE TOTAL COOLING SYSTEM

The piping layout of a conventional cooling system is illustrated in Figure 8.13a. In order to develop a completely generalized method for controlling and optimizing such systems, the duplication of equipment will be disregarded, and all chiller systems will be treated as if they were configured as shown in Figure 8.13b. Here any number of cooling towers, pumps, or chillers are represented by single units, because variations in their numbers will not affect the overall optimization strategy.

In this generalized cooling system, the first heat transfer step is when the cooling load from the process is transferred to the chilled water. In the second step, the heat carried by the chilled water is transferred to the refrigerant in the evaporator. The refrigerant takes the heat to the condenser, where in the third step the heat is transferred to the cooling tower water, so that it might finally be rejected to the ambient air.

This heat pump operation involves four heat transfer substances (chilled waster, freon or other refrigerant, cooling tower water, and air) and four heat exchanger devices (process heat exchanger, evaporator, condenser, cooling tower). The total system operating cost is the sum of the costs of circulating the four heat transfer substances ($M1$, $M2$, $M3$, and $M4$ in Figure 8.13b).

In the traditional (unoptimized) control systems, such as the one illustrated in Figure 8.12f, each of these four systems were operated independently in an uncoordinated manner. In addition, the conventional control systems did not vary the speed of the four transportation devices ($M1$ to $M4$). By operating them at constant speeds, they introduced more energy than was needed for the circulation of refrigerant, air, or water and therefore they had to waste that excess energy.

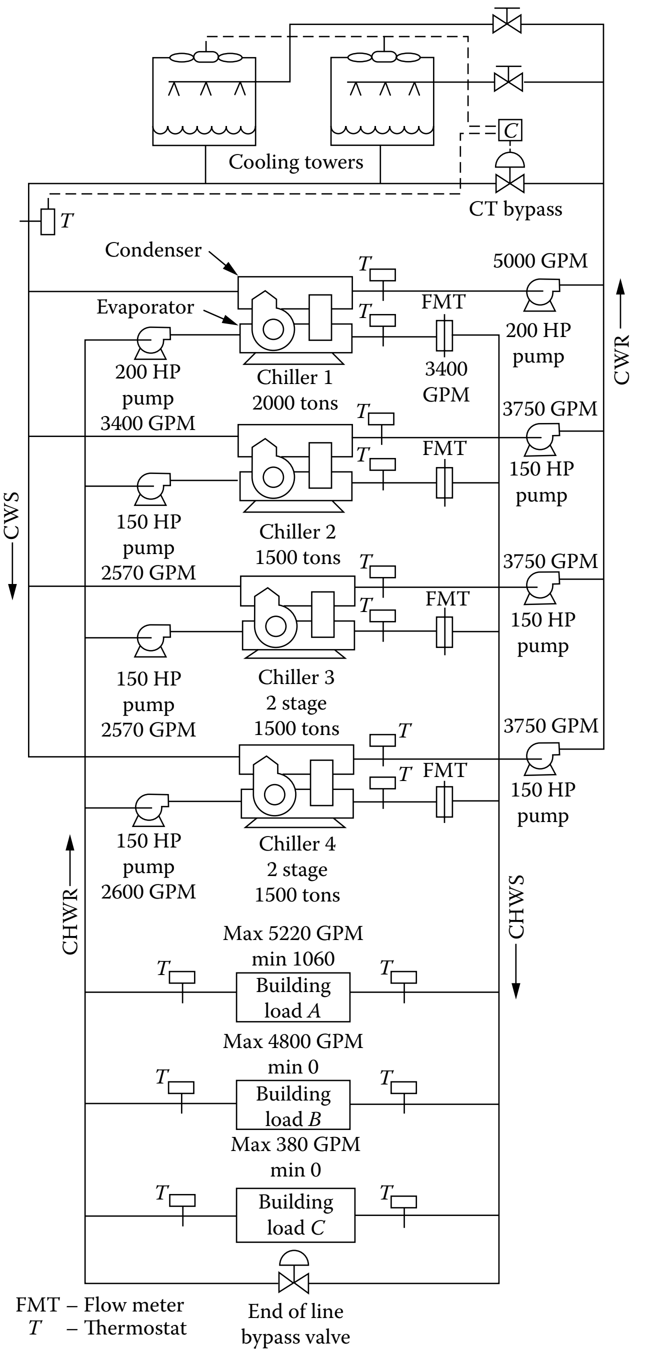
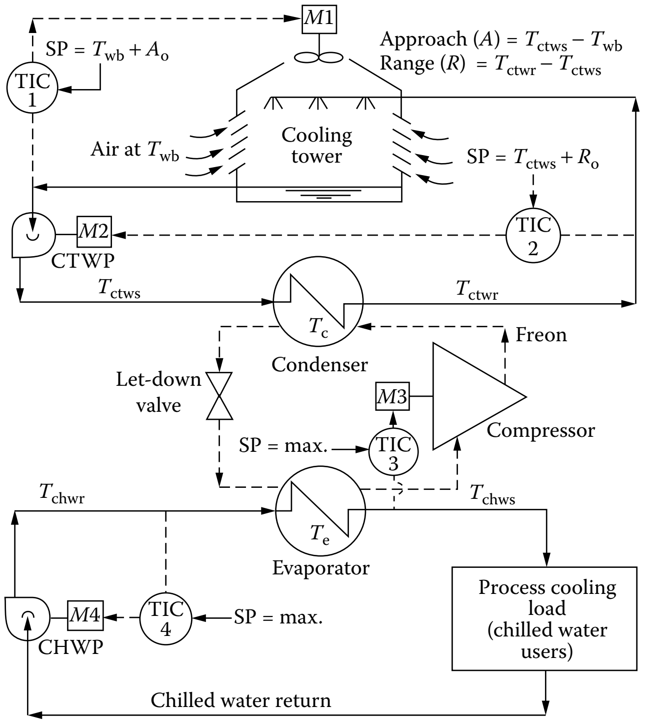


FIG. 8.13a

A typical overall cooling system utilizes four heat transfer substances (CWS, CWR, CHWS, CHWR) and four heat exchange devices (cooling towers, chiller condenser, chiller evaporator, process heat exchangers).[1]



Abbreviations:

A_o : Optimum approach
R_o : Optimum range
T_{wb} : Wet bulb temp.
T_{ctws} : Cooling tower water supply temp.
T_{ctwr} : Cooling tower water return temp.
T_{chws} : Chilled water supply temp.
T_{chwr} : Chilled water return temp.
CTWP: Cooling tower water pump
CHWP: Chilled water pump
T_c : Freon temperature in condenser
T_e : Freon temperature in evaporator

FIG. 8.13b

In order to optimize a cooling system, the cooling tower(s), the pumping stations, the chiller(s), and the process equipment should be treated as an integrated single system.[2]

OPTIMIZING MECHANICAL REFRIGERATION SYSTEMS

Load-following optimization eliminates the waste resulting from constant speed operation by operating the aforementioned four systems as a coordinated single process, with the goal to keep the cost of operation at a minimum. In this control system, the controlled variables are the supply and return temperatures of chilled and cooling tower waters, and the manipulated variables are the flow rates of chilled water, refrigerant, cooling tower water, and air.

If water temperatures are allowed to float in response to load and ambient temperature variations, the waste associated with keeping them at arbitrarily selected fixed values is eliminated, and the operating cost of the cooling system is drastically reduced.

In order to control the total refrigeration system depicted in Figure 8.13b, four control loops must be configured. In these four loops the controlled (measured) variables are the four water temperatures, and the four manipulated variables are the four motor speeds ($M1$ to $M4$) that drive the four transportation devices. Table 8.13c lists the controlled and the manipulated variables in each of these loops and also indicates the optimization criteria for determining the set points of the four controllers.

TABLE 8.13c
Control Configuration and Optimization Criteria for the Four Loops Controlling a Chiller

Loop No.	Controlled Variable	Manipulated Variable	Optimization Criteria (Set Points of the Temperature Controllers)
1	CT water supply temperature (T_{ctws})	Air flowrate obtained by level of fan operation ($M1$)	Optimum approach (A_o) is selected to keep the sum $M1 + M2 + M3$ to a minimum. Therefore, TIC set point becomes: $T_{\text{ctws}} = T_{\text{wb}} + A_o$
2	CT water return temperature (T_{ctwr})	Rate of CT water pumping ($M2$)	The optimum range (R_o) of CT is found as a function of the A_0. Therefore, the TIC set point becomes: $T_{\text{ctwr}} = T_{\text{ctws}} + R_o$
3	Chilled water supply temperature (T_{chws})	Rate of chiller compressor operation ($M3$)	The optimum TIC set point is the maximum T_{chws} temperature, which will satisfy *all* the process loads
4	Chilled water return temperature (T_{chwr})	Rate of chilled water pumping ($M4$)	The optimum TIC set point is the maximum T_{chwr} temperature that will satisfy all the process loads

Minimizing the Operating Cost

The yearly cost of operating the total cooling system can typically be broken down as follows:

$M1$ (Fans)	10%
$M2$ (CT Pumps)	15%
$M3$ (Compressors)	60%
$M4$ (CH Pumps)	15%
TOTAL:	100%

These costs vary a great deal. The proportion of $M1$ increases in warm weather regions. $M2$ and $M4$ increase when water transport lines are long and the proportion of $M3$ is lowered as the maximum allowable chilled water temperature rises. But regardless of these proportions in a particular installation, the goal of optimization is to find the minimum chilled water and cooling tower water temperatures that will result in meeting the needs of the process at minimum cost.

The overall control system can be reviewed in two steps: First one can look at the chilled water side (the evaporator side), then at the cooling tower water side (condenser side).

The Chilled Water Side The optimization of the lower (evaporator) portion of Figure 8.13b is easily comprehended, because the sum of $M3$ and $M4$ will be the minimum when both the chilled water supply and return temperatures (T_{chws} and T_{chwr}) are as high as the process will permit. This is true under all load conditions, because the amount of work that the chiller compressor has to do is reduced as the suction pressure rises, and this is the case whenever the evaporator temperature (T_{chws}) rises.

Therefore, the optimum set point of TIC-3 is the allowable maximum chilled water supply temperature. Consequently, the minimum cost of operation will be achieved when the chilled water supply temperature has been maximized, but it can be maximized only to the point where it can still provide all the cooling that is required by any part of the process. The method of finding that maximum value will be discussed later.

The chilled water pumping cost ($M4$) depends only on the ΔT across the process users. The higher this ΔT, the less water needs to be pumped to remove the required amount of heat. As the chilled water supply temperature is set by the load, this ΔT can be maximized only by maximizing the chilled water return temperature (T_{chwr}). Therefore, the optimum set point for TIC-4 is the allowable maximum. By increasing this return temperature, the following benefits result:

1. The power consumed by $M3$ is lowered by 1.5% per °F rise.
2. The heat transfer efficiency of the evaporator is improved by increasing the ΔT through it.
3. Assuming that on the average the ΔT is 15°F, increasing it by 1°F will lower $M4$ by about 6%.

As $M3$ is about 60% and $M4$ about 15% of the total operating cost, a 1°F decrease in ΔT will lower the yearly cost of operation by about $(1 \times 60)/100 + (6 \times 15)/100 = 1.5\%$.

The Cooling Tower Water Side Finding the optimum cooling tower water temperatures is more complicated, because changing the cooling tower water supply and return temperatures will increase some costs, while it will also lower some others. For example (as it is shown in Figure 8.17d), an increase in the cooling tower water supply temperature (T_{ctws}) will increase the tower's approach and the cost of chiller operation but will reduce the amount of work the cooling tower fans need to do.

Therefore, the optimum temperature in not necessarily the minimum temperature (or approach) that the towers are capable of providing, but rather the temperature that can meet the particular load at the minimum total cost of operation of all the equipment ($M1 + M2 + M3$). Figure 8.13d illustrates that this is a function of the load and of the weather. When the conditions are such that a lowering of the approach (or T_{ctws}) would require a large increase in fan operating costs ($M1''$), the minimum point on the total cost curve (A_0'') will be above the minimum attainable approach (A_0').

Inversely, if the approach can be reduced with a small increase in fan operating costs ($M1'$), the optimum operating point for the total system will correspond to the minimum T_{ctws} temperature and therefore to the minimum attainable approach (A_0'). Whichever is the applicable total cost curve

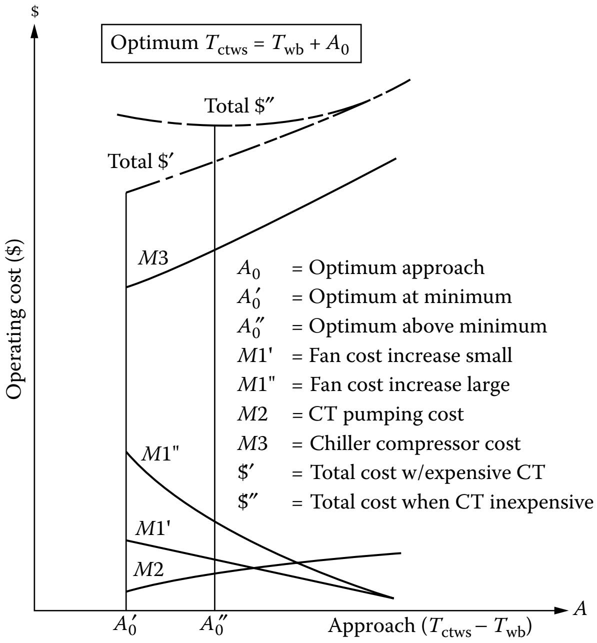


FIG. 8.13d
The empirically determined operating costs are shown as a function of the approach at two cooling loads. The optimum approach at each load is the one that corresponds to the minimum total cost of operation.

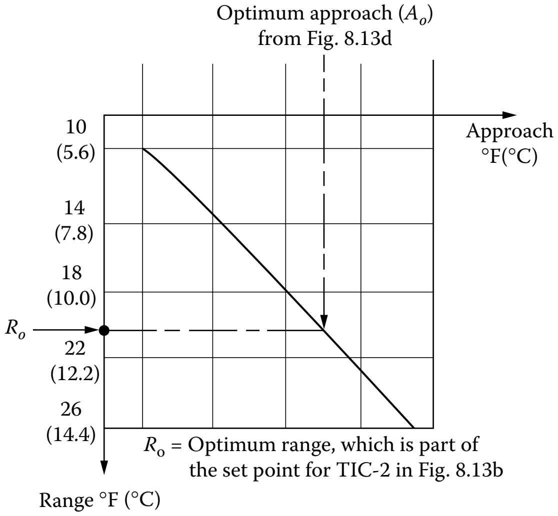


FIG. 8.13e
The optimum range for a particular cooling tower is a function of the approach at which it is operating. Therefore, R_0 can be found on the basis of A_0.

($′ or $″), the T_{ctws} temperature that corresponds to the minimum point becomes the set point for TIC-1 in Figure 8.13b.

The curves in Figure 8.13d can be based on the measurement of the actual total operating cost, on a projected cost based on past performance, or on some combination of the two. The continuous storing and updating of the operating cost history as a function of load, ambient conditions, and equipment configuration will not only provide the curves required for optimization but can also be used to signal the need for maintenance.

In Figure 8.13d the optimum T_{ctws} is found by summing $M1$, $M2$, and $M3$. In most installations the major cost element is $M3$, which increases by about 1.5% of compressor operating cost (or 1% of total cooling system cost) for each 1°F reduction in approach.

The cost-vs.-approach curve is not necessarily a smooth one, because if constant-speed machines are stopped and started or if positive-displacement compressors are loaded and unloaded, there will be steps in the $M3$ curve. This also is true for constant speed fans ($M1$) or pumps ($M2$). The pumping cost also increases as approach rises, but $M2$ is usually only a fraction of $M3$, unless the piping is very long or undersized.

The fan cost $M1$ tends to drop off as approach rises. If the load is low relative to the size of the cooling towers, the rate of increase in $M1$ with a reduction in approach will also be small. In such cases ($M1'$) the optimum approach (A_0') corresponds to the safe lowest temperature that the tower can generate. If, on the other hand, the load is high, the fan cost of lowering approach will also increase ($M1''$). In that case, the optimum approach (A_0') is at some intermediate value.

The optimum return water temperature back to the cooling tower (T_{ctwr}) can be found by determining the optimum range (R_o) for the particular set of operating conditions. The relationship between operating approach and range is a function of the particular cooling tower design. Therefore, if one has determined the optimum approach for a particular set of conditions A_o (from Figure 8.13d), that value can be used to arrive at the optimum range (R_o) in Figure 8.13e.

The set point for TIC-2 in Figure 8.13b then becomes the sum of the already-determined T_{ctws} and the optimum range:

$$T_{\text{ctwr}} = T_{\text{ctws}} + R_o \qquad \textbf{8.13(1)}$$

Chilled Water Supply Temperature Optimization

The yearly operating cost of a chiller is reduced by about 1.5% for each 1°F (0.6°C) reduction in the temperature difference across this heat pump. In order to minimize this difference, T_H must be minimized and T_L maximized. Therefore, the optimum value of T_{chws} is the maximum temperature that will still satisfy all the loads. Figure 8.13f illustrates a method of continuously finding and maintaining this maximum value.

It should be noted that an energy-efficient refrigeration system cannot be provided by instrumentation alone. Equipment sizing and selection also play an important part in the overall result. For example, the evaporator heat transfer area should be maximized so that T_L is as close as possible to the average chilled water temperature in the evaporator.

Similarly, the refrigerant flow should be made to match the load by adjusting the compressor motor speed and not by refrigerant throttling. TCV-1 and TCV-2 in Figure 8.12f in the previous section represent sources of energy waste, whereas

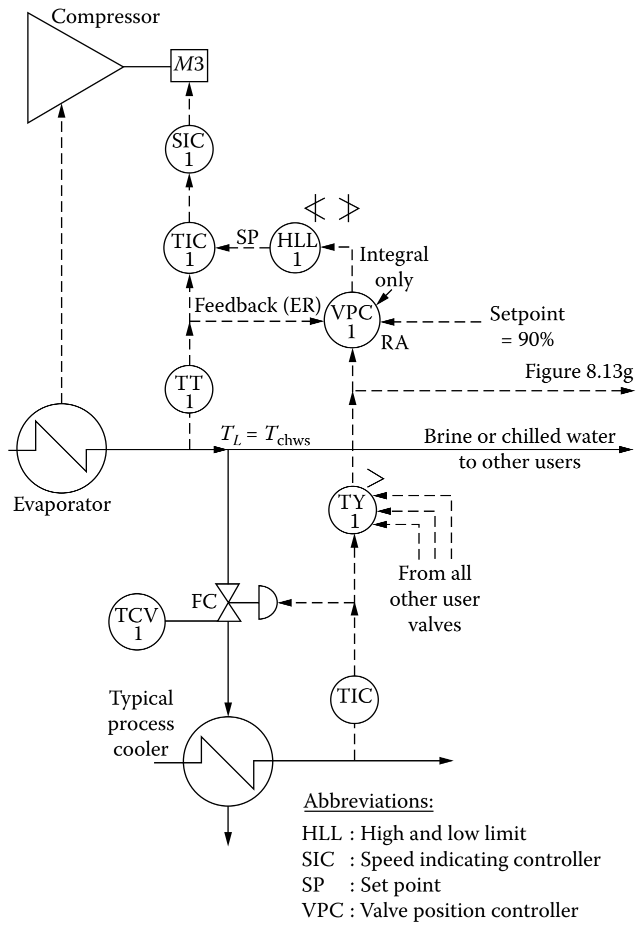


FIG. 8.13f

Maximizing the chilled water supply temperature by load-following floating control is a means of continuously finding and maintaining this temperature at its optimum value.

Figure 8.13f shows the energy-efficient technique of motor speed control by eliminating TCV-2. This load-following mode of operation can be achieved by using variable-speed drives on electric motors or by using steam turbine drives.

If several constant-speed motors are used, then all compressors except one should be driven to the loading level at which they are most efficient (TCV-1 and TCV-2 in Figure 8.12f fully open). The remaining one compressor should be used to match the required load by throttling.

Maximizing T_{chws} Figure 8.13f shows the proper technique of maximizing the chilled water supply temperature in a load-following, floating manner. The optimization control loop guarantees that all chilled water users in the plant will always be satisfied while the chilled water temperature is maximized. This is done by selecting (by TY-1) the most open chilled water valve in the plant and comparing that opening with the 90% set point of the valve position controller, VPC-1. If even the most open valve is less than 90% open, the set point of TIC-1 is increased; if the valve opening exceeds 90%, the TIC-1 set point is decreased.

This technique allows all the process users to obtain more cooling (by further opening their chilled water supply valves) if needed, while the header temperature is continuously maximized. The VPC-1 set point of 90% is adjustable. Lowering it gives a wider safety margin, which might be required if some of the processes that are being cooled are very critical. On the other hand, increasing the set point will maximize energy conservation at the expense of a reduced safety margin.

An additional benefit of this load-following optimization strategy is that because all chilled water valves in the plant are opened as T_{chws} is maximized, valve cycling is reduced and pumping costs are lowered. This is so because when all chilled water valves are opened they require less pressure drop and therefore less discharge pressure from the pumps. Similarly, valve cycling is eliminated, because by increasing the opening of all user valves, they are moved away from the unstable region near their closed position.[2]

In order for the control system in Figure 8.13f to be stable, it is necessary to use an integral-only controller for VPC-1, with an integral time that is tenfold that of the integral setting of TIC-1 (usually several minutes). This control mode selection is needed to allow the optimization loop to be stable when the valve opening signal selected by TY-1 is either cycling or noisy.

The high/low limits (HLL-1) on the set-point signal to TIC-1 guarantee that VPC-1 will not drive the chilled water temperature to unsafe or undesirable levels. Because these limits can block the VPC-1 output from affecting T_{chws}, it is necessary to protect VPC-1 against reset windup. This is done through the external feedback signal shown in Figure 8.13f.

Chilled Water Return Temperature Optimization

The combined cost of operating the chilled water pumps and the chiller compressor ($M4 + M3$) is a function of the temperature drop across the evaporator ($T_{chwr} - T_{chws}$). Because an increase in this ΔT decreases compressor operating costs (suction pressure rises) while also decreasing pumping costs (the higher the ΔT the less water needs to be pumped), the aim of this optimization strategy is to maximize this ΔT.

Maximizing T_{chwr} This ΔT will be the maximum when the chilled water flow rate across the chilled water users is the minimum. As T_{chws} is already controlled, this ΔT will be maximized when T_{chwr} is maximum. This goal can be reached by evaluating the opening of the most-open chilled water valve in Figure 8.13f and, if even the most-open chilled water valve is not yet fully open, making adjustments to further open it.

The steps that can be taken include increasing the chilled water supply temperature (set point of TIC-1 in Figure 8.13f), or increasing the temperature rise across the process users by lowering the ΔP across the users (set point of PDIC-1 in Figure 8.13g). Both methods can also be used simultaneously.

Increasing the chilled water supply temperature reduces the yearly compressor operating cost ($M3$) by approximately 1.5% for each °F of temperature increase, whereas lowering

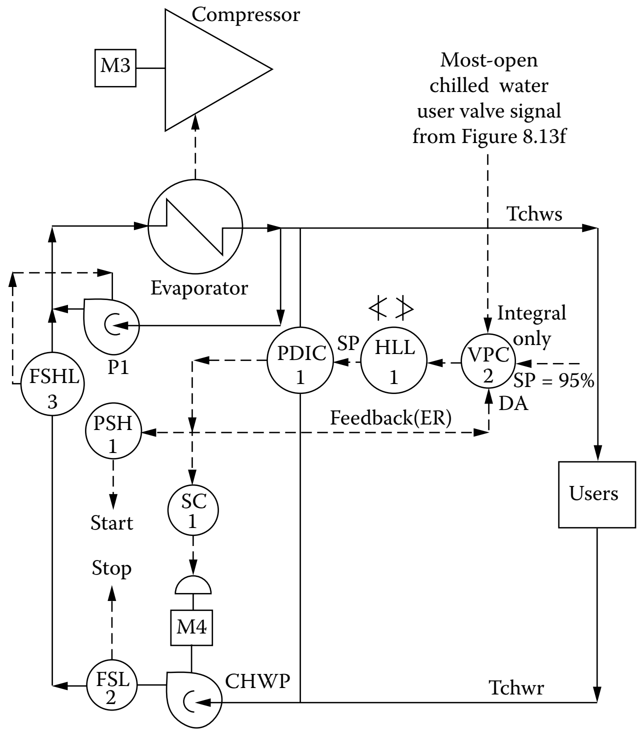


FIG. 8.13g

Adjusting the chilled water flow rate such that the most open user valve is nearly full will keep the evaporator ΔT and chilled water return temperature always at their maximum (optimum) values.

the ΔP across the users reduces the yearly pump operating cost ($M4$) by approximately 50 cents/gpm for each psid.

The set points of the two valve position controllers (VPC-1 in Figures 8.13f and VPC-2 in 8.13g) will determine if these adjustments are to occur in sequence or simultaneously. If both set points are adjusted to be the same, simultaneous action will result, while if one adjustment is economically more advantageous or safer than the other, the set point of the corresponding VPC should be set lower than the other.

This will result in sequencing, which means that the more cost-effective or safer correction will take place before the less effective one is started. In Figures 8.13f and 8.13g, it was assumed that increasing T_{chws} is the more cost-effective step. This is not always the case, and even when it is, there might be process reasons that make it undesirable to allow the floating of T_{chws} up or down.

If the settings are as shown, the system will function as follows: If the most open valve is less than 90% open, VPC-2 in Figure 8.13g will lower the set point of PDIC-1, and VPC-1 in Figure 8.13f will increase the set point of TIC-1. When the opening of the most-open valve reaches 90%, VPC-1 in Figure 8.13f will slowly start lowering the TIC set point, while VPC-2 in Figure 8.13g will continue to lower the set point of PDIC-1.

If VPC-1 in Figure 8.13f did not lower the TIC set point sufficiently (or if the lowering was not fast enough) and the most-open user valve continues to open, and at 95% the VPC-2 in Figure 8.13g will take fast corrective action by quickly raising the set point of PDIC-1. Thus, no user valve will ever be allowed to open fully and go out of control as long as the pumps and chillers are sized so as to be capable of meeting the load.

Pump Speed Control VPC-2 in Figure 8.13g is the cascade master of PDIC-1, which guarantees that the pressure difference between the chilled water supply and return is always high enough to motivate water flow through the users but never so high as to exceed their pressure ratings. The high and low limits are set on HLL-1, and VPC-2 is free to float this set point within these limits to keep the operating cost at a minimum.

In order to protect against reset windup when the output of VPC-2 reaches one of these limits, an external feedback is provided from the PDIC-1 output signal back to VPC-2. The VPC is an integral-only controller, which is tuned to be much more responsive (shorter integral time) than is VPC-1 in Figure 8.13f.

The high speed of response of the VPC in Figure 8.13g is also important from a safety point of view. This is because without it, in a conventional control system, the response can be too slow. In such a case, when the demand for cooling suddenly increases (because of some process difficulty), once the valve (TCV-1 in Figure 8.13f) is fully open, the amount of cooling provided cannot be increased faster than the rate at which the chiller can lower the chilled water supply temperature. This does take time, and therefore, if the cooling load evolves faster (runaway reaction in a reactor), accidents can occur.

This is not the case when the control system shown in Figure 8.13g is implemented, because as soon as the most-open chilled water valve has opened to 95%, VPC-2 will quickly speed up the chilled water pump (CHWP). Because water is incompressible, this will immediately increase the chilled water supply pressure in the main supply header, therefore, the water supply pressure to TCV-1 in Figure 8.13f will also rise, thereby immediately increasing the water flow and therefore the amount of cooling of the process.

This practically instantaneous action, utilizing the already-chilled water stored in the distribution headers, can be fast enough to arrest a runaway reaction. Naturally, once the chiller has had time to respond to the increase in load by lowering the chilled water supply temperature, this emergency action will no longer be needed and the water supply pressure will be lowered back to its normal setting.

Multiple and Bypass Pumps When the chilled water pump station consists of several pumps, only one of which is variable-speed, additional pump increments are started when PSH-1 in Figure 8.13g signals that the pump speed controller set point is at its maximum. When the load is dropping, the excess pump increments are stopped on the basis of flow, detected by FSL-2. In order to eliminate cycling, the excess pump increment is turned off only when the actual total flow has dropped to less than 90% of the capacity of the remaining pumps.

The load-following optimization loop in Figure 8.13f and 8.13g will float the total chilled water flow to achieve maximum overall economy. In order to maintain efficient heat transfer and the required turbulence within the evaporator, a small local circulating pump (*P*1) is provided at the evaporator. This pump is started and stopped by FSHL-3 in Figure 8.13g, guaranteeing that the water velocity in the evaporator tubes will never drop below the adjustable limit of about 4 fps (1.2 m/s).

Cooling Tower Supply Temperature Optimization

As discussed in more detail in Section 8.17, minimizing the temperature of the cooling tower water is one of the most effective contributors to chiller optimization. Conventional control systems in the past produced constant cooling tower temperatures of 75°F (23.9°C) or higher.

It is an enemy of efficiency and therefore of optimization to maintain a utility at a constant value. Each 10°F (5.6°C) reduction in the cooling tower water temperature will reduce the yearly operating cost of the compressor by approximately 15%. For example, if a compressor is operating at 50°F (10°C) condenser water instead of 85°F (29.4°C), it will meet the same load while consuming half as much power. Operation at condenser water temperatures of 50°F (10°C) or even less can be quite practical during the winter months. Savings exceeding 50% have been reported.[3]

Minimizing T_{ctws} As shown in Figure 8.13h, an optimization control loop is required in order to maintain the cooling tower water supply continuously at an economical minimum temperature. This minimum temperature is a function of the wet-bulb temperature of the atmospheric air. The cooling tower cannot generate a water temperature that is as low as the ambient wet bulb, but can approach it. The temperature difference between T_{ctws} and T_{wb} is called the *approach*, as was shown in Figure 8.13d.

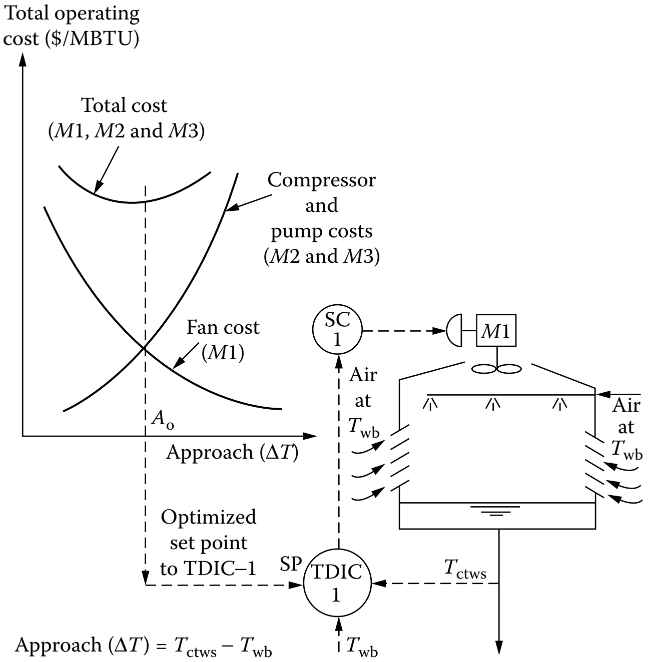


FIG. 8.13h
Optimizing (minimizing) the cooling tower water supply temperature requires that the fan speed be so modulated as to keep the approach ($T_{ctws} - T_{wb}$) at a value that corresponds to the minimum cost of operation.

Figure 8.13h illustrates the fact that as the approach and the cooling tower supply temperature increases, the cost of cooling tower fan operation drops, while the costs of pumping and of compressor operation increase. Therefore, the total operating cost curve has a minimum point that identifies the optimum approach, which corresponds to the minimum cooling tower temperature that will allow operation at an overall minimum cost. This approach (ΔT) automatically becomes the set point of TDIC-1, and this optimum approach will increase if the load on the cooling tower rises or if the ambient wet bulb increases.

If the cooling tower fans are centrifugal units or if the blade pitch is variable, the optimum approach is maintained by continuous throttling. If the tower fans are two-speed or single-speed units, the output of TDIC-1 will start and stop the fan units incrementally in order to maintain the optimum approach. In cases in which a large number of cooling tower cells constitute the total system, it is also desirable to balance the water flows to the various cells automatically as a function of the operation of the associated fans.

The water flows to all cells whose fans are at high speed should be controlled at equal and relatively high rates, while the cells with fans operating at low speeds should receive water at equal low flow rates. Cells with their fans off should be supplied with water at equal minimum flow rates. Subjects such as these and the cooling towers in general are discussed in more detail in Sections 8.16 and 8.17.

Cooling Tower Return Temperature Optimization

The optimum setting for T_{ctwr} is based on the optimum approach (A_o) obtained from Figure 8.13h, using the characteristic curve of the cooling tower depicted in Figure 8.13e. This curve relates the approach to the range for the particular tower design, and therefore once A_o is known, R_o can be obtained from it.

As shown in Figure 8.13i, the ΔT corresponding to this optimum R_o automatically becomes the set point of TDIC-1 in the optimized control loop. This controller is the cascade master of PDIC-1, which guarantees that the pressure difference between the cooling tower supply and return water flows is always high enough to provide the required flow through the process users but never so high as to cause damage. The high and low limits are set on HLL-1. TDIC-1 floats the PDIC-1 set point within these ΔP limits, to keep the operating cost at a minimum.

In order to protect against reset windup (when the output of TDIC-1 reaches one of these limits), an external feedback is provided from the output of PDIC-1 to TDIC-1.

When the cooling tower water pump station consists of several pumps, only one of which is variable-speed, additional

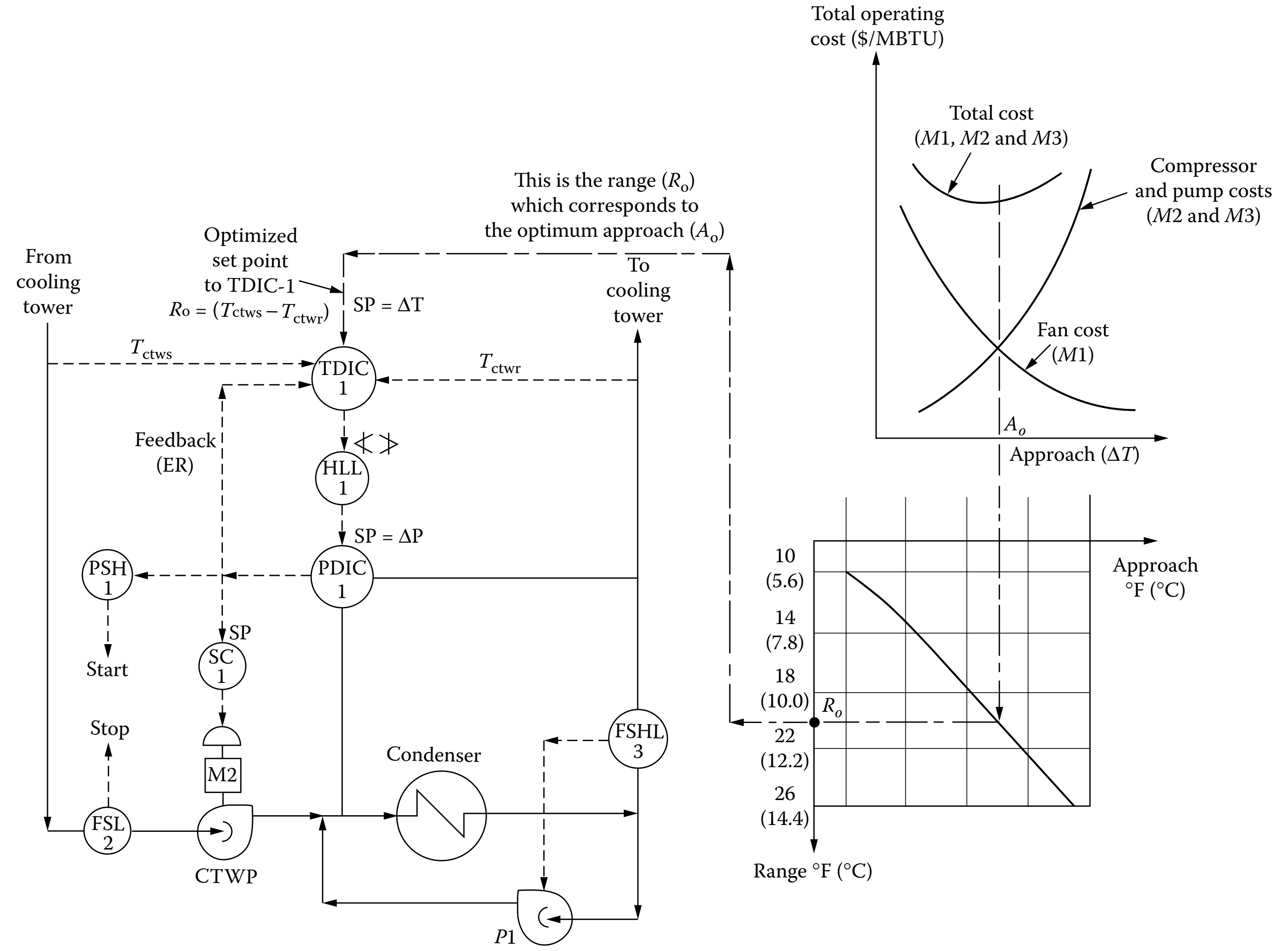


FIG. 8.13i

The cooling tower water flow rate is modulated to keep the ΔT of the chiller's condenser (R_0, the range of the cooling tower) at the value that corresponds to the optimum value of the approach.

pump increments are started when PSH-1 signals that the pump speed controller set point is at its maximum. When the load is dropping, the excess pump increments are stopped on the basis of *flow,* which is detected by FSL-2. In order to eliminate cycling, the excess pump increment is only turned off when the actual total flow is less than 90% of the capacity of the remaining pump(s).

The load-following optimization loop shown in Figure 8.13i will float the total cooling tower water flow to achieve maximum overall economy. In order to maintain efficient heat transfer and appropriate turbulence within the condenser, a small local circulating pump ($P1$) is provided at the chiller's condenser. This pump is started and stopped by FSHL-3, guaranteeing that the water velocity in the condenser tubes will never drop below the adjustable limit of 4 fps (1.2 m/s).

Heat Recovery Optimization

Figure 8.13j depicts the optimizing control loop required when the heat pumped by a chiller is recovered in the form of hot water and the hot water temperature is continuously optimized in a load-following floating manner.

If, at a particular load level, it is sufficient to operate with 100°F (37.8°C) instead of 120°F (48.9°C) temperature hot water, this technique will allow the chiller to meet the same tonnage of refrigeration load at 30% lower operating cost. The reason for this operating cost reduction is that the required discharge pressure of the chiller compressor depends on the hot water temperature in the split condenser.

The optimization control loop in Figure 8.13j guarantees that all hot water users in the plant will always obtain enough heat while the hot water temperature set point of TIC-1 is minimized. In this system TY-1 selects the most open hot water valve in the plant, and VPC-1 compares that with its 90% set point. If even the most open valve is less than 90% open, the set point of TIC-1 is decreased, and if this opening exceeds 90%, the TIC-1 set point is increased. This allows all users to obtain more heat (by further opening their supply valves) if needed, while the header temperature is continuously optimized (minimized).

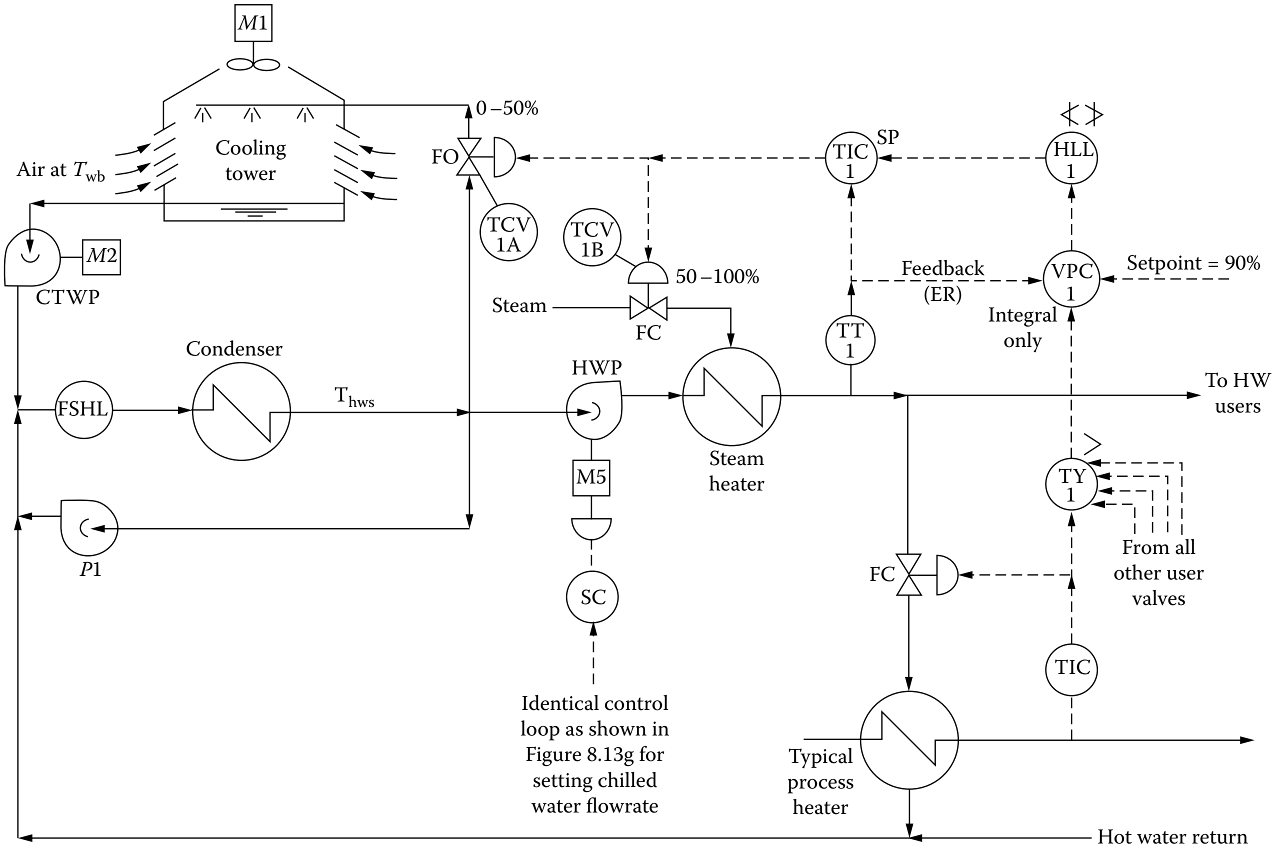


FIG. 8.13j

When the heat pumped by a chiller is recovered as hot water, a load-following cascade master controller (VPC-1) can be used to set the hot water temperature controller's (TIC-1) set point.

Figure 8.13j also shows that an increasing demand for heat will cause the TIC-1 output to rise. An increase in the heat load of the plant will cause a decrease in the heat spill to the cooling tower through TCV-1A when the TIC-1 output is between 0 and 50%. At 50% (0.6 bar) output, all the available cooling load is being recovered and TCV-1A is fully closed. If the heat load continues to rise (TIC-1 output signal rises over 50%), this will result in the partial opening of the "pay heat" valve, TCV-1B. In this mode of operation, the steam heat is used to supplement the freely available recovered heat to meet the prevailing heat load of the plant.

A local circulating pump, P1, is started by FSHL whenever the flow velocity is low. This prevents the formation of deposit in the condenser tubes. P1 is a small 10–15 hp pump that is operating only when the flow is low. The main cooling tower pump (usually larger than 100 hp) is stopped when TCV-1A is closed.

OPTIMIZATION BY OPERATING MODE SELECTION

The cost-effectiveness of heat recovery is a function not only of the outdoor temperature, but also of the unit cost of energy from the alternative heat source and the percentage of the cooling load that can be used in the form of recovered heat.

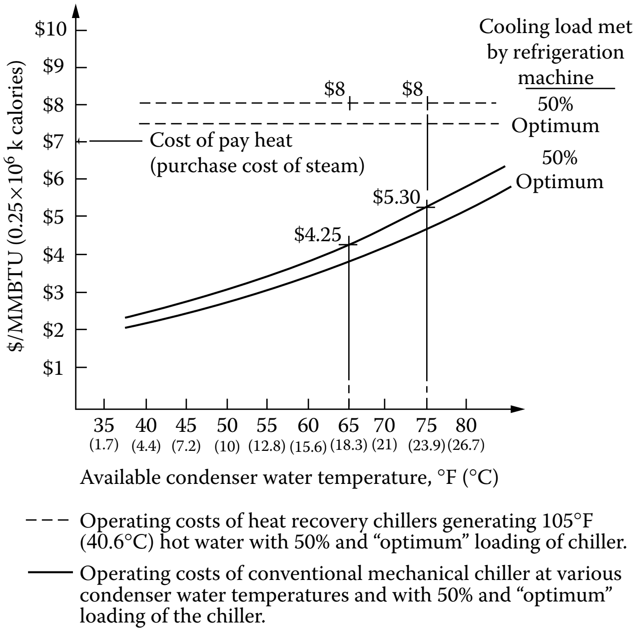
When the outside air temperature is below 65°F (18°C) and steam is available at $7/MMBTU while only half of the cooling load is required in the form of hot water, it is more cost-effective to operate the chiller in the regular cooling mode than in heat-recovery. In that case, according to the data in Figure 8.13k, it is cost-effective to use steam as the heat source.

Conversely, when the outdoor temperature is above 75°F (23.9°C), the penalty for operating the split condenser at hot water temperatures is no longer excessive; therefore, the plant should automatically switch back to the heat-recovery mode of operation. This cost-benefit analysis can be a continually used and automatic element of the overall optimization scheme.

Artificial Heat Sources

In locations such as the southern United States, where there is no alternative heat source, another problem can arise because all the heating needs of the plant must be met by recovered heat from heat pumps. It is possible that during cold winter days there might not be enough recovered heat to meet this load. Whenever the heat load exceeds the cooling load and there is no alternative heat source available, an artificial cooling load must be placed on the heat pump.

This artificial heat source can in some cases be the cooling tower water itself. A direct heat exchanger between the cooling



Example[*]

Temperature of Condenser Water	*Cost Components*	*Mechanical Refrigeration Mode*	*Heat Recovery Mode*
65°F (18.3°C)	Cost of Cooling	$4.25	$8.00
	Cost of Heating	(0.5)(7.0) = $3.50	$0.00
	Total	$7.75	$8.00
75°F (23.9°C)	Cost of Cooling	$5.30	$8.00
	Cost of Heating	(0.5)(7.0) = $3.50	$0.00
	Total	$8.80	$8.00

[*] This example is based on the following assumptions:

a. The actual cooling load is 50% of chiller capacity (CL = 0.5 CAP).

b. The heating load (the demand for hot water) is 50% of cooling load. (HL = 0.5 CL)

FIG. 8.13k

The decision of operating in either the free cooling mode or in the heat recovery mode can be made automatically on the basis of cost-effectiveness.

tower and chilled water streams is also advantageous in the case when there is no heat load but there is a small cooling load during the winter. At such times, the chiller can be stopped and the cooling load can be met directly by the cooling tower water (Figure 8.13l).

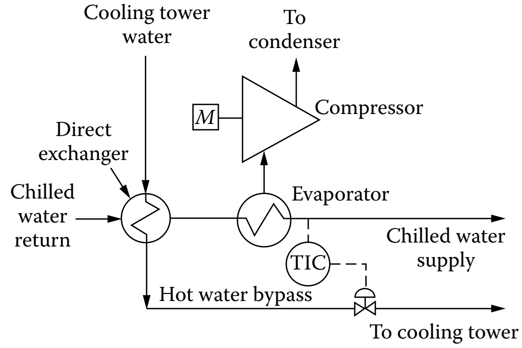


FIG. 8.13l

When the chiller compressor is on, the cooling tower water can be used as an artificial heat source; with the compressor off, it can be used as a means of direct cooling.[4]

OPTIMIZATION BY SYSTEM RECONFIGURATION

In most plants, coolant can be provided from many sources, so another approach to optimization is to reconfigure the system in response to changes in loads, ambient conditions, and utility costs. For example, during some operating and ambient conditions, the cooling tower water may be cold enough to meet the cooling load of the process directly.

Alternatively, if the cooling tower water temperature is below the chilled water temperature required by the process, the chillers can be operated in a free cooling or thermosiphon mode, which will be described in more detail later in connection with Figure 8.13o. In this case, the refrigerant circulation is driven by the temperature differential rather than by the compressor, because the cooling tower water in the condenser is colder than the chilled water in the evaporator. In this thermosiphon mode of operation, the chiller capacity drops to about 10% of its rating.

When the compressor of a chiller or heat pump is in operation, the process is in the so-called *mechanical refrigeration mode* (Figure 8.13m).

When the process is cooled directly by the cooling tower water, without the use of chillers or refrigeration machines, it is in the *free cooling mode.* Most modern cooling systems are provided with the capability of operating in several different modes as a function of load, ambient conditions, and utility costs.

The decision to switch from one operating mode to another is usually based on economic considerations. Such a control system will automatically select the mode of operation that will allow the meeting of the load at the lowest total cost. Figure 8.13k illustrates the decision-making process by which one can decide if the heat recovery or the mechanical refrigeration mode will provide the better cost-effectiveness for the prevailing conditions. The same kind of logic is applied when the choice is between free cooling or mechanical refrigeration.

Once it is decided to switch modes, the actual reconfiguration of the associated piping and valve positions are selected automatically. In larger, more complex cooling systems, there can be dozens of different modes of operation (such a system was designed by the author for IBM corporate headquarters at 590 Madison Avenue in New York). In this chapter, only four basic modes will be discussed.

The cooling system illustrated in Figure 8.13m is configured as a mechanical refrigeration system. The controls for this configuration have been described in Figures 8.13g

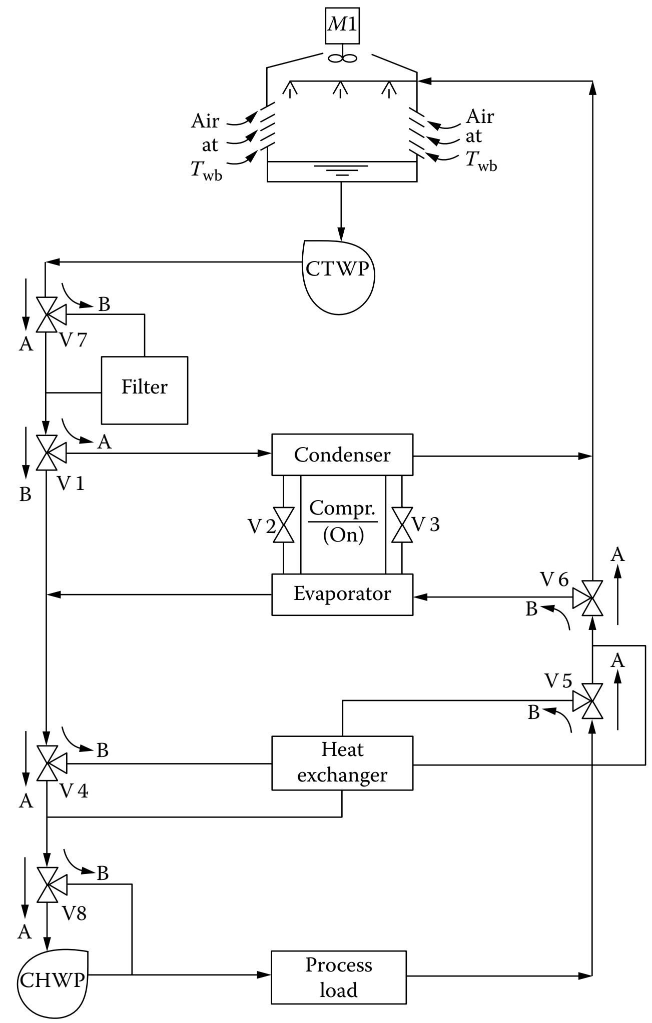


FIG. 8.13m
This mechanical refrigeration system can be automatically reconfigured to operate in any of the three "free cooling" configurations, which are listed in Table 8.13n.

and 8.13j. The heavily drawn pipelines show the flow paths that are active in this mode of operation.

When the load drops off or when the ambient temperature decreases, it is possible to switch to one of the free-cooling modes after the compressor has been stopped. Table 8.13n lists the equipment status and valve positions for each of the operating modes that the system in Figure 8.13m can support.

Indirect Free Cooling by Thermosiphon

Free cooling can be direct or indirect. When indirect free cooling is used, the compressor is off and the refrigerant transports the heat to the cooling tower water from the chilled water by evaporating in the "condenser" and condensing in the "evaporator." This is mode 2 in Table 8.13n. In this indirect free-cooling configuration the compressor is off and the heat is transferred from the evaporator to the condenser through the natural migration of refrigerant vapors (Figure 8.13o).

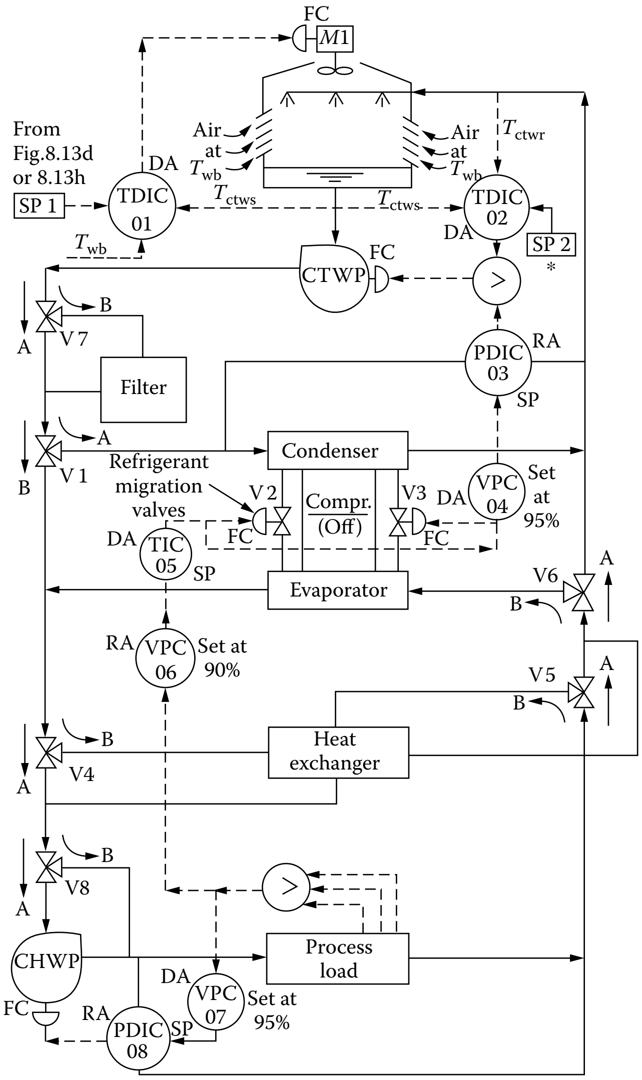
Opening the refrigerant migration valves (*V*2 and *V*3) equalizes the evaporator and condenser pressures. Because the condenser is at a lower temperature, the refrigerant that is vaporized in the evaporator is recondensed and returns to the evaporator by gravity flow. The cooling capacity of a chiller is about 10% of full load in this mode of operation.

The cooling system is operated in this mode when the load is low and the cooling tower water temperature is about 10°F (5.6°C) below the required chilled water temperature. This usually means 40–45°F (4.4–7.2°C) cooling tower and 50–60°F (10–15.6°C) chilled water temperatures. Such high chilled water temperatures are not unrealistic in the winter, because the air tends to be dry and dehumidification is not required.

This (mode 2 in Table 8.13n) can also be implemented without controls. Under manual controls both sets of pumps are operated and the temperatures are allowed to float as a function of load and ambient conditions.

TABLE 8.13n
Equipment Status in Four Operating Modes of a Refrigeration System

Equipment	*Mode 1 (Mechanical refrigeration)*	*Mode 2 (Vapor migration based indirect free cooling)*	*Mode 3 (Indirect free cooling by the use of heat exchanger)*	*Mode 4 (Direct free cooling with full filtering)*
Compressor	On	Off	Off	Off
Cooling-tower pumps	On	On	On	On
Chilled-water pumps	On	On	On	Off
Valve V1	A	A	B	B
Valve V2 and 3	Closed	Open	Closed	Closed
Valve V4	A	A	B	A
Valve V5	A	A	B	A
Valve V6	A	B	A	A
Valve V7	A	A	A	B
Valve V8	A	A	A	B



* From Figure 8.31i.

FIG. 8.13o

In the free-cooling mode (mode 2 in Table 8.13o), the compressor is off. The refrigerant evaporates in the "condenser" and freely migrates to the evaporator, where it condenses, because the condenser is colder than the evaporator.

Optimized Operation To optimize this operation, the control strategy shown in Figure 8.13o has to be implemented. Here TDIC-01 maintains the optimum approach and TDIC-02 maintains the optimum range, unless these controls are overridden by VPC-04 when the refrigerant migration valves approach their full openings, signaling that increased heat transfer is needed at the condenser in order to meet the load.

The migration valves $V2$ and $V3$ are throttled by TIC-05. The set point of TIC-05 is maximized by VPC-06 by allowing all user valves to open until the most-open valve reaches 90% opening. The rate of chilled water pumping is minimized by VPC-07, which modulates the set point of PDIC-08.

As VPC-06 normally keeps the most-open user valve at 90% opening, therefore the set point of VPC-07 is not reached under normal conditions. Therefore, under these conditions, the PDIC-08 set point is kept at its allowable minimum. This in turn results in minimizing the chilled water pumping cost so long as all the users are satisfied. If any of the user valves would open to over 95%, VPC-07 will quickly increase the pumping rate to guarantee sufficient coolant to that valve.

When this occurs, the system is using up the cooling capacity stored in the circulated chilled water. Once this is exhausted, control will be lost. Therefore, it is advisable to switch to a different mode of operation when this condition is detected.

Indirect Free Cooling by Heat Exchanger

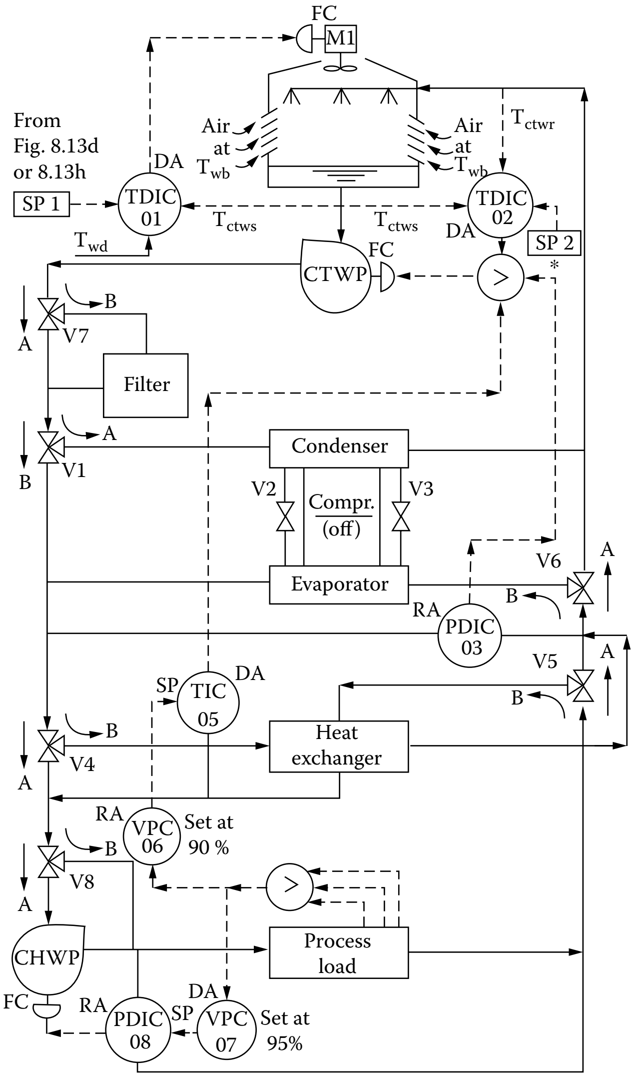
Another method of indirect free cooling is to transfer the heat directly between the cooling tower water and the chilled water through a heat exchanger (mode 3 in Table 8.13n). This allows the complete bypassing of the refrigeration machine. This system has no heat transfer capacity limitation, because any cooling load can be handled as long as the plate-type heat exchanger is large enough to handle it. Therefore, the main advantage of mode 3 over mode 2 is that it is not restricted to loads of 10% or less of chiller capacity. Its main disadvantage is the need for an additional major piece of equipment, namely the heat exchanger.

Indirect free cooling using the plate-type heat exchanger can be operated manually, without any controls. In that case, the chilled water temperature will float as a function of the process load and the ambient temperature, while the pump stations operate at full capacity.

Optimized Operation If it is desired to optimize the operation of mode 3, the control system described in Figure 8.13p can be implemented. Here, on the cooling tower, TDIC-01 maintains the optimum approach and TDIC-02 maintains the optimum range. TDIC-02 can be overridden by PDIC-03 or by TIC-05. The pumping rate of cooling tower water circulation will therefore be set by the highest of the three controller outputs. The purpose of PDIC-03 is to guarantee the minimum pressure differential required for the plate-type exchanger, while TIC-05 can override TDIC-02 and PDIC-03 if more cooling is required.

The set point of TIC-05 is set by VPC-06 to prevent the most-open user valve from exceeding a 90% opening. Under normal conditions, the pumping rate of chilled water is kept at a minimum by PDIC-08, which maintains the minimum ΔP required across the process load. When VPC-06 is unable to keep the most-open user valve at a 90% opening and this opening rises to 95%, VPC-07 will start raising the set point of PDIC-08, thereby increasing the pumping rate.

This is only a temporary cure, because the added cooling capacity is available only at the expense of heating up the stored chilled water in the pipe distribution system. Therefore, it is advisable to detect if this condition occurs for more than a minute or so and, when it does, automatically switch the system to a cooling mode that can handle the increased



* From Figure 8.13i.

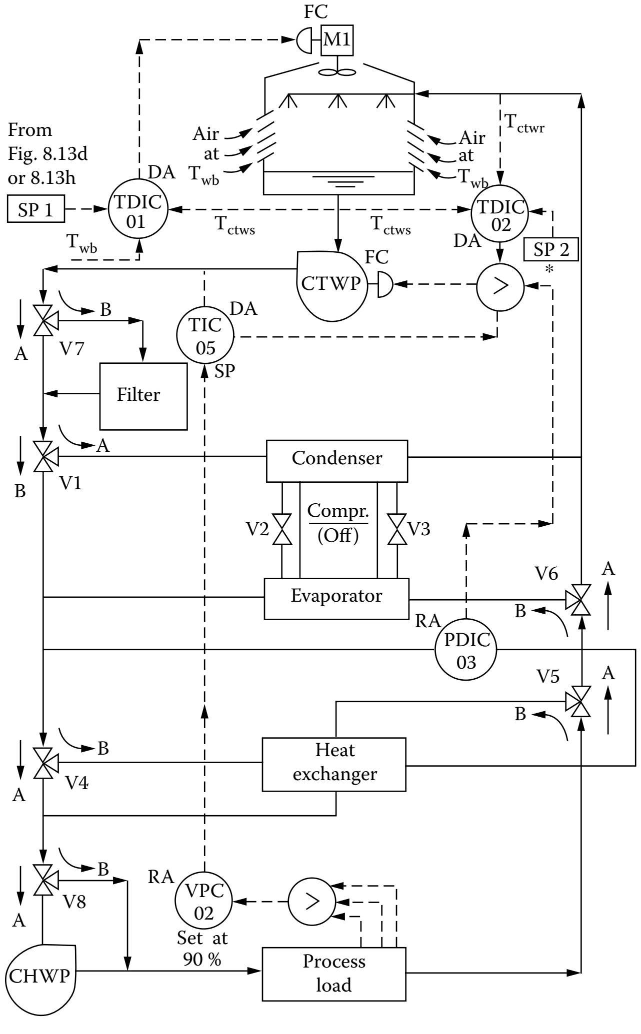
FIG. 8.13p
Indirect free cooling is obtained by sending the chilled water that is returning from the process to a heat exchanger in which it is cooled by the cooling tower water (mode 3).

load. This can be mechanical refrigeration (mode 1) or free cooling through interconnection (mode 4).

Direct Free Cooling

In direct free cooling, the cooling tower water is piped directly to the process load, as shown in Figure 8.13q. This method (mode 4) is the most cost-effective mode of cooling, because both the compressor and the chilled water pump station are off. In the winter, this mode of operation can handle high process loads, as it is limited only by the size of the towers and their pumps.

The main disadvantage of direct free cooling is that it brings potentially dirty cooling tower water to the process users, causing plugging and build-up on the heat transfer surfaces. This problem can be solved either by full flow filtering (also called strainer cycle), which is shown in Figure 8.13q, or by the use of closed-circuit, evaporative cooling towers. In such “noncontact” or “closed-loop” cooling towers, the water has no opportunity to pick up contaminants from the air.



* From Figure 8.13i.

FIG. 8.13q
In the direct free-cooling mode, the cooling tower water is sent directly to the process load.

This configuration is also frequently operated manually, without automatic controls. In that case, the cooling tower water temperature floats as the load and ambient conditions vary, and the fan and pumping rates are not optimized.

Optimized Operation If optimization is desired, the controls shown in Figure 8.13q can be implemented. In this figure, TDIC-01 serves to keep the cooling tower approach at an optimum value, and TDIC-02 optimizes its range. The range controller can be overridden by PDIC-03 or by TIC-05 when either of these controllers require a higher pumping rate than does TDIC-02, the range controller. The set point of TIC-05 is optimized to keep the most-open user valve from exceeding a 90% opening.

Operating Mode Reconfiguration

The number of possible modes of operation can be rather high, if the cooling system of the plant includes optional storage, heat recovery, and alternate types of motor drives. Switching from one mode to another is not as simple as it might first appear, because reconfiguration necessitates the stopping and starting of equipment and because the reconfiguration of control loops often requires the modification of pump discharge heads.

For example, when switching to mode 4, illustrated in Figure 8.13q, the water circulation loop served by the cooling tower pumps becomes much longer. This in turn shifts the operating point of the pumps and can lower their efficiency if the system is not carefully designed and evaluated for each operating mode.

The control loop configuration requirements for all four modes are tabulated in Table 8.13r. When the control loops are automatically reconfigured and their manipulated variables are changed, it is important to also revise their tuning constants, because the time constants of the loops are also changed. This should preferably be done automatically to minimize the potential for human error.

Whenever the process can tolerate an interruption of a few minutes in its cooling system, the switching from one mode to another should be done in a stagnant state, while the equipment is turned off. This usually is acceptable for all motors except for the pumps, which are serving the process load. These pumps usually can be left running, and during switching they can utilize the coolant storage capacity of the water distribution piping. Transferring to another mode while the pumps are running can be accomplished, because the slowly diverting three-way valves will never completely block the pump discharge but will only gradually change the destination of the water.

TABLE 8.13r
Listing of the Manipulated Variables That Are Assigned To Be Throttled by the Various Controllers in the Four Modes of Operation

	Manipulated Variables Controlled by the Controller			
Controller	*Mode 1*	*Mode 2*	*Mode 3*	*Mode 4*
TDIC-01	Fan	Fan	Fan	Fan
TDIC-02	CTWP	CTWP	CTWP	CTWP
PDIC-03	CTWP	CTWP	CTWP	CTWP
VPC-04	—	PDIC-03 Set point	—	—
TIC-05	Compressor	V2/3	CTWP	CTWP
VPC-06	TIC-05 Set point	TIC-05 Set point	TIC-05 Set point	TIC-05 Set point
VPC-07	PDIC-08 Set point	PDIC-08 Set point	PDIC-08 Set point	—
PDIC-08	CHWP	CHWP	CHWP	—

If the mode changes are frequent or if the coolant capacity of the piping headers is very small (insufficient to meet the process load for even a few minutes), all systems must be switched while running. Dynamic switching requires more planning and a higher level of automation, because the automatic starting of certain pieces of equipment (such as chillers driven by a steam turbines) requires a more comprehensive set of safety interlocks.

Automatic operating mode reconfiguration is one of the most powerful tools of optimization, because it provides flexibility to a previously rigid system. The mode reconfiguration technique is not limited to cooling systems. It can be effective in any unit operation in which the system must adapt to changing conditions.

RESPONSIVENESS AND PRIORITIZING

In some processes it is critical that the chillers be able to respond to fast and drastic changes in the process cooling load. In other processes, the plant's cooling system simultaneously serves high- and low-priority users, and it is much more important to meet the needs of the high-priority users than the others. The control systems described in the following paragraphs are designed to improve responsiveness and to consider user priorities in chilled water distribution.

More Responsive Control

Feedforward anticipation can increase the responsiveness of the control system while also providing more precise temperature control. Such precision is desired when, for example, the goal of optimization is to maximize chilling without freezing.

When in a heat exchanger the heat transfer area is fixed (such is the case in an evaporator with constant refrigerant level), the rate of heat transfer (load) is a function of the temperature difference between the process fluid and the boiling refrigerant. Therefore, as the load changes, the boiling temperature (and pressure) must also change.

Figure 8.13s shows the main components of such a responsive control system. In this case the instantaneous load is calculated by multiplying the flow rate of the cooled process fluid (in FY-05) by the drop in temperature (TY-04) that it should experience. Based on this flow rate and the desired process temperature (set point of TIC-01), FY-06 determines the desired refrigerant temperature. The value of K in this summing device is adjustable to reflect the actual slope of load vs. refrigerant temperature. PY-07 serves to convert the desired boiling temperature to the corresponding vapor pressure set point.

Dynamic compensation is provided by FY-03 to match the response of process temperature to refrigerant pressure.[5] A valve position controller can be added to detect the refrigerant control valve opening and to keep that valve over 90% open by varying the speed of the compressor, in order to minimize the operating cost of the compressor.

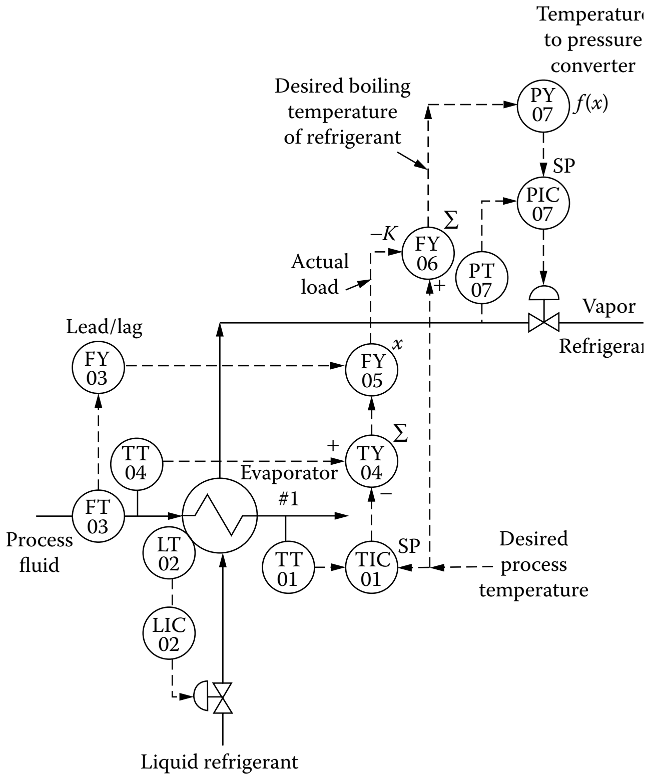


FIG. 8.13s

Increased responsiveness can be obtained by the feedforward adjustment of refrigerant pressure. (Adapted from Reference 5.)

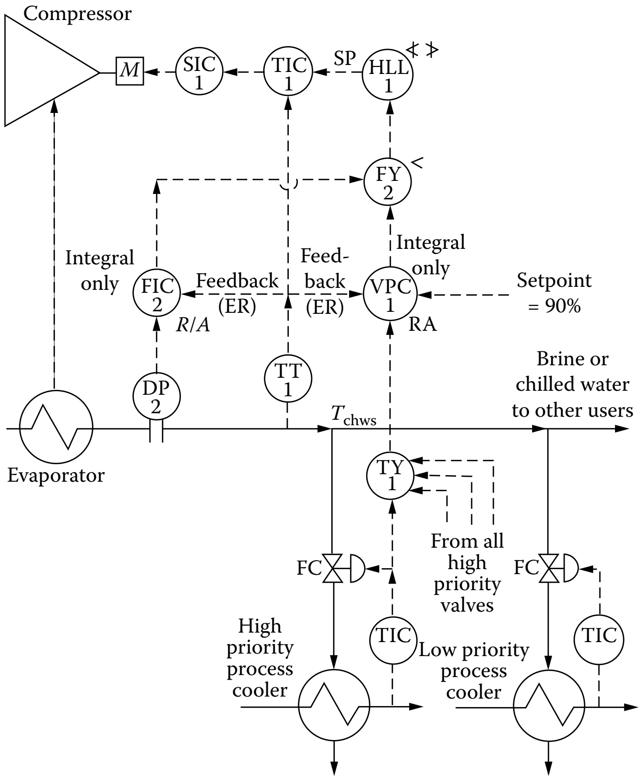


FIG. 8.13t

The chilled-water supply temperature can be optimized by keeping the most-open high priority user valve at 90%. (Adapted from Reference 2.)

Water Distribution Optimization

Figure 8.13f illustrates the ideal water distribution system, in which the individual users are served by two-way valves and the optimum supply temperature is found by keeping the most-open valve at 90% opening. Although this is a very effective control configuration, it is not always used when some of the chilled water users are not critical.

If the number of user valves is very high and if the valves are distributed over a large area, or if the loops are not properly tuned and the valves have a tendency to cycle from fully closed to fully open, the control shown in Figure 8.13f is not practical. In such situations, the concept illustrated in Figure 8.13f can still be applied by selecting a few representative user valves that are not cycling and by basing the optimization on those.

If the chilled water user valves are grouped into high- and low-priority categories or into other groups, the water distribution and supply temperature can still be optimized. Figure 8.13t illustrates a control system in which the high-priority users are treated the same way as they were in Figure 8.13f. The low-priority users, on the other hand, are grouped together, and their demand is detected through the measurement of total flow.

If all low-priority valves are the two-way type and if supply temperature is constant, flow will vary directly with cooling load. If valve-position-based optimization were applied, it would attempt to maximize valve opening, which in turn would maximize flow by raising T_{chws} to the highest acceptable level. This same goal is achieved by FIC-2 in Figure 8.13t. As the flow drops off as a result of a reduction in demand for cooling, FIC-2 will raise the supply temperature, which in turn will increase the total flow, as the valves open up. Therefore, the total flow is kept constant, and load variations result in supply temperature variations. VPC-1 and FIC-2 should both be integral-only controllers, with time constants of several minutes.

The control of total flow alone in Figure 8.13t can satisfy the average user, but not all the users. Therefore, the critical users cannot be treated in this manner, but must be protected by valve position control. Such control is provided by VPC-1, which overrides the flow controller whenever a high-priority valve reaches 90% opening.

Three-Way Valves to Optimize HVAC Systems In many older or air-conditioning-type applications, three-way control valves are used. This is a highly undesirable practice, because the use of three-way valves unnecessarily increases the required amount of pumping and it also lowers return water temperature. Yet, these badly designed systems can also be optimized. A control system that will do that is shown in Figure 8.13u.

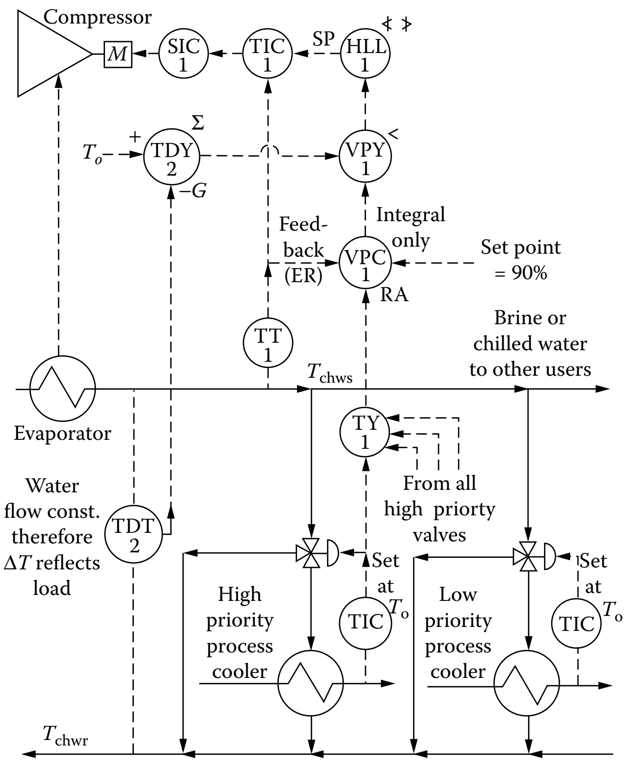


FIG. 8.13u

The optimization of chilled water supply temperature on the basis of keeping the most open high priority three-way user valve 90% open. (Adapted from Reference 4.)

When three-way valves are used, the total flow is relatively constant, and it is the return water temperature that reflects the variations in the cooling load. Therefore, the control system in Figure 8.13u determines the set point for TIC-1 by relating the temperature rise ($T_{chwr} - T_{chws}$) and the desired space temperature set points (T_o) as follows:[5]

$$\text{TIC-1 setpoint} = T_o - G(T_{chwr} - T_{chws}) \qquad \textbf{8.13(2)}$$

If the thermostats are set at T_o = 80°F (27°C) and if, at full load, T_{chws} = 50°F (10°C) and T_{chwr} = 65°F(18°C), the value of G can be determined as follows:

$$G = (T_o - \text{setpoint})/(T_{chwr} - T_{chws}) = (80 - 50)(65 - 50) = 2 \qquad \textbf{8.13(3)}$$

TDY-2 in Figure 8.13u therefore calculates the desired set point for TIC-1 at any load. For example, if the load drops to 50%, the space temperature T_o will begin to fall, and the thermostats will divert more coolant into the return line. This will reduce T_{chwr} from 65°F (18°C) to 57.5°F (14°C). In conventional systems, this then would become the new steady state. With the control system shown in Figure 8.13u, TDY-2 will respond by revising the TIC-1 set point as follows:

$$\text{set point} = 80 - 2(57.5 - 50) - 65°\text{F} \qquad \textbf{8.13(4)}$$

As T_{chws} is slowly increasing, the space thermostats will divert less and less water until they come to rest at the same percentage of water diverted as at full load. The benefit of this optimization strategy is in reducing the temperature differential across the chiller, which in turn lowers the operating cost by 15% for each 10°F (5.5°C).

The setting of G determines the percentage of coolant diverted by the average valve. If G is set too low, some valves will be 100% open to the chilled water supply and therefore out of control. If G is set too high, the chilled water supply temperature will be lower than needed, and compressor energy will be wasted.

In order to protect the high-priority users from running out of coolant, when G is set low, a valve-position-based override is used in Figure 8.13u. Whenever a high-priority valve is more than 90% open, VPC-1 will override the set point of TIC-1 and will lower it until all high-priority valves are less than 90% open.

OPTIMIZING STORAGE AND LOAD ALLOCATION

The total cost of cooling can also be reduced by making the coolant when it is less expensive to make it and placing it in storage until it is needed. When a variety of equipment is available to meet the cooling load of the plant, the total cost of operation can also be minimized by always using the most efficient combination of equipment. These are the topics of the next paragraphs.

Optimized Storage

If storage tanks are available, it is cost-effective to generate the daily brine or chilled water needs of the plant at night, when it is the least expensive to do so, because ambient temperatures are low and night-time electricity is less expensive in some areas.[6]

When demand is low, operating costs can be reduced by operating the chillers part of the time at peak efficiency rather than continuously at partial loading. Efficiency tends to be lower at partial loads because of losses caused by friction drop across suction dampers, prerotation vanes, or steam governors.

Cycling is practical if the storage capacity of the distribution headers is enough to avoid frequent stops and starts. When operation is to be intermittent, data such as the amount of heat to be removed and the characteristics of the available chillers are needed to determine the most economical operating strategy.

When the chiller is cycled, the thermal capacity of the chilled water distribution system is often used to absorb the load while the chiller is off. For example, if the pipe distribution network has a volume of 100,000 gallons (378,000 l), this represents a thermal capacity of approximately 1 million BTUs for each °F of temperature rise (1.9×10^6 J/°C).

So, if one can allow the chilled water temperature to float 5°F (2.8°C) (such as from 40°F to 45°F, or from 4.4°C to 7.2°C) before the chiller is restarted, this represents the equivalent of approximately 400 tons (1405 kW) of thermal capacity. If the load happens to be 200 tons (704 kW), the chiller can be turned off for 2 hours at a time. If the load is 1000 tons (3,514 kW), the chiller will be off for only 24 minutes. This illustrates the natural load-following, time-proportioning nature of this scheme.

If the chiller needs to be off for a longer period than the thermal capacity of the distribution system can provide, three options are available:

1. Storage tanks can be added to increase the water volume.
2. A second chiller can be started (not the one that was just stopped).
3. The load can be distributed among several chillers of different sizes, keeping some in continuous operation while cycling others.

Optimized Load Allocation

Continuous measurement of the actual efficiency ($/ton) of each chiller can enable the plant to meet all loads with the most efficient combination of machines. In plants with multiple refrigerant sources, the cost per ton of cooling can be calculated from direct measurements, and it can be used to select the most efficient combination of units to meet any existing or anticipated loads. The differences between the cost per ton characteristics and the efficiency-vs.-load characteristics of the individual coolant sources can be combined and used as in Figure 8.13v.

In simple load allocation systems, only the starting and stopping of the chillers is optimized. In such systems, when the load is increasing, the most efficient idle chiller is started, and when the load is dropping, the least efficient one is stopped. In more sophisticated systems, the load distribution between operating chillers is also optimized. In such systems, the real-time efficiency of each chiller is calculated. This determines the incremental cost for the next load change for each chiller.

If the load increases, the incremental increase is sent to the set point of the most cost-effective chiller. If the load decreases, the incremental decrease is sent to the least cost-effective unit (Figure 8.13w). Software packages are readily available that have proven capabilities for continuous load balancing through the predictions of efficiencies and operating costs.[7]

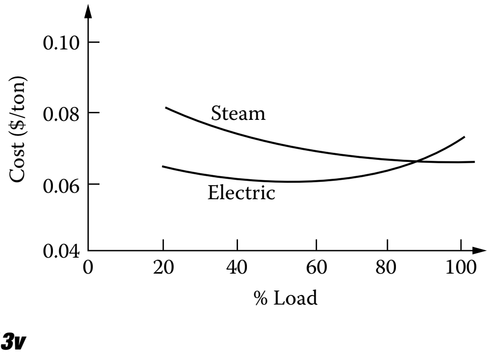


FIG. 8.13v
The unit cost of refrigeration is a function of both the load and of the characteristics of each chiller. Installed systems should be used as the basis of obtaining the operating cost curves.[6]

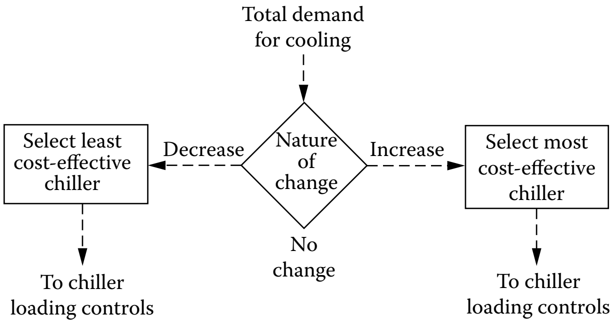


FIG. 8.13w
Optimized, computer-based load allocation will direct load increases to the most cost-effective chiller, while sending the load decreases to the least cost-effective chiller.

With the strategy described in Figure 8.13w, the most efficient chiller will either reach its maximum loading or will enter a region of decreasing efficiency and will no longer be the most efficient. When the loading limit is reached on one chiller, or when a chiller is put on manual, the computer will select another unit as the most efficient one for future load increases.

The least efficient chiller will accept all decreasing load signals until its minimum limit is reached. Its load will not be increased unless all other chillers are at their maximum load or when it is in manual. As shown in Figure 8.13v, some chillers can have high efficiency at normal loads while being less efficient at low load. Such units are usually not allowed to shut down, but are given a greater share of the load by the control system.

If all chillers are identical, some will be driven to the capacity corresponding to maximum efficiency while others will be shut down by this strategy, and only one chiller will be placed at an intermediate load.

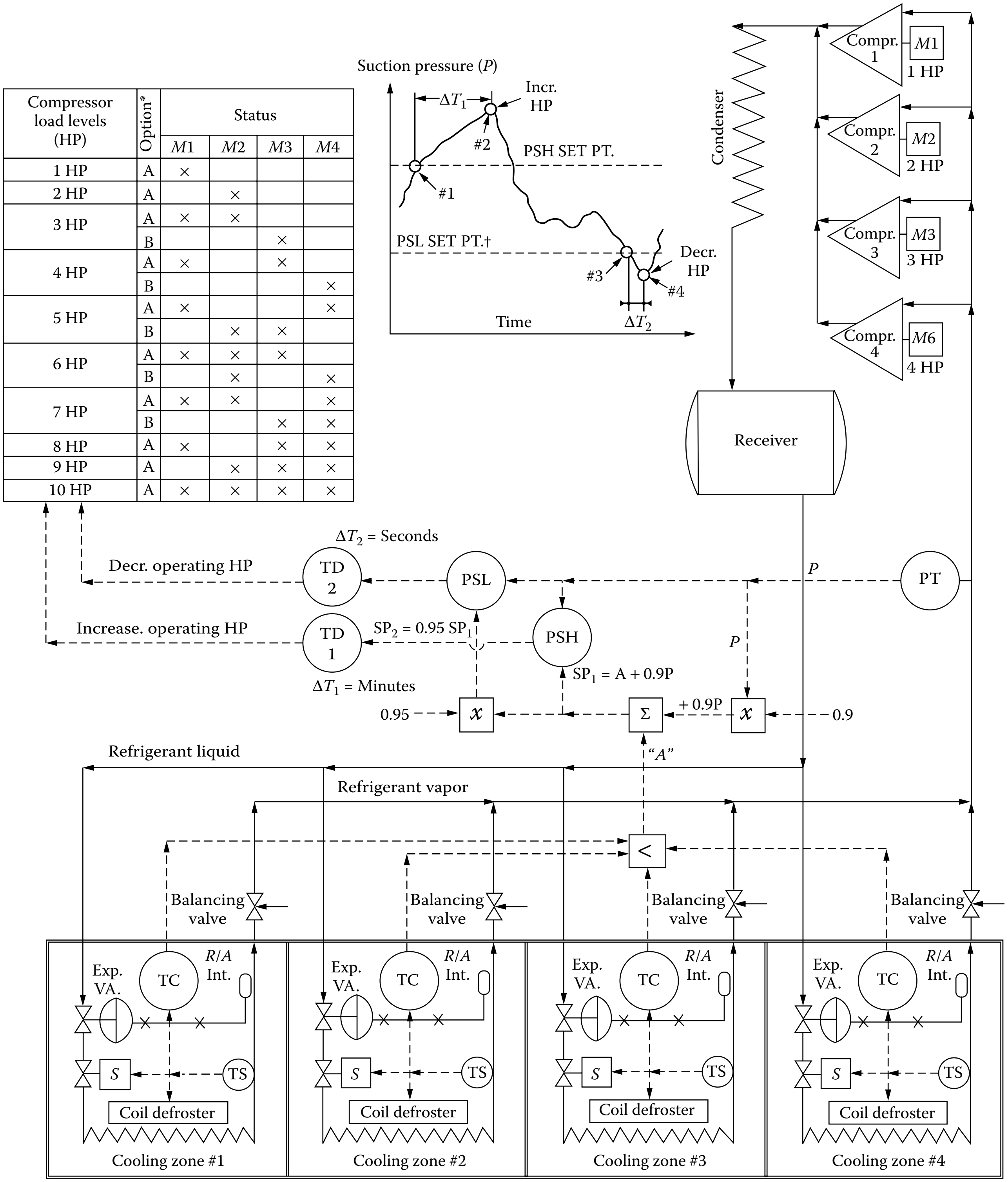
In starting up chiller stations, the optimization system "knows" how many BTUs need to be removed before start-up and the size and efficiency of the available chillers. Therefore, the length of the "pull-down" period can be minimized and the energy cost of this operation can also be optimized.

OPTIMIZATION OF SPECIAL CHILLERS

In the following paragraphs the optimization of some special chiller configurations will be discussed. These will include the commercial chillers that utilize a number of positive displacement refrigerant compressors. Also discussed will be the multistage refrigeration units that are needed to generate very low temperature coolants and the multiple-user systems.

Optimized Commercial Chillers

In supermarkets, warehouses, and other commercial facilities, there usually are a large number of loads (zones, each served by an evaporator coil), and there are a number of



Compressor load levels (HP)	Option*	Status			
		M1	M2	M3	M4
1 HP	A	×			
2 HP	A		×		
3 HP	A	×	×		
	B			×	
4 HP	A	×		×	
	B				×
5 HP	A	×			×
	B		×	×	
6 HP	A	×	×	×	
	B		×		×
7 HP	A	×	×		×
	B			×	×
8 HP	A	×		×	×
9 HP	A		×	×	×
10 HP	A	×	×	×	×

FIG. 8.13x

Here a control system is shown that will minimize the energy cost, while equalizing the operating times of multiple chiller compressors, which are serving an unlimited number of evaporator coils.[8]

positive-displacement refrigerant compressors, which are turned on or off to follow their loads. The goals of optimization are to minimize the cost of operation (to meet any cooling load with a minimum of compressor horsepower) and to select the operating compressors in such a way that the total run time of each machine is about the same.

Figure 8.13x illustrates such a control system, which incorporates some of the ideas contained in the patents listed in Reference 8. In Figure 8.13x each of the cooling zones can be a separate room, freezer, or other cooled area, which is provided with its own temperature controls (TC). Each of the zones is cooled by an evaporator coil, which is kept cold by its boiling liquid refrigerant.

The liquid refrigerant is admitted into each coil by an expansion valve, which controls the superheat of the evaporated vapors by detecting both the pressure and the temperature of those vapors. The expansion valve admits as much refrigerant liquid as is required to keep the evaporator superheat (the temperature above the boiling point at the operating pressure of the refrigerant) constant.

A manual balancing valve is also provided for each zone. These valves are located in the refrigerant vapor lines leaving the zones. If all zones are to be kept at the same temperature, all balancing valves can be kept fully open. This will minimize the pressure drop on the suction side of the compressors and thereby will maximize their efficiency.

If some of the zone temperatures are to be controlled (by their TCs) at a higher temperature than the others, the balancing valves on these higher temperature zones must be throttled to increase the vaporization pressure and thereby the boiling temperature of the refrigerant in the coils of those zones.

Defrosting and Sequencing Controls If the zones are operated below or near to the freezing temperature (below 35°F, or 2°C), it is also necessary to provide defrosting controls because of the ice build-up on the cold evaporator coils. The defrosters are controlled by coil surface temperature detector switches (TS), one in each zone (TS). These switches are continuously detecting the coil surface temperatures, and when that temperature drops below some low limit for some preset time period, the switch first shuts off the refrigerant liquid supply by closing a solenoid valve.

Once all refrigerant has been vaporized, the switch turns on a coil defroster heater for a short time, to remove the ice. During the defrosting cycle, the zone temperature controller (TC) is disabled by setting its measurement and set point to be equal and thereby causing its output signal to stay constant.

The nonoptimized method of turning the refrigerant compressors on and off is by sequential controls, where each added compressor unit is turned on when the suction pressure to the overall compressor station has dropped to an even lower level. This approach is inefficient, because the reaching of these lower levels of suction pressure requires added compressor horsepower.

Optimized Controls The control system shown in Figure 8.13x eliminates the waste of added compressor horsepower by turning added compressors on (PSH) or turning unnecessary compressors off (PSL) at the pressure settings of PSH and PSL. This is achieved by both pressure switches actuating a time delay (TD), and whenever the pressure switch is actuated for the time period set on the time delay, it changes the operating horsepower level of the system. TD-1 increases it while TD-2 decreases the horsepower level.

TD-1 is set in minutes, while TD-2 is set in seconds. Therefore, a response to a load increase is relatively slow (minutes), while the response to a load decrease is fast (seconds), which serves energy conservation.

Further optimization is obtained by pushing the control gap between the pressure switches (PSH and PSL) as high as the loads will allow. This continuous floating of the controlling suction pressure gap is achieved by feedforward action based on zone temperature control. As any one of the zone temperatures rises above the zone thermostat (TC, a reverse-acting, integral-only controller) set point, the output signal of that TC drops. A low-signal selector selects the lowest of the TC outputs (A) and thereby identifies the zone with the highest cooling load (the critical zone, which, if satisfied, will result in all others being satisfied).

The set point of PSH is obtained by taking 90% of the suction pressure of the compressor station (0.9P) and adding to it the output signal (A) of the critical TC. The set point of PSL is 95% of the PSH set point. Both the 90% and the 95% values are adjustable and can be varied from one installation to the next.

The energy consumption of the system can be minimized and its operation optimized by maximizing these set points. This is because whenever all the zone temperatures are on set point (or below), the suction pressure of the operating compressor station is increased, and therefore the amount of work the compressors need to do is reduced.

The graph of suction pressure (P) vs. time in Figure 8.13x illustrates the operation of the system. As the suction pressure rises, at point #1 the set point of PSH is reached and the time delay TD-1 is actuated. At point #2 the time delay that was set on TD-1 has run out, which increases the operating horsepower level by one unit. As a result, the suction pressure starts to drop.

If at this point the operating compressor capacity exceeds the cooling load, the suction pressure will continue to drop, and when it reaches the set point of PSL (point #3), TD-2 will be energized. After the time delay of TD-2 has run out (point #4), the operating horsepower level of the system is reduced by one unit.

Minimum Maintenance In addition to energy optimization, the system is also optimized from a maintenance point of view by keeping all compressor running times approximately the same. This is accomplished by the use of the horsepower table in the top left of Figure 8.13x. Let us assume that the system is operating at the 4 hp level when PSH calls for an increase to 5 hp. The 5 hp level can be met by either operating compressors 1 and 4 (option A) or by operating compressors 2 and 3 (option B). In this situation the least used compressor pair will be started.

If at some time later PSL calls for reducing the operating level to 4 hp, there again are two options. Option A is to run compressors 1 and 3, while option B is to run compressor 4.

Here, the choice is made in such a manner that the most used compressor will be turned off. The net result is a continuous balancing of running times between machines. According to Reference 8, the total energy savings obtained by this control system amounts to about 20%.

Multistage Chillers

It is not practical to obtain a compression ratio outside the range of 3:1 to 8:1 when using the compressors usually installed in the process industry. This places a limitation on the minimum temperature that a single-stage refrigeration unit can provide. For example, in order to maintain the evaporator at –80°F (–62°C) while the condenser is operating at 86°F (30°C), the required compression ratio for Freon-12 is calculated as

$$\text{compression ratio} = \frac{\text{refrigerant presure at } 86^\circ\text{F}}{\text{refrigerant presure at } -80^\circ\text{F}} = \frac{108}{2.9} = 37 \qquad \textbf{8.13(5)}$$

For a single-stage machine, such a compression ratio is not practical, and therefore, a multistage system is required to provide it. Figure 8.13y illustrates a multistage refrigeration system, with ethylene as the lower and propane as the higher temperature refrigerant. In order to minimize the total operating cost, the loading of the two stages must be coordinated to balance the total work between the stages.

The higher the interstage temperature (evaporator #2), the more work is done by stage #1 and the less work remains for stage #2. If the compression ratios (and therefore the temperature differences) of the two stages are the same, the work will be nearly equally distributed. If true equality is to be achieved, the differences in the properties of the refrigerants and the added load on the higher stage caused by the work that is introduced by the lower stage compressor must also be considered.

Optimized Controls One of the goals of optimization is to maximize the rangeability of the multistage refrigeration units. This rangeability will be maximum when the two stages approach their surge limits at the same time. This can be guaranteed by redistributing the load between the stages in such a manner that they maintain equal distance from their respective surge lines as the load drops.

In Figure 8.13y, TIC-01, the temperature controller of the chilled utility supply to the process, sets the speed of the first stage compressor #1 by modifying the set point of its suction pressure controller PIC-01. The speed of compressor #2 is set in speed ratio to that of compressor #1 by PY-01. The set point of PY-01 is optimized by an envelope algorithm that attempts to minimize the operating cost and to maximize rangeability, as it is moving the PY-01 set point to its optimum.

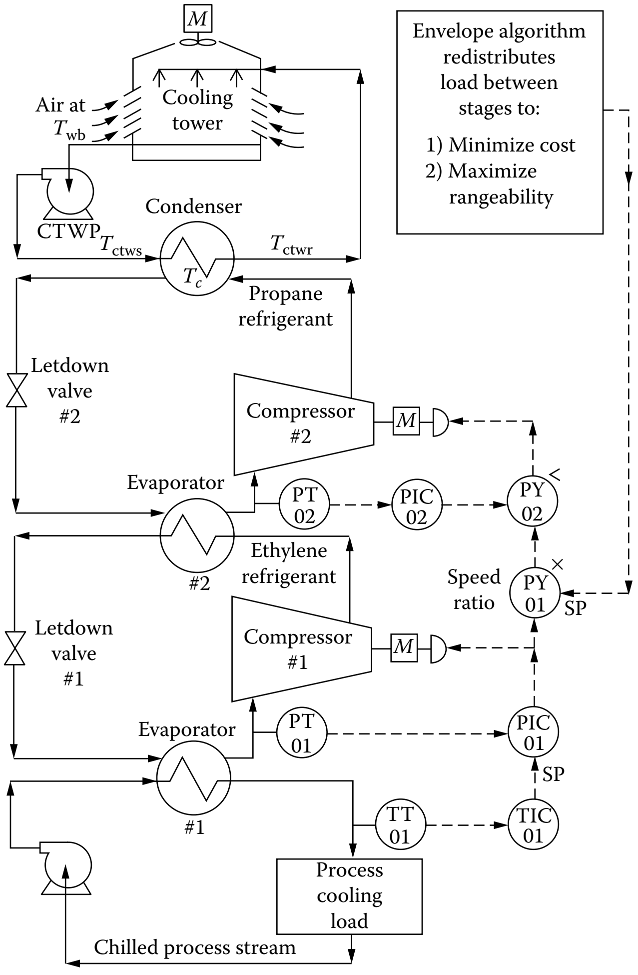


FIG. 8.13y

The operation of multistage refrigeration units can be optimized by automatically balancing the work load between the stages. (Adapted from Reference 5.)

As long as both stages are far from surge, the algorithm redistributes the load to minimize total operating cost. When one of the stages is nearing surge, the envelope algorithm increases its loading to keep the system out of surge. Therefore, at high loads the goal of load distribution between the stages is to minimize the total operating cost, and at low loads the goal is to keep the system out of surge.

As ambient temperatures drop, the cascaded stages will operate against a lower total temperature difference, and therefore, each of the stages will require less work. As the condenser temperature drops, suction pressure of compressor #2 will also drop. This in turn will cause the temperature in evaporator #2 to drop, which will result in the lowering of the suction pressure of compressor #1.

As the suction pressures drop, PIC-01 and PIC-02 will reduce the speeds of their respective compressors, thereby reducing the total work introduced by the heat pumps. Using these controls, this multistage refrigeration system will respond to ambient variations in a flexible and efficient manner.

Optimized Control of Multiple Users

When the same compressor serves several evaporators in parallel, either the evaporator temperature or the refrigerant level in the evaporators can be varied to follow the load. Figure 8.13z shows the controls for a configuration, where the evaporator levels (their wetted heat transfer areas) are kept constant, and therefore the load variations result in changes in the evaporator temperature. In the control system shown, TY-03 selects the most open TCV valve opening and keeps that valve from opening to more than 90% by speeding up the compressor when the valve opening reaches 90%.

The resulting reduction in compressor suction pressure lowers the evaporator temperature and increases the rate of heat transfer. Thus, no user is ever allowed to run out of coolant, and the compressor operating costs are kept to a minimum. This type of load-following optimization system, where compressor speed varies with the requirements of the most heavily loaded user, is very sensitive to dynamic upsets. Therefore, VPC-03 must be tuned for slow adjustment of the compressor speed to avoid upsetting the other users that are not selected for control.

If one of the user valves is consistently more open than the others, it might be possible to repipe the other loads to the interstage of the compressor. This might provide a better balance between users.

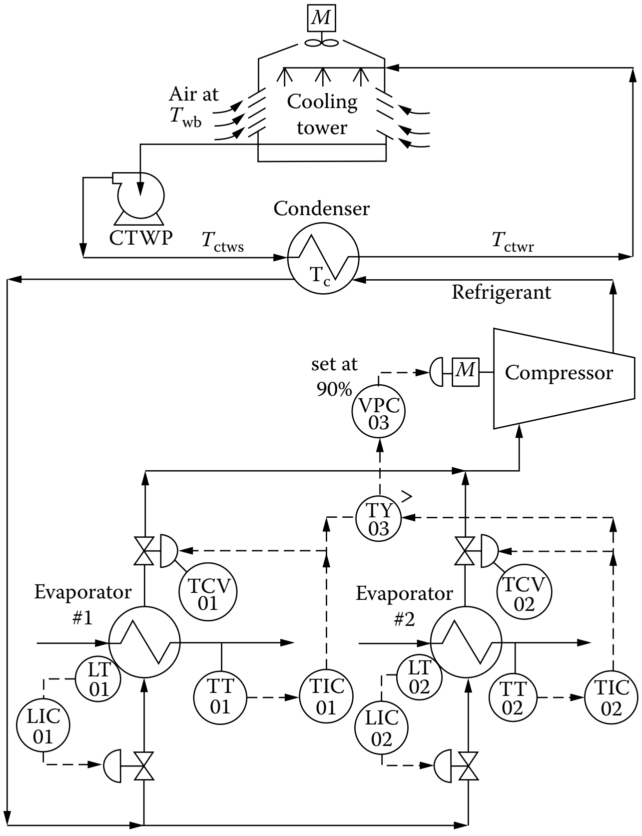


FIG. 8.13z

When a chiller is serving a number of users, the speed of the chiller can be automatically modulated so as to keep the temperature control valve of the most demanding user at 90% opening.

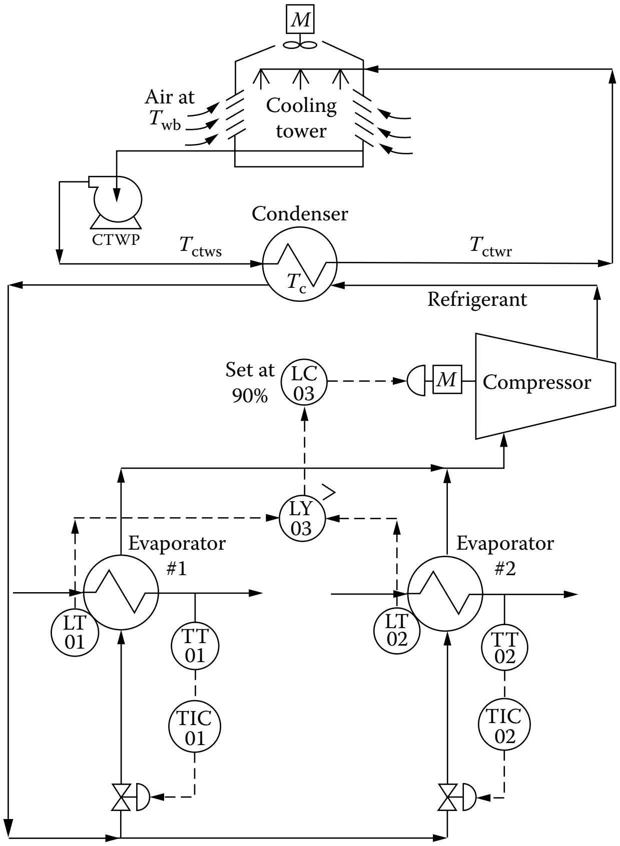


FIG. 8.13aa

The vapor-side control valves on multiple users can be eliminated by automatically adjusting the compressor speed, so as to keep the highest level in the evaporators at 90%. (Adapted from Reference 5.)

Eliminating the Vapor-Side Valves Figure 8.13aa illustrates a control system in which the vapor-side control valves have been eliminated. This advantage of the illustrated configuration must be balanced against a number of disadvantages. In the illustrated system, the heat transfer area is modulated and is exposed to boiling, and the user with the highest load will have the highest refrigerant level.

LY-03 selects the most flooded evaporator for control by LC-03, which, when the level rises above 90%, increases the speed of the compressor and lowers it when the level is below 90%.

This method of load following has some drawbacks. One problem is that the heat transfer surfaces within the evaporators are not fully utilized. Another difficulty is that the relationship between level and heat transfer area can be confused by foaming, which keeps the tubes wet even when the level has dropped.[5] For these reasons, tuning of such control system can be difficult, necessitating the sacrifice of responsiveness for stability.

Therefore, if responsiveness is critical, it is better to keep the refrigerant level constant and manipulate the evaporating pressure, as shown in Figure 8.13z.

RETROFIT OPTIMIZATION

In new plants it is easy to install variable-speed pumps, to provide the thermal capacity required for chiller cycling, or to locate chillers and cooling towers at the same elevation and near each other so that the cooling tower water pumping costs will be minimized. In existing plants, one has to accept the inherent design limitations and the system must be optimized without major changes to the equipment. Even with those limitations, one can still produce savings up to 50%,[3] but certain precautions are needed.

In optimizing existing chillers, it is important to give careful consideration to the constraints of low evaporator temperature, economizer flooding, steam governor rangeability, surge, and piping/valving limitations.

Surge Protection

A detailed discussion of surge controls is given in this chapter under Compressor Controls. Therefore, here it should suffice to just mention that surge occurs at low loads when not enough refrigerant is circulated. A surge condition can cause violent vibration and damage.

Old chillers might not have automatic surge controls and have only vibration sensors for shutdown. If the chillers will operate at low loads, it is necessary to add an antisurge control loop. Surge protection is always provided at the expense of efficiency. To bring the machine out of surge, the refrigerant flow must be artificially increased if there is no real load on the machine. The only way to provide this increase in flow is to add artificial and wasteful loads (for example, hot gas or hot water bypasses).

Therefore, it is much more economical either to cycle a large chiller or to operate a small one than to meet low loads by running the machine near its surge limit.

Low Evaporator Temperature

Low temperatures can occur in the evaporator when an old chiller is optimized—for example, when one that has been designed for operation at 75°F (23.9°C) condenser water is run in the winter using 45 or 50°F (7.2 or 10°C) condenser water. This phenomenon is exactly the opposite of surge, because it occurs when refrigerant is being vaporized at an excessively high rate. Such vaporization occurs because the chiller is able to pump twice the tonnage for which it was designed as a result of the low compressor discharge pressure.

In such a situation, the evaporator heat transfer area becomes the limiting factor. Furthermore, the only way to increase heat flow is to increase the temperature differential across the evaporator tubes. This shows up as a gradual lowering of refrigerant temperature in the evaporator until it reaches 32°F (0°C), at which point the machine shuts down to protect against ice formation.

There are two ways to prevent this phenomenon from occurring in existing chillers. The first is to increase the evaporator heat transfer area (a major equipment modification). The second is to prevent the refrigerant temperature in the evaporator from dropping below 33°F (0.6°C) by not allowing the cooling tower water to cool the condenser to its own temperature. The latter solution requires only the addition of a temperature control loop (Figure 8.13bb). This prevents the chiller from taking full advantage of the available cold water from the cooling tower by throttling its flow rate, thereby causing its temperature to increase.

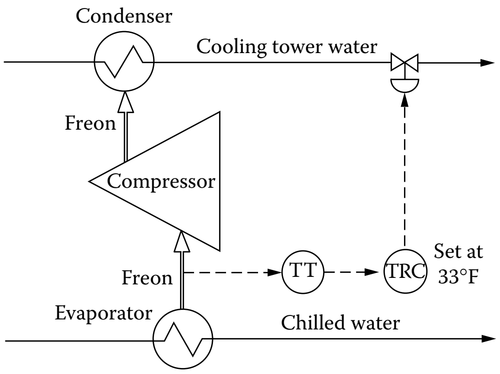


FIG. 8.13bb
Protection against evaporator freeze-up can be provided by the evaporator temperature controller throttling the cooling tower water flow.[4]

Economizer and Steam Governor

In existing chillers, the sizing of the economizer level control valves (LCV-1 and LCV-2 in Figure 8.13cc) is often based on an assumption. This assumption is that the refrigerant vapor pressure in the condenser ($P3$) is constant and corresponds to a condenser water temperature of 75 or 85°F (23.9 or 29.4°C). Naturally, when such units are operated with 45 or 50°F (7.2 or 10°C) condenser water, $P3$ is much reduced, as is the available pressure differential across LCV-1 and LCV-2.

If this occurs—when the freon circulation rate is high—the control valves will be unable to provide the necessary flow rate, and flooding of the economizer will occur, because the flow is higher and the ΔP is lower than was the basis of valve sizing. The solution is to install larger valves, preferably external, and throttle them using controllers with proportional and integral control modes.

The proportional control mode alone cannot maintain the set point as load changes. The addition of the integral mode is necessary to eliminate the offset. This is important in machines that were not initially designed for optimized, low-temperature condenser water operation, because otherwise the liquid refrigerant that overflows from the flooded evaporator can damage the compressor.

To optimize a steam-turbine-driven compressor, one must be able to modulate its rotational velocity over a reasonably wide range. This is not possible with old, existing machines, if they are provided with quick-opening steam governor valves. When such governors are throttled, a slight increase in lift from the fully closed position results in a substantial

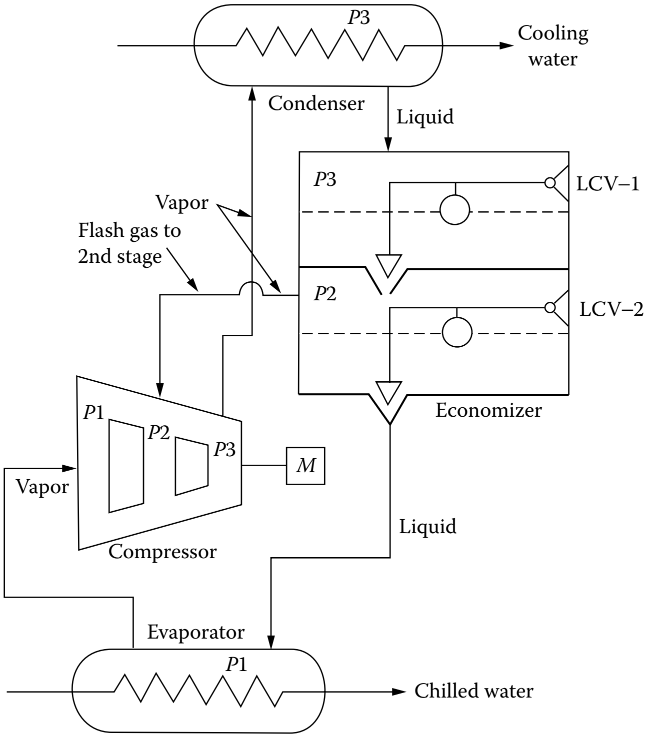


FIG. 8.13cc
To protect against economizer flooding, the internal valves (LCV-1 and LCV-2) can be supplemented with larger external valves.

steam flow and therefore a substantial rotational velocity. Consequently, the valves become unstable and noisy.

The characteristics of the steam governor valves can be changed from quick-opening to linear at minimal cost. The linear characteristics and the desired wide rangeability can be obtained by welding two rings with V-notches to the seats of the existing steam governor valves. This is shown in Figure 8.13dd.

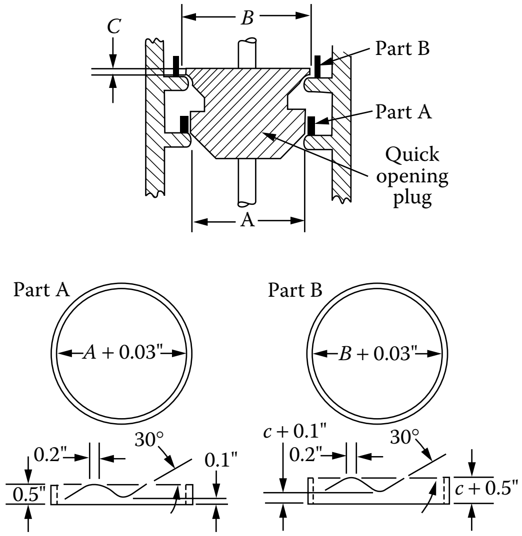


FIG. 8.13dd
A double-seated steam governor valve can be rebuilt to provide the required rangeability for optimized variable-speed service, if two notched rings are installed in the valve seat to modify its characteristics.[4]

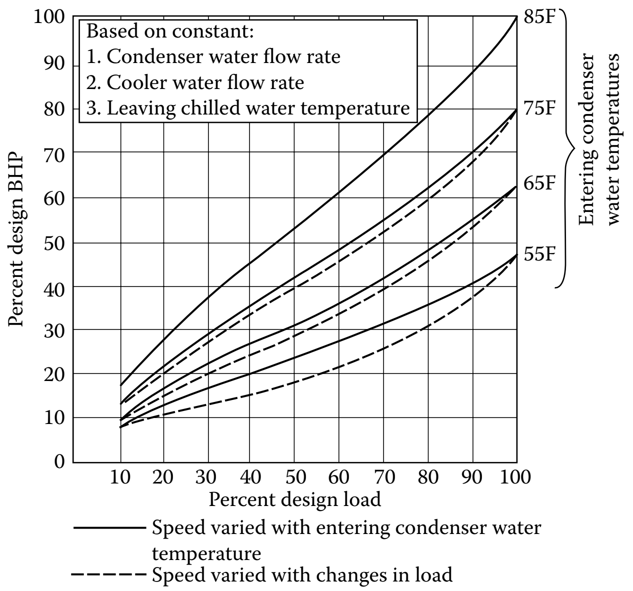


FIG. 8.13ee
The actual power consumption of a variable-speed compressor is reduced by lowering either condenser water temperature or the loading (speed). (Adapted from Reference 9.)

CONCLUSIONS

The goals of chiller optimization include:

- Minimizing the temperature difference across the chiller by minimizing the cooling tower water temperature and maximizing the chilled water temperature
- Minimizing pumping costs by transporting only as much water as is required to meet the load and by using variable-speed pumping
- Minimizing cooling tower operating costs by approach optimization
- Operating chillers at the load at which their efficiency is the highest and by meeting partial loads with part-time operation at maximum efficiency
- Making coolant when least expensive, if night storage is available
- Making load allocation on the basis of maximizing operating efficiency
- Initiating maintenance if a drop in efficiency is detected

As shown in Figure 8.13ee, when using variable-speed compressors, either a reduction in speed or a reduction in condenser water temperature will reduce the power consumption. For example, when the design load is 50% and the condenser water temperature is 55°F, the power consumption can be less than 20%. This compares with 53% if guide vanes are used on a constant-speed machine operating with 85°F condenser water.

The operating costs of chiller stations that operate throughout the year and are located in the northern regions can be cut in half[3] if the potentials of all of the optimization strategies that were discussed in this section are fully exploited. The payback period on investment for chiller optimization controls is usually under 1 year.

With the availability of inexpensive solid-state sensors and microprocessors, it is possible to incorporate all the optimization strategies in a single chiller controller. Such multivariable control software makes it possible to replace the uncoordinated control of flows and temperatures with optimized load following that always corresponds to the lowest possible cost of operation.

References

1. Lipták, B. G., "Cutting in Half the Cost of Cooling," *Control,* November 2003.
2. Lipták, B. G., "Optimizing Controls for Chillers and Heat Pumps," *Chemical Engineering,* October 17, 1983.
3. Romita, E., "A Direct Digital Control for Refrigeration Plant Optimization," *ASHRAE Transaction,* Vol. 83, Part 1, 1977.
4. Lipták, B. G., "Optimizing Plant Chiller Systems," *InTech,* September 1977.
5. Shinskey, F. G., *Energy Conservation Through Control,* New York: Academic Press, 1978.
6. Lipták, B. G., "Save Energy by Optimizing Your Boilers, Chillers, and Pumps," *InTech,* March 1981.
7. Zimmer, H., "Chiller Control Using On-Line Allocation for Energy Conservation," Paper No. 76-522, *Advances in Instrumentation*, Part 1, ISA/76 International Conference, Houston, TX, 1976.
8. Alsenz, R. H., U.S. Patent Numbers 4,612,776 (dated September 23, 1986) and 4,628,700 (dated December 16, 1986).
9. Carrier Corporation, "Centrifugal Liquid Chillers," *Application Data,* 17-IXA, 1975.

Bibliography

ASHRAE Handbook and Product Directory, Systems Volume, Chapter 7, New York: ASHRAE, Inc.,

Backus, A. O., "Energy Savings through Improved Control of Heat Pump Setback," *ASHRAE Journal,* February 1982.

Bower, J. R., "The Economics of Variable-Speed Pumping with Speed Changing Drives," *Pumpworld,* Mountainside, NJ: Worthington Pump.

Brownell, D. L., "Handbook of Applied Thermal Design," London: Taylor and Francis, March 1999.

Carrier System Design Manual, Part 7, Syracuse, NY: Carrier Corporation.

Chiller Control, Installation and Operation Manual, Moeller Refrigeration Products, April 2002.

CIBSE, *Building Control Systems,* Butterworth-Heinemann, August 2000.

Cooper, K. W., and Erth, R. A., "Centrifugal Water Chilling Systems: Focus on Off-Design Performance," *Heating/Piping/Air Conditioning,* January 1978, p. 63.

Cooper, K. W., and Erth, R. A., "Utilizing Water Chillers Efficiently," International Conference on Energy Use Management, Tucson, AZ, Vol. 1, New York: Pergamon Press, 1977, p. 233.

Edwards, T. C., "A New Air Conditioning Refrigeration and Heatpump Cycle," *ASHRAE Transactions,* Vol. 84, Part 2, 1978.

Frayne, C., *Cooling Water Treatment,* Chemical Publishing, January 2001.

Hartman, T., "Ultra-Efficient Cooling with Demand-Based Control," *HPAC Engineering,* December 2001.

Lipták, B. G., "Optimizing Plant Chiller Systems," *InTech,* September 1977.

Lipták, B. G., "Optimizing Controls for Chillers and Heat Pumps," *Chemical Engineering,* October 17, 1983.

Lipták, B. G., "Envelope Optimization," *Control,* November 1998.

Lipták, B. G., "Cutting in Half the Cost of Cooling," *Control,* November 2003.

Luyben, W. L., *Process Modeling, Simulation, and Control,* New York: McGraw-Hill, 1989

McMillan, G., "Good Tuning: A Pocket Guide," *ISA,* June 2000.

Null, H. R., "Heat Pumps in Distillation," *Chemical Engineering Progress,* July 1976.

Polimeros, G., "There Is Energy for Use in Refrigeration Waste Heat," *Specifying Engineer,* January 1980.

Rishel, J. B., "Cut Energy Costs with Improved Controls for Pumping Systems," *Instruments and Control Systems,* August 1976.

Roper, M. A., *Energy Efficient Chiller Control,* BSRIA Ltd., September 2000.

Schwedler, M., "Multiple Chiller System Design and Control," TRANE, 2001.

Shinskey, F. G., "Energy Conservation," *Instruments and Control Systems,* January 1977.

Stair, W. S., "Air Conditioning System," U.S. Patent Numbers 2,715,514 and 2,715,515, issued August 16, 1955.

Turpin, J., "How to Control Chillers," *ES: Engineering Systems,* February 2001.

Werden, R. G., "Heat Pump Gets by with Less Energy by Making Ice Summer and Winter," *Architectural Record,* November 1976.

Zimmer, H., "Chiller Control for Energy Conservation," 1976 ISA Conference, Reprint No. 76-522, Research Triangle Park, NC: Instrument Society of America, 1976.

8.14 Clean-Room Controls and Optimization

B. G. LIPTÁK (1985, 1995, 2005)

INTRODUCTION

Clean-rooms are used for experimental studies of a wide variety. Testing and analysis laboratories, used in the medical, military, and processing industries, are all designed to operate as clean-rooms. The control and optimization of clean-rooms, therefore, is an important subject to many industries.

The traditional method of maintaining air cleanliness is by filtering. Standard HEPA filters are capable of removing up to 99.97% of impurities at 0.3 micrometers, while ULPA filters are efficient up to 99.9995% at 12 micrometers. In this section, besides filtering, other air cleanliness control methods will also be discussed.

The production of semiconductors is one of several processes that require a clean-room environment. Optimization of the clean-room control system will drastically reduce both the energy cost of operation and the cost associated with the manufacturing of off-spec products, because plant productivity is maximized if drafts are eliminated in the clean work aisle. This is because drafts can stir up the dust in the workstation area, and if the dust settles on the product, production losses result.

SEMICONDUCTOR MANUFACTURING

The process, or production area of a typical semiconductor manufacturing laboratory is between 100,000 and 200,000 ft^2 (9,290 and 18,580 m^2). The market value of the daily production from this relatively small area is well over $1 million. Plant productivity is increased if the following conditions are maintained in the process area:

1. No drafts (+0.02 in. H_2O ±0.005 in. H_2O, or 5 Pa ±1.3 Pa)
2. No temperature gradients (72°F ±1°F, or 22°C ±0.6°C)
3. No humidity gradients (35% RH ±3%)
4. No airflow variations (60 air changes/hr ±5%)

Therefore, the goal is to continuously maintain these conditions through the accurate control of these parameters. The secondary goal is to conserve energy. The sensors and control loop configurations required to meet both of these goals will be described in this section.

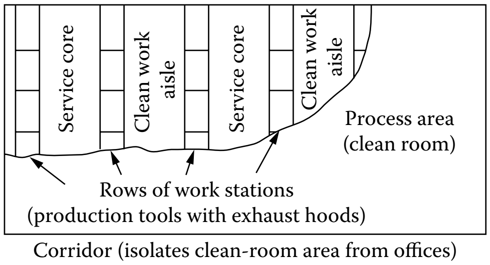


FIG. 8.14a
The configuration of a semiconductor manufacturing clean-room with rows of workstations and a corridor surrounding the whole area.

In order to prevent contamination through air infiltration from the surrounding spaces, the pressure in the clean-room must be higher than that of the rest of the building. As shown in Figure 8.14a, the clean-room is surrounded by a perimeter corridor. In order to protect against the in-leakage of dirty air from the corridor, the clean-room pressure is controlled above it.

The clean-room is made up of rows of workstations. These rows are also referred to as zones. A typical semiconductor producing plant has approximately 200 workstations, which are also called subzones. Each subzone faces the clean work aisle, and behind it is the service core.

Subzone Optimization

As shown in Figure 8.14b, air is supplied to the subzone through filters located in the ceiling in front of the workstation. The total flow and the temperature of this air supply are both controlled. The evacuated exhaust air header is connected to the lower hood section of the workstation. It pulls in all the air and chemical fumes that are generated in the workstation and safely exhausts them into the atmosphere.

In order to make sure that none of the toxic fumes will spill into the work aisle, clean air must enter the workstation at a velocity of about 75 fpm (0.38 m/s). The air that is not pulled in by the exhaust system enters the service core, where some of it is recirculated back into the workstation by a local fan. The rest of the air is returned from the service core by the return air header. A damper in this header (RAR-1) is modulated to control the pressure (DPS-1) in the clean work aisle.

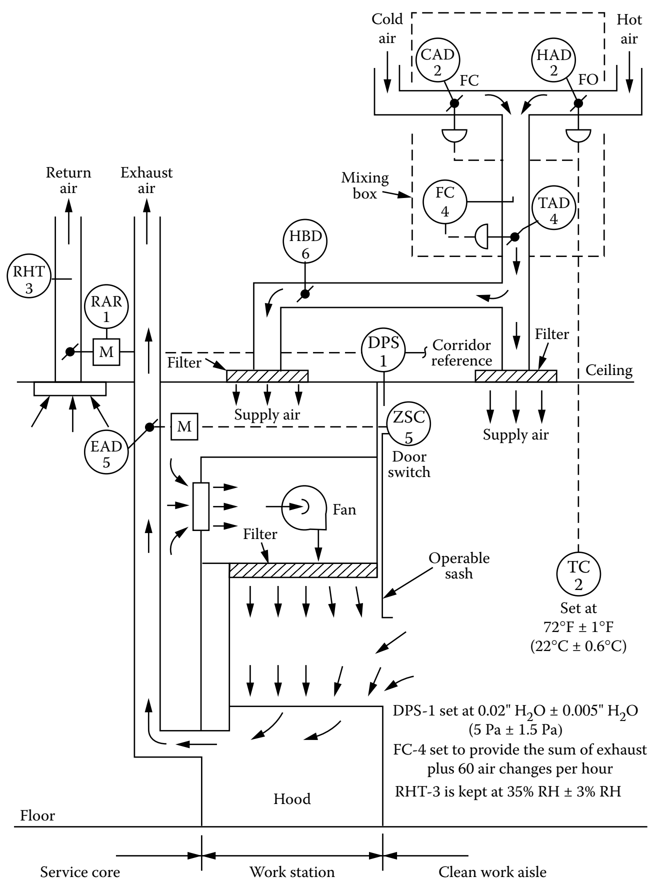


FIG. 8.14b

Workstation controls maintain comfort while guaranteeing that all toxic chemical fumes are removed from the work area.

The filter shown in Figure 8.14b can be a standard HEPA filter, which is efficient up to 99.97% at 0.3 micrometers, or it can be an ULPA filter, which is efficient up to 99.9995% at 12 micrometers.

Pressure Controls The pressure in the corridor is the reference for the clean-room pressure controller DPS-1 in Figure 8.14b. This controller is set to maintain a few hundredths of an inch of positive pressure relative to the corridor. The better the quality of building construction, the higher this set point can be, but even with the lowest quality buildings, a setting of approximately 0.02 in. H_2O (5 Pa) can easily be maintained.

At such near-atmospheric pressures, air behaves as if it were incompressible, so the pressure control loop shown in Figure 8.14b is both fast and stable. When the loop is energized, DPS-1 quickly rotates the return air control damper (RAR-1) until the preset differential is reached. At that point, the electric motor stops rotating the damper, and it stays at its last opening. This position will remain unaltered as long as the air balance in the area remains the same:

$$\begin{aligned}\text{return airflow} = \text{supply airflow} - (&\text{exhaust airflow}\\ &+ \text{pressurization loss})\end{aligned} \tag{8.14(1)}$$

When this airflow balance is altered (for example, as a result of a change in exhaust airflow), it will cause a change in the space pressure, and DPS-1 will respond by modifying the opening of RAR-1.

Elimination of Drafts Plant productivity is maximized if drafts are eliminated in the clean work aisle. Drafts would stir up the dust in this area, which in turn would settle on the product and cause production losses. In order to eliminate drafts, the pressure at each workstation must be controlled at the same value. Doing so eliminates the pressure differentials between stations and therefore prevents drafts. When a DPS-1 unit is provided to control the pressure at each workstation, the result is a uniform pressure profile throughout the clean-room.

PDS-1 usually controls the pressure in the clean work aisle on the "process" side of the workstation. Yet, it is important to make sure that all points, including the service aisle, are under positive pressure. Because the local circulating fan within the workstation draws the air in from the service core and discharges it into the clean work aisle, the pressure in the service core will always be lower than that on the process side (see Figure 8.14b). On the other hand, it is possible for localized vacuum zones to evolve in the service core. This could cause contamination by allowing air infiltration into the service core. To prevent this from happening, several solutions have been proposed.

One possibility is to have DPS-1 control the service core pressure. This is not recommended, because a draft-free process area can be guaranteed only if there are no pressure gradients on the work aisle side; this can only be achieved by locating the DPS-1 units on the work aisle side of the workstations.

Another possibility is to leave the pressure controls on the work aisle side but raise the set point of DPS-1 until all the service cores in the plant are also at a positive pressure. This solution cannot be universally recommended either, because the quality of building construction might not be high enough to allow operation at elevated space pressures. The pressurization loss in badly sealed buildings can make it impossible to reach the elevated space pressure.

Yet another solution is to install a second DPS controller, which would maintain the service core pressure by throttling the damper HBD-6. This solution will give satisfactory performance but will also increase the cost of the control system by the addition of a few hundred control loops (one per workstation).

A more economical solution is shown in Figure 8.14b. Here, a hand-operated bypass damper (HBD-6) is manually set during initial balancing. This solution is reasonable for most applications, because the effect of the workstation fan is not a variable, and therefore a constant setting of HBD-6 should compensate for it.

Temperature Controls The temperature at each workstation is controlled by a separate thermostat, TC-2 in Figure 8.14b.

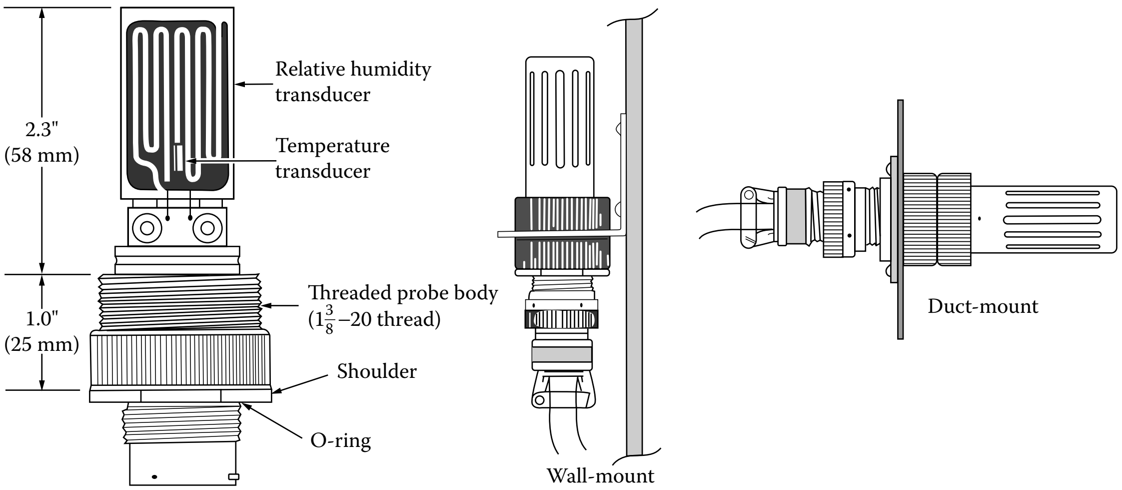


FIG. 8.14c
Polystyrene surface conductivity humidity element. (Courtesy of General Eastern.)

This temperature control adjusts the ratio of cold air to hot air within the supply air mixing box to maintain the space temperature.

Unfortunately, conventional thermostats cannot be used if the temperature gradients within the clean-room are to be kept within ±1°F of 72°F (–0.6°C of 22°C). Conventional thermostats cannot meet this requirement for either measurement accuracy or control quality. Even after individual calibration, it is unreasonable to expect an error of less than ±2 or 3°F (±1 or 1.7°C) in the overall loop performance if HVAC-quality thermostats are used. Part of the reason for this is the fact that the "offset" cannot be eliminated in plain proportional controllers such as thermostats (Section 7.9).

Thus, in order for the thermostat to move its output from the mid-scale value (50–50% mixing of cold and hot air), an error in room temperature must first develop. This error is the permanent offset. The size of this offset error for TC-2 in Figure 8.14b for the condition of maximum cooling can be determined as follows:

$$\text{offset error} = \frac{\text{spring range of CAD}}{2(\text{thermostat gain})} \qquad \mathbf{8.14(2)}$$

Assuming that CAD-2 has an 8–13 PSIG (55–90 kPa) spring (a spring range of 5 PSI, or 34.5 kPa) and that TC-2 is provided with a maximum gain of 2.5 psi/°F (31 kPa/°C), the offset error is 1°F (0.6°C). Under these conditions, the space temperature must permanently rise to 73°F (22.8°C) before CAD-2 can be fully opened. The offset error will increase as the spring range increases or as the thermostat gain is reduced. Sensor and set-point dial error are always additional to the offset error.

Therefore, in order to control the clean-room temperature within ±1°F (±0.6°C), it is necessary to use an RTD-type or a semiconductor transistor-type temperature sensor and a proportional-plus-integral controller, which will eliminate the offset error. This can be most economically accomplished through the use of microprocessor-based shared controllers that communicate with the sensors over a pair of telephone wires that serve as a data highway.

Humidity Controls The relative humidity sensors are located in the return air stream (RHT-3). In order to keep the relative humidity in the clean-room within 35% RH ±3% RH, it is important to select a sensor with an error lower than ±3% RH. The repeatability of most relative humidity sensors (see Section 8.32 in the first volume of this handbook) is approximately ±1% RH.

Figure 8.14c illustrates the polystyrene surface conductivity-type humidity element. These units are available for duct mounting and can be used for clean-room control applications if they are individually calibrated for operation at or around 35% RH. Without such individual calibration, they will not perform satisfactorily, because their off-the-shelf inaccuracy, or error, is approximately ±5% RH.

The controller associated with RHT-3 is not shown in Figure 8.14b, because relative humidity is not controlled at the workstation (subzone) level but at the zone level (as mentioned earlier, a zone is a row of workstations). The control action is based on the relative humidity reading in the combined return air stream from all workstations within that zone.

Flow Controls The proper selection of the mixing box serving each workstation is of critical importance. Each mixing box serves the dual purpose of providing accurate control of the total air supply to the subzone and of modulating the ratio of "cold" and "hot" air to satisfy requirements of the space thermostat TC-2.

The total air supply flow to the subzone should equal 60 air changes per hour plus the exhaust rate from subzone. This total air supply rate must be controlled within ±5% of actual flow by FC-4, over a flow range of 3:1. The rangeability of 3:1 is required because as processes change, their associated exhaust requirements will also change substantially.

FC-4 in Figure 8.14b can be set manually, but this setting must change every time a new tool is added to or removed from the subzone. The setting of FC-4 must be done by individual in-place calibration against a portable hot wire anemometer reference (Sections 2.2 and 2.13 in Volume 1). Settings based on the adjustments of the mixing box alone (without an anemometer reference) will not provide the required accuracy.

Some of the mixing box designs available on the market are not acceptable for this application. Such unacceptable designs include the following:

- "Pressure-dependent" designs, in which the total flow will change with air supply pressure. Only "pressure-independent" designs can be considered, because both the cold and the hot air supply pressures to the mixing box will vary over some controlled minimum.
- Low rangeability designs. A 3:1 rangeability with an accuracy of ±5% of *actual* flow is required.
- Selector or override designs. In these designs either flow or temperature is controlling on a selective basis. Such override designs will periodically disregard the requirements of TC-2 and will thus induce upsets, cycling, or both.

If the mixing box is selected to meet the aforementioned criteria, both airflow and space temperature can be accurately controlled.

Optimization of the Zones

Each row of workstations shown in Figure 8.14a is called a zone, and each zone is served by a cold deck (CD), a hot deck (HD), and a return air (RA) subheader. These subheaders are frequently referred to as *fingers*. The control devices serving the individual workstation (subzones) will be able to perform their assigned control tasks only if the "zone finger" conditions make it possible for them to do so.

For example, RAR-1 in Figure 8.14b will be able to control the subzone pressure as long as the ΔP across the damper is high enough to remove all the return air without requiring the damper to open fully. As long as the dampers are throttling (neither fully open nor completely closed), DPS-1 in Figure 8.14b is in control.

PIC-7 in Figure 8.14d is provided to control the vacuum in the RA finger and thereby to maintain the required ΔP across RAR-1. PIC-7 is a nonlinear controller with a fairly wide neutral band. This protects the CD finger temperature (TIC-6 set point) from being changed until a sustained and substantial change takes place in the detected RA pressure.

Similarly, the mixing box in Figure 8.14b will be able to control subzone supply flow and temperature as long as its dampers are not forced to take up extreme positions. Once a damper is fully open, the associated control loop is out of control. Therefore, the purpose of the damper position controllers (DPCs) in Figure 8.14d is to prevent the corresponding dampers from having to open fully.

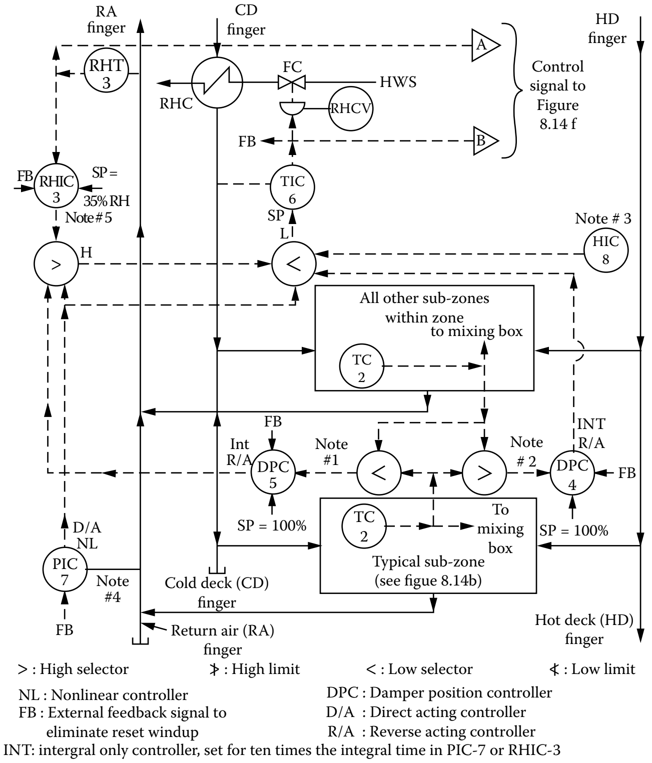


FIG. 8.14d

Typical zone control system that is under envelope control. Note 1: This is the opening of the most-open hot air damper. HAD-2 (shown in Figure 8.14b) which is fully open when the TC-2 output is minimum. Note 2: This is the opening of the most-open cold air damper. CAD-2 (shown in Figure 8.14b) is fully open when the TC-2 output is maximum. Note 3: HIC-8 sets the maximum limit for the CD temperature at approximately 70°F (21°C). Note 4: As vacuum increases (and pressure decreases), the output of PIC-7 also increases. Note 5: The controller action is D/A in summer, R/A in winter.

Lastly, the relative humidity in the return air must also be controlled within acceptable limits. Therefore, at the zone level there are five controlled, or limit variables but only one manipulated variable:

Limit or Controlled Variables	*Manipulated Variable*
RA pressure	TIC-6 set point
RA relative humidity	
Max. CAD opening	
Max. HAD opening	
Max. TIC-6 set point	

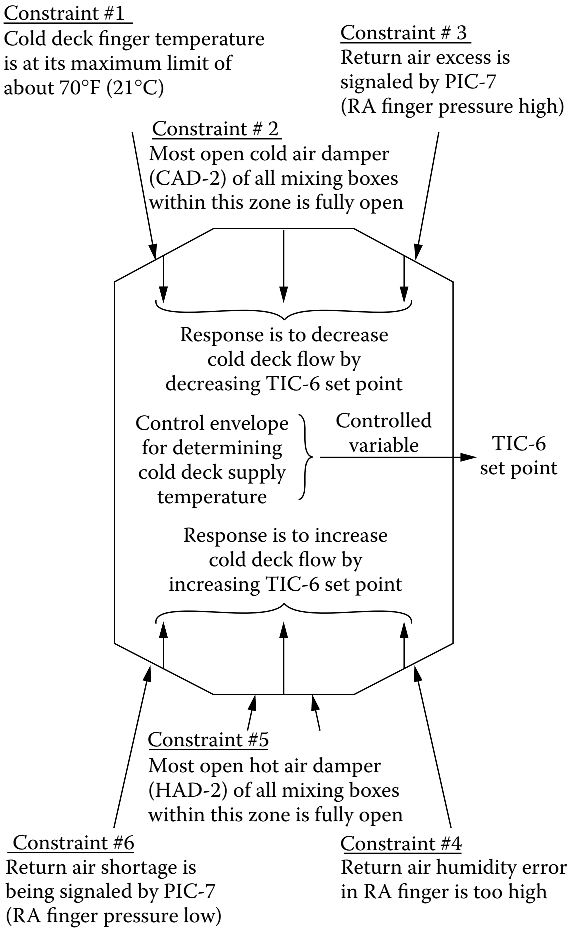


FIG. 8.14e

For optimized control of the cold deck temperature, a constraint envelope-based control system can be used.

Figure 8.14d shows the control loop configuration required to accomplish the aforementioned goals. It should be noted that dynamic lead/lag elements are not shown and that this loop can be implemented in either hardware or software.

Envelope Optimization Whenever the number of control variables exceeds the number of available manipulated variables, it is necessary to apply multivariable envelope control. This means that the available manipulated variable (TIC-6 set point) is not assigned to serve a single task but is selectively controlled to keep many variables within acceptable limits—within the "control envelope" shown in Figure 8.14e.

By adjusting the set point of TIC-6, it is possible to change the cooling capacity represented by each unit of CD air. Because the same cooling can be accomplished by using less air at lower temperatures or by using more air at higher temperatures, the overall material balance can be maintained by manipulating the set point of the heat balance controls.

Return air humidity can similarly be affected by modulating the TIC-6 set point, because when this set point is increased, more CD air will be needed to accomplish the required cooling. Increasing the ratio of humidity-controlled CD air in the zone supply (HD humidity is uncontrolled) also brings the zone closer to the desired 35% RH set point.

In this envelope control system, the following conditions will cause an increase in the TIC-6 set point:

- Return air shortage detected by a drop in the RA finger pressure (increase in the detected vacuum) measured by PIC-7. An increase in the TIC-6 set point will increase the CD demand (opens all CAD-2 dampers in Figure 8.14b), which in turn lowers the HD and, therefore, the RA demand.
- Hot air damper in mixing box (HAD-2 in Figure 8.14b) is near to being fully open. This, too, necessitates the aforementioned action to reduce HD demand.
- RA humidity does not match the RHIC-3 set point. This condition also necessitates an increase in the proportion of the CD air in the zone supply. Because the moisture content of the CD air is controlled, an increase in its proportion in the total supply will help control the RH in the zone.

On the other hand, the following conditions will require a decrease in the TIC-6 set point:

- Return air excess detected by a rise in RA finger pressure (drop in the detected vacuum) measured by PIC-7.
- Cold air damper in mixing box (CAD-2 in Figure 8.14b) is fully open.
- CD finger temperature exceeds 70°F (21°C). This limit is needed to keep the CD always cooler than the HD.

Plantwide Optimization

In conventionally controlled semiconductor plants, both the temperature and the humidity of the main CD supply air are fixed. This can severely restrict the performance of such systems, because as soon as a damper or a valve is fully open (or closed), the conventional system is out of control.

The optimized control system described here does not suffer from such limitations, because it automatically adjusts both the main CD supply temperature and the humidity so as to follow the load smoothly and continuously, never allowing the valves or dampers to lose control by fully opening or closing. The net result is not only increased productivity but also reduced operating costs.

A semiconductor production plant might consist of two dozen zones. Each of the zones can be controlled as shown in Figure 8.14d. The total load represented by all the zones is followed by the controls shown in Figure 8.14f. The overall control system is hierarchical in its structure: The subzone controls in Figure 8.14b are assisted by the zone controls in

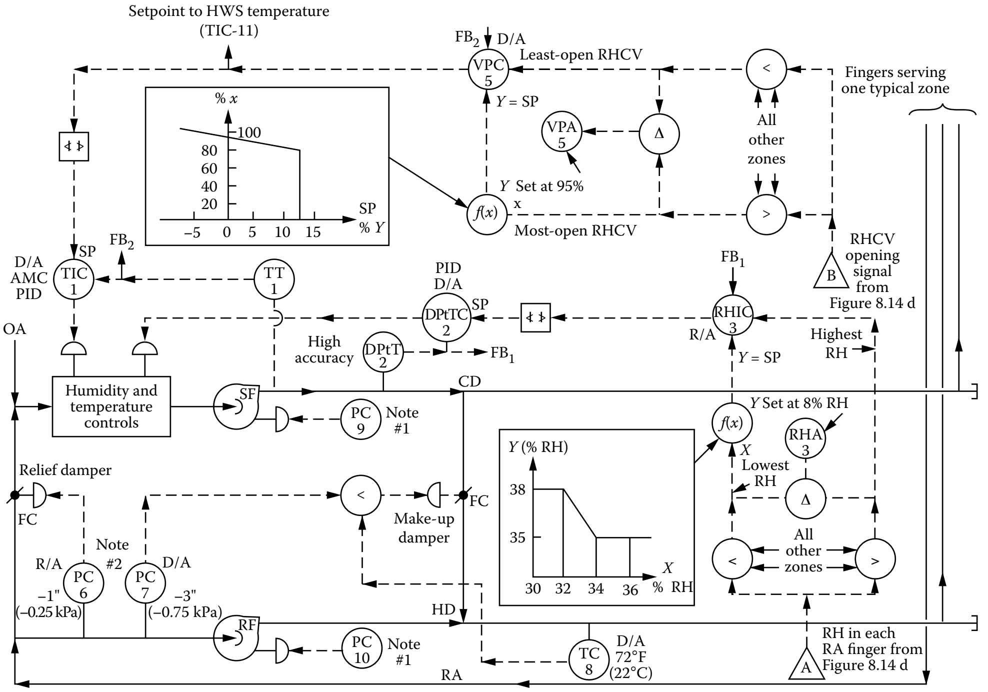


FIG. 8.14f

The plantwide control system handles the total load, the sum of the loads of all the zones in an optimized manner. Symbols: AMC = Controller with positions for automatic-manual-cascade modes of operation. DPtIC = Dew point controller. INT = Controller with integral control action only. Its setting is to be 10 times the integral time setting of the associated process controller(s). PID = Controller with proportional, integral, and derivative control modes. VPC = Valve position controller. Note 1: This pressure controller keeps the lowest of all finger pressures at a value above some minimum. It modulates the fan volume or speed and starts extra fan units to meet the load. When the load drops, the unnecessary fans are stopped by flow (not pressure) control. Note 2: If the pressure rises above −1 in. H_2O (−0.25 kPa), the relief damper starts to open, and if the pressure drops below −3 in. H_2O (−0.75 kPa), the makeup damper opens. This opening is limited by TC-8, which prevents the HD temperature from dropping below 72° (22°C). Between the settings of PC-6 and PC-7, both dampers are tightly closed and the suction pressure floats.

Figure 8.14d, and the plantwide controls in Figure 8.14f guarantee that the zone controllers can perform their tasks.

The interconnection among the levels of the hierarchy is established through the various valve and damper position controllers (VPCs and DPCs). These guarantee that no throttling device, such as mixing boxes or RHC valves, anywhere in this plant will ever be allowed to reach an extreme position. Whenever a control valve or a damper is approaching the point of losing control (nearing full opening), the load-following control system at the next higher level modifies the air or water supply conditions so that the valve or damper in question will not need to open fully. The hierarchical control scheme is depicted in Figure 8.14g.

The overall plantwide control system shown in Figure 8.14f can be viewed as a flexible combination of material balance and heat/humidity balance controls. Through the load-following optimization of the set points of TIC-1 and TIC-11, the heat balance controls are used to assist in maintaining the material balance around the plant. For example, if the material balance requires an increase in airflow and the heat balance requires a reduction in the heat input to the space, both requirements will be met by admitting more air at a lower temperature.

Material Balance Controls The plantwide material balance is based on pressure control. PC-9 and PC-10 modulate the variable-volume fans so as to maintain a minimum supply pressure in each of the CD and HD fingers. The suction side

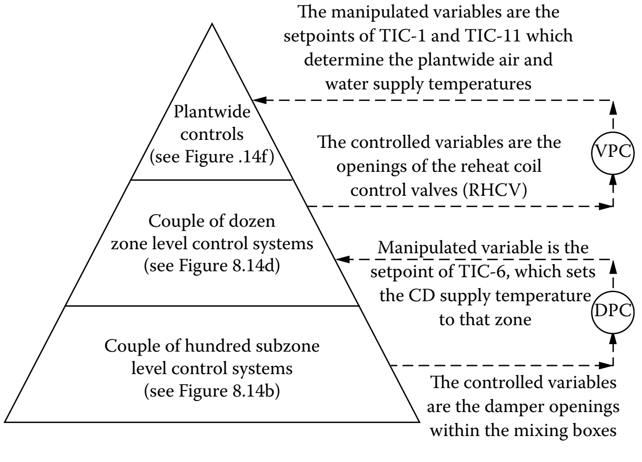


FIG. 8.14g

The overall control architecture in a semiconductor manufacturing plant is hierarchical in structure.

of the CD supply fan station is open to the outside and will draw as much outside air as the load demands (Figure 8.14f).

The suction pressure of the HD supply fan station is an indication of the balance between RA availability and HD demand. This balance is maintained by PIC-7 in Figure 8.14d, at the zone level. Because this controller manipulates a heat transfer system (RHCV), it must be tuned for slow, gradual action. This being the case, it will not be capable of responding to sudden upsets or to emergency conditions, such as the need to purge smoke or chemicals.

Such sudden upsets in material balance are corrected by PC-6 and PC-7 in Figure 8.14f. PC-6 will open a relief damper if the suction pressure is high (vacuum is low), and PC-7 will open a makeup damper if it is low (vacuum is high). In between these limits, both dampers will remain closed and the suction pressure will be allowed to float.

Heat Balance Controls Heat balance also requires a multivariable envelope control system (similar to that shown in Figure 8.14e), because the number of controlled variables exceeds the number of available manipulated variables. Therefore, the plantwide air and water supply temperature set points (TIC-1 and TIC-11) are not adjusted as a function of a single consideration but are selectively modulated to keep several variables within acceptable limits inside the control envelope (Figure 8.14e).

By adjusting the set points of TIC-1 and TIC-11, it is possible to adjust the cooling capacity represented by each unit of CD supply air and the heating capacity of each unit of hot water. This then provides not only for load-following control but also for minimization of the operating costs. Cost reductions are accomplished by minimizing the HWS temperature and by minimizing the amount of simultaneous cooling and reheating of the CD air.

This control envelope is so configured that the following conditions will decrease the TIC set points (both TIC-1 and TIC-11):

1. The least-open RHCV is approaching full closure.
2. CD supply temperature is at its maximum limit.

On the other hand, the following conditions will require an increase in the TIC set points:

1. The most-open RHCV is approaching full opening.
2. The CD supply temperature is at its minimum limit.

The set point of VPC-5 in Figure 8.14f is produced by a function generator, $f(x)$. The dual purposes of this loop are 1) to prevent any of the reheat coil control valves (RHCVs) from fully opening and thereby losing control, and 2) to force the least-open RHCV toward a minimum opening, so as to minimize wasteful overlap between cooling and reheat. All this is done while keeping any of the RHCVs from fully closing as long as possible.

These purposes are accomplished by $f(x)$, which keeps the least-open valve at approximately 12% opening as long as the most-open valve is less than 90% open. If this 90% opening is exceeded, $f(x)$ prevents the most-open valve from fully opening. It does so by lowering the set point of VPC-5, which in turn will increase the set points of the TICs.

Increasing the set point of TIC-1 reduces the load on the most-open RHCV, whereas increasing the TIC-11 set point increases the heating capacity at the reheat coil. Through this method of load-following and through the modulation of main supply air and water temperatures, all zones in the plant will be kept under stable control if the loads are similar in each zone.

Mechanical Design Limitations All control systems — including this one—will lose control when the design limits of the associated mechanical equipment are reached or if the dissimilarity in load distribution is greater than that which was anticipated by the mechanical design. Therefore, if a condition ever arises in which one zone requires large amounts of cooling (associated RHCV closed) while at the same time some other zone requires its finger reheat coil valve to be fully open, then the control system cannot control both.

In such a case, the control system must decide which condition it is to correct. The control system in Figure 8.14f is so configured that it will give first priority to preventing the RHCV valve from fully opening. Therefore, if as a result of mechanical design errors or because of the misoperation of the plant there is no CD supply temperature that can keep all valves from fully closing or from fully opening, then the control system will allow some RHCV valves to close while preventing all of them from fully opening.

Whenever such a condition is approaching (that is, when the difference between the opening of the most-open and the least-open valves reaches 95%), a valve position alarm (VPA-5) is actuated. This allows the operator to check the causes of such excessively dissimilar loads between the zones and to take corrective action by revising processes, relocating tools, modifying air supply ducts, or adding or removing mixing boxes.

By keeping all RHCV valves from being nearly closed, this control system simultaneously accomplishes the following goals:

1. Eliminates unstable (cycling) valve operation by not allowing valves to operate nearly closed
2. Minimizes pumping costs by minimizing pressure losses through throttling valves
3. Minimizes heat pump operating costs by minimizing the required hot water temperature
4. Provides a means of detecting and thereby smoothly following the variations in the plantwide load

Humidity Controls At the zone level, the RA humidity is controlled by RHIC-3 in Figure 8.14d. In addition, the dew point of the main CD air supply must be modulated. The main CD supply temperature is already being modulated to follow the load. This is accomplished by measuring the relative humidity in all RA fingers (RHT-3 in Figure 8.14d) and selecting the one finger that is farthest away from the control target of 35% RH.

The control system is similar to the TIC set point optimization loop discussed earlier. The first elements in the loop are the selectors shown in Figure 8.14f. They pick out the return air fingers with the highest and the lowest relative humidity readings. The purpose of the loop is to "herd" all RH transmitter readings in such a direction that both the highest and the lowest readings will fall within the acceptable control gap limits of 35% ±3% RH. This is accomplished by sending the highest reading to RHIC-3 in Figure 8.14f as its measurement and also sending the lowest reading to this RHIC-3 as its modified set point.

A humidity change in either direction can be recognized and corrected through this herding technique. The set point of RHIC-3 is produced by the function generator, $f(x)$. Its dual purposes are to prevent the most humid RA finger from exceeding 38% RH and to keep the driest RA finger humidity from dropping below 32% RH, as long as this can be accomplished without violating the first goal.

These purposes are accomplished by keeping the set point of RHC-3 at 35% as long as the driest finger reads 34% RH or more. If it drops below that value, the set point is raised to the limit of 38% RH in order to overcome this low-humidity condition, without allowing excessive humidity in the return air of some other zone.

This control strategy and the RHIC-3 controller in Figure 8.14d will automatically respond to seasonal changes and will give good RH control as long as the loads in the various zones are similar. If the humidity loads are substantially dissimilar, this control system is also subject to mechanical equipment limitations. In other words, the mechanical equipment is so configured that addition or removal of moisture is possible only at the main air supply to the CD. Consequently, if some zones are moisture-generating and others require humidification, this control system can respond to only one of these needs.

Therefore, if the lowest RH reading is below 32% while the highest is already at 38%, the low-humidity condition will be left uncontrolled, and the high-humidity zone will be controlled to prevent it from exceeding the 38% RH upper limit. Allowing the minimum limit to be temporarily violated while maintaining control on the upper limit is a logical and safe response to humidity load dissimilarities between zones. It is safe because the intermixing of the return airs will allow self-control of the building by transferring moisture from zones with excess humidity to ones with humidity deficiency.

Whenever the difference between minimum and maximum humidity readings reaches 8% RH, the RHA-3 alarm is actuated. This will alert operators that substantial moisture load dissimilarities exist so that they can seek out and eliminate the causes.

Exhaust Air Controls The exhaust air controls at the subzone level are shown in Figure 8.14b. At this level, the control element is a two-position damper, EAD-5. When the workstation is functioning, its operable sash is open. This condition is detected by the door position limit switch, ZSC-5, which in that case fully opens the two-position damper. When the workstation is out of service, its operable sash is closed, and therefore the door switch closes EAD-5 to its minimum position.

This minimum damper position still provides sufficient air exhaust flow from the workstation to guarantee the face velocity needed for operator safety. In other words, when the sash is closed, the air will enter the workstation over a smaller area; therefore, less exhaust airflow is required to provide the air inflow velocity that is needed to keep the chemical fumes from leaking out. Through this technique, the safety of operation is unaffected, and operating costs are lowered because less outside air needs to be conditioned if the exhaust airflow is lowered.

In order for EAD-5 to maintain the required exhaust air flow accurately, it is necessary to keep the vacuum in the exhaust air collection ductwork at a constant value. This pressure control loop is depicted in Figure 8.14h. In each zone exhaust finger, the PC-1 shown in Figure 8.14h keeps the vacuum constant by throttling the EAD-1 damper as required.

In order for these dampers to stay in control while the exhaust airflow varies, it is important to keep them from fully opening. This is accomplished first by identifying the most-open finger damper (low selector) and then by comparing its opening with the set point of the damper position controller, DPC-2.

The task of this controller is to limit the opening of the most-open EAD-1 to 80% and to increase the vacuum in the main EA header if this opening would otherwise exceed 80%. Therefore, if the measurement of DPC-2 exceeds 80%, the vacuum set point of PC-3 is increased (pressure setting lowered). This in turn will increase the operating level of the exhaust fan station (EAF).

The limits on the set point of DPC-2 will prevent mechanical damage, such as collapsing of the ducts as a result of excessive vacuum or manual misoperation. The reasons for integral action and external feedback are the same as in the case of all other damper and valve position controllers discussed earlier.

Figure 8.14h also shows a glycol-circulating heat recovery loop that can be used as shown to preheat the entering outside air or that can be used as a heat source to a heat pump in the winter (not shown). In either case, the operating cost of the plant in the winter is lowered by recovering the heat content of the air before it is exhausted.

The discharging of chemical vapors into the atmosphere is regulated by pollution considerations. The usual approach is to remove most of the chemicals by adsorption and scrubbing prior to exhaustion of the air. An added measure of safety is provided by exhausting the air at high velocity, so

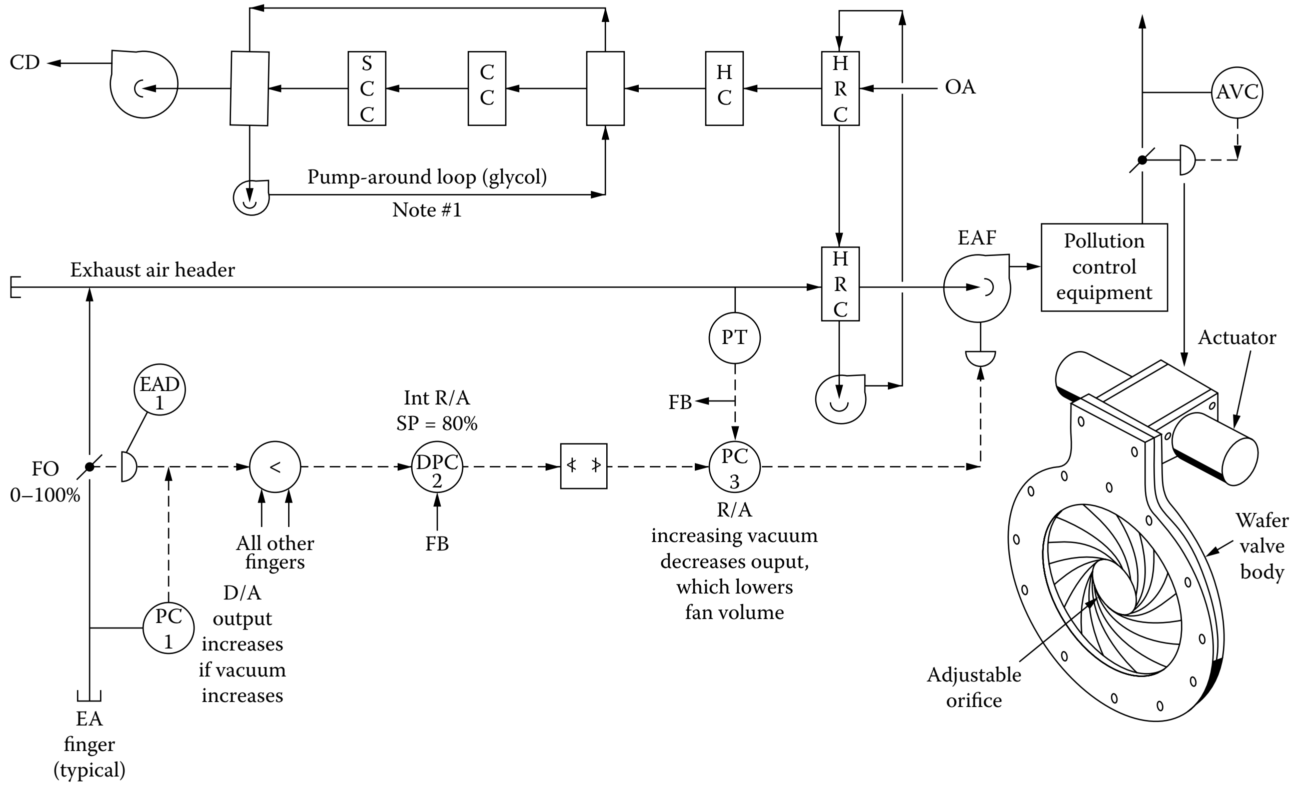


FIG. 8.14h

Exhaust air optimization controls are used to keep the vacuum in the exhaust air collection ductwork at a constant value and to maximize heat recovery. Note 1: When the need for dehumidification in the summer results in excessive overcooling of the CD air supply, and therefore in increased demand for reheat at the CD fingers, the pump-around economizer loop is started to lower operating cost. It also increases efficiency by reducing the degree of overcooling, thereby lowering the need for reheat.

as to obtain good dispersion in the atmosphere. Because the volume of air being exhausted varies, an air velocity controller (AVC in Figure 8.14h) is used to maintain the velocity of discharge constant. This is accomplished by modulation of a variable orifice iris damper, also illustrated in Figure 8.14h.

CONCLUSIONS

The productivity of semiconductor manufacturing plants can be greatly increased and the operating costs can be lowered through the control and optimization methods described in this section and by such added control system features as:

- Low leakage dampers (0.5 CFM/ft^2 at 4 in. H_2O ΔP, or 2.5 1/s/m^2 at 1 kPa ΔP)
- Accurate airflow metering for material balance and pressurization loss control
- Pump-around economizers (see Figure 8.14h)

The initial cost of the previously described control system is not higher than the cost of conventional systems, because the added expense of more accurate sensors is balanced by the reduced installation cost of distributed shared controls. Therefore, the benefit of increased productivity is a result not of higher initial investment but of better control system design.

It should be emphasized that the described load-following optimization envelope control strategy is far from being typical of today's practices in the semiconductor manufacturing industry. In fact, it has never been fully implemented, and many plants are still operated under conventional HVAC controls.

Therefore, the conclusion that yields will increase and operating costs will drop when such improved control systems are installed is an assumption. This appears to be supported by the results of partial system implementations that the author implemented at IBM plants and by the experiences in other industries but will remain an assumption until a semiconductor manufacturing plant that includes all the features of such modern controls is in operation.

Bibliography

ACHIH, *Air Sampling Instrumentation of Atmospheric Contaminants,* 9th ed., Cincinnati, OH: American Conference of Governmental Industrial Hygienists (ACHIH), 2001.

AIHA, *Workplace Environmental Exposure Level Guide Series on Toxic Chemicals,* American Industrial Hygiene Association (AIHA), 2002.

Carlson, D. R., “Temperature Measurement in Process Control,” *InTech,* October 1990.

Cooper, F. G., “Low Pressure Sensing and Control,” *InTech,* September 1992.

Daryanani, S., “Design Engineer’s Guide to Variable Air-Volume Systems,” *Actual Specifying Engineer,* July 1974.

DHO/Atlanta Corp., “Conserving Energy,” *Powered Induction Unit, Data Sheet #1,* May 1, 1974.

Kerin, T. W., and Katz., E. M., “Temperature Measurement in the 1990s,” *InTech,* August 1990.

Kusuda, T., “Intermittent Ventilation for Energy Conservation,” ASHRAE Symposium #20, Paper #4.

Lipták, B. G. (Ed.), “Carbon Dioxide Sensors,” in *Environmental Engineer’s Handbook,* Vol. II, Section 4.6, Radnor, PA: Chilton Books, 1975.

Lipták, B. G., “Reducing the Operating Costs of Buildings by the Use of Computers,” *ASHRAE Transactions,* Vol. 83, Part 1, 1977.

Lipták, B. G., “Savings through CO_2-Based Ventilation,” *ASHRAE Journal,* July 1979.

Lipták, B. G., “Save Energy by Optimizing Your Boilers, Chillers, and Pumps,” *Instrumentation Technology,* March 1981.

Nordeen, H., “Control of Ventilation Air in Energy Efficient Systems,” ASHRAE Symposium #20, Paper #3.

Shih, J. Y., “Energy Conservation and Building Automation,” ASHRAE Paper #2354.

Spielvogel, L. G., “Exploding Some Myths about Building Energy Use,” *Architectural Record,* February 1976.

Stillman, R. B., *Systems Simulation Engineering Report,* prepared for IBM-RECD, December 10, 1971.

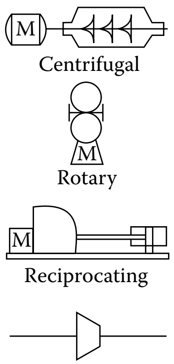
Warmoth, D., “Temperature Measurement and Control,” *Measurements and Control,* February 1986.

Woods, J. E., “Impact of ASHRAE Ventilation Standards 62–73 on Energy Use,” ASHRAE Symposium #20, Paper #1.

Yan, W. B., and Mallon, T., "Breakthrough Technology Measures Parts-per-Trillion Moisture in Gases," *Chem. Res*., March 2002.

8.15 Compressor Control and Optimization

F. B. HOROWITZ (1970) **B. P. GUPTA** (1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

INTRODUCTION

The transportation of vapors and gases is an important portion of the total operating cost of processing plants. The optimization of this unit operation can substantially lower the total operating cost of the plant. The goal of this section is to describe the strategies available to increase the safety and energy efficiency in these systems. Prior to the discussion of state-of-the-art advanced controls, the basic equipment and the conventional control strategies will be described.

Three types of compressors will be discussed: rotary, reciprocating, and centrifugal compressors. The emphasis will be on the centrifugal units, whose load, surge, and override controls will be described in some detail. Multiple compressor installations will also be described.

Compressors are gas transportation machines that perform the function of increasing the gas pressure by confinement or by kinetic energy conversion. Methods of capacity control for the principal types of compressors are listed in Table 8.15a. The method of control to be used is a function of process requirements, type of the driver, and cost considerations.

TABLE 8.15a
Capacity Control Methods Used on the Different Compressor Types

Compressor Type	*Capacity Control Method*
Centrifugal	Suction throttling Discharge throttling Variable inlet guide vanes Speed control
Rotary	Bypassing Speed control
Reciprocating	On/off control Constant-speed unloading Speed control Speed control and unloading

Because the driver constitutes half of the cost of the compressor installation, careful selection must be made in order to ensure trouble-free performance. Variable speed control can be accomplished by the use of steam turbines, gas turbines, or gasoline or diesel engines. Electric motors are well suited for both constant- and variable-speed applications.

The flow and discharge pressure ranges of the various compressor designs are shown in Figure 8.15b.

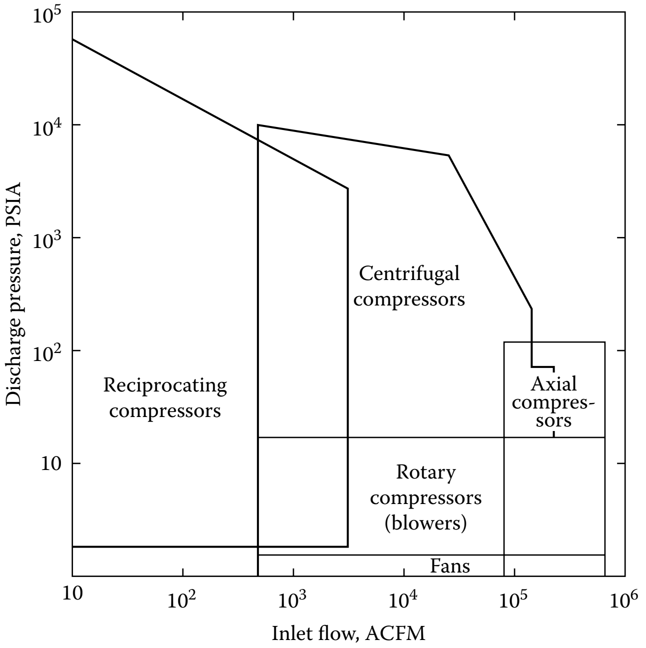
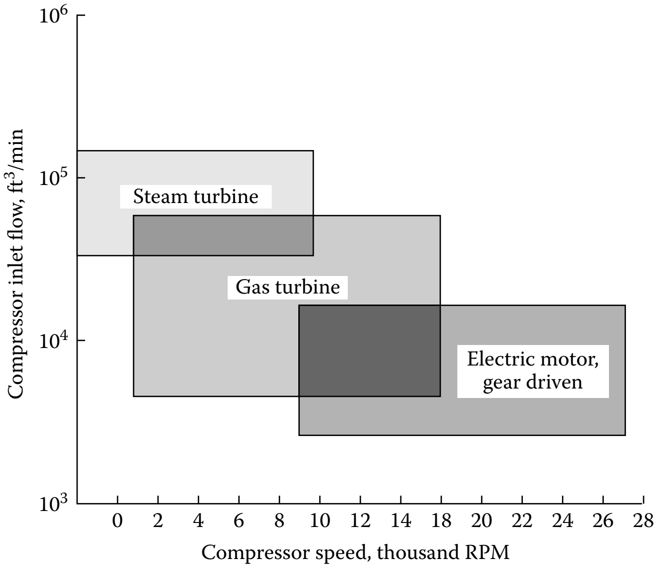


FIG. 8.15b
The range of flow capacities and discharge pressures that the different compressor designs can handle.[1]



Speed Governor Classification

Governor class	*Regulation*	*Variation*
A	10%	¾%
B	6%	½%
C	4%	¼%
D	½%	¼%

FIG. 8.15c
The speed and flow capacity ranges of standard compressor drives (top) and the classification of speed governors (bottom).

The operating ranges of the most popular drives are shown in Figure 8.15c. Also given in the lower part of the figure is the definition of some two-speed governors. Regulation error is the percentage of the difference in speed between the speed at zero power output and at the rated output. Variation error is the percentage variation that can be expected around the speed set point.

CENTRIFUGAL COMPRESSORS

The Compression Process

The centrifugal compressor is a machine that converts the momentum of gas into a pressure head.

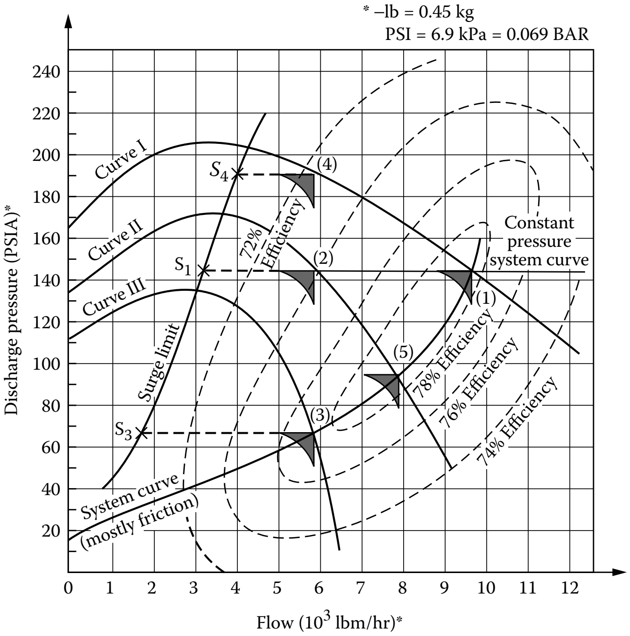
$$H = \frac{\tau\omega}{w} = \frac{n}{n-1} ZRT_I, \qquad [(P_D/P_I)^{\frac{n}{n-1}} - 1] \qquad \mathbf{8.15(1)}$$

Equation 8.15(1) is the basis for plotting the compressor curves and for understanding the operation of capacity controls. The nomenclature for the symbols in the equation is as follows:

h = differential head (ft or m)
H = polytrophic compressor head (ft or m)
$K_{1,2,3}$ = flow constant
m = mass flow (lb/hr or kg/hr) = $c\sqrt{h\zeta}$
n = polytrophic coefficient
P_D = discharge pressure (psia or Pa)
P_I = inlet pressure (psia or Pa)
Q = volume flow rate (ACFH or m^3/hr) = $c\sqrt{h\zeta}$
R = gas constant
T_1 = inlet temperature
u = rotor tip speed (ft/s or m/s)
W = weight flow (lbm/hr or kg/hr)
Z = gas compressibility factor
τ = motor torque (ft lb_f or J)
ψ = head coefficient
ω = angular velocity (radians/hr)
ζ = density of gas (lb/ft^3 or kg/m^3)

Characteristic Curves

In Equation 8.15(1), the pressure ratio (P_D/P_I) varies inversely with mass flow (W). For a compressor running at constant speed (ω), constant inlet temperature (T_1), constant molecular weight (implicit in R), and constant n, τ, and Z, the discharge pressure may be plotted against weight flow as in Figure 8.15d



*For SI units see Section A.1

FIG. 8.15d
The operating point is located where the centrifugal compressor curves cross the system curve of the process. The system curve can be a constant pressure one (horizontal line), a mostly friction one, or any other.

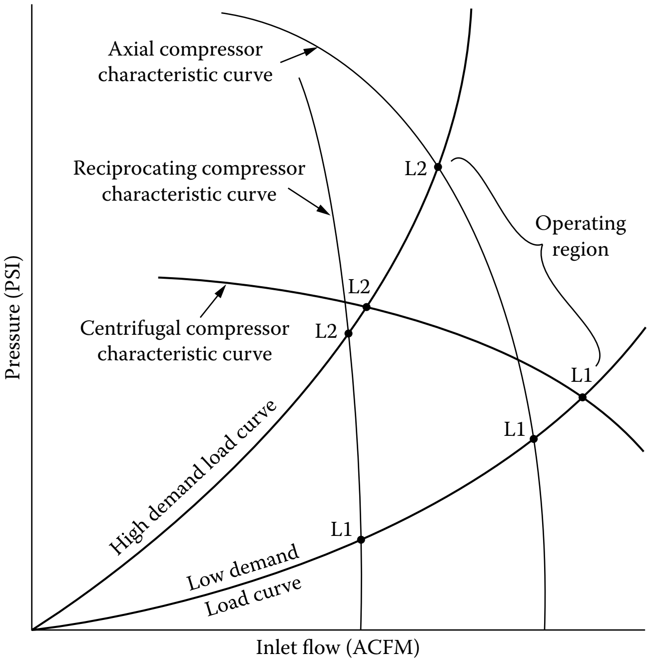


FIG. 8.15e
The characteristic curve of the axial compressor is steep, while that of the centrifugal compressor is flat. (Adapted from Reference 1.)

(curve I). The design point (1) is located in the maximum efficiency range at design flow and pressure.

Positive-displacement compressors pressurize gases through confinement. Dynamic compressors pressurize them by acceleration. The axial compressor moves the gas parallel to the shaft. In the case of the centrifugal compressor, the gas receives a radial thrust toward the wall of the casing where it is discharged.

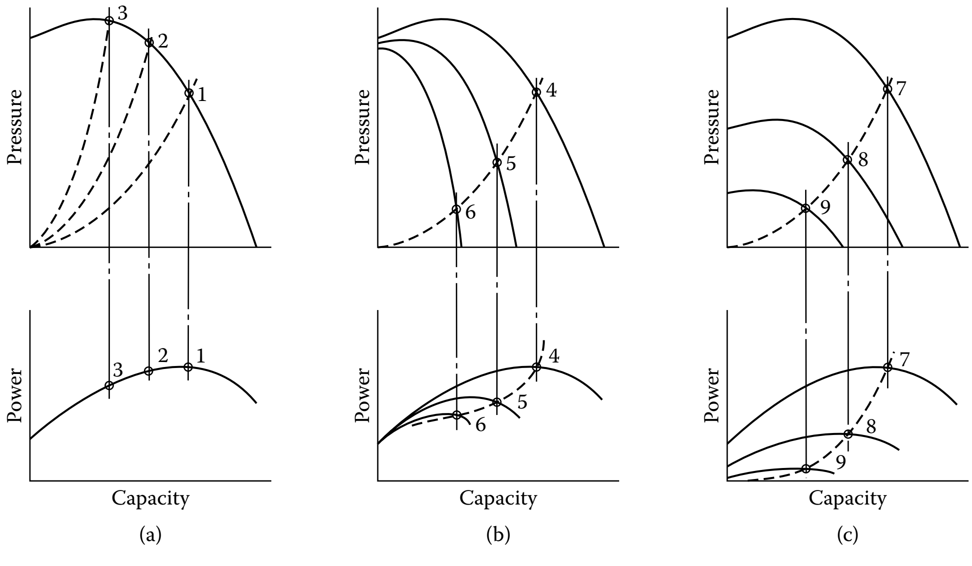
The axial compressor is better suited for constant flow applications, whereas the centrifugal design is more applicable for constant pressure applications. This is because the characteristic curve of the axial design is steep, and that of the centrifugal design is flat (Figure 8.15e). The characteristic curve of a compressor plots its discharge pressure as a function of flow, and the load curve relates the system pressure to the system flow. The operating points (*L*1 or *L*2 in Figure 8.15e) are the intersections of these curves. The normal operating region falls between the low and the high demand load curves in Figure 8.15e.

Axial compressors are more efficient; centrifugal ones are better suited for dirty or corrosive services.

Compressor Throttling

Compressor loading can be reduced by throttling a discharge or a suction valve, by modulating a prerotation vane, or by reducing the speed. As is shown in Figure 8.15f, discharge throttling is the least energy efficient and speed modulation is the most energy efficient method of turndown. Suction throttling is a little more efficient and gives a little better turndown than discharge throttling, but it is still a means of wasting that transporting energy that should not have been introduced in the first place.

Guide vane positioning, which provides prerotation or counter-rotation to the gas, is not as efficient as speed modulation, but it does provide the greatest turndown. As is shown in Figure 8.15f, speed control is the most efficient, as small



(a) (b) (c)

FIG. 8.15f
The efficiencies of discharge throttling (left), suction throttling (center), and variable speed control (right).

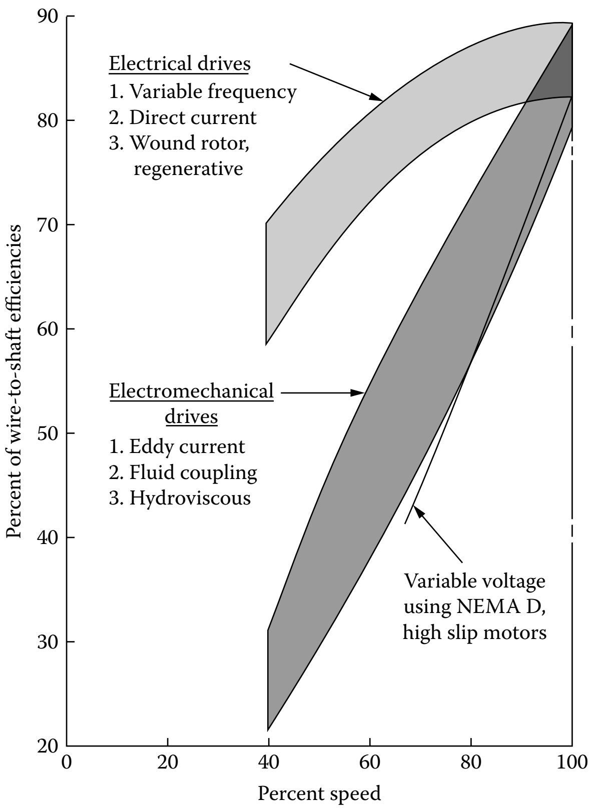


FIG. 8.15g
Wire-to-shaft efficiencies of electrical drives (top) are higher than those of the electromechanical drives.[2]

speed reductions result in large power savings because of the cubic relationship between speed and power.

If the discharge pressure is constant, flow tends to vary linearly with speed. If the discharge head is allowed to vary, it will change with the square of flow and, therefore, with the square of speed as well. This square relationship between speed and pressure tends to limit the speed range of compressors to the upper 30% of their range.[3]

Constant-speed steam turbine governors can be converted into variable-speed governors by revising the quick-opening characteristics of the steam valve into a linear one. The efficiency of variable-speed drives varies substantially with their design (Figure 8.15g). Electric governors tend to eliminate the dead bands that are present in mechanical designs. They also require less maintenance because of the elimination of mechanical parts. Electric governors also give better turndowns and are quicker and simpler to interface with surge or computer controls.

The error in following the load increases as the speed of process disturbances increase or as the speed control loop speed is reduced. Therefore, it is desirable to make the speed loop response as fast as possible. On turbines, this goal is served by the use of hydraulic actuators, and the motor response is usually increased by the use of tachometer feedback.

For the purposes of the control systems shown in this chapter, it will be assumed that the compressor throughput is controlled through speed modulation with tachometer feedback.

Suction Throttling One can control the capacity of a centrifugal compressor by throttling a control valve in the suction line, thereby altering the inlet pressure (P_I). From Equation 8.15(1) it can be seen that the discharge pressure will be altered for a given flow when P_I is changed, and a new compressor curve will be generated. This is illustrated in Figure 8.15d (curves II and III).

Consider first that the compressor is operating at its normal inlet pressure (following curve I) and is intersecting the "constant pressure system" curve at point (1) with a design flow of 9600 lbm/hr (4320 kg/hr) at a discharge pressure of 144 psia (1 MPa) and 78% efficiency. If it is desired to change the flow to 5900 lbm/hr (2655 kg/hr) while maintaining the same discharge pressure, it would be necessary to shift the compressor from curve I to curve II.

The new intersection with the "constant pressure system" curve is at the new operating point (2), at 74% efficiency. In order to shift from curve I to curve II, one must change the discharge pressure of 190 psia (1.3 MPa) at the 5900 lbm/hr (2655 kg/hr) flow on curve I to 144 psia (1 MPa) on curve II. If the pressure ratio is 10 (P_D/P_I), then it would be necessary to throttle the suction by only $\Delta P_I = 46/10 = 4.6$ psi (32 kPa) to achieve this shift.

It is also important to consider how close the operating point (2) is to the surge line. The surge line represents the low-flow limit for the compressor, below which its operation will become unstable as a result of momentary flow reversals. Methods of surge control will be discussed later in this section. At point (2) the flow is 5900 lbm/hr (2655 kg/hr), and at the surge limit (S_1) it is 3200 lbm/hr (1440 kg/hr). Thus, the compressor is operating at 5900/3200 = 184% of surge flow. This may be compared with curve I at point (1), where prior to suction throttling the machine was operating at 9600/3200 = 300% of surge flow.

The same method of suction throttling may be applied in a "mostly friction system" also shown in Figure 8.15d. In order to reduce the flow from 9600 lbm/hr (4320 kg/hr) to 5900 lbm/hr (2655 kg/hr), it is necessary to alter the compressor curve from curve II to III, so that the intersection with the "mostly friction system curve" is at the new operating point (3), at 77% efficiency.

In order to do this, one must change the discharge pressure from 190 psia (1.3 MPa—on curve I) to 68 psia (0.5 MPa—on curve III). Thus, $\Delta P_D = 190 - 68 = 122$ psi (0.8 MPa), and the amount of inlet pressure throttling for a machine with a compression ratio of 10 is $\Delta P_I = 122/10 = 12.2$ psi (84 kPa). The corresponding surge flow is at 1700 lbm/hr (765 kg/hr), which means that the compressor is operating at 5900/1700 = 347% of surge flow. Therefore,

TABLE 8.15h

Compressor Performance Parameters as a Function of Throttling Method

	Control Valve ΔP(PSI)	*Compressor Efficiency*	*Operation Above Surge By*
Suction throttling "constant pressure system"	4.6	74%	184%
Suction throttling "mostly friction system"	12.2	77%	347%
Discharge throttling "mostly friction system"	122	72%	148%

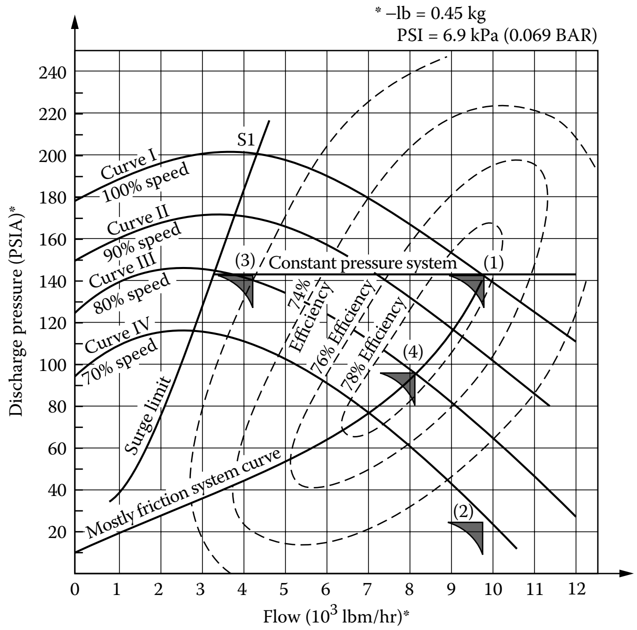
surge is less likely in a "mostly friction system" than in a "constant-pressure system" under suction throttling control.

Discharge Throttling A control valve on the discharge of the centrifugal compressor may also be used to control its capacity. In Figure 8.15d, if the flow is to be reduced from 9600 lbm/hr (4320 kg/hr) at point (1) to 5900 lbm/hr (2655 kg/hr), the compressor must follow curve I and therefore operate at point (4), at 190 psia (1.3 MPa) discharge pressure and 72% efficiency. However, the "mostly friction system" curve at this capacity requires only 68 psia (0.5 MPa) and the rest of the pressure is wasted through the discharge control valve. The surge flow (at S_4) is 4000 lbm/hr (1800 kg/hr), and the compressor is therefore operating at 5900/4000 = 148% of surge. Thus, surge is more likely to occur in a mostly friction system when discharge throttling is used than when suction throttling is used.

Table 8.15h compares the valve pressure drops, efficiencies, and surge margins of suction and discharge throttling.

Inlet Guide Vanes This method of control uses a set of adjustable guide vanes on the inlet to one or more of the compressor stages. By prerotation or counter-rotation of the gas stream relative to the impeller rotation, the stage is unloaded or loaded, thus lowering or raising the discharge head. The effect is similar to suction throttling as illustrated in Figure 8.15d (curves II and III), but less power is wasted because pressure is not throttled directly. Also, the control is two-directional, since it may be used to raise as well as to lower the band. It is more complex and expensive than throttling valves but may save 10 to 15% on power and is well suited for use on constant-speed machines in applications involving wide flow variations.

The guide vane effect on flow is more pronounced in constant discharge pressure systems. This can be seen in Figure 8.15d (curve II), where the intersection with the "constant pressure system" at point (2) represents a flow change from the normal design point (1) of 9600 − 5900 = 3700 lbm/hr (1665 kg/hr). The intersection with the "mostly friction system" at point (5) represents a flow change of only 9600 − 7800 = 1800 lbm/hr (810 kg/hr).



*For SI units see Section A.1

FIG. 8.15i

Control of centrifugal compressor capacity by speed variation.

Variable Speed The pressure ratio developed by a centrifugal compressor is related to the tip speed in the following manner:

$$\psi u^2/2g = \frac{ZRT_I}{(n-1/n)}[(P_D/P_I)^{(n-1/n)} - 1] \qquad \mathbf{8.15(2)}$$

From this relation the variation of discharge pressure with speed may be plotted for various percentages of design speed, as shown in Figure 8.15i. The obvious advantage of speed control from a process viewpoint is that both suction and discharge pressures can be specified independently of the flow.

The normal flow is shown at point (1) for 9700 lbm/hr (4365 kg/hr) at 142 psia (0.98 MPa). If the same flow is desired at a discharge pressure of 25 psia (173 kPa), the speed is reduced to 70% of design, shown at point (2). In order to achieve the same result through suction throttling with a pressure ratio of 10:1, the pressure drop across the valve would have to be (142 – 25)/10 = 11.7 psi (81 kPa), with the attendant waste of power, as a result of throttling. This is in contrast with a power saving accomplished with speed control, because power input is reduced as the square of the speed.

One disadvantage of speed control is apparent in constant-pressure systems, in which the change in capacity may be overlay sensitive to relatively small speed changes. This is shown at point (3), where a 20% speed change gives a flow change of (9600 − 4300)/9600 = 55%. The effect is less pronounced in a "mostly friction system," in which the flow change that results from a 20% speed change at point (4) is (9600 − 8100)/9600 = 16%.

Surge Control

The design of compressor control systems is not complete without consideration of surge control, because it affects the stability of the machine. Surging begins at the positively sloped section of the compressor curve. In Figure 8.15i this occurs at S_1 on the 100% speed curve at 4400 lbm/hr (1980 kg/hr). If the flow never drops below this limit, that will ensure safe operation for all speeds, but some power will be wasted at speeds below 100% because the surge limit decreases at reduced speed.

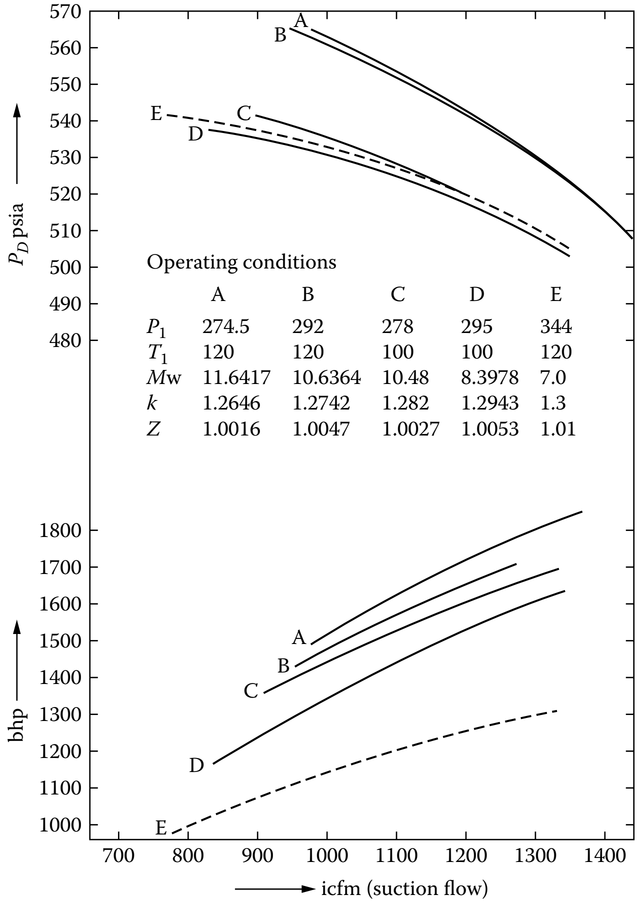
Even for a compressor running at a constant speed, the surge point changes as the thermodynamic properties vary at the inlet. This is shown in Figure 8.15j. Although inaccurate control of the surge point can put the compressor into deep surge, a conservatively set surge point results in useless recycling and wasted energy.

Various schemes to control surge are outlined in the following paragraphs. These include:

1. Compressor pressure rise ($\Delta P = P_D - P_I$) vs. differential across suction flow meter (h)
2. Pressure ratio (P_D/P_I) vs. actual volumetric flow (Q)
3. Break horsepower vs. mass flow (m)
4. Pressure ratio (P_D/P_I) vs. Mach number squared
5. Incipient surge
6. Surge spike detection

The Phenomenon of Surge In axial or centrifugal compressors, the phenomenon of momentary flow reversal is called surge. During surging, the compressor discharge pressure drops off and then is reestablished on a fast cycle. This cycling, or surging, can vary in intensity from an audible rattle to a violent shock. Intense surges are capable of causing complete destruction of compressor parts, such as blades and seals.

The characteristic curves of compressors are such that at each speed they reach a maximum discharge pressure as the flow drops (Figure 8.15k). A line connecting these points (*A* to *F*) is the surge line. If flow is further reduced, the pressure generated by the compressor drops below that which is already existing in the pipe, and momentary flow reversals occur. The frequency of these oscillations is between 0.5 and 10 Hz. The surge frequency of most compressor installations in the processing industries is slightly less than 1 Hz.[5] Surge is usually preceded by a stall condition, which is caused by localized flow oscillations around the rotor at frequencies of 50–100 Hz.



	A	B	C	D	E
P_1	274.5	292	278	295	344
T_1	120	120	100	100	120
*M*w	11.6417	10.6364	10.48	8.3978	7.0
k	1.2646	1.2742	1.282	1.2943	1.3
Z	1.0016	1.0047	1.0027	1.0053	1.01

FIG. 8.15j

Discharge pressure generated and break horsepower used at various suction flows by a typical centrifugal compressor, under five different operating conditions.

At the beginning of surge, the total flow drops off within 0.05 seconds, and then it starts cycling rapidly at a period of less than 2 sec.[1] This period is usually shorter than that of the flow control loop, which controls the capacity of the compressor. If the flow cycles occur faster than the control loop can respond to them, this cycling will pass through undetected as uncontrollable noise. Therefore, fast sensors and instruments are essential for this loop.

As is shown in Figure 8.15k, the surge line is a parabolic curve on a plot of pressure rise (discharge pressure minus suction pressure) vs. flow. This function shows as increasing nonlinearity as the compression ratio increases. If the surge line is plotted as *Pd* − *Pi* vs. the square of flow (orifice differential = *h*), it becomes a straight line (Figure 8.15l) if the compression ratio is low (less than 4:1).

On a plot of ΔP vs. volumetric flow (Q), the following changes will reduce the safety margin between the operating

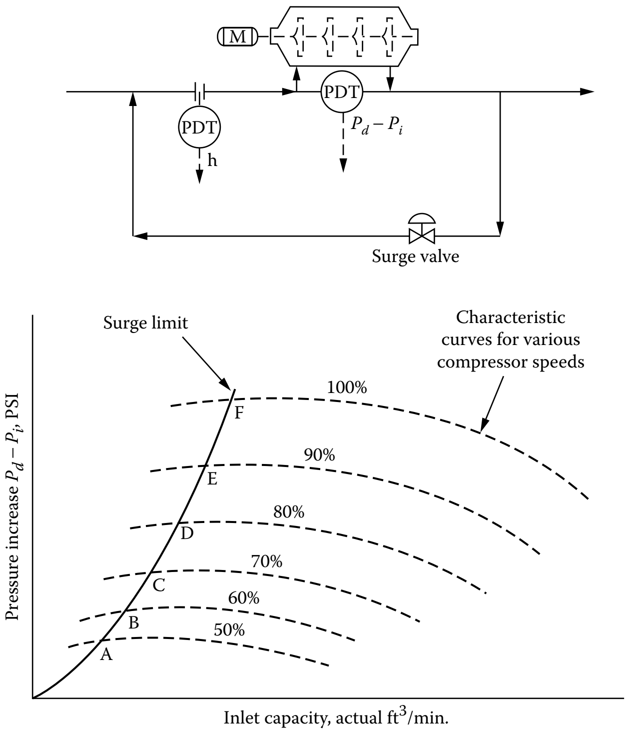


FIG. 8.15k

The location of the surge control valve (top) and the parabolic surge line of a speed-controlled centrifugal compressor (bottom).[4]

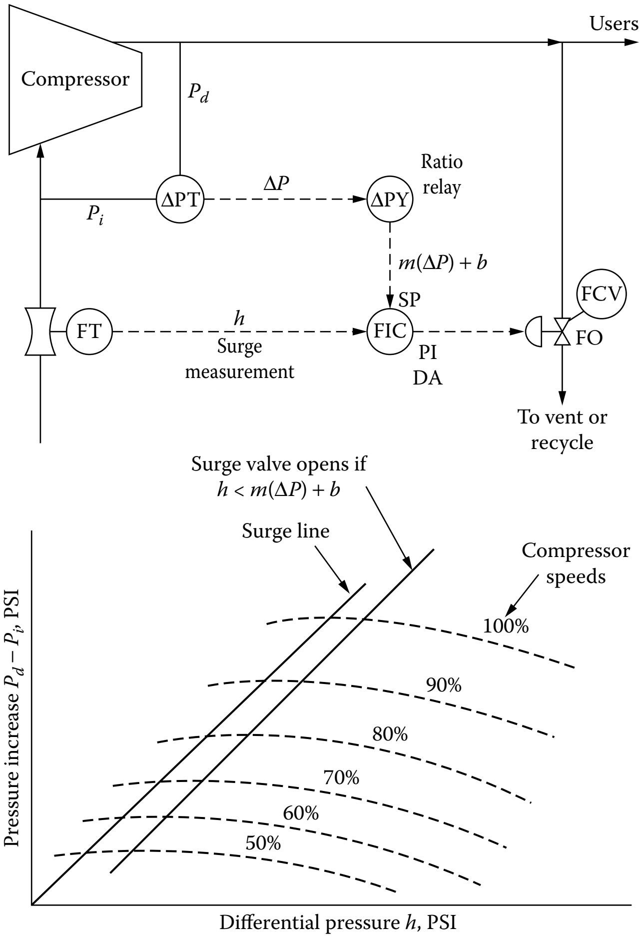


FIG. 8.15l

On a plot of ΔP vs. h, the surge curve becomes a straight line. (Adapted from Reference 4.)

point and the surge line: 1) a decrease in suction pressure; 2) an increase in suction temperature; 3) a decrease in molecular weight; 4) a decrease in specific heat ratio. These conditions will also increase the probability of surge (Figure 8.15k).

On a plot of ΔP vs. h (Figure 8.15l), these effects are most favorable. A decrease in suction pressure moves the surge line in the safe direction, temperature has no effect, and the effect of the other variables is also less pronounced. Therefore, although the ΔP vs. h plot is accurate only at low compression ratios, it does have the advantage of being independent from the effects of composition and temperature changes. However, suction pressure should be included in the model in order to be exact; most ΔP vs. h plots disregard it.

Variations in the Surge Curve Figure 8.15k shows the surge and speed characteristic curves of a centrifugal compressor; Figure 8.15m shows these curves for an axial compressor. The characteristic curves of the axial compressor are steeper, which makes it better for constant-flow services. The centrifugal design is better for constant pressure control.

The effect of guide vane throttling is also shown for both centrifugal and axial compressors (Figure 8.15n). As can be seen, the shape of the surge curve varies with the type of equipment used. The surge curve of speed-controlled centrifugal compressors bends up (Figure 8.15k), whereas for axial (Figure 8.15m) and vane-controlled machines (Figure 8.15n), the surge lines bend over. It is this negative slope of the axial compressor's surge curve that makes it sensitive to speed variations, because an increase in speed at constant flow can quickly bring the unit into surge.

As can be seen from the above information, the shape of the surge curve varies with compression ratio and with equipment design. It should also be noted that in case of multistage compressors, the surge line is discontinuous. If the compressor characteristics are as shown in Figure 8.15o, the addition of a compressor stage causes a break point in the surge curve. With more stages, more break points would also be added, and the resulting net effect is a surge curve that bends over instead of bending up, as does the surge curve for the single-stage compressor in Figure 8.15k.

Figure 8.15o also shows the choke curve. This curve connects the points at which the compressor characteristic lines become vertical. Below this curve, flow will stay constant

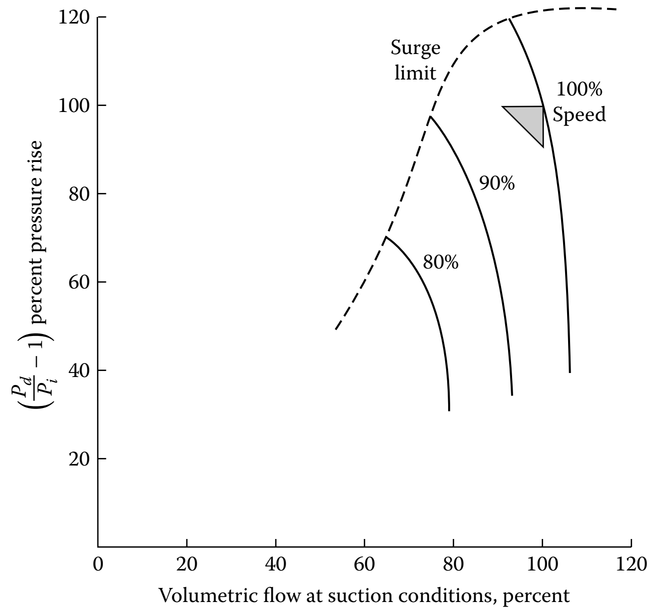


FIG. 8.15m
The surge line for an axial compressor is steep at low flows and flat at high flows.[3]

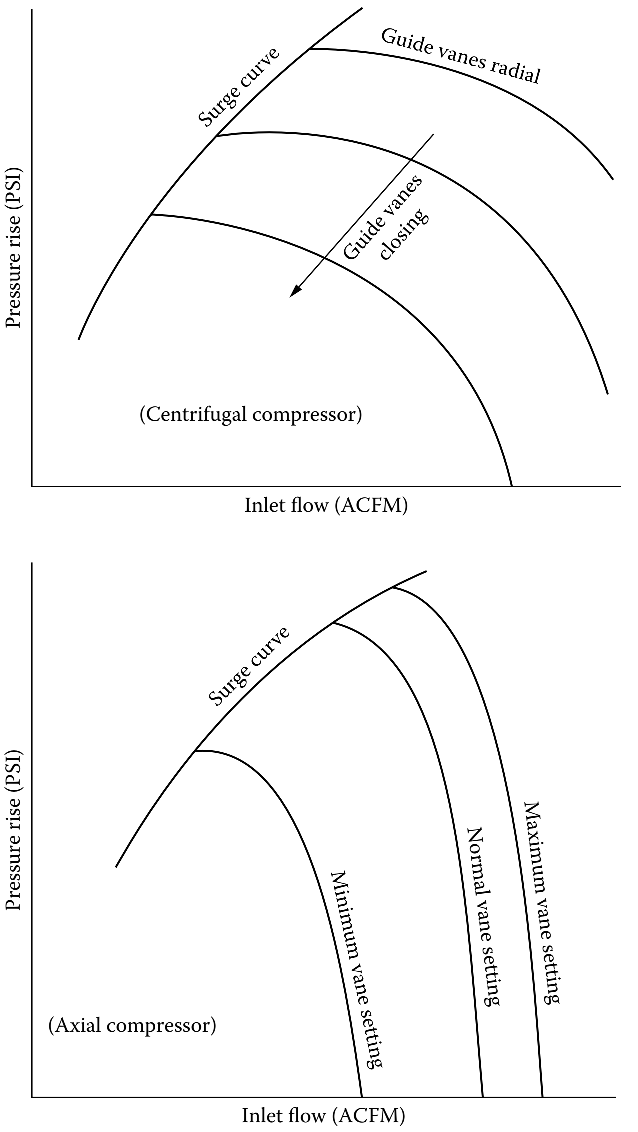


FIG. 8.15n
When guide vanes are used for throttling, both the centrifugal and the axial compressor's surge curve bends over.[1]

even if pressure varies, as long as the compressor is at a constant speed.[1] As speed is reduced, the surge and choke lines intersect. Below this intersection, the traditional methods of surge protection (venting, recycling) are ineffective; only a quick raising of the compressor speed can bring the machine out of surge.

This is similar to the situation when the compressor is being started up. As the operating point is moving on the load curve (Figure 8.15p), it must pass through the unstable region on the left of the surge curve as fast as possible in order to avoid damage from vibration.

Flow Measurement The two critical components of a surge control loop are the flow sensor and the surge valve. Both must be fast, accurate, and reliable.

Flow oscillations under surge conditions occur on a cycle of little more than a second. The flow transmitter should be fast enough to detect these. The time constants of various transmitter designs are as follows:[1]

Pneumatic with damping: Up to 16 sec
Electronic *d/p*: 0.2–1.7 sec
Diffused silicone *d/p*: Down to 0.005 sec

Only the diffused silicone-type sensor design is fast enough to follow the precipitous flow drop that occurs at the beginning of surge or the oscillations during surge. Measurement noise is another serious concern, because it necessitates a greater margin between the surge and the control lines. Noise can be minimized by the use of 20 pipe diameters of upstream and 5 diameters of downstream straight runs around the streamlined flow tube-type sensor. Noise will also be reduced if the low-pressure tap of the *d/p* cell is connected to a piezometric ring in the venturi-type flow tube (Figure 8.15q). The addition of straightening vanes will also contribute to the reduction of noise.

Antisurge control usually requires a flow sensor on the suction side of the compressor. If good, noise-free flow measurement cannot be obtained on that side, a corrected discharge side differential pressure reading (*hd*) can be substituted. Equation 8.15(3) can be used to obtain the suction side differential pressure (*hs*) from readings of (*hd*) and the suction plus discharge pressures and temperatures (*Ps*, *Pd*, *Ts*, and *Td*):

$$hs = hd(Pd/Ps)(Ts/Td) \qquad \mathbf{8.15(3)}$$

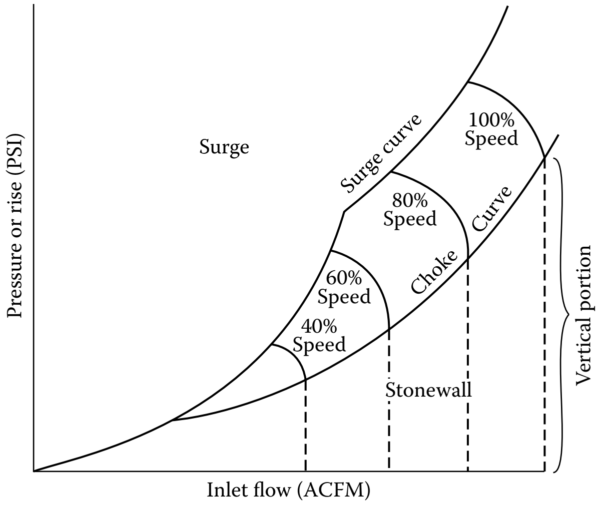


FIG. 8.15o
With multistage compressors, operation must be confined to the area between the surge and choke curves.[1]

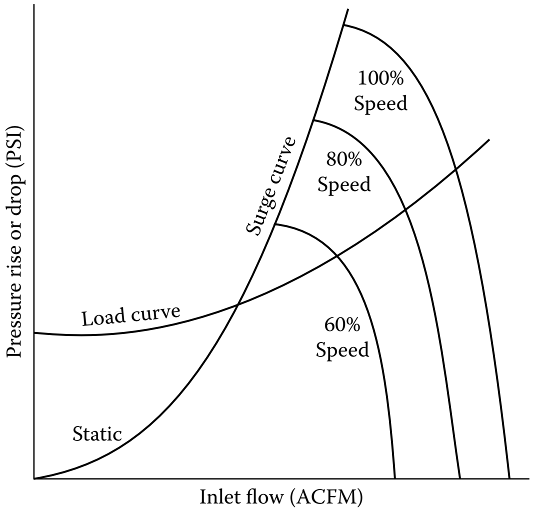


FIG. 8.15p
The speed and surge curves are both characteristics of the compressor, while the load curve is a function of the process load.[1]

Some of the performance characteristics of a number of flow tubes and flow nozzles are given in Table 8.15r.

Surge Control Valves Surge valves are usually fail-open, linear valves that are tuned for fast and precise throttling. Because the valve is the weak link in the total surge protection system, and because testing and maintenance cannot be done on-line if only one valve is used, total redundancy is recommended. Each of the two parallel surge valves should be sized for the full flow of the compressor but for only 70% of the discharge pressure. This pressure reduction is caused by the flow reversals during surge.

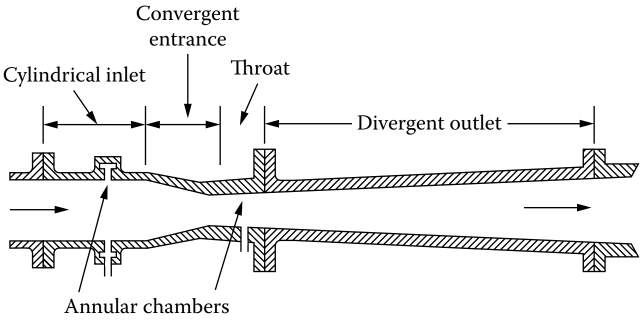


FIG. 8.15q
Herschel venturi with annular pressure chamber.

TABLE 8.15r
Venturi, Flow Tube, and Flow Nozzle Inaccuracies (Errors) in Percent of Actual Flow for Various Ranges of Beta Ratios and Reynolds Numbers

Flow Sensor		*Line size, inches (1 in. = 25.4 mm)*	*Beta Ratio*	*Pipe Reynolds Number Range for Stated Accuracy*	*Inaccuracy, Percent of Actual Flow*
Herschel standard[1]	Cast	4–32	0.30–0.75	2×10^5 to 1×10^6	±0.75%
	Welded	8–48	0.40–0.70	2×10^5 to 2×10^6	±1.5%
Proprietary true venturi[2]	Cast	2–96	0.30–0.75	8×10^4 to 8×10^6	±0.5%
	Welded	1–120	0.25–0.80	8×10^4 to 8×10^6	±1.0%
Proprietary flow tube[3]	Cast	3–48	0.35–0.85	8×10^4 to 1×10^6	±1.0%
ASME flow nozzles[4]		1–48	0.20–0.80	7×10^6 to 4×10^7	±1.0%

[1] No longer manufactured because of long laying length and high cost.
[2] Badger Meter Inc.; BIF Products; Fluidic Techniques Inc.; Primary Flow Signal Inc.; Tri-Flow Inc.
[3] ABB Instrumentation; Badger Meter Inc.; BIF Products; Preso Industries.
[4] BIF Products; Daniel Measurement and Control.

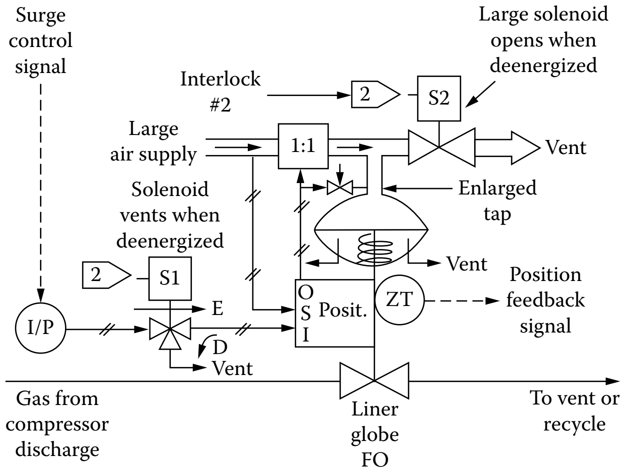


FIG. 8.15s
A well-designed surge control valve will have a positioner, a large air supply with a one-to-one booster repeater with bypass, vent solenoids, and position feedback. (Adapted from Reference 1.)

Globe valves (Section 6.19) are preferred to rotary designs because of hysteresis and breakaway torque considerations. Boosters (Figure 8.15s) can cut the sum of prestroke dead time and full-scale stroking to under 1 sec. Actually, valves as large as 18 in. have been made to throttle to any position in under a half-second.[1]

Digital valves (Section 6.18) are even faster: They can stroke in 0.1 sec without any overshoot. Their limitations are in the plugging of the smaller ports and in the difficulty of inspecting them.

Positioners are frequently required to reduce hysteresis (packing friction), dead band (boosters), disc flutter (butterfly valves), and overshoot; they are also often needed just to provide the higher air pressure required by piston actuators.

Figure 8.15s shows some of the main components of a surge control valve.[1] The electronic control signal is converted to a pneumatic one by the I/P converter, and this pneumatic signal is sent to the positioner through a three-way solenoid ($S1$). The output signal from the positioner is sent simultaneously into a booster relay (1:1) and an adjustable bypass. The output from the booster relay is connected by short- and large-diameter tubes to the valve actuator with an enlarged pressure tap and to the large venting solenoid ($S2$). Some of the operational features of the system are as described below.

When interlock #2 is de-energized, requiring instantaneous full opening of the surge valve (because of discharge line blockage or other reasons), $S1$ vents the positioner inlet and $S2$ vents the actuator top-work. Because $S2$ is large, air removal is fast and the valve opens quickly. On return to normal, both solenoids are energized. This closes $S2$ and opens $S1$ to the control signal.

Depending on the size or restriction in the line from $S1$, air might enter the system more slowly and the valve might close more slowly than it opened. Such fast-opening, slow-closing designs can respond to the first precipitous drop in flow and thus can prevent the second surge cycle from developing. It is important not to slow the speed of valve opening too much, because quick throttling is still required.

The valve must also respond quickly to the control signal and throttle quickly in either direction. In order to speed up the air movement into and out of the actuator, the vent and signal ports on the actuator are both drilled out, an air booster is installed, and a large-capacity path is established between the booster and the actuator. In order to reduce the dead band of the 1:1 booster, it is advisable to use a signal range of 6–30 PSIG instead of the usual 3–15 PSIG.

The bypass needle valve around the booster in Figure 8.15s is required because without it the volume on the outlet of the positioner would be much smaller than on the outlet of the booster. This would allow the positioner to change the input to the booster faster than the booster could change its output, resulting in a limit cycle. This limit cycle is eliminated by the addition of the adjustable bypass.

All surge valves should be throttle-tested before shipment. It is also desirable to monitor the surge valve opening through the use of a position transmitter.

Surge Control Curves As was shown in Figure 8.15l, a parabolic surge curve with a positive slope will appear as a straight line on a ΔP vs. h plot. The purpose of surge control is to establish a surge control line to the right of the actual surge line, so that corrective action can be taken before the machine goes into surge. Such a control system is shown in Figure 8.15t.

The biased surge control line is implemented through a biased ratio relay (ΔPY), which generates the set point for FIC as follows:

$$\mathrm{SP} = m(\Delta P) + b \qquad \mathbf{8.15(4)}$$

where

SP = desired value of h in inches H_2O
m = slope of the surge line at the operating point
ΔP = compressor pressure rise in psi
b = bias of the surge set point in inches H_2O

The offset between the surge and the control lines should be as small as possible for maximum efficiency, but it must be large enough to give time to correct upsets without violating the surge line.[6] The slower the upsets and the faster the control loop, the less offset is required for safe operation. In a good design, the bias b is about 10%, but in bad ones, as disturbances get faster or instrument slower, it can grow to 20%.

A second line of defense is the backup interlock FSL. It is normally inactive, as the value of h is normally above the FIC set point SP. When the surge controller FIC is not fast enough to correct a disturbance, h will drop below SP. FSL

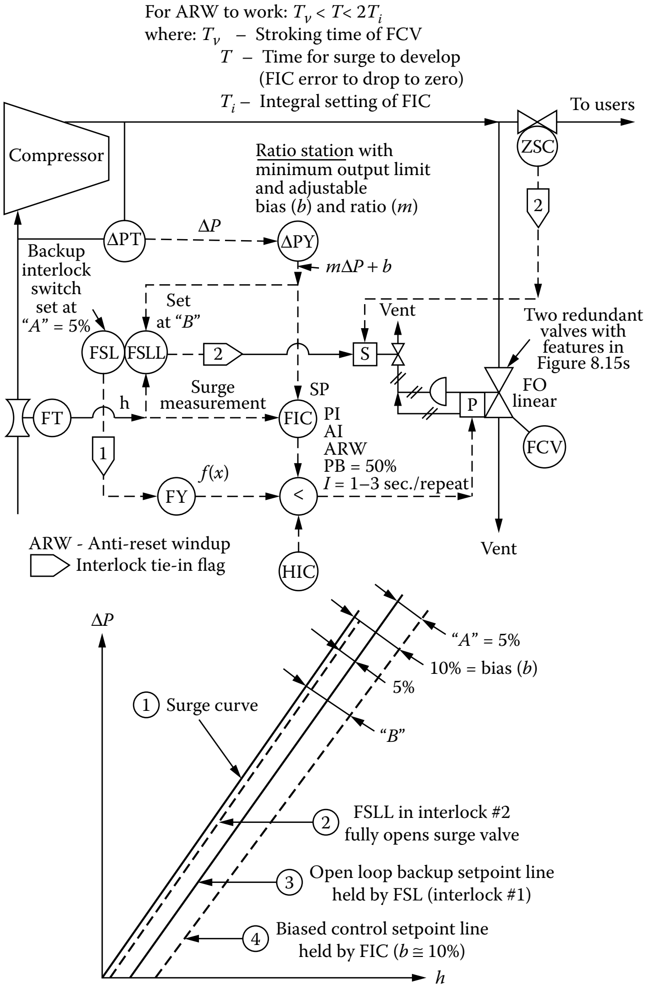


FIG. 8.15t

The surge control system (top) and the surge margin settings for controller set point and interlocks (bottom) for compressors operating at low compression ratios. (Adapted from Reference 1.)

is set to actuate when h has fallen 5% below SP, and at that point it fully opens the surge valve through interlock #1.

The shape of the control and backup lines shown in Figure 8.15t is applicable only when the surge curve is parabolic. As was shown in the surge curves in Figures 8.15m, 8.15n, and 8.15o, the surge curves can also be linear or have negative slopes and discontinuities. If the surge line is not parabolic, the following techniques can be considered for generating the surge control lines.

If the surge curve is linear, the signal *h* should be replaced by a signal representing flow. This can be obtained by adding a square root extractor to the FT in Figure 8.15t. If the surge curve bends down, the use of two square root extractors in series[1] can be used to approximate its shape.

On multistage machines with high compression ratios, it might be necessary to substitute a signal characterizer for ΔPY in Figure 8.15t. On variable vane designs, it has been proposed that the flowmeter differential (*h*) could be corrected to be a function of actual vane position to better approximate the surge curve[3] (see Figure 8.15u). If a flowmeter differential cannot be obtained from either the suction or the discharge side of the compressor, the surge curve can also be approximated on the basis of speed, power, or vane position measurements.[3]

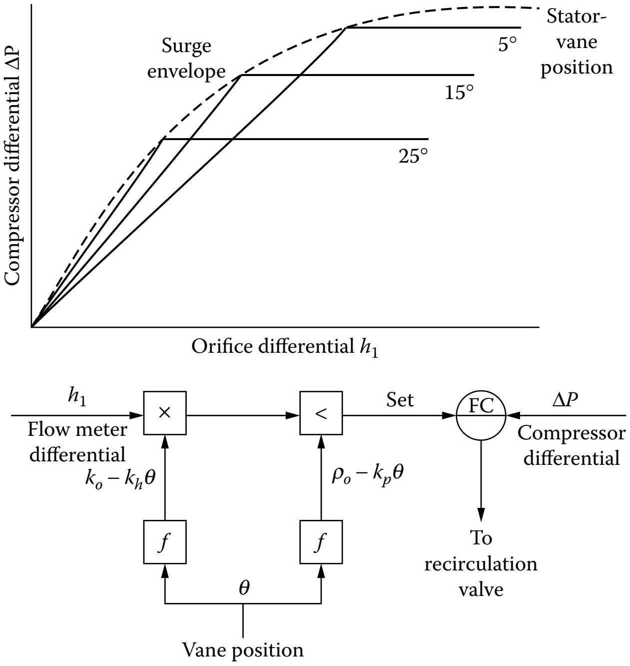


FIG. 8.15u

If the surge line of the compressor bends over, the control loop should measure both the flow meter differential and the vane position. (Adapted from Reference 3.)

The Surge Flow Controller The surge controller (FIC) in Figure 8.15t is a direct-acting controller with proportional and integral action and with antireset windup (AWR) features. It is a flow controller with a remote set point and a narrower proportional band than would be used for flow control alone. In a well-tuned surge controller, both the proportional band (PB) and the integral settings must be minimized (PB ~ 50%, 1–3 sec/repeat). With this level of responsiveness, the FIC set point can be less than 10% from the surge curve, as shown in Figure 8.15t.

Under normal conditions, the actual flow (*h*) is much higher than the surge control limit (SP), and therefore the integral mode of the FIC has a tendency to wind up on its positive error toward a saturated maximum output. If this was allowed to happen, the controller could not respond to a surge condition (a quick drop in *h*), because it would first need to develop a negative error equal to that which caused the saturation.

The aim of an antireset windup feature in this controller is to hold the FIC output under normal conditions at around

105%, so that the valve is closed but the signal is just lingering above the 100% mark without saturation.

As soon as *h* starts dropping toward the set point, the valve should start to open, and it should reach full opening before the set point is reached. As soon as *h* starts dropping, the proportional contribution to the FIC output decreases. If the approach is slow, the increase in the reset contribution will be greater than the decrease caused by the proportional contribution. Therefore, if *h* does not drop to SP, the valve slowly returns to the closed position. The effect of the internal feedback in the FIC is to activate the integral action only when the valve is not closed, while keeping the proportional action operational all the time.

The main goal of a feedback surge controller is to protect the machine from going into surge. Once the compressor is in surge, the FIC is not likely to be able to bring it out of it, because the surge oscillations are too fast for the controller to keep up with them. This is why backup systems are needed.

The surge controller is usually electronic, and the more recent installations tend to be microprocessor-based. Such digital units can memorize complex, nonlinear surge curves and can also provide adaptive gain, which is a means of increasing controller gain as the operating point approaches the surge curve. In digital controllers, it is desirable to set the sample time at about one quarter of the period of surge oscillation, or at 0.3 sec, whichever is shorter. If a flow-derivative backup control (Figure 8.15v) is used, the sample time must be 0.05 sec or less, because if it is slower, the backup system would miss the precipitous drop in flow as surge is beginning.

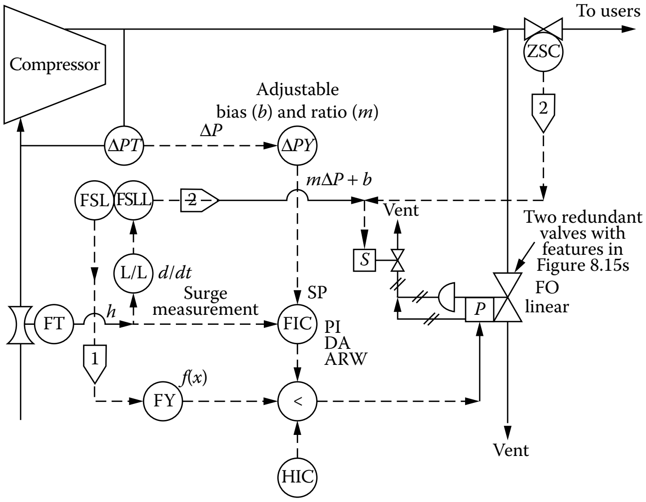


FIG. 8.15v
In this configuration the interlock backup (FSLL) is triggered by measuring the rate of flow change (L/L). If this is done by a digital control system, the sample time must be 0.05 sec or less. (Adapted from Reference 1.)

Tuning the Controller The *closed loop ultimate oscillation* method cannot be used for tuning the surge loop, because there is a danger of this oscillation triggering the start of surge. The *open loop reaction curve* technique of tuning is used with both the surge and the throughput controller in manual. The output of each controller is changed by 10%, and the resulting measurement response is analyzed on a high-speed recorder, yielding the time constant and the dead time of the process (Figure 8.15w).

The tuning settings of the surge controller can be tested by first increasing the bias (*b* in Figure 8.15t) until the surge valve opens and then analyzing the loop response. The FIC must be so tuned that its overshoot is less than the distance between the FSL and FIC set points (*A* in Figure 8.15t). The throughput controller should be kept in automatic while the surge controller is tested so that interaction problems will be noted. Once the surge controller is tuned, it should not be switched to manual but should remain always in automatic.

The antireset windup feature is not effective on slow disturbances, because the integral contribution that increases the FIC output outweighs the proportional contribution that lowers it; thus, the valve stays closed until the surge line is crossed. In order for the ARW to be effective, the time for the controller error to drop from its initial value to zero (*T*) must fall within two limits. It must be slower than the stroking time of the surge valve (*Tv*), but it must be faster than two integral times of the controller:[1]

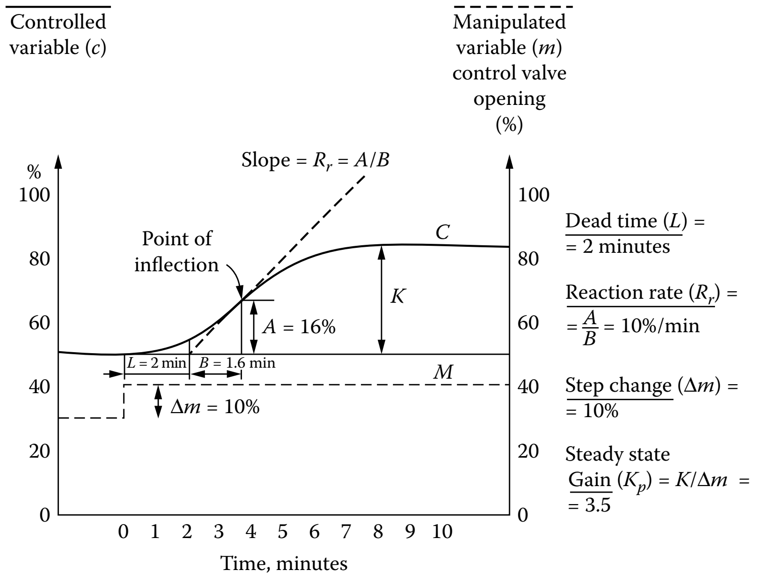
$$Tv < T < 2Ti \qquad \textbf{8.15(5)}$$

If the above requirement cannot be satisfied—because, for example, one of the potential disturbances is a slow closing discharge valve—an "anticipator" control loop should be added (Figure 8.15z).

Surge Protection Backup The purpose of the FIC in Figure 8.15t was to protect the compressor from going into surge. The feedback surge controller is usually not fast enough to bring the machine out of surge once it has developed, because the surge oscillations are too fast for this controller. Yet even with the best feedback surge controller design, the machine will go into surge.

This can be intentionally initiated to verify the surge curve, it can be caused by errors or shifts in the surge curve as the unit ages, and it can occur because of misoperation or failure of either the controller or downstream block and check valves. For these reasons, a second line of defense—a backup system—is required in addition to the surge FIC.

The backup system usually consists of two subsystems, identified as interlocks #1 and #2 in Figure 8.15t. In interlock #2, FSLL detects conditions that will cause the complete stopping of all forward flow from the compressor; when such a condition is detected, FSLL instantaneously and fully opens



Recommended Ziegler-Nichols settings (noninteracting)

Settings / Type of controller	Proportional	Integral (min/repeats)	Derivative
Proportional	$100\, R_r L/\Delta m = 200\%$	–	–
PI controller	$111\, R_r L/\Delta m = 222\%$	$3.33\, L = 6.66$	–
PID controller	$83\, R_r L/\Delta m = 166\%$	$2\, L = 4.0$	$0.5\, L = 1.0$

FIG. 8.15w
An example of open loop tuning where the settings are determined on the basis of the dead time (L) and reaction rate (R_r) of the controlled process.

the surge valve. One such condition in Figure 8.15t is signaled by a closed position limit switch (ZSC) on the block valve.

Interlock #1 in Figure 8.15t provides open-loop backup as follows: The FSL detects an approach to surge that is closer than the FIC set point; when it detects such a condition, it takes corrective action. Under normal conditions, FSL senses an *h* value that is higher than SP, meaning that the operating point is to the right of curve ④. As surge approaches, the operating point crosses the control line (curve 4) as *h* drops below SP. The FSL is set to actuate interlock #1 when *h* has crossed curve ③. This occurs when the flow has dropped below the FIC set point by the amount "*A*" (usually 5%).

When interlock #1 is actuated, it triggers the signal generator FY to drop its output signal to zero and then, when the operating point has returned to the right of curve ③, gradually increase it back to 100%. When the FY output drops to zero, the low signal selector (in the output of the FIC) will select it and send it to the valve. This causes the surge valve to open fully in less than a second, followed by a slow closure according to the program in the signal generator FY. As the FY output rises, it will reach the output of the FIC, at which point control is returned from the backup system to the feedback controller.

This type of backup is called the *operating point method.* Its advantage is that it takes corrective action *before* the surge curve (curve ①) is reached. Its disadvantages are that it is not usable on multistage machines or on systems with recycle and that the protection it provides is lost if Δ*PY* in Figure 8.15t fails by dropping its output to zero. For this reason, it is desirable to provide a minimum limit on the output of Δ*PY.*

The manual loader input (HIC) to the low-signal selector is used during start-up. Because its output passes through the low selector, it can increase but cannot decrease the valve opening.

Flow-Derivative Backup The difference between the operating point technique of backup (Figure 8.15t) and the flow-derivative method (Figure 8.15v) is that the first method acts before the surge curve is reached, whereas the second is activated by the beginning of surge. Because flow drops off very quickly at the beginning of surge, the instruments that measure the rate of this drop must be very fast.

If implemented digitally, the sample time of the flow derivative loop must be 0.05 sec or less in order not to miss the initial drop in flow. The lead-lag station (*L/L*) that detects the rate of flow reduction is adjusted so that the lag setting will filter out noise and the lead setting will give a large output when the flow drops.

The function of interlock #1 is the same as was described in connection with Figure 8.15t. When FSL detects the drop in flow, it causes the signal generator (FY) output to drop to zero, fully opening the valve within a second. If this action succeeds in arresting the surge, the signal generator output slowly rises and control is returned to the feedback FIC.

If interlock #1 does not succeed in bringing the machine out of surge, then after a preset number of oscillations FSLL is actuated. This triggers interlock #2, which keeps the surge valve fully open until surge oscillations stop.

The flow derivation method of backup has the advantage of providing protection even if the compressor pressure-rise instruments (ΔPT, ΔPY) fail, but it also has the disadvantage of not being usable when the flow signal is noisy.

Optimized Adaptations of Surge Curve The surge curve shifts with wear and with operating conditions. Figure 8.15x describes a surge control loop that recognizes such shifts and automatically adapts to the new curve. The adaptation subroutine consists of two segments, the set point adaptation section (blocks ① to ⑥) and the output backup section (blocks ⑦ to ⑪).

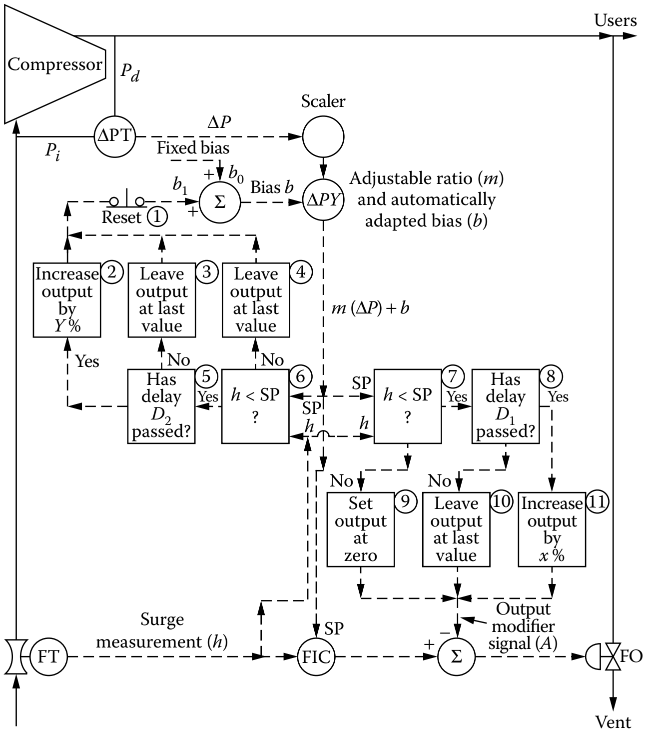


FIG. 8.15x
This control system recognizes changes in the surge curve and adapts the surge controller to the new curve. (Adapted from Reference 1, describing U.S. Patent 856,302, owned by Naum Staroselsky.)

The purpose of the output backup section is to recognize the approach of a surge condition that the feedback controller (FIC) was unable to arrest and to correct such a condition when it occurs. As long as the surge measurement (*h*) is not below the set point (SP), no corrective action is needed, and therefore blocks ⑦ and ⑨ will set the FIC output modifier signal (*A*) to zero. Once *h* is below SP, the operating point is to the left of curve ④ in Figure 8.15t, and the output of block ⑦ in Figure 8.15x is switched to "Yes."

Next, block ⑧ checks if the adjustable time delay D1—having typical values of 0.3 to 0.8 sec[1]—has passed. If it has not, signal *A* remains at its last value. If D1 has passed, *A* is increased by an increment *X*. Typical values of the *X* increment range from 15 to 30%.[1] As *A* is subtracted from the FIC output signal, this backup loop will open up the surge valve at a speed of 15–30% per 0.3–0.8 sec. When *h* is restored to above set point, the signal *A* is slowly returned to zero, allowing the surge valve to reclose.

The purpose of the set point adaptation loop (blocks ① to ⑥) is to recognize shifts in the surge curve and, as a response, to move the control line ④ in Figure 8.15t to the right by increasing to total bias *b* of the ratio relay ΔPY. The logic of blocks ② to ⑥ is similar to that described for blocks ⑦ to ⑪, except that the resulting variable bias signal (b_1) is added to the fixed bias (b_0) to arrive at the adapted new bias (*b*). The speed of set point adaptation does not need to be as fast as the opening of the surge valve. Therefore, the time delay D2 tends to be longer than D1, and the increment *Y* is smaller than *X*. The purpose of the reset button in block ① is to provide a means for the operator to reset the b_1 signal back to zero.

Override Controls

Figures 8.15t and 8.15v show backup and overrides implemented by a low-signal selector on the outlet of the feedback FIC. Figure 8.15y shows some additional overrides that might

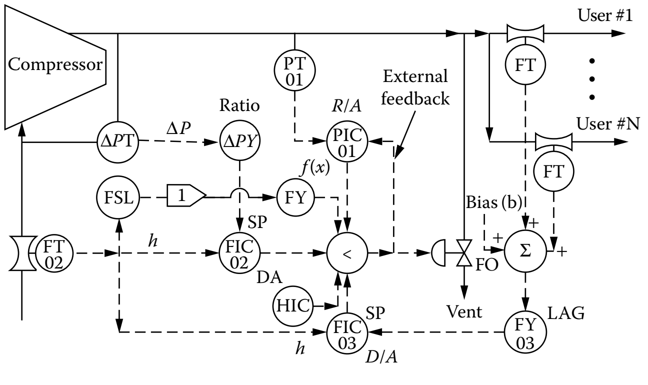


FIG. 8.15y
The low-signal selector provides this control system with a high-pressure (PIC-01) and a user shutdown (FIC-03) override. (Adapted from Reference 1.)

be desirable in some installations. Under normal conditions all override signals are above 100%, and they leave the feedback FIC in control. Sending the override signals through the low-signal selector guarantees that they can only increase the opening of the surge valve, not decrease it.

The high-pressure override controller (PIC-01 in Figure 8.15y) protects downstream equipment from overpressure damage. Normally its measurement is far below its set point, and therefore its output signal is high. In order to prevent reset windup, external feedback is used.

The other override loop (FIC-03) guarantees that the shutdown of a user or a sudden change in user demand will not upset the mass balance (balance between inlet and outlet flow) around the compressor. This is an important feature, because without it the resulting domino effects could shut down other users or the whole plant.

In Figure 8.15y, when a user is shut down, FT-03 detects the drop in total flow within milliseconds, if a fast flow transmitter is used. This drop in flow reduces the measurement signal (*h*) to FIC-03, while its set point is still unaltered because of the lag (FY-03). Therefore, the surge valve is opened to maintain the mass balance. As FY-03 slowly lowers the set point to the new user flow rate (plus bias), FIC-03 recloses the valve. The bias *b* is used to keep the set point of FIC-03 above its measurement, thus keeping the surge valve closed when conditions are normal.

Figure 8.15z illustrates another method of maintaining mass balance as users are suddenly turned off. The velocity limiter (VY-03) limits the rate at which the surge signal (*h*) can be lowered. If the surge signal drops off suddenly, the VY-03 output remains high and, through the high signal selector, keeps the surge controller set point up. This opens the surge valve to maintain the mass balance until it is reclosed by the limiter. This method is called an *anticipator algorithm,* because it opens the surge valve sooner than the time when the operating point (*h*) would otherwise drop to the set point, computed by the ratio station (ΔPY).

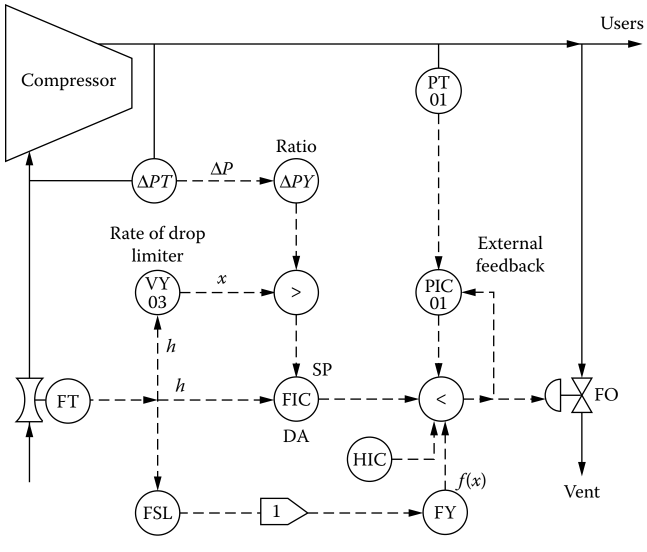


FIG. 8.15z
The velocity limiter serves to protect against sudden load changes by temporarily opening the surge vent valve when the load is quickly dropping.[7] This is achieved by providing VY-03, which, in case of a sudden drop in h, does not immediately lower x and therefore it (x) is selected by the high selector as the set point.

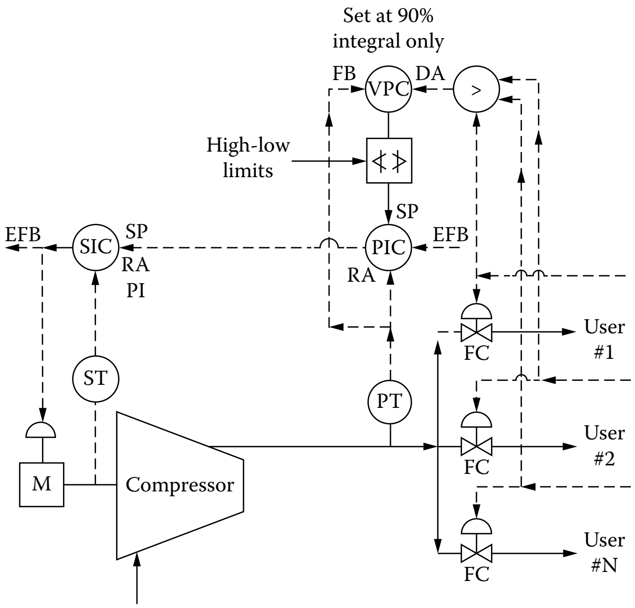


FIG. 8.15aa
By keeping the compressor's discharge pressure just high enough so that even the most open user valve does not open beyond 90%, the needs of all users are satisfied with a minimum investment of energy.

Optimized Load-Following

When a compressor is supplying gas to several parallel users, the goal of optimization is to satisfy all users with the minimum investment of energy. The minimum required header pressure is found by the valve position controller (VPC) in Figure 8.15aa. It compares the highest user valve opening (that identifies the needs of the most demanding user) with a set point of 90%, and if even the most open valve is not yet 90% open, it lowers the header pressure set point.

This supply-demand matching strategy not only minimizes the use of compressor power but also protects the users from being undersupplied, because it protects any and all supply valves from the need to open fully. On the other hand, as the VPC causes the user valves to open farther, these valves not only reduce their pressure drops but also become more stable.

The VPC is an integral-only controller, with its integral time set for about 10 times that of the PIC. This guarantees

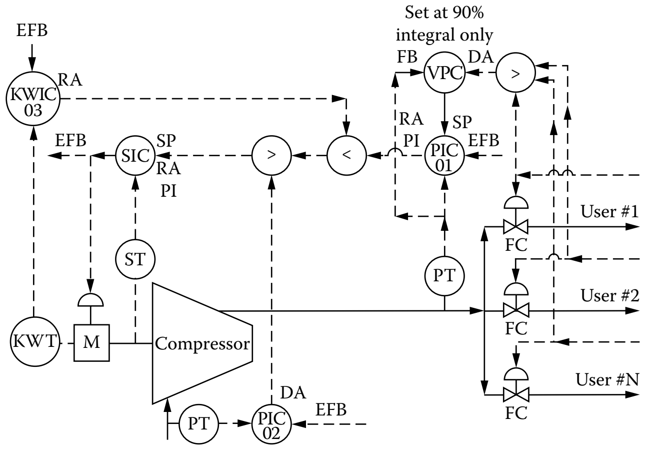


FIG. 8.15bb
Protective overrides can be added to optimized load-following controls, so that the system is protected against excessively low pressures on the suction side of the compressor or from overloading the compressor's motor drive.

that the VPC will act more slowly than all the user controllers, thus giving stable control even if the user valves are unstable. The external feedback (FB in Figure 8.15aa) protects the VPC from reset windup when its output is limited or when the PIC has been switched to manual.

In addition to following the load, it is also necessary to protect the equipment. In Figure 8.15bb, one protective override prevents the development of excessively low suction pressures (PIC-02), which could result in drawing oil into the compressor. The other override (KWIC-03) protects from overloading the drive motor and thereby tripping the circuit breaker.

In order to prevent reset windup when the controller output is blocked from affecting the SIC set point, external feedback (EFB) is provided for all three controllers. This arrangement is typical for all selective or selective-cascade control systems.

Interaction and Decoupling

Both the load and the surge control loops are shown in Figure 8.15cc. The manipulated variable of both of these loops is the compressor throughput. Under normal conditions, there is no problem of interaction; because the surge loop is inactive, its valve is closed.

Under abnormal conditions, when point "*A*" in Figure 8.15dd is reached, FIC-02 quickly opens the surge valve, which causes the discharge pressure to drop off as the flow increases. PIC-01 responds by increasing the vane setting or speed of the machine.

The faster the PIC-01 loop instrumentation is in Figure 8.15cc and the higher its gain (the narrower the proportional band), the larger and faster will be the increase in the speed of the compressor. If the load curve is steeper than the surge curve (as in Figure 8.15dd), the action of PIC-01 will bring the operating point closer to the surge line. In this case, a conflict exists between the two loops, because as FIC-02 acts to correct an approaching surge situation, PIC-01 responds by worsening it. The better tuned (narrower proportional band) and faster (electronic or hydraulic speed governors) the PIC-01

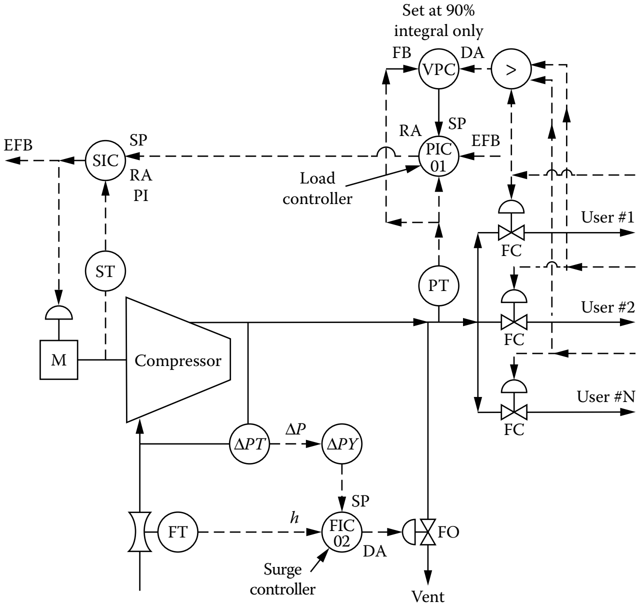


FIG. 8.15cc
The load controller (PIC-01) and surge controller (FIC-02) both affect the compressor throughput.

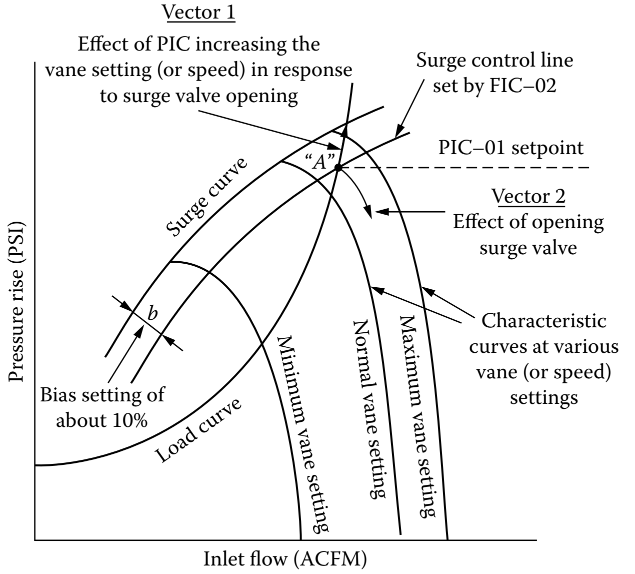


FIG. 8.15dd
If the load curve is steeper than the surge curve and if the controlled variable is pressure, the load and surge control loops will fight each other. (Adapted from Reference 1.)

TABLE 8.15ee
Type of Interactions between Load and Surge Controller

Controlled Variable	*Load Curve Steeper than Surge Curve*	*Nature of Interaction*
Pressure	Yes	Conflict
	No	Assist
Flow	Yes	Assist
	No	Conflict

TABLE 8.15ff
The Nature of the Interaction as a Function of the RG Value

RG Value	*Effect of Other Loop*
0 to 1.0	Assists the primary loop
Above 1.0	Conflicts with the primary loop
Below 0	Conflict that also reverses the action of primary loops

loop is, the more dangerous is its effect of worsening the approach of surge.

The throughput controller (PIC-01 in Figure 8.15cc) moves the operating point on the path of the load curve. The surge controller (FIC-02 in Figure 8.15cc) moves the operating point on the path of the characteristic curve of the compressor at constant speed. If the characteristic curve is steep, the action of the surge controller will cause a substantial upset in the discharge pressure controlled by PIC-01 (Figure 8.15cc). If the load curve is flat, the effect of the pressure controller on the surge controller will be the greatest.

If the throughput is under flow control, the opposite effects will be observed. The effect of the surge controller on the flow controller will be the greatest when the characteristic curves of the compressor are flat, and the effect of the flow controller on the surge controller will be greatest when the load curve is steep.

As shown in Table 8.15ee, whether the loops will assist or conflict with each other, both the relative slopes of the load and surge curves and the variable selected for process control need to be considered.

As both loops try to position the operating point on the compressor map, the resulting interaction can cause the type of inverse response that was described in Figure 8.15dd, or it can cause oscillation and noise. The oscillating interaction is worst if the proportional bands and time constants or periods of oscillation are similar for the two loops. Therefore, if tight control is of no serious importance, one method of reducing interactions is to reduce the response (widen the proportional band) of the load controller. Similarly, the use of slower actuators (such as pneumatic ones) to control compressor throughput will also reduce interaction, but at the cost of less responsive overall load control.

Relative Gain In evaluating the degree of conflict and interaction between the load and surge loops, it is desirable to calculate the relative gain between the two loops. The relative gain is the ratio between the open-loop gain when the other loop is in manual, divided by the open-loop gain when the other loop is in automatic. The open-loop gain of PIC-01 (Figure 8.15cc) is the ratio of the change in its output to a change in its input that caused it. Therefore, if a 1% increase in pressure results in a 0.5% decrease in compressor speed, the open-loop gain is said to be –0.5. Assuming that the open-loop gain of PIC-01 is –0.5 when FIC-02 is in manual and the switching of FIC-02 into automatic causes the PIC-01 loop gain to drop to –0.25, the relative gain is 0.5/0.25 = 2. Table 8.15ff lists the correct interpretations of the relative gain (RG) values.

If the calculated relative gain values are put in a 2-by-2 matrix, the best pairing of controlled and manipulated variables is selected by choosing those that will give the least amount of conflict. These are the RG values between 0.75 and 1.5, preferably close to 1.0. The regions of inhibition, reinforcement, and reversal are shown in Figure 8.15gg.

Decoupling Decoupling is the means of reducing the interaction between the surge and the load control loops. If the goal of decoupling is to maintain good pressure control even during a surge episode, the system shown in Figure 8.15hh can be used. In this system, as the antisurge controller opens the vent valve, a feedforward signal (X) simultaneously increases the speed of the compressor in proportion. The negative sign at the summing device is necessary because the surge valve fails to open. If the two vectors shown in Figure 8.15dd are correctly weighed in the summer, the end

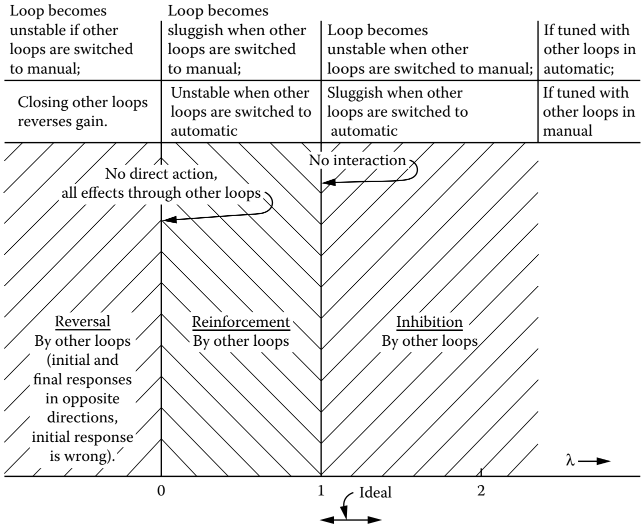


FIG. 8.15gg
The relative gain spectrum.

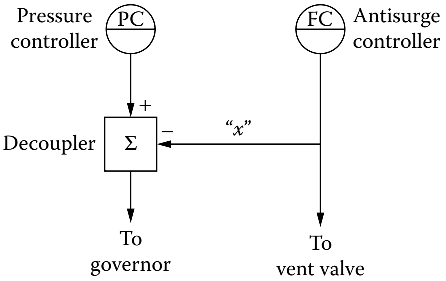


FIG. 8.15hh
The compressor discharge pressure can be kept unaffected by a surge episode, if the opening of the surge vent valve is converted into a proportional increase in speed setting of the governor.[3]

result of their summation will be a horizontal vector to the right, and PIC-01 will stay on set point.

On the other hand, if the main goal of decoupling is to temporarily reduce the size of vector #1 in Figure 8.15dd so that it will not contribute to the worsening of the surge condition, then the half-decoupling configuration shown in Figure 8.15ii can be considered. Here the decoupling summer also receives a feedforward signal (X) corresponding to the opening of the surge valve, but it acts to slow down (not speed up) the machine.

As the output of FIC-02 drops, the compressor speed is temporarily lowered, because the added value of X is reduced. This brings the operating point farther away from the surge line, as shown in Figure 8.15dd. The lagged signal Y later on eliminates this bias, because when its time constant has been reached, the values of X and Y will be equal and will cancel each other. Therefore, this decoupler will serve to temporarily desensitize a fast and tightly turned PIC-01 loop, which otherwise might worsen the situation by overreacting to a drop in pressure when the vent valve opens.

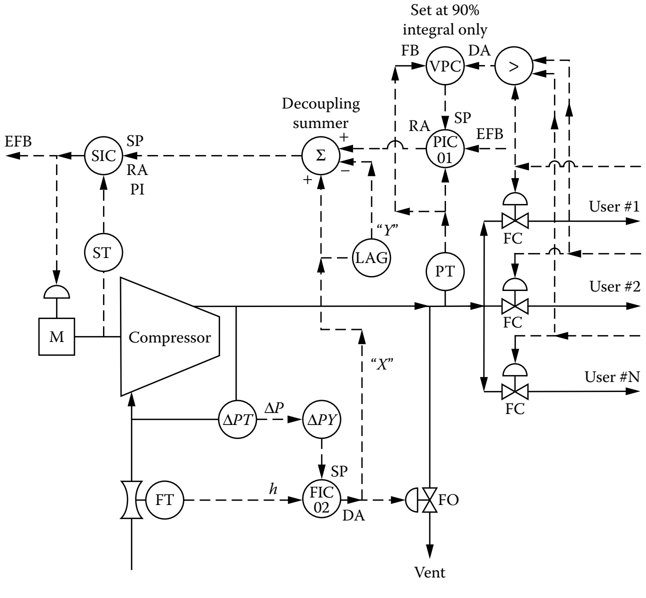


FIG. 8.15ii
The interaction between surge and load control loops can also be decoupled by temporarily reducing the speed of the compressor simultaneously with the opening of the surge valve. (Adapted from Reference 1.)

Multiple Compressor Systems

Compressors can be connected in series to increase their discharge pressure (compression ratio) or they can be connected in parallel to increase their flow capacity. Series compressors on the same shaft can usually be protected by a single antisurge control system.

When driven by different shafts, they require separate antisurge systems, although an overall surge bypass valve can be common to all surge controllers through the use of a low-signal selector.[3] This eliminates the interaction that otherwise would occur between surge valves, as the opening of a bypass around one stage not only would increase the flow through the higher stages but would also decrease it through the lower ones. In some installations, this interaction has been found to be of less serious consequence than the time delay caused by the use of a single overall bypass surge valve, which cannot quickly increase flow in the upper stages.[1] In such installations it is best to duplicate the complete surge controls around each compressor in series.

If streams are extracted or injected between the compressors, which are in series and their flows are not equal, a control loop needs to be added to keep both compressors away from their surge lines by automatically distributing among them the total required compression ratio. Such a control system is shown in Figure 8.15jj.

In this system, PIC-01 controls the total discharge pressure by adjusting the speed of compressor #1. The speed of

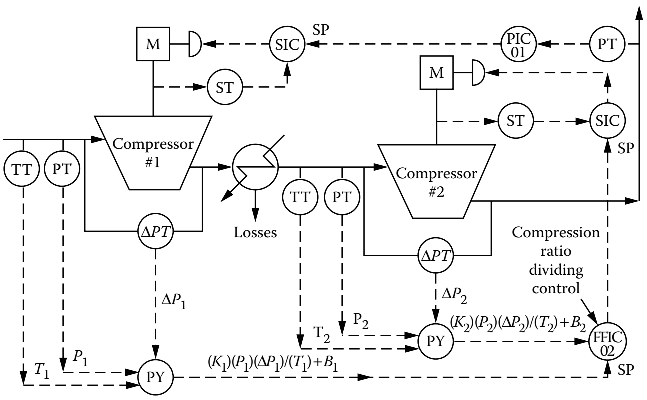


FIG. 8.15jj
When two compressors operate in series, one can be dedicated to maintain the total discharge pressure (PIC-01) from the pair, while the speed of the other can be manipulated to keep them at equal distance from their surge curves. This is achieved by FFIC-02 controlling the distribution of the total compression ratio between them. (Adapted from Reference 1.)

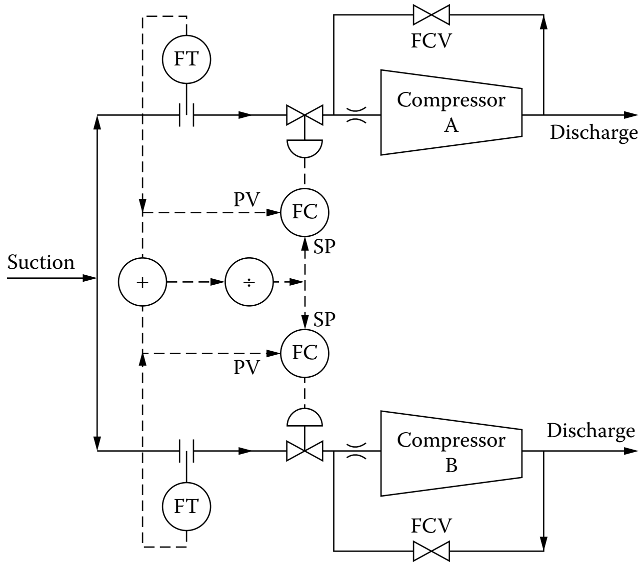


FIG. 8.15kk
When operating multiple compressors in parallel, the total flow (load) can be so distributed among the machines that one is fully loaded and the other handles the variations in demand.

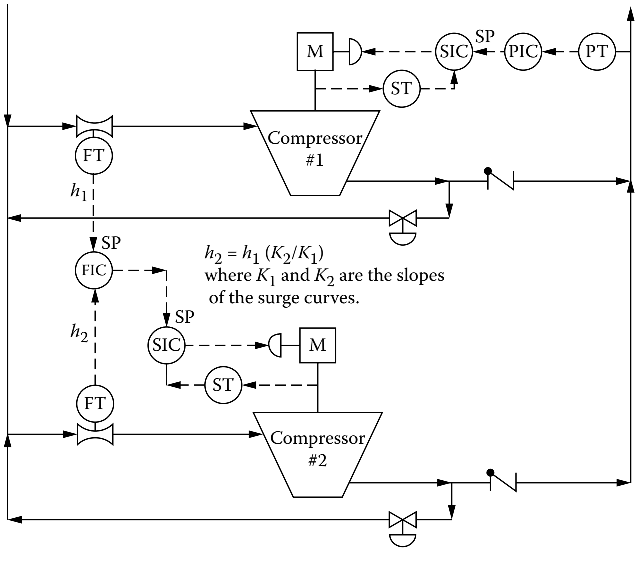


FIG. 8.15ll
Two parallel compressors can be so loaded as to keep both of them at equal distance from their surge curves. (Adapted from Reference 1.)

compressor #2 is set by FFIC-02; both compressors are thus kept at equal distances from their respective surge lines. This is accomplished by maintaining the following equality:

$$(K_1(P_1)(\Delta P_1)/(T_1)) + B_1 = (K_2(P_2)(\Delta P_2)/(T_2)) + B_2 \quad \mathbf{8.15(6)}$$

In this equation, K is the slope and B is the bias of the surge set points of the respective compressors.

Parallel Control of Compressors Controlling two or more compressors operating in parallel and having identical characteristics would be relatively simple. It is very difficult, if not impossible, to find two compressors having identical performance characteristics. Slight variations in flow can cause one compressor to be fully loaded. The parallel machine then has useless recycle. The control scheme shown in Figure 8.15kk alleviates that problem. Typically, care is exercised to ensure that the suction valve that receives the lower flow is kept 100% open. This prevents both suction valves from going fully closed to balance the flow. For an example of flow balancing controls, see Figure 8.17e.

Figure 8.15ll illustrates how two compressors can be proportionally loaded and unloaded, while keeping their operating points at equal distance from the surge curve. The lead compressor (#1 in Figure 8.15ll) is selected either as the larger unit or as the one that is closer to the surge curve when the load rises or is further from it when the load drops.

In Figure 8.15ll it is assumed that the compressors were so chosen that their ratio of bias to slope (b/K) of the surge set point is equal. In that case,

$$h_2 = h_1(K_2/K_1) \quad \mathbf{8.15(7)}$$

where h is the flow meter differential and K is the surge set point slope of the respective compressors.

Because of age, wear, or design differences, no two compressors are identical. A change in load will not affect them equally and each should therefore be provided with its own antisurge system.

Another reason for individual surge protection is that check valves are used to prevent backflow into idle compressors. Therefore, the only way to start up an idle unit is to let it build up its discharge head while its surge valve is partially open. If this is not done and the unit is started against the head of the operating compressors, it will surge immediately. The reason why the surge valve is usually not opened fully during start-up is to protect the motor from overloading.

Improper distribution of the load is prevented by measuring the total load (summer #9 in Figure 8.15mm) and assigning an adjustable percentage of it to each compressor by adjusting the set points of FFIC-01 and FFIC-02.

Optimization of Efficiency The load distribution can be computer-optimized by calculating compressor efficiencies (in units of flow per unit power) and loading the units in the order of their efficiencies. The same goal can be achieved if the operator manually adjusts the ratio settings of FFIC-01 and -02.

In the control system of Figure 8.15mm, the pressure controller (PIC-01) directly sets the set points of SIC-01 and -02 while the balancing controllers (FIC-01 and -02) slowly bias those settings. This is a more stable and responsive configuration than a pressure-flow cascade, because the time constants of the two loops are similar.

The output of PIC-01 must be corrected as compressors are started or stopped. One method of handling this is illustrated

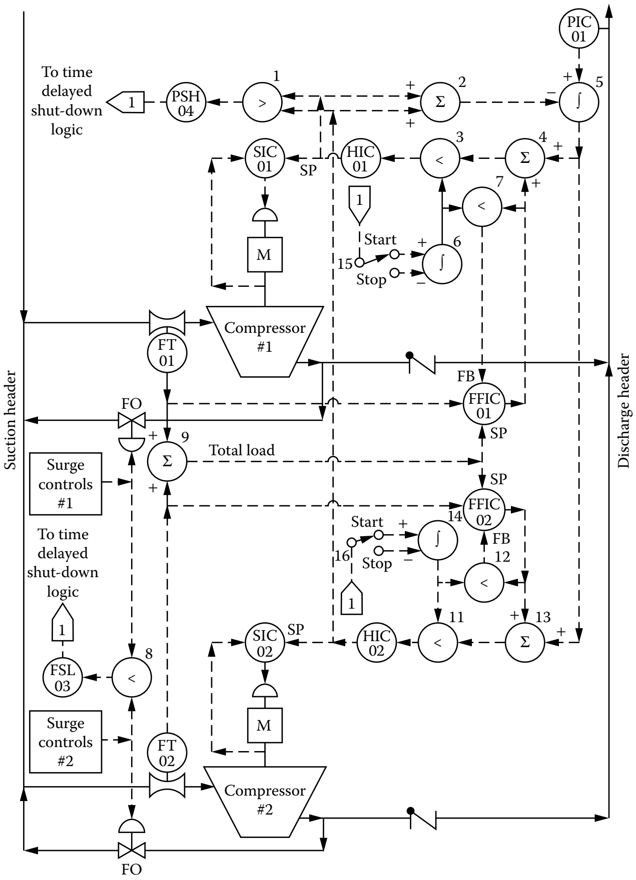


FIG. 8.15mm

A flow-balancing bias can be superimposed on direct pressure control. This control system can distribute the load between the two machines in the order of their efficiencies. (Adapted from Reference 3.)

in Figure 8.15mm. Here a high-speed integrator (item #5) is used on the summed speed signals to ensure a correspondence between the PIC output signal and the number of compressors (and their loading) used. The integrator responds in a fraction of a second and therefore does not degrade the speed of response of the PIC loop.

Figure 8.15mm also illustrates the automatic starting and stopping of individual compressors as the load varies. When the total flow can be handled by a single compressor or when any of the surge valves open, FSL-03 triggers the shutdown logic interlock circuit #1 after a time delay.

Operation with an open surge valve would be highly inefficient, because the recirculated gas is redistributed among all the operating units. When a compressor is to be stopped, item #15 (or #16) is switched to the stop position, causing the integrator #6 (or #14) to drive down until it overrides the control signal in low selector #3 (or #11) and reduces the speed until the unit is stopped.

Automatic starting of an additional compressor is also initiated by interlock #1 when PSH-04 signals that one of the compressors has reached full speed. When a compressor is to be started, interlock #1 switches item #15 (or #16) to the start position, causing the integrator #6 (or #14) to drive up (by applying supply voltage to the integrator) until PIC-01 takes over control through the low selector #3 (or #11). The integrator output will continue to rise and then will stay at maximum, so as not to interfere with the operation of the control loop.

The ratio flow controllers (FFIC-01 and -02) are protected from reset windup by receiving an external feedback signal through the low selector #7 (or #12), which selects the lower of the FFIC output and the ramp signal.

Interlock #1 is also provided with "rotating sequencer" logic, which serves to equalize run times between machines and protects the same machine from being started and stopped frequently. A simple approximation of these goals is achieved if the machine that operated the longest is stopped and the one that was idle the longest is started.

If only one of the compressors is variable-speed, the PIC-01 output signal can be used in a split-range manner. For example, if there were five compressors of equal capacity, switches would be set to start an additional constant-speed unit as the output signal rises above 20, 40, 60, and 80% of its full range. The speed setting of the one variable-speed compressor is obtained by subtracting from the PIC output the sum of the flows developed by the constant-speed machines and multiplying the remainder by 5. The gain of 5 is the result of the capacity ratio between that of the individual compressor and the total.

Multi-Inlet Compressor Control

These compressors most often are refrigeration compressors. The main problem here is to keep the process temperatures, which are controlled by the evaporators, within tolerable limits. This is achieved by controlling the vapor temperature (by quenching) in a feedforward manner, based upon the amount of recycle flow or recycle valve position. A typical control scheme to control a three-inlet compressor is shown in Figure 8.15nn.

Installation

A check valve in the discharge line as close to the compressor as possible will protect it from surges. On motor-driven compressors, it is helpful to close the suction valve during starting to prevent overload of the motor. After the unit is operating, it should be brought to stable operating range as soon as possible to prevent overheating.

The recycle valve control should be fast opening and slow closing to come out of surge quickly and then stabilize the flow. The transmitter ranges should be such that the recycle valve is able to open in a reasonable amount of time with reasonable proportional and integral control constants of the antisurge controller. The importance of proper controller tuning cannot be underestimated. Typical settings are PB = 50%, I = 1–3 sec/repeat.

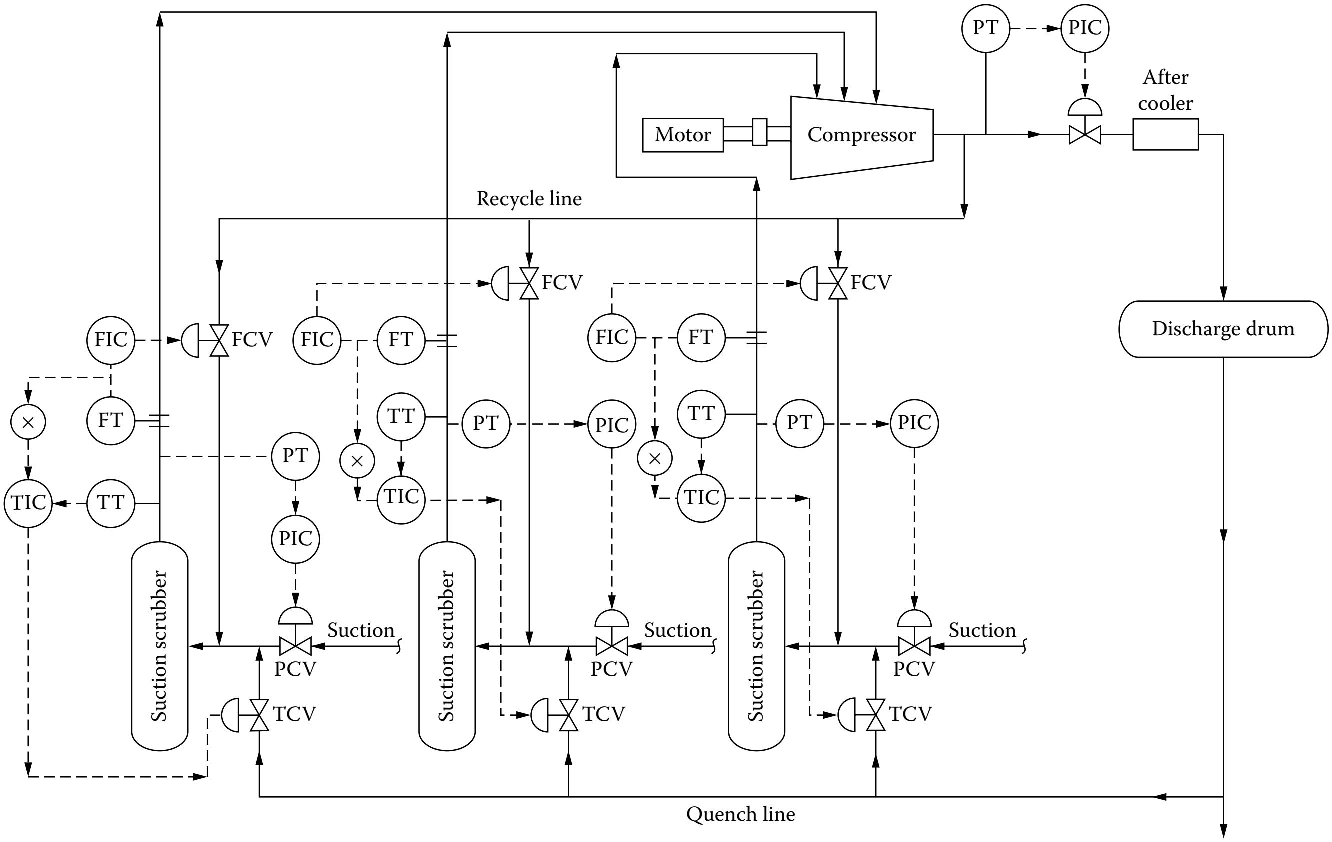


FIG. 8.15nn
Control strategy for a multiple inlet compressor.

Lube and Seal Systems A typical lube oil system is shown in Figure 8.15oo. Dual pumps are provided to ensure the uninterrupted flow of oil to the compressor bearings and seals. Panel alarms on low oil level, low oil pressure (or flow), and high oil temperature are provided. A head tank provides oil for coasting down in case of a power failure.

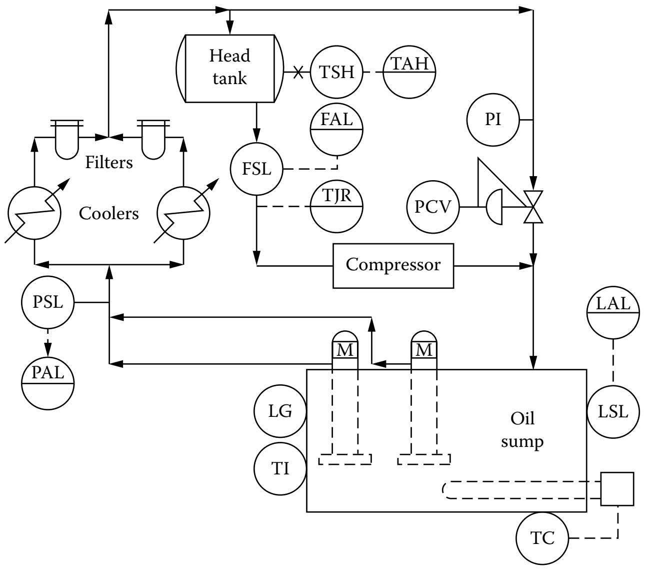


FIG. 8.15oo
Lube oil controls for the compressor's bearings and seals.

The design of these systems is critical, because failure of the oil supply could mean a shutdown of the entire process. In cases in which the process gas cannot be allowed to contact the oil, an inert gas seal system may be used. This is shown in Figure 8.15pp for a centrifugal compressor with balanced seals.

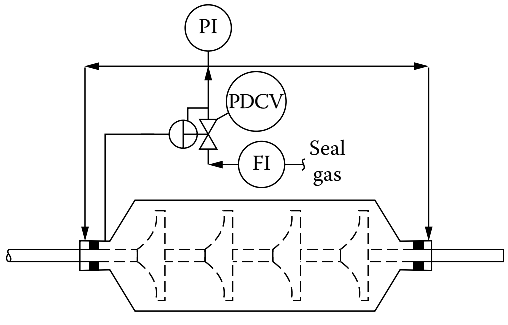


FIG. 8.15pp
Balanced seal controls on a centrifugal compressor.

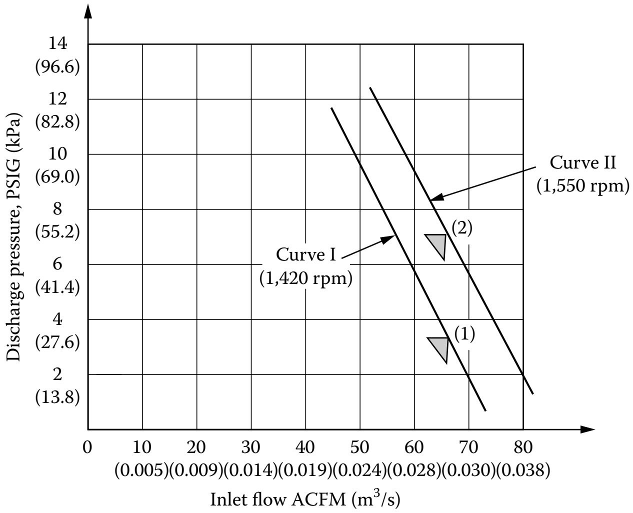


FIG. 8.15qq

The rotary compressor is a positive displacement machine. At constant inlet flow, the discharge pressure rises if the speed is increased, while at constant speed, the flow drops as the discharge pressure rises.

ROTARY COMPRESSORS, BLOWERS

The rotary compressor is essentially a constant-displacement, variable-discharge pressure machine. Common designs include the helical-screw, the lobe, the sliding vane, and the liquid ring types. The characteristic curves for a lobe-type unit are shown in Figure 8.15qq. As shown by curves I and II, the inlet flow varies linearly with the speed of this positive-displacement machine.

The small decrease in capacity at constant speed with an increase in pressure is a result of gas slippage through impeller clearances. It is necessary to compensate for this by small speed adjustments as the discharge pressure varies. For example, when the compressor is operating at point 1, it delivers the design volume of 66 ACFM (1.9 m^3/m) and 3.5 PSIG (24 kPa). In order for the same flow to be maintained when the discharge pressure is 7 PSIG (48 kPa), the speed must be increased from 1420 rpm to 1500 rpm at point 2 by the flow controller in the discharge line.

In addition to speed variation, the discharge flow can also be adjusted by throttling the suction, the bypass, or the vent line from the rotary blower. Vent throttling of the excess gas is shown in Figure 8.15rr in a process where the discharge is throttled by a temperature controller. This can be a dryer application, where the temperature of the outlet gas is controlled to prevent product degradation and to provide the proper product dryness. In other systems, instead of venting, the gas is returned to the suction of the blower under pressure control.

An important application of the liquid ring rotary compressor is in vacuum service. The suction pressure is often the independent variable and is controlled by bleeding gas into the suction on pressure control. This is shown in Figure 8.15ss, where suction pressure control is used on a rotary filter, maintaining the proper drainage of liquor from the cake on the drum.

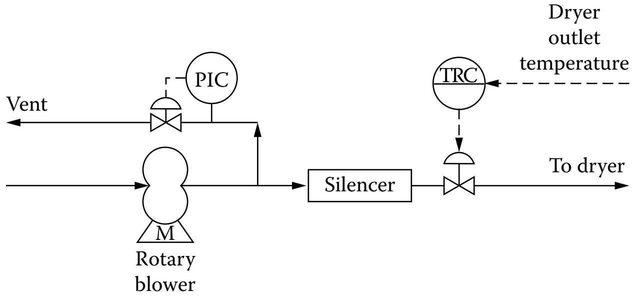


FIG. 8.15rr

The blower discharge pressure can be controlled by venting the excess gas that is not required by the process.

The optimization of rotary compressors will be discussed later, together with that of reciprocating compressors.

RECIPROCATING COMPRESSORS

The reciprocating compressor is a constant-volume, variable-discharge pressure machine. A typical compressor curve is shown in Figure 8.15tt, for constant-speed operation. The curve shows no variation in volumetric efficiency in the design pressure range, which may vary by 8 PSIG (53.6 kPa) from unloaded to fully loaded.

The volumetric inefficiency is a result of the clearance between piston end and cylinder end on the discharge stroke. The gas that is not discharged reexpands on the suction stroke, thus reducing the intake volume.

The relationship of speed to capacity is a direct ratio, because the compressor is a displacement-type machine. The

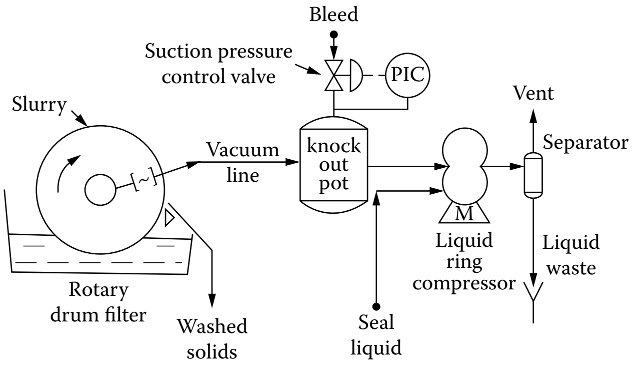


FIG. 8.15ss

The suction pressure control of a liquid ring-type rotary compressor is illustrated here.

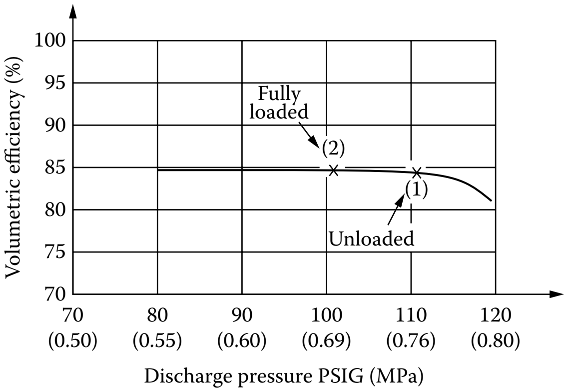


FIG. 8.15tt
Reciprocating compressor curve.

typical normal turndown with gasoline or diesel engine drivers is 50% of maximum speed, in order to maintain the torque within acceptable limits.

On/Off Control

For intermittent demand, where the compressor would waste power if run continuously, the capacity can be controlled by starting and stopping the motor. This can be done manually or by the use of pressure switches. Typical switch settings are on at 140 PSIG (1 MPa), off at 175 PSIG (1.2 MPa). This type of control would suffice for processes in which the continuous usage is less than 50% of capacity, as shown in Figure 8.15uu, where an air mix blender uses a rapid series of high-pressure air blasts when the mixer becomes full. The high-pressure air for this purpose is stored in the receiver.

Constant-Speed Unloading

In this type of control, the driver operates continuously, at constant speed, and one varies the capacity in discrete steps by holding suction valves open on the discharge stroke or opening clearance pockets in the cylinder. The most common schemes are three- and five-step unloading techniques. The larger number of steps saves horsepower because it more closely matches the compressor output to the demand.

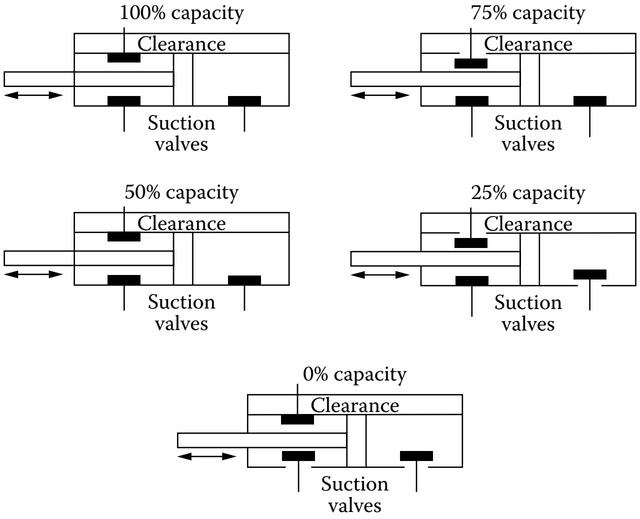


FIG. 8.15vv
The five steps in a constant-speed, positive displacement compressor with five-step unloading.

In three-step unloading, capacity increments are 100, 50, or 0% of maximum flow. This method of unloading is accomplished by the use of valve unloading in the double-acting piston. At 100%, both suction valves are closed during the discharge stroke. At 50%, one suction valve is open on the discharge stroke, wasting half the capacity of the machine. At 0%, both suction valves are held open on the discharge stroke, wasting total machine capacity.

For five-step unloading, a clearance pocket is used in addition to suction valve control. The capacity can be 100, 75, 50, 25, or 0% of maximum flow. This is shown in Figure 8.15vv. At 100%, both suction valves and the clearance pocket are closed. At 75%, only the clearance pocket is open. At 50%, only one suction valve is open on the discharge stroke. At 25%, one suction valve and the clearance pocket are open. At 0%, both suction valves are opened during the discharge stroke.

The use of step unloading is most common when the driver is inherently a constant-speed machine (Figure 8.15ww),

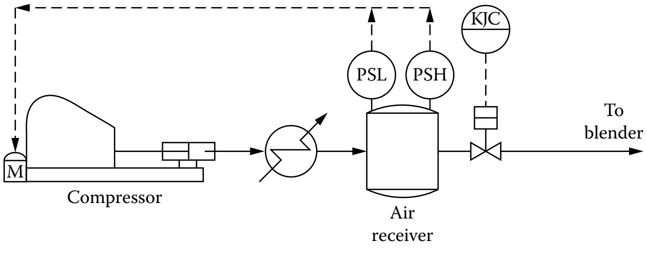


FIG. 8.15uu
On/off control of a reciprocating compressor.

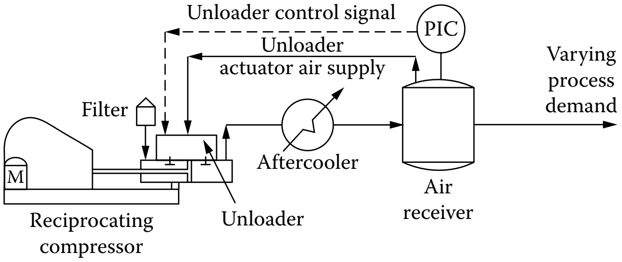


FIG. 8.15ww
Constant-speed capacity control of a reciprocating compressor, provided with pneumatic unloading.

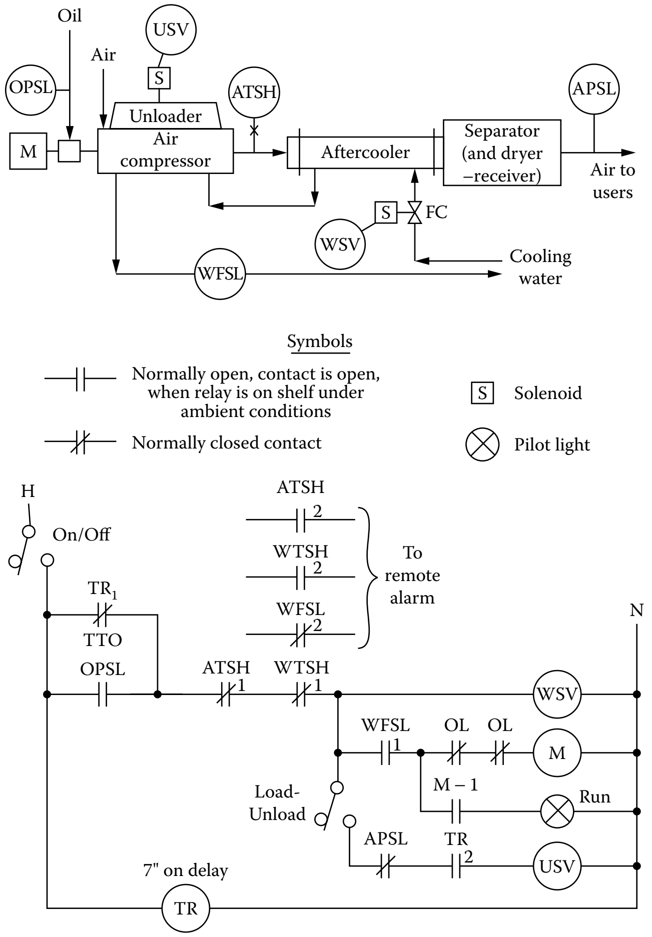


FIG. 8.15xx
The layout and control logic of a typical stand-alone air compressor.[2]

such as an electric motor. The pressure controller signal from the air receiver operates a solenoid valve in the unloader mechanism. The action of the solenoid valve directs the power air to lift the suction valve or to open the clearance port, or both.

For three-step unloading, two pressure switches can be used. The first switch loads the compressor to 50% if the pressure falls slightly below its design level, and the second switch loads the compressor to 100% if the pressure falls below the setting of the first switch.

For five-step unloading, a pressure controller is usually substituted for the four pressure switches otherwise required, and the range between unloading steps is reduced to not more than 2 PSI (13.8 kPa) deviation from the design level, keeping the minimum pressure within 8 PSI (55 kPa) of design. In cases in which exact pressure conditions must be met, a throttling valve is installed, which bypasses the gas from the discharge to the suction of the compressor. This device smoothes out pressure fluctuations and, in some cases, eliminates the need for a gas receiver in this service. This can prove economical in high-pressure services above 500 PSIG (3.45 MPa), in which vessel costs become significant.

The Stand-Alone Air Compressor

Abbreviations and Terminology

APSL Low air pressure switch with contact that closes as pressure drops below setting.

ATA Air temperature alarm, actuated on high temperature.

ATSH High air temperature switch with two contacts.

ATSH-1 First contact of ATSH is normally closed; it opens on high temperature.

ATSH-2 Second contact of ATSH is normally open; it closes on high temperature.

H Fused, 120 V, hot power supply of the stand-alone compressor control circuit.

HO Fused, 120 V, hot power supply of the integrated remote controls.

M Motor.

N Neutral wire of the stand-alone compressor control circuit.

NO Neutral wire of the integrated remote control system.

OFF DELAY This type of time delay introduces a delaying action when the relay is de-energized. A 2 min off-delay means that for the first 2 min period after de-energization, the relay contacts will remain as if the relay was energized. When the delay expires, the contacts will switch to their de-energized state.

ON DELAY This type of time delay introduces a delaying action when the relay is energized. A 3 sec on-delay means that for the first 3 sec after it is energized, the relay contacts will remain as if the relay was still de-energized. When the delay expires, the contacts will switch to their energized state.

TR Time delay relay. The delay time setting is marked next to it.

TTC Time to close action. This contact is open, then the TR relay is de-energized. This contact stays open when the TR relay is de-energized until the delay time setting expires. Then it closes.

TTO Time to open action. This contact is closed when the TR relay is de-energized. This contact stays closed when the TR relay is energized until the delay time setting expires. Then it opens.

USV Unloading solenoid. The compressor is loaded when this solenoid is energized.

WFA Water flow alarm. This alarm is energized if the flow drops below the minimum allowable.

WFSL Low water flow switch, with two contacts.

WFSL-1 This contact opens on low flow.

WFSL-2 This contact closes on low flow.

WSV Cooling water solenoid valve, opens when energized.

WTA Water temperature alarm, energized on high temperature.
WTSH High water temperature switch, with two contacts.
WTSH-1 This contact opens on high temperature.
WTSH-2 This contact closes on high temperature.

Operation A typical air compressor, together with the controls that are normally provided by its manufacturer, is shown in Figure 8.15xx.

Such a stand-alone compressor usually operates as follows: When the operator turns the control switch to "On," this will energize the time delay (TR) and will open the water solenoid valve (WSV) if neither the air temperature (ATSH) nor the cooling water temperature (WTSH) is high.

After 7 sec, TR-1 opens. Assuming that this time delay was sufficient for the oil pressure to build up, the opening of TR-1 will have no effect, because OPSL will have closed in the meantime.

If the opening of WSV resulted in a cooling water flow greater than the minimum setting of the low water flow switch, then WFSL closes and the compressor motor (M) is started. Whenever the motor is on, the associated "Run" pilot light is energized. This signals to the operator that oil (OPSL), water (WTSH, WFSL), and air (ATSH) conditions are all acceptable and, therefore, the unit can be loaded.

When the operator turns the other control switch to "Load," the TR-2 contact is already closed because the 7 sec have passed. Therefore, the APSL contact will determine the status of the machine. If the demand for air at the users is high, the pressure at APSL will drop, which in turn will cause the APSL contact to close and the compressor to load. As a result of loading the compressor, the pressure will rise until it exceeds the control gap of APSL, causing its contact to reopen and the machine to unload.

The compressor continues to load and unload automatically as a function of plant demand. As the demand rises, the loaded portion of the cycle will also rise. Once the load reaches the full capacity of the compressor, the unit stops cycling and stays in the loaded state continuously. If the demand rises beyond the capacity of this compressor, it is necessary to start another one. The following paragraphs describe how this is done automatically, requiring no operator participation.

If at any time during operation the cooling water flow (WFSL) drops too low, the motor will stop. If either water or air temperature rises to a high valve, this condition not only will stop the motor but will also close the water solenoid valve (WSV).

The above three conditions (WFSL, WTSH, ATSH) will also initiate remote alarms, as shown in Figure 8.15xx, to advise the operator of the possible need for maintenance. If the oil pressure drops below the setting of OPSL, it will also cause the stoppage of the motor and the closure of WSV, but after such an occurrence, the compressor will not be allowed to restart automatically when the oil pressure returns to normal. In order to restart the machine, the operator will have to go out to the unit and reset the system by turning the control switch to "Off" and then to "On" again to repeat the complete start-up procedure.

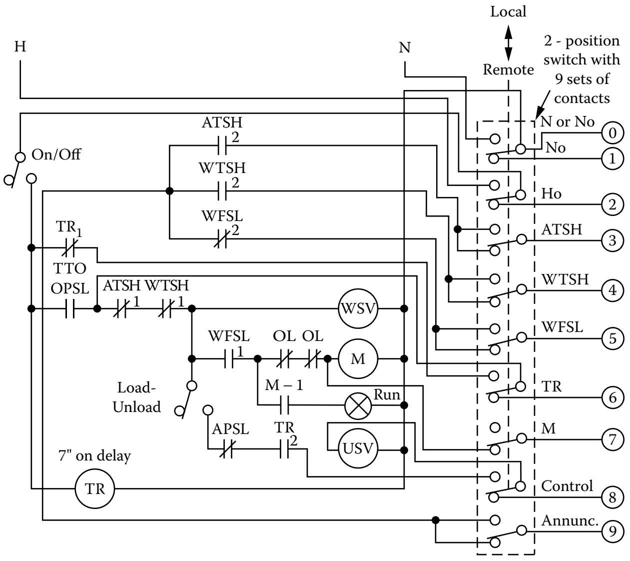


FIG. 8.15yy
The local/remote switch is wired to allow stand-alone or integrated compressor operation.[2]

Local/Remote Switch The first step in integrating several compressors into a single system is the addition of a local/remote switch at each machine. As shown in Figure 8.15yy, when this switch is turned to "Local," the compressor operates in the stand-alone mode, as was described in connection with Figure 8.15xx. When this switch is turned to "Remote," the compressor becomes a part of the integrated plantwide system, consisting of several compressors.

As shown in Figure 8.15yy, this two-position switch has nine sets of contacts and is mounted near the compressor. It can be installed in a few hours, and once installed, the compressor can again be operated in the "Local" mode. Only ten wires need to be run from each compressor to the remote controls. These ten wires serve the following functions:

#0 The working neutral, N in local, No in the remote mode of operation
#1 The common neutral (No) of the integrated controls
#2 The common hot (Ho) of the integrated controls
#3 The high air temperature (ATSH) alarm
#4 The high water temperature (WTSH) alarm
#5 The low water flow (WFSL) alarm
#6 The 7 sec time delay (TR) of the integrated controls
#7 The motor (M) status indication

#8 The load/unload control signal from the remote system

#9 The common hot for the remote annunciator (Annunc.)

When integrating several compressors into a single system, it is advisable to number these ten wires in a consistent manner, such as:

Compressor #1	Wire #10 to #19
Compressor #2	Wire #20 to #29
Compressor #3	Wire #30 to #39, and so on

In this system, the first digit of the wire number indicates the compressor, and the second digit describes the function of the wire. Immediately knowing, for example, that wire number 45 comes from compressor #4 and serves to signal a low water flow condition on that machine simplifies checkout and start-up.

Annunciator Of the ten wires from each of the compressors, six are used for remote alarming. Figure 8.15zz shows a remote annunciator for two compressors. This alarm system can be expanded to serve any number of compressors.

The position of the local/remote switch in Figure 8.15yy does not affect the operation of the annunciator. It provides the following remote indications for each compressor:

"Run" light
High air temperature light
High water temperature light
Low water flow light
Audible alarm bell with silencer
Alarm reset buttons

The only time these circuits are deactivated is when the associated compressor is off.

Lead-Lag Selector As plants grow, their compressed air requirements also tend to increase. As a result of such evolutionary growth, many existing plants are served by several uncoordinated compressor stations. When, because of space limitations, the new compressors are installed in different locations, the manual operation of such systems becomes not only inefficient but also unsafe.

The steps involved in integrating such stand-alone compressor stations into an automatically operating, load-following single system are described here. In such integrated systems, the identity of the "lead" and "lag" compressors, or the ones requiring maintenance ("Off"), can all be quickly and conveniently altered, while the system continues to efficiently meet the total demand for air. Thus, air supply shortage or interruption is eliminated, together with the need for continuous operator's attention.

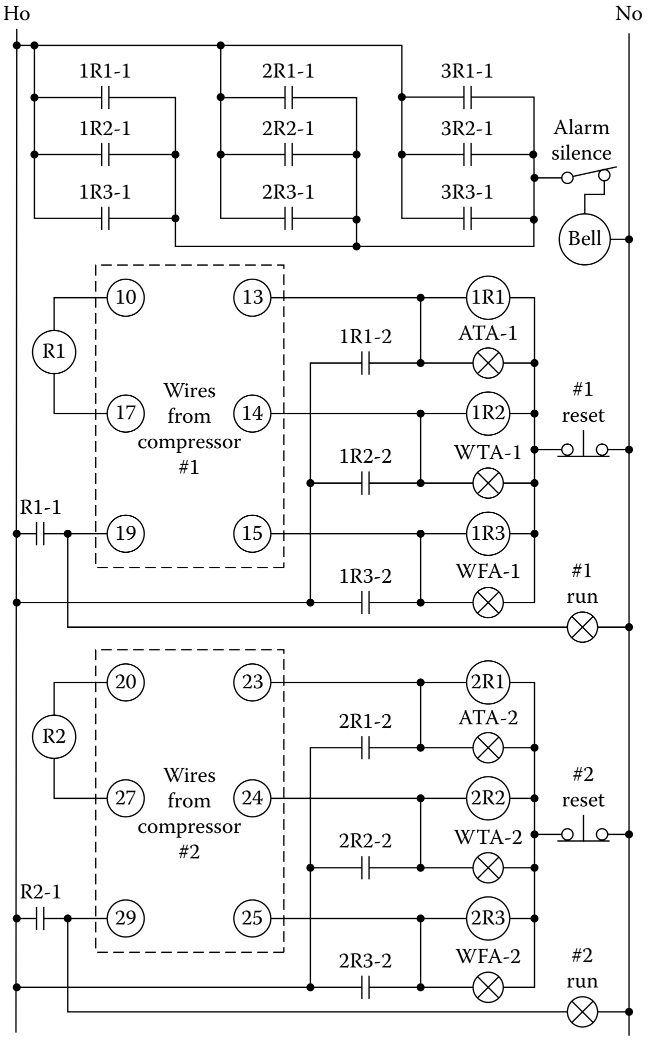


FIG. 8.15zz
The wiring requirements for an annunciator serving two compressors.[2]

When it is desirable to combine two compressors into an integrated lead-lag station, all that needs to be added are two pressure switches, as shown in Figure 8.15aaa.

APSL-A is the lead and APSL-B is the lag control pressure switch. To maintain a 75 psig air supply at the individual users, these normally closed pressure switches can be set as shown in Table 8.15bbb.

With the above settings, the system performance will be as shown in Table 8.15ccc.

When two stand-alone compressors are integrated into such a lead-lag system, only five wires need to be brought from each compressor to the lead-lag switch, shown in Figure 8.15ddd. Turning this single switch automatically reverses the lead-lag relationship between the two compressors. In addition to the two-position lead-lag switch, running lights are also provided; they show whether either or both compressors are unloaded (standby) or loaded.

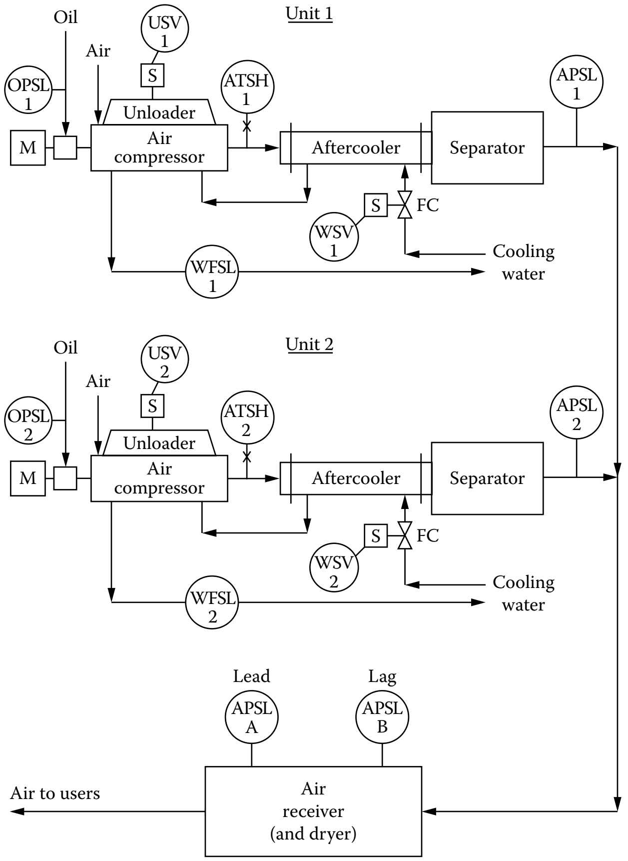


FIG. 8.15aaa

By adding two pressure switches (APSL-A and APSL-B), two compressors can be integrated into a single system.[2]

TABLE 8.15bbb

Set Pressures and Differential Gaps for Integrating the Controls of Two Compressors

Pressure Control Switch	*Pressure Below which Switch Will Close (psig)*	*Differential Gap (psi)*
ASPL-A	90	10
ASPL-B	85	10

As the demand for air increases, the lead machine will load and unload to meet that demand. If the lead machine is continuously loaded and the demand is still rising, the lag compressor will be started automatically. The interlocks provided are as follows:

The on-time delays 2TR and 3TR in Figure 8.15ddd are provided for stabilizing purposes only. They guarantee that a system configuration (or reconfiguration) will be recognized only if it is maintained for at least 3 sec. Responses to quick changes are thus eliminated.

The on-time delays ITR and 4TR are provided to give time for the oil pressure to build up in the system. If after 7 sec the oil pressure is not yet established, and therefore OPSL in Figure 8.15xx is still open, the contacts ITR-1 and 4TR-1 in Figure 8.15ddd will open and the corresponding compressor will be stopped.

The off-time delay 5TR guarantees that the lag machine will not be cycled on and off too frequently. Once started, the lag compressor will not be turned off (but will be kept on standby) until the off delay of 2 min has passed.

Off Selector Three stand-alone compressors can be combined into an integrated load-following single system, by the

TABLE 8.15ccc

System Performance when ASPL-A Is Set to Close at 90 psig, APSL-B Is Set to Close at 85 psig, and the Differential Gap Is Set at 10 psid

Pressure in Air Receiver (psig)	*APSL-A Contact*	*APSL-B Contact*	*Lead Compressor*	*Lag Compressor*
Over 100	Open	Open	Off	Off
100	Open	Open	Off	Off
95	Open	Open	Off	Off
90	Closed	Open	On	Off
85	Closed	Closed	On	On
80	Closed	Closed	On	On
85	Closed	Closed	On	On
90	Closed	Closed	On	On
95	Closed	Open	On	Off
100	Open	Open	Off	Off

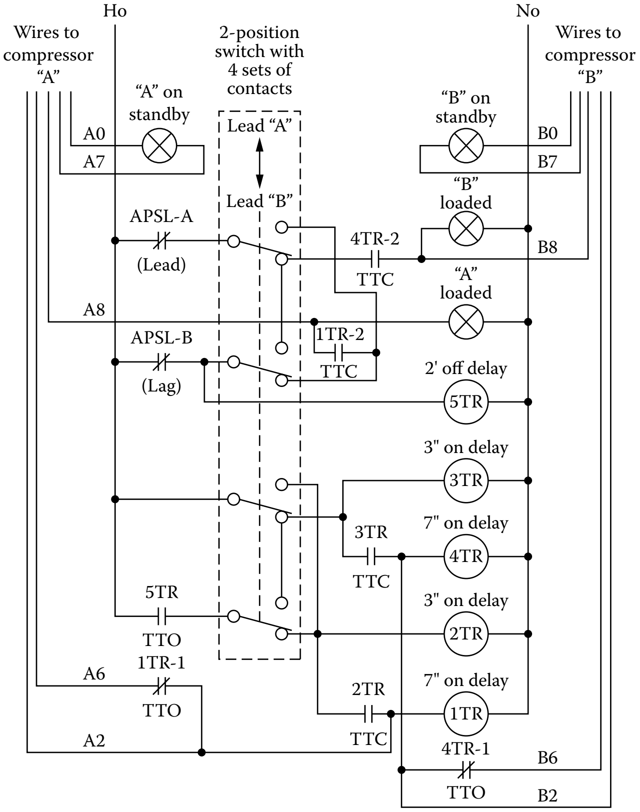


FIG. 8.15ddd

The wiring of a lead-lag selector switch that can automatically reverse the lead-lag relationship between two compressors.[2]

addition of a three-position "off" selector switch, illustrated in Figure 8.15eee. With the addition of the off-selector, any of the compressors can be selected for load, lag, or off duties, just by turning these conveniently located switches.

Depending on the position of the off selector, the identities of Compressors A and B in Figure 8.15ddd will be as follows:

Selected Off	*Compressor A*	*Compressor B*
1	3	2
2	1	3
3	1	2

For an integrated three-compressor system, only 30 wires need to be run if the remote annunciator is included. Without remote alarms, only 15 wires are needed (5 per compressor). The front face of a remote control cabinet of a three-compressor lead-lag system is illustrated in Figure 8.15fff.

Large Systems From the building blocks discussed in the previous paragraphs, an integrated remote controller can be configured for every compressor combination. Figure 8.15ggg

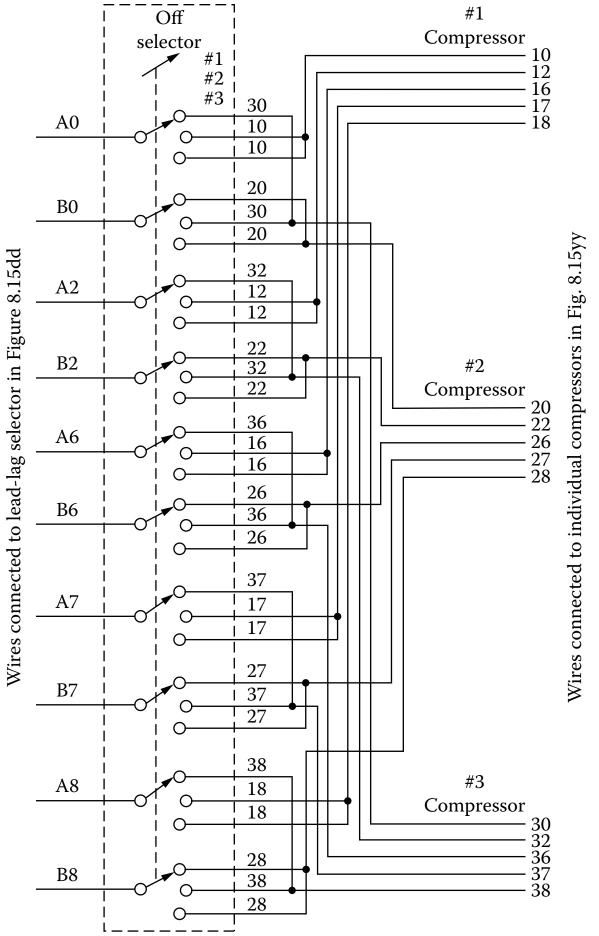


FIG. 8.15eee

The wiring of an "off" selector switch that can be used to identify the common spare compressor, or the compressor that is undergoing maintenance from among a set of three compressors.[2]

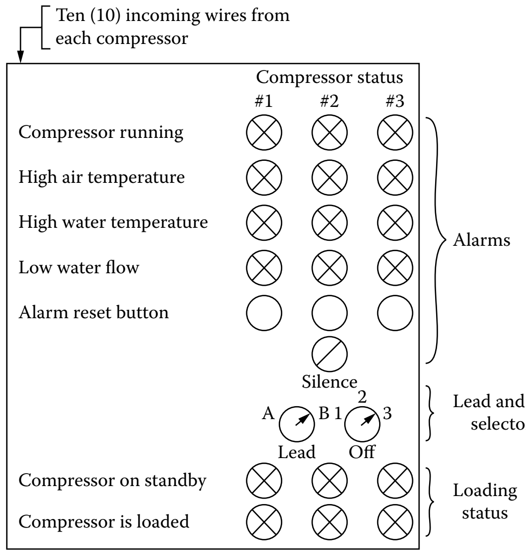


FIG. 8.15fff

The faceplate of the control panel that can be used to operate an integrated three-compressor system.[2]

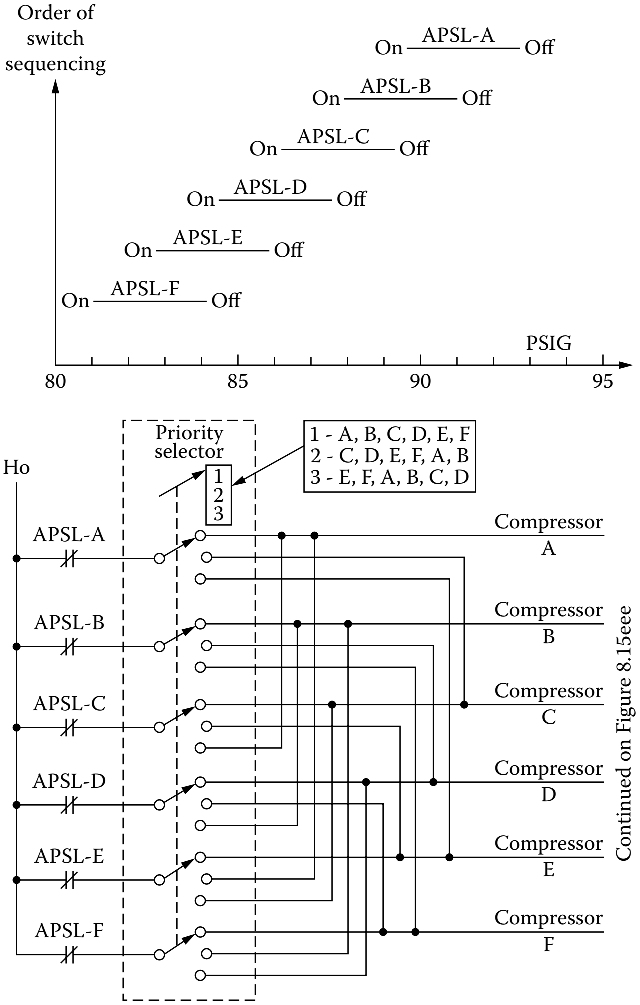


FIG. 8.15ggg

In large, multiple-compressor systems, a priority selector can be used to change the order of priorities of the compressors within the system.[2]

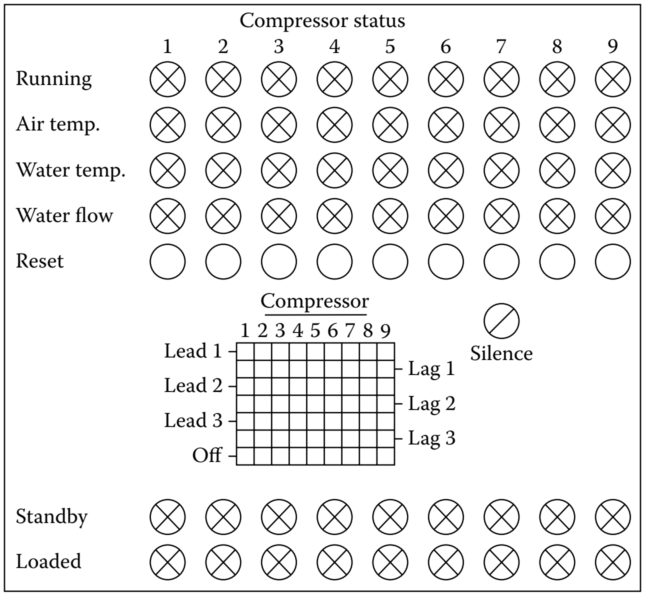


FIG. 8.15hhh

A control cabinet can provide full flexibility in remote control for an integrated nine-compressor system. Such a mechanically interlocked pushbutton station or its microprocessor controlled digital equivalent can be so designed that only one button can be pressed in each column or row, except in the bottom one. In the bottom row any or all buttons can be simultaneously pressed.[2]

illustrates how six stand-alone compressors might be integrated into a single load-following system.

As the system pressure drops below 90 psig, APSL-A closes, starting the first compressor. As the load rises, causing the system pressure to drop further, more switches will close, until at 80 PSIG all six switches will be closed and all six compressors will be running.

The "priority selector" shown in Figure 8.15ggg is provided for the convenient reconfiguration of compressor sequencing. When the selector is in position #1, as the load increases, the compressors will be started in the A, B, C, D, E, F order. When switched to position #2, the sequence becomes C, D, E, F, A, B. In position #3, the sequence is E, F, A, B, C, D.

If a system consists of nine compressors, and full flexibility in integrated remote control is desired, the faceplate of a control cabinet might look like Figure 8.15hhh. With this system, the operator can set the lead, lag, or off status for all nine compressors and can also select the order in which they are to come on.

Wired or Digital Bus From the previous discussion it can be seen that the building-block approach to compressor control system design is very flexible. Any number of stand-alone compressors can be integrated into automatic load-following control systems, with complete flexibility for priority, lead-lag, or off selection. The logic blocks described can be implemented either in hardware or in software. All wires and terminals can be prenumbered to minimize installation errors.

The potential benefits of compressor control system integration are listed below:

1. Unattended, automatic operation relieves operators for other tasks.
2. Automatic load-following eliminates the possibility of accidents caused by the loss of the air supply.
3. Because the supply and the demand are continuously and automatically matched, the energy cost of operation is minimized.

If priority or lead-lag switching is under computer control, this would also make the digital implementation of the system more desirable. On the other hand, if a dedicated, hard-wired system is preferred, because it is more familiar

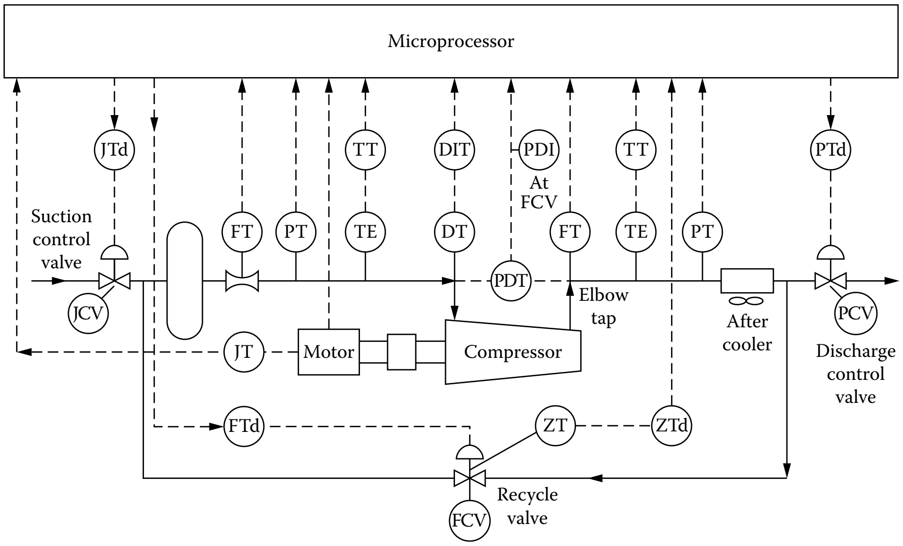


FIG. 8.15iii

A digital unit controller that is connected to the compressor I/O by a digital bus is well suited for both compressor optimization and for integrating the compressor controls into the total control system of the plant.

to the average operator, hardware implementation can be the proper choice.

CONCLUSIONS

As was shown in Figures 8.15g and 8.15j, the energy cost of operating a single centrifugal compressor at 60% average loading can be cut in half if optimized variable-speed control is applied. In the case of multiple compressor systems, similar savings can be obtained by the use of optimized load-following and supply-demand matching control strategies.

The full automation of compressor stations — including automatic start-up and shutdown — not only will reduce operating cost but will also increase operating safety as human errors are eliminated. Figure 8.15iii illustrates the use of a microprocessor-based digital unit controller. Such units can be provided with all the features that were discussed in this section in the form of software algorithms. Such digital systems provide flexibility by allowing changes in validity checks, alteration of limit stops, or reconfiguration of control loops.

References

1. McMillan, G. K., *Centrifugal and Axial Compressor Control,* Research Triangle Park, NC: Instrument Society of America, 1983.
2. Lipták, B. G., “Integrating Compressors into One System,” *Chemical Engineering*, January 19, 1987.
3. Shinskey, F. G., *Energy Conservation through Control*, Academic Press, 1978.
4. Magliozzi, T. L., “Control System Prevents Surging,” *Chemical Engineering*, May 8, 1967.
5. Baumeister, T. (ed.), *Mark’s Standard Handbook for Mechanical Engineers*, 8th edition, New York: McGraw-Hill, 1978, pp. 14–53.
6. Waggoner, R. C., “Process Control for Compressors,” paper given at Houston Conference of ISA, October 11–14, 1976.
7. Baker, D. F., “Surge Control for Multistage Centrifugal Compressors,” *Chemical Engineering,* May 31, 1982, pp. 117–122.

Bibliography

Adamski, R. S., “Improving Reliability of Rotating Machinery,” *InTech,* February 1982.

American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE), *ASHRAE Handbook,* Equipment Volume, Atlanta, GA: ASHRAE, 1983.

API Standard 614, “Lubrication, Shaft-Sealing, and Control-Oil Systems for Special-Purpose Applications,” Washington, D.C.: American Petroleum Institute, 1973.

API Standard 617, “Centrifugal Compressors for General Refinery Service,” 3rd ed., Washington, D.C.: American Petroleum Institute, 1973.

API Standard 618, “Reciprocating Compressors for General Refinery Service,” 2nd ed., Washington, D.C.: American Petroleum Institute, 1974.

Arant, J. B., “Compressor Instrumentation Systems,” *Instrumentation Technology,* April 1974.

Austin, W. E., “Packaged Air System Can Give Peak Operation, Cost Savings,” *Hydraulics and Pneumatics,* August 1979.

Baker, D. F., “Surge Control for Multistage Centrifugal Compressors,” *Chemical Engineering,* May 31, 1982, pp. 117–122.

Baumeister, T. (Ed.), *Mark’s Standard Handbook for Mechanical Engineers,* New York: McGraw-Hill, 1978.

Bently, D. E., “Machinery Protection Systems for Various Types of Rotating Equipment,” Minden, NV, Bently Nevada Corp., 1980.

Bloch, H. P., and Hoefner, J. J. *Reciprocating Compressors*, Gulf Professional Publishing, 1996.

Bogel, G. D., and Rhodes, R. L., “Digital Control Saves Energy in Gas Pipeline Compressors,” *InTech,* July 1982, pp. 43–45.

Braccini, G., “Electronic System for the Antisurge Control of Centrifugal Compressors,” *Quoderni Pignone*, No. 16.

Buzzard, W. S., "Controlling Centrifugal Compressors," *Instrumentation Technology,* November 1973.

Byers, R. H., Snider, M., and Brownstein, B., "Simulator to Test Compressor Research Facility Control System Software," *Military Electronics Defense Expo* '80, Geneva, Switzerland: Interavia, 1980, pp. 140–153.

Carrier Corp., *Air Conditioning News,* Syracuse, NY: Carrier Corp., May 12, 1980, pp. 3, 12.

"Centrifugal Compressors," *Bulletin 8282-C*, Woodcliff Lake, NJ: Ingersoll-Rand Co., 1972.

Cheremisinoff, N. P., *Compressors and Fans*, Prentice Hall, 1992.

"Compressibility Charts and their Application to Problems Involving Pressure-Volume-Energy Relations for Real Gases," *Bulletin P-7637*, Mountainside, NJ: Worthington-CEI Inc., 1949.

Compressor Handbook for the Hydrocarbon Processing Industries, Houston, TX: Gulf Publishing Co., 1979.

Cooke, L. A., "Saving Fuel at Pipeline Compressor Station," *Instrumentation Technology,* November 1977.

Davis, A., Keegan, P. J., and Peltzman, E., "Variable Capacity Compressor Controller," *Report No. DOE/CS/35224-T1,* U.S. Dept. of Energy, 1979.

Dussourd, J. L., Pfannebecker, G., Singhania, S. K., and Tramm, P. C., "Considerations for the Control of Surge in Dynamic Compressors Using Close Coupled Resistances," in *Centrifugal Compressors and Pump Stability, Stall, and Surge*, 1976, pp. 1–28.

Dwyer, J. J., "Compressor Problems: Causes and Cures," *Hydrocarbon Processing,* January 1973.

"Elliott Multistage Compressors," *Bulletin P-25,* Jeannette, PA: Elliott Division, Carrier Corp., 1973.

Engineering Data Book, 9th ed., Tulsa, OK: Natural Gas Processors Suppliers Association, 1972.

Fallin, H. K., and Belas, J. J., "Controls for an Axial Turboblower," *Instrumentation Technology,* May 1968.

Fehervari, W., "Asymmetric Algorithm Tightens Compressors Surge Control," *Control Engineering,* October 1977.

Filippini, V., "Surge Condition and Antisurge Control of a Centrifugal Compressor," *Quaderni Pignone,* No. 5.

"Gas Properties and Compressor Data," *Form 3519-C,* Woodcliff Lake, NJ: Ingersoll-Rand Co., 1967.

Gravdahl, J. T., et.al., *Compressor Surge and Stall,* Springer Verlag, May 1999.

Gaston, J. R., "Centrifugal Compressor Control," Research Triangle Park, NC: ISA Chemical and Petroleum Instrumentation Symposium Proceedings, 1974.

Gaston, J. R., "Improved Flow/Delta-P Antisurge Control System," Compressor-Engine Workshop, Pacific Energy Association, March 1981, pp. 1–13.

Gupta, B. P., and Jeffrey, M. F., "Compressor Controls Made Easy with Microprocessors," *Instrument Maintenance and Management,* Vol. 13, 1979, pp. 63–67.

Gupta, B. P., and Jeffrey, M. F., "Optimize Centrifugal Compressor Performance," *Hydrocarbon Processing, Computer Optimization,* June 1979.

Hallock, D. C., "Centrifugal Compressors—The Cause of the Curve," *Air and Gas Engineering,* January 1969.

Hansen, K. E., et al., "Experimental and Theoretical Study of Surge in a Small Centrifugal Compressor," *Journal of Fluids Engineering,* September 1981, pp. 391–395.

Hassenfuss, F., "Compressor Controls Coordinated for Carbon Dioxide Line," *Oil and Gas Journal*, July 31, 1972, pp. 91–95.

Hiller, C. C., and Glicksman, L. R., *Improving Heat Pump Performance via Compressor Capacity Control—Analysis and Test*, Vol. 1, Massachusetts, January 1976.

Kolnsberg, A., "Reasons for Centrifugal Compressor Surging and Surge Control," *Journal of Engineering for Power*, ASME, January 1979, pp. 79–86.

Langill, A. W., Jr., "Microprocessor-Based Control of Large Constant-Speed Centrifugal Compressors," *Advances in Test Measurement*, Vol. 19, Research Triangle Park, NC: ISA, 1982.

Lipták, B. G., "Compressor Control," *Control Magazine,* June 1999.

Moellenkamp, G., Scott, J., and Farmer, F., "Compressor Station Minicomputer System for Control, Monitoring, and Telemetry," 13th World Gas Conference, Paper IGU/C, London, England: International Gas Union, 1976.

Neerken, R. F., "Compressors in Chemical Process Industries," *Chemical Engineering,* January 20, 1975.

Nisenfeld, A. E., and Cho, C. H., "Parallel Compressor Control," *Australian Journal of Instrumentation and Control,* December 1977, pp. 108–112.

Nisenfeld, A. E., and Cho, C. H., "Parallel Compressor Control: What Should be Considered," *Hydrocarbon Processing,* February 1978, pp. 147–150.

Rammler, R., "Advanced Centrifugal Compressor Control," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Rammler, R., "Energy Savings through Advanced Centrifugal Compressor Performance Control," *Advances in Instrumentation,* Vol. 37, 1982.

Rammler, R., and Langill, A. W., Jr., "Centrifugal Compressor Performance Control," *Joint Automatic Control Conference,* AIChE, 1979, pp. 261–265.

Roberts, W. B., and Rogers, W., "Turbine Engine Fuel Conservation by Fan and Compressor Profile Control," Symposium on Commercial Aviation Energy Conservation Strategies, National Technical Information Service Document PC A16/MF A01, April 2, 1981, pp. 231–256.

Rutshtein, A., and Staroselsky, N., "Some Considerations on Improving the Control Strategy for Dynamic Compressors," *ISA Transactions,* Vol. 16, 1977, pp. 3–19.

Shinskey, F. G., *Process Control, Application, Design, and Tuning*, 4th ed., New York: McGraw-Hill Professionals, 1996.

Shinskey, F. G., "Interaction between Control Loops, Part II: Negative Coupling," *Instruments and Control Systems,* June 1976.

Smith, G., "Compressor Surge Control Achieved Using Unique Flow Sensor," *Pipeline Gas Journal*, December 1978, pp. 42, 46–47.

Staroselsky, N., "Better Efficiency and Reliability for Dynamic Compressors Operating in Parallel or in Series," Paper 80-PET-42, New York: ASME, 1980.

Staroselsky, N., and Ladin, L., "Improved Surge Control for Centrifugal Compressors,"*Chemical Engineering,* May 21, 1979, pp. 175–84.

Van Ormer, H. P., Jr., "Air Compressor Capacity Controls: A Necessary Evil, Vol. 1. Reciprocating Compressors," *Hydraulics and Pneumatics,* June 1980, pp. 67–70.

Waggoner, R. C., "Process Control for Compressors," *Advances in Instrumentation,* Vol. 31, 1976.

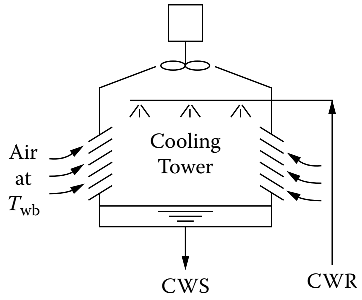
Warnock, J. D., "Are Your Compressors Wasting Energy?" *Instruments and Control Systems,* March 1977, pp. 41–45.

Warnock, J. D., "Typical Compressor Control Configurations," *Advances in Instrumentation,* Vol. 31, 1976.

White, M. H., "Surge Control for Centrifugal Compressor," *Chemical Engineering,* December 25, 1972.

8.16 Cooling Tower Control

B. G. LIPTÁK (1985, 1995, 2005)



Flow sheet symbol

Types of Designs:	a) Mechanical draft towers a1. Induced (fan at air outlet) cross-flow a2. Induced counter-flow a3. Forced (fan at air inlet) cross-flow a4. Forced counter-flow b) Natural draft towers c) Evaporative condensers
Water Losses:	Blowdown: –0.5 to 3.0% of circulated flow rate Evaporation: –1% for each 12.5°F (7°C) of cooling range Drift*:* 0.02% to 0.1% of circulated flow rate
Fan Types:	Induced and forced draft fans, axial-flow propellers
Heat Load:	BTU = 500(gpm)(water ΔT)
Approach:	5 to 10°F (2.8 to 5.6°C) for mechanical draft 20°F (11°C) for evaporative condensers
Fan Power Savings at Reduced Speeds:	50% at 80% speed 85% at 50% speed
Costs:	Initial investment: \$25 to \$40 per gpm of capacity Energy cost of operation: 0.01 BHP/gpm, or about \$6 to \$10 per year per gpm

INTRODUCTION

This section is devoted to the discussion of the control of cooling towers, while the next section deals with their optimization. Other related sections in this chapter describe the control and optimization of HVAC systems and of chillers.

Cooling towers are water-to-air heat exchangers that are used to discharge waste heat into the atmosphere. Their cost of operation is a function of the water and air transportation costs. In conventional cooling towers, the air and water flows are often constant. In optimized cooling towers the water and air flows are variable and are adjusted as a function of the load.

In this section the types of cooling tower design are described and the various techniques of load-following capacity controls are discussed. Cooling tower winterizing and blowdown controls are also covered.

DEFINITIONS

In order to make the discussion of the control aspects of cooling towers completely clear, first some related terms will be defined:[1]

APPROACH The difference between the wet-bulb temperature of the entering ambient air and of the leaving water temperature from the tower. The approach is a function of cooling tower capacity; a large cooling tower will produce a closer approach (colder leaving water) for a given heat load, flow rate, and entering air condition (Figure 8.16a). (Units: °F or °C.)

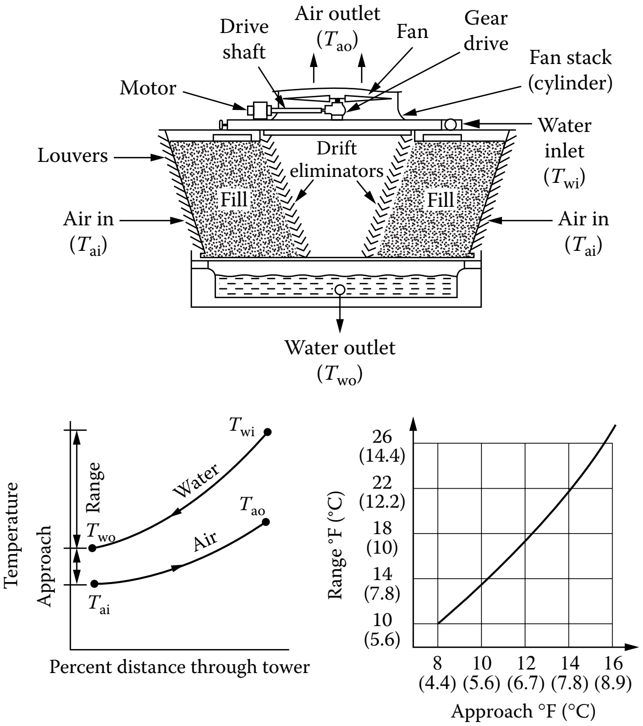


FIG. 8.16a

The mechanical cooling tower is a water-to-air heat exchanger. A direct relationship exists between water and air temperatures in a counterflow cooling tower. An increase in range will cause an increase in approach, if all other conditions remain unaltered.

BAY The area between two bents of lines of framing members; usually longitudinal.

BENT A line of structural framework composed of columns, girts, or ties; a bent may incorporate diagonal bracing members; usually transverse.

BLOWDOWN Water discharged to control concentration of impurities in circulated water. (Units: percentage of circulation rate.)

BRAKE HORSEPOWER (BHP) The actual power output of an engine or motor.

COLD WATER BASIN A storage device underlying the tower to receive the cold water from the tower and direct its flow to the pump suction line or sump.

COUNTERFLOW COOLING TOWER A cooling tower in which air flows in the opposite direction to the fall of water through the cooling tower.

CROSS-FLOW COOLING TOWER A cooling tower in which the direction of the airflow through the fill is perpendicular to the plane of the falling water.

DEW-POINT TEMPERATURE (TDP) The temperature at which condensation begins, if air is cooled under constant pressure (see Figure 8.16b).

DOUBLE-FLOW COOLING TOWER A cross-flow tower with two fill sections and one plenum chamber that is common to both fill sections.

DRIFT The water loss due to liquid droplets entrained and removed by the exhaust air. Usually under 0.2% of circulated water flow rate.

DRIFT ELIMINATOR An assembly constructed of wood or honeycomb materials that serves to remove entrained moisture from the discharged air.

DRY-BULB TEMPERATURE (TDB) The air temperature measured by a normal thermometer (see Figure 8.16b).

EVAPORATION LOSS Water evaporated from the circulating water into the atmosphere in the cooling process. (Unit: percentage of total GMP.)

FAN-DRIVE OUTPUT Actual power output (BHP) of drive to shaft.

$$\mathrm{BHP} = \frac{(\text{motor efficiency})(\text{amps})(\text{volts})(\text{power factor})1.73}{746} \tag{8.16(1)}$$

FAN PITCH The angle that a fan blade makes with the plane of rotation. (Unit: degrees from horizontal.)

FAN STACK (CYLINDER) Cylindrical or modified cylindrical structure in which the fan operates. Fan cylinders are used on both induced draft and forced draft axial-flow propeller-type fans.

FILLING That part of an evaporative tower consisting of splash bars, vertical sheets of various configurations, or honeycomb assemblies that are placed within the tower to effect heat and mass transfer between the circulating water and the air flowing through the tower.

FORCED DRAFT COOLING TOWER A type of mechanical draft cooling tower in which one or more fans are located at the air inlet to force air into the tower.

HEAT LOAD The heat removed from the circulating water within the tower. Heat load may be calculated from the range and the circulating water flow. (Unit: BTU per hour = gpm 500 [$T_{ctwr} - T_{ctws}$].)

LIQUID GAS RATIO (L/G) Mass ratio of water to airflow rates through the tower. (Units: lb/lb or kg/kg.)

LOUVERS Assemblies installed on the air inlet faces of a tower to eliminate water splash-out.

MAKEUP Water added to replace loss by evaporation, drift, blowdown, and leakage. (Unit: percentage of circulation rate.)

MECHANICAL DRAFT COOLING TOWER A tower through which air movement is effected by one or more fans. There are two general types of such towers: those that use forced draft with fans located at the air inlet and those that use induced draft with fans located at the air exhaust.

NATURAL DRAFT WATER COOLING TOWER (AIR MOVEMENTS) A cooling tower in which air movement is essentially dependent upon the difference in density between the entering air and internal air. As the heat of the

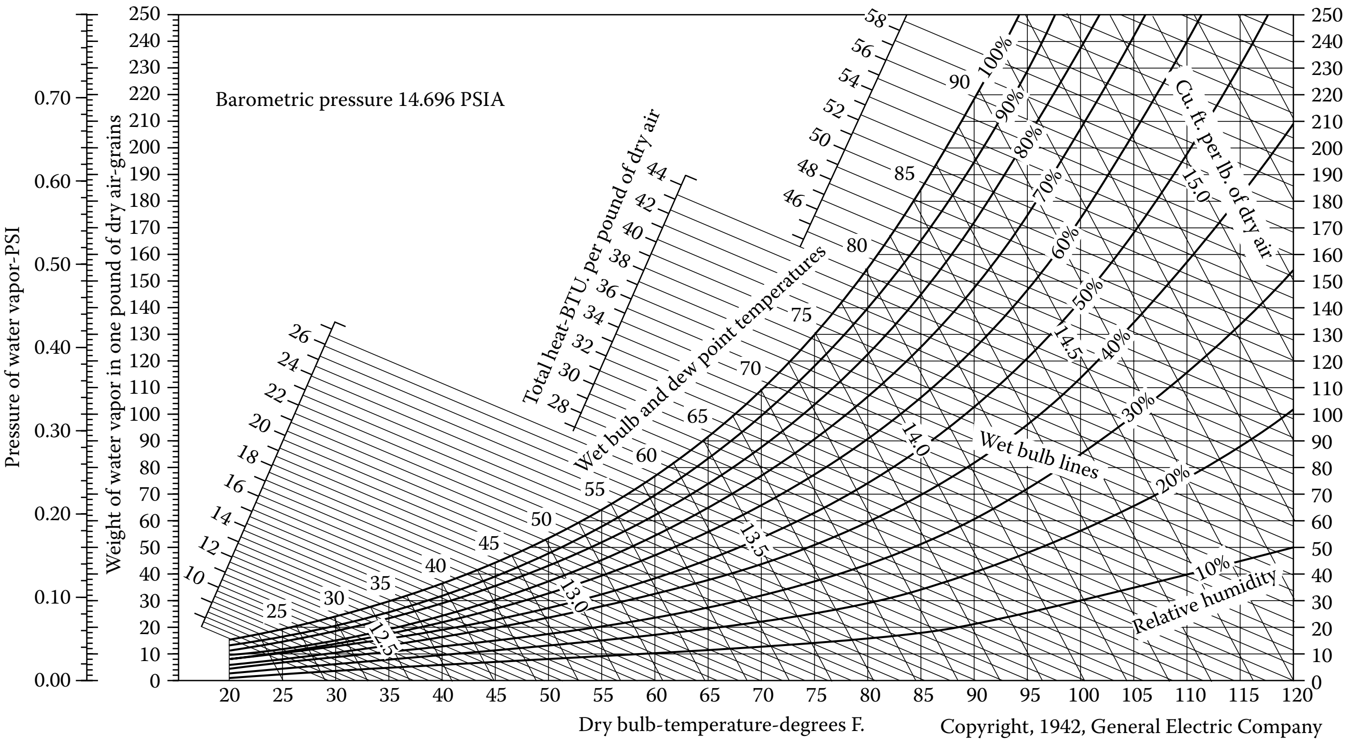


FIG. 8.16b

The psychrometric chart describes the following properties of air: Water vapor pressure, weight of water per pound of air, dry-bulb temperature, wet-bulb temperature, heat content per pound of air, volume per pound of air. The dew point temperature is the wet-bulb temperature when the relative humidity is 100%.

water is transferred to the air passing through the tower, the warmed air tends to rise and draw in fresh air at the base of the tower.

NOMINAL TONNAGE One nominal ton corresponds to the transfer of 15,000 BTU/hr (4.4 kW) when water is cooled from 95 to 85°F (35 to 29.4°C) by ambient air having a wet-bulb temperature of 78°F (25.6°C) and when the water circulation rate is 3 gpm (11.3 lpm) per ton.

PERFORMANCE The measure of a cooling tower's ability to cool water. Usually expressed in gallons per minute cooled from a specified hot water temperature to a specified cold water temperature with specific wet-bulb temperature. Typical performance curves for a cooling tower are shown in Figure 8.16c. This is a "7500 gpm at 105-85-78" tower, meaning that it will cool 7500 gpm of water from 105 to 85°F when the ambient wet-bulb temperature is 78°F.

PLENUM The enclosed space between the eliminators and the fan stack in induced draft towers, or the enclosed space between the fan and the filling in forced draft towers.

POWER FACTOR The ratio of true power (watts) to the apparent power (amps × volts).

PSYCHROMETER An instrument used primarily to measure the wet-bulb temperature.

RANGE (COOLING RANGE) The difference between the temperatures of water inlet and outlet, as shown in Figure 8.16a. For a system operating in a steady state, the range is the same as the average water temperature rise through the process loads. Accordingly, the range is determined by the heat load and water flow rate, not by the size or capability of the cooling tower. On the other hand, the range does affect the approach, and as the range increases, a corresponding increase in the approach will occur if all other factors remain unaltered (see Figure 8.16a).

RISER Piping that connects the circulating water supply line from the level of the base of the tower or the supply header to the tower inlet connection.

SPEED REDUCER A device for changing the speed of the driver in order to arrive at the desired fan speed.

STANDARD AIR Dry air having a density of 0.075 lb/ft^3 at 70°F and 29.92 in. Hg.

STORY The vertical dimension or area between two lines of horizontal framework ties, girts, joists, or beams. In wood frame structures, the story will vary from 5 to 7 ft (about 2 m) in height.

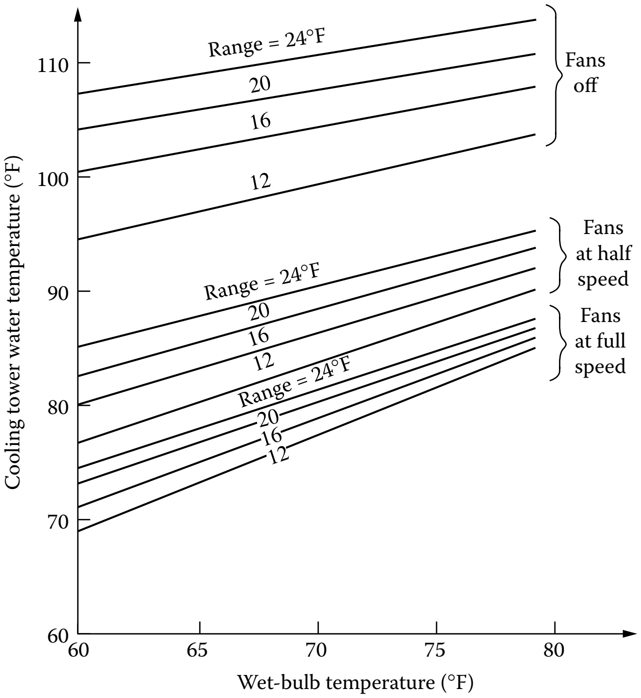


FIG. 8.16c

Cooling tower performance curves give the temperature of the generated cooling water as a function of the wet-bulb temperature of the air, the fan speed, and the range (the difference between the inlet and outlet temperatures of the tower water).

SUMP (BASIN) Lowest portion of the basin to which cold circulating water flows: usually the point of suction connection.

TONNAGE *See* Nominal Tonnage.

TOTAL PUMPING HEAD The total head of water, measured above the basin curb, required to deliver the circulating water through the distribution system.

TOWER PUMPING HEAD Same as the total pumping head minus the friction loss in the riser. It can be expressed as the total pressure at the centerline of the inlet pipe plus the vertical distance between the inlet centerline and the basin curb.

WATER LOADING Water flow divided by effective horizontal wetted area of the tower. (Unit: gpm/ft^2 or m^3/hr · m^2.)

WET-BULB TEMPERATURE (TWB) If a thermometer bulb is covered by a wet, water-absorbing substance and is exposed to air, evaporation will cool the bulb to the wet-bulb temperature of the surrounding air. This is the temperature read by a psychrometer. If the air is saturated with water, the wet-bulb, dry-bulb, and dew-point temperatures will all be the same. Otherwise, the wet-bulb temperature is higher than the dew-point temperature but lower than the dry-bulb temperature (see Figure 8.16b).

THE COOLING PROCESS

In a cooling tower, heat and mass transfer processes combine to cool the water. The mass transfer due to evaporation does consume water, but the amount of loss is only about 5% of the water requirements for equivalent once-through cooling by river water.

Cooling towers are capable of cooling within 5–10°F (2.8–5.6°C) of the ambient wet-bulb temperature. The larger the cooling tower for a given set of water and airflow rates, the smaller this approach will be. The mass transfer contribution to the total cooling is illustrated in the simplified psychrometric chart in Figure 8.16d.

Points A and D illustrate the conditions in which the ambient air might enter. After it transferred heat and absorbed mass from the evaporating water, the air leaves in a saturated condition described to point B. The total heat transferred in terms of the difference between entering and leaving enthalpies ($h_B - h_A$) is the same for both the AB and the DB processes.

The AC component of the AB vector represents the heat transfer component (sensible air heating), and the CB vector represents the evaporative mass transfer component (latent air heating). If the air enters at condition D, the two component vectors do not complement each other, because only one of them heats the air (EB). Therefore, the total vector DB results from the latent heating (EB) and sensible cooling (DE) of the air.

In terms of the water side of the DB process, the water is being sensibly heated by the air and cooled by evaporation. Therefore, as long as the entering air is not saturated, it is possible to cool the water with air that is warmer than the water.

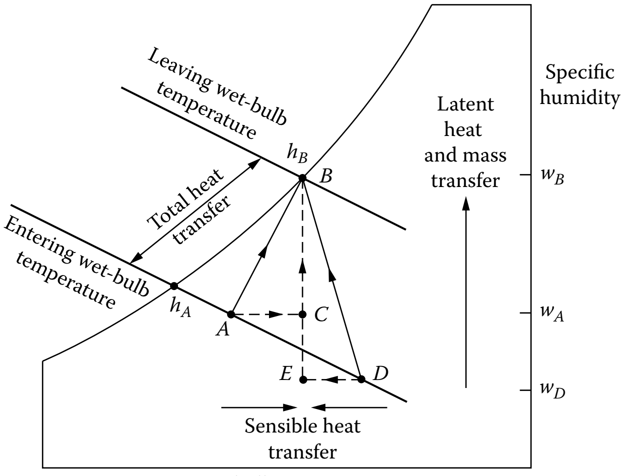
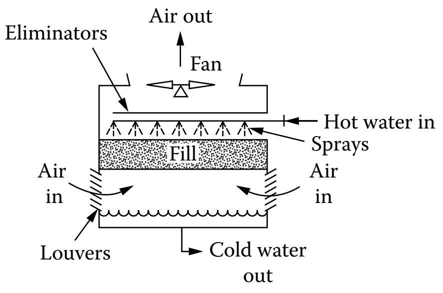
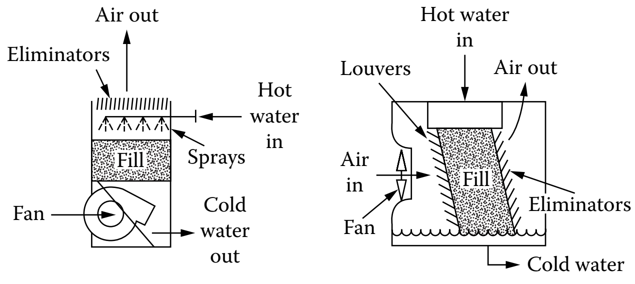


FIG. 8.16d

Heat and mass transfer processes combine to cool water. Total cooling is the sum of the latent cooling caused by the evaporation of water and the sensible heat transfer, which can either add to or reduce the latent cooling as a function of the relative temperatures of the entering air and water.[2]

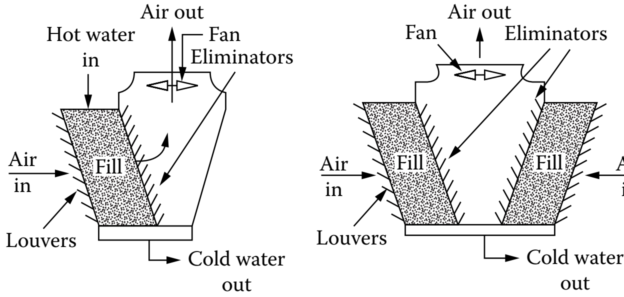


Induced-draft counterflow



Forced-draft counterflow

Force-draft crossflow



Induced-draft crossflow (single air entry)

Induced-draft crossflow (double air entry)

FIG. 8.16e

Mechanical draft cooling towers are distinguished by the location of the air fan (forced or induced draft), by the direction of the air flow (cross- or counterflow), and by the nature of the air entrance (single or double entry).[2]

MECHANICAL DRAFT COOLING TOWERS

The basic mechanical draft cooling tower designs are described in Figure 8.16e. For operational purposes, tower characteristic curves for various wet-bulb temperatures, cooling ranges, and approaches can be plotted against the water flow to airflow ratio in accordance with following equation:

$$\frac{KaV}{L} \sim \frac{(L)^n}{(G)} \qquad \textbf{8.16(2)}$$

where

K = overall unit of conductance, mass transfer between saturated air at mass water temperature, and main air stream: lb per hour (ft^2)(lb/lb)

a = area of water interface per unit volume of tower: ft^2 per ft^3

V = active tower volume per unit area: ft^3 per ft^2

L = mass water flow rate: lb per hour

G = airflow rate: lb dry air per hour

n = experimental coefficient: varies from -0.35 to -1.1 and has an average value of -0.55 to -0.65

Characteristic Curves

From the characteristic curves of the cooling tower it is possible to determine tower capability, effect of wet-bulb temperature, cooling range, water circulation, and air delivery, as shown in Figure 8.16f. The ratio KaV/L is determined from test data on hot water and cold water, wet-bulb temperature, and the ratio of the water flow to the airflow.

For air conditioning applications the thermal capability of cooling towers is usually stated in terms of nominal tonnage. This tonnage is defined as the heat dissipation of 15,000 BTU/h (4.4 kW) per condenser ton (kW) and a water circulation rate of 3 gpm/ton (0.054 l/s per kW) cooled from 95 to 85°F (35 to 29.4°C) at 78°F (25.6°C) wet-bulb temperature.

It may be noted that the subject tower would be capable of handling a greater heat load (flow rate) when operating in

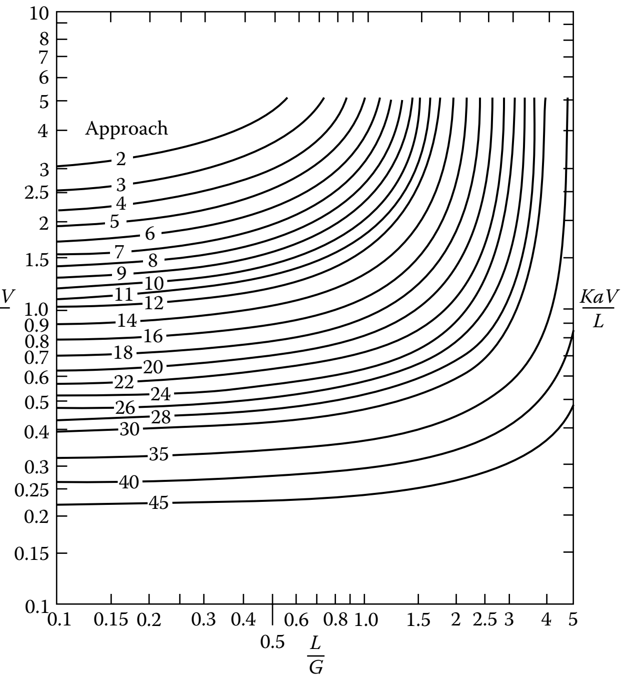


FIG. 8.16f

Characteristic curves of cooling towers (based on 64°F [17.8°C] air wet-bulb temperature and 18°F [10°C] cooling water temperature drop range).

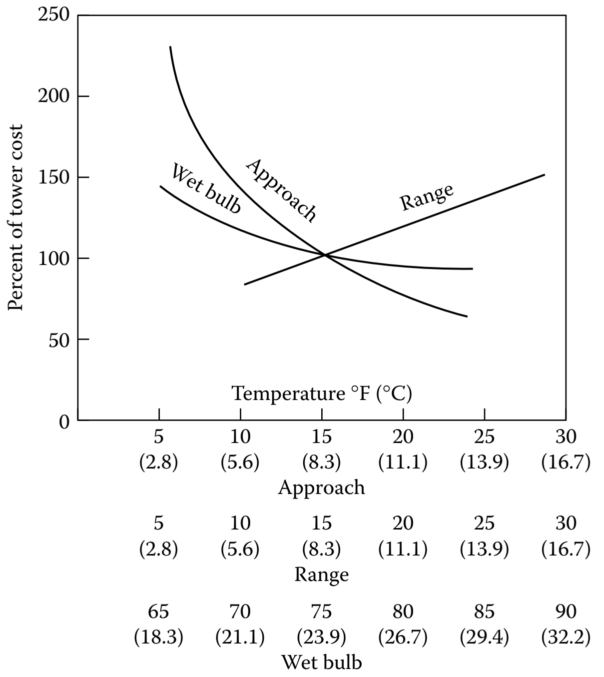


FIG. 8.16g

The smaller the cooling tower (the lower the initial investment), the higher will be the operating cost. If the tower cost is considered to be 100 when it is sized for an approach and a range of 15°F when the wet bulb is 75°F, the data in this figure give the comparative cooling tower costs for a range of other sizing basis. (Courtesy of Foster Wheeler Corp.)

a lower ambient wet-bulb region. For operation at other flow rates, tower manufacturers usually provide performance curves covering a range of at least 2 to 5 gpm per nominal ton (0.036 to 0.09 lps/kW).

The initial investment required for a cooling tower is essentially a function of the required water flow rate, but this cost is also influenced by the design criteria for approach, range, and wet-bulb. As shown in Figure 8.16g, cost tends to increase with range and tends to decrease with a rise in approach or wet-bulb. The initial investment is about \$25–\$40 per gpm of tower capacity, whereas the energy cost of operation is approximately 0.01 BHP/gpm, or about \$6–\$10 per year per gpm. This means that the cost of operation reaches the cost of initial investment in 5–10 years, depending on energy costs.

CONTROLS

On HVAC-type applications, the cooling tower controls usually consist of on/off or sequential load controls, safety interlocks, and blowdown and winter operation controls. In industrial applications, the controls tend to be continuous and are usually controlling sophisticated, variable-speed fans (Figure 8.16h) as discussed in the next section.

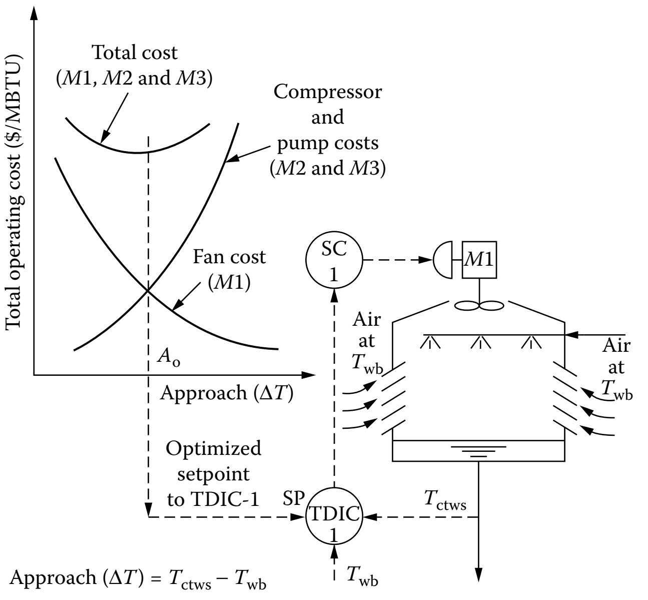


FIG. 8.16h

In order to minimize the total cost of operation, the speed of the cooling tower's air fan can be so throttled so that the optimum approach (the approach corresponding to the minimum cost of operation) is maintained.

Load Controls

The cooling tower is an air-to-water heat exchanger. The controlled variables are the supply and return water temperatures, while the manipulated variables are the air and water flow rates. Manipulation of these flow rates can be continuous, through the use of variable-speed fans and pumps, or can be incremental, through the cycling of single- or multiple-speed units.

If the controlled variable is the temperature of the water supplied by the cooling tower and the manipulated variable is the fan speed, a change in cooling load will result in a change in the operating level of the fans. If the fans are single-speed devices, the airflow rate is changed by cycling the fan units on or off as a function of load. This is illustrated in Figure 8.16i and can be done by simple sequencers or by more sophisticated digital systems.

In most locations, the need to operate all fans at full speed is required for only a few thousand hours every year. Fans running at half speed consume approximately one-seventh of the design air horsepower but produce over 50% of the design air rate (cooling effect). Therefore in 98% of the cases, two-speed motors are a wise investment for minimizing operating costs.

Operating Interlocks

Figure 8.16j describes a simple fan starter. A three-position "hand/off/automatic" switch controls the fan. When the switch is placed in "automatic" (contacts 3 and 4 connected),

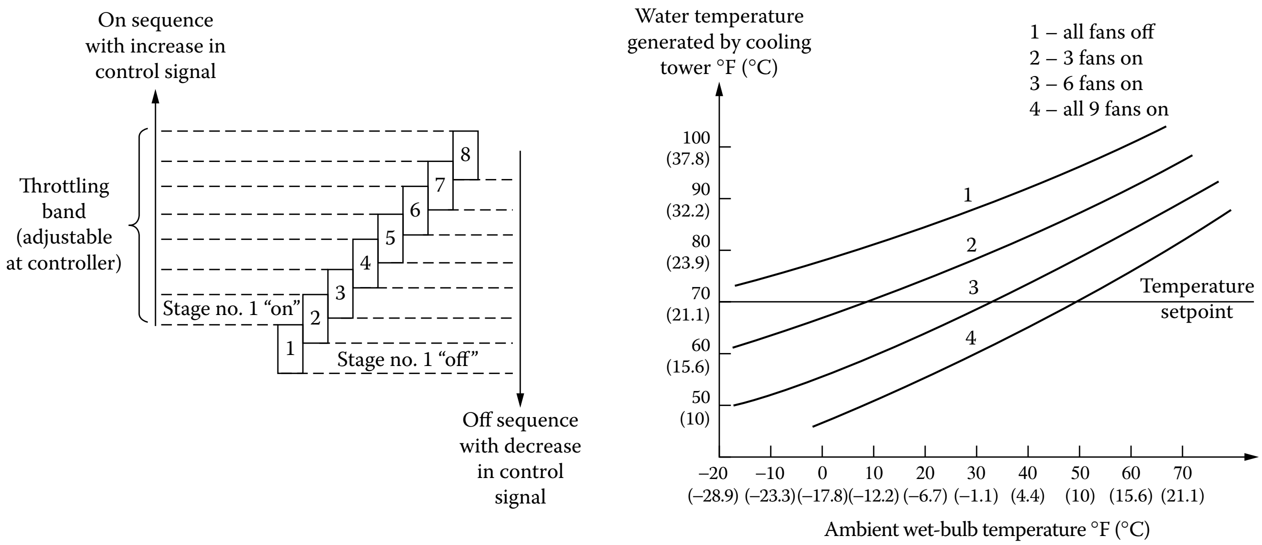


FIG. 8.16i

Illustrations of load-following control accomplished by fan cycling. The figure on the left illustrates how the output signal from the approach controller (TDIC-1 in Fig. 8.16h) sequences the fan stages on and off as a function of the load. The figure on the right shows the effect of fan cycling on a nine-cell tower that is designed for a full capacity of 160,000 gpm (608,000 l/m) water loading with an 84°F (29°C) inlet and 70°F (21°C) outlet temperature, operating at 50% load.[3]

the status of the interlock contact *I* determines whether or not the circuit is energized and the fan is on.

The sequencing of this contact *I* is described in Figure 8.16i. The purpose of the auxiliary motor contact *M* is to energize the running light *R* whenever the fan motor is on. The parallel hot lead to contact 6 allows the operator to check quickly to see whether the light has burned out.

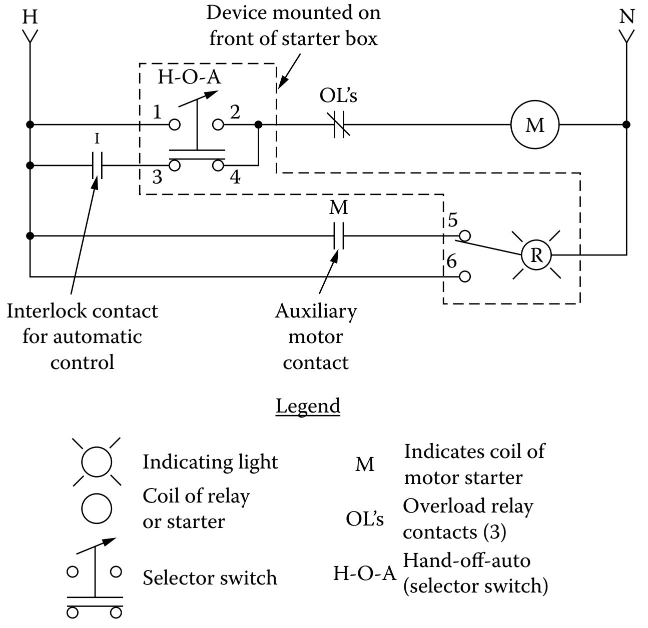


FIG. 8.16j

The wiring diagram of a simple fan starter circuit, manually controlled by a three-position (hand-off-automatic) switch.

The amount of interlocking is usually greater than that shown in Figure 8.16j. Some of the additional interlock options are described in the paragraphs below.

Safety Interlocks

Most fan controls include some safety overrides in addition to the overload (OL) contacts shown in Figure 8.16j. These overrides might stop the fan if fire, smoke, or excessive vibration is detected. They usually also provide a contact for remote alarming.

Most fan controls also include a reset button that must be pressed after a safety shut-down condition is cleared before the fan can be restarted.

When a group of starters is supplied from a common feeder, it may be necessary to make sure that the starters are not overloaded by the inrush currents of simultaneously started units. If this feature is desired, a 25 sec time delay is usually provided to prevent other fans from starting until that time has passed.

Larger fans should be protected from overheating caused by too frequent starting and stopping. This protection can be provided by a 0–30 min time delay guaranteeing that the fan will not be cycled at a frequency faster than the delay time setting.

When two-speed fans are used, added interlocks are frequently provided. One interlock might guarantee that even if the operator starts the fan to operate at high speed, it will operate for 0–30 sec in low before advancing automatically to high. This makes the transition from off to high speed more gradual.

Another interlock might guarantee that when the fan is switched from high to low, it will be off the high speed for 0 to 30 sec before the low-speed drive is engaged. This will give time for the fan to slow before the low gears are engaged.

When the fan can also be operated in reverse, further interlocks are needed. One interlock should guarantee that the fan is off for 0–2 min before a change in the direction of rotation can take place. This gives time for the fan to come to rest before making a change in direction. Another interlock will guarantee that the fan can operate only at slow speed in the reverse mode and that reverse operation cannot last more than 0–30 min. This limitation is desirable because when the fan is in reverse in the winter, ice tends to build up on the blades and the gears can wear excessively.

If the same fans can be controlled from several locations, interlocks are needed to resolve the potential problems with conflicting requests. One method is to provide an interlock that determines the location that is in control. This can be a simple local/remote switch or a more complicated system.

When many locations are involved, conflicts are resolved by interlocks that either establish priorities between control locations or select the lowest, highest, safest, or other specified choice from the control requests. In installations with multiple control centers, it is essential that feedback be provided so that the operator is always aware not only of the actual status of all fans but also of any conflicting requests that are possibly coming from other operators or computers.

EVAPORATIVE CONDENSERS

The evaporative condenser is a special cooling tower type, shown in Figure 8.16k. In this unit the induced air and sprayed water flow concurrently downward, and the process vapors travel upwards inside the multipass condenser tubes. A pressure controller (PIC) modulates the amount of cooling, so that it will match the load. The continuous water spray on that makes the tube surfaces wet, improves heat transfer and maintains tube temperatures.

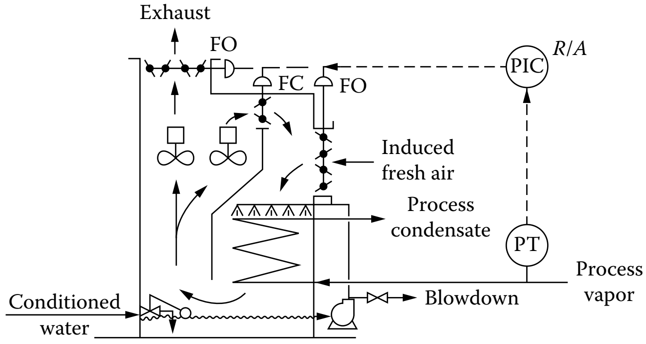


FIG. 8.16k
In evaporative condensers, the pressure of the process vapors can be controlled by modulating the proportion of the humid air, which is recirculated. (Adapted from Reference 4.)

The approach of these cooling towers is around 20°F (11°C), and the wet-bulb temperatures of the cooling air can be varied by the recirculation of humid (or saturated) air. The resulting condensing temperatures are lower than the condensers, which are cooled by dry air. The higher heat capacity and lower resistance to heat transfer make these units more efficient than the dry-air coolers.

The turndown capability of these units is also very high, because the fresh-air flow can be reduced all the way to zero. Both the exhaust and the fresh air dampers can be closed while the recirculation damper simultaneously opens to maintain the internal air flow rate constant. As more humid air is recirculated, the internal wet-bulb temperature rises and the heat flow drops off. In Figure 8.16k the heat flow is effectively linear with fresh-air flow.[4]

SECONDARY CONTROLS

In addition to the previously discussed load controls and safety interlocks, controls are also needed for winter operation, water quality, and blowdown controls and other miscellaneous purposes.

Winter Operation

In case of a power failure during subfreezing weather, the tower basin should be electrically heated and should be provided with an emergency draining system. Another method of freeze protection is to provide bypass circulation illustrated in Figure 8.16l. In this system, a thermostat (TSL) detects the outlet temperature of the cooling tower water, and

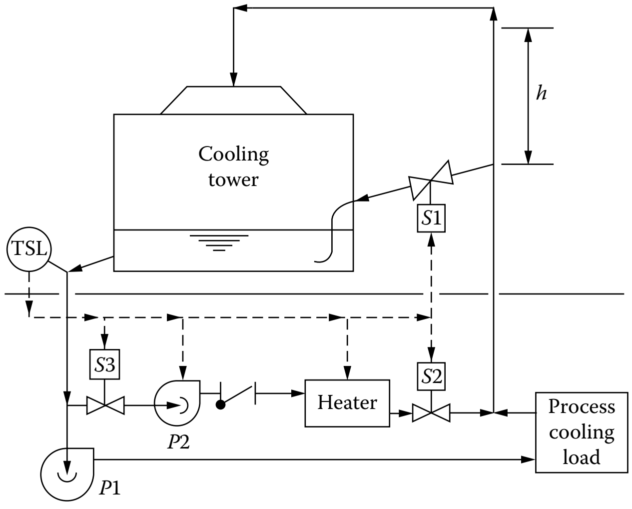


FIG. 8.16l
Protection against freezing can be provided in the winter by starting a bypass pump if the temperature drops to near freezing (TSL) and by sending this bypassed water through a heater and then back to the tower.

when it drops below 40°F (4.4°C), the bypass circulation (located in a protected indoor area) is started. This circulation is terminated when the water temperature rises to approximately 45°F (7.2°C).

The thermostat is wired to open the solenoids *S*1, *S*2, and *S*3 first and then to start the circulating pump (*P*2), which usually is a small, $^1/_4$ or $^1/_3$ hp (186.5 or 248.7 W) unit. The bypass line containing solenoid *S*1 is sized to handle the flow capacity of *P*2, with a pressure drop that is less than the height of the riser pipe (*h*). This makes sure that when *P*1 is off, no water will reach the top of the column.

In the northern regions, in addition to the *S*1 bypass shown in Figure 8.16l, there is a full-size bypass (not shown in Figure 8.16l). When the main (P1) pump is started in the winter, this full bypass is opened; it does not allow the water to be sent to the top of the tower until its temperature is approximately 70°F (21°C). Once the water reaches that temperature, the bypass is closed because the water is warm enough to be sent to the top of the tower without danger of freezing. As the water temperature rises further, fans are turned on to provide added cooling.

When open cooling towers are operated at freezing temperatures, the induced draft fans must be reversed periodically in order to deice the air intakes. The need for deicing can be determined through visual observation, through remote closed-circuit TV inspection, or by comparing the load and ambient conditions to past operating experience, as have already been discussed under interlocks.

All exposed pipes that will contain water when the tower is down should be protected by electric tape or cable tracing and by insulation. The sump either should be drainable to an indoor auxiliary sump or should be provided with auxiliary steam or electric heating.

Closed cooling towers should either operate on antifreeze solutions or be provided with supplemental heat and tracing. If the second method is used, the system must be drained if power failure occurs.

Blowdown Controls

The average water loss by blowdown is 0.5–3.0% of the circulating water rate. The loss is a function of the initial quality of the water and the amount and concentration of the dissolved natural solids and chemicals added for protection against corrosion or build-up of scale on the heat transfer surfaces. Because the cooling tower is a highly effective air scrubber, it continuously accumulates the solid content of the ambient air on its wet surfaces, which are then washed off by the circulated water.

The “normal” condition of the circulated water can be defined arbitrarily as:

pH	6–8
Chloride as NaCl	under 750 ppm
Total dissolved solids	under 1500 ppm

The blowdown requirement can be determined as follows:

$$B = \frac{([E+D]/N) - D}{1-(1/N)} \qquad \mathbf{8.16(3)}$$

where

B = blowdown in percentage (typically, for *N* = 2, 0.9%; for *N* = 3, 0.4%; for *N* = 4, 0.24%)
E = evaporation in percentages (typically 1% for each 12.5°F [7°C] of cooling range)
D = drift in percentage (typically 0.1%)
N = number of concentration relative to initial water quality

Problems and Solutions In cooling towers, the major causes for concern are delignification, which is the loss of the binding agent of the cellulose. It is caused by the use of oxidizing biocides, such as chlorine. It can also be caused by excessive bicarbonate alkalinity; biological growth, which can clog the nozzles and foul the heat exchange equipment; corrosion of the metal components (this corrosion should be less than 3 mils per year without pitting); general fouling by a combination of silt, clay, oil, metal oxides, calcium and magnesium salts, organic compounds, and other chemical products that can cause reduced heat transfer and enhanced corrosion; and scaling by crystallization and precipitation of salts or oxides (mainly calcium carbonate and magnesium silicate) on surfaces.

The treatment techniques to prevent these conditions from occurring are listed in Table 8.16m. The blowdown must meet the water quality standards for the accepting stream and must not be unreasonably expensive. It is also possible to make the water system a closed-cycle system (no blowdown) by ion-exchange treatment.

Miscellaneous Controls

The minimum basin level is maintained by float-type, makeup level control valves. The maximum basin level is guaranteed by overflow nozzles, which are sized to handle the total system flow. In critical systems, additional safety is provided by the use of high/low level alarm switches.

Torque and vibration detectors can be used to safeguard the operational safety of the fans and of the fan drives. Low-temperature alarm switches can also be furnished as warning devices signaling the failure of the freeze protection controls.

The need for flow balancing among multiple cells is another reason for using additional valves and controls. (Advanced water distribution controls will be discussed in the next section.) In conventional systems, the warm water is usually returned to the multicell system through a single riser, and a manifold is provided at the top of the tower for water distribution to the individual cells. Balancing valves should be installed to guarantee equal flow to each operating cell. If two-speed fans are used, further economy can be gained by lowering the water flow when the fan is switched to low speed. It is also advisable to install equalizer lines

TABLE 8.16m
Potential Cooling Tower Problems and their Solutions

Problem	*Factors*	*Causative Agents*	*Corrective Treatments*
Wood deterioration	Microbiological Chemical	Cellulolytic fungi Chlorine	Fungicides Acids
Biological growths	Temperature Nutrients pH Inocula	Bacteria Fungi Algae	Chlorine Chlorine donors Organic sulfurs Quatenary ammonia
General fouling	Suspended solids Water Velocity Temperature Contaminants Metal oxides	Silt Oil	Polyelectrolytes Polyacrylates Lignosulfonates Polyphosphates
Corrosion	Aeration pH Temperature Dissolved solids Galvanic couples	Oxygen Carbon dioxide Chloride	Chromate Zinc Polyphosphate Tannins Lignins Synthetic organic compounds
Scaling	Calcium Alkalinity Temperature pH	Calcium carbonate Calcium sulfate Magnesium silicate Ferric hydroxide	Phosphonates Polyphosphates Acid Polyelectrolytes

between tower sumps to eliminate imbalances caused by variation in flow or pipe layouts.

Another reason for considering the use of two-speed fans is to have the ability to lower the associated noise level at night or during periods of low load. Switching to low speed will usually lower the noise level by approximately 15 dB.

CONCLUSIONS

One of the least understood and most neglected unit operations in the processing industries is the cooling tower. Its pumps are frequently constant speed with three-way or bypass valves used to circulate the excess water. This excess should not have been pumped in the first place, because the process does not require it.

Meeting a variable load with a constant supply by wasting the excess is also frequently practiced in fan operation. In some installations, fan speeds cannot be changed at all; in others, they can be changed only manually, which can require seasonal adjustments in some extreme cases.

As the rating of the tower fans and pumps usually adds up to several hundred horsepower, their yearly operating cost is in the hundreds of thousands of dollars. As optimization can cut this in half, the added costs for controls are justified. This is the subject of the next section.

References

1. Dickey, J. B., *Evaporative Cooling Towers,* Marley Cooling Towers Co., 1978.
2. ASHRAE, "Cooling Towers," in *ASHRAE Handbook—HVAC Systems and Equipment,* New York: ASHRAE, 2004.
3. Wistrom, G., "Cold Weather Operation of Crossflow Cooling Towers," *Plant Engineering,* September 8, 1975.
4. Lipták, B. G. (ed.), "Cooling Towers: Their Design and Application," in *Environmental Engineer's Handbook,* Lewis Publishers, 1997.

Bibliography

ASHRAE, "Chilled and Dual-Temperature Water Systems," in *ASHRAE Handbook,* Equipment Volume, Atlanta, GA: ASHRAE, 1983.

ASHRAE, "Cooling Towers," in *ASHRAE Handbook-HVAC Systems and Equipment,* New York: ASHRAE, 2004.

ASHRAE, "Heating, Ventilating, and Air Conditioning Systems," in *ASHRAE Handbook,* June 2004.

Baechler, R. H., Blew, J. O., and Duncan, C. G., "Cause and Prevention of Decay of Wood in Cooling Towers," ASME Petroleum Division Conference, New York: ASME, 1961.

Baker, D. R., "Durability of Wood in Cooling Towers," *Technical Bulletin R-62-W-1,* Kansas City, MO: The Marley Co., 1962.

Brownell, D. L., *Handbook of Applied Thermal Design,* Taylor and Francis, 1999.

Donohue, J. M., "Cooling Towers — Chemical *Treatment,*" *Industrial Water Engineering,* May 1970.

Frayne, C., *Cooling Water Treatment,* Chemical Publishing Co., January 2001.

Hartman, T., "Ultra-Efficient Cooling with Demand-Based Control," *HPAC Engineering,* December 2001.

Kern, D. Q., *Process Heat Transfer,* New York: McGraw-Hill, 1950, p. 563.

Koch, J., *Untersuchung und Berenchnung und Kuehlwerken,* VDI Forschungsheft no. 404, Berlin, 1940.

Lipták, B. G. (ed.), "Cooling Towers: Their Design and Application," in *Environmental Engineer's Handbook,* Lewis Publishers, 1997.

Lipták, B. G., "Cutting in Half the Cost of Cooling," *Control,* November 2003.

Luyben, W. L., *Process Modeling, Simulation, and Control,* New York: McGraw-Hill, 1989.

Mathews, R. T., "Some Air Cooling Design Considerations," *Proceedings of the American Power Conference,* 1970.

McMillan, G., "Good Tuning: A Pocket Guide," *ISA,* June 2000.

Nottage, H. B., "Merkel's Cooling Diagram as a Performance Correlation for Air-Water Evaporative Cooling Systems," *ASHRAE Transactions,* Vol. 47, p. 429, 1941.

Nussbaum, O. J., "Dry-Type Cooling Towers for Packaged Refrigeration and Air Conditioning Systems," *Refrigeration Science and Technology,* 1972.

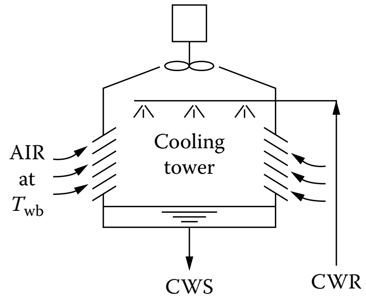
Roper, M. A., *Energy Efficient Chiller Control,* BSRIA Ltd., September 2000.

Walker, W. H., Lewis, W. K., McAdams, W. H., and Gilliand, E. R., *Principles of Chemical Engineering,* New York: McGraw-Hill, 1937, p. 480.

Witt, M. D., "Successful Instrumentation and Control Design," *ISA,* 2003.

8.17 Cooling Tower Optimization

B. G. LIPTÁK (1985, 1995, 2005)



Flow sheet symbol

Savings through Optimization:	A 20% reduction in either air or water flow will result in a 50% reduction in fan or pump operating costs. A 50% speed reduction cuts operating cost by 85%.
Water Losses:	Blowdown:- 0.5 to 3.0% of circulated flow rate Evaporation: 1% for each 12.5°F (7°C) of cooling range Drift*:* 0.02% to 0.1% of circulated flow rate
Heat Load:	BTU = 500 (gpm)(water ΔT)
Approach:	5 to 10°F (2.8 to 5.6°C) for mechanical draft 20°F (11°C) for evaporative condensers
Costs:	Initial investment: $25 to $40 per gpm of capacity Energy cost of operation: 0.01 BHP/gpm, or about $6 to $10 per year per gpm.

INTRODUCTION

The plant's demand for cooling water determines the load on the cooling tower. The most efficient way of meeting this variable demand is to use variable-speed water pumps (Figure 8.17a). As the cooling tower is a water-to-air heat exchanger, when a drop in load causes the water flow rate to drop, the air flow on the other side of this exchanger can also be reduced.

In optimized cooling towers, the water and air flows are both controlled by variable-speed devices. The goal of cooling tower optimization is to maximize the amount of heat discharged into

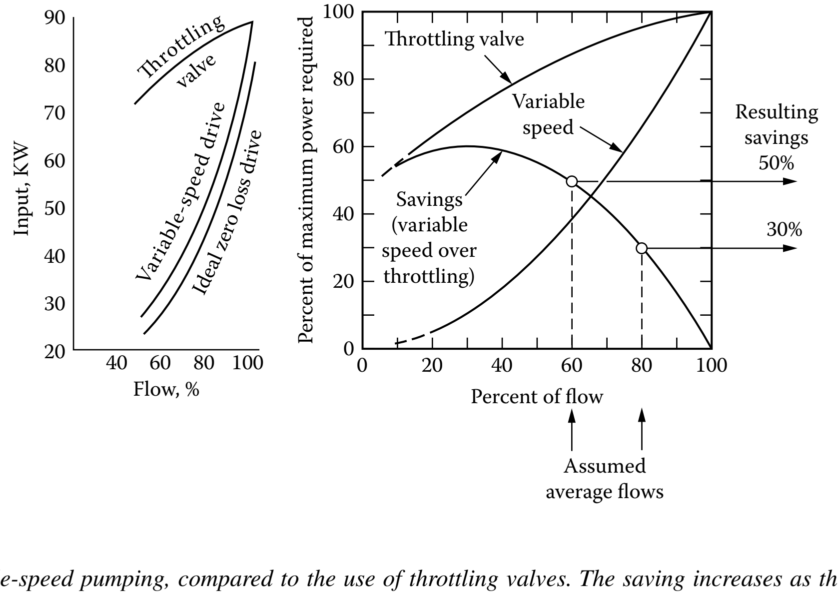


FIG. 8.17a
The savings due to variable-speed pumping, compared to the use of throttling valves. The saving increases as the flow rate drops.

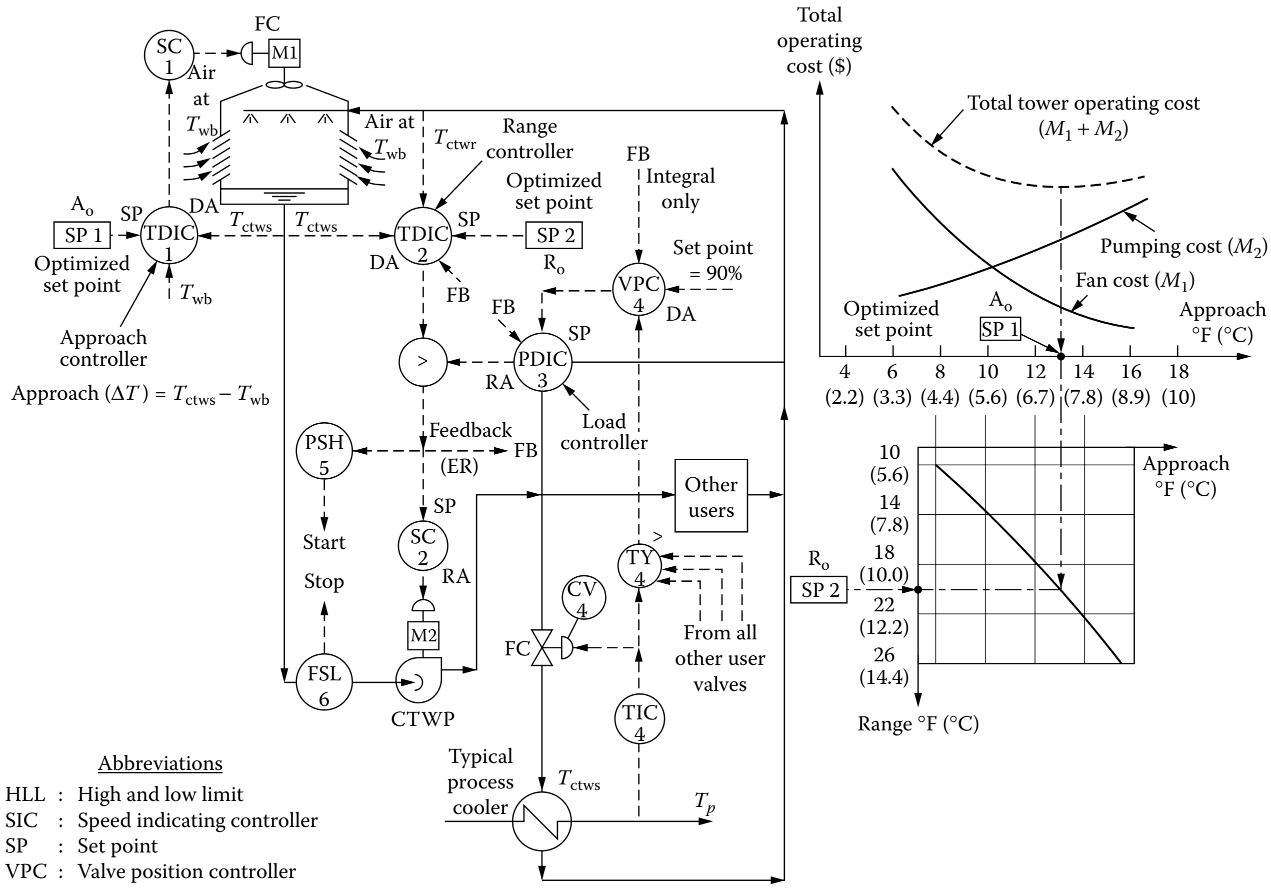


FIG. 8.17b

Optimization minimizes the unit cost of cooling by minimizing the operating speeds of both the cooling tower fans and pumps.

the atmosphere per unit of operating cost invested. The optimization strategies discussed in this section are applicable not only to standard mechanical draft cooling towers but to evaporative condensers and natural draft cooling towers.

The optimization of cooling towers as parts of chiller systems has been discussed in Section 8.13. The discussion here will concentrate on cooling systems, where the cooling water from the cooling towers is not used to generate chilled water but is pumped directly to cool the process.

MINIMIZING OPERATING COST

The cost of fan operation can be reduced by allowing the cooling tower water temperature (T_{ctws} in Figure 8.17b) to rise, thereby increasing the approach ($T_{ctws} - T_{wb}$) at which the tower operates. As shown in Figure 8.16c, the approach can be increased to a point at which the fans are off and their operating cost is zero. Under most load conditions, however, this would not produce a low enough cooling water temperature for the process.

The Optimum Approach

As shown in Figure 8.17b, as the approach, and therefore T_{ctws}, rises, the temperature difference across all process coolers ($T_p - T_{ctws}$) is reduced. This will cause the process temperature (T_p) to rise and its controller (TIC-4) to further open the coolant valve, CV-4. In order to reduce the process temperature, more and more water must be pumped. Therefore, an increase in approach will result in an increase in water transportation, and consequently the pumping costs will rise.

If the actual total operating cost is plotted against the approach (as in the right side of Figure 8.17b), the fan costs tend to drop and the pumping costs tend to rise with an increase in approach. If the operating cost of some of the users is also affected by cooling water temperature (this was the case when the water is used in chiller condensers, as in Figure 8.17c), then the total cost model should also consider that effect. Once a total operating cost curve is obtained, the optimum approach can be found by looking for the minimum point on that curve.

Cost Curves

The data for the cost curves are empirically collected and are continually updated through the actual measurement of fan and pump operating costs. Consequently, for any combination of load and ambient conditions there is a reliable prediction

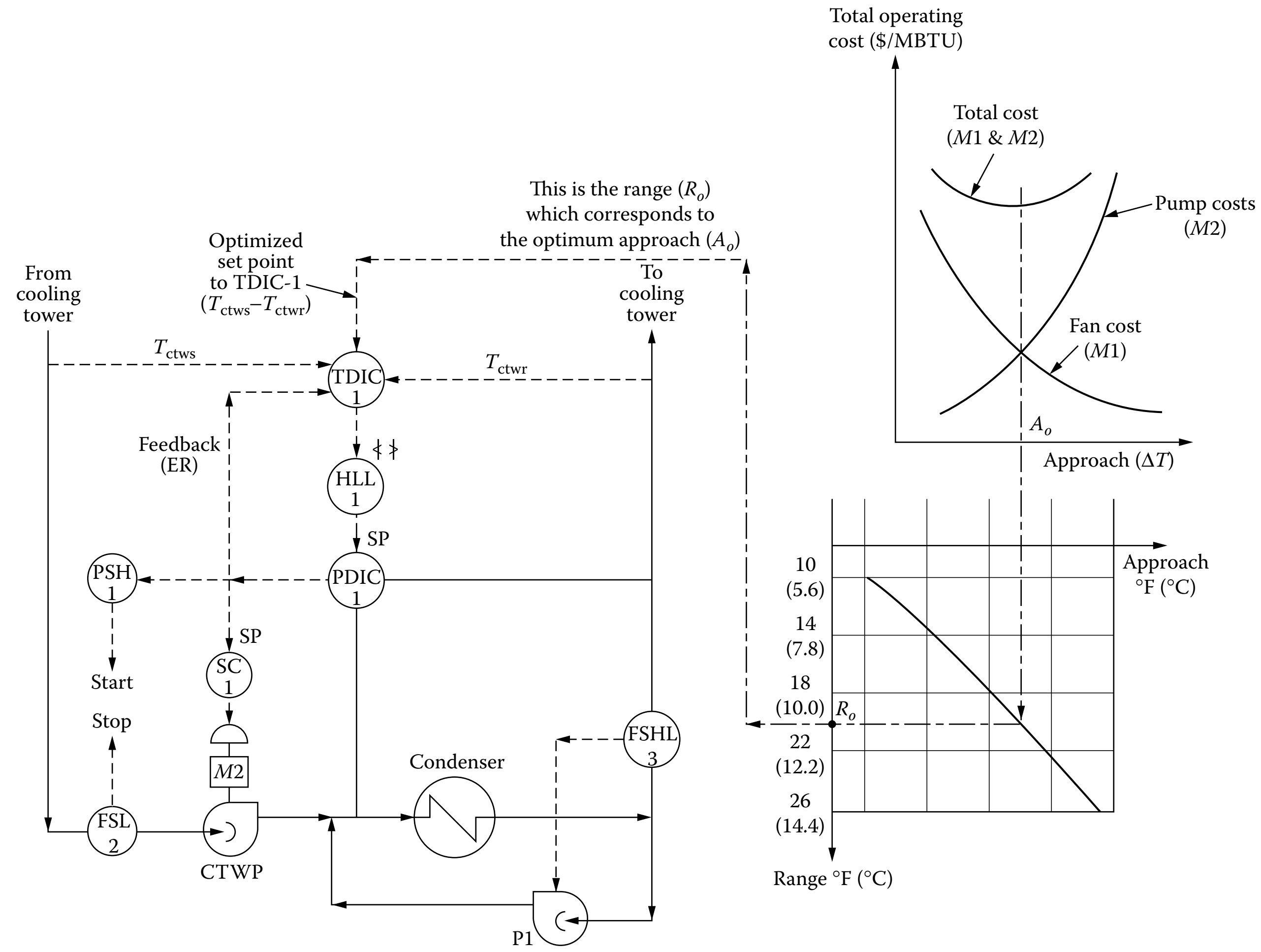


FIG. 8.17c

Total operating cost can be minimized by TDIC-1 controlling the range of the cooling tower (the difference between the return and supply temperatures) at the value that corresponds to the minimum cost of operation.

of optimum approach setting (A_o in Figure 8.17c). Once the initial prediction is set, that setting can be refined by adjusting it in 0.5°F increments, if the total operating cost is lowered by such adjustments.

As is also shown in the right side of Figure 8.17b, there is an empirical relationship between the values of approach and range, if all other conditions remain unaltered. Therefore, if the optimum value of approach (SP-1-A_o) has been determined, the corresponding optimum range value (SP-2-R_o) under the prevailing load conditions can be obtained from this curve. This is how the optimized set points (SP-1 and SP-2) of the approach and range controllers are determined and updated.

Supply Temperature Optimization

As shown in Figure 8.17b, an optimization control loop is required in order to maintain the cooling tower water supply continuously at an economical minimum temperature. This minimum temperature is a function of the wet-bulb temperature of the atmospheric air. The cooling tower cannot generate a water temperature that is as low as the ambient wet-bulb but it can approach it (hence the term "approach").

As is shown in Figure 8.17b, as the approach increases, the cost of operating the cooling tower fans drops, but the cost of pumping increases. The optimum approach is the one that will result in a minimum total cost operation. In Figure 8.17b, this ΔT is the set point of TDIC-1. This optimum approach A_o will increase if the load on the cooling tower increases or if the ambient wet-bulb temperature decreases.

If the cooling tower fans are centrifugal units or if the blade pitch is variable, the optimum approach can be obtained by continuous throttling. If the tower fans are two-speed or single-speed units, the output of TDIC-1 will incrementally start and stop the fan units in order to maintain the optimum approach, as was described in Figure 8.16i.

Return Temperature Optimization

Figure 8.17b also shows the controls needed for optimizing the return water temperature. TDIC-2 is the range controller, which is set by the optimized set point of SP-2. This is the range value corresponding to the optimum approach (R_o). TDIC-2 throttles the water circulation rate in order to maintain the range at its optimum value. The output of TDIC-2 sets the speed of the cooling tower water circulating pump, if none of the process user valves are nearly full open. This is guaranteed by PDIC-3.

The outputs of both TDIC-2 and PDIC-3 are sent to a high-signal selector, which guarantees that when one of the user valves is nearly full open and therefore nearly out of control, the needs of the users take priority over range control. The load is detected by PDIC-3, and if its output is higher than that of the range controller (TDIC-2), it will be selected to set the pump speed. This load-following optimization loop guarantees that all cooling water users in the plant will always be satisfied while the range is being optimized.

This is done by selecting (with TY-4) the set point of PDIC-3 so that if the most-open cooling water valve opening is less than 90%, the PDIC-3 set point is decreased; however, if the valve opening exceeds 90%, the PDIC-3 set point is increased.

In this way, a condition is maintained that allows all users to obtain more cooling (by further opening their supply valves) if needed, while the differential pressure of the water is continuously optimized. The VPC-4 set point of 90% is adjustable. Lowering it gives a wider safety margin, which might be required if some of the cooling processes served are very critical. Increasing the set point maximizes energy conservation at the expense of the safety margin.

Benefits of Optimization

An additional benefit of this load-following optimization strategy is that because all cooling water valves in the plant are opened up as the water ΔP across the users is minimized, valve cycling is reduced and pumping costs are lowered. The reduction in pumping costs is a direct result of the opening of all cooling water valves, which reduces the pressure drop across them. A side benefit of opening the valves is that valve cycling is eliminated when the valve openings are moved away from the unstable region near the closed position.[1]

In order for the control system in Figure 8.18b to be stable, it is necessary to use an integral-only controller for VPC-4, with an integral time that is tenfold that of the integral setting of PDIC-3. This control mode selection is needed to allow the optimization loop to be stable when the valve opening measurement signal that is selected by TY-4 is either cycling or noisy.

A high limit setting on the output of VPC-4 (not shown in Figure 8.17b) is also recommended to make sure that it will not drive the cooling water pressure to unsafe or undesirably high values. Because this limit can block the VPC-4 output from affecting the pump speed, it is necessary to protect against reset windup in VPC-4. This is done through the external feedback signal (FB), which protects the controllers TDIC-2, PDIC-3, and VPC-4 from reset windup.

A side benefit of the optimization system shown in Figure 8.17b is that it brings attention to design errors. For example, if PDIC-3 is in control most of the time, that shows that the control valves and water pipes are undersized in the plant. Similarly, if TY-4 consistently selects the same user valve, it shows that the water supply to that user is undersized. When such design errors are corrected, either by adding local booster pumps or by replacing undersized valves and pipes, control will automatically return to TDIC-2.

Therefore, in a well-designed water distribution network, the range optimizer TDIC-2 will be in control most of the time and PDIC-3 will override it only during process upsets that cause excessive cooling loads. Thus, PDIC-3 should operate as a safety override that becomes active only under emergency conditions and guarantees that no user will ever run out of cooling water.

Starting Additional Pumps

When the cooling tower water pump station consists of several pumps, only one of which is variable-speed, additional pump increments are started when PSH-5 signals that the pump speed controller (SC-2) set point is at its maximum. When the load is dropping, the excess pump increments can be stopped on the basis of flow, detected by FSL-6. In order to eliminate pump cycling, the excess pumping increment is only turned off when the actual total flow corresponds to less than 90% of the capacity of the remaining pumps.

The load-following optimization system described in Figure 8.17b will float the total flow rate from the cooling tower to achieve maximum overall economy.

Return Water Distribution and Balancing

In cases where a large number of cooling tower cells are available to receive the returning cooling water, it is also desirable to automatically distribute the water flows to the various cells as a function of the operation of their associated fans. In other words, the water flows to all cells whose fans are at high speed should be equal and high. Similarly, the water flows to cells whose fans are operating at low speeds should receive water at equal low flow rates. Cells with their fans off should receive water at equal minimum flow rates.

The normal water flow rate ranges from 2 gpm to 5 gpm per ton when the fan is at full speed. Figure 8.17d illustrates the relationship between water flow and water temperature (T_{ctws}) or approach. Under the illustrated conditions, a 20% increase in water flow rate will increase the approach from 7 to 9°F, while a 20% decrease in flow will reduce it from 7 to 5°F.

Water distribution balancing is often done manually, but it can also be done automatically, as shown in Figure 8.17e. Here, the total flow is used as the set point of the ratio flow controllers (FFICs). If the ratio settings are the same, the total

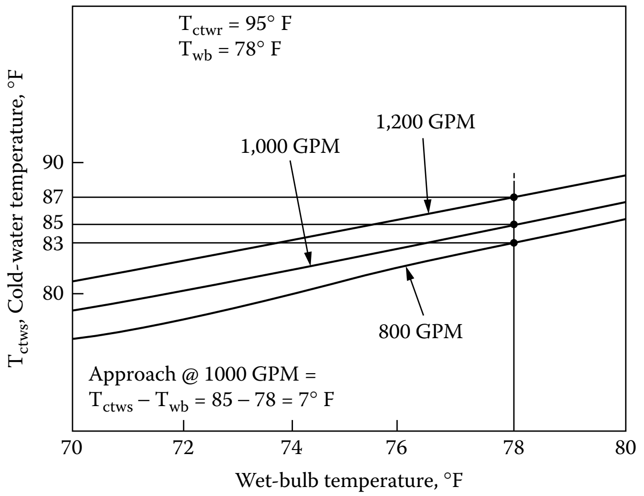


FIG. 8.17d

This figure shows the characteristic curves of a particular cooling tower that, when receiving a return water flow rate of 1000 gpm/cell at a temperature of 95°F and air at a wet-bulb temperature of 78°F, will generate a supply water temperature of 85°F (approach of 7°F and range of 10°F). In this case, if the demand for cold water rises to 1200 gpm/cell, the cold water temperature generated by the tower will rise by 2°F, while if the flow drops to 800 gpm/cell, it drops by 2°F. (From Reference 2.)

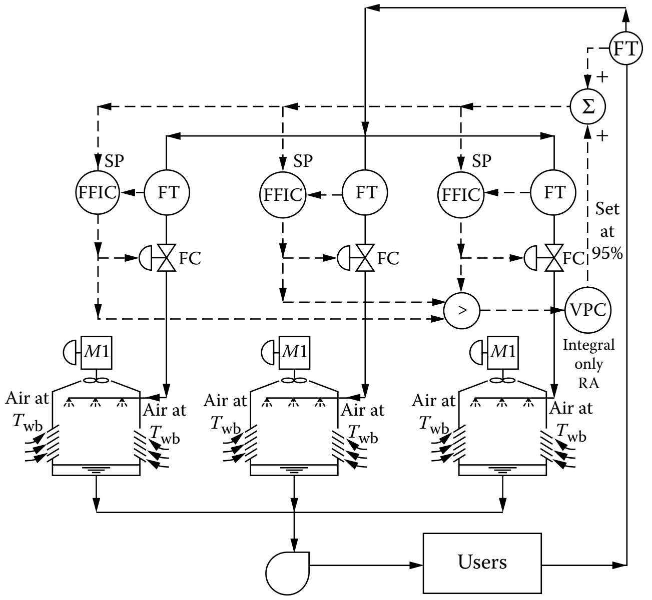


FIG. 8.17e

The distribution of the return water among several cooling tower cells can satisfy the dual goals of sending a preset percentage to each and to do that at a minimum pumping cost. The second goal is served by the valve position controller (VPC), which increases the measurement of all flow ratio controllers (FFICs) until the most open valve opens to 90%.

flow is equally distributed. The ratio settings can be changed manually or automatically to reflect changes in fan speeds. Naturally, the total of the ratio settings must always be 1. Therefore, if one ratio setting is changed, all others should also be modified. This, too, can be done automatically.

The purpose of the control system in Figure 8.17e is not only to distribute the returning water between the cells correctly but also to make sure that this is done at minimum cost. The cost in this case is the energy cost of pumping, which will be minimum when the pressure drop through the distribution control valves is minimum.

The valve position controller (VPC) in Figure 8.17e serves to keep this pressure drop to a minimum. It does that because as long as even the most open valve is not nearly fully open, the VPC adds a positive bias to all the set point signals of all the flow ratio controllers (FFICs). As a result, all valves will open and keep opening until the most open valve reaches the desired 90 or 95% opening. This technique provides correct flow distribution, while keeping the cost of pumping at a minimum.

When cells are manually balanced, it is not unusual to find all balancing valves throttled. It is also often the case that the same water flow is being sent to the tower cells, no matter if their fan is on or off. Both of these conditions will increase operating costs. The savings from automatic balancing can more than justify the control instruments required.

CONCLUSIONS

Cooling towers are simple devices, and possibly for that reason, their optimization is often neglected. Yet, the motors of the cooling tower fans and pumps in a plant can add up to many hundreds of horsepower. The operation of these motors can cost several hundred thousand dollars each year. That operating cost can be cut in half by optimization. Optimization is achieved by meeting the variable cooling load of the plant by the minimum water and air flows that are needed. Optimization also can include the cost-effective balancing of the distribution of the returning water among the tower cells.

A desirable side effect of optimization is the automatic indication of design defects in pipe and valve sizing and the increased level of safety, by making sure that no process cooling load is ever neglected.

References

1. Lipták, B. G., "Optimizing Controls for Chillers and Heat Pumps," *Chemical Engineering,* October 17, 1983.
2. Baumeister, T. (ed.), *Mark's Standard Handbook for Mechanical Engineers,* 8th edition, New York: McGraw-Hill, 1978, pp. 9–74.
3. Lipták, B. G., "Envelope Optimization," *Control Magazine,* November 1998.

Bibliography

Blevins, T. L. et al., "Advanced Control Unleashed," *ISA*, October 2002.

Frayne, C., *Cooling Water Treatment*, Chemical Publishing Co., January 2001.

Hartman, T., "Ultra-Efficient Cooling with Demand-Based Control," *HPAC Engineering,* December 2001.

Lipták, B. G., "Envelope Optimization," *Control Magazine,* November 1998.

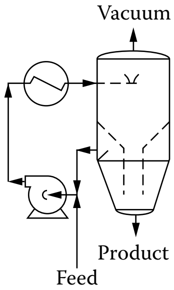
Lipták, B. G., "Cutting in Half the Cost of Cooling," *Control Magazine,* November 2003.

Luyben, W. L., *Process Modeling, Simulation, and Control,* New York: McGraw-Hill, 1989.

McMillan, G., "Good Tuning: A Pocket Guide," *ISA,* June 2000.

8.18 Crystallizer Controls

C. J. SANTHANAM (1970, 1985) **B. G. LIPTÁK** (1995)
B. G. LAKATOS (2005)



Flow sheet symbol

INTRODUCTION

This section describes the control of industrial crystallizers. Crystallization requires supersaturation, which can be achieved by cooling, evaporation, vacuum cooling, dilution, and chemical reaction. The control of multiple-effect crystallizers is also covered.

Crystallization is a widely used technique in the process industry and is used for the purpose of purification and separation of substances, as well as for the production of good-quality crystals. It is a practical method to obtain concentrated chemical substances in particulate solid form that have desirable properties such as good flow characteristics, handling convenience, suitability for packaging, and pleasant appearance.

Crystallization can be carried out from solution, vapor, or melt. Here, only solution crystallization is discussed. The most commonly used methods for creating supersaturation in the solution phase include cooling, evaporation, vacuum cooling, dilution, or chemical reaction. The choice of the method depends on the solubility of the solute, feed solution composition, product specifications, and other engineering considerations.

Most crystallizers in industrial use operate on a continuous basis, but batch and semibatch crystallizers are also widely used to produce fine chemicals and special-effect high-added-value materials such as pharmaceuticals, pigment, agrochemicals, or catalysts.

This section is devoted to the subject of controlling basic continuous crystallizers. It will describe the basic crystallization techniques, while focusing on the measurement and control practices used in this industry.

THE CRYSTALLIZATION PROCESS

The fundamental driving force for crystallization is the difference in chemical potential between the crystallization substance in the solid and liquid phases, but the common engineering practice is to use supersaturation as the driving force of concentration. The crystallization process has three basic steps: (1) generation of supersaturation, (2) formation of nuclei, and (3) growth of nuclei into crystals.

Supersaturation is usually expressed as the supersaturation ratio

$$S = \frac{c}{c'} \qquad \textbf{8.18(1)}$$

or the concentration difference

$$\Delta c = c - c' \qquad \textbf{8.18(2)}$$

where

c = concentration of the crystallizing substance in the solution
c' = equilibrium concentration (solubility) of the crystallizing substance in the solution

In most cases, the solubility c' of a solute increases with the increase of temperature, but there are a few exceptions to this rule. The concept of supersolubility and the existence of the metastable zone are useful in understanding the behavior of the crystallization process. Figure 8.18a shows a temperature-solubility plot for a typical salt; the plot is divided into three zones: (1) stable under saturated zone, where crystallization is not possible; (2) metastable (supersaturated) zone between the solubility and supersolubility, where spontaneous crystallization is not possible; and (3) labile (supersaturated) zone, where spontaneous crystallization is probable.

The rate of nucleation and the crystal growth rate are controlled by supersaturation. The ideal process would be a stepwise procedure, but nucleation (i.e., formation of new crystals), because of the secondary nucleation, cannot be eliminated in a growing mass of crystals. The influence of supersaturation on nucleation and growth rates is shown in Figure 8.18b. Whereas the dependence of growth rate on supersaturation is linear or moderately nonlinear, the nucleation rate increases exponentially with supersaturation.

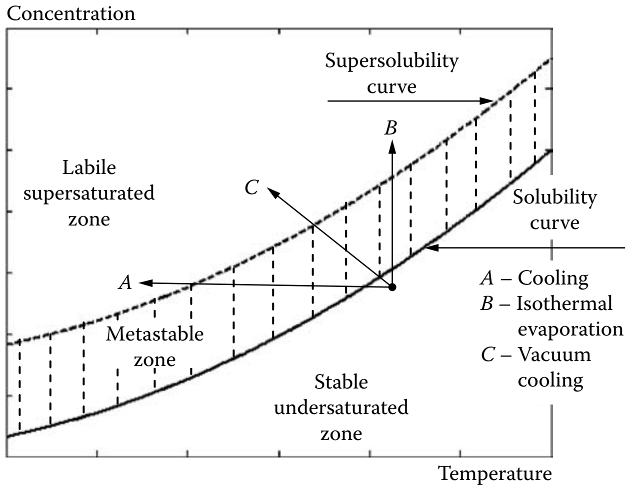


FIG. 8.18a
The zones of saturation and solubility plots and the methods by which the state of supersaturation can be achieved.

Figure 8.18b shows three regions of supersaturation, corresponding, in principle, to the zones in Figure 8.18a: (1) metastable, where nucleation is very low and growth predominates; (2) intermediate, where nucleation becomes larger, but growth is still significant; and (3) labile, where nucleation predominates.

Crystal Size Distribution

Industrial crystallizers usually yield a crystalline product that has a wide crystal size distribution (CSD). This CSD affects the behavior of the product in succeeding operations such as filtration, drying, transport, and storage, and is also defined by customer specifications. Because very small crystals are difficult to filtrate, dry, or handle, crystallizer design and control usually are directed toward reasonably large crystals by minimizing nucleation.

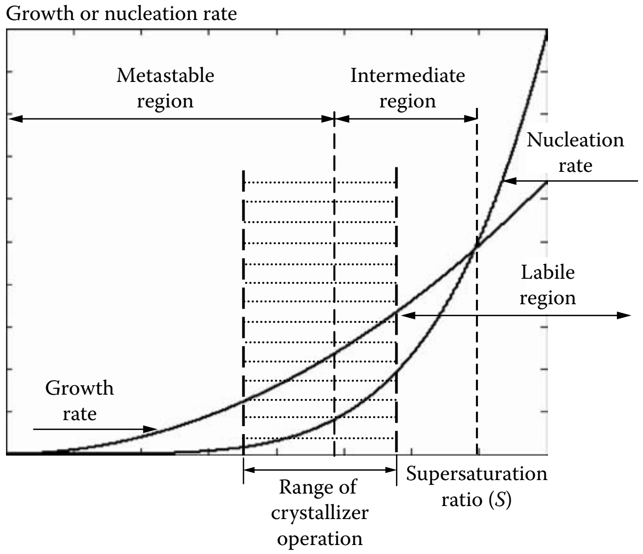


FIG. 8.18b
The influence of level of supersaturation on nucleation and the crystal growth.

Usual industrial practice involves sufficiently low supersaturation to minimize supersaturation while being adequate for reasonable growth. At the same time, among others in the pharmaceutical industry, there is also an increasing need to produce high specific-surface micro particles. Therefore, because the desirable properties of the crystalline product may change within wide limits, the main purpose is to control the crystal size distribution.

The CSD, determined by the rates of nucleation and growth, is influenced by the supersaturation, thus control of supersaturation becomes the basic element of crystallizer system control. Unfortunately, the workable degree of supersaturation is usually 0.5–1% so that it is so small that it is hard to measure directly. Besides, the on-line measurement of CSD, although it is also possible by laser diffraction, is restricted to slurries that contain solids at low concentration. As a consequence, the vast majority of crystallizers use indirect means of control.

An important exception to this rule is sugar, in the case of which the level of supersaturation can be greater than usual in industrial crystallizers, and correlations can be used to measure supersaturation directly. Estimations of sugar supersolubilities are based on these empirical correlations.

Degrees of Freedom

To define the maximum number of controllers, we consider the degrees of freedom for a crystallization system. The variables are

1. Temperature and flow of the process fluid
2. Temperature and flow of the cooling or heating medium
3. Level of supersaturation
4. Ratio of mother liquor to crystals, which can be changed by varying recycles of mother liquor to feed stream
5. Removal and dissolution rate of fine crystals

The first principles-based governing equations are

1. Energy balance equation
2. Mass balance equation for the crystallizing substance
3. Balance equation for fines

Because the system has seven variables and three equations, it has four degrees of freedom. Thus, the maximum number of automatic controllers permissible in a crystallizer system, without overdefining it, is four. The exceptions are the dilution and reaction crystallizers, where the flow of the diluent and ratio of reactants form additional degrees of freedom.

The amount of fines can be estimated by using density measurement. In crystallization operations, the density sensor (differential pressure or any other type), which gives an indication of the amount of crystals in the crystal slurry, appears to be an important instrument. This measurement is based on the fact that the density of the clear liquor is constant. Hence, the measurement of the differential pressure between two points in the vessel is a measure of the amount of crystals in the suspension, because any change in density is caused by a change in the amount of crystals in the crystal slurry.

EVAPORATIVE CRYSTALLIZERS

This type of crystallizer produces supersaturation and, hence, crystals, by loss of solvents induced by one of the three methods: (1) indirect heating, (2) submerged combustion, and (3) spray evaporation. The first two are the dominant types and will be discussed here.

Indirectly Heated Crystallizers

Circulating-magma crystallizers with indirect heating are by far the most important type of crystallizers in use today: The forced-circulation (FC) and the draft-tube baffle (DTB) designs belong to this class.

A typical FC crystallizer is shown in Figure 8.18c. In this design the feed is introduced into the recirculation loop. The critical design parameters in FC crystallizers are the internal recirculation rate and velocity, the crystallizer hold volume, and the speed of the circulating pump. At internal circulation less than the optimum, excessive flashing can occur at the boiling surface, causing a high level of supersaturation. If this supersaturation cannot be reduced by deposition of the solute because of the lack of adequate crystal surface area in the suspension, intensive nucleation will occur, producing a large amount of fines and causing build-up of solids on the walls of the crystallizer.

In general, recirculation rates aim at restricting the flashing at the crystallizer walls or at other boiling surfaces to approximately 1.7–4.5°C (3–8°F) whereas a magma-density range of 15–25% is typical, even though the exact optimum depends on the particular crystal system.

Various designs of crystallizers are available, but the approaches of controlling the processes are similar. A possible control system is shown in Figure 8.18c where four control loops are involved:

1. Feed clear liquor is fed on level control to a feed tank. Here it is mixed with the mother liquor from the centrifuge of the crystalline product.
2. The mixed liquor is fed to the suction side of the crystallizer recirculating pump. This feed is adjusted by a level controller. (For methods of protecting the level sensor from plugging and material build-up, refer to Chapter 3 of the first volume of this handbook.)
3. The steam flow to the heat exchanger is on flow control. Once the steam rate is fixed, the production rate is also fixed, provided that the feed composition does not change.
4. Temperature control in the vessel may be achieved by controlling the evaporator chamber pressure by an air bleed.

Refinements to this basic system are possible, as an interlock between the steam and circulating pump, or the addition of a density recorder. The magma density can also be used for control, rather than just for monitoring. In the control configuration shown in Figure 8.18d, the density controller

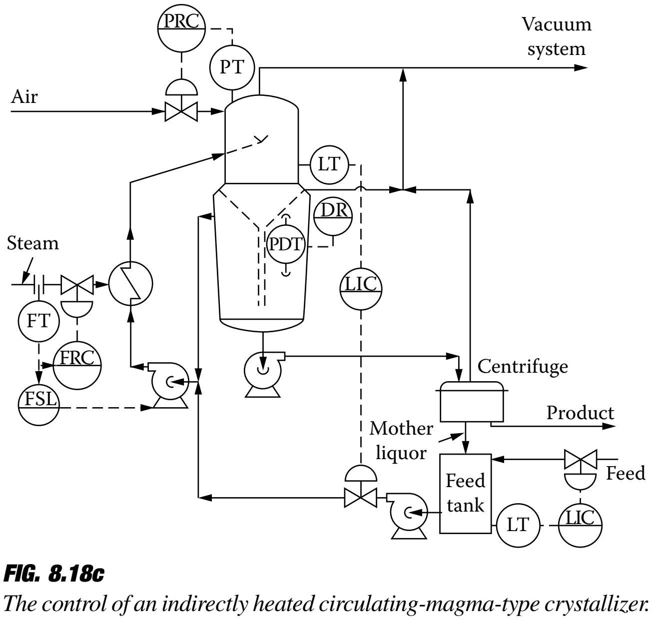


FIG. 8.18c
The control of an indirectly heated circulating-magma-type crystallizer.

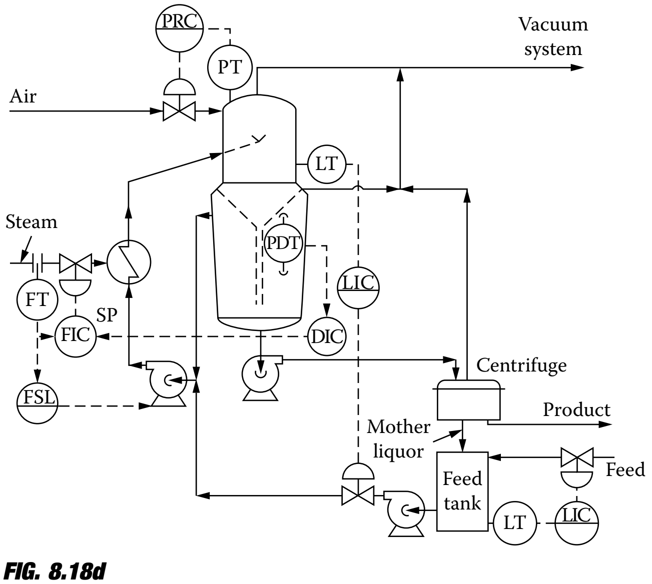


FIG. 8.18d
Cascade control of crystal concentration (measured by density) by throttling heat input.

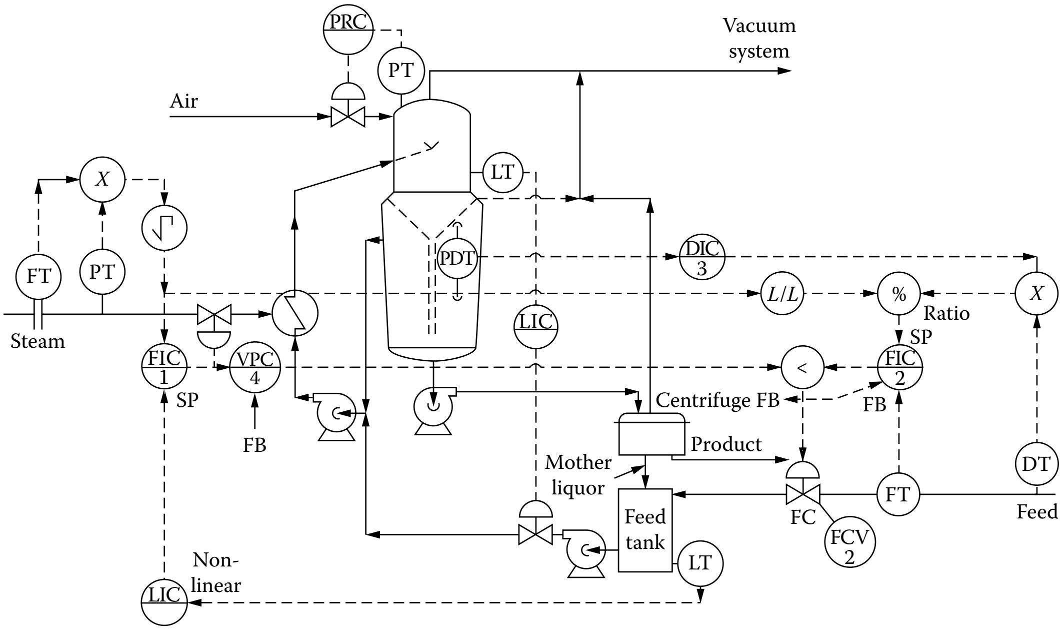


FIG. 8.18e
In this control system, feedforward, dynamic compensation and selective control of the feed are added to the basic controls of a circulating-magma crystallizer.

serves as the cascade master of the steam flow controller, and it varies the heat input in order to keep the crystal concentration constant.

Advanced Controls

Figure 8.18e illustrates how a number of additional features and flexibilities can be incorporated into the controls of a circulating-magma crystallizer. In this configuration the steam flow is measured and controlled on a mass basis (FIC-1) so that it might be directly related to the feed flow (FIC-2). The density of the feed stream is used to estimate the concentration of the solute. It can also be used to signal the need to increase the steam-to-feed ratio in a feedforward whenever the solvent content of the feed stream rises, i.e., when it becomes more dilute. The steam-to-feed ratio is trimmed by the feedback controller DIC-3 so as to guarantee consistent product concentration in the crystallizer.

In Figure 8.18e, in order to take full advantage of the surge capacity of the feed tank, its level is not held constant but is allowed to float up and down. This level variation slowly adjusts the steam flow rate (FIC-1), which in turn changes the feed flow set point, and thereby keeps the feed tank level within limits.

The lead-lag relay (L/L) provides dynamic compensation for the time constants of the process. The valve position controller (VPC-4) serves to guarantee that the crystallizer will not be starved for steam. Therefore, whenever the steam valve opening approaches 100%, the low-signal selector on the feed valve (FCV-2) blocks the control signal from FIC-2 and allows VPC-4 to reduce the feed flow as required to match the availability of the steam. Because VPC-4 and FIC-2 control the FCV-2 valve in a selective manner, they are both provided with external feedback (FB) to make sure that their integrals will not wind up when the other controller is manipulating the valve.

Draft-Tube Baffle Crystallizer

The draft-tube baffle produces larger crystals than the FC crystallizers under equivalent conditions. It consists of a closed vessel with an inner baffle forming a partitioned settling area, inside which a tapered vertical-draft tube surrounds the agitator, which enters from the top or the bottom. The agitator is of the axial-flow type and operates at low speeds. If additional classification of the crystals is desired, an elutriating leg can be fitted to the bottom cone. The draft tube is centered by support vanes to prevent body swirl and to minimize turbulence in the circulating magma. Supersaturation may be generated by either evaporation, cooling, or vacuum cooling.

The baffle controls the crystal size by permitting the separation of unwanted fine crystals. Figure 8.18f shows the control of the crystal size distribution in a draft-tube baffle-type crystallizer using fines removal and their dissolution.

Multiple-Effect Operation

Evaporative crystallizers are often used as multiple-effect systems, because such configurations improve the product size distribution because of the narrow residence time distributions. Different strategies can be employed to improve the product CSD. A successful strategy is, for instance, to permit nucleation in the first stage and only growth in the subsequent

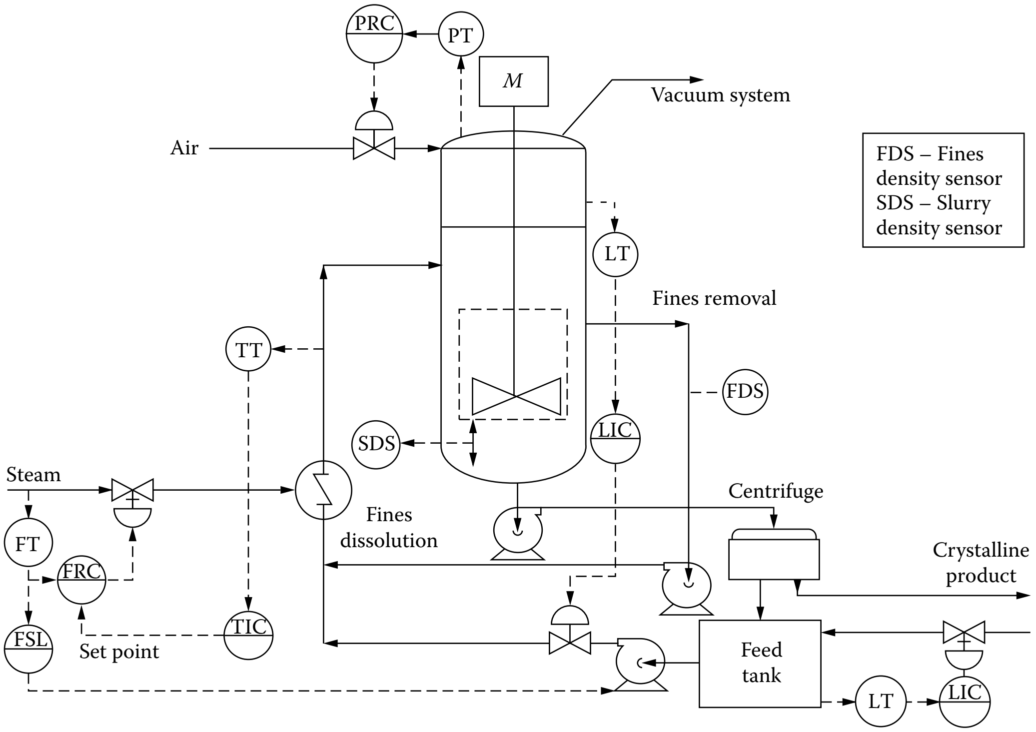


FIG. 8.18f
The crystal size in a draft-tube-type baffle crystallizer can be controlled by removing and redissolving the fines.

stages. A simplified control scheme on a triple-effect unit is shown in Figure 8.18g.

This complex system has three important features:

1. Level in each unit is an important process variable, because it determines the residence time. Level is usually controlled by throttling the makeup of the mixed liquor.
2. Steam flow to the first unit is usually on flow control.
3. Feed enters the feed tank on level control, where it is mixed with the mother liquor from the centrifuge of the crystalline product. Temperature control in the last unit is obtained by pressure control of the air bleed.

In these control systems, density recorders can also be used. If boiling point elevation is sufficiently large, the detected density can also be used to directly control the effluent liquor concentration, as was the case in Figure 8.18e.

Submerged-Combustion Crystallizers

Submerged-combustion crystallizers are used on corrosive applications or in cases where the salts have inverted solubility. One possible control configuration is shown in Figure 8.18h. Here, the clear liquor is fed under flow control, while the burner fuel gas can be under either flow or pressure control, in which case a bypass controls the flow of combustion air to the burner.

In developing the overall control system, both auxiliaries and safety interlocks must be included and designed to meet the requirements of the Fire Insurance Association (FIA).

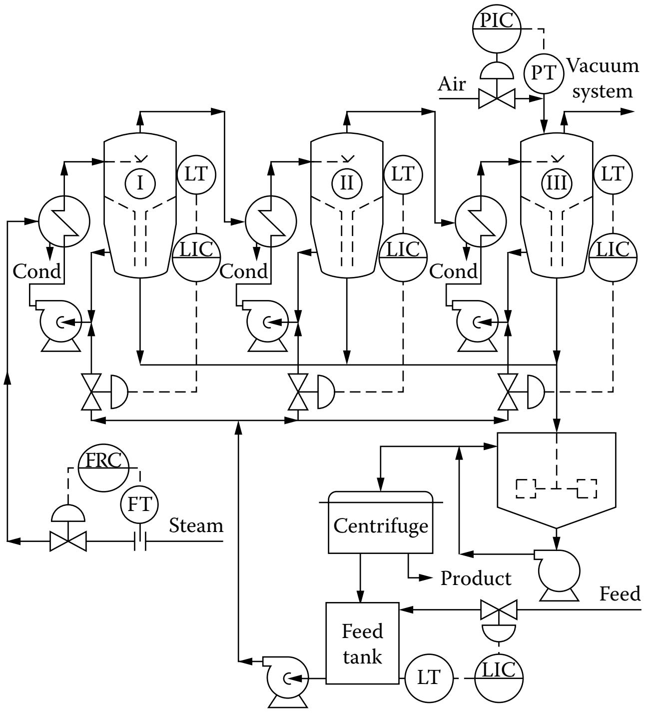


FIG. 8.18g
The basic controls of a three-effect crystallizer.

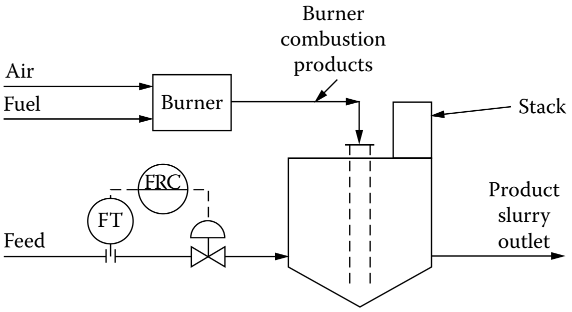


FIG. 8.18h
Submerged combustion-type evaporator crystallizer.

COOLING CRYSTALLIZERS

In principle, cooling crystallizers operate at atmospheric pressure, and their heat is transferred to a cooling medium or to air by either indirect or direct contact.

There are many types of cooling crystallizers, but the majority falls into three categories: (1) controlled-growth magma crystallizers, (2) classifying crystallizers, and (3) direct-contact crystallizers.

Controlled-Growth Magma Crystallizers

A variety of cradle crystallizers and scraped-surface units belong to this category. The cradle types are used in small applications and involve little instrumental control. The various scraped-surface crystallizers are used in crystallization from high-viscosity liquors or in open-tank crystallizers, as coolers to induce nucleation. Two designs of controlled-growth magma crystallizers and their controls are shown in Figures 8.18i and 8.18j. In Figure 8.18i an evaporating refrigerant cools the shell of the unit, while the crystallization process takes place in the tubes. Usually several scraped-surface tubes are used in series.

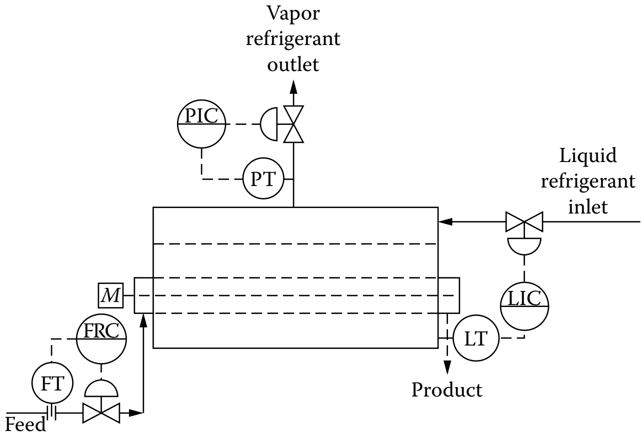


FIG. 8.18i
The basic controls of a refrigerant-cooled crystallizer.

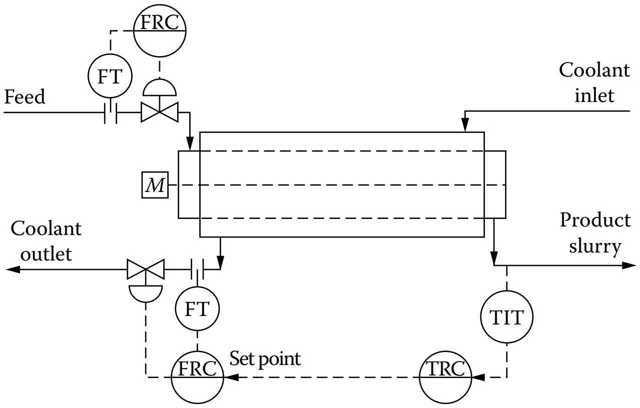


FIG. 8.18j
The controls of a cooling crystallizer that is cooled by a liquid coolant.

The control system operation is as follows:

1. Clear liquor is fed under flow control. Because the residence time in the tubes is large and because it is required in order to obtain the required crystal size, product quality-based feedback control from the product outlet is not practical. Yet, in some applications, feedback control with dead-time compensation has been applied.
2. In Figure 8.18i, liquid refrigerant enters under level control and leaves the system as a vapor under pressure control. In other applications, the refrigerant liquid is introduced under flow control provided by a metering pump. Both two- and three-mode controllers have been used to provide pressure control.

In other cooling crystallizer applications, liquid coolant is used as the heat transfer fluid. Figure 8.18j illustrates the control system of such a unit, where the cooling fluid flow is controlled in cascade by the outlet temperature of the product slurry.

Classifying Crystallizers

In the Oslo “Krystal” cooling crystallizer the supersaturation is generated entirely by cooling. In this design the feed is introduced into a circulating loop, which is also being cooled. The control configuration for a cooling-type classifying crystallizer is shown in Figure 8.18k.

The functions of the control loops are as follows:

1. Flow control of feed liquor is provided to maintain constant throughput.
2. Flow of coolant is controlled in cascade to maintain the outlet temperature of the process fluid at the outlet of the heat exchanger. One refinement is to interlock the coolant flow with the circulation pump motor.

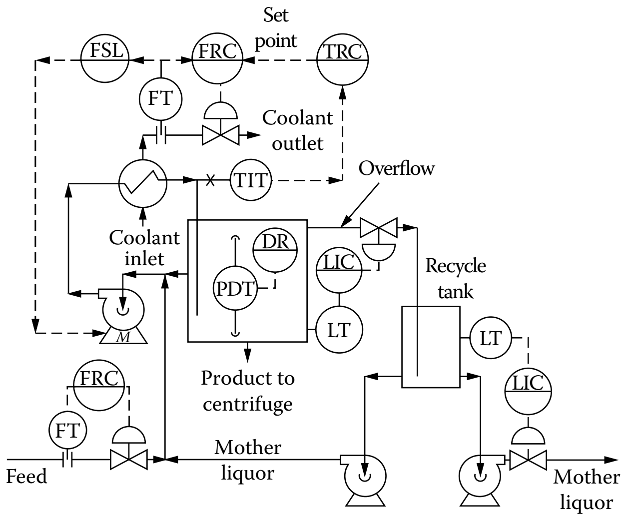


FIG. 8.18k
The controls of a classifying crystallizer with external circulation for cooling.

3. Overflow from the crystallizer is on level control. Differential pressure-type level measurement is acceptable for this application.
4. Mother liquor outflow from the recycle vessel is on level control.

Figure 8.18l illustrates a design that improves the crystal size distribution of the product by providing external classification in a cooling crystallizer. Here, the cooling occurs through a jacket under cascade control. The undersize fraction of crystals leaving the hydrocyclone with the mother liquor is returned to the crystallizer. Such operational configuration is very effective to control the size of the crystals, but it also has a tendency to cause steady-state oscillation. To minimize such oscillations requires more sophisticated control systems.

Direct-Contact Crystallizers

In such units, the coolant is an evaporating refrigerant or brine, and it is in direct contact with the process slurry. Basic control methods are similar to those in controlled-growth crystallizers (Figure 8.18i). These involve constant feed rate and flow controls of the evaporating refrigerant, based on process fluid outlet temperature or on tank level.

VACUUM CRYSTALLIZERS

In vacuum crystallizers, heat input to produce adiabatic evaporation comes entirely from the sensible heat of the feed liquor and from the heat of crystallization of the crystalline product. Thus, supersaturation is generated by a combination of cooling and concentrating the liquor. The forced-circulation and draft-tube baffle crystallizer designs are both used in this operation mode.

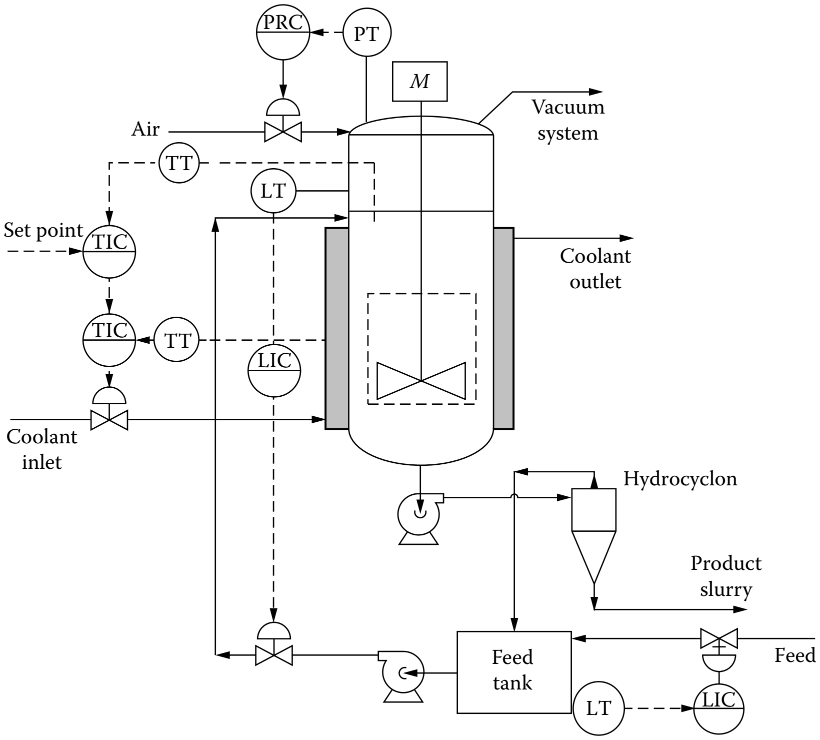


FIG. 8.18l
The controls of a jacketed cooling crystallizer with classified product removal.

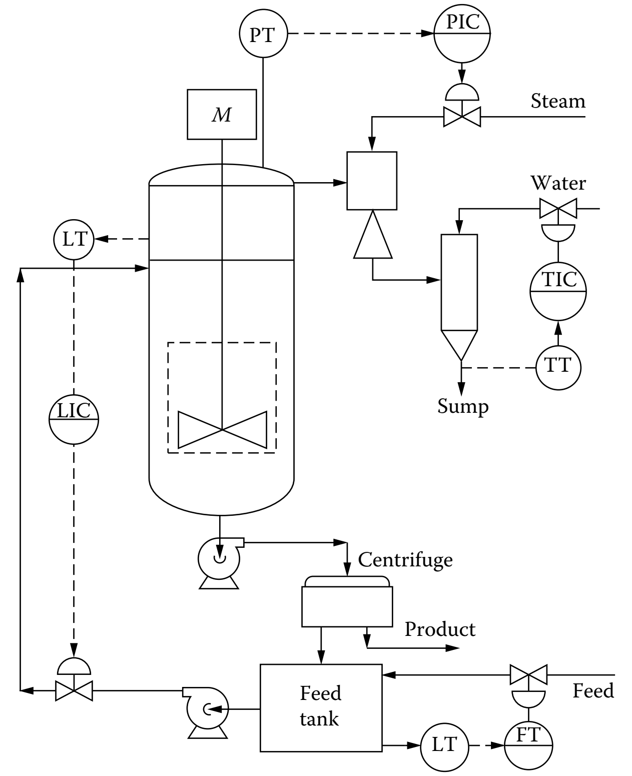


FIG. 8.18m
Draft-tube-type baffle crystallizer, provided with vacuum control and mother liquor recirculation.

The control strategy for a draft-tube baffle crystallizer with vacuum cooling is shown in Figure 8.18m. Here, the vacuum is on control with steam jet ejector. Air bleed is on pressure control, and both the steam and water supplies are on automatic control, which allows achieving the minimum cost of utilities.

Another control configuration is shown in Figure 8.18n. Here, the vacuum in the crystallizer is maintained by throttling the coolant outlet. This vacuum indirectly controls the process temperature.

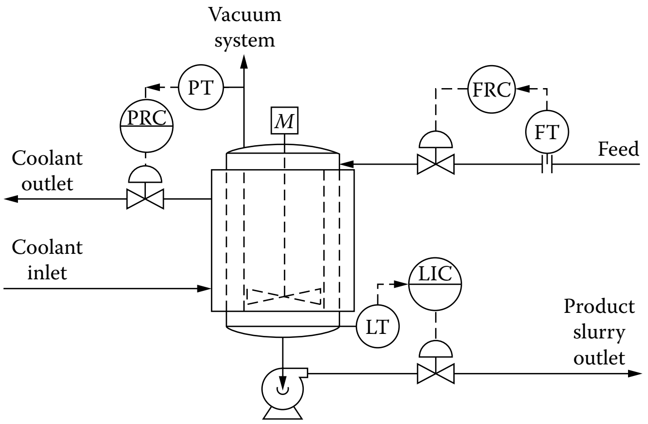


FIG. 8.18n
Draft-tube baffle crystallizer with combination of cooling and vacuum control.

The control system consists of (1) feed flow control, (2) product removal control based on level measurement, and (3) temperature control accomplished indirectly by manipulation of jacket coolant. Use of suspension density recorder and other refinements can also be used.

REACTION CRYSTALLIZERS

In reaction crystallizers, crystallization is associated with a chemical reaction that usually produces the solute directly. Chemical reactions between two components, however, may also produce some diluents, in turn causing “salting-out.” In a reaction crystallizer, once the reaction mixture is saturated with respect to the crystallizing substance, the reaction rate would determine the rate of supersaturation. In some cases, the heat of exothermic reaction may be used in evaporating the solvent, thereby producing additional supersaturation during the course crystallization.

A control configuration for a reaction crystallizer is shown in Figure 8.18o, where one of the reactants is fed into the crystallizer in the gaseous phase while the second reactant is fed in the form of aqueous solution. Here, the aqueous solution feed is on flow control, the total liquid feed is on level control, while the gaseous reactant is added on pH-cascaded control. The pH metering lines should be continuously flushed when this system is used.

In the case when both reactants A and B are liquids, they can be charged into the crystallizer under ratio control to maintain either the stoichiometric or any other predetermined ratio of these reactants. As shown in Figure 8.18p, the total

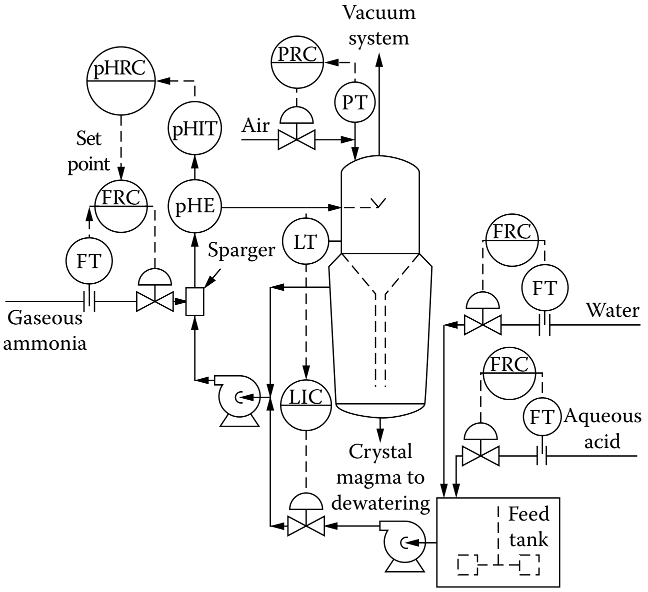


FIG. 8.18o
The controls of a reaction crystallizer that is utilizing both gaseous and liquid reactants.

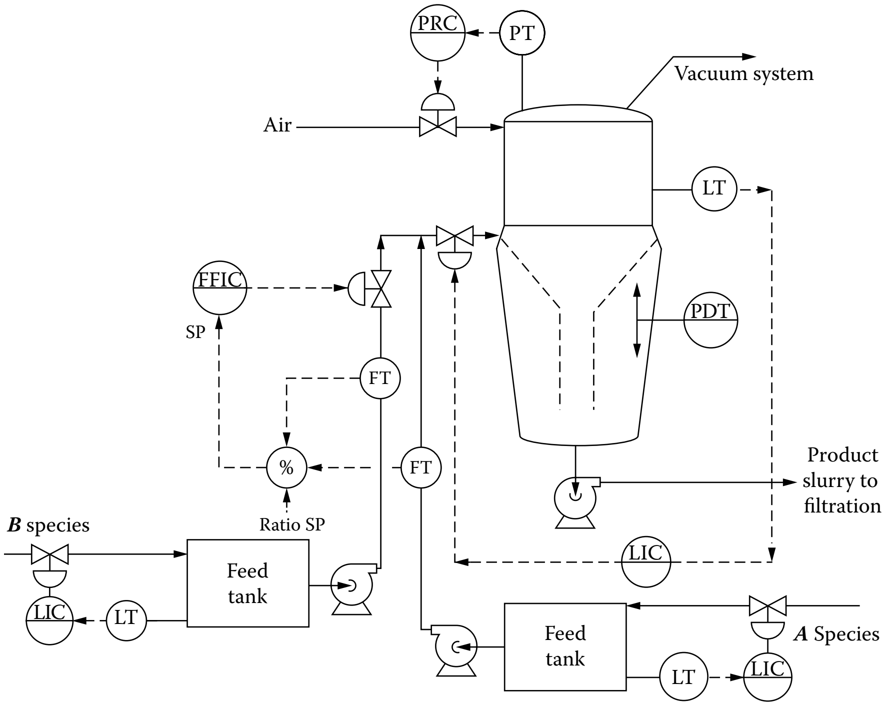


FIG. 8.18p
Reaction crystallizer controls provide ratio control of the reactants A and B.

flow into the reactor crystallizer can be on level control and the temperature can be controlled by bleeding in air to adjust the pressure.

AUXILIARY EQUIPMENT

The control requirements of crystallizers are also a function of the equipment associated with them. There are four subsystems whose control should be considered in particular:

1. The feed system, including the feed liquor, recycle, and wash streams.
2. Vacuum control to maintain predetermined pressure in the system as a common means of crystallizer temperature control.
3. Dewatering system control. This includes filters or centrifuges.
4. In the case of external classification equipment, in order to improve the crystal size distribution of the product, hydrocyclones, vibrating screens, and other external classification devices can be used.

CONCLUSIONS

In this section some of the common methods of crystallizer control have been outlined. In the crystallization process, because of the phase changes and because of the dispersed nature of the crystalline product, strong interactions exist between the process variables. Because of the complex, multiple-input/multiple-output nature of this process, the control system development should be based on a step-by-step analysis.

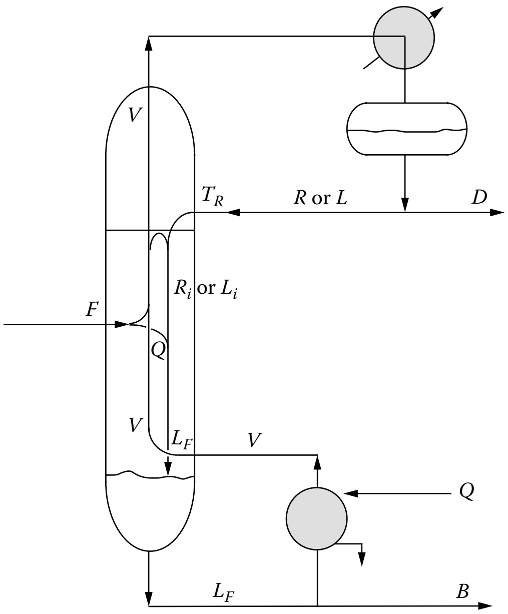
Pilot plant studies are often recommended, because small variations in the feed liquor compositions can have tremendous influence upon nucleation and crystal growth. Also, minor differences in operating conditions can produce crystals showing significantly different properties. The developer of a control system should take these factors into consideration.

Bibliography

Bamford, A. W., *Industrial Crystallization,* New York: Macmillan, 1966.
Mullin, J. W., *Crystallization,* London: Butterworth & Co., 1972.
Nyvlt, J., *Industrial Crystallization*: *The Present State of the Art,* Wienheim: Verlag Chemie, 1978.
Perry, R. H., and Chilton, C. H., *Chemical Engineering Handbook,* New York: McGraw-Hill.
Randolph, A. D., and Larson, M. A., *The Theory of Particulate Processes,* New York: Academic Press, 1988.
Tavare, N. S., *Industrial Crystallization. Process Simulation, Analysis, and Design.* New York: Plenum Press, 1995.

8.19 Distillation: Basic Controls

H. L. HOFFMAN, D. E. LUPFER (1970)
L. A. KANE (1985) **B. A. JENSEN** (1995)
B. A. JENSEN, B. G. LIPTÁK (2005)



Flow sheet symbol

INTRODUCTION

Distillation is the most common class of separation processes and one of the better understood unit operations. It is an energy-separating-agent equilibrium process that uses the difference in relative volatility, or differences in boiling points, of the components to be separated. It is the most widely used method of separation in the process industries. The distillation process will most often be the choice of separation unless the following conditions exist:

- Thermal damage can occur to the product.
- A separation factor is too close to unity.
- Extreme conditions of temperature or pressure are needed.
- Economic value of products is low relative to energy costs.

Control involves the manipulation of the material and energy balances in the distillation equipment to affect product composition and purity. Difficulties arise because of the multitude of potential variable interactions and disturbances that can exist in single-column fractionators and in the process that the column is a part of.

Even seemingly identical columns will exhibit great diversity of operation in the field. Therefore, this section will not attempt to provide control strategies that can be applied to columns in a "cookbook" fashion. Instead, discussion will begin with a basic description of the distillation process and equipment, followed by techniques used to derive a mathematical column model.

The presentation in this section will then describe methods to evaluate interactions and alternative control strategies; control models used for some product quality, pressure, and feed flow control strategies; and finally some common feedforward advanced regulatory control strategies commonly used in the regulation of fractionators.

The goal of this section is to provide the process control engineer with the tools necessary to design unique control strategies that will match the specific requirements of distillation columns.

General Considerations

Distillation separates a mixture by taking advantage of the difference in the composition of a liquid and that of the vapor formed from that liquid. In the processing industries, distillation is widely used to isolate and purify volatile materials. Thus, good process control of the distillation process is vital to maximize the production of satisfactory purity end products.

Although engineers often speak of controlling a distillation tower, many of the instruments actually are used to control the auxiliary equipment associated with the tower. For this reason, the equipment used in distillation will be discussed.

DISTILLATION EQUIPMENT

There are some basic variations to the distillation process. One such basic difference is between continuous and batch distillation. The main difference between these processes is that in continuous distillation the feed concentration is relatively constant, while in batch distillation it is rich in light components at the beginning and lean in light components at the end. While batch distillation is also described in this section, the emphasis is on the continuous processes.

Another basic difference is in the way the condenser heat is handled. The more common approach is to reject that heat into the cooling water and thereby waste it. This necessitates the use of "pay heat" at the reboiler, which usually is a large part of the total operating cost of the column. An alternate approach, also discussed in this section, is "vapor recompression" (Figure 8.19a), in which the heat taken out by the condenser is reused at the reboiler after a heat pump (compressor) elevates its temperature. While vapor recompression controls are also discussed in this section, the emphasis is on the traditional air- or water-cooled condenser designs.

The Column

The primary piece of distillation equipment is the main tower. Other terms for this piece of equipment are *column* and *fractionator*, and all three terms are used interchangeably. The tower, column, or fractionator has two purposes: First, it separates a feed into a vapor portion that ascends the column and a liquid portion that descends; second, it achieves intimate mixing between the two countercurrent flowing phases. The purpose of the mixing is to get an effective transfer of the more volatile components into the ascending vapor and corresponding transfer of the less volatile components into the descending liquid. The other equipment associated with the column is shown schematically in Figure 8.19b.

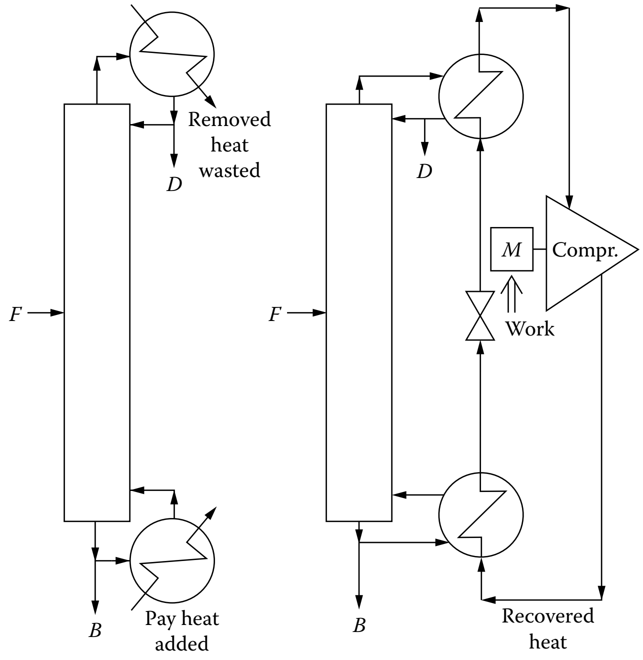


FIG. 8.19a
In contrast with conventional distillation, the vapor recompression system uses recovered heat.

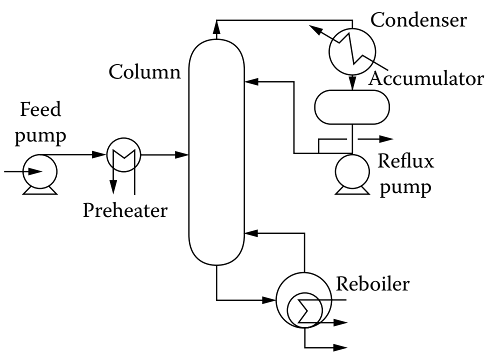


FIG. 8.19b
Distillation equipment.

In continuous distillation, the feed is introduced continuously into the side of the distillation column. If the feed is all liquid, the temperature at which it first starts to boil is called the *bubble point*. If the feed is all vapor, the temperature at which it first starts to condense is called the *dew point*. The feed entering the column is normally operated in a temperature range that is intermediate to the two extremes of dew point and bubble point. However, some optimization strategies may call for designs where the feed is either superheated or subcooled. For effective separation of the feed, it is important that both vapor and liquid phases exist throughout the column.

The separation of phases is accomplished by differences in vapor pressure, with the lighter vapor rising to the top of the column and the heavier liquid flowing to the bottom. The portion of the column above the feed is called the *rectifying* section and below the feed is called the *stripping* section.

Packing and Trays The intimate mixing is obtained by one or more of several methods. A simple method is to fill the column with lumps of an inert material, or *packing*, that will provide surface for the contacting of vapor and liquid. Another effective way is to use a number of horizontal plates, or *trays*, which cause the ascending vapor to be bubbled through the descending liquid (Figure 8.19c).

Tray designs are numerous and varied.[1] Tray designs include bubble cap plate unit, valve, sieve plate, tunnel, dual-flow, chimney, disc-and-donut, turbogrid trays, v-grid, Perform-Kontakt, Haselden baffle tray, Kittel trays, and other specialty-type units. Dualflo® trays, Flexitray®, Varioflex®, Bi-Frac®, Max-Frac®, NYE Trays®, Superfrac® trays, Super-Flux® trays, and Ultra-Frac® trays are specialty registered tray designs from different manufacturers that are variations of the aforementioned tray designs. Bubble caps

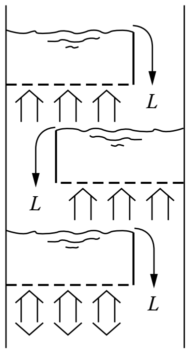


FIG. 8.19c
Intimate contact and therefore equilibrium is obtained as the vapor bubbles ascend through the liquid held up on each tray, as the liquid descends down the column.

and sieve trays are the most common designs used in distillation applications.

Many different types of packings are available.[2] They are normally classified as random or stacked. Random packings are those that are dumped into the containing shell. Raschig rings, Berl saddles, Intalox saddles, and Pall rings are the most common random packings and come in various sizes from $^1/_2$ to $3^1/_2$ in (1.25 to 9 cm).

Stacked packings, also known as grid or stacked packing, include large-sized Raschig rings and Lessing rings. Packings generally give lower pressure drops at the cost of higher installation costs. They are made of ceramic, plastic, or metal, depending upon the type of packing and the intended application. Other packings such as Maspac®, HyPak®, Tellerette®, IMTP® FLEXIPAC® KATAMAX®: FLEXIGRID®-2, -3, and -4, and KOCH-GLITSCH GRID® EF-25A are specialty registered packings from different manufacturers that are just variations of the aforementioned packings.

When deciding between the use of trays and packing, the following factors should be considered:[3]

- Because of liquid dispersion difficulties in packed towers, the design of plate towers is considerably more reliable and requires less safety factor when the ratio of liquid mass velocity to gas mass velocity is low.
- Towers using trays can be designed to handle wider ranges of liquid rates without flooding.
- Towers using trays are more accessible for cleaning.
- Towers using trays are preferred if interstage cooling or heating is needed because of lower installation costs of delivery piping.
- Towers using trays have a lower total dry weight, though total weight with liquid hold-up is probably equal.
- Towers using trays are preferred when large temperature changes are expected because of thermal expansion or when contraction may crush packing.
- Design information for towers using trays is generally more readily available and more reliable.
- Packed towers are cheaper and easier to construct than plate towers if highly corrosive fluid must be handled.
- Diameters of packed towers are generally designed to be less than 4 ft, while plate tower diameters are designed to be more than 2ft.
- Packed towers are preferred if the liquids have a large tendency to foam.
- The amount of liquid hold-up is considerably less in packed towers.
- The pressure drop through packed towers may be less than for plate towers performing the same service, making packed towers desirable for vacuum distillation.

Thus, generally, trays work better in applications requiring high flow, such as those encountered in high-pressure distillation columns, such as depropanizers, debutanizers, xylene purification columns, and the like. Packing works best at lower flow parameters, as the low-pressure drop of structured packing makes it very attractive for use in vacuum columns or ethylbenzene recycle columns of styrene plants.

The contacting between the vapor and liquid in a single-stage contacting device will not produce total equilibrium. The relationship between ideal and actual performance is the efficiency that translates the number of ideal separation stages into actual finite stages that must be used to accomplish the desired final separation. Efficiency varies, not only with the type of mixing method used (e.g., packing or trays), but also with fluid rates, fluid properties, column diameter, and operating pressure.

The influence of plate efficiency in the operation of the distillation tower becomes important in the control of the overhead composition. Because plate efficiencies increase with increased vapor velocities, the influence of the reflux-to-feed ratio on overhead composition becomes a nonlinear relationship.

Dynamics Dynamic considerations due to liquid hold-up on the trays comes into play when discussing distillation control. Because the liquid on each tray must overflow its weir and work its way down the column due to tray or packing hydraulics, this change will not be seen at the bottoms of the tower until some time has passed. The exact dynamics depend on column size, type of tray, number of trays, and tray spacing. The hold-up at each tray as shown in Figure 8.19c can be modeled by the LaPlace transform of the form

$$KG(s) = \frac{K}{(T_1 s + 1)} \tag{8.19(1)}$$

where

KG(s) = transfer function
K = system gain
T_1 = time constant
S = LaPlace transfer operator

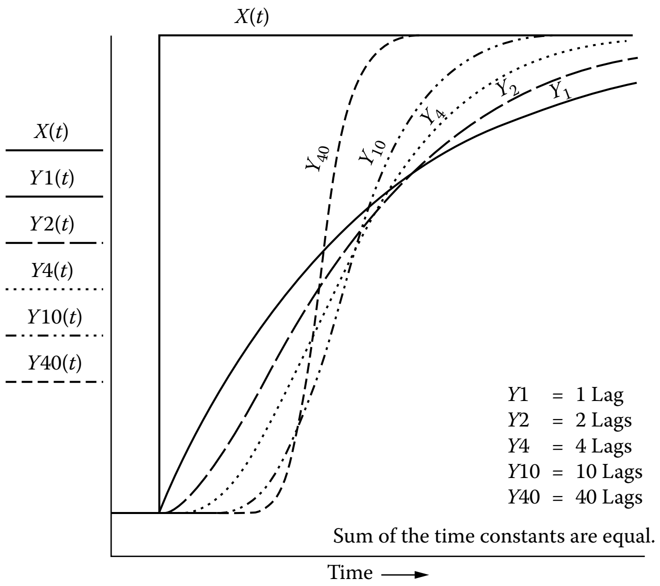


FIG. 8.19d
Response of nth-order lags to unit step change.

These lags are cumulative as the liquid passes each tray on its way down the column. Thus, a 30-tray column could be approximated by 30 first-order exponential lags in series of approximately the same time constant.

$$KG(s) = \frac{K}{(T_1 s + 1)^n} \quad \textbf{8.19(2)}$$

where
$n = 30$ for a 30-tray column

Figure 8.19d shows the response of nth order lags to a unit step change. The effect of increasing the number of lags in series is to increase the apparent dead time and increase the response curve slope. Thus, the liquid traffic within the distillation process is often approximated by using a second-order lag plus dead time as modeled by the LaPlace transform:

$$KG(s) = \frac{Ke^{-ts}}{(T_1 s + 1)(T_2 s + 1)} \quad \textbf{8.19(3)}$$

where
$e = e$ of log to the base e
$\phi =$ dead time
T_1, $T_2 =$ time constants

Condensers

The overhead vapor leaving the column is sent to a condenser and is collected as a liquid in a receiver, or accumulator. A part of the accumulated liquid is returned to the column as reflux. The remainder is withdrawn as overhead product or distillate. In many cases, complete condensation is not accomplished. In that case, the condensers are called partial condensers. In this instance, a vapor product is normally withdrawn as well as a liquid product.

A total condenser is usually designed for accumulator pressures up to 215 psia (1.48 MPa) at an operating temperature of 120°F (49°C).[4] A partial condenser is used from 215 psia to 365 psia (1.48 to 2.52 MPa), and a refrigerant coolant is used for the overhead condenser if the pressure is greater than 365 psia (2.52 MPa).

Common condensers include fin fans and water coolers. However, in order to improve efficiency of heat recovery, heat exchange with another process stream is often performed.

Propane is the most common refrigerant used. A pressure drop of 5 psia (34.4 KPa) across the condenser is often assumed if no measurements are available. The condenser and accumulator are the key pieces of equipment with respect to controlling pressure in the column.

Reboilers

The liquid leaving the bottom of the column is reheated in a reboiler. A reboiler is a special heat exchanger that provides the heat necessary for distillation. Part of the column bottoms liquid is vaporized and the vapors are injected back into the column as boil-up. The remaining liquid is withdrawn as a bottom product or as residue.

As shown in Figure 8.19e, reboilers come in widely varying designs. They can be internal, but most are external to the column. They can use natural or forced circulation.

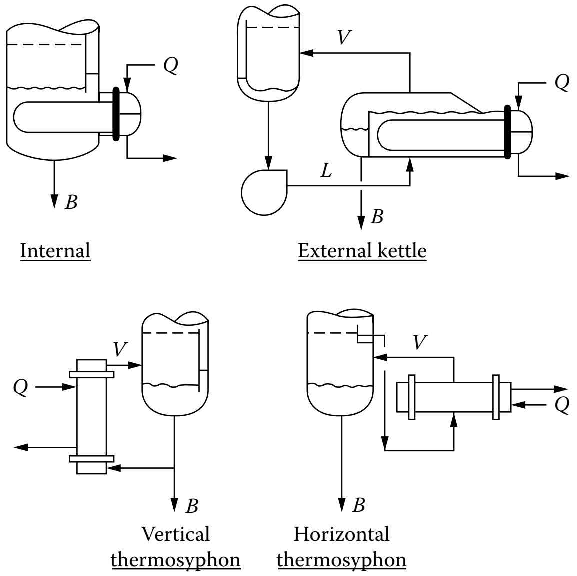


FIG. 8.19e
Reboiler design variations. External kettle reboilers often use forced circulation (pump), while the thermosyphon designs depend on natural circulation. The horizontal thermosyphon reboiler takes its liquid from the bottom tray, while the others take it from the column bottoms.

The kettle reboiler is the most common external forced circulation design.

Vertical and horizontal thermosiphon reboilers operate by natural circulation. In these, flow is induced by the hydrostatic pressure imbalance between the liquid inside the tower and the two-phase mixture in the reboiler tubes. In forced circulation reboilers, a pump is used to ensure circulation of the liquid past the heat transfer surface. Reboilers may be designed so that boiling occurs inside vertical tubes, inside horizontal tubes, or on the shell side.

A newer development in reboiler design is the concept self-cleaning shell-and-tube heat exchangers for applications where heat exchange surfaces are prone to fouling by the process fluid. Common heat sources include hot oil, steam, or fuel gas (fired reboilers). Cases where simple heat exchange with another process stream is used for efficiency of heat recovery are common. Thus, the choice of instrumentation to control heat addition to the tower depends upon the type of reboiler used.

Interheaters/Intercoolers

In some cases, additional vapor or liquid is withdrawn from the column at points above or below the point at which the feed enters. All or a portion of this sidestream can be used as intermediate product. Sometimes, economical column design dictates that the sidestream be cooled and returned to the column to furnish localized reflux. The equipment that does this is called a sidestream cooler, or intercooler. Multiproduct fractionators often have these intercoolers in a pump-around stream.

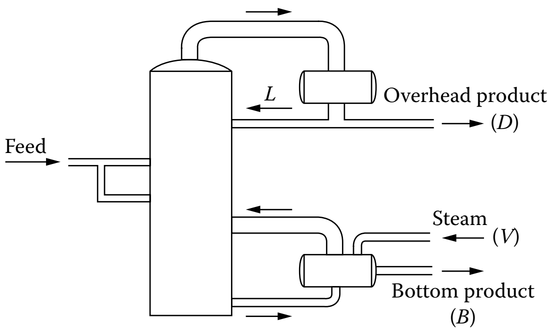
At other times, localized heat is required. Then, some of the liquid in the column is removed and passed through a sidestream reboiler, or interheater, before being returned to the column. Interheaters are usually utilized in cryogenic demethanizers.

Often the feed is preheated before entering the column. Common preheat mediums include the bottoms product or low-pressure steam. Preheating is often a convenient method to recover heat that would otherwise be wasted.

Column Variables

Controlling a fractionator requires the identifying of the controlled, manipulated, and load variables (Figure 8.19f). Controlled variables are those variables that must be maintained at a precise value to satisfy column objectives. These normally include product compositions, column temperatures, column pressure, and tower and accumulator levels.

Manipulated variables are those variables that can be changed in order to maintain the controlled variables at their desired values. Common examples include reflux flow, coolant flow, heating medium flow, and product flows. Load variables are those variables that provide disturbances to the column. Common examples include feed flow rate and feed composition. Other common disturbances are steam header pressure, feed enthalpy, environmental conditions (e.g., rain, barometric pressure, and ambient temperature), and coolant temperature.



Apparent variables:	Independent variables
C_1 = overhead temperature	2
C_2 = overhead pressure	
C_3 = overhead composition	
C_4 = overhead flow rate	1
u_1 = bottom temperature	2
u_2 = bottom pressure	
u_3 = bottom composition	
u_4 = bottom flow rate	1
u_5 = feed temperature	2
u_6 = feed pressure	
u_7 = feed composition	
u_8 = feed per cent vapor	1
u_9 = feed flow rate	1
m = steam flow rate (heat input)	1
	11

FIG. 8.19f
In a binary distillation process the number of independent variables is eleven (11) and the number of defining equations is two (2). Therefore, the number of degrees of freedom is nine (9), which is the maximum number of automatic controllers that can be used on such a process.

To handle these disturbances, column controls can be so designed as to make the column insensitive to these disturbances, or secondary controls can be designed to eliminate the disturbances. It is also important to evaluate the expected magnitude and duration of the likely disturbances, so that proper control system scaling and tuning can be achieved.

Feedforward controls are designed to compensate for these disturbance variables and are discussed later in this section. There are other advanced control or optimization methods that can be designed to compensate for these disturbance variables. They are discussed in Section 8.21.

Pairing of Variables The variables that should be controlled are usually obvious. They are normally identified when process objectives are defined and understood. Load variables are also easily identified. But identification of the manipulated variables can be more difficult. The general guidelines for identifying which manipulated variables to associate with which controlled variables are

- Manipulate the stream that has the greatest influence on the associated controlled variable.
- Manipulate the smaller stream if two streams have the same effect on the controlled variable.
- Manipulate the stream that has the most nearly linear correlation with the controlled variable.
- Manipulate the stream that is least sensitive to ambient conditions.
- Manipulate the stream least likely to cause interaction problems.

Unfortunately, the decision on pairing controlled and manipulated variables is complicated by the fact that the above rules may sometimes result in conflicting recommendations. Section 8.20 provides information on relative gain calculations, which can help to optimize the pairing of controlled and manipulated variables. Once the pairings are completed, the equations are then solved for the manipulated variables in terms of the controlled and load variables. In that form, the equations are the mathematical representations of the control systems.

MODELING AND CONTROL EQUATIONS

The primary application of instruments in distillation is to control the product purity, and secondarily, to minimize upsets to the unit caused by a change in process inputs. The instruments calculate the effects of the input changes and determine the corrective action needed to counteract them. The control actions are implemented by direct manipulation of the final control elements or by alteration of the set points of lower level controllers.

A careful analysis of limits and operating constraints is essential to the successful control of distillation columns. If the system is not designed to provide limit checks and overrides to handle operating limits, frequent operator intervention will be required during upsets. This is likely to result in a lack of confidence in the control system and will cause the operators to remove the column from automatic control more often than necessary, thereby not only reducing the effectiveness of the system, but also reducing safety.

The first step in the design of a good control system is the derivation of a process model. Knowing the defining equations, the manipulated variables can be selected, and the operating equations for the control system can be developed. The instrumentation is then selected for the correct solution of these equations.

The final control system can be relatively simple or can be a complex, interacting, multicomponent, computer-based system. In the discussion that follows, the procedures for designing distillation controls is followed by examples of the more common applications in distillation column control. A more detailed discussion of alternative strategies and advanced distillation column controls will be presented in Section 8.21.

Steady-State Model

The first step in the design of a control system must be the development of a process model. Frequently omitted in simple distillation columns, this step is essential to minimize the need for field reconfiguration of control strategies. Even with easily reconfigurable process automation systems (PASs), the development of the model is essential to fully understanding the process.

The model defines the process with equations developed from the material and energy balances of the unit. A common simplifying assumption is that all components of the feed have equal heats of vaporization, which leads to the assumption of equimolal overflow. Most shortcut fractionation calculations are based upon this underlying assumption.

The model is kept simple by the use of one basic rule: The degrees of freedom limit the number of controlled variables (product compositions) specified in the equations, as was illustrated in connection with Figure 8.19f. Some of the variables that can be manipulated to control a column are shown in Figure 8.19g.

Material Balance For example, for a given feed rate only one degree of freedom is available for material balance control. If overhead product (distillate) is a manipulated variable (controlled directly to maintain composition), then the bottom product cannot be independent but must be manipulated to close the overall material balance according to the following equations:

$$F = D + B \qquad \textbf{8.19(4)}$$

$$\text{Accumulation} = \text{Inflow} - \text{Outflow} \qquad \textbf{8.19(5)}$$

$$\text{Accumulation} = F - (D + B) \qquad \textbf{8.19(6)}$$

Because accumulation is zero at steady state, B is dependent upon F and D, as expressed by Equation 8.19(4):

$$B = F - D \qquad \textbf{8.19(7)}$$

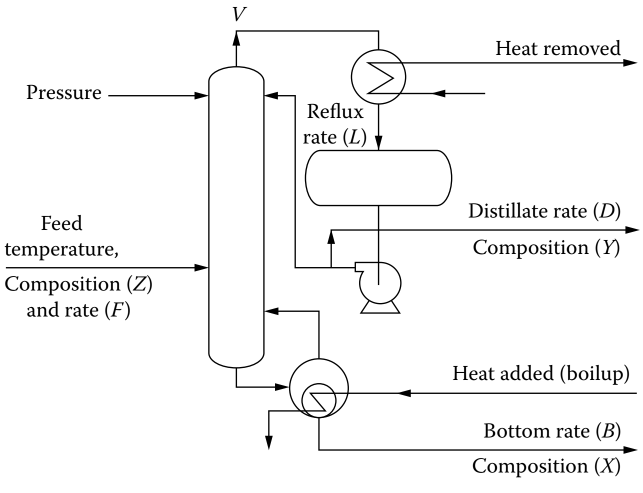


FIG. 8.19g
Variables that fix the distillation operation.

or if the bottoms product is the manipulated variable:

$$D = F - B \qquad \textbf{8.19(8)}$$

where:

F = feed rate (the inflow)
D = overhead rate (an outflow)
B = bottoms rate (an outflow)

If the compositions of the feed, distillate product, and bottoms product are known, then the component material balance can be solved:

$$100 = \%LLK_D + \%LK_D + \%HK_D \qquad \textbf{8.19(9)}$$

$$D \times \%LLK_D = F \times \%LLK_F \qquad \textbf{8.19(10)}$$

$$F \times \%LK_F = D \times \%LK_D + B \times \%LK_B \qquad \textbf{8.19(11)}$$

where:

$\%LLK_F$ = lighter than light key in the feed (mol%)
$\%LK_F$ = light key in the feed (mol%)
$\%LLK_D$ = lighter than light key in the distillate product (mol%)
$\%LK_D$ = light key in the distillate product (mol%)
$\%HK_D$ = heavy key in the distillate product (mol%)
$\%LK_B$ = light key in the bottoms product (mol%)

In the most general case, the feed might have four components, having the concentrations of LLK_F, LK_F, HK_F, and HHK_F. Three of these components appear in each of the bottom and overhead products. The separation of the column is fixed by specifying the heavy key component in the overhead product $\overline{HK}_D$ and the concentration of the light key component in the bottom product $\overline{LK}_B$.

Equations 8.19(9) to 8.19(11) assume no heavier than heavy key is found in the distillate and that no lighter than light key is found in the bottoms. Rearranging Equation 8.19(11) gives

$$\%LK_D = (F \bullet \%LK_F - B \bullet \%LK_B)/D \qquad \textbf{8.19(12)}$$

Substituting Equation 8.19(8) into Equations 8.19(10) and 8.19(12) gives

$$\%LLK_D = (F \bullet \%LLK_F)/(F - B) \qquad \textbf{8.19(13)}$$

$$\%LK_D = (F \bullet \%LK_F - B \bullet \%LK_B)/(F - B) \qquad \textbf{8.19(14)}$$

Substituting Equations 8.19(13) and 8.19(14) into Equation 8.19(9) to eliminate $\%LLK_D$ and $\%LK_D$:

$$B/F = \frac{(100 - \%HK_D - \%LLK_F - \%LK_F)}{(100 - \%HK_D - \%LK_B)} \qquad \textbf{8.19(15)}$$

For a given feed composition and desired product compositions, only one bottoms-to-feed ratio, B/F (product split), will satisfy the overall and component material balances. By fixing the bottoms flow, the distillate flow will be fixed. However, fixing a value of product split does not fix either the distillate or bottoms composition because many combinations of $\%LLK_F$, $\%LK_F$, $\%LK_B$, and $\%HK_D$ could yield the same value of B/F.

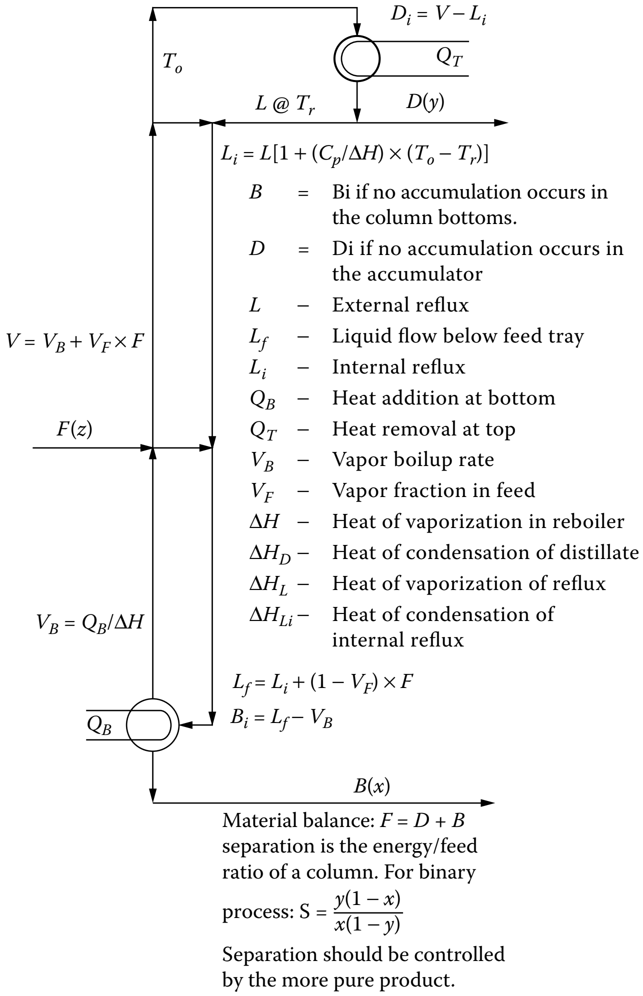


FIG. 8.19h
Energy balance equations can be used to describe the steady-state heat flow model of a distillation column.

Energy Balance The energy balance and the separation obtained are closely related. Conceptually, product composition control can be thought of as a problem of the rate of heat addition Q_B at the bottom of the fractionator and the rate of heat removal Q_T at the top of the column. A series of energy balances produces additional equations. Figure 8.19h shows a steady-state internal model of these equations.[5]

The vapor boil-up rate V_B equals the heat Q_B added by the reboiler divided by the heat of vaporization (ΔH) of the bottoms product:

$$V_B = Q_B/\Delta H \qquad \textbf{8.19(16)}$$

The vapor rate V above the feed tray equals the vapor boil-up rate plus the vapor entering with the feed (feed rate

F times vapor fraction V_F, provided the feed is neither subcooled nor superheated):

$$V = V_B + F \times V_F \qquad \textbf{8.19(17)}$$

The internal reflux rate, that is, the liquid at the top tray of the column is derived by a heat balance around the top of the tower. Assuming a steady-state heat balance where the heat into the tower equals the heat out:

$$\begin{aligned} &D \times (\Delta H_D + C_{p_D} \times T_t) + L_I \times (\Delta H_{L_I} + C_{p_{R_I}} \times T_t) \\ &\quad + L \times (C_{p_L} \times T_r) = D \times (\Delta H_D + C_{p_D} \times T_o) \\ &\quad + L \times (\Delta H_L + C_{p_L} \times T_o) + L_I \times (C_{p_{L_I}} \times T_t) \end{aligned} \qquad \textbf{8.19(18)}$$

where

C_p = specific heat

T_o = overhead vapor temperature (vapor at its dew point)

L = external reflux

T_r = external reflux temperature

L_I = internal reflux

T_t = top tray temperature (liquid at its bubble point)

Equation 8.19(18) reduces to:

$$\begin{aligned} &D \times C_{p_D} \times (T_t - T_o) + L_I \times \Delta H_{L_I} - L \times \Delta H_L \\ &\quad + L \times C_{p_L} \times (T_r - T_o) = 0 \end{aligned} \qquad \textbf{8.19(19)}$$

Making a simplifying assumption that the tray temperature equals overhead vapor temperature (i.e., the dew point of the vapor equals the bubble point of the liquid; $T_t = T_o$) produces:

$$L_I \times \Delta H_{L_I} = L \times \Delta H_L + L \times C_{p_L} \times (T_o - T_r) \qquad \textbf{8.19(20)}$$

or

$$\frac{L_i}{L} = \frac{\Delta H_L}{\Delta H_{L_i}} \cdot \left[\left(1.0 + \frac{C_{p_L}}{\Delta H_L} \cdot (T_o - T_r) \right) \right] \qquad \textbf{8.19(21)}$$

resulting in the equation

$$L_{pI}/L = K_2 \times [1 + K_1 \times (T_{po} - T_{pr})] \qquad \textbf{8.19(22)}$$

If a total condenser is employed, the composition of the internal reflux and external reflux are the same, i.e., $\Delta H_{L_I} = \Delta H_L$, so the constant $K_2 = 1.0$. Thus,

$$L = \frac{L_I}{[1 + K_1(T_O - T_r)]} \qquad \textbf{8.19(23)}$$

or

$$z_I = L \times [1 + K_1 \times \Delta T] \qquad \textbf{8.19(24)}$$

Note: This equation is valid for whatever units are used for C_{p_L} or ΔH_L. Because specific heat and heat of vaporization are nearly always in mass units, care must be taken to account for density differences whenever volume units are being used by the control equation. Also, C_{p_L} and ΔH_L should be calculated near the existing pressure and temperature of the external reflux.

The liquid rate, L_F, below the feed tray equals the internal reflux plus the liquid in the feed:

$$L_F = L_I + (1 - V_F) \times F \qquad \textbf{8.19(25)}$$

The distillate rate, D, equals the vapor rate, V, above the feed tray minus the internal reflux:

$$D = V - L_I \qquad \textbf{8.19(26)}$$

The bottoms rate, B, equals the liquid rate, L, minus the boil-up, V_B:

$$B = L - V_B \qquad \textbf{8.19(27)}$$

The criterion for separation is the ratio of reflux (L) to distillate (D) flows vs. the ratio of boil-up (V) to bottoms (B) flow rates. Manipulating reflux affects separation equally as well as manipulating boil-up, albeit in opposite directions. Consequently, only one degree of freedom exists to control separation. Thus, for a two-product tower, two equations define the process. One is an equation describing separation, and the other is an equation for material balance.

Dynamic Model

Because the tower doesn't always operate at steady state, it is essential to also account for the dynamics of the process. This necessitates extending the steady-state internal flow model and requires additional considerations. Figure 8.19i shows the internal flow model that includes dynamics.[6]

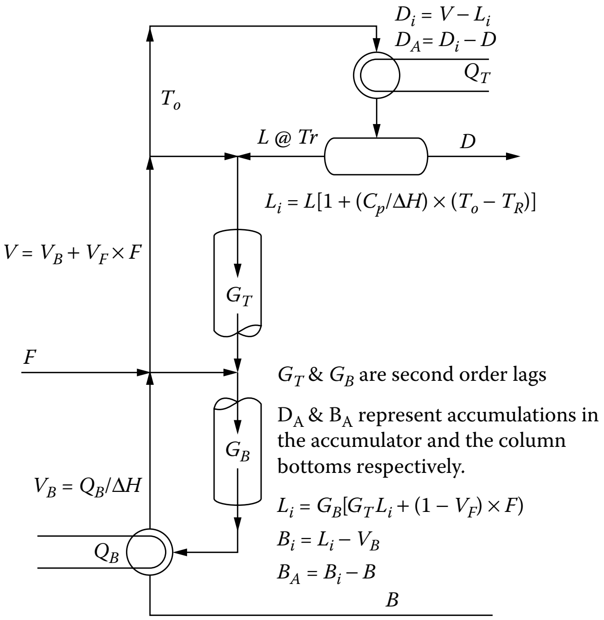


FIG. 8.19i
Dynamic internal flow model.

Because a change in the reflux rate must work its way down the column due to tray or packing hydraulics, this change will not be seen at the reboiler until some time has passed. The holdup at each tray has previously been modeled by the LaPlace transform of Equation 8.19(1). This Laplace transform can be converted to a simple first-order exponential lag equation of the form, which describes the response to a step change in input:

$$L_{\text{lag}} = L\,(1 - e^{-t}) \qquad \textbf{8.19(28)}$$

where

L is the liquid incoming to the tray

L_{lag} is the liquid leaving the tray

t is the time constant

These lags are cumulative as the liquid passes each tray on its way down the column. However, implementation of multiple first-order lags is impractical. Fortunately, it can be shown that multiple lags in series can be approximated by a dead time and a second-order exponential lag as shown by the LaPlace transform of Equation 8.19(3). For this reason, two dynamic terms (G_T and G_B) are included in Figure 8.19i. Equation 8.19(25) is then rewritten as

$$L = G_B[G_T L_I + (1 - V_F) \times F] \qquad \textbf{8.19(29)}$$

where

$G_B = \phi_1\,(1 - e^{-t1})\,(1 - e^{-t2})$

$G_T = \phi_2\,(1 - e^{-t3})\,(1 - e^{-t4})$

ϕ_1 and ϕ_2 are the dead times

G_B and G_T are the solution to the LaPlace transform of Equation 8.19(3).

Changes in boil-up rates are observed at the condenser in a matter of seconds. Normally, no dynamic terms are necessary for vapor streams, as the value of use of computing resources to that of the benefits by compensating for the dynamics is negligible.

The liquid inventory in the condenser or associated accumulator will change during unsteady-state actions. In the unsteady state, the difference $D_I - D$ is the rate of accumulation of material in the accumulator. Similarly for the liquid inventory at the bottom of the tower (the kettle), the difference $B_I - B$ is the rate of accumulation:

$$D_A = D_{II} - D \qquad \textbf{8.19(30)}$$

$$B_A = B_I - B \qquad \textbf{8.19(31)}$$

where

D_A is the accumulation in the overhead accumulator

B_A is the accumulation in the tower bottoms

Separation Equations

The control of product compositions for a fractionator is primarily a matter of control of the internal flows. In considering product separation, the degree of separation and the orientation of separation are important. The degree of separation is

$$\text{Degree of Separation} = \ln_e \frac{(\%LK_D \times \%HK_B)}{(\%HK_D \times \%LK_B)} \qquad \textbf{8.19(32)}$$

while the orientation of separation for a given degree of separation is defined as

$$\text{Orientation of Separation} = \frac{\%HK_D}{\%LK_B} \qquad \textbf{8.19(33)}$$

The relationship between x (the light key component) and the energy balance was developed by Shinskey[7] as a function of separation S:

$$S = \frac{y(1-x)}{x(1-y)} \qquad \textbf{8.19(34)}$$

where

x = mole fraction of the key light component the distillate ($\%LK_D$)

y = mole fraction of the key light component in the bottoms, (LK_B)

The relationship between separation (S) and the ratio of boil-up to feed (V/F) over a reasonable operating range is

$$V/F = a + bS \qquad \textbf{8.19(35)}$$

where a and b are functions of the relative volatility, the number of trays, the feed composition, and the minimum V/F. The control system therefore computes V based on the equation:

$$V = F\left[a + b\left(\frac{y(1-x)}{x(1-y)}\right)\right] \qquad \textbf{8.19(36)}$$

Because y is held constant, the bottom composition controller adjusts the value of the parenthetical expression if an error should appear in x. Let $V/F = y(1 - x)/(1 - y)$, and the control equation becomes:

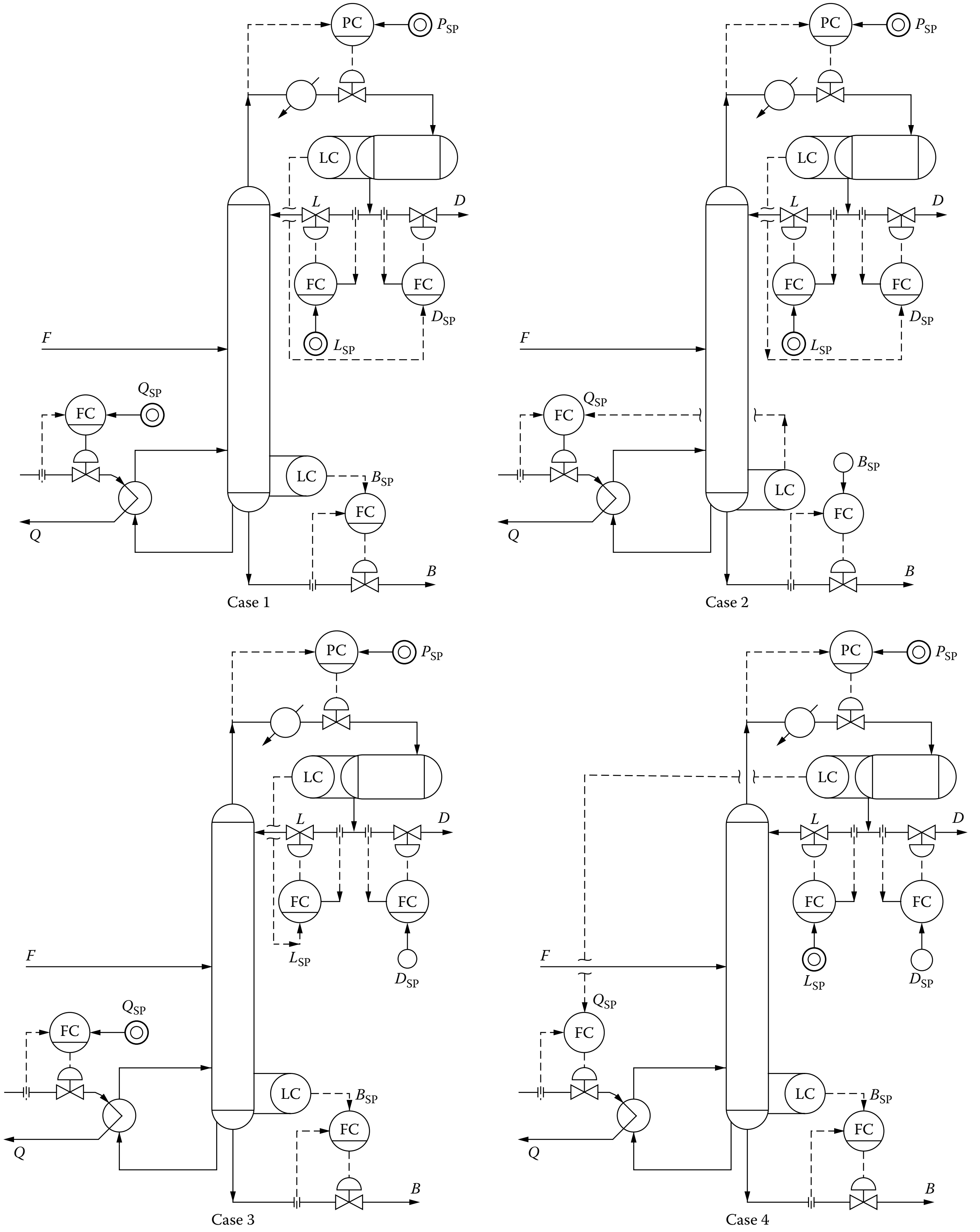
$$V = F(a + b[V/F])/x \qquad \textbf{8.19(37)}$$

where $[V/F]$ = the desired ratio of boil-up to feed.

Figure 8.19j illustrates four of the most common basic controls for the flows and levels of a two-product fractionator, where it is assumed that feed flow and tower pressure are held constant. A different set of the above control equations for controlling internal product flow rates will apply, depending upon the configuration of instrumentation used.

Scaling

The form of the control system equations influences the computing functions required. Boolean operands, such as high and low selectors, and dynamic functions, such as dead times, lead, and lag function, are also used. Most process automation systems have these basic computing function blocks. Implementation in a distributed control system (DCS), programmable



Case 1

Case 2

Case 3

Case 4

FIG. 8.19j
Four cases of conventional distillation control configurations.

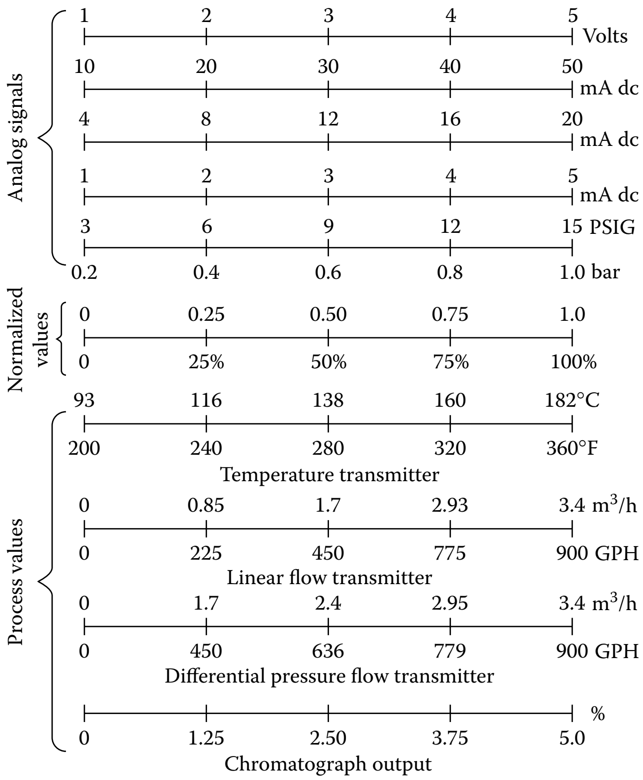


FIG. 8.19k

Common analog signals and their relationship to process variables.

logic controller (PLC), or multivariable digital controllers is vendor-specific.

The terms of the equations are sometimes scaled because most analog instruments and some PAS systems act on normalized numbers (0–100%) rather than on actual process values. With digital instrumentation and today's process automation systems, those occurrences are rare. The calculations become easier for those systems operating in engineering units.

Analog, and many digital, transmitters also operate on normalized values of the process variables. That is, the measurement signal will vary from 0 to 100% as the process variable shifts from 0 to its maximum value. Figure 8.19k illustrates the relationship among the various forms of analog signals and some typical process measurements.

The actual value of a process measurement is found by multiplying the analog signal by the calibrated full-scale value (*meter factor*) of the process variable. In the examples of Figure 8.19k, the temperature, represented by a 75% analog signal, is 320°F (160°C), the linear flow is 775 gph (2.93 m^3/h), the output of the differential pressure transmitter (flow squared) is 779 gph (2.95 m^3/h), and the composition is 3.75%.

Example As an example, let us review a flow ratio system in which the load stream, L, has the range of 0 to 1000 gpm (0 to 3.79 m^3/h); the manipulated stream, M, has a range of 0 to 700 gpm (0 to 2.65 m^3/h); and the ratio range, R, is 0 to 0.8 ($R = M/L$).

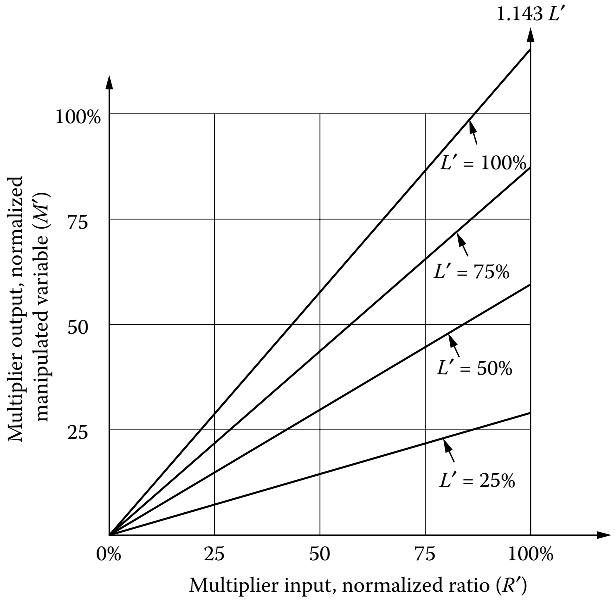


FIG. 8.19l

Multiplier output for the solution of Equation 8.19 (39).

$$700M' = (1000L')(0.80R') \qquad \mathbf{8.19(38)}$$

Reducing to the lowest form,

$$M' = 1.143(L')(R') \qquad \mathbf{8.19(39)}$$

The number 1.143 is the scaling factor. M' is plotted as a function of L' and R' in Figure 8.19l.

In applications such as the constant separation system, exact scaling is not critical. Exact scaling is when scaling constants must be used as calculated from instrument spans. The alternative is flexible scaling, where exact ranges are not needed but some arbitrary range is used to allow internal calculations to remain within range.

The flexible scaling cannot be used (1) when compensation for feed composition is part of the model, (2) when narrow spans must be used for reasons of stability, and (3) when transmitter calibrations are inconsistent with material balance ratios. Exact scaling techniques must be used for these cases.

MULTIPLE COMPONENT DISTILLATION

With binary mixtures, only two products are removed in the distillation column. However, most separations involve multiple components. Even then, most distillations remove only two liquid products. In other applications a vapor product is removed, or multiple liquid products are drawn from the tower. Sometimes only one product is withdrawn at a time.

Columns with Sidedraw

Having a sidestream product in addition to the overhead and bottom products adds a degree of freedom to a control system. The source of this extra degree of freedom can be seen from the overall material balance equation:

$$F = D + C + B \tag{8.19(40)}$$

where C is the sidestream flow rate. Two of the product streams can be manipulated for control purposes, and the material balance can still be closed by the third product stream.

The presence of this added degree of freedom makes the careful analysis of the process even more essential to avoid mismatching of the manipulated and controlled variables. As in the case of the previously discussed columns, the development of a control system for sidedraw applications also involves developing the process model and determining the relationship among the several controlled and manipulated variables.

In this case, for a constant feed rate and column pressure, five degrees of freedom exist: three composition specifications and two levels that can manipulate three product flows, and two heat balances (V and L). Several possible combinations of variables are available and should be explored.

The possible combinations of manipulated variables for the column in which the bottom composition and the sidestream composition must be controlled are

Distillate and sidestream flows
Distillate and bottom flows
Distillate flow and heat input
Sidestream and bottom flows
Sidestream flow and heat input
Bottom flow and heat input

Similarly, the possible combinations of manipulated variables for the column in which the distillate composition and the sidestream composition must be controlled are:

Distillate and sidestream flows
Distillate and bottom flows
Distillate flow and heat input
Sidestream and bottom flows
Sidestream flow and reflux
Bottom flow and reflux

The equations are

$$D = F\left[\frac{z_1 - c_1}{y_1 - c_1}\right] \tag{8.19(41)}$$

$$C = F\left[\frac{z_2 - x_2}{c_2 - x_2}\right] \tag{8.19(42)}$$

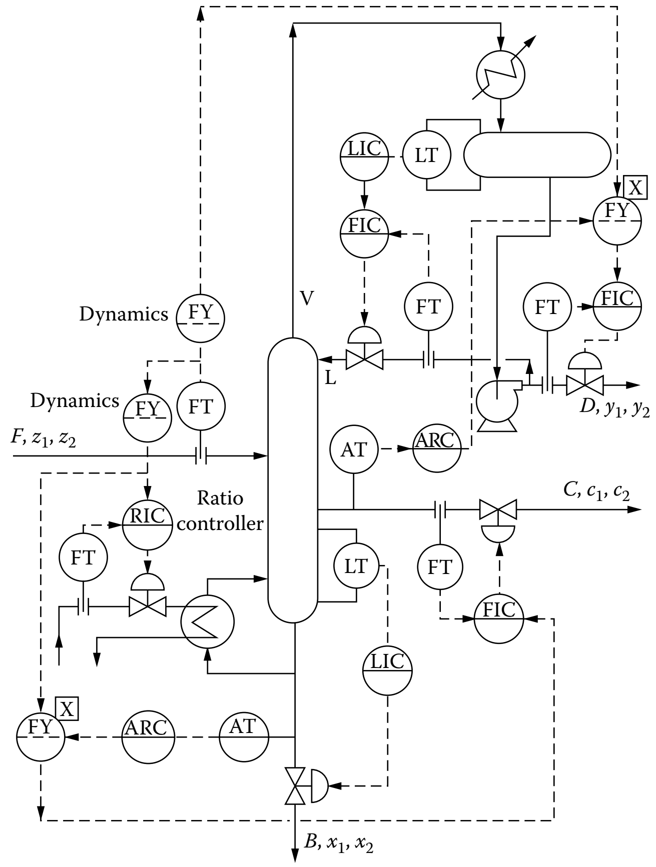


FIG. 8.19m
Control of composition in two product streams with a sidedraw.

The symbols z_1, y_1, and c_1 refer to the concentrations in the feed, distillate, and sidestream of the component under control in the sidestream. The concentrations of the key component in the bottom are respectively expressed by z_2, x_2, and c_2 for the feed, the bottoms, and the sidestream.

The resulting control system is shown in Figure 8.19m. Note that in this configuration the ratio of heat input to feed (and, therefore, boil-up to feed) is held constant. Separate dynamic elements are used for the distillate loop and for the heat input and sidestream loops.

Multiproduct Fractionators

Multiproduct fractionators are most common in the refining industry where multicomponent streams are separated into many fractions. Examples of multiproduct fractionators are crude towers, vacuum towers, and fluidized catalytic cracking unit (FCCU) main fractionators.

Product quality controls are used to adjust local column temperatures and sidedraw flow rates to control distillate properties related to the product specifications. An example is true boiling point (TBP) cut points. TBP cut points approximate the composition of a hydrocarbon mixture and are numerically similar to the American Society for Testing and

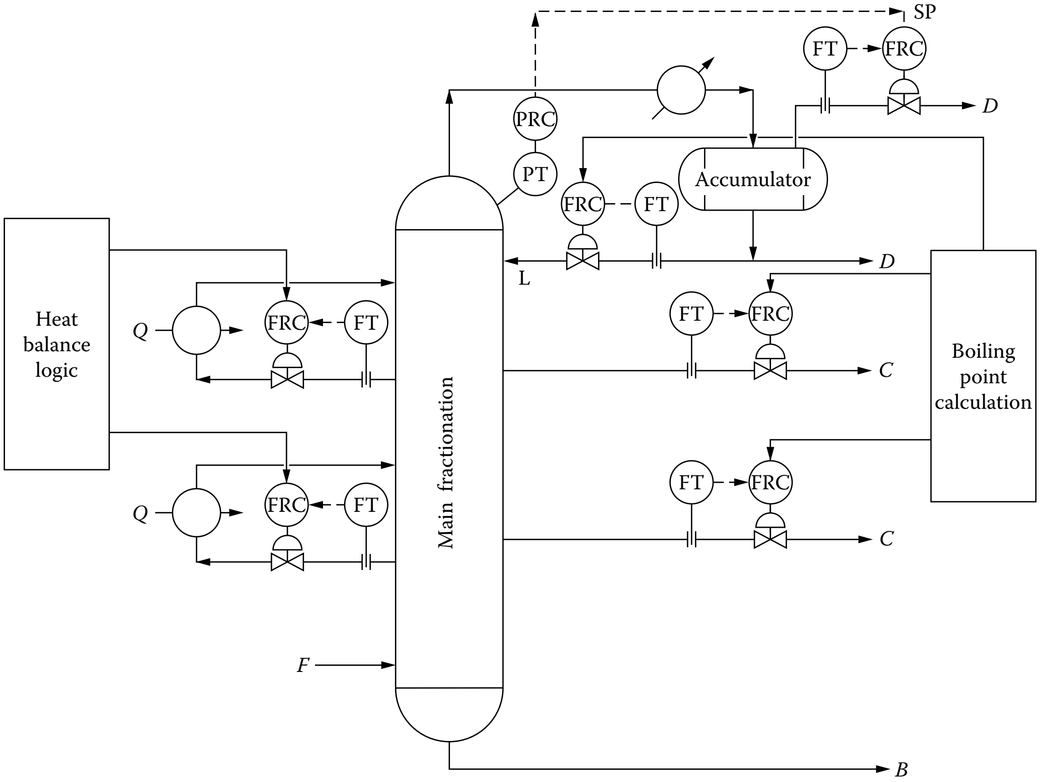


FIG. 8.19n
Control of product flows and pump-around refluxes.

Materials' (ASTM's) 95%. The ASTM laboratory distillate evaluation method is the standard used in the petroleum refining industry for determining the value (composition) of the distillation products.

A computer is required to calculate the product boiling point specification, such as 95% boiling point or TBP cut point on the basis of local temperature, pressure, steam flow, and reflux data. Local reflux is derived from internal liquid and vapor flows, as discussed previously, and the remaining variables are measured.

Boiling point analyzers can be used to provide the measurement signals. If there is no analyzer, the calculated boiling points can be used by themselves, or if there is one, they can be used as a fast inner loop with analyzer trim. Because of the volume of liquid/vapor loads within most multiproduct fractionators, the manipulated variables that provide the greatest sensitivity and the quickest response are generally the product flows.

Adjustment of reflux flows, as shown in Figure 8.19n, is an example of a heat balance control. The goal is to maximize heat exchange to feed, subject to certain limits[8] (limits and constraints are discussed as part of the subject of the optimization of distillation towers in Section 8.21). The task of maximizing the heating of the feed often simplifies to recovering heat at the highest possible temperature, which means recovering it as low as possible in the column.

Superfractionators

The term *superfractionator* is applied to towers that are physically large. These distillation units separate streams having their light and heavy key relative volatilities quite close to each other. Included in this classification are deisobutanizers, which separate isobutane from normal butane; propylene splitters, which separate propane from propylene; ethylbenzene towers, which separate ethylbenzene from xylene; and xylene splitters, which separate para- and ortho-xylene from meta-xylene.

Sometimes, the number of trays and subsequent height make it necessary to physically divide these towers into two or even three sections. Superfractionators have tremendous internal vapor-liquid rates in order to achieve the separation. Reflux-to-distillate ratios are very high, as are vapor-to-bottoms ratios.

A large pressure drop through the tower also exists. Long dead times and lag times are experienced before any response is seen to feed rate or reflux changes. Generally, distillate compositions of superfractionators have to be controlled with material balance equations due to the lack of sensitivity of response.

Batch Distillation

In batch distillation (see Figure 8.19o), an initial charge of liquid is fed to a vessel, and the distillation process is initiated

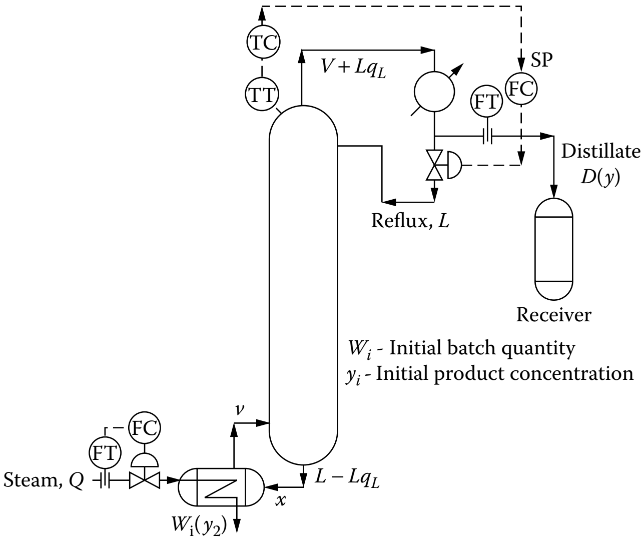


FIG. 8.19o
Batch distillation.

by turning on the heating and cooling systems. During the distillation process, the initial charge in the vessel continually depletes while building up the overhead product in the distillate receiver.

Batch distillations are more common in smaller, multiproduct plants where the various products can only be manufactured at different times, and where a number of different mixtures may be handled in the same equipment. Equation 8.19(43) is the basic equation that describes this operation:

$$W = W_i - Dt \qquad \textbf{8.19(43)}$$

where

W = amount remaining in the bottoms
W_i = the initial charge
D = distillate rate
t = time period of operation

The basic objective of the control system of this type of separation is to keep the composition of the distillate constant. Other goals include keeping the distillate flow constant or maximizing the total distillate production. The main goal of a batch distillation is to produce a product of specified composition at minimum cost. This often means that operating time must be reduced to some minimum while product purity or recovery is maintained within acceptable limits.

If product removal is too fast, separation and the quantity of the product are reduced. Conversely, if the product is withdrawn to maintain separation, its withdrawal rate is reduced and operating time is increased. However, the set point to a composition controller can be programmed so that the average composition of the product will still be within specifications while withdrawal rate is maximized.[9]

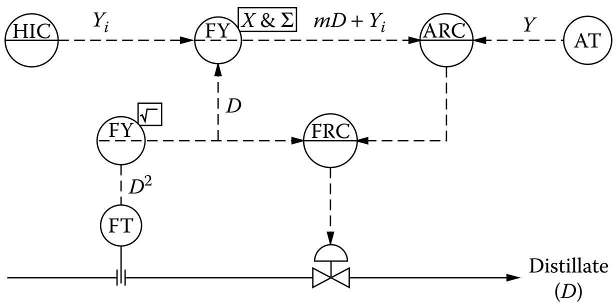


FIG. 8.19p
Control system for batch distillation.

Figure 8.19p shows the control system that will accomplish this when the vapor rate from the batch column is maintained constant. The equation describing this operation is:

$$Y = mD + y_i \qquad \textbf{8.19(44)}$$

where

y = the fraction of key component in the product
m = the rate of change of y with respect to the distillate (D)
y_i = the initial concentration of the product

The only adjustment required is the correct setting of m. The higher its value, the faster y will change and the smaller will be the quantity of material recovered.

CONTROL OBJECTIVES AND STRATEGIES

Operating objectives include the composition specifications for the top and bottom product streams. Other objectives can include increasing throughput, enhancing column stability, and operating against equipment constraints. Yet other considerations include what product composition is considered most important to maintain during disturbances, what are acceptable variations in product specifications, and what are relative economic values of the product streams and cost of energy used in the separation.[10]

The column operating objectives are ultimately governed by economic benefits that are measurable, significant, and achievable.[11] Economics of individual fractionators may continually change throughout the life of the plant. Prices and costs may determine that energy savings are important at one particular time but that recovery is more important at some other time.

The economic benefits of fractionator control include shifting of less profitable components into more profitable products, energy conservation, and increased throughput. Other benefits arise, including minimum disturbances propagated to downstream units, minimum rework or recycle of off-spec products, and more consistent product quality. Thus, a given column's operating economics and, therefore, its objectives may change with time.

When minimization of fractionator utilities is an objective, the following guidelines are recommended:

- Implement control to achieve composition control on all products of the fractionator
- Operate the fractionator to produce minimum overseparation
- Ascertain that the reduction in energy usage is reflected in the energy inflow to the production complex
- Minimize energy waste from blending of overseparated products

Alternative Control Strategies

Many choices confront the design engineer when selecting the control variables for a column. The first decision involves configuration of the top or bottom control loops, which directly determines product compositions. Once these strategies are tentatively determined, the control strategies for the remaining variables (e.g., pressure or levels) become easier to select.

Pairings of controlled and manipulated variables are normally made according to the single-input single-output (SISO) method. Multivariable control, where multiple-input and multiple-output (MIMO) variables are paired, are discussed in Section 8.21. In these multivariable strategies, although a controlled variable can be affected by several manipulated variables, only one manipulated variable is used to directly affect the controlled variable. The minimum number of controlled variables for a fractionator tower is four. These include:

Controlled Variables	*Manipulated Variables*
Overhead composition	Reflux flow
Bottoms composition	Reboiler heating media flow
Accumulator level	Distillate flow
Bottoms level	Bottoms flow

This allows for 24 possible configurations (4 factorial). Of course, most towers include pressure as a controlled variable, with condenser flow or vapor bypass as a manipulated variable. Additional manipulated variables can include feed flow and enthalpy. If a tower includes a sidedraw stream, another control pair is added to the possible combinations.

In fact, additional control variables increase the number of possible control configurations factorially (e.g., six variables produce 720 possible configurations).

The pairing of controlled and manipulated variables can follow three general control structures: energy balance control, material balance control, and ratio control.[12] Energy balance control uses reflux and reboiler heating media flow to control compositions, thus fixing the energy inputs.

Material balance control uses the distillate and bottoms product flows to control compositions, thus fixing the overall material balance. Ratio control utilizes a ratio of any two flow rates at each end of the column. The two common examples of ratio control are the control of reflux-to-distillate ratio and the boil-up-to-bottoms ratio. These control configurations perform quite differently depending upon the fractionator characteristics.

CONTROL LOOP INTERACTION

The selection of which product composition to control (or both, if control of both can be controlled) and the decision on which variables will give better control can be aided by calculation of a relative gain array. The concept of relative gain[13,14] provides a measure of the interaction that can be expected between control loops. This subject is covered in more detail in Chapter 2 in Section 2.12 and in Section 8.20. The concept may be used to find the control configurations that will have the least amount of interaction. Therefore, relative gain analysis should be considered the first step in evaluating alternative composition control strategies.

In addition, some pairings can be made heuristically from operating experience and on the basis of a general understanding of column dynamics (Table 8.19q).

The following are general rules used to reject some possible control pairings:[15,16]

1. Overhead composition and bottoms composition should not both be controlled with material balance equations if the objective is to control product specifications at both ends of the fractionator.

Because of lack of dynamic response the following loops should not be paired:

1. Accumulator level should not be controlled with reboiler heat if the reboiler is a furnace.
2. Bottoms level should not be controlled with reboiler heat if the reboiler is a furnace.
3. Bottoms level should not be controlled with distillate flow.
4. Accumulator level should not be controlled with bottoms product flow.
5. Overhead composition should not be controlled with bottoms product flow.
6. Bottoms composition should not be controlled with distillate flow.
7. Bottoms level should not be controlled with reflux flow.
8. Bottoms composition should not be controlled with reflux flow if the number of trays is greater than a minimum limit (approximately 20).
9. Bottoms level should not be controlled with reboiler heat if the diameter of the column is greater than a minimum limit (approximately 15–20 ft (4.5–6 m), indicating a high volume of liquid in the bottoms).

TABLE 8.19q
Dynamic Response and Sensitivity Limitations on the Pairing of Distillation Control Variables[4]
(Both compositions should not be controlled by material balance (B,D) if both specifications are important)

Controlled Variable / *Manipulated Variable*	*Distillate Flow (D)*	*Bottoms Product Flow (B)*	*Vaporization Rate (V) or Heat Input at Reboiler (O)*	*Reflux Flow Rate (L)*
Composition of Overhead Product (AC_y)	OK if $L/D \geqq 6$ Note 3		Notes 1 and 2	Note 2
Composition of Bottoms Product (AC_x)		Note 3	Notes 1 and 2	OK if trays $\leqq 20$
Accumulator Level (LC_a)	OK if $L/D \leqq 6$		Not good with furnace OK if $V/B \geqq 3$	OK if $L/D \geqq 0.5$
Bottoms Level (LC_b)		OK if $V/B \leqq 3$	Not good if furnace is used OK if diameter at bottom $\leqq 20$ ft	

Notes: 1. Control that concentration (x or y) which has the shorter residence time by throttling vapor flow (v).
2. More pure product should control separation (energy).
3. Less pure product should control material balance.
4. When controlling both x and y, the only choices for possible pairings are:
  a. Control y by D and x by V.
  b. Control y by D and x by L.
  c. Control y by L and x by V.
  d. Control y by B and x by L.

Of these, choice *d* is not recommended because *a* y/B combination is not responsive dynamically.

10. Accumulator level should not be controlled with reboiler heat if the control objective is to maintain overhead product specification and the V/B ratio is less than a minimum limit (approximately 3).

Because of lack of sensitivity, these loops should not be paired:

1. Overhead composition should not be controlled with reflux flow if the reflux ratio (L/D) is less than a minimum value (approximately 6).
2. Accumulator level should not be controlled with distillate flow if the reflux ratio (L/D) is less than a maximum value (approximately 6).
3. Accumulator level should not be controlled with reflux flow if the reflux ratio (L/D) is less than a maximum value (approximately 0.5).
4. Bottoms composition should not be controlled with sidedraw flow if the sidedraw is a vapor phase.
5. Overhead composition should not be controlled with sidedraw flow if the sidedraw is a liquid phase.
6. Bottoms composition should not be controlled with sidedraw flow if the sidedraw is a liquid phase and the sidedraw tray number is greater than a minimum number (approximately 20).
7. Sidedraw composition should not be controlled with reflux or distillate flow if the difference between the total number of trays and the number of the sidestream tray is greater than a minimum value (approximately 20).
8. Bottoms level should not be controlled with sidedraw flow if the difference between the bottoms and the number of the sidestream tray is greater than a minimum value (approximately 100).
9. Bottoms level should not be controlled with bottoms flow if the V/B ratio is greater than a minimum limit (approximately 3).

Choices for controlling product compositions include (1) controlling top or bottom composition only (generally suitable for constant separation conditions, where specifications for one product are loose or where effective feedforward/feedback systems can be designed to compensate for load changes) and (2) controlling of both product compositions (minimizes energy use and provides tight specification top and bottom products for columns in which the problems of interaction are small).

These choices can be broken down further into considerations such as manipulation of distillate-boil-up, *DV* configuration (generally suitable for high reflux columns) or manipulation of reflux-boil-up, *LV* configuration (generally suitable for low reflux columns), and so forth.

Further considerations include the use of decoupling control schemes (can present practical problems, such as insensitive control, operating problems, and high sensitivity to errors) and the use of temperature measurements to infer composition or analyzers to measure composition directly (generally an economic decision based on how well a temperature-sensitive control point can be determined and the costs of analyzer hardware and maintenance). These choices are based on operating objectives of the column, expected disturbance variables, and the degree of control loop interaction.

PRODUCT QUALITY CONTROL

Conceptually, product control is a problem of making precise adjustments to the rate of heat addition and the rate of heat removal from the tower. Heat removal determines the internal reflux flow rate, and the internal reflux as measured on the top tray is a direct reflection of the composition of the distillate. Heat added determines the internal vapor rate. These internal vapor and liquid flow rates determine the circulation rate, which in turn determines the degree of separation between two key components.

Once interaction of the various variable pairings has been established, and the column's operating objectives and disturbance variables are considered, the primary composition control loops of the column can be selected. Measurement of these control variables can be either direct or inferred.

Inferring Composition from Temperature

If the cost of on-line analyzer hardware and maintenance is prohibitive, or if backup is desired in case of analyzer failure or maintenance, and because the results of laboratory analysis take too long to be usable for effective control, temperature measurement often can be used to infer composition.

Because distillation separates materials according to their difference in vapor pressures, and because vapor pressure is a temperature-controlled function, temperature measurement has historically been used to indicate composition. This presumes that the column pressure remains constant, or that the temperature measurement is compensated for pressure changes, and that feed composition is constant. Then, any change in composition within a column will be detected as a temperature change.

The best point to locate the temperature sensor cannot be established from generalizations. The important consideration is to measure the temperature on a tray that strongly reflects the changes in composition. When composition of the bottom product is important, it is desirable to maintain a constant temperature in the lower section. This can be done by letting the temperature measurement manipulate the reboiler steam supply by resetting the steam flow controller set point (Figure 8.19r).

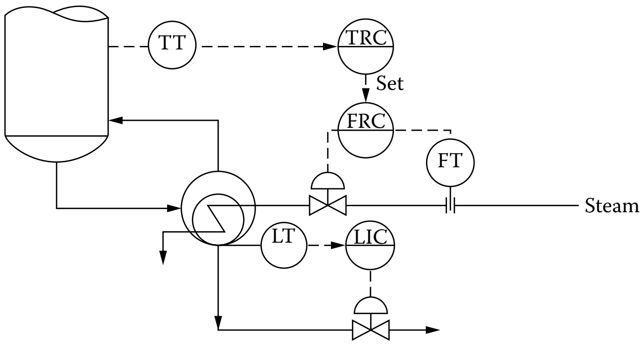


FIG. 8.19r
In this configuration the reboiler heat input is throttled by a temperature controller to keep the bottoms product composition constant.

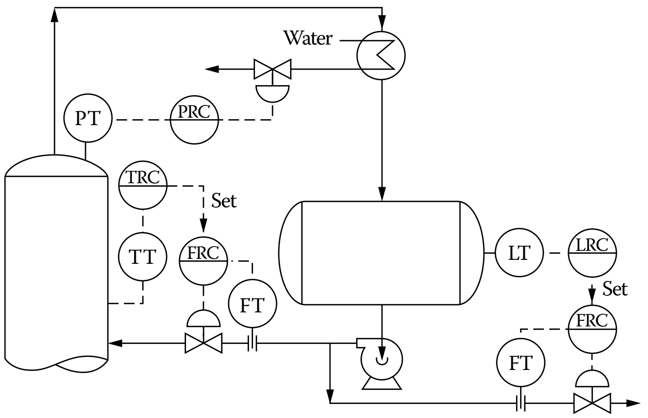


FIG. 8.19s
If overhead composition is to be controlled, the reflux flow to the column is throttled by a temperature controller.

When composition of the distillate is more important, it is desirable to maintain a constant temperature in the upper section, as in Figure 8.19s. In this configuration the sensing point for column pressure control should be located near the temperature control point. Keeping the sensor locations close to each other helps to fix the relation between temperature and composition at this particular point.

If column temperature profiles caused by small positive and negative changes in manipulated variables, such as a ±1% change in distillate flow (Figure 8.19t), can be generated, the

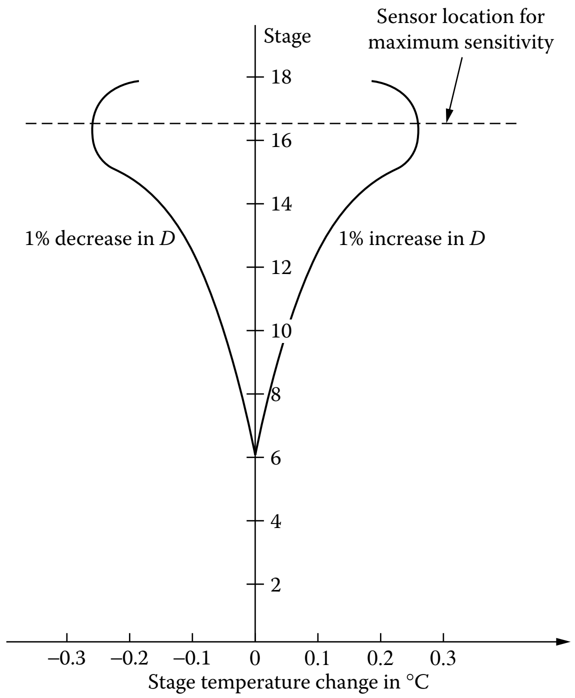


FIG. 8.19t
Example of column temperature profiles resulting from a 1% increase and from a 1% decrease in distillate flow.

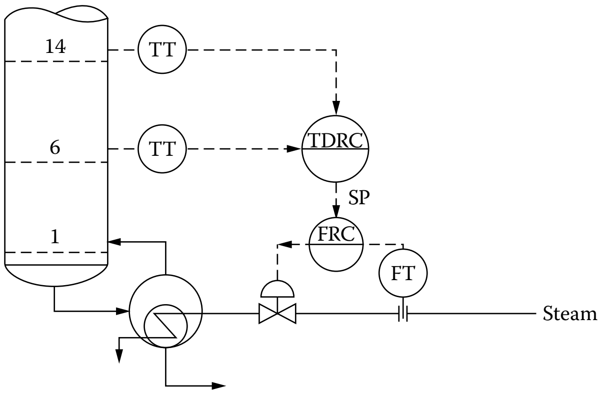


FIG. 8.19u
Heat input controlled by temperature difference.

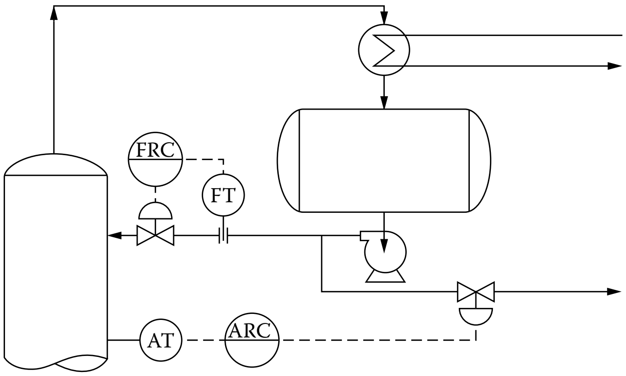


FIG. 8.119v
Distillate withdrawal controlled by chromatograph.

following criteria may be helpful in selecting sensor locations:[17] (1) The sensitivity of the temperature-manipulated variable pairing should be in the range of 0.1 to 0.5°C/% and (2) equal temperature changes should result when increasing and when decreasing the manipulated variable.

For a two-product fractionator, distillation temperature is an indication of composition only when column pressure remains constant or if the temperature measurement is pressure-compensated. When separation by distillation is sought between two compounds having relatively close vapor pressures, temperature measurement, as an indication of composition, is not satisfactory.

Fixing two temperatures in a column is equivalent to fixing one temperature and the pressure. Thus, by controlling two temperatures, or a temperature difference, the effect of pressure variations can be eliminated. The assumption used here is that the vapor pressure curves for the two components have constant slopes.

Controlling two temperatures is not equivalent to controlling a temperature difference. A plot of temperature difference vs. bottom product composition exhibits a maximum. Thus, for some temperature differences below the maximum it is possible to get two different product compositions.

Separation of normal butane and isobutane (in the absence of other components, such as pentanes and heavier substances) can be accomplished very well by using temperature difference control. Figure 8.19u illustrates how the heat input to such a column can be controlled by a temperature difference controller.

Control by Analyzers

Analytical or composition control is a way to sidestep the problems of temperature control. Although additional investment is needed for the analytical equipment, savings from improved operation usually results. Several types of instruments are available for composition analysis. Of these, the gas chromatograph is the most versatile and most widely used. (For details, refer to Chapter 8 of Volume 1 of this handbook.)

Once, the time required for a chromatographic analysis (several minutes) was a great barrier to its use for automatic control. Since then, the equipment has been enhanced so that analyses can now be made in less than 5 min, and in many cases for low-volatility hydrocarbons, the analysis can be made continuous.

With careful handling, the under 5 min sampling rate will permit closed-loop distillate control. In fact, fractionators are successfully controlled with cycle times as long as 7–10 min by applying dead time compensation algorithms.

Light ends fractionators have been satisfactorily controlled by the use of chromatography. Figure 8.19v illustrates the controls of a superfractionator designed to separate isobutane and normal butane. In this case, the chromatograph continuously analyzes a sample from one of the intermediate trays, and this measurement is used by the analyzer controller to modulate the product draw-off valve.

Overhead and bottoms analyzers typically measure the loss of a valuable product or the presence of impurities. Impurity components are chosen because small concentration variations can be measured more precisely and with better repeatability, and can provide a more sensitive measure of separation. For example, the change of an impurity from 1.0 to 1.1% can be measured with greater precision than a change of the major component from 99 to 98.9%.

When composition analyzers are used in feedback control, several configurations can be considered. These include 1) direct control of a manipulated variable, 2) cascade control adjusting the set point of a slave temperature controller, and 3) analysis control in parallel with temperature control in a selective control configuration. The configuration used depends on the control objective, sensitivity of control, and analysis dead time.

Direct Control by Analyzers Analyzer controllers in a feedback configuration can be considered when the dead time of each analysis update is less than the response time of the

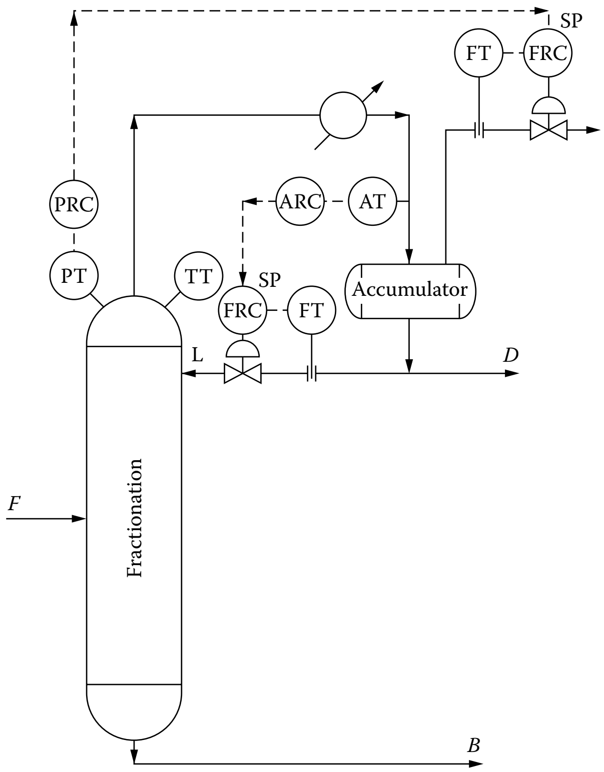


FIG. 8.19w
Direct control of overhead product composition by an analyzer controller (ARC) throttling the set point of a reflux flow controller (FRC).

process. Because it is the control of the composition of the product, which is often the objective, direct control by an analyzer controller would seem to be better than indirect control by temperature.

The composition controller provides feedback correction in response to feed composition changes, pressure variations, and variations in tower efficiencies. Figure 8.19w shows the configuration of a control system, in which a chromatograph analyzes a liquid sample from the condenser rundown line.

A sample probe gathers the liquid sample and the sampling system conditions and vaporizes the liquid sample to provide a representative vapor sample to the chromatograph. The analyzer controller (ARC) uses the chromatographic measurement to manipulate the reflux flow by adjusting the set point to the reflux flow controller (FRC).

Smith Predictor Often the analyzer is so slow that it introduces a significant delay time that degrades the controllability of the process. In that case, some type of dead time compensation is used (see Section 2.19 in Chapter 2). A Smith predictor compensator can serve to model the process to predict what the analyzer measurement should be between analysis updates. When the actual measurement is completed, the model's prediction is compared to the actual measurement and the input to the controller is biased by the difference.

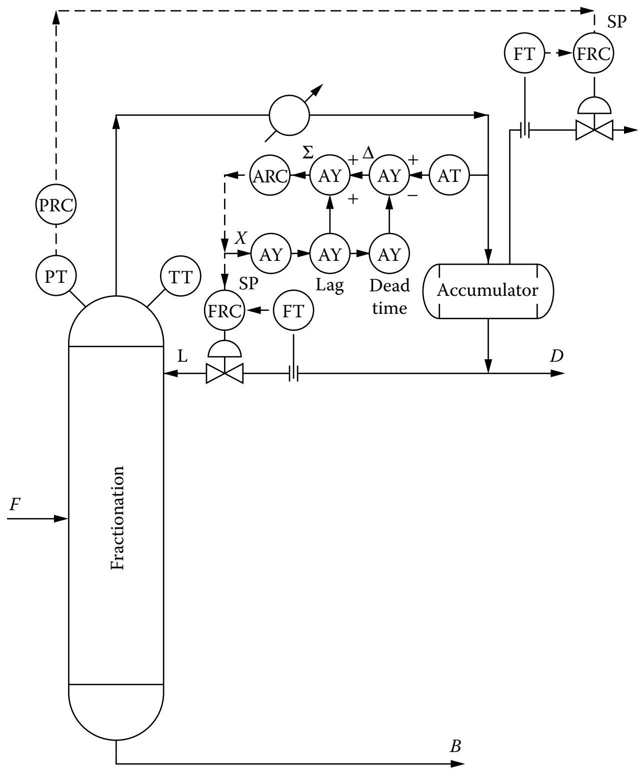


FIG. 8.19x
Analyzer controller with dead-time compensation cascaded to reflux flow control.

Figure 8.19x shows the same configuration as did Figure 8.19w except that the analyzer controller is equipped with a first-order Smith predictor that provides dead time compensation.

In Figure 8.19x, the multiplier, lag, and dead time calculations (*AY*) provide the predicted analysis. (The lag represents the first-order process.) This predicted response is subtracted from the actual measurement to give a differential of the actual process from its own model. This delta is added to the model without dead time to provide a modified pseudomeasurement to the analyzer controller. Thus, the analyzer measurement, which has a significant dead time due to sampling and cycle times, provides a trim to the predicted measurement of the model.

Triple Cascade and Selective Control Analyzer control cascaded to temperature control can be used when stable temperature on a particular tray is desired and the tower operates at a constant, maintainable, and controllable pressure. An example is cascading the analyzer controller to the overhead temperature of a tower, which in turn is cascaded to the reflux flow rate. Because temperature is an indicator of composition at this pressure, the analyzer controller only serves as a trim correcting for variations in feed composition. Figure 8.19y shows this triple cascade configuration of an analyzer controller setting the temperature controller setting the reflux flow controller.

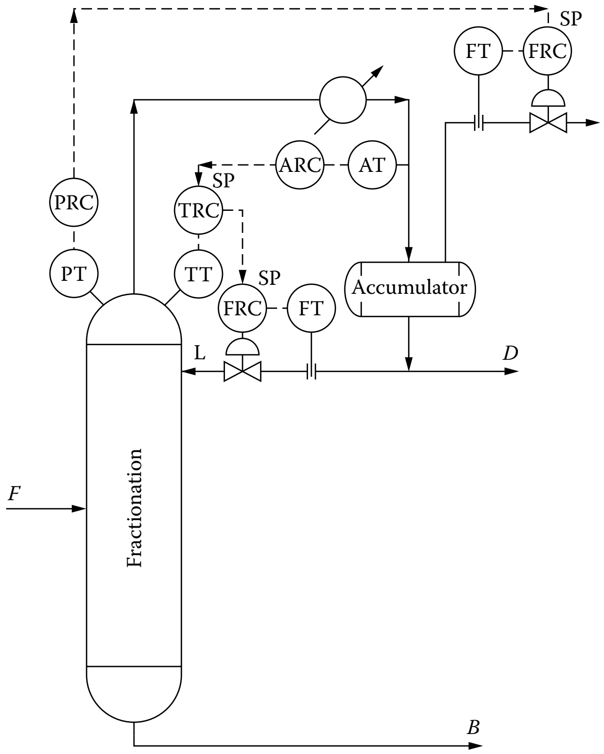


FIG. 8.19y
Triple cascade configuration of overhead composition control.

Analyzer controls can be used in a high or low select configuration in combination with temperature when a high or low limit based on temperature is important. The temperature controller is a constraint controller (see Section 2.28 in Chapter 2 for details) serving to prevent the temperature from exceeding a limit.

An example of using this control configuration is the control of the bottoms of an absorber stripper. Here, the temperature should not exceed a certain value, as no additional stripping of the light component in the bottoms of the column could be accomplished. Even though an analyzer controller may call for more heat, this heat would only increase the bottoms temperature of the recycled oil to the absorber without removing the impurity, thereby reducing the absorption capability at the absorber.

Figure 8.19z depicts an analyzer controller in a low select configuration with the temperature constraint controller.

Note that both cascade and selective control configurations require external feedback to protect them from reset windup. Figure 8.19aa illustrates how the external feedback (EF) is applied to the master controller in a cascade configuration (TIC) and to both controllers (FIC and PIC) in a selective control configuration. For more details on external feedback, refer to Section 2.28 in Chapter 2.

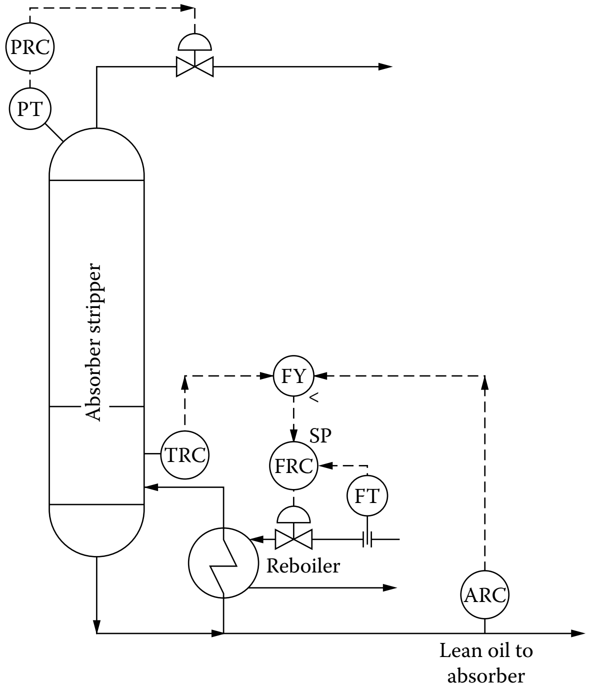


FIG. 8.19z
Analyzer control in a low select configuration with a temperature constraint controller.

Fractionator Trains A controller may use as its measurement the analysis of a single component, or may use the ratio of two components. A ratio (e.g., ethane-to-propane, C_2/C_3) is often used when the fractionator is not the final step in the separation sequence.[18] This often occurs in a natural gas liquids separation train where a de-ethanizer, a depropanizer, a debutanizer, and a deisobutanizer (butane splitter) produce the products ethane, propane, butane, isobutane, and gasoline as shown in Figure 8.19bb.

Specifications for the primary overhead products may include limitations on the amount of both light and heavy impurities. For example, the propane product from the overhead of the depropanizer would have limitations on ethane as well as isobutane. The problem is that the light impurity (lighter than light key) cannot be controlled in the tower that produces that product. Rather, it must be controlled in an upstream tower.

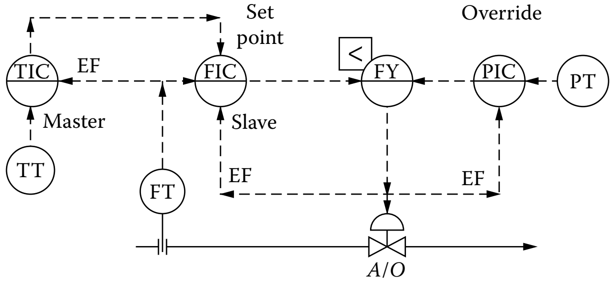


FIG. 8.19aa
In a cascade configuration the external feedback signal (EF) is the slave measurement, while in selective control configurations, it is the signal that is throttling the control valve.

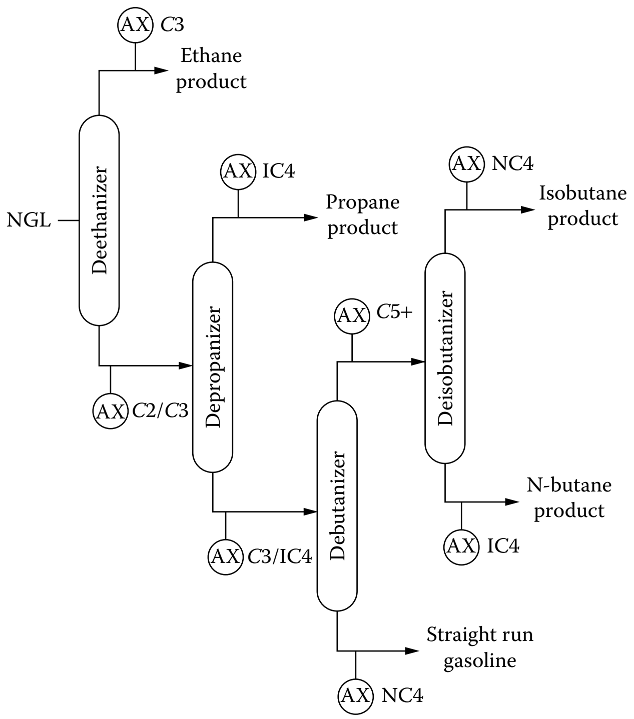


FIG. 8.19bb
Analyzer placement in a fractionator train.

The lighter than light key specification in the distillate of the downstream tower can be controlled more easily by controlling the ratio in the bottoms of the upstream tower. That is, the ethane content in the propane product (depropanizer distillate) is maintained by controlling the C_2/C_3 ratio in the bottoms of the de-ethanizer. Measuring the C_2/C_3 ratio in the bottoms requires an additional analyzer but eliminates the dead time of obtaining the concentration in the overhead of a downstream tower.

A feed analyzer is sometimes included as a part of feedforward control. The feed analysis is used in predicting internal reflux/overhead flow and bottoms/heat input. However, when feed composition changes slowly or when results from the analyzer cannot be obtained faster than the dynamics of the tower, this analyzer is omitted and the burden is placed on feedback control from the product analyzers.

In practice, a feed analyzer is the exception rather than the rule. Its use is mainly when the analyzer is already in place, because it is controlling an upstream tower. For example, the NGL separation train in Figure 8.19bb has a deethanizer bottoms analyzer that could also be considered the depropanizer feed analyzer.

Analyzer Selection The choice of analyzer control depends upon the analytical equipment available and on the type of separation desired. Each type of separation requires a compromise between the controllability and the delay of the control system. For example, the NGL train (Figure 8.19bb) was studied to determine the best analyzer system. In the depropanizer (where isobutane was to be measured in the presence of ethane, propane, and normal butane) and in the deisobutanizer (where isobutane was to be measured in the presence of normal butane and isopentane), an infrared analysis was to be preferred.

However, in the debutanizer the goal was to measure the combined isopentane plus normal pentane concentrations in the presence of isobutane and normal butane to control the butane-pentane separation. Here, investigation revealed that gas chromatography provides the best solution.

Some boiling point analyzers are reliable enough to be used for on-line control (see Section 8.50 in Chapter 8 in Volume 1 of this handbook). Normally, cut points between overhead products and side-cuts are maintained by temperature controllers. These controllers generally influence reflux rate or product draws to achieve the desired results. Laboratory distillation results are used to adjust the set points to the temperature controllers. This method of control, however, is cyclical because of the time lags involved in temperature control.

To avoid exceeding the target cut points and to meet required product specifications, the cut point is set below specification. This results in downgrading the more valuable product to the stream of lesser value. This downgrading can be minimized through the use of on-line boiling point analyzers. Justification of a boiling point analyzer depends upon the value of the products, how much downgrading is occurring, and the cost of analyzer maintenance. Figure 8.19cc illustrates an end-point analyzer.

Viscosity is another property that can be measured continuously to give faster control corrections. In vacuum distillation, the viscosimeter monitors each of the streams for which viscosity is a specification. Any deviation from the desired viscosity is corrected by a change in the set point of the control loop involved.

Deviation from the desired viscosity and the subsequent downgrading of the product can occur because of frequent variations in tower operating conditions and feed composition. In addition to normal operation, the use of a viscosity analyzer minimizes downgrading during major upsets and large feed compositions changes. With such an arrangement, low viscosity vacuum bottoms can be detected quickly and diverted to recoverable feed for profitable reprocessing.

Once again, profitability determination requires a thorough analysis of column operation and an assessment of the engineering, operating, and maintenance capabilities at the location where this type of control is to be implemented.

Many other analytical instruments are being moved out of the laboratory and into the processing area. Mobile units containing several different kinds of analyzers can be used to learn the best place to locate on-stream analyzers. In cases in which permanent analyzers cannot be justified, the mobile unit is connected to the process long enough to find the best operating conditions. Then, the mobile unit can be moved elsewhere.

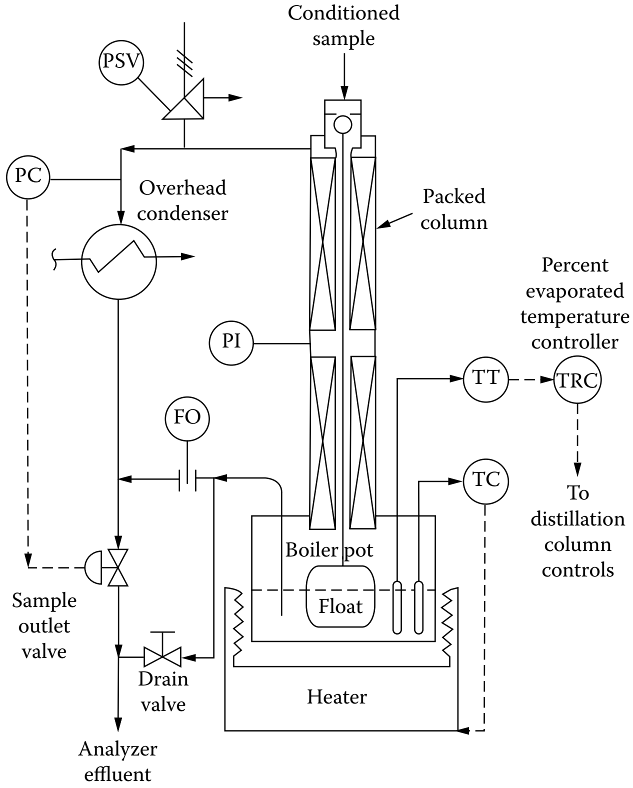


FIG. 8.19cc
The end-point distillation analyzer is a miniature version of the process column.

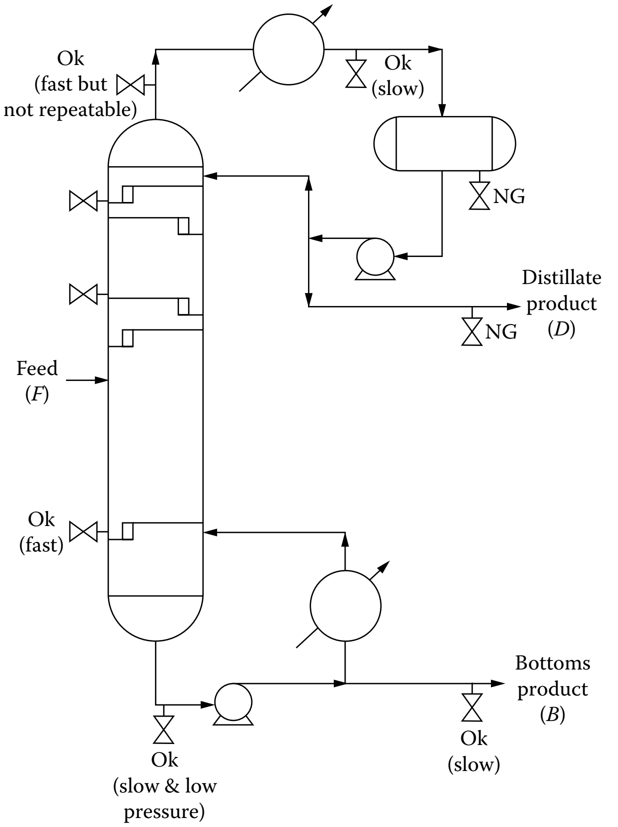


FIG. 8.19dd
The choices of sampling point locations for analyzers used in distillation column control.

Sampling Proper sampling of material in a column is necessary if analyzers are to control effectively. A poor sampling system often is responsible for the unsatisfactory performance of plant analyzers. For details on sampling system design, refer to Chapter 8, Section 8.2, in Volume 1 of this handbook.

The sampling points for composition analysis should be at, or very near, the column terminals for the following reasons: (1) freedom from ambiguity in the correlation of sample composition with terminal composition, and (2) improved control loop behavior as a result of reduction of transport lag (dead time) and of the time constants (lags) describing the sampling point's compositional behavior. This assumes that the controller applies its manipulation at the same terminal (steam or reflux) where the controlled variable is measured.

The factors favoring moving the point of sampling nearer to the feed entry point are (1) improved terminal composition behavior as a result of earlier recognition of composition transients as they proceed from the feed entry toward the column terminals, and (2) less stringent analytical requirements as a result of (a) analyzing the control component at a higher concentration and over a wider range, and (b) simplifying the multicomponent mixture, because nonkey components tend to exhibit constant composition zones in the column.

Figure 8.19dd shows some of the typical sample locations of a distillation tower. Most analyzers are designed to accept a clean, dry, noncorrosive sample at low temperatures, pressures, and flow rates. Such conditions seldom exist in the process, so the sampling system must be designed and operated to overcome the difference between the conditions in the process and the conditions required by the analyzer.

The sample system must provide a current and representative sample of the stream being analyzed. It must transport the sample from the sample point to the analyzer with a minimum of transport lag (preferably less than 30 sec and definitely not greater than 1 min). Transportation times are minimized using high flow rate bypass streams taken from the process sample point and returned to the process at a lower pressure. The sample system must condition the sample to remove traces of foreign materials through filtering, maintain pressure and temperature, and maintain or change phase for introduction into the analyzer.

For chromatographs, liquid sample points are generally preferred (Figure 8.19ee). This is because vapor streams have historically not provided representative samples. Vapor samples do not tend to produce repeatable values as consistently and reliably[19] because of condensation at the sample probe and in the sample lines when hydrocarbons of high boiling

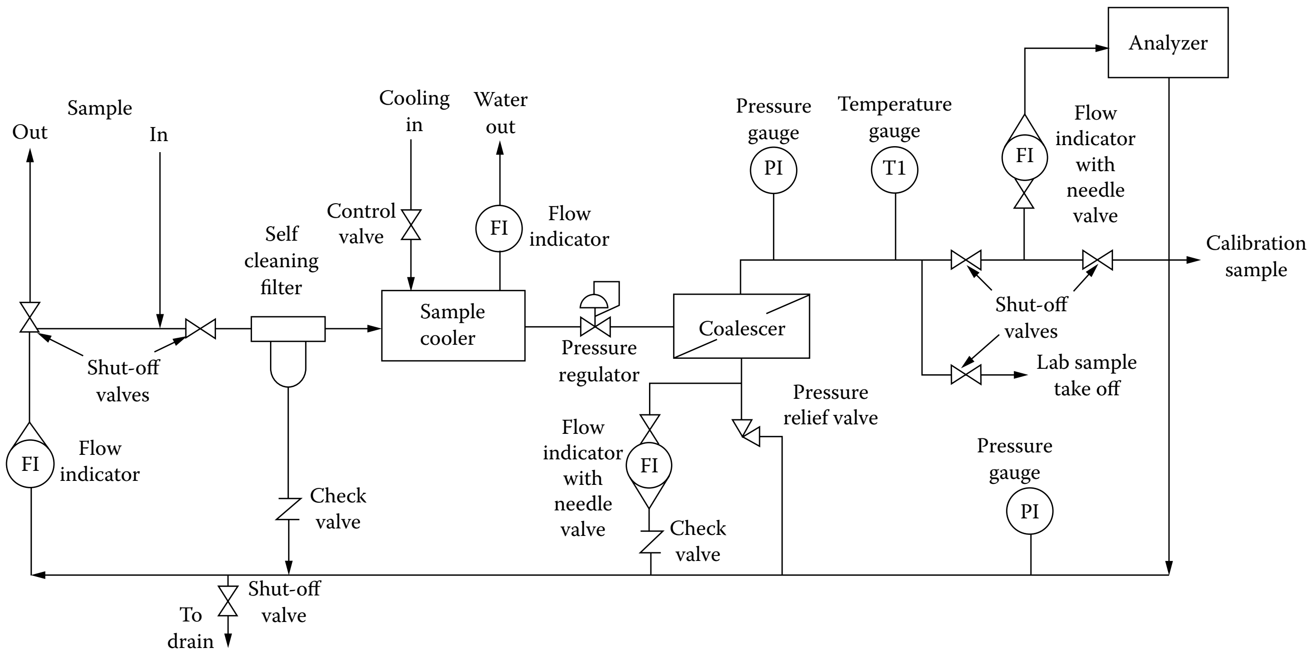


FIG. 8.19ee
Sampling system for a liquid product in a refinery application.

points are present in the sample. When the sample lines are long, some separation between components can also occur. However, vapor samples can be used when warranted and if the proper care can be taken.

A satisfactory point for measuring bottoms product composition is at the point of highest pressure. This approach will ensure a representative sample and will provide the pressure drop to return the sample bypass. The point of highest pressure is generally immediately after the product pump. However, if liquid holdup in the reboiler and kettle is large, a long lag is introduced, which slows the transient response of the measurement and control system. Alternative sample points such as a bottoms tray or seal pan may be used, but may require extra expense for the sample system.

A satisfactory sample-point location for measuring the distillate is the outlet liquid of the overhead vapor condenser. Sampling the overhead accumulator liquid after the reflux or distillate pump should be avoided because of the tremendous process lag it introduces. Sampling the overhead vapor reduces the process lag of sampling after the condenser if a repeatable, representative sample can be obtained.

PRESSURE CONTROL

Most distillation columns are operated under constant pressure. However, floating-pressure operation can have advantages in many processes. One reason for the resistance to the use of floating-pressure control is based on the fact that temperature is sensitive to pressure changes, and therefore, it requires pressure compensation if the pressure varies. As analyzers are increasingly replacing temperature-based controls, the argument favoring constant pressure operation is also lessening. However, even when temperature control is used, the temperature measurements can be compensated for pressure variations.

The primary advantage of floating-pressure control is the ability to operate at the minimum column pressure within the constraints of the system. Lower pressure reduces the volatility of distillation components, thereby reducing the heat input required to effect a given separation. Other advantages include increased reboiler capacity and reduced reboiler fouling due to lower tower temperatures.

In the following paragraphs, floating-pressure control strategies will be described for the following conditions: (1) liquid distillate withdrawn when noncondensables are present, (2) vapor distillate withdrawn when noncondensables are present, and (3) liquid distillate withdrawn when the amount of noncondensables is negligible.

Liquid Distillate and Inerts

In some separation processes the problem of pressure control is complicated by the presence of large percentages of inert gases. The noncondensables must be removed, or they will accumulate and blanket off the condensing surface, thereby causing loss of column pressure control.

The simplest method of handling this problem is to bleed off a fixed amount of gases and vapors to a lower pressure unit, such as to an absorption tower, if one is present in the system.

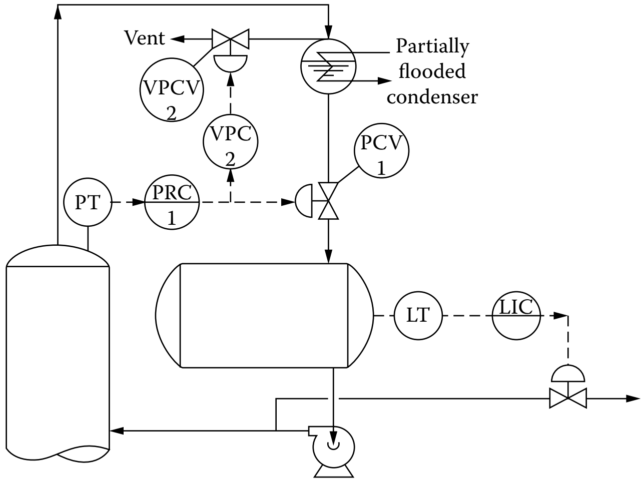


FIG. 8.19ff
Column pressure control with inerts present.

If an absorber is not present, it is possible to install a vent condenser to recover the condensable vapors from this purge stream. Often in refinery applications such noncondensables go to the fuel gas system or to flare.

It is recommended that the fixed continuous purge be used whenever economically possible; however, when this is not permitted, it is possible to modulate the purge stream. This might be desirable when the amount of inerts is subject to wide variations over time.

As the noncondensables build up in the condenser, the pressure controller will tend to open the control valve (PCV-1 in Figure 8.19ff) to maintain the proper rate of condensation. The controller signal that is throttling PCV-1 could also be used to start opening the purge control valve (VPCV-2), when the opening of PCV-1 reaches some preset limit. This can be done by means of a calibrated valve positioner or by using a valve position controller (VPC-2) in Figure 8.19ff.

Vapor Distillate and Inerts

In the case where the distillate is in the vapor phase and inerts are present, the overhead product is removed under pressure control as shown in Part A of Figure 8.19gg. In this configuration the system pressure will quickly respond to changes in the distillate vapor flow. In this control system a level controller is installed on the overhead receiver to regulate the cooling water to the condenser, so that it will condense only enough condensate to provide the column with reflux.

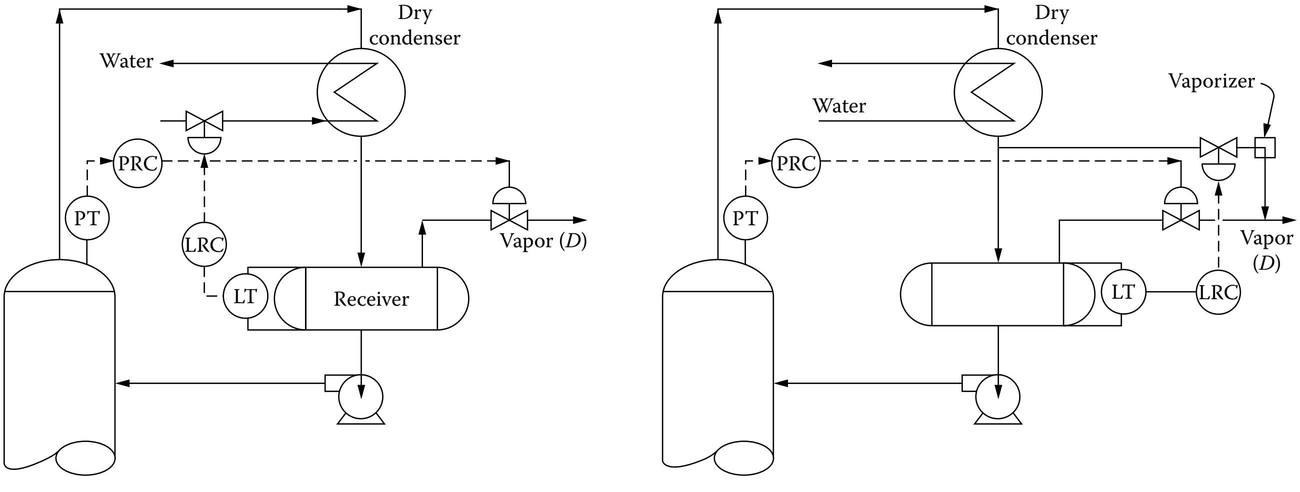
This control system will operate properly only if the condenser is designed to provide a short residence time for the coolant, which will minimize the level control time lag. If this is not the case, the cooling water flow should be maintained at a constant rate.

In this case (Part B in Figure 8.19gg), the level controller can regulate the flow of condensate through a small vaporizer and mix it with the vapor from the pressure control valve.

If the cooling water has fouling tendencies, it is preferable to use the control system shown in Figure 8.19hh, where a pressure controller regulates a vapor bypass around the condenser.

Liquid Distillate with Negligible Inerts

In distillation processes where the distillate is in the liquid phase and the amount of inerts is negligible, the column pressure is usually controlled by modulating the rate of condensation in the condenser. The method of controlling the



A. Variable water flow

B. Constant water flow

FIG. 8.19gg
Column pressure control when the distillate is in the vapor phase and contains inerts for variable (A) and constant condenser water flow configurations (B).

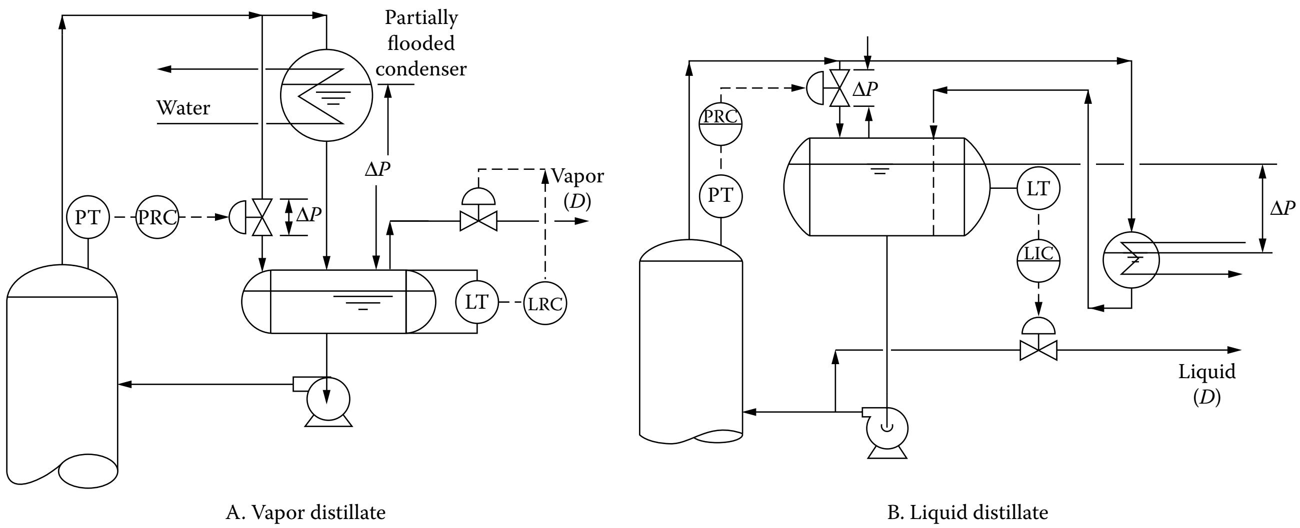


FIG. 8.19hh
Column pressure controlled by hot gas bypass throttling in case of vapor (A) and liquid (B) distillate processes.

rate of condensation depends upon the mechanical construction of the condensing equipment.

Controlling the Cooling Water Flow Figure 8.19ii describes a control configuration where the column pressure is controlled by throttling the cooling water flow from the condenser. This method of control is recommended only when the cooling water is treated with chemicals that prevent the fouling of the tubes in the event of high temperature rise across the condenser tubes. In such configuration, the maintenance costs are low, because the control valve is on the water side and the control performance is acceptable, provided the condenser is properly designed.

The best condenser for this service is a bundle-type unit with the cooling water flowing through the tubes. This water should be flowing at a rate of more than 4.5 ft/s (1.35 m/s), and the water should have a residence time of less than 45 sec. The shorter the residence time of the water, the better will be the quality of control obtained, owing to the decrease in dead time or lag in the system.

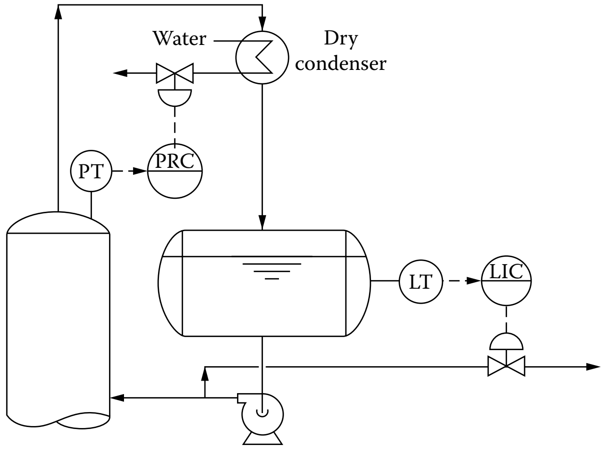


FIG. 8.19ii
Column pressure control by throttling condenser water.

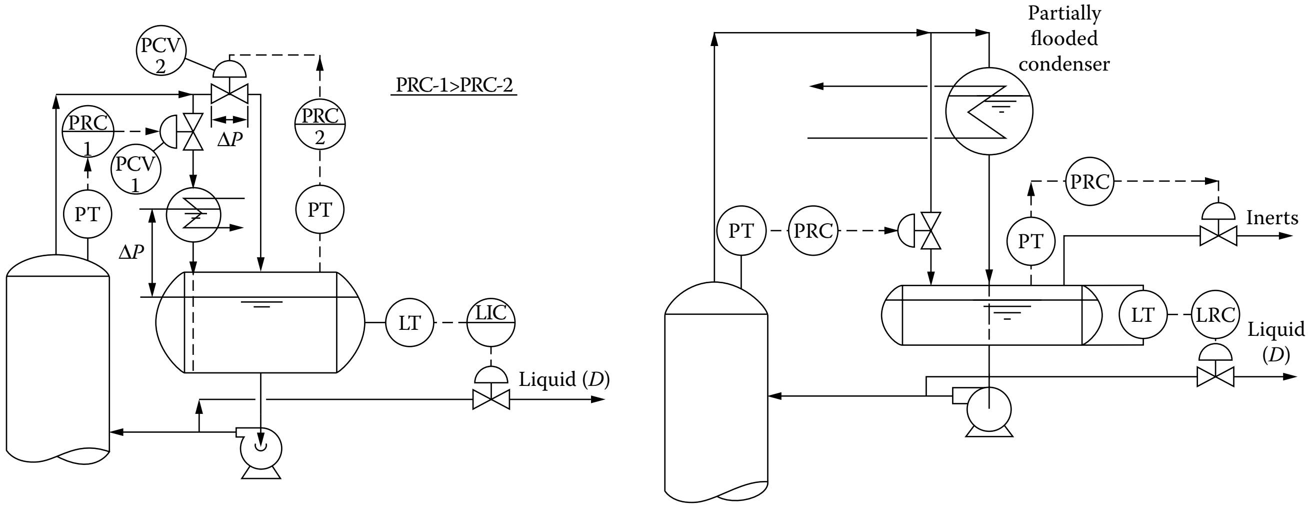
With a properly designed condenser, the pressure controller needs only proportional control, because a narrow throttling range is sufficient. However, as the residence time of the water increases, the time lag of the system will increase, and consequently the controller will require a wider throttling range and will need automatic reset to compensate for the load changes.

The control obtained by using a wide proportional band is not satisfactory for precision distillation columns because of the length of time required for the system to recover from an upset. Also, the dead time varies with load, and therefore the integral setting of the PRC should be set to match the variation in residence time.

Therefore, it is unacceptable to use this control system on a condenser box with submerged tube sections, because there would be a large time lag in the system due to the large volume of water in the box. It would require the passage of a significant amount of time before a change in water flow rate would change the temperature of the water in the box and finally would affect the rate of condensation.

Controlling the Condensate Flow To reduce such unfavorable time lags, it becomes necessary to use a different type of control system, one that permits the water flow rate to remain constant and controls the amount of surface exposed to the condensing vapors. This is done by modulating the flow of condensate from the condenser.

When the column pressure is dropping, this condensate throttling valve reduces the condensate flow, causing it to



A. No inerts

B. With inerts

FIG. 8.19jj
High-speed column pressure control.

build up and flood more tube surface and, consequently, to reduce the condensing surface exposed to the vapors. Thereby, the condensing rate is reduced, and the pressure in the column rises. In such designs, a vent valve should be installed to purge the noncondensables from the top of the condenser if it is expected that noncondensables could build up and blanket the condensing surface.

Control of Hot Vapor Bypass A third possible control configuration for applications with liquid distillate containing negligible inerts is used when the condenser is located below the receiver. This is frequently done to make the condenser available for servicing and to save on steel work. It is usual practice to elevate the bottom of the accumulator 10–15 ft (3–4.5 m) above the suction of the pump in order to provide a positive suction head on the pump.

In this type of installation the control valve is placed in a bypass of the vapor line to the accumulator (see Figure 8.19hh, Part B). When this valve is open, it equalizes the pressure between the vapor line and the receiver, causing the condensing surface to become flooded with condensate because of the 10–15 ft of head in the condensate line from the condenser back to the receiver.

The flooding of the condensing surface causes the pressure to build up because of the decrease in the rate of condensation. Under normal operating conditions, the subcooling that the condensate receives in the condenser is sufficient to reduce the vapor pressure in the receiver. The difference in pressure permits the condensate to flow up the 10–15 ft of pipe between the condenser and the accumulator.

A modification of this latter system controls the pressure in the accumulator by throttling the condenser bypass flow (Figure 8.19jj, Part A). The column pressure is maintained by throttling the flow of vapor through the condenser. Controlling the rate of flow through the condenser provides faster pressure regulation for the column.

Part A of Figure 8.19jj shows the operation if there are no inerts, as follows: If column pressure rises, PRC-1 opens PCV-1. This increases the vapor pressure in the condenser, which pushes some of the condensate out of it and increases the condensing surface area exposed to the vapors. Therefore, the rate of condensation is increased, and thereby the column pressure is lowered back to the set point of PRC-1.

At this higher rate of condensation, the pressure drop (ΔP) across PCV-2 is also reduced (the valve opens). If the column pressure drops, the opposite sequence occurs: PCV-1 closes and the flooding of the condenser increases, reducing the rate of condensation and increasing the pressure drop (ΔP) across PCV-2 by slightly closing it. The setting of PRC-1 must always be above that of PRC-2.

The most common pressure control configuration is shown in Part B of Figure 8.19jj. Here, the column pressure controller is throttling the hot vapor bypass, as was the case in Part A of Figure 8.19hh, but in addition a second pressure controller is utilized on the accumulator. This PRC is set at about 5 PSIG below the required tower pressure and is used to vent the inert gases that may build up in the system.

Vacuum Systems

For some liquid mixtures, the temperature required to vaporize the feed would need to be so high that decomposition would result. To avoid this, it is necessary to operate the column at pressures below atmospheric. Steam jet ejectors are often used to create vacuum in distillation systems. These

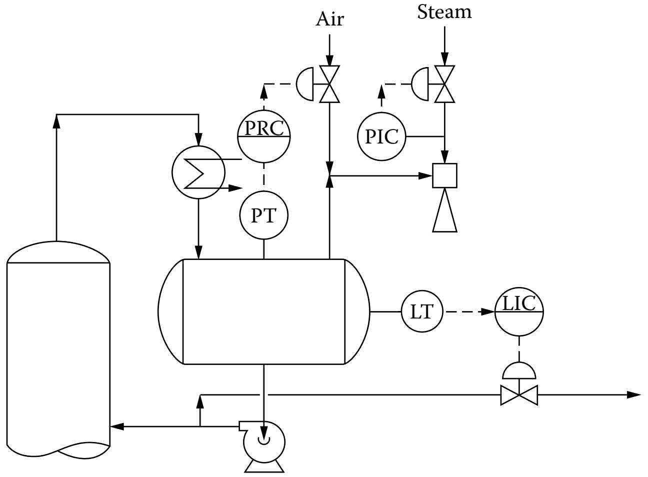


FIG. 8.19kk
Vacuum column pressure control.

can be used singly or in stages, when a wide range of vacuum conditions are required. The acceptance of steam jet ejectors is due to their having no moving parts and requiring very little maintenance.

Most ejectors are designed for a fixed capacity and work best at one steam condition. Increasing the steam pressure above the design point will not usually increase the capacity of the ejector; instead, it will sometimes decrease the capacity because of the choking effect of the excess steam in the diffuser throat.

Steam pressure below a critical value for a jet will cause the ejector operation to be unstable. Therefore, it is recommended that a pressure controller be installed on the steam to keep it at the optimum pressure required by the ejector.

The recommended control system for pressure control in vacuum distillation applications is shown in Figure 8.19kk. Here, a controlled rate of air or gas is bled into the vacuum line just ahead of the ejector. Closing this bleed valve makes the maximum capacity of the ejector available to handle any surges or upsets in the process load. A control valve regulates the amount of bleed air used to maintain the pressure on the reflux accumulator. Using the pressure of the accumulator for control involves less time lag than if the column pressure were used as the control variable.

Because ejectors are fixed capacity, the variable load is met by air bleed into the system. At low loads this represents a substantial waste of steam. Therefore, if substantial load variations are expected, operating costs can be lowered by installing a larger and a smaller ejector. This makes it possible to automatically switch to the small unit when the load drops off, thereby reducing the steam demand.

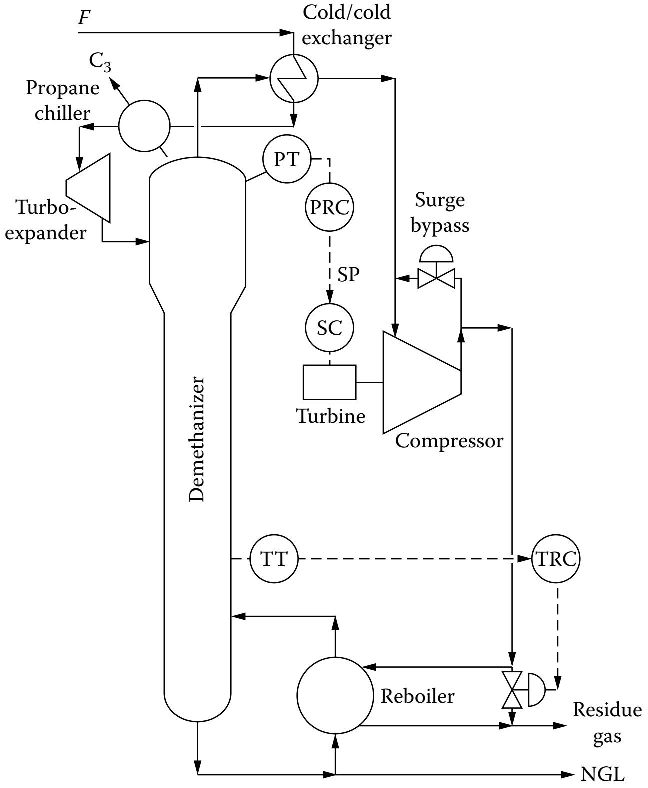


FIG. 8.19ll
Vapor recompression pressure control.

Vapor Recompression

Vapor recompression is another means of improving energy efficiency of the operation. The overhead vapor from the distillation column is compressed to a pressure where its condensation temperature is greater than the boiling point is at the pressure of the tower bottoms. The heat of condensation of the overhead can then be used as the source of heat for reboiling the bottoms.

This scheme is known as vapor recompression. It is used fairly often when the distillation involves a relatively close-boiling mixture, and the boiling points of the top and bottom products are similar. In cryogenic demethanization processes, illustrated in Figure 8.18ll, the column pressure is controlled by the throttling of the speed of the recompression compressors.

The heat of condensation of the overhead is also used as the heat for the reboilers in propylene fractionators. Figure 8.19mm shows the pressure controls needed for the operation of this vapor recompression via the heat pump on this particular tower.

FCCU main fractionators and crude towers make use of compressors to “draw” vapors from the tower because operation is essentially at atmospheric pressures. The pressure control system used in this case is shown in Figure 8.19nn. In this configuration, the tower pressure is maintained by controlling the speed of the compressor. This is accomplished by the manipulation of the steam used to drive the compressor turbine.

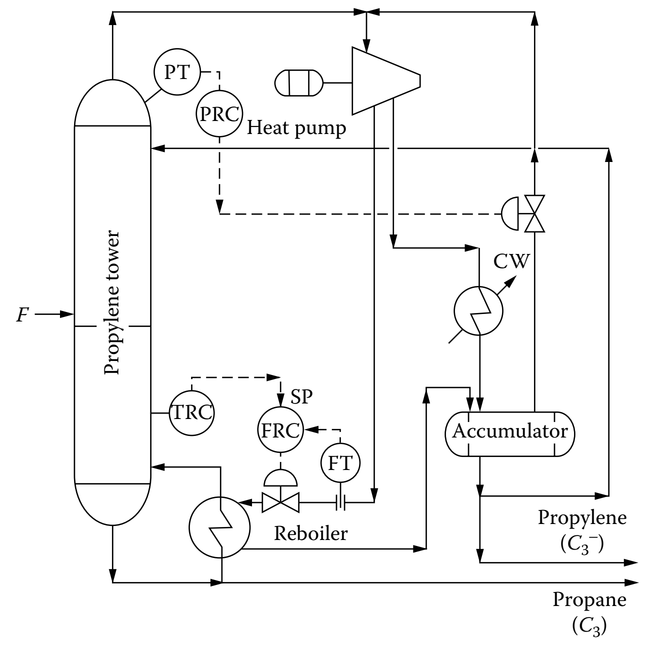


FIG. 8.19mm
Propylene tower vapor recompression pressure control.

FEED CONTROLS

One of the best means of stabilizing the operation of almost any continuous-flow-process, including distillation, is to hold the flow rates and operating temperatures constant. Therefore, whenever possible, a flow controller should be used on the feed to maintain a constant rate of flow.

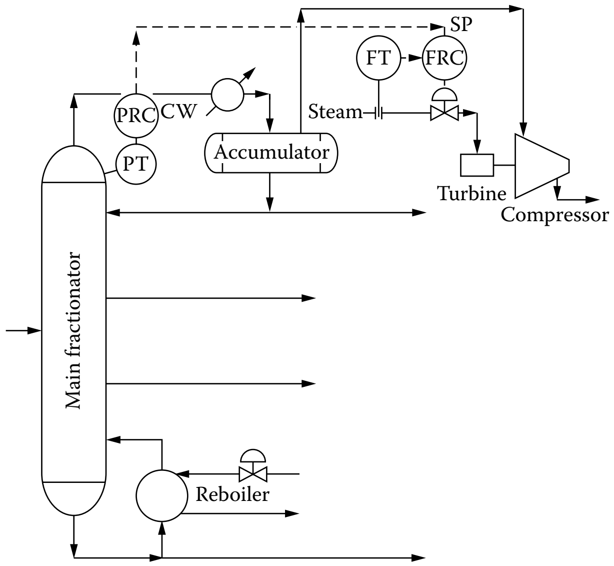


FIG. 8.19nn
Main fractionator pressure control.

Feed Flow Control

A flow controller in the feed line can maintain a constant flow rate. In some instances, the feed pump of a distillation unit is a steam-driven pump instead of an electrically driven one. In this case, the controller modulates the steam to the driver.

Feed composition has a great influence upon the operation of a distillation unit. Unfortunately, feed composition is seldom subject to adjustment. For this reason, it is necessary to make changes elsewhere in the operation of the column in order to compensate for the variations in feed composition. The corrective steps are discussed later. The discussion below assumes a constant feed composition.

Variable Column Feed

Having constant feed conditions simplifies the amount of control required to achieve stable operation. However, the distillate product is often fed to a second column. Then, any changes that occur in the first column are reflected in the quantity and composition of the feed to the second. If the feed flow to the column is controlled by a liquid level controller of the previous column, that controller can be tuned with a low gain, so that the level can swing over a wide range without drastically upsetting the flow of product.

Nevertheless, the second column will receive a varying flow of feed if it is linked to the first column. One way to iron out temporary variations caused by liquid level changes is to cascade the level to a flow controller in the product lines. Flow controllers also serve to smooth out the pressure fluctuations caused by the distillate/reflux pump.

With variable feed rates and variable feed compositions, cascade controls are justified. If the feed rate and composition are relatively constant, resetting the major control loop manually is sometimes adequate. In other cases the flow controller is arranged as the cascade slave of the level controller (Figure 8.19oo). The control algorithm for the level controller in Figure 8.19oo is usually selected to be nonlinear to allow the level to float in the surge tank without changing the FRC set point, which would upset the feed to the next column.

Therefore, the nonlinear controller is so configured that as long as the level in the surge tank is between 25 and 75%,

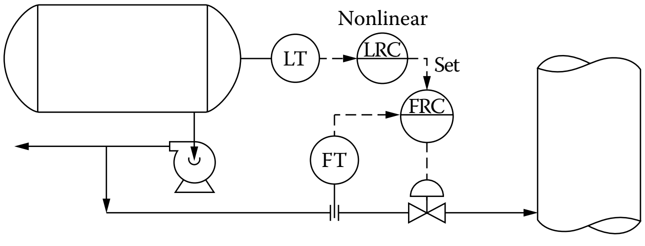


FIG. 8.19oo
Feed flow to the next column is kept relatively constant by the use of a nonlinear level controller (LRC) on the surge tank acting as the cascade master of the slave flow controller (FRC).

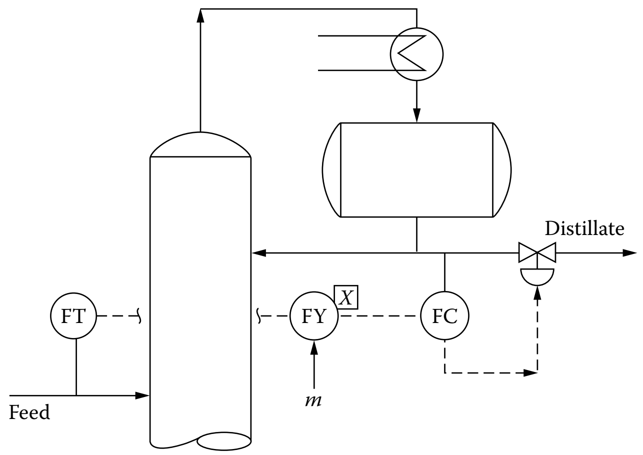


FIG. 8.19pp
Feedforward control minimizes feed rate disturbances.

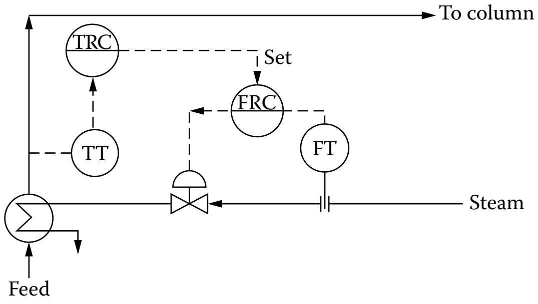


FIG. 8.19qq
The use of a temperature-flow cascade loop improves the column feed temperature controls provided by a preheater.

the set point to the FRC remains constant. This will allow the surge tank to fulfill its purpose and smooth out the load variations between the related processes. If the level drops below 25% or rises above 75%, the FRC set point is reduced or increased respectively to protect it from draining or flooding the tank.

If feed rate disturbances must be accepted by the column, a feedforward control system as shown in Figure 8.19pp can be used to minimize the impact of these disturbances.[10] The ratio, *m*, is selected by performance of a simple material balance around the column. Changing the product flow in proportion to the feed flow minimizes internal column transients and, thus, the quantity of off-spec material during recovery.

The value of *m*, however, is accurate only for one feed composition and will have to be readjusted either manually or automatically for different feed compositions. Dynamic compensation, which will be discussed in more detail in Section 8.21, is also recommended here.

Another method of minimizing feed rate disturbances is to use adaptive tuning or other nonlinear level control techniques (see Section 2.36 in Chapter 2) on the level controller. The key is to allow the accumulator, or the tower kettle, to utilize its capacity to accommodate transient material balance accumulations and act as a "surge drum" to minimize feed flow changes to the next unit.

Feed Temperature Control

The thermal condition of the feed determines how much additional heat must be added to the column by the reboiler. For efficient separation, it usually is desirable to have the feed preheated to its bubble point when it enters the column. Unless the feed comes directly from some preceding distillation step, an outside source of heat is required to achieve that.

Steam may be used to heat the feed, and a thermocouple inside a thermowell can detect the temperature inside the feed line. In this configuration, the temperature of the feed leaving this preheater controls the steam flow into the preheater. In such a cascade configuration, the temperature master is usually a three-mode controller. On start-up, the initially large correction provided by rate action of this three-mode controller helps to get the unit lined out faster. A full discussion of the advantages of cascade loops is provided in Section 2.6 in Chapter 2. Figure 8.19qq describes such a control system on a preheater application.

An alternate feed preheating configuration is to use an economizer on the feed stream. An economizer is a heat exchanger designed to take advantage of the waste heat to preheat the feed. Often, if the bottoms product is just sent to storage, it must be cooled anyway. Therefore, exchanging its heat content with the feed stream accomplishes both the objective of feed preheating and that of product cooling.

Two control configurations are common. If heat from the bottoms product stream is not sufficient, a second exchanger using steam is also installed to augment the heating of the feed, and on this second exchanger, a temperature control system like that previously shown in Figure 8.19qq is used. If the available bottoms product heat is more than sufficient, temperature control is achieved by manipulation of a bypass valve around the economizer, as shown in Figure 8.19rr.

Constant temperature feed does not necessarily mean constant feed quality. If feed composition varies, its bubble point also varies. It is common practice to set the temperature control at a point that is equivalent to the bubble point of the heaviest feed. As the feed becomes lighter, some of it will vaporize, but this variation can be handled by subsequent controls.

FEEDFORWARD CONTROLS

This section discusses the basic single-input–single-output feedback control loops, serving to control the product qualities, feed rate and temperature, and tower pressure. Such feedback systems are capable of compensating for deviations and disturbances only after they have occurred and have been

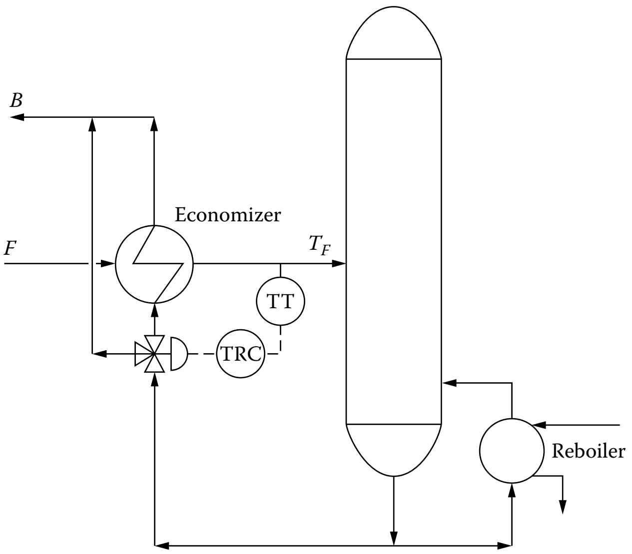


FIG. 8.19rr
The heat content of the bottoms products can be utilized to preheat the feed to the column in an economizer preheater.

detected. When using these simple control schemes, the operators are required to manually adjust the set point of these SISO loops in response to changing plant conditions as they occur. This approach is usually sufficient to keep the distillation column in operation, but it is not sufficient to achieve optimal performance.

As will be discussed in Section 8.21 in more detail, feedforward strategies attempt to compensate for process disturbances in the shortest time possible by accounting for process dynamics, dead times, time delays, and loop interactions.[1] The benefits of better control are:[11]

- Increased throughput
- Increased product recovery
- Energy conservation
- Reduced disturbances to other processing units
- Minimum rework or recycle of off-spec products
- Reduced operating personnel
- Increased plant flexibility

It has been reported that feedforward-based product composition control of distillation can give energy savings of 5–15%.[5]

While feedback-based basic distillation controls only aim at running the processing unit at current conditions, the objective of feedforward control is not only to do that, but also to account for conditions that can be anticipated. The challenge is to utilize the technique, the tools, and available resources to design unique feedforward control strategies that will match the specific objectives for the distillation columns. The choice between any of these control techniques depends upon factors such as preference and familiarity, complexity, degree of compensation, hardware for application, and number of variables monitored and controlled by a single strategy.

Often, additional instrumentation is not necessary when building upon basic feedback control designs to implement feedforward control. However, in many cases, if key measurements are not available and are needed for the feedforward calculation or compensation, the installation of new in-line sensors is also required.

Unlike basic feedback control, where much of the control could be implemented by simple analog control devices, feedforward control strategies generally require more sophisticated level computing systems. In this section, the common applications of feedback-based distillation column controls are discussed. Section 8.21 will discuss feedforward, model-based, and other advanced control systems, including optimization.

CONCLUSIONS

This section deals with some of the more basic control configurations for distillation towers. Section 8.20 describes the calculation of relative gains and Section 8.21 is devoted to the more advanced and optimized control strategies. The separation between these three areas is not very sharp, and some overlap does exist.

In this section, the control strategies for some of the more common distillation problems have been described. Although many other system configurations can exist, they usually are combinations of those presented. Control strategies today are no longer hardware dependent. Most modern microprocessor-based systems are designed with control function modules to execute a variety of the basic strategies that were discussed in this section. Multivariable unit operations controllers of both the model-predictive and the model-free variety are also slowly becoming available and will be discussed in Section 8.21.

It is important to emphasize that control by feedback methods alone cannot approach the quality of control possible by predictive (feedforward) techniques. This is true even though it is likely that the predictive control equations may need to be updated by feedback. In effect, predictive control tends to substantially reduce the size of the errors that are left to be handled by feedback. Further discussion of feedforward strategies and of other techniques for optimization are provided in more detail in Section 8.21.

References

1. Lockett, M. J., *Distillation Tray Fundamentals,* Cambridge, MA: Cambridge Press, 1986.
2. Strigle, R. J., Jr., *Random Packings and Packed Towers, Design, and Applications,* Houston, TX: Gulf Publishing Company, 1987.
3. Peters, M. S., and Timmerhaus, K. D., *Plant Design and Economics for Chemical Engineers,* 3rd ed., New York: McGraw-Hill Book Company, 1980.

4. Henley, E. J., and Seader, J. D., *Equilibrium-Stage Separation Operations in Chemical Engineering,* New York: John Wiley & Sons, Inc., 1981.
5. Smith, D. E., Stewart, W. S., and Griffin, D. E., "Distill with Composition Control," *Hydrocarbon Processing,* February 1978.
6. Fenske, M. R., "Fractionation of Straight-Run Pennsylvania Gasoline," *Industrial Engineering Chemistry,* May 1932.
7. Shinskey, F. G., *Process-Control Systems, Application, Design, Adjustment,* 3rd ed. New York: McGraw-Hill Book Company, 1988.
8. Bannon, R., et al., "Heat Recovery in Hydrocarbon Distillation," *Chemical Engineering Progress,* July 1978.
9. Converse, A. O. and Gross, G. D., "Optimal Distillate Rate Policy in Batch Distillation," *Industrial Engineering Chemistry,* August 1963.
10. Thurston, C. W., "Computer-Aided Design of Distillation Column Controls: Part 1," *Hydrocarbon Processing,* July 1981.
11. Jensen, B. A., and Collins, P. L., "Incentives For Tighter Fractionator Control," *Control,* November 1990.
12. Yang, D. R., Seborg, D. E., and Mellichamp, D. A., "Combined Balance Control Structure for Distillation Columns," *Industrial & Engineering Chemistry Research,* September 1991, pp. 2159–2168.
13. Gordon, L. M., "Practical Evaluation of Relative Gains: The Key to Designing Dual Composition Controls," *Hydrocarbon Processing,* December 1982.
14. Shinskey, F. G., "Predicting Distillation Column Response Using Relative Gains," *Hydrocarbon Processing,* May 1981.
15. Jensen, B. A., Harper, T., and Likins, M. R., "An Expert System in Fractionator Control System Design," *Proceedings of the AIChE Spring Meeting,* Orlando, FL, March 1990.
16. Shinskey, F. G., "An Expert System for the Design of Distillation Controls," *Proceedings of the Third International Conference on Chemical Process Control,* Asilomar, CA, January 1986.
17. Thurston, C. W., "Computer-Aided Design of Distillation Column Controls: Part 2," *Hydrocarbon Processing,* August 1981.
18. Griffin, D. E., "Fractionator Controls Can Save Energy," *Oil & Gas Journal,* March 1978.
19. Oglesby, M. W., and Hobbs, J. W., "Chromatograph Analyzers for Distillation Control," *Oil and Gas Journal,* January 10, 1966.
20. Bhullar, R. S., "Advanced Distillation Control," *Control,* May 1990.
21. Van Kampen, J. A., "Automatic Control by Chromatograph of a Distillation Column," Convention on Advances in Automatic Control, Nottingham, England, April 1961.
22. Luyben, W. L., "Distillation Decoupling," *AIChE Journal,* Vol. 16, No. 2, pp. 198–203, March 1970.
23. Shinskey, F. G., "The Stability of Interacting Loops with and without Decoupling," presented at the IFAC Symposium on Multivariable Control, Fredericton, New Brunswick, July 4–8, 1977.
24. Shinskey, F. G., *Distillation Control for Productivity and Energy Conservation,* 2nd ed., New York: McGraw-Hill Book Company, 1984.
25. Ryskamp, C. J., "New Control Strategy Improves Dual Composition Control," *Hydrocarbon Processing,* June 1980.

Bibliography

1960

Boyd, D. M., "Fractionator Pressure Control," *Petroleum Refiner.,* February 1947, 113–15.

Boyd, D. M., "Fractionator Instrumentation and Control Part 1," *Petroleum Refiner.,* October 1948, 115–536.

Hollander, L., "Pressure Control of Light-Ends Fractionators," *ISA Journal,* May 1957, 185–187.

Kirschbaum, E, *Distillation and Rectification,* Brooklyn, NY: Chemical Publishing Co., 1948.

Whistler, A. M., "Locate Condensers at Ground Level," *Petroleum Refiner.,* 33(3): 173–174, March 1954.

1960–1969

Denbigh, K., *The Principles of Chemical Equilibrium,* London: Cambridge University Press, 1961, pp. 63 and 116.

Van Kampen, J. A., "Automatic Control by Chromatograph of a Distillation Column," Convention on Advances in Automatic Control, Nottingham, England, April 1961.

Smuck, W. W., "Operating Characteristics of a Propylene Fractionating Unit," *Chemical Engineering Progress,* June 1963.

Lupfer, D. E. and Johnson, M. L., "Automatic Control of Distillation Columns to Achieve Optimum Operations," *ISA Transactions,* April 1964.

Rijnsdorp, J. E., "Interaction in Two-Variable Control Systems for Distillation Columns," in *Automatica*, Vol. 1, London: Pergamon Press, 1965.

Forman, E. R., "Control Systems for Distillation," *Chemical Engineering*, November 8, 1965.

Perry, E. S., and Weissberger, A., *Distillation*, 2nd ed., New York: Interscience, 1965.

Bristol, E. H., "On a New Measure of Interaction for Multivariable Process Control," *IEEE Transactions on Automatic Control,* January 1966.

Luyben, W. L., "10 Schemes to Control Distillation Columns with Sidestream Drawoff," *ISA Journal,* Vol. 13, No. 7, pp. 37–42, 1966.

Van Winkle, M., *Distillation,* New York: McGraw-Hill Book Company, 1967.

Wood, C. E., "Method of Selecting Distillation Column Tray Location for Temperature Control," presented at the Joint IMIQ-AIChE Meeting, Mexico City, September 24–27, 1967.

Owens, W. R., and Maddox, R. N., "Shortcut Absorber Calculations," *Industrial Engineering Chemistry,* December 1968.

Hengstebeck, R. J., "An Improved Shortcut for Calculating Difficult Multicomponent Distillations," *Chemical Engineering* (NY), January 13, 1969.

Buckley, P. S., "Controls for Sidestream Drawoff Columns," *Chemical Engineering Progress,* Vol. 65, No. 5, 1969, p. 45.

Luyben, W. L., "Feedback Control of Distillation Columns by Double Differential Temperature Control," *Industrial and Engineering Fundamentals,* Vol. 8, November 1969.

McNeill, G. A., and Sacks, J. D., "High-Performance Column Control," *Chemical Engineering Progress,* March 1969.

1970

Luyben, W. L., "Distillation Decoupling," *AIChE Journal,* Vol. 16, No. 2, pp. 198–203, March 1970.

Wahl, E. F., and Harriott, P., "Understanding and Prediction of the Dynamic Behavior of Distillation Columns," *I&EC Process Design and Development,* Vol. 9, No. 3, pp. 396–407, 1970.

Maarleveld, A., and J. E. Rijnsdrop, "Constraint Control of Distillation Columns," *Automatica,* 1970, pp. 51–58.

1971

Doig, I. D., "Variation of the Operating Pressure to Manipulate Distillation Processes," *Australian Chemical Engineering,* July 1971.

Hall, E. B., "Distillation Control Problems in the Chemical and Petroleum Industries," *Proceedings of the Twelfth Annual ISA Chemical and Petroleum Instrumentation Symposium,* Houston, TX, April 1971.

Holt, A. B., "Improved Fractionation thru Computer Control," *Instrumentation Technology,* January 1971, pp. 43–48.

King, C. J., *Separation Processes*, New York: McGraw-Hill Book Company, 1971.

Shinskey, F. G., "Avoiding Reset Windup in Cascade Systems," *Instrument and Control Systems,* August 1971.

Shinskey, F. G., “When to Use Valve Positioners,” *Instrument and Control Systems,* September 1971.

Shinskey, F. G., “When You Have the Wrong Valve Characteristic,” *Instrument and Control Systems,* October 1971.

Shunta, J. P., and Luyben, W. L., “Dynamic Effects of Temperature Control Tray Location in Distillation Columns,” *AIChE Journal,* Vol. 17, No. 1, pp. 92–96, January 1971.

1972

Toijala, K., and Fagervik, K., “A Digital Simulation Study of Two-Point Feedback Control of Distillation Columns,” *Kem. Teollisuus,* January 1972.

Shinskey, F. G., “What to Know about Interaction between Control Loops,” *Canadian Controls Instrumentation,* March 1972.

McFarlane, A. G. J., “A Survey of Some Recent Results in Linear Multivariable Feedback Theory,” *Automatica,* Vol. 8, No. 4, pp. 455–492, 1972.

Shinskey, F. G., “Effective Control for Automatic Start-up and Plant Protection,” *Canadian Controls Instrumentation,* April 1972.

Shinskey, F. G., “Controlling Surge-Tank Level,” *Instrument and Control Systems,* September 1972.

Shinskey, F. G., “Controlling Variable Period Once-Through Processes,” *Instrument and Control Systems,* November 1972.

White, M. H., “Surge Control for Centrifugal Compressors,” *Chemical Engineering* (NY), December 25, 1972.

1973

Evans, F. L., *Equipment Design Handbook for Refineries and Chemical Plants*, Vol. I, Houston, TX: Gulf Publishing Co., 1973, p. 89.

Perry, J. H., *Chemical Engineers’ Handbook*, 5th ed., New York: McGraw-Hill Book Company, 1973.

Houghton, J., and McLay, J. D., “Turboexpanders Aid Condensate Recovery,” *Oil and Gas Journal,* March 5, 1973.

Sayles, J. H., “Computer Control Maximizes Hydrocracker Throughput,” *Instrumentation Technology,* May 1973.

Frank, O., and Prickett, R. D., “Designing Vertical Thermosyphon Reboilers,” *Chemical Engineering* (NY), September 3, 1973.

Sisson, B., “How to Determine MEA Circulation Rates,” *Chemical Engineering* (NY), November 26, 1973.

O’Conner, J., and Illing, H., “The Turbine Control Valve. A New Approach to High-Loss Applications,” *Instrumentation Technology,* December 1973.

1974

Ackley, W. R., “Feedforward Control Strategy for Distillation Towers,” *Instrumentation Technology,* February 1974.

Adiutori, E. F., *The New Heat Transfer,* Cincinnati, OH: Ventuno Press, 1974, Chapter 7.

Kline, P. E., “Technical Task Force Approach to Energy Conservation,” *Chemical Engineering Progress,* February 1974.

Shinskey, F. G., “Values of Process Control,” *Oil and Gas Journal,* February 18, 1974.

Ellerbe, R. W., “Steam Distillation Basics,” *Chemical Engineering* (NY), March 4, 1974.

Lupter, D. E., “Distillation Column Control for Utility Economy,” presented at 53rd Annual GPA Convention, Denver, March 25–27, 1974.

Waller, K. V. T., “Decoupling in Distillation,” *AIChE Journal,* May 1974.

Farwell, W., “More Cooling at Less Cost,” *Plant Engineering,* August 1974.

Gallier, P. W., and McCune, L. C., “Simple Internal Reflux Control,” *Chemical Engineering Progress,* September 1974.

Rathore, R. N. S., VanWorme, K. A., and Powers, G. J., “Synthesis of Distillation Systems with Energy Integration,” *AIChE Journal,* September 1974.

Buckley, P. S., “Material Balance Control in Distillation Columns,” presented at AIChE Workshop on Industrial Process Control, Tampa, FL, November 11–13, 1974.

1975

Bojnowski, J. H., Crandall, J. W., and Hoffman, R. M., “Modernized Separation Systems Saves More than Energy,” *Chemical Engineering Progress,* October 1975.

Boyd, D. M., “Fractionation Column Control,” *Chemical Engineering,* June 1975.

Buckley, P. S., Cox, R. K., and Rollins, D. L., “Inverse Response in a Distillation Column,” *Chemical Engineering Progress,* June 1975.

Franzke, A., “Save Energy with Hydraulic Power Recovery Turbines,” *Hydrocarbon Process,* March 1975.

Luyben, W. L., “Steady-State Energy Conservation Aspects of Distillation Column Control Design,” *I&EC Fundam.,* Vol. 14, No. 4, pp. 321–325, 1975.

Monroe, E. S., “Vacuum Pumps Can Conserve Energy,” *Oil and Gas Journal,* February 3, 1975.

Rademaker, O., Rijnsdorp, J. E., and Maarleveld, A., *Dynamics and Controls of Continuous Distillation Units,* Amsterdam: Elsevier, 1975.

Shinskey, F. G., “Controlling Unstable Processes, Part II: A Heat Exchange,” *Instrument and Control Systems,* January 1975.

Sternlicht, B., “Low-Level Heat Recovery Takes on Added Meaning as Fuel Costs Justify Investment,” *Power,* April 1975.

Tyreus, B., and Luyben, W. L., “Control of a Binary Distillation Column with Sidestream Drawoff,” *I&EC Process Design and Development*, Vol. 14, No. 4, p. 391, 1975.

Tyreus, B., and Luyben, W. L., “Two Towers Cheaper than One?” *Hydrocarbon Processing,* Jun., 1975.

Wolf, C. W., Weiler, D. W., and Ragi, E. G., “Energy Costs Prompt Improved Distillation,” *Oil and Gas Journal,* September 1, 1975.

1976

ISA Standard S51.1, “Process Instrumentation Terminology,” Research Triangle Park, NC: Instrument Society of America, 1993.

Shinskey, F. G., “Energy-Conserving Control for Distillation Units,” *Chemical Engineering Progress,* May 1976.

Wright, R. M., “A Better Approach to Distillation Control,” *Instruments and Control Systems,* June 1976.

Van Horn, L. D., and Latour, P. R., “Computer Control of a Crude Still,” *Instrumentation Technology,* November 1976, pp. 33–39.

VanWinkle, M., *Distillation*, New York: McGraw-Hill, 1976.

1977

Hadley, K., “Control Objectives Analysis,” National Petroleum Refiners Association Computer Conference, New Orleans, 1977.

Hobbs, J. W., “Unifying Concepts for Fractionator Product Composition Control,” *Proceedings of the Automatic Control Conference,* San Francisco, 1977.

Evans, L. B., “Impact of the Electronics Revolution on Industrial Process Control,” *Science,* Vol. 195, No. 4283, March 18, 1977.

Hughart, C. L., “Designing Distillation Units for Controllability,” *Instrumentation Technology,* May 1977.

Lieberman, N., “Instrumenting a Plant to Run Smoothly,” *Chemical Engineering,* September 12, 1977, pp. 140–154.

Shinskey, F. G., “The Stability of Interacting Loops with and without Decoupling,” presented at the IFAC Symposium on Multivariable Control, Fredericton, New Brunswick, July 4–8, 1977.

1978

Buckley, P. S., Cox, R. K., and Luyben, W. L., "How to Use a Small Calculator in Distillation Column Design," *Chemical Engineering Progress,* June 1978, pp. 49–55.

Buckley, P. S., "Distillation Column Design Using Multivariable Control," *Instrumentation Technology,* September/October 1978.

Buford, B. N., Bush, B. A., and Staten, H. W., "Computers Conserve Energy in NGL Fractionators," *Oil & Gas Journal,* December 1978.

Doukas, N., "Control of Sidestream Columns Separating Ternary Mixtures," *Instrumentation Technology,* June 1978.

Doukas, N. and Luyben, W. L., "Economics of Alternative Distillation Configurations for the Separation of Ternary Mixtures," *I&CE Process Design and Development,* Vol. 17, No. 272, 1978.

Griffin, D. E., Parsons, J. R., and Smith, D. E., "The Use of Process Analyzers for Composition Control of Fractionators," *Proceedings of the ISA Spring Joint Conference,* Houston, TX, May 1978.

Latour, P. R., "Composition Control of Distillation Columns," *Instrumentation Technology,* July 1978.

Lamb, M. Y., "Computer Control of a Propylene Upgrading Unit," 85th National AIChE Meeting, Philadelphia, PA, 1978.

McCoy, R. D., "Adding Capabilities to Process Chromatography with Microprocessor-Based Programmers," *Proceedings of the ISA Joint Spring Conference,* Houston, TX, May 1978.

Mix, T. J., Dweck, J. S., Weinberg, M., and Armstrong, R. C., "Energy Conservation in Distillation," *Chemical Engineering Progress,* April 1978.

Painter, J. W., and Gonnella, J. L., "Improved Control of a Distillation Column Using a Minicomputer and an On-Line Gas Chromatograph," Texas A&M Instrumentation Symposium, January 1978.

Tolliver, T. L., and McCune, L. C., "Distillation Column Control Design Based on Steady State Simulation," *ISA Transactions,* Vol. 17, No. 3, pp. 3–10, 1978.

1979

Watkins, R. N., *Petroleum Refinery Distillation*, 2nd ed., Houston, TX: Gulf Publishing Company, 1979.

Douglas, J. M., Jafarey, A., and McAvoy, T. J., "Shortcut Techniques for Distillation Column Design and Control, Part 1: Column Design," *I&EC Process Design and Development,* Vol. 18, pp. 197–202, April 1979.

Van Horn, L. D., "Computer Control. How to Get Started," *Hydrocarbon Processing,* September 1979.

Chin, T. G., "Guide to Distillation Pressure Control Methods," *Hydrocarbon Processing,* October 1979, 145–153.

1980

Treybal, R. E., *Mass-Transfer Operations,* 3rd ed., New York: McGraw-Hill Book Company, 1980.

Van Horn, L. D., "Crude Unit Computer Control. How Good Is It?" *Hydrocarbon Processing,* April 1980.

Ryskamp, C. J., "New Control Strategy Improves Dual Composition Control," *Hydrocarbon Processing,* June 1980.

Soderstrom, E. D., "Computer Control for Energy Savings," *Chemical Engineering Progress,* August 1980, pp. 60–62.

Black, C., "Distillation Modeling of Ethanol Recovery and Dehydration Process for Ethanol and Gasohol," *Chemical Engineering Progress,* September 1980.

Andrews, A. J., and Griffin, D. E., "Performance Audits Evaluate Two Distillation Control Projects," *Oil & Gas Journal,* December 8, 1980.

1981

Griffin, D. E., "Tighten Distillation Column Control and Save Energy," *Instruments and Control Systems,* March 1981.

Black, J. W., "Model Estimates Analyzer Payouts," *Hydrocarbon Processing,* September 1981.

Shinskey, F. G., "Controlling Distillation Processes for Fuel-Grade Alcohol," *Instruments and Control Systems,* December 1981.

Thurston, C. W., "Computer-Aided Design of Distillation Column Controls," *Hydrocarbon Processing,* Part 2, August 1981, p. 5.

1982

Rinne, R., Sunnel, H., Latour, P. R., and Payntex, K. K., "Experience with Distillation Unit Computer Control," *Hydrocarbon Processing,* March 1982.

Roffel, B., and Rijnsdorp, J. E., *Process Dynamics, Control, and Protection,* Ann Arbor, MI: Ann Arbor Science, 1982.

DiBano, R. J., "Advanced Control: General Purpose Design vs. Working It Out in the Field," *Hydrocarbon Processing,* August 1982.

Dweck, J. S. and Mix, T. W., *Conserving Energy in Distillation*, MIT Press, 1982.

1984

Mix, P. E., *The Design and Application of Process Analyzer Systems,* New York: John Wiley & Sons, Inc., 1984.

Stephanopoulos, G., *Chemical Process Control, An Introduction to Theory and Practice,* Englewood Cliffs, NJ: Prentice Hall, 1984.

Barduhn, A. J., "Setting the Pressure at which to Conduct a Distillation," *Chemical Engineering Education,* Vol. 18, No. 1, pp. 38–40, 1984.

1985

Deshpande, P. B., *Distillation Dynamics and Control,* Research Triangle Park, NC: Instrument Society of America, 1985.

Buckley, P. S., Luyben, W. L., and Shunta, J. P., *Design of Distillation Column Control System,* Research Triangle Park, NC: Instrument Society of America, 1985.

Gani, R., Romagnoli, J. A., and Stephanopoulos, G., "Control Studies in an Extractive Distillation Process: Simulation and Measurement Structure," *Chem. Engng. Commun.,* 40, pp. 281–302, 1985.

1986

Tsai, T. H., Lane, J. W., and Lin, C. S., *Modern Control Techniques for the Process Industries*, Vol. 23, Marcel Dekker, 1986.

Hansen, T. T., and Jørgensen, S. B., "Optimal Open Loop Control of Binary Batch Distillation," *Chem. Eng. J.,* 33 , pp. 151–155, 1986.

1987

Christensen, F. M., and Jørgensen, S.B., "Optimal Open Loop Control of Binary Batch Distillation with Recycled Waste Cut," *Chem. Eng. J.,* 34, pp. 57–64, 1987.

1988

Balchen, J. G., *Process Control, Structures, and Applications*, New York: Van Nostrand Reinhold, 1988.

Birky, G. J., McAvoy, T. J., and Tyreus, B. D., "Expert System for Design of Distillation Controls," *Proceedings of the ISA/88 International Conference and Exhibit,* Houston, TX, October 1988.

Murrill, P. W., *Application Concepts of Process Control Theory,* Research Triangle Park, NC: Instrument Society of America, 1988.

Nichols, G. D., *On-Line Process Analyzers,* New York: John Wiley & Sons, Inc., 1988.

Roat, S. D., Moore, C. F., and Downs, J. J., "A Steady-State Distillation Column Control System Sensitivity Analysis Technique," *Proceedings IEEE Southeast Con.,* 1988, pp. 296–300.

1989

Finco, M. V., Luyben, W. L., and Polleck, R. G., "Control of Distillation Columns with Low Relative Volatilities," *Ind. Eng. Chem. Res.*, Vol. 28, January 1989, pp. 75–83.

Wassick, J. M. and Tummala, R. L., "Multivariable Internal Model Control for a Full-Scale Industrial Distillation Column," *Control Systems Magazine,* January 1989, pp. 91–96.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* New York: John Wiley & Sons, Inc., 1989.

Hall, G. F., "Improve Process Performance by Using Chromatograph-Directed Control," *Proceedings of the ISA/89 International Conference and Exhibit,* Part 1, Philadelphia, PA, October 1989.

Christie, D. A., "The Top-Down Approach to Successful Process Control Projects," *Control*, October 1989.

Luyben, W. L., *Process Modeling, Simulation, and Control for Chemical Engineers,* 2nd ed., New York: McGraw-Hill Book Company, 1989.

Kister, H. Z., *Distillation Operation,* New York: McGraw-Hill Publishing Company, 1989.

Riggs, J. B., Sinha, R., "High Purity Distillation Control Using Nonlinear Process Model-Based Control," ISA89, Philadelphia, PA, October 1989.

Riggs, J. B., "Nonlinear Process Model-Based Control of a Distillation Column with a Sidestream Draw-Off," presented at the annual AIChE meeting, San Francisco, CA, November 1989.

Riggs, J. B., Sinha, R., and McDaniel, R., "Comparison of Control Techniques for High Purity Distillation Columns," AIChE Spring National Meeting, Houston, TX, April 1989.

1990

Papastathopoulou, H. S., and Luyben, W. L., "Turning Controllers on Distillation Columns with the Distilled Bottoms Structure," *Industrial and Engineering Chemistry Research*, Vol. 29, September 1990, pp. 1859–1868.

Chien, I.-L., and Fruehauf P. S., "Consider IMC Tuning to Improve Controller Performance," *Chemical Engineering Progress,* Vol. 86, No. 10, pp. 33–41, October 1990.

Cingara, A., Jovanovic, M., and Mitrovic, M., "Analytical First-Order Dynamic Model of Binary Distillation Column," *Chemical Engineering Science,* Vol. 45, No. 12, pp. 3585–3592, 1990.

Kister, H. Z., *Distillation Operation,* New York: McGraw-Hill Book Company, 1990.

Li, R., Olson, J. H., and Chester, D. L., "Dynamic Fault Detection and Diagnosis Using Neural Networks," *Proceedings of the Fifth IEEE International Symposium on Intelligent Control,* Philadelphia, PA, September 1990, pp. 1169–1174.

McGreavy, C., *Dynamics and Control of Chemical Reactors, Distillation Columns, and Batch Processes*, Pergamon Press, 1990.

Papastathopoulou, H. S. and Luyben, W. L., "Potential Pitfalls in Ratio Control Schemes," *Industrial & Engineering Chemistry Research,* October 1990, pp. 2044–2053.

Pitt, M. J., *Instrumentation and Automation in Process Control*, New York: Horwood, 1990.

Skogestad, S., Lundstrom, P., and Jacobsen, E. W., "Selecting the Best Distillation Control Configuration," *AIChE Journal,* Vol. 36, pp. 753–764, 1990.

Skogestad, S., Jacobsen, E. W., and Morari, M., "Inadequacy of Steady-State Analysis for Feedback Control. Distillate. Bottom Control of Distillation Columns," *Industrial & Engineering Chemistry Research*, December 1990, pp. 2339–2346.

Pandit, H. G., and Rhinehart, R. R., "Process Model-Based Control of a Nonideal Binary Distillation Column," *Proceedings of the Annual AIChE Meeting,* Chicago, IL, November 1990.

Pedersen, N. H., and Jørgensen, S. B., "A GC Subsystem for Fast On-Line Concentration Profile Measurement for Advanced Distillation Control," *Analytica Chemica Acta,* 238, pp. 139–148, 1990.

Riggs, J. B., Watts, J., and Beauford, M., "Advanced Model-Based Control for Distillation," National Petroleum Refinery Association Meeting, Seattle, WA, October 1990.

Riggs, J. B., Watts, J., and Beauford, M., "Industrial Experience with Applying Nonlinear Process Model-Based Control to Distillation Columns," ISA90, New Orleans, LA, October 1990.

Riggs, J. B., "Advanced Model-Based Control of a Sidestream Draw Column," at ISA 90, New Orleans, LA, October 1990.

1991

Coughanowr, D. R., *Process Systems Analysis and Control,* 2nd ed., New York: McGraw-Hill Book Company, 1991.

Jensen, B. A., "Improve Control of Cryogenic Gas Plants," *Hydrocarbon Processing,* May 1991.

Papastathopoulou, H. S., and Luyben, W. L., "Control of a Binary Sidestream Distillation Column," *Industrial & Engineering Chemistry Research,* April 1991, pp. 705–713.

Rijnsdorp, J. E., *Integrated Process Control and Automation,* Amsterdam: Elsevier, 1991.

Sandelin, P. M., Haeggblom, K. E., and Waller, K. V., "Disturbance Rejection Properties of Control Structures at One-Point Control of a Two-Product Distillation Column," *Industrial & Engineering Chemistry Research,* June 1991, pp. 1182–1186.

Sandelin, P. M., Haeggblom, K. E., and Waller, K. V., "Disturbance Sensitivity Parameter and its Application to Distillation Control," *Industrial & Engineering Chemistry Research,* June 1991, pp. 1187–1193.

1992

Fruehauf, P. S., and Mahoney, D. P. "Distillation Column Control Design using Steady-State Models, Usefulness and Limitations," in *Advances in Instrumentation and Control,* Vol. 47, Part 1, Research Triangle Park, NC: Instrument Society of America, 1992, pp. 92–120.

Skogestad, S., "Dynamics and Control of Distillation Columns: A Critical Survey," Preprints IFAC Symposium, DYCORD +92, College Park, MD, pp. 1–25.

Luyben, W. L., *Practical Distillation Control,*" New York: Van Nostrand Reinhold, October 1992.

Koggersbøl, A., and Jørgensen, S. B., "Dynamics and Control of a Distillation Column with a Sidestream," IChemE Symposium Series, 128, pp. A429–A449, 1992.

1993

Fruehauf, P. S., and Mahoney, D. P., "Distillation Column Control Design Using Steady State Models: Usefulness and Limitations," *ISA Transactions*, Research Triangle Park, NC. 1993.

Balchen, J. G., *Dynamics and Control of Chemical Reactors, Distillation Columns, and Batch Processes,* Pergamon Press, April 1993.

Ganguly, S., "Model Predictive Control of Distillation," ISA/93 Technical Conference, Chicago, IL, September 1993.

1994

Gokhale, V., Shukla, N., and Munsif, H., "Analysis of Advanced Distillation Control on a C_3 Splitter and a Depropanizer," 1994 AIChE National Annual Meeting, San Francisco, CA, November 1994.

Fruehauf, P. S., and Mahoney, D. P., "Improve Distillation Column Control Design," *Chemical Engineering Progress*, March 1994.

1995

Fleming, B., and Sloley, A. W., "Feeding and Drawing Products: The Forgotten Part of Distillation," *Proceedings of the ChemShow and Exposition,* New York, December 1995.

Musch, H. E., and Steiner, M., "Robust PID Control for an Industrial Distillation Column," *Control System Magazine,* Vol. 15, No. 4, 46–55, 1995.

Hurowitz, S. E., and Gokhale, V., "A Dynamic Model of a Superfractionator: A Test Case for Comparing Distillation Control Techniques," DYCORD '95, 4th IFAC Symposium, Helsingor, Denmark, June 1995.

Lundstrom, P., and Skogestad, S., "Opportunities and Difficulties with 5 × 5 Distillation Control," *J. Process Control,* Vol. 5, 249–261, 1995.

Rawlings, J. B., "Dynamics and Control of Chemical Reactors, Distillation Columns, and Batch Processes (Dycord+'95)," a Postprint Volume from the 4th IFAC Symposium on Dynamics and Control of Chemical Reactors, Distillation Columns and Batch Processes (DYCORD '95), Helsingor, Denmark, June 1995.

Diwekar, U. M., "Batch Distillation: Simulation, Optimal Design, and Control (Series in Chemical and Mechanical Engineering)," Taylor & Francis, September 1995.

Banerjee, A., and Arkun, Y., "Control Configuration Design Applied to the Tennessee Eastman Plantwide Control Problem," *Computers & Chem. Eng.,* 19(4), 453–480, 1995.

1996

Koggersbøl, A., Andersen, R., Nielsen, J. S., Jørgensen, S., "Control Configuration for Energy Integrated Distillation," *Computers & Chem. Eng*., 20 (supplement), pp. S853–S858, 1996.

1997

Linsley, J., "New, Simpler Equations Calculate Pressure-Compensated Temperatures," *Oil & Gas Journal,* May 24, 1997, 58–64.

Ming T. Tham, "Distillation," Base Document URL: http://lorien.ncl.ac.uk/ming/distil/distil0.htm, October 1997.

Anderson, N. A., *Instrumentation for Process Measurement and Control,* 3rd edition, Boca Raton, FL: CRC Press, October 1997

Hurowitz, S. E. and Anderson, J. J., "Distillation Configuration Selection for Dual Composition Control," AIChE Spring National Meeting, Houston, TX, April 1997.

Hurowitz, S. E. and Anderson, J. J., "Control of High Purity Distillation Columns," Control 97 Conference, Sydney, Australia, October 1997.

Mahoney, D. P. and Fruehauf, P. S., "An Integrated Approach for Distillation Column Control Design Using Steady-State and Dynamic Simulation," Aspentech technical articles, March 1997. www.aspeutech.com/corporate/press/publications.

Skogestad, S., "Dynamics and Control of Distillation Columns: A Tutorial Introduction," *Trans. IChemE.,* Vol. 75, Part A, pp. 539–562, 1997.

1998

Betlem, B. H. L., Krijnsen, H. C., and Huijnen, H., "Optimal Batch Distillation Control Band on Specific Measures," *Chemical Engineering Journal,* 71, pp. 111–126, 1998.

Riggs, J. R., "Improve Distillation Column Control," *Chemical Engineering Progress,* October 1998, 31–47.

Stichlmair, J. G. and Fair, J. R., *Distillation: Principles and Practices*, New York: John Wiley & Sons, 1998.

1999

Eker, I. and Sakthivel, K., "Automation & Lube Oil Additives Blending Plant Using an S88.01 Consistent Batch Software: A Case Study," *Proceedings of the World Batch Forum*, San Diego, CA, April 1999.

Hurowitz, S., Anderson, J., Duvall, M., and Riggs, J.B., "An Analysis of Controllability Statistics for Distillation Configuration Selection," presented at the AIChE Annual Meeting, Dallas, TX, November 1999.

2000

Andrew W. Sloley, "Steady Under Pressure: Distillation Pressure Control," presented at the American Institute of Chemical Engineers Spring Meeting, March 6–9, 2000.

Betlem, B.H.L. "Batch Distillation Column Low-Order Models for Quality Control Program," *Chemical Engineering Science,* 55, pp. 3187–3194, 2000.

Roffel, B., Betlem, B. H. L., and De Ruijter, J. A., "Modeling and Control of a Cryogenic Distillation Column," *Computers and Chemical Engineering,* 24, pp. 111–123, 2000.

Roffel, B., "Distillation: Instrumentation and Control Systems," in *Encyclopedia of Separation Science*, London: Academic Press, 2000.

Willis, M. J., "Selecting a Distillation Column Control Strategy (a basic guide)," 2000. http://lorien.ud.cc.uk/ming/control/g

2002

Florez, M., "Batch Distillation: Practical Aspects of Design and Control," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April, 2002

Cook, B., Engel, M., Landis, C., Tedeschi, S., and Zehnder, A., "Synthesis of Optimal Batch Distillation Sequences," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2002.

2003

Kralj, F., "Application of the S88 Model in the Control of Continuous Distillation Facilities," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

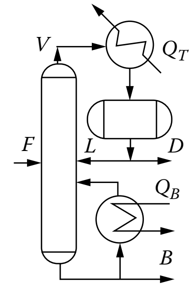
Jones, M., and Kilian, A., "Tricky Pressure Control in Distillation Column," July, 2003. http://instrumentation.co.za/regulator

2004

Hurowitz, S., Anderson, J., Duvall, M., and Riggs, J. B., "Distillation Control Configuration Selection," submitted to *J. Process Control,* March 2004.

8.20 Distillation: Calculations of Relative Gains

L. A. KANE (1985) **S. OCHIAI** (1995) **S. OCHIAI, F. G. SHINSKEY** (2005)



Flow sheet symbol

INTRODUCTION

A distillation column has anywhere from four to six control loops, which need to be optimally configured if the column is to be stable and meet its performance objectives. Four single loops can be configured in 24 different ways, five single loops in 120 different ways, and six loops in 720 different ways. Some of these configurations can be ruled out on the basis of impossible dynamic response—for example, one would not attempt to control column base level by manipulating the flow of distillate product.

But even after eliminating those from consideration, many possibilities remain. Furthermore, a simple configuration of single loops may not be as effective in controlling product compositions as one that manipulates a logical combination of variables, such as reflux-to-distillate ratio. Logical mathematical combinations of manipulated variables expand the possible number of configurations manyfold.

It has also been observed that a configuration that successfully controls one column will not necessarily be successful on another, requiring each column to be considered on its own merits. Because of the complexity of the configuration problem, and the unique requirements of each individual column, a methodology is required for matching the configuration to the column. That methodology is relative gain analysis (RGA). The basics of RGA are described in detail in Section 2.25. Here, RGA is specially adapted to the particular multivariable control problem represented by distillation columns.

In addition, from the numerical values of relative gains, the open-loop gains are readily calculated. The open-loop gains can help control engineers to conduct plant tests to collect transient responses of distillation columns that are required to implement so-called advanced process control (APC).

RGA FOR DISTILLATION

Each product leaving a distillation column may have the concentration of one of its components controlled. Most columns have two products—distillate and bottom—but many also have a sidestream as well, and the composition loops tend to interact with each other. There are usually three other loops on a column: liquid levels in the base and in the reflux accumulator, and column pressure. (A column operating with a flooded condenser may also have a flooded accumulator, thereby eliminating one level loop.)

All of the economic value in operating a column, however, resides in the composition of its products. There is no inherent value in controlling liquid level or pressure more precisely; therefore, their control should be considered secondary to controlling product compositions. In fact, the configuration of the level loops especially should be based on what will provide the best composition control, and not the best level control. They must function, obviously, or product compositions cannot be controlled, but they must be subservient to the composition loops.

In this light, level and pressure loops should be arbitrarily assigned in such a way as to minimize the interaction among the composition loops. They are faster than the composition loops by an order of magnitude or more, and so cannot be upset by the composition controllers. In the definition of relative gain, the open-loop gain of a pair of manipulated and controlled variables with all other loops open is divided by what it would be with all loops closed. But in a distillation column, the level and pressure loops are always closed, so considering what the compositions would do with level and pressure loops open has no significance.

Two-by-Two Subsets

The interaction between the composition loops is of utmost importance because they are the slowest loops on the column and similar in dynamic response, and have the most economic significance. In a two-product column, the interaction exists between the key impurities in the distillate and bottom products. The controllable key component in the distillate is the heavy key, and in the bottom it is the light key, the column separating the light from the heavy key.

Petroleum fractionators are named after their light key, such as depropanizer, debutanizer, or deisobutanizer.

Each sidestream withdrawn from a column adds a manipulated variable — its flow — and therefore a controlled variable—another composition. However, manipulation of the smallest of the product streams in a single-sidestream column can be arbitrarily assigned to one of the composition loops in a logical procedure described later. This leaves the remaining two composition loops to be configured on the basis of relative gain.

The advantage of limiting consideration to two interacting loops is that only one relative gain needs to be calculated. In Section 2.25, it was explained that relative gain arrays consist of pure numbers whose rows and columns add up to 1.0. A two-by-two array then consists of diagonal elements having values of λ and off-diagonal elements having values of $1 - \lambda$:

$$\Lambda_{12} = \begin{array}{ccc} & m_1 & m_2 \\ y & \lambda_{y1} & 1-\lambda_{y1} \\ x & 1-\lambda_{y1} & \lambda_{y1} \end{array} \qquad \textbf{8.20(1)}$$

For a distillation column, y is the distillate-composition variable, x is the bottom-composition variable, m_1 and m_2 are the manipulated variables selected to control them, and Λ_{12} identifies the two-by-two array that applies those manipulated variables to composition control. Because $\lambda_{x2} = \lambda_{y1}$, the latter alone is enough to identify the entire array, and will be used as such throughout this section, in the form $\lambda_{y1}(\Lambda_{12})$. In practice, λ_{y1} can have different values depending on the choice of the second manipulated variable—for example, $\lambda_{y1}(\Lambda_{12})$ will not equal $\lambda_{y1}(\Lambda_{13})$.

Choice of Manipulated Variables

Figure 8.20a shows a two-product distillation column, identifying all streams and controlled variables. The five controlled variables are the levels in the reflux accumulator and column base, pressure, and the composition of distillate and bottom streams (often sensed as tray temperatures).

The five manipulated variables are distillate flow D, bottom flow B, reflux flow L, vapor boil-up V (as heat input Q_B), and heat removal Q_T. (Feed rate F is an independent variable.) Almost without exception, column pressure is controlled by heat removal to close the heat balance, as condenser cooling tends to be quite variable and difficult to measure. This leaves D, B, L, and V as flows that can be manipulated for composition control, although D and B are not independent of each other, for the overall material balance requires that

$$F = D + B \qquad \textbf{8.20(2)}$$

Two flow ratios are also possibilities as manipulated variables: reflux ratio L/D and boil-up ratio V/B. (Another possibility exists in L/B, but because these streams are at opposite ends of the column, their coordination presents serious dynamic problems, and so this ratio is not used.) A material balance across the reflux accumulator gives

$$V = L + D \qquad \textbf{8.20(3)}$$

which renders the ratios D/V and L/V dependent on L/D.

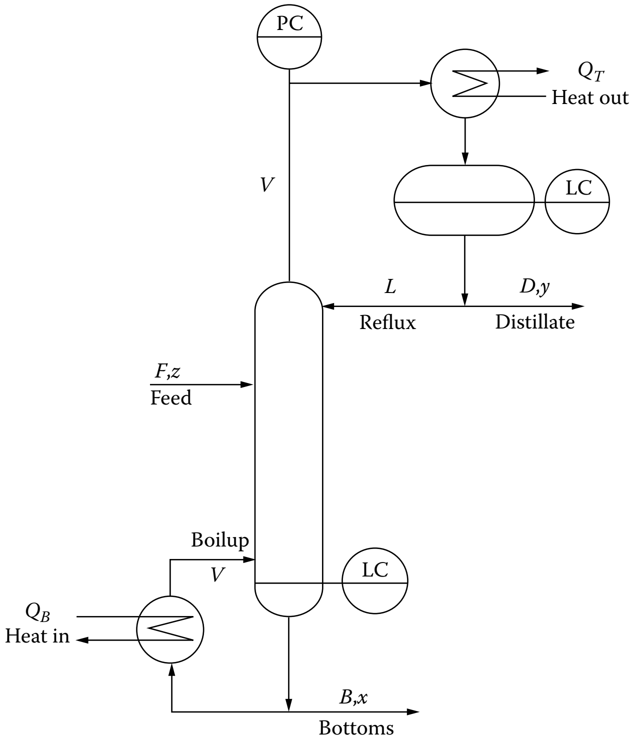


FIG. 8.20a
A typical distillation column has five manipulated and five controlled variables.

So, the independent manipulated variables that can be used for controlling product compositions are D or B, L, V, L/D, and V/B. Only those variables located at the top of the column can be manipulated successfully to control top composition y, and are listed as choices for m_y; those located at the bottom can be manipulated successfully to control bottom composition x, and are listed as choices for m_x. The choices can be displayed in a table such as Table 8.20b.

Each of the Λ terms in the table represents a two-by-two relative gain array. Note that there are only eight, because that in the upper left corner representing $\Lambda_{D/F,B/F}$ would be an array of ±∞, flow rates D and B being dependent on each other. So, there are basically eight different ways to control a two-product column, and therefore eight different two-by-two relative gain arrays.

Finally, each two-by-two array can be represented by a single relative gain number, so each space in the table needs only one number (rather than four) to identify its array. The

TABLE 8.20b
Table of Independent Relative Gain Arrays for Two-Product Distillation

m_y/m_x	D	L	L/D
B	—	Λ_{LB}	$\Lambda_{L/D,B}$
V	Λ_{DV}	Λ_{LV}	$\Lambda_{L/D,V}$
V/B	$\Lambda_{D,V/B}$	$\Lambda_{L,V/B}$	$\Lambda_{L/D,V/B}$

complementary numbers in the off-diagonal elements of each two-by-two array are not useful. For example, in the Λ_{LV} array, the diagonal elements λ_{yL} and λ_{xV} are useful pairs, whereas λ_{yV} and λ_{xL} will have negative values, and their loops are not dynamically viable.

COLUMN MODEL

The method of partial derivatives described in Section 2.25 under Equation 2.25(13) will be used to extract the relative gains. But this requires a differentiable column model. For this purpose, the column model described in Reference 1 is used, as expanded for multicomponent mixtures in Reference 2.

Material Balance

The model begins with the overall material balance of Equation 8.20(2) and continues with a component balance:

$$Fz_i = Dy_i + Bx_i \qquad 8.20(4)$$

where flow rates are in mols/hr, z represents the composition of the feed with all compositions in mol fraction, and subscript i refers to any particular component. When these two equations are combined, we have a single relationship between the two product compositions and the fractional product flow rates:

$$\frac{z_i - x_i}{y_i - x_i} = \frac{D}{F} = 1 - \frac{B}{F} \qquad 8.20(5)$$

Separation

A second relationship is required to reach a solution, and it is derived from the Fenske equation,[3] used for estimating the minimum number of trays required to separate a mixture at total reflux:

$$\frac{y_i/y_j}{x_i/x_j} = \alpha_{ij}^{nE} \qquad 8.20(6)$$

where i and j are any two components in a volatile mixture and α_{ij} is their relative volatility, n is the number of trays and E their fractional efficiency, so that nE is the number of theoretical trays. This relationship can be extended to operating conditions other than total reflux, and applied to the separation of the light l and heavy h key components:

$$\frac{y_l/y_h}{x_l/x_h} = S \qquad 8.20(7)$$

In Reference 1, the separation factor S is related to α^{nE}, and also to the reflux ratio L/D and the light components z in the feed:

$$S = \left(\frac{\alpha}{\sqrt{1 + D/Lz}} \right)^{nE} \qquad 8.20(8)$$

where α is the average relative volatility between the key components across the column.

This model is tractable for a feed stream having four components: the light and heavy keys, a lighter key ll, and a heavier key hh. The lighter is defined as exiting entirely with the distillate and the heavier as exiting entirely with the bottom product; each product stream then contains three components. In the expression above, $z = z_l + z_{ll}$.

To find the product compositions for any given set of feed conditions, distillate, and reflux flow rates, the following quadratic equation must be solved:

$$y_h = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad 8.20(9)$$

where
$$a = (S-1)D/F$$
$$b = Sz_l - (S-1)(D/F - z_{ll}) + z_h$$
$$c = -z_h(1 - Fz_{ll}/D)$$

After solving for y_h, then

$$x_h = \frac{z_h - y_h D/F}{1 - D/F} \qquad 8.20(10)$$

After the heavy-key components in the products are determined, the others are readily calculated from the material balance and separation factor:

$$y_{ll} = \frac{z_{ll}}{D/F} \qquad 8.20(11)$$

$$y_l = 1 - y_h - y_{ll} \qquad 8.20(12)$$

$$x_l = \frac{x_h y_l}{S y_h} \qquad 8.20(13)$$

$$x_{hh} = 1 - x_h - x_l \qquad 8.20(14)$$

Operating Curves

We are primarily interested in controlling the key impurities, which are the heavy key in the distillate y_h and the light key in the bottom x_l. It is revealing to plot these two controlled variables against each other while holding selected manipulated variables constant. A set of such operating curves is shown in Figure 8.20c for a column separating a feed consisting of 1.2% lighter-than-light key, 38.0% light key, 47.0% heavy key, and 13.8% heavier-than-heavy key in 24 theoretical trays. The distillate is to be controlled at 1.5% heavy key and the bottoms at 0.3% light key; the reflux ratio (L/D) is 2.6 and α is 2.11.

The focal point of the curves (where they all cross) is the given set of controlled key impurities. Holding any one of the manipulated variables constant and changing one of the others will move both compositions along the path representing the variable held constant.

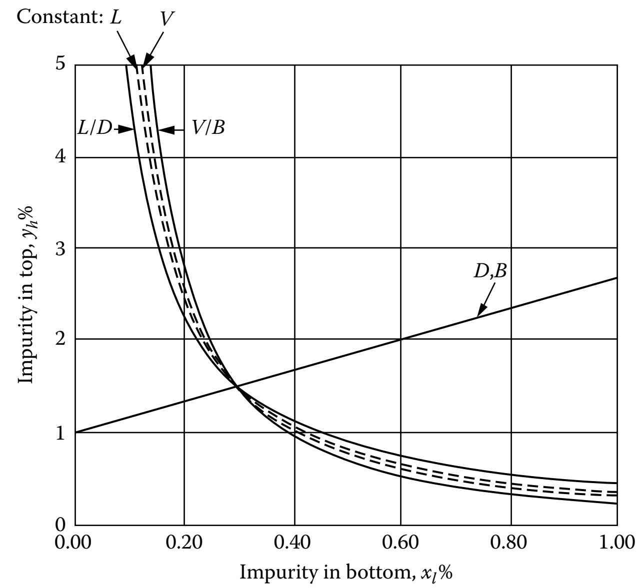


FIG. 8.20c
Each curve shows the relationship between top and bottom composition when that particular manipulated variable is held constant.

For example, if distillate (and hence bottom) flow is held constant and either reflux or boil-up is increased, both product impurities will decrease, moving along the straight line to the left. If reflux were held constant and distillate flow increased, compositions would move upward to the left along the curve represented by constant L, y_h increasing sharply and x_l decreasing to a smaller extent. (Recognize that feed rate is considered constant in this plot, or that L, V, D, and B are held in constant ratio to F.)

The material-balance line where B and D are held constant is given by Equation 8.20(5).

The line in the figure has a positive slope because y_h is plotted against x_l, whereas the Equation 8.20(5) uses the same component in distillate and bottoms. The slope of the material-balance line is always opposite in sign to that of the four hyperbolic curves. The curves are plotted by iterating D/F and calculating y_h and x_l with each of the other manipulated variables held constant.

Observe how closely spaced the four curves are, particularly the two representing constant reflux L and constant boil-up V, which are always the closest together. The curves representing constant reflux ratio and constant boil-up ratio are always on the outside of the envelope. The particular separation described in this example is considered mild. For columns making purer products, or with more trays or a higher reflux ratio, the curves are even more tightly packed. As will be seen later, this increases loop interaction.

Some information can be gleaned from the curves themselves. If they are widely spaced, interaction is slight and almost any control system will work satisfactorily. At the other extreme, if they are tightly packed, a material-balance control system manipulating one or the other product flow must be used if control over both compositions is to be attained, and any of the other manipulated variables can be used with it. Finally, the slope of the material-balance line is instructive. If it is nearly horizontal, then holding distillate flow constant (in ratio to feed rate) will tend to regulate distillate composition; if it is nearly vertical, then holding bottom flow constant (in ratio to feed rate) will tend to regulate bottom composition. In these cases, pairing manipulated and controlled variables is intuitive. For all other cases, however, relative gains should be calculated.

RELATIVE GAIN CALCULATIONS

The relative gains for controlling two compositions in a distillation column is a function of the ratio of the two slopes of the selected operating curves at the focal point:

$$\lambda_{y1} = \frac{1}{1 - \dfrac{(\partial y/\partial x)_2}{(\partial y/\partial x)_2}} \qquad \textbf{8.20(15)}$$

In this expression, the slope of the curve representing a constant manipulated variable m_1, proposed to control top composition, is divided by the slope of the curve representing a constant manipulated variable m_2, proposed to control bottom composition. If either the first approaches zero, or the second approaches infinity, the relative gain will approach 1.0—a favorable but unlikely result in distillation. If the two slopes have opposite signs—in which case one of the manipulated variables would have to be a product flow—the denominator will exceed unity and the relative gain will therefore fall in the range of 0–1. If the two slopes have the same sign—a nonmaterial-balance system—the resulting relative gains will lie outside of the 0–1 range.

Slope Calculations

The slope of the material-balance line is the partial derivative of Equation 8.20(5):

$$\left.\frac{\partial y_i}{\partial x_i}\right|_{D/F} = -\frac{y_i - z_i}{z_i - x_i} = -\frac{B}{D} \qquad \textbf{8.20(16)}$$

However, it needs to be placed in terms of the key impurities. At constant D/F, the partials of the key components have opposite signs: $\partial y_h = -\partial y_l$. Therefore,

$$\left.\frac{\partial y_h}{\partial x_l}\right|_{D/F} = \frac{y_l - z_l}{z_l - x_l} = \frac{B}{D} = \delta \qquad \textbf{8.20(17)}$$

It is given the designation δ because it will appear again in the formulas for the other slopes, which are much more complex.

The slope of the constant-reflux-ratio curve is given the designation σ because it, too, will be part of later formulas:

$$\left.\frac{\partial y_h}{\partial x_l}\right|_{L/D} = \frac{(1-x_{hh})y_l y_h}{(1-y_{ll})x_h x_l} = \sigma \qquad \textbf{8.20(18)}$$

The slope of the constant-reflux curve is a combination of δ, σ, and another term designated ε:

$$\varepsilon = \frac{nEy_l y_h}{2[1+(z_l+z_{ll})L/D](y_l-x_l)} \qquad \textbf{8.20(19)}$$

$$\left.\frac{\partial y_h}{\partial x_l}\right|_{L/F} = \frac{\sigma-\delta\varepsilon}{1-\varepsilon} \qquad \textbf{8.20(20)}$$

The slope of the constant-boil-up curve is similar:

$$\left.\frac{\partial y_h}{\partial x_l}\right|_{V/F} = \frac{\sigma-\delta\varepsilon(1+D/L)}{1-\varepsilon(1+D/L)} \qquad \textbf{8.20(21)}$$

Finally, the slope of the constant-boil-up ratio curve is another extension of these:

$$\left.\frac{\partial y_h}{\partial x_l}\right|_{V/B} = \frac{\sigma-\delta\varepsilon(1+D/L)(\delta-1)}{1-\varepsilon(1+D/L)(\delta-1)/\delta} \qquad \textbf{8.20(22)}$$

Because of the common factors spread throughout the formulas, it is not surprising that the operating curves always fall in the same sequence.

Table of Relative Gains

Using the format given in Table 8.20b, the relative gains will be displayed for the column described by the operating curves of Figure 8.20c.

Again, each of the numbers represents the diagonal pair in a two-by-two array. For example, 0.793 is the relative gain when distillate flow is assigned to control its own composition and boil-up is assigned to control bottom composition.

The off-diagonal pair in each array is not useful and so is not displayed. The numbers in the first row and first column always fall in the 0–1 range because they are the result of material-balance configurations. The other four numbers are always >1; the opposite pairs in these arrays are negative and therefore not useful.

Interpreting the Results

From the set of eight possible relative gains, there will be two on either side of the ideal value of 1.0—one higher and one lower. For relative gains in the 0–1 range, the worst case of interaction would be represented by 0.5, because configurations having lower numbers should be rejected by the designer. In general, relative gains are >0.5 for a less-pure product controlled by its own flow and a more-pure product controlled by one of the other manipulated variables. In the above example, distillate is less pure and is more favorably controlled by its own flow (relative gain of approximately 0.8 vs. 0.2).

TABLE 8.20d
Table of Independent Relative Gain Arrays for the Column of Figure 8.20c

m_y/m_x	D	L	L/D
B	—	0.125	0.236
V	0.793	21.02	6.408
V/B	0.813	6.400	3.963

Of the set of relative gains >1, the higher the number, the worse the interaction. Relative gains in the 1–10 range are controllable, but configurations having numbers >10 should be rejected. For composition loops on a distillation column, relative gain values of 0.5 and 10 give about equivalent levels of interaction.[4]

High relative gain values pose additional problems, however, one of which is increased sensitivity to disturbances associated with high *steady-state* gains. Consider the case for the *LV* subset in Table 8.20d, whose relative gain is 21. Relative gain in a two-by-two system is defined as the ratio of one open-loop gain with the other loop open (i.e., in terms of the other manipulated variable) to the open-loop gain with the other loop closed (i.e., in terms of the other controlled variable). The diagonal elements of Λ_{LV} are

$$\lambda_{xV} = \lambda_{yL} = \frac{(\partial x/\partial V)_L}{(\partial x/\partial V)_y} = 21.0 \qquad \textbf{8.20(23)}$$

However, if reflux ratio were chosen instead of reflux to control top composition, then we would be working with the elements of $\Lambda_{L/D,V}$:

$$\lambda_{xV} = \lambda_{yL/D} = \frac{(\partial x/\partial V)_{L/D}}{(\partial x/\partial V)_y} = 6.41 \qquad \textbf{8.20(24)}$$

Now the denominators in these two expressions are identical because the bottom-composition loop is unchanged, but its steady-state gain is different—it is a function of the configuration chosen for the top loop! Reconfiguring the top-composition loop from manipulating reflux to manipulating reflux ratio actually lowers the steady-state gain of the bottom-composition loop by a factor of 21/6.4 or approximately 3.3.

This allows a 3.3-fold reduction in the proportional band (increase in gain) of the bottom-composition controller, with an accompanying 3.3-fold decrease in peak deviation and integrated error following any disturbance to the column. This improvement has been witnessed following reconfiguration of the control-system structure on distillation columns.

Another reason to avoid configurations having high relative gains is the difficulty in applying decoupling to them. Reference 5 estimates the minimum error in a decoupler that would force the decoupled relative gain of the two loops to infinity—where all control is lost. For a relative gain of 21, the error in the decoupler—that is, its failure to match the parameters of the column perfectly—would have to be only 2.5% to produce a decoupled relative gain of infinity.

This "ill-conditioning" of the basic regulatory control structure has led to the failure of some multivariable predictive controllers to successfully decouple their processes.[5]

The lowest of the relative gains >1 is always the combination of controlling y with reflux ratio and x with boil-up ratio. In the example given in Table 8.20d, the relative gain of 3.96 should give acceptable control, significantly better than the other choices of 6.4. This is also the most complex control system, but has the advantage of built-in feedforward control.

Following an increase in liquid feed, the base level will start to rise, causing the level controller to increase bottom flow and boil-up in their existing ratio. The increased boil-up will raise column pressure and/or level in the reflux drum, causing the pressure or level controller to increase reflux and distillate flows in their ratio. Therefore, all four streams will have increased in ratio to the feed rate, at the right time, without a feedforward calculation having been made. The dynamic balance is achieved without a dynamic compensator required to model the hydraulics of the column below the feed tray.

The material-balance choices have distillate flow manipulated to control its own composition, with relative gains of about 0.8. Performance of these systems will be comparable to the 3.96 relative gain, but they require feedforward control in the way of distillate flow set in ratio to feed rate. To achieve equivalent dynamic response, changes in distillate flow must be converted into opposite changes in reflux flow as follows:

$$L = m_L - kD \qquad \textbf{8.20(25)}$$

where m_L is the output of the level controller on the reflux drum (or the pressure controller if the drum and condenser are flooded). Coefficient k has at least the value of L/D, and can be increased further to speed the response of the top-composition loop.

SIDESTREAM COLUMNS

Each sidestream adds a degree of freedom—a manipulated variable that can be used to control either its own composition or the concentration of an off-key impurity in one of the other products. This treatise will focus on single-sidestream columns as being the most common. The method can also be extended to some columns having two sidestreams, but it is not intended for those having multiple sidestreams, such as crude-oil fractionators.

The problem of assigning control loops to a single-sidestream column can be reduced to the simpler problem of interaction in a two-product column by deductive logic. Most single-sidestream columns will have one product flow that is much smaller than the other two, and it should be placed under material-balance control, as described for three different classes of column described below.

Pasteurization Columns

A pasteurization column has a main product withdrawn a few trays from the top, with a small stream taken overhead, usually as a vapor. An ethylene fractionator operates this way, its major product—ethylene—being withdrawn as a liquid sidestream whose lighter impurity, methane, is purged as a vapor overhead. Control-loop assignment for this column begins by using methane-product flow to control the methane content of the ethylene sidestream. This is an obvious choice, because whatever methane in the feed that is not withdrawn from the top will exit with the ethylene stream—none goes out the bottom.

With this loop assigned, relative gains may be calculated for controlling the heavy key (ethane) in the ethylene product and the light key (ethylene) in the bottom product, as if it were a two-product column. In the slope calculations, use the number of trays between the sidestream and the bottom as n, and the ratio of internal reflux (below the sidestream) to the sidestream flow, as the reflux ratio.

Nonmaterial-balance relative gains for ethylene fractionators tend to be very high, often over 1000, due to a low relative volatility and high product purities. As a result, a material-balance control system is needed, manipulating either the sidestream flow to control its own quality or bottoms flow to control its own composition, depending on the set of relative gains in the 0–1 range. An example is given in Reference 6.

Heavy-Ends Columns

The opposite of a pasteurization column is the heavy-ends column, where a major product is withdrawn as a sidestream vapor a few trays above the bottom. A small stream of heavy ends is taken from the bottom as a liquid to limit their concentration in the vapor product. The top section of the column is normal.

Here again, logic dictates that the flow of the smallest stream—the bottom—should be used to control the heavier impurity in the sidestream. This reduces the problem to one of interaction between the sidestream and distillate composition loops, and their relative gains can be used to select the best manipulated variables. The slope calculations should use the number of trays between the vapor sidestream and the top of the column—no other modifications are necessary.

Small Sidestreams

When a sidestream is the smallest of the flows, it is being used to withdraw a middle-boiling contaminant in the feed. Boiling between the major products, this substance is so volatile that it vaporizes before reaching the bottom, and yet condenses before reaching the top. If it is not intentionally removed as a sidestream, it will accumulate until it occupies most of the trays in the column, thereby reducing the separation between the main products.

This loop assignment is straightforward—sidestream flow is manipulated to control either its own composition or the middle-key impurity in the nearer major product. Then the interaction between the major products is treated in the same way as any two-product column.

An ethanol-water column has two liquid sidestreams—ethanol leaving a few trays below the top, and fusel oil (higher alcohols) withdrawn a few trays above the bottom. A small stream of heads (aldehydes) is removed from the top, and water from the bottom. This is essentially a pasteurization column in that a major product (ethanol) is taken as an upper sidestream.

Heads flow determines the aldehyde content in the ethanol product. Because the ethanol content in the water leaving the bottom is in the parts-per-million range, and the sidestream is not very concentrated, relative gains for material-balance control are typically in the 0.95–0.99 range. The preferred configuration then has ethanol flow controlling its own water content (proof), and boil-up controlling the trace level of ethanol in the water leaving the column base. A small stream of fusel oil is withdrawn to control its own water content—it is typically subcooled and decanted, with the aqueous phase returned to the column.

OPEN-LOOP GAINS IN TERMS OF RELATIVE GAINS

During the last 10 years process control software, advanced process control has been extensively used in the process industries including computer control of distillation columns operation. To develop control systems for distillation columns with it, dynamic characteristics of the column operations have to be identified by intentionally changing inputs to obtain responses of the outputs.

Based on the data of the inputs and outputs, the dynamic models of the columns are developed by using a part of the APC software. The size of an input change must be large enough to produce the sufficient change of the outputs that are clearly above the noise level. However, the input change should not be too large to cause the column upsets that are not only costly but also prolong the test periods. Without theoretical values of eventual changes of the outputs responding to a given input, a trial-and-error approach tends to require excessive time for model building for the APC usage. This aspect is particularly significant for distillation in which time constants are long compared with most "unit operations."

In addition, for most "self-regulating" processes, in which a step change in an input produces outputs that approach some fixed steady-state values, open-loop gains are readily calculated.[7] This is another reason why it is significant to calculate open-loop gains for distillation operations.

As stated above, to obtain the theoretical values of eventual changes of the outputs, we pay attentions to open-loop gains. An open-loop gain is a numerator in the definition of a relative gain; therefore, it is readily expressed in terms of a relative gain. As an example, let the distillation column shown in Figure 8.20a represent a methanol recovery column whose steady-state condition is given in Table 8.20e.

In Table 8.20e, a steady-state condition of a methanol recovery column, which fits to Figure 8.20a in its configuration, is illustrated. However, the performance of the column represented by this figure is different from that treated earlier in Figure 8.20c and Table 8.20b.

Using the definition of relative gain as it was defined in Equations 2.25(1) and 2.25(2) in Section 2.25, Relative Gain Calculations, open-loop gain $\partial y/\partial\, m_1|m_2$ is the sensitivity of y to m_1, whereas the $m_2 - x$ loop is open (control valve opening is constant) and the two level loops are closed (in automatic, and therefore maintaining their set points). Instead of computing the open-loop gains on the basis of their definitions, they will be readily calculated from the relative gains $\lambda_{yD}(\Lambda_{DV})$ and $\lambda_{yL}(\Lambda_{LV})$ by the formulas in Figure 8.20f.

Because the equation

$$\Delta D = -\,\Delta B \qquad \textbf{8.20(26)}$$

holds, derivatives with respect to B are not included in the figure.

Reducing Matrix Models to 2 × 2 Subsets

As described earlier in the paragraph titled RGA for Distillation, we set up 2 × 2 subsets that set x and y against possible

TABLE 8.20e

Steady-State Condition of a Methanol Recovery Column

Flow Rate	*Feed (F)*	*Distillate (D)*	*Residue (B)*	*Reflux (L)*	*Vapor (V)*
lb-mol/h	882.0	38.6	843.4	1253.5	1292.1
(kg-mol/h)	(400.0)	(17.5)	(382.5)	(568.5)	(586.0)
Mole Fraction	*In Feed (F)*	*In Distillate (D)*	*In Bottoms (B)*		
Methanol	0.046 = z	0.986 = y	0.0032 = x		
Formaldehyde Water	$\left.\begin{matrix}0.196\\0.758\end{matrix}\right\}=1-z$	$\left.\begin{matrix}0.005\\0.009\end{matrix}\right\}=1-y$	$\left.\begin{matrix}0.205\\0.7918\end{matrix}\right\}=1-x$		

Equation **8.20(43)**

	i \ j	Manipulated variables fixed (j)		
		D (1)	V (2)	L (3)
Manipulated variables varied (i)	D (1)	—	$\left.\frac{\partial y}{\partial(D/F)}\right\|_V$	$\left.\frac{\partial y}{\partial(D/F)}\right\|_L$
	V (2)	$\left.\frac{\partial y}{\partial(V/F)}\right\|_D$	—	$\left.\frac{\partial y}{\partial(V/F)}\right\|_L$
	L (3)	$\left.\frac{\partial y}{\partial(L/F)}\right\|_D$	$\left.\frac{\partial y}{\partial(L/F)}\right\|_V$	—

Equation **8.20(44)**

	i \ j	Manipulated variables fixed (j)		
		D (1)	V (2)	L (3)
Manipulated variables varied (i)	D (1)	—	$\left.\frac{\partial x}{\partial(D/F)}\right\|_V$	$\left.\frac{\partial x}{\partial(D/F)}\right\|_L$
	V (2)	$\left.\frac{\partial x}{\partial(V/F)}\right\|_D$	—	$\left.\frac{\partial x}{\partial(V/F)}\right\|_L$
	L (3)	$\left.\frac{\partial x}{\partial(L/F)}\right\|_D$	$\left.\frac{\partial x}{\partial(L/F)}\right\|_V$	—

Matrices for Open Loop Gains

	i \ j	Manipulated variables fixed (j)		
		D (1)	V (2)	L (3)
Manipulated variables varied (i)	D (1)	—	$-\frac{y-x}{D/F}\lambda_{yD}(\Lambda_{DV})$	$\frac{y(1-y)}{\beta}[1-\lambda_{yL}(\Lambda_{LV})]$
	V (2)	$\frac{y(1-y)}{\beta}[1-\lambda_{yD}(\Lambda_{DV})]$	—	$\frac{y(1-y)}{\beta}[1-\lambda_{yL}(\Lambda_{LV})]$
	L (3)	$\frac{y(1-y)}{\beta}(1-\lambda_{yD}(\Lambda_{DV}))$	$\frac{y-x}{D/F}\lambda_{yD}(\Lambda_{DV})$	—

	i \ j	Manipulated variables fixed (j)		
		D (1)	V (2)	L (3)
Manipulated variables varied (i)	D (1)	—	$-\frac{y-x}{B/F}[1-\lambda_{yD}(\Lambda_{DV})]$	$\frac{x(1-x)}{\beta}\lambda_{yL}(\Lambda_{LV})$
	V (2)	$-\frac{x(1-x)}{\beta}\lambda_{yD}(\Lambda_{DV})$	—	$\frac{x(1-x)}{\beta}\lambda_{yL}(\Lambda_{LV})$
	L (3)	$\frac{x(1-x)}{\beta}\lambda_{yD}(\Lambda_{DV})$	$\frac{y-x}{B/F}[1-\lambda_{yD}(\Lambda_{DV})]$	—

Formulas for Open Loop Gains

Where

$$\lambda_{yD} = \lambda_{yD}(\Lambda_{DV}) \qquad \textbf{8.20 (45)}$$

$$\lambda_{yL} = \lambda_{yL}(\Lambda_{LV}) \qquad \textbf{8.20 (46)}$$

$$\begin{bmatrix} — & -3.7 & -3.6 \\ 0.079 & — & -3.6 \\ 0.079 & 3.7 & — \end{bmatrix} \qquad \begin{bmatrix} — & -0.858 & 0.862 \\ -0.0036 & — & 0.862 \\ 0.0036 & 0.858 & — \end{bmatrix}$$

Matrices of Calculated Open Loop Gains
for Example #2

FIG. 8.20f

Open-loop gain matrices: y vs. m_i, left, Equation 8.20(43); x vs. m_i, right, 8.20(44). Top row of matrices give the defining equations, the middle row provides the formulas for calculation, and the bottom row gives the results for Example 2.

pairs of manipulated variables. The open-loop gains (numerators) for the pairs in these matrices are then evaluated with only the other composition controller in manual. For example, the subset that assigns m_1 and m_2 to control y and x is denoted by

$$\Lambda_{m_1m_2} = \begin{array}{c|cc} & m_1 & m_2 \\ \hline y & \lambda_{ym_1} & \lambda_{ym_2} \\ x & \lambda_{xm_1} & \lambda_{xm_2} \end{array} \qquad \textbf{8.20(27)}$$

By using similar notations as employed in Equation 2.25(1) in Section 2.25, Relative Gain Calculations, the modified relative gain for controlling y by manipulating m_1 is given by

$$\lambda_{ym_1}(\Lambda_{m_1m_2}) = \frac{\left.\frac{\partial y}{\partial m_1}\right|_{m_2} \text{(level loops closed and } m_2 - x \text{ loop open)}}{\left.\frac{\partial y}{\partial m_1}\right|_{x} \text{(level loops and } m_2 - x \text{ loop closed)}} \qquad \textbf{8.20(28)}$$

Simplified Model for Separation

To compute the relative gains, separation (S) defined in Equation 8.20(7) is used. With notation $y_l = y$, $y_h = 1 - y$,

$x_l = x$, and $x_h = 1 - x$, separation is expressed as

$$S = \frac{y(1-x)}{x(1-y)} \qquad \textbf{8.20(29)}$$

The separation may be thought of as the molar ratio of the light component to the heavy component in the distillate, $y/(1 - y)$, divided by the same ratio, $x/(1 - x)$ in the bottom product.

To keep this discussion and the following discussions more comprehensive, the following equation is used to represent separation:

$$V/F = \beta(\ln S) \qquad \textbf{8.20(30)}$$

where F is the feed flow rate to the column and β is a constant. This representation, outlined by Shinskey[8] in the 1970s, states that the separation is a function of vapor flow rate V. The constant β may be thought of as a factor that determines the separation in terms of the vapor flow rate for a given feed rate. For a given column, β can be obtained by dividing an observed V/F by the natural logarithm of the calculated separation.

The equation does not have the feature as the one that would be derived from Equation 8.20(8). However, the relative gains derived from Equations 8.20(29) and 8.20(30) are simpler and easier to use. Simplification is justified because our objective here is to estimate open-loop gains for the purpose that was described earlier.

Formulas for Relative Gain

Relative gain $\lambda_{yD}(\Lambda_{DV})$ is calculated from Equation 8.20(28) by computing the partial derivatives:

$$\lambda_{yD}(\Lambda_{DV}) = \frac{\partial y/\partial D|V}{\partial y/\partial D|x} = \frac{\partial y/\partial (D/F)|V}{\partial y/\partial (D/F)|x} \qquad \textbf{8.20(31)}$$

According to Equation 8.20(30), in which it is assumed that only V and S are variables, the numerator of the first equation is equal to $\partial y/\partial D|S$. Equations 8.20(29) and 8.20(31), together with material and component balance equations

$$F = D + B \qquad \textbf{8.20(32)}$$

$$Fz = Dy + Bx \qquad \textbf{8.20(33)}$$

will lead to:[9]

$$\lambda_{yD}(\Lambda_{DV}) = \frac{1}{1 + \dfrac{(y-z)x(1-x)}{(z-x)y(1-y)}} \qquad \textbf{8.20(34)}$$

where z is molar fraction of the light component in the feed. As described, the relative gain array has the number in every row and column summing to 1.0. Therefore,

$$\Lambda_{DV} = \begin{array}{c|cc} & D & V \\ \hline y & \lambda_{yD}(\Lambda_{DV}) & 1-\lambda_{yD}(\Lambda_{DV}) \\ c & 1-\lambda_{yD}(\Lambda_{DV}) & \lambda_{yD}(\Lambda_{DV}) \end{array} \qquad \textbf{8.20(35)}$$

Similarly, relative gain array Λ_{LV} can be written as

$$\Lambda_{LV} = \begin{array}{c|cc} & L & V \\ \hline y & \lambda_{yL}(\Lambda_{LV}) & 1-\lambda_{yL}(\Lambda_{LV}) \\ x & 1-\lambda_{yL}(\Lambda_{LV}) & \lambda_{yL}(\Lambda_{LV}) \end{array} \qquad \textbf{8.20(36)}$$

In the same way Equation 8.20(34) was derived, the value of $\lambda_{yL}(\Lambda_{LV})$ can be determined as

$$\lambda_{yL}(\Lambda_{LV}) = \left[1 + \frac{\beta(y-x)^2}{y(1-y)(z-x)}\right]\lambda_{yD}(\Lambda_{DV}) \qquad \textbf{8.20(37)}$$

Relative gain array for Λ_{DL}, which (as it was described previously) is equal to Λ_{BL}, is given by Equation 8.20(38):

$$\Lambda_{DL} = \begin{array}{c|cc} & D & L \\ \hline y & \lambda_{yD}(\Lambda_{DL}) & 1-\lambda_{yD}(\Lambda_{DL}) \\ x & 1-\lambda_{yD}(\Lambda_{DL}) & \lambda_{yD}(\Lambda_{DL}) \end{array} \qquad \textbf{8.20(38)}$$

Equations 8.20(34) and 8.20(37) will be used to obtain numerical values of matrices for open-loop gains y vs. m_i, which are contained in Equations 8.20(43) and 8.20(44) in Figure 8.20f.

Example 1

In Table 8.20e, pertinent steady-state conditions for a methanol recovery column[10] are shown. As illustrated in Table 8.20e, the column separates methanol from a methanol-formaldehyde-water mixture. Methanol is the light component and its molar fraction in the feed is z.

For simplicity, formaldehyde and water will be lumped as a heavy component whose "combined" molar fraction in the feed is $1 - z$. The same applies to the distillate and bottom products listed in Table 8.20e. By substituting the values of x, y, and z from Table 8.20e into Equation 8.20(34), and from Equations 8.20(35), 8.20(36), and 8.20(38), the relative gain arrays Λ_{DV}, Λ_{LV}, and Λ_{BL} are

$$\Lambda_{DV} = \begin{array}{c|cc} & D & V \\ \hline y & 0.165 & 0.835 \\ x & 0.835 & 0.165 \end{array} \qquad \textbf{8.20(39)}$$

$$\Lambda_{LV} = \begin{array}{c|cc} & L & V \\ \hline y & 39.6 & -38.6 \\ x & -38.6 & 39.6 \end{array} \qquad \textbf{8.20(40)}$$

$$\Lambda_{DL} = \Lambda_{BL} = \begin{array}{c|cc} & B(or\,D) & L \\ \hline y & 0.162 & 0.838 \\ x & 0.838 & 0.162 \end{array} \qquad \textbf{8.20(41)}$$

With reference to Equation 8.20(41),

$$\lambda_{yL}(\Lambda_{BL}) = \lambda_{xB}(\Lambda_{BL}) = 0.838 \qquad \textbf{8.20(42)}$$

Equations 8.20(43) to 8.20(46) are given in Figure 8.20f.

Open-Loop Gains

The open-loop gains, which are the numerators of the definition of the relative gains, should be considered. Open-loop gain $\partial y/\partial m_1|m_2$ is the sensitivity of y to m_1, while the $m_2 - x$ loop is open (control valve opening is constant) and the two level loops are closed (in automatic, and therefore maintaining their set points). Even when the relative gains are used in selecting the pairing of the controlled and manipulated variables, the designer should make sure that the open-loop gains are not too small. Instead of computing the open-loop gains on the basis of their definitions, they are readily calculated from the relative gains $\lambda_{yD}(\Lambda_{DV})$ and $\lambda_{yL}(\Lambda_{LV})$ by the formulas in Figure 8.20f.[10] Because Equation 8.20(26) holds, derivatives with respect to B are not included in the figure.

Example 2

If, for the distillation process of the previous example, the top product composition were more important than the bottom, the designer might use open-loop gains instead of relative gains to select variable pairing. The open-loop gains listed below have been calculated from the equations in the left-side matrices in Figure 8.20f. From Equation 8.20(43) in the figure,

$$\left.\frac{\partial y}{\partial (D/F)}\right|_v = -\left.\frac{\partial y}{\partial (L/F)}\right|_v = -3.7 \qquad \textbf{8.20(47)}$$

This equation means that for a fixed vapor flow rate V, or approximately for a fixed flow rate of the heating medium to the base reboiler (Q_B), an incremental increase in distillate-to-feed-flow ratio (or corresponding decrease of reflux-to-feed-flow ratio) will decrease the molar fraction of methanol in the top product by a factor of 3.7. Similarly, for a fixed reflux flow rate L, from Equation 8.20(43) in Figure 8.20f,

$$\left.\frac{\partial y}{\partial (D/F)}\right|_L = \left.\frac{\partial y}{\partial (V/F)}\right|_L = -3.6 \qquad \textbf{8.20(48)}$$

and for a fixed top product flow, from Equation 8.20(43),

$$\left.\frac{\partial y}{\partial (V/F)}\right|_D = \left.\frac{\partial y}{\partial (L/F)}\right|_D = 0.079 \qquad \textbf{8.20(49)}$$

The absolute values of the open-loop gains in Equations 8.20(47) and 8.20(48) are about the same, whereas that of Equation 8.20(49) is much smaller. Therefore, only control actions corresponding to Equations 8.20(47) and 8.20(48) should be considered. Based on this and on the preference for material-balance control[8] over energy balance, the recommended choice of pairing the controlled and manipulated variables as shown in Figure 8.20g might be obvious to many readers.

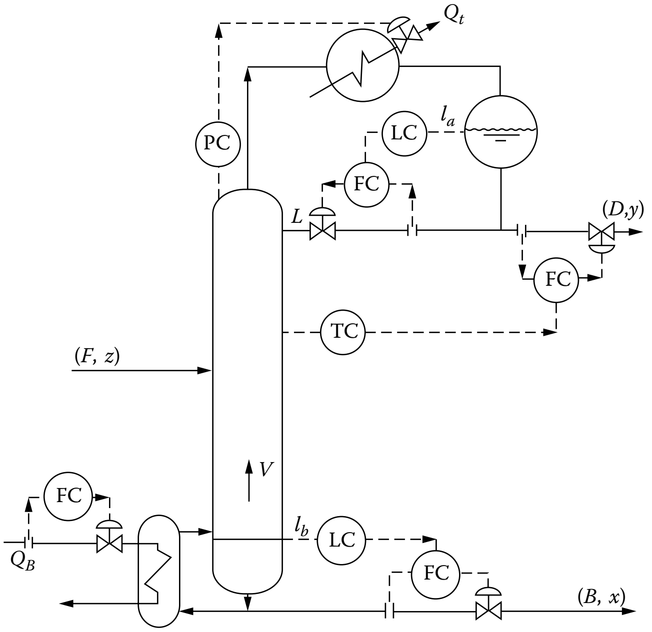


FIG. 8.20g
Pairing of controlled variables with manipulated variables as selected in Example 2.

A step-by step approach for arriving at the same conclusion is given in Reference 10.

CONCLUSIONS

A distillation column has anywhere from four to six control loops, which need to be optimally configured if the column is to be stable and meet its performance objectives. Four single loops can be configured in 24 different ways, and logical mathematical combinations of manipulated variables expand the possible number of configurations manyfold.

Because of the complexity of the configuration problem, and the unique requirements of each individual column, a methodology is required for matching the configuration to the column. That methodology is relative gain analysis. The basics of RGA, described in detail in Section 2.25, were specially adapted in this section to the particular multivariable control problem represented by distillation columns.

In addition, from numerical values of relative gains, open-loop gains were also readily calculated. The open-loop gains will help control engineers to conduct plant tests to effectively collect transient responses of distillation columns that are required to implement so-called advanced process control.

NOMENCLATURE

B	Bottom product flow rate
D	Distillate flow rate
E	Tray efficiency

F	Feed flow rate
h	Suffix indicating heavy key component
hh	Suffix indicating heavier key component
k	Coefficient defined in the descriptions that follow Equation 8.20(25)
L	Reflux flow rate
l	Suffix indicating light key component
ll	Suffix indicating lighter key component
l_a	Liquid level in overhead receiver
l_b	Liquid level in column base
m_i	Manipulated variable
m_L	Output of level controller on the reflux drum (or output of pressure controller if the drum and condenser are flooded)
n	Number of trays
P	Column pressure
Q_B	Heat input rate to base reboiler
Q_T	Heat removal rate at overhead condenser
S	Separation
V	Vapor flow rate
x	Molar fraction of light component in bottom product
y	Molar fraction of light component in distillate
z	Molar fraction of light component in feed

GREEK LETTERS

α	Average relative volatility between the key components across the column
β	Positive constant defined by descriptions that follow Equation 8.20(30)
∂	Partial derivative
δ	Constant defined in Equation 8.20(17)
Δ	Deviation from a steady-state condition
ε	Constant defined by Equation 8.20(19)
Λ	Relative gain array
λ	Relative gain
σ	Slope of constant-reflux-ratio curve defined by Equation 8.20(18)

References

1. Douglas, J. M., Jafarey, A., and McAvoy, T. J., "Short-Cut Techniques for Distillation Column Design and Control, I: Column Design," *Ind. Eng. Chem. Process Des. Dev.,* Vol. 18, No. 2, pp. 197–202, 1979.
2. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, pp. 340–347.
3. Fenske, M. R., "Fractionation of Straight-Run Pennsylvania Gasoline," *Ind. Eng. Chem.,* May 1932.
4. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, p. 350.
5. Hugo, A., "Limitations of Model Predictive Controllers," *Hydrocarbon Proc.,* January 2000, p. 86.
6. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, p. 368.
7. Ochiai, S., "Calculating Process Control Parameters from Steady State Operating Data," *ISA Transactions,* Vol. 36, No. 4. pp. 313–320, 1998.
8. Shinskey, F. G., *Distillation Control for Productivity and Energy Conservation,* 1st edition, New York: McGraw-Hill, 1977, pp. 35–63, pp. 289–321.
9. Shinskey, F. G., *Distillation Control for Productivity and Energy Conservation,* 2nd edition, New York: McGraw-Hill, Inc., 1984, pp. 146–147.
10. Ochiai, S., "A Guide for Quickly Designing Distillation Column Control Systems," *Proceedings of Joint Automatic Control Conference,* San Francisco, CA, Vol. 1, Paper No. TA7-B, 1980.

Bibliography

Gordon, L. M., "Practical Evaluation of Relative Gains: The Key to Designing Dual Composition Controls," *Hydrocarbon Processing*, December 1982, pp. 87–92.

McAvoy, T. J., *Interaction Analysis*, Research Triangle Park, NC: Instrument Society of America, 1983.

McAvoy, T. J., "Steady-State Decoupling of Distillation Columns," *Ind. Eng. Chem. Fundam.*, Vol. 18, No. 3, pp. 269–273, 1979.

Ryskamp, C. J., "New Control Strategy Improves Dual Composition Control," *Hydrocarbon Processing*, June 1980.

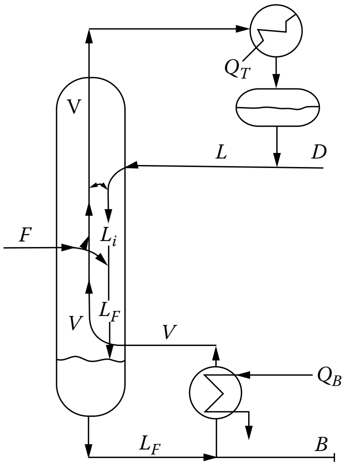
Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control*, New York: John Wiley & Sons, 1989, pp. 689–695.

Shinskey, F. G., *Distillation Control for Productivity and Energy Conservation*, 2nd edition, New York: McGraw-Hill, 1984, pp. 60–89, pp. 131–166.

Shinskey, F. G., "Predict Distillation Column Response Using Relative Gains," *Hydrocarbon Processing*, May 1981, pp. 196–200.

8.21 Distillation: Optimization and Advanced Controls

H. L. HOFFMAN, D. E. LUPFER (1970) **L. A. KANE** (1985)
B. A. JENSEN (1995) **B. G. LIPTÁK** (2005)



Flow sheet symbol

INTRODUCTION

Section 8.19 described the basic, single-input single-output (SISO) distillation control systems. These simple control schemes do keep the operation stable, but they cannot optimize it and they do necessitate that the operator, as plant conditions change, periodically readjust the set points of these SISO loops.

In Section 8.20, it was noted that a two-product distillation tower has five controlled and five manipulated variables. Because pressure is usually controlled to close the heat balance and the two levels are controlled to close the material balance around the column, eight configurations are possible to control product compositions (Table 8.20b). Interaction always exists between the material and energy balances in a distillation column. Section 8.20 describes how the interaction between the two composition control loops can be minimized by calculating the eight corresponding relative gain (RG) values and selecting the pairing, which gives an RG closest to 1.0.

Control of distillation towers involves the manipulation of the material and energy balances in the distillation equipment to affect the composition of the products. This section builds upon the previous two, while focusing on optimization and on the use of multivariable advanced process controls (APC).[1] In today's competitive market, it is necessary to push equipment to operating limits to maximize production rate or minimize the energy cost of production.

Advanced process controls are usually distinguished from regulatory SISO controls by being multivariable in nature (multiple input/multiple output) and by utilizing some model of the process. The APC products on today's market can be distinguished on the basis of their approach to modeling the process. They can be grouped into three categories: The *white box* models apply to well understood processes, such as distillation, where theoretical dynamic models of the process can be derived based on mass, energy, and momentum balances of the process.

The *fuzzy logic* and *black box* models are used for processes that are poorly understood or when it is acceptable to use a complete mechanistic empirical model constructed solely from *a priori* knowledge. Because of the well-understood nature of distillation, this section will give emphasis to the white box approach to modeling.

The goal of this section is to provide instrument engineers with the tools necessary to design *unique* advanced control strategies that will match the requirements of the specific distillation columns they encounter. The section will first discuss the various APC control strategies, and after that it will describe a variety of optimization schemes. After a listing of APC-related definitions, the discussion of APC in distillation will first discuss the black box and the fuzzy logic techniques, which are less applicable to this well-understood process. After this brief treatment, a more detailed discussion of the development of the white box models will be presented.

Definitions

ARTIFICIAL NEURAL NETWORKS (ANN): ANNs can learn complex functional relations by generalizing from a limited amount of training data; hence, they can thus serve as black box models of nonlinear, multivariable static and dynamic systems and can be trained by the input/output data of these systems. ANNs attempt to mimic the structures and processes of biological neural systems. They provide powerful analysis properties such as complex processing of large input/output information arrays, representing complicated nonlinear associations among data, and the ability to generalize or form concepts-theory.

BLACK BOX MODEL: *See* EMPIRICAL MODEL.

EMPIRICAL MODEL: This type of model can be used for processes for which no physical insight is available or used. This model structure belongs to families that are known to have good flexibility and have been "successful in the past." The parameters of the models are identified based on measurement data. A complete mechanistic model is constructed from *a priori* knowledge.

FUZZY LOGIC MODELING: This type of model is used for processes that are not fully understood. It is a linguistically interpretable rule-based model that is based on the available expert knowledge and measured data.

MODEL-BASED CONTROL (MBC): In model-based control, a process model is used to make control decisions. The controller uses this model of the process to calculate a value for the manipulated variable, which should make the controlled variable behave in the desired way. The "inverse" nomenclature arises from how the model is used. In a normal modeling approach, one specifies the process input, and the model predicts the process output response. By contrast, MBC determines the process input (manipulated variable) that will cause a desired process output response (controlled variable value) to occur. This is the model inverse.

MODEL PREDICTIVE CONTROL (MPC): is a model-based control technique that uses process output prediction and calculates consecutive controller moves in order to satisfy control objectives.

OPEN-LOOP GAIN: The steady-state gain of a control loop when the other control loop(s) is (are) in manual. (Their control valve opening is constant.)

RELATIVE GAIN: RG is the ratio of the steady-state gain of the loop with other loops in manual, divided by the steady-state gain of the loop when the other loops are in automatic.

RELATIVE GAIN ARRAY: A matrix of dimensionless gain ratios giving one RG value for each pairing of manipulated and controlled variables.

WHITE BOX MODELING: This type of modeling is feasible if a good understanding of the process exists. In such cases, the dynamic models are derived based on mass, energy, and momentum balances of the process.

ADVANCED PROCESS CONTROL

Fuzzy logic- and black box-type model-free expert systems can be compared to the behavior of tennis players. The players do not necessarily understand Newton's laws of motion or the aerodynamic principles that determine the behavior of a tennis ball, but they have simply memorized the results of a large number of past responses. This is also the basis of human learning. All the neural network software packages on the market mimic this method of learning.

Neural networks, fuzzy logic, and statistical process control are all such methods, which can be used without the need for knowing the mathematical model of the process. The major difference between fuzzy logic and neural networks is that the latter can only be trained by data, but not with reasoning. Fuzzy logic is superior from this perspective, because it can be modified both in terms of the gain (importance) and also in terms of the functions of its inputs.

The main limitations of all model-free expert systems is their long learning period (which can be compared to the growing up of a child) and the fact that their knowledge is based solely on past events. Consequently, they are not prepared to handle new situations, and therefore if the process changes, they require retraining, because they are not well suited to anticipation.

Model-based control, model predictive control, and internal model control (IMC) are all based on white box modeling and are all suited for the optimization of such unit processes that are well understood, such as heat transfer or distillation. Their performance is superior to that of the model-free systems (fuzzy logic and black box), because they are capable of anticipation and, thereby, can respond to new situations. In this sense their performance is similar to that of feedforward control systems, while the model-free systems behave in a feedback manner only.

In this section, the APC control strategies that are based on fuzzy logic and black box models will be discussed first. This discussion will be followed by a more in-depth explanation of the white box model-based controls.

The Goals of APC Advanced control strategies attempt to compensate for process deviations in the shortest time possible by accounting for process dynamics, dead times, time delays, and loop interactions. The benefits of better control are:[2]

- Increased throughput
- Increased product recovery
- Energy conservation
- Reduced disturbances to other processing units
- Minimum rework or recycle of off-spec products
- Reduced operating personnel
- Increased plant flexibility

For example, good product composition control of distillation towers can save 5–15% of the energy required to achieve the required separation.[3]

The goal of basic distillation controls is to keep the unit running. The objective of advanced control is to keep it running at maximum profitability. The techniques available to implement advanced control include feedforward control; optimization, including constraint control; and model-based and multivariable control (MVC).

The challenge is to utilize the technique, the tools, and the available resources to design *unique* advanced control strategies that will match the specific objectives for the distillation columns. The choice between any of these control techniques depends upon factors such as preference and familiarity, complexity of scheme, degree of optimization, hardware for application, and number of variables monitored and controlled by single strategy.

Often, additional instrumentation is not needed when implementing advanced controls by building upon basic control designs. However, in many cases, new measurements are needed for calculation or compensation in order to implement an advanced control strategy. These must be retrofitted to the process.

Unlike basic distillation control, in which much of the control can be implemented by analog control systems, advanced control strategies usually require the use of higher-level computing systems. Optimization programs and model-based controls require large amounts of computing power. It is for this reason that APC control systems can be distributed over a variety of control equipment types in some kind of hierarchical or distributed fashion.

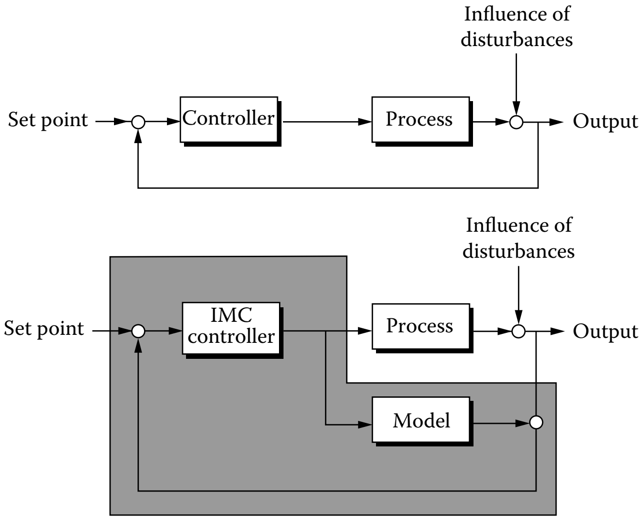


FIG. 8.21a
The configuration of a PID control loop (top) and an internal model controller loop (bottom).

Model-Based Control

The strategies presented in Sections 8.19 and 8.20 implement distillation control using PID controllers. Efforts have been made to improve PID performance by considering the dynamic nature of the fractionator, the nonlinearity of the system, and the decoupling of interactions.

Model-based controls have been gaining increasing popularity and have been discussed in detail in Sections 2.13 to 2.18 in Chapter 2. These use alternatives to the PID algorithms such as the internal model controller,[12] model algorithmic control,[13] dynamic matrix control,[14] and neural controllers.[15] Process model-based control uses an approximate process model directly for control in order to overcome the coupling effects in the distillation tower.

Most of these methods are nonlinear, all are predictive, and many are multiple–input multiple-output (MIMO). All depend upon the availability of some process model. Once a process model has been established, it is possible to build the inverse of that model, which can be used as a controller. In that sense, the PID controller is a linear inverse model of a single loop.

All control design is basically a model-based activity. This is true even with the PID controller, which uses first- and second-order lag approximations of the process to determine tuning parameters. An alternative to the PID controller is a linear model built into the controller. A simple model-based controller is the internal model controller. The difference between the PID and IMC controller is shown in Figure 8.21a.

Note that the IMC looks like it has the same structure as a Smith predictor in Figure 8.19x in Section 8.19. The difference is that the process model is explicitly an internal part of the controller model in the IMC. For a first-order system with dead time, a Smith predictor (Figure 8.19x) with a PI controller is equivalent to an IMC.

Nonlinear approximate models include algebraic representation of the McCabe-Theile diagram for both rectifying and stripping sections, short-cut fractionator calculations, and others.[16] These methods require the power of a computer to solve the equations. A number of control strategies also exist once the process model is known. For dual composition control, one method is the generic model control (GMC),[17] whose control law is described by the following equations:

$$y_{ss} = y_o + K_{1,1}(y_{sp} - y_o) + K_{2,1}\int (y_{sp} - y_o)\,dt \qquad \textbf{8.21(1)}$$

$$x_{ss} = x_o + K_{1,2}(x_{sp} - x_o) + K_{2,2}\int (x_{sp} - x_o)\,dt \qquad \textbf{8.21(2)}$$

where

x_{sp} and y_{sp} = the target set points for bottoms and overhead products
x_{ss} and y_{ss} = the specifications for bottoms and overhead products
x_o and y_o = the current compositions
K = tunable parameters for disturbance rejection

For example, if $K_{1,2} = 2.5$, $K_{2,2} = 0$, $x_{sp} = 2\%$, and $x_o = 1\%$, then $x_{ss} = 3.5\%$. x_{ss} is then used in the process model as the basis to compute *V*/*B* or any other output (manipulated) variable. Because the same action can be performed for y_{ss} with y_{ss} being substituted into a process model equation, such as *D*/*L*, the model-based control can be multivariable, handling nonlinearity, disturbances, and coupling, by tuning the *K* values.

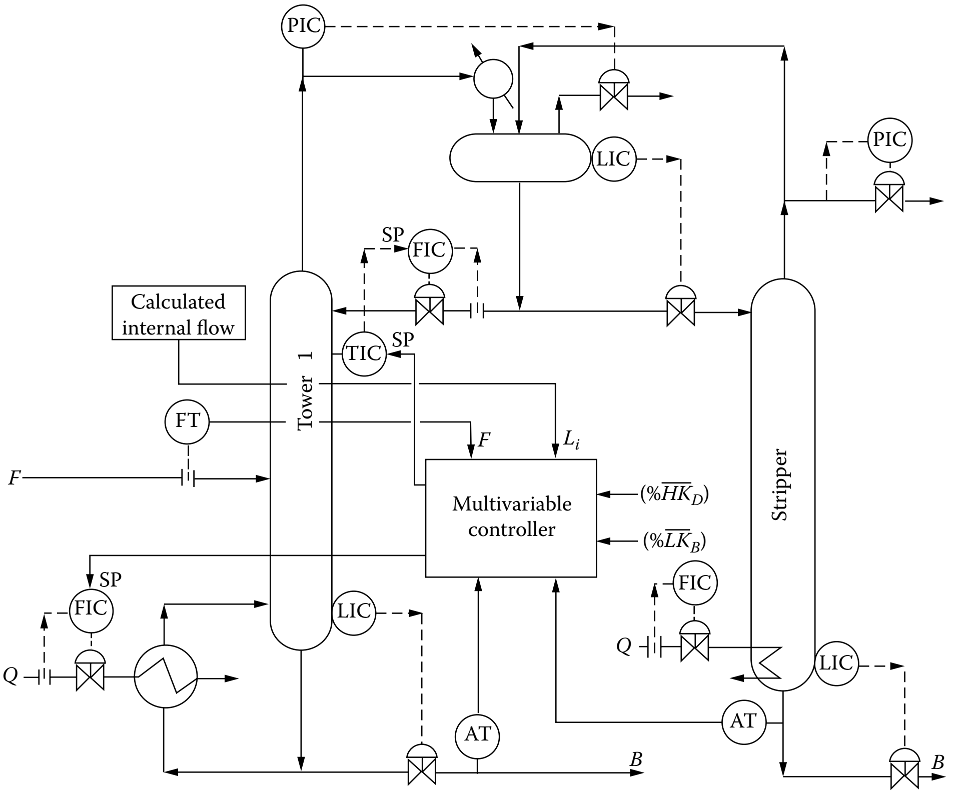


FIG. 8.21b
Fractionator control using a multivariable controller.

Multivariable Control

Multivariable control is a technique that services multiple-input, multiple-output algorithms simultaneously as opposed to the single-input, single-output ones. MVC is particularly well suited for highly interactive multivariable fractionators, where several control loops need to be decoupled. In general, the more difficult the process, the greater are the benefits of multivariable control. Multivariable control techniques can take safety constraints, process lags, and economic optimization factors all into consideration.

Like the model-based controller, the MVC-type controller is a predictive controller that uses information from the past plus dynamic models of the process to predict future behavior. Based upon predicted responses, the controller plans future moves to manipulated variables that will minimize the errors in each dependent controlled variable (Figure 8.21b).

The control diagram shown in Figure 8.21b illustrates an application of a multivariable controller. In the example, two products and an impurity stream are separated using two towers. The objective is to control the composition of both products. The two composition control loops are coupled so that when any single control action is taken to control one composition, that action also affects the other composition.

In this example, the controlled variables are the two product compositions as measured by process analyzers. Feed flow rate is a disturbance variable. The steam to the first column and the temperature at the top of that column are the manipulated variables. A constraint variable is an internal flow as calculated from other tower temperatures and flows.

The multivariable controller will take the appropriate steps to control both compositions, subject to the calculated constraint, by adjusting the two manipulated variables while accounting for the dead time caused by the stripper.

The identification and command (IDCOM)[13] method is a type of multivariable model algorithmic control. It is based on a process impulse response, which utilizes a predictive heuristic scenario technique to calculate the manipulated variable. The technique is to use a dynamic model to determine future values of the controlled variables. These calculated future values are compared to a desired reference set point trajectory. The manipulated variables are then adjusted to force future controlled variable values to follow the desired reference trajectories.

The technique of multivariable control requires the development of dynamic models based upon fractionator testing and data collection. Multivariable control applies the dynamic models and historical information to predict future fractionator characteristics. Predicted fractionator responses result in planned controller actions on the manipulated variables to minimize error for the dependent controlled variable, while considering constraints in the present and the future.

This controller is similar to a PID controller, except that the multivariable controller accepts several controlled variable set points and load variable measurements and, subject to constraints, outputs several manipulated variables.

All multivariable techniques require some sort of process model. Differences between various multivariable techniques lie in their calculation of internal models (whether nonlinear or linear), their method of predicting the future, their method of constraint handling, and their method for minimizing the controller's error.

Multivariable control may be considered to be an "overkill" and at worst a poor controller, if simpler techniques are adequate. However, for towers that are subject to constraints, towers that have severe interactions, and towers with complex configurations, multivariable control can be a valuable tool.

Dynamic Matrix Control

A multivariable predictive controller is based on dynamic matrix control.[14] DMC is a predictive control technique that uses a set of linear differential equations to describe the process. The DMC method is based upon a process step response and calculates manipulated variable moves via an inverse model. Coefficients for the linear equations describing the process dynamics are determined by process testing. A series of tests are conducted whereby a manipulated or load variable is perturbed and the dynamic response of all controlled variables is observed. This identification procedure is time-consuming and requires local expertise because of the experimentation involved. Once the models are obtained, the controller design can be designed.

The least-squares approach is taken to minimize the error of the controlled variables from their set points. Weighting constants scale controlled variable errors and influence which controlled variables are allowed to deviate from their set points if a constraint is encountered. The controller considers constraints in its plan for both present and future moves in each manipulated variable. Other factors affecting the response of the DMC controller are parameters that govern the relative amount of movement in the manipulated variables and the rate at which errors are reduced. This is analogous to the tuning parameters in a PID controller.

Artificial Neural Networks

As was discussed in detail in Section 2.18 in Chapter 2, one of the tools used in building internal models is the Artificial Neural Network, which can usually be applied under human supervision or integrated with expert or fuzzy logic systems. Figure 8.21c shows a three-layer, back-propagation ANN that serves to predict the manipulated variables of a column. Such predictive ANN models can be valuable, because they overcome the limitations of analyzers, which include both availability and dead time.

The process model's knowledge is stored in the ANN by the way the processing elements (nodes) are connected and the importance that is assigned to each node (weight). The ANN is "trained" by example, and therefore it contains the adaptive mechanism for learning from examples and to adjust its parameters based on the knowledge that is gained through this process of adaptation. During the "training" of these networks, the weights are adjusted until the output of the ANN matches that of the real process. Naturally, these networks do need "maintenance," because process conditions change, and when they do, the network requires retraining. The hidden layers help the network to generalize and even to memorize.

The ANN is capable to learn input/output relationships and inverse relationships, and hence it is useful in building internal model control based on the ANN-constructed plant models and their inverses. In a neural controller (Figure 8.21d), the ANN is used in calculating the control signal.

Neural Control The PID controller is the basic feedback mechanism for correcting errors between the current condition (measurement) and what is desired (set point). The PID assumes a linear process. Adaptive control and other techniques are used when nonlinearities are encountered (see Section 2.19 in Chapter 2). However, because of the structure of neural networks with their distributed representation, the neural controller promises the ability of adaptation, learning, and generalization to nonlinear problems.[18]

In the single-input, single-output configuration, instead of utilizing the basic PID equation, the network builds an internal nonlinear model, relating the controlled and corresponding manipulated variable. It builds this model by learning or "training" from a data set of known measurements and process responses. Often, a primary disturbance variable is included in this model. The dynamic response is recorded for the training data set. This makes the neural controller more useful and more robust than the standard PID.

These controllers most often use the back-propagation method of training to relate controlled, manipulated, and load variables (see Section 2.18 in Chapter 2). Figure 8.21c illustrates the simple back-propagation neural network used to create the nonlinear model.

Because the neural network paradigm can accommodate multiple inputs and multiple outputs, an entire fractionator model can be built into a single controller. The neural controller can be thought of in the same terms as model-based control algorithms, whereby the neural network is used to obtain the inverse of the process model. A back-propagation network can be trained to obtain the inverse model by considering load and controlled variables in its input vector and manipulated variables in its output vector. An example of a neural network controller on a distillation tower control application is shown in Figure 8.21e.

Building the Neural Model To build such a model, all inputs and outputs must first be normalized based upon expected minimum and maximum values and are presented to the network as the training set. All weights and processing element offsets are initially set to small random values. A recursive algorithm starting at the output processing elements

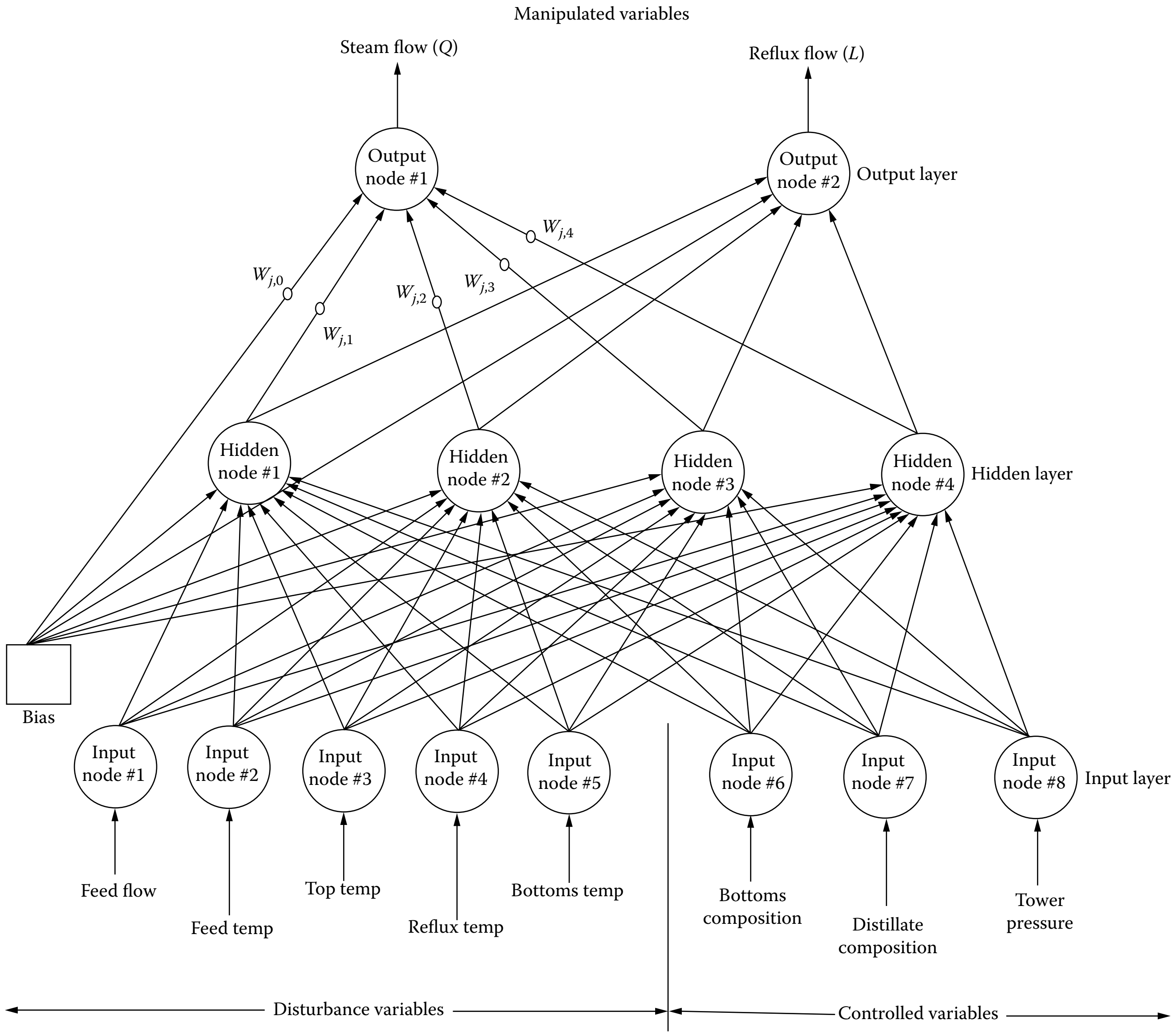


FIG. 8.21c
Back-propagation neural network.

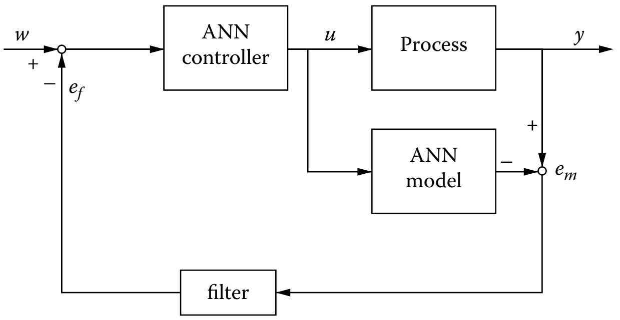


FIG. 8.21d
The configuration of artificial neural network (ANN) being used as an internal model controller (IMC).

is used and repeated until the input processing elements are reached. The weights are adjusted by

$$W_{ij}(t+1) = W_{ij}(t) + \eta \delta_j x_i \tag{8.21(3)}$$

where

W_{ij} (t) = the weight from hidden node i or from an input to node j at time t
x_i = either the output of node i or is an input
η = a gain term
δ_j = an error term for node j

The error term is

$$\delta_j = dE/dx_j \tag{8.21(4)}$$

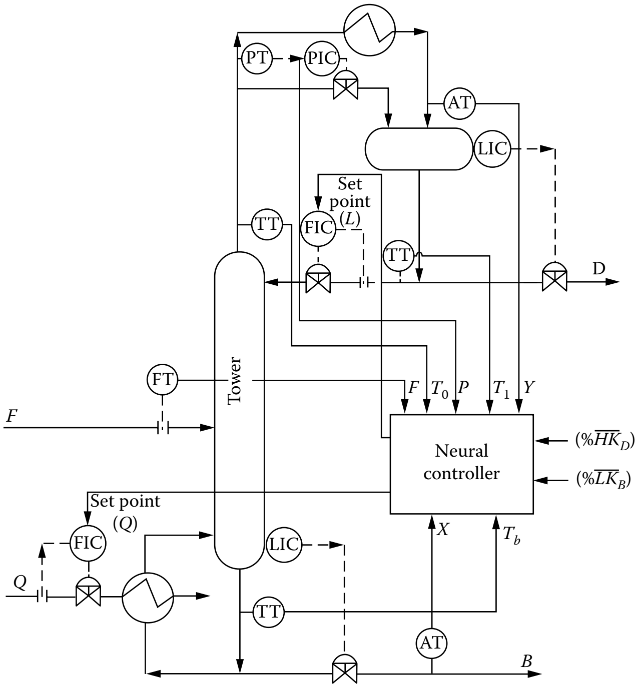


FIG. 8.21e
Fractionator control using a back-propagation neutral network.

where E is the difference between the desired output and the actual output.

A transfer function (also known as a squashing function) is applied to the weighted sum of the normalized inputs at each processing element to calculate each processing element's output. An often-used transfer function known as the sigmoid is

$$f(x) = 1/(1 + e^{-x})^{-1} \qquad \textbf{8.21(5)}$$

giving

$$dE/dx_j = x_j(1 - x_j) \qquad \textbf{8.21(6)}$$

If node j is an output node, then

$$\delta_j = y_i(1 - y_i)(d_j - v_i) \qquad \textbf{8.21(7)}$$

where d_j is the desired output of node j and y_i is the actual output.

If node j is a hidden node, then:

$$\delta_j = x_j(1 - x_j)\Sigma\delta_k w_{jk} \qquad \textbf{8.21(8)}$$

where k is over all nodes in the layers before node j.

Convergence is sometimes faster if a momentum term is added and weight changes are smoothed by a filter:

$$W_{ij}(t+1) = W_{ij}(t) + \eta\delta_j x_i + \alpha[W_{ij}(t) - W_{ij}(t-1)] \qquad \textbf{8.21(9)}$$

where $0 < \alpha < 1$.

As with any gradient descent method, back-propagation could find a local minimum instead of the global minimum. The momentum term is designed to help the training algorithm overcome the small valleys of local minima.

The learning procedures require that the change in weights be proportional to rate of change of error with respect to changes in weights. The constant of proportionality is called the learning rate, η (or learning coefficient). The larger the value of η, the faster the learning rate. Convergence is reached when the root mean square (RMS) error reaches a defined threshold value.

By using the same historical data required for the multivariable controller, the network can be trained and a nonlinear internal model can be created. In fact, the single neural controller is just a subset of the overall network used to build the entire fractionator model.

During the recall mode of operation, the network responds to the current values of the load and of the controlled variables by adjusting all manipulated variables accordingly. Each node sums the values of its weighted inputs and applies a transfer function. Thus, each output is attained by

$$I_j = \Sigma W_{ji}x_i \qquad \textbf{8.21(10)}$$

$$y_j = 1/(1 + e^{-Ij})^{-1} \qquad \textbf{8.21(11)}$$

The network's ability to do the prediction of the dynamics of the fractionator improves as more data become available for training. This approach assumes no explicit feedforward or feedback control actions because the control is totally integrated as part of the internally generated model (Figure 8.21e).

Thus, the neural controller can be considered a specific type of nonlinear, multivariable, model-based control algorithm. Instead of creating the nonlinear process model with explicit equations that are dependent upon various sets of assumptions (such as equimolal overflow, constant relative volatilities at differing conditions, and constant efficiencies), the neural controller builds its own process model from actual tower operation.

Because the neural controller is an empirical model as opposed to a theoretical model, it is susceptible to errors if operated outside the conditions of the training set. Data for the training set need to be continually gathered and the network retrained whenever novel conditions occur in order to increase the robustness of the neural controller throughout its life of operation.

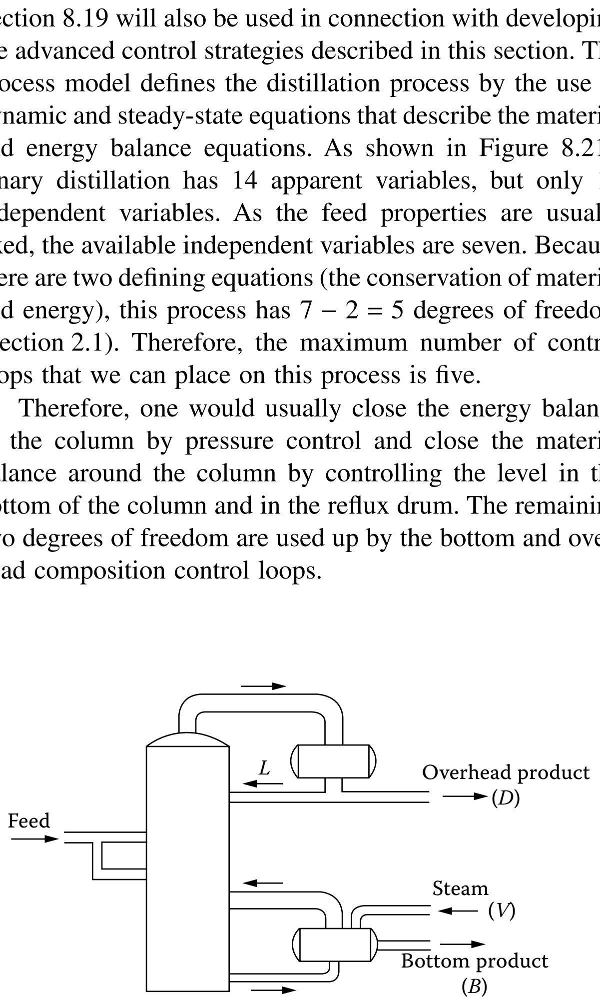
SISO CONTROL ADVANCES

Before proceeding to the subject of distillation optimization, some of the advances in single-input, single-output control strategies will be reviewed. These generally PID-based strategies involve the development of a process model and the use of feedforward and supervisory control techniques to achieve better control quality and localized optimization goals.

Process Model

The process model equations that were developed in Section 8.19 will also be used in connection with developing the advanced control strategies described in this section. The process model defines the distillation process by the use of dynamic and steady-state equations that describe the material and energy balance equations. As shown in Figure 8.21f, binary distillation has 14 apparent variables, but only 11 independent variables. As the feed properties are usually fixed, the available independent variables are seven. Because there are two defining equations (the conservation of material and energy), this process has 7 − 2 = 5 degrees of freedom (Section 2.1). Therefore, the maximum number of control loops that we can place on this process is five.

Therefore, one would usually close the energy balance of the column by pressure control and close the material balance around the column by controlling the level in the bottom of the column and in the reflux drum. The remaining two degrees of freedom are used up by the bottom and overhead composition control loops.



Apparent variables:		Independent variables
c_1	= Overhead temperature	2
c_2	= Overhead pressure	
c_3	= Overhead composition	
c_4	= Overhead flow rate	1
u_1	= Bottom temperature	2
u_2	= Bottom pressure	
u_3	= Bottom composition	
u_4	= Bottom flow rate	1
u_5	= Feed temperature	2
u_6	= Feed pressure	
u_7	= Feed composition	
u_8	= Feed percent vapor	1
u_9	= Feed flow rate	1
m	= Steam flow rate (heat input)	1
		11

FIG. 8.21f
A binary distillation process has five degrees of freedom, and therefore five of its process variables can be independently controlled: one pressure, two levels, and two compositions.

Control Equations Listed below are some of the key material and energy balance equations that define the distillation model, as they have been developed in Section 8.19:

$$F = D + B \tag{8.21(12)}$$

If the feed flow is uncontrolled, B is dependent upon F and D:

$$B = F - D \tag{8.21(13)}$$

or if the bottoms product is the manipulated variable:

$$D = F - B \tag{8.21(14)}$$

where
F = feed rate (the inflow)
D = overhead rate (an outflow)
B = bottoms rate (an outflow)

If the compositions of the feed, distillate product, and bottoms product are all known, then the component materials balance can be solved:

$$B/F = \frac{(100 - \%HK_D - \%LLK_F - \%LK_F)}{(100 - \%HK_D - \%LK_B)} \tag{8.21(15)}$$

For a given feed composition and desired product compositions, only one bottoms-to-feed ratio, *B*/*F* (product split), will satisfy both the overall and the component material balances.

A series of energy balances produce additional equations. The vapor boil-up rate V_B equals the heat Q_B added by the reboiler divided by the heat of vaporization (ΔH) of the bottoms product:

$$V_B = Q_B/\Delta H \tag{8.21(16)}$$

The vapor rate V above the feed tray equals the vapor boil-up rate plus the vapor entering with the feed (feed rate F times vapor fraction V_F, provided the feed is neither subcooled nor superheated):

$$V = V_B + (F)(V_F) \tag{8.21(17)}$$

The internal reflux rate (L_i), or the liquid at the top tray, of the column is derived by a heat balance around the top of the tower. If a total condenser is employed, this gives the equations:

$$L = \frac{L_i}{[1 + K_1(T_o - T_r)]} \tag{8.21(18)}$$

or

$$L_i = (L) \cdot [1 + (K_r)(\Delta T)] \qquad \textbf{8.21(19)}$$

The liquid rate, L_f, below the feed tray equals the internal reflux plus the liquid in the feed:

$$L_f = L_i + (1 - V_F)(F) \qquad \textbf{8.21(20)}$$

The distillate rate, D, equals the vapor rate, V, minus the internal reflux:

$$D = V - L_i \qquad \textbf{8.21(21)}$$

The bottoms rate, B, equals the liquid rate, L_f, minus the boil-up, V_B:

$$B = L_f - V_B \qquad \textbf{8.21(22)}$$

The criterion for separation is the ratio of reflux (L) to distillate (D) vs. the ratio of boil-up (V) to bottoms (B). Manipulating reflux affects separation equally as well as manipulating boil-up, albeit in opposite directions. Thus, for a two-product tower, two equations define the process: One is an equation describing separation and the other is an equation for material balance.

During the unsteady state of upsets, the process model must account for the dynamics of the process. This extends the steady-state internal flow model and requires additional consideration. For this reason two dynamic terms, G_T and G_B, are included, which provides a dynamic model for the tower based on its dead time and second-order lag, giving

$$L = G_B[G_T L_i + (1 - V_F) \cdot F] \qquad \textbf{8.21 (23)}$$

where

$$G_B = ta\,(1 - e^{-t_1})(1 - e^{-t_2})$$

$$G_T = tb\,(1 - e^{-t_3})(1 - e^{-t_4})$$

where
ta, tb = dead times

Scaling When using process models, it is very important that the measurements be correctly represented, that all I/O values be properly scaled. This was less of a challenge in the analog age, when a 9 PSIG or a 12 mA signal always meant 50%, no matter which supplier, industry, or continent was involved. This is not necessarily the case in the present digital age, with its multiple protocols and the need for interfacing translators when connecting them.

Scaling (the conversion from engineering units to fractions or percentages) is required in order to make the various transmitter signals meaningful to the DCS, PLC, or other central control system. A simple example of this type of conversion has already been given for zero-based signals in connection with Figure 8.19l in Section 8.19. In this section scaling will be illustrated on more complex systems, involving several nonzero-based transmitter signals.

The value of a transmitted signal in engineering units can be obtained from the normalized (scaled) transmitter signal and from the zero and range of the transmitter as follows:

$$\text{measurement in engineering units} = \text{zero} + \text{range (\% signal)} \qquad \textbf{8.21(24)}$$

Inversely, the percentage transmitter signal (scaled equivalent) corresponding to an engineering measurement can be obtained as

$$\text{\% signal} = \frac{\text{measurement in engineering units} - \text{zero}}{\text{range}} \qquad \textbf{8.21(25)}$$

Most DCS and other electronic controllers require that all signals conform to 0–100%. As mentioned previously, scaling is done to convert engineering unit inputs and outputs into normalized values that these DCS and electronic systems can use. Figure 8.21g shows how the internal reflux rate of a distillation column is calculated.

The calculation for internal reflux is given by the equation below:

$$L_I = L.[1 + (CP/\Delta H)\Delta T\,] \qquad \textbf{8.21(26)}$$

The subtracter is scaled first. Assuming $\Delta T_{\max}$ of 50°F (27.8°C), the span of T_o between 150°F and 250°F (65.6°C and 121°C), and the span of T_r between 125°F and 225°F (51.7°C and 107°C), the equation for the subtracter is written first in engineering units as

$$\Delta T = T_o - T_r \qquad \textbf{8.21(27)}$$

Now, converting from engineering to scaled units and denoting the scaled transmitter signal values as T_o' and T_r', the scaled equivalent of Equation 8.21(27) is

$$0 + 50\,\Delta T' = (150 + 100\,T_o') - (125 + 100\,T_r') \qquad \textbf{8.21(28)}$$

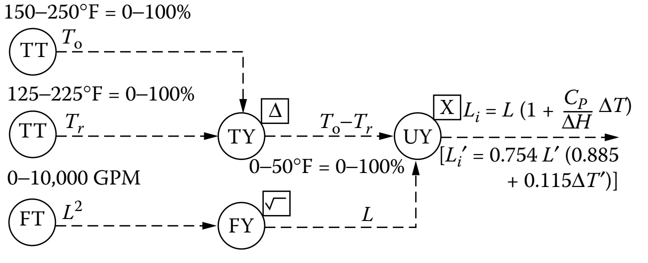


FIG. 8.21g
The steps in the calculation required to determine the internal reflux flow rate of a distillation column.

This reduces to the scaled equations:

$$\Delta T' = 2(T_o' - T_r' + 0.25) \qquad \mathbf{8.21(29)}$$

If the following assumptions are made:

$L_{I\,\max} = 15{,}000$ gpm ($0.95\ m^3/s$)
$L_{\max} = 10{,}000$ gpm ($0.63\ m^3/s$)
$Cp = 0.65$ BTU/lb°F (0.65 kcal/kg°C)
$\Delta H = 250$ BTU/lb (450 kcal/kg)

the equation for the multiplier then becomes

$$L_i' = \frac{10{,}000L'}{15{,}000}\left[1+\left(\frac{0.65}{250}\right)50\,\Delta T'\right] \qquad \mathbf{8.21(30)}$$

Equation 8.21(30) then reduces to

$$L_i' = 0.667L'(1 + 0.13\Delta T') \qquad \mathbf{8.21(31)}$$

When $\Delta T'$ is zero, the internal reflux equals 0.667 times the external reflux. The number 1 within the parentheses, therefore, sets the minimum internal reflux. When $\Delta T'$ is 100%, the ratio of internal reflux to external reflux is at a maximum.

The expression within the parentheses must be normalized. This is done by dividing both terms by the total numerical value, that is 1.13. To preserve the equality, the coefficient of L' is multiplied by 1.13. The scaled equation becomes

$$L_i' = 0.754L'(0.885 + 0.115\,\Delta T') \qquad \mathbf{8.21(32)}$$

Internal reflux systems are designed to compensate for changes outside the column, such as reflux temperature that is affected by ambient conditions. It should be understood that a change within the column can introduce positive feedback. Figure 8.21h shows a typical internal reflux application and its response to an upset within the column. The control system reacts in the same way to an increase in overhead vapor temperature and to a decrease in reflux liquid temperature, but the required control actions are in the opposite direction.

FEEDFORWARD SYSTEMS

Feedforward controls represented the first steps on the road towards multivariable model-based process control. They were first applied in well-understood processes such as heat transfer and distillation, where the material and heat balance equations made it possible to predict and anticipate the consequences of the process outputs to changes in the inputs, before they had time to evolve.

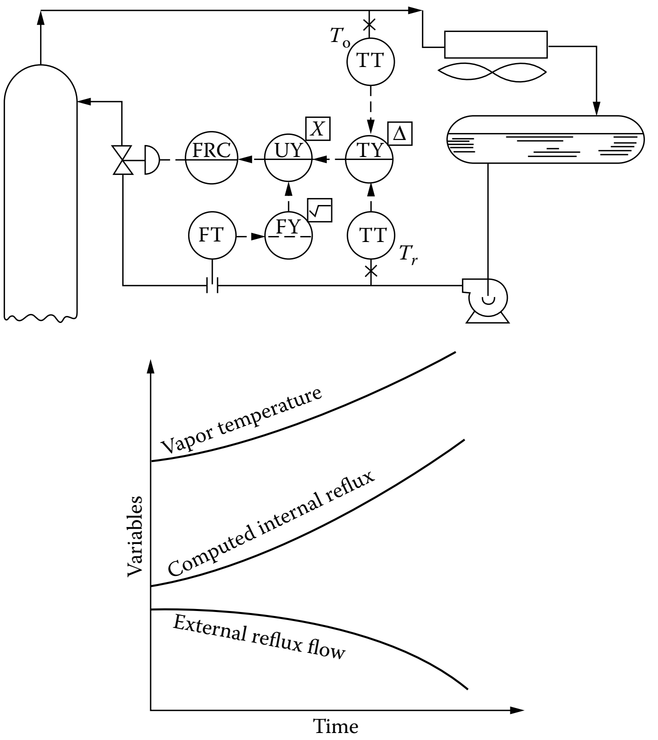


FIG. 8.21h
Response of internal reflux control system to an increase in the concentration of heavy components in the overhead vapors.

Feedforward control techniques react to variations in disturbance variables, predict the disturbance's effects, and take corrective action before the tower is significantly affected. Feedback control attempts to maintain the set point of a controlled variable by measuring its value at the outlet of the tower. In most cases, a combination of feedforward and feedback techniques can correct process deviations in the shortest time. This correction is accomplished by considering process dynamics (dead times and time lags), the nonlinearities between separation efficiency and column loading, loop interactions, and process measurements.

The types of disturbances that feedforward control is most often used to compensate for include (1) feed flow rate, (2) ambient temperature (top and reflux temperatures), and (3) reflux flow rate. Other variables that can be compensated for, but to a lesser degree include the disturbances are (4) tower pressure, (5) feed composition, (6) feed temperature or enthalpy, and (7) reboiler heat. The application of feedforward techniques involves the use of the models and equations described in Section 8.19, but dynamically tuned to approximate the response of the distillation tower.

Literally dozens of different feedforward control strategies have been proposed for distillation column control, and many of the more successful ones will be described and analyzed in this section. While feedforward control is common, it cannot be considered to be a universal solution for all columns.

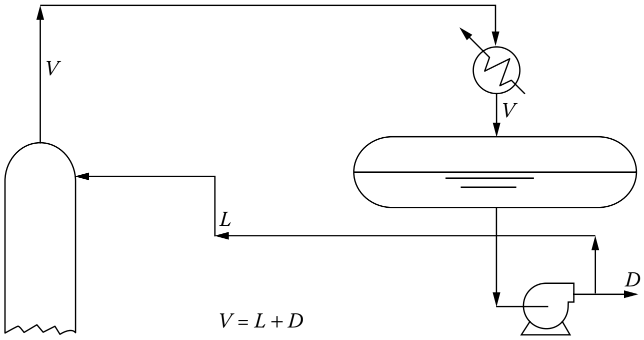


FIG. 8.21i
Reflux accumulator material balance.

Flow Control of Distillate

The column interactions that otherwise might necessitate the use of an internal reflux control system can be eliminated in some cases when the flow of distillate product draw-off is controlled and reflux is put under accumulator level control. This is a slower system than one in which flow controls the reflux, and its response is not always adequate. If necessary, the response can be speeded up by reduction of the accumulator lag.[4]

The steady-state material balance around the accumulator (Figure 8.21i) is expressed by

$$V = L + D \qquad \textbf{8.21(33)}$$

where
- V = boil-up (vapor rate)
- L = reflux rate
- D = distillate rate

To overcome the accumulator lag, the reflux rate, L, must be manipulated in direct response to a change in distillate rate, D, rather than by waiting for the response of a level controller. If V is constant (k), Equation 8.21(33) can be solved for L, which is the manipulated variable in this part of the system.

$$L = k - D \qquad \textbf{8.21(34)}$$

For this equation to be satisfied, L must be decreased one unit for every unit D is increased, and vice versa.

If V is indeed constant and both the computations and the flow manipulations are perfectly accurate, no level controller is needed. If these conditions cannot be met, a trimming function is introduced. The system equation becomes

$$L = m - KD \qquad \textbf{8.21(35)}$$

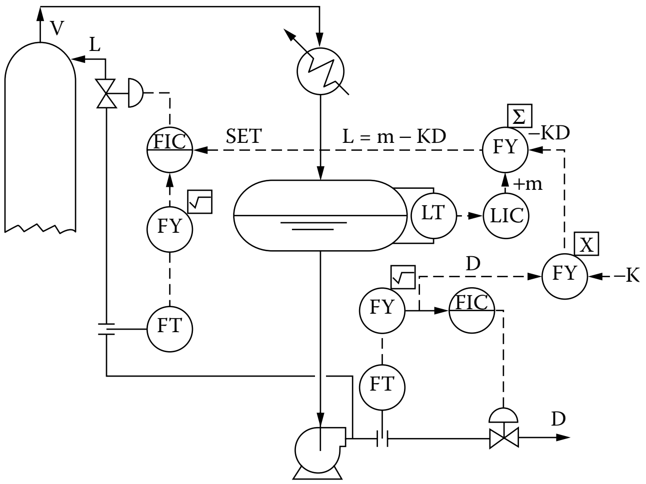


FIG. 8.21j
Reflux rate control system for overcoming accumulator lag.

where m is the output of the level controller and K is an adjustable coefficient. The resulting control system is shown in Figure 8.21j.

The range of coefficient K should be broad enough to allow scaling and adjustment to be done during commissioning. The level controller trims the computation, the scaling and the value of K but do not alter the steady-state value of external reflux, L, because these factors affect the transient response only. The response of the reflux flow to changes in distillate for several values of K is given in Figure 8.21k. The full-scale values of reflux, L_{max}, and distillate, D_{max}, flows in this case are 1000 gpm (3.79 m^3/min) and 500 gpm (1.89 m^3/min), respectively.

When $K = 0$, the reflux is adjusted by the level controller. In other cases, the reflux flow is immediately altered by some percentage for a change in distillate, and the level controller forces the balance of the change. The response is a first-order lag.

If $K = 0.5$, the reflux flow is changed to the exact new steady-state value, because K equals the ratio of D_{max}/L_{max}, and therefore the computation is exact; the lead equals the lag and the net effect is no dynamic contribution. If $K = 1.0$,

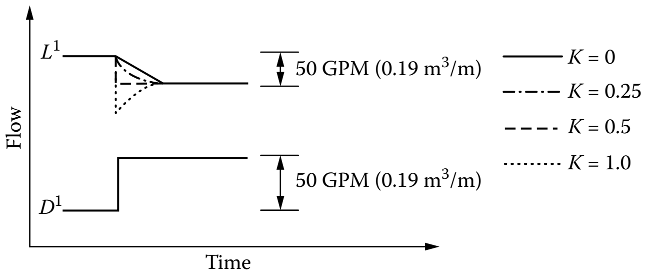


FIG. 8.21k
Reflux response as function of K.

the initial response is a first-order lead-lag function. In this case, the reflux is greater than required for the new steady state, and the level controller eventually corrects the flow.

The value of K does not change the steady-state flow. It affects the transient response only, and therefore it can be used to adjust the dynamics of the loop. The greater the value of K, the faster the response. Care must be taken to prevent increasing the response to the point of instability. A rule of thumb is

$$K_{\max} = 1.5(D_{\max}/L_{\max}) \qquad \textbf{8.21(36)}$$

Therefore, in this example, K should not be set greater than 0.75.

In some implementations, the range of adjustability of K is limited, and scaling is necessary. For the values used in the illustration (reflux full scale value equals 1000 gpm and distillate full scale equals 500 gpm), Equation 8.21(35) becomes

$$1000L' = 1000m' - 500KD' \qquad \textbf{8.21(37)}$$

where L' and D' are the normalized values of L and D. The maximum value of m is equal to the maximum value of L, because the level controller by itself can cause the level control valve to open fully.

The scaled equation is

$$L' = m' - 0.5KD' \qquad \textbf{8.21(38)}$$

K must be adjustable over a range of ±10% for satisfactory tuning flexibility.

Flow Control of Bottoms

Similarly to the feedforward systems described for distillate flow control, a similar system can be used on control the column bottom flows, if the bottom product is flow controlled and the bottoms level is maintained by manipulation of the heat input or boil-up (V). The equation for that system is

$$V = m - KB \qquad \textbf{8.21(39)}$$

where V is the boil-up, B is the bottom product flow, m is the output of the bottoms level controller, and K is the same kind of adjustable coefficient as in Equation 8.21(38).

Another commonly used model for feedforward compensation involves the bottoms-to-feed ratio (B/F). The bottoms product draw, ratioed with the feed rate, is a function of the overhead and bottoms composition targets at a given feed composition. The implementation of the bottoms-to-feed ratio control usually requires dynamic (dead time and lag) compensation of the feed rate. The dynamically compensated (lagged) feed rate, F_L, is then multiplied by the desired bottoms-to-feed ratio to obtain the target for bottoms flow rate.

$$B = (B/F)(F_L) \qquad \textbf{8.21(40)}$$

so that

$$V = m - K(B/F)(F_L) \qquad \textbf{8.21(41)}$$

Because these models are only approximations of the real process, inaccuracies do exist. For this reason, the bottoms-to-feed ratio target obtained by the feedforward calculation should be trimmed by analysis-based feedback control. In the majority of feedforward applications, their purpose is not to replace feedback but to minimize the amount of work that the feedback part of the loop has to do. This requires that the advanced control must be able to measure and quantify the disturbance, then react before the fractionator separation can be upset in the first place.

Constant Separation

A distillation column operating under constant separation conditions has one fewer degree of freedom than others, because its energy-to-feed ratio is constant. At a given separation, for each concentration of the key component in the distillate, a corresponding concentration exists in the bottoms.

In other words, for a constant-feed composition, holding the concentration of a component constant in one product stream fixes it in the other.[4] Figure 8.21l shows an example of a constant separation feedforward system in which distillate is the manipulated variable.

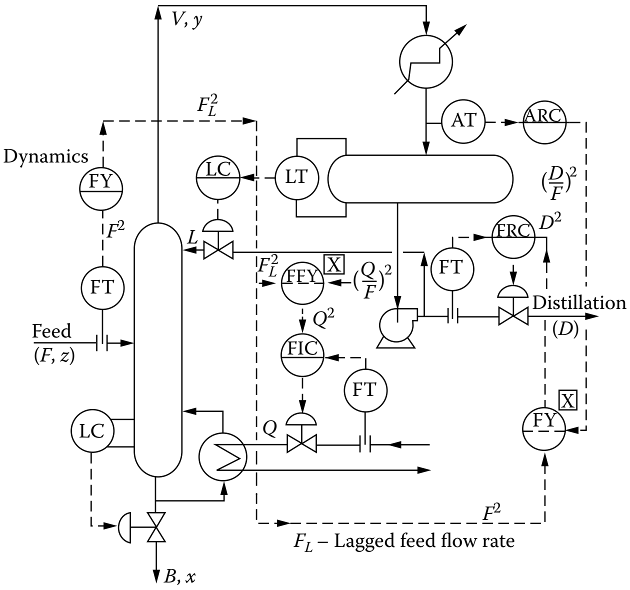


FIG. 8.21l
Feedforward distillation control system with constant separation.

A material balance on the light key component gives

$$Fz = Dy + Bx = Dy + (F - D)x \tag{8.21(42)}$$

$$D = F\left(\frac{z-x}{y-x}\right) = F\left(\frac{D}{F}\right) \tag{8.21(43)}$$

If the flow measurements are of the differential pressure type, then:

$$D^2 = F^2\left(\frac{z-x}{y-x}\right)^2 = F^2\left(\frac{D}{F}\right)^2 \tag{8.21(44)}$$

Because boil-up must change in proportion to feed rate, a second feedforward loop is obtained for setting heat input:

$$Q = F(Q/F) \quad or \quad Q^2 = F^2[Q/F]^2 \tag{8.21(45)}$$

where

z, y, x = mole fraction of the key light component in feed, overheads, and bottoms, respectively
D/F = required distillate-to-feed ratio
Q/F = required energy-to-feed ratio

No scaling is required of this equation if an adjustable ratio is used for both D/F and Q/F.

Normal design practice for scaled systems calls for the output of the trim analyzer controller ARC to be at 50% when the design or normal distillate-to-feed ratio is required. If the gain of the multiplier is set at 2, the output tracks the load when this normal distillate-to-feed ratio occurs.

In a linear system, the gain of the multiplier equals the scaling factor. In this system, however, the gain of the multiplier equals the square root of the scaling factor. When this rule is applied to the example, the scaled form of Equation 8.21(44) is

$$D^{2\prime} = 4.0(F^2)'[(D/F)^2]' \tag{8.21(46)}$$

where $D^{2\prime}$, $F^{2\prime}$, and $[(D/F)^2]'$ are the normalized values of the respective terms in Equation 8.21(44).

The block labeled "dynamics" in Figure 8.21l is a special module designed to influence the transient response. This is because the time response of the distillate to a feed rate change must be dynamically matched. The dynamic block is generally a dead time module and a lead-lag module in series. In the steady state, its output equals its input. Figure 8.21m illustrates the temporary modifications that various dynamic compensators can introduce to match the "dynamic personality" of the process. For a discussion of dead time compensation, refer to Section 2.9 in Chapter 2.

Maximum Recovery

In many distillations, one product is worth much more than the other, and the control system is designed to maximize the recovery of the more valuable stream. One such equation for this type of system is[4]

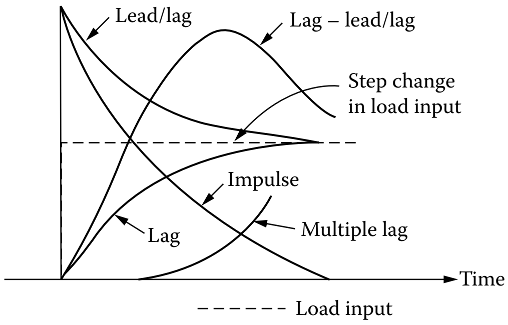


FIG. 8.21m
Dynamic compensators introduce temporary modifications into the value of their output signals which match the dynamic "personality" of the controlled process.

$$D = m(KF + K_2F^2) \tag{8.21(47)}$$

where

D = distillate rate
F = feed rate
K = adjustable coefficient
$K_2 = 1 - K$
m = feedback trim

This equation assumes that energy is free and that the distillate product is worth more than the bottoms. Distillate product flow is not linear with feed rate when boil-up is held constant. The control diagram for this maximum recovery system is shown in Figure 8.21n. Note that the distillate-to-reflux loop for accelerated response is also used. The summing block (FY-1) used to compute $(KF + K_2F^2)$ needs no special scaling.

The values of m can be computed from the feed composition. A typical range for m is 0.35–0.65. This is the output signal range of ARC-2, the feedback controller. Although the coefficients can be calculated in advance with reasonable accuracy, on-line adjustment is quite easy (these coefficients are accessible in most DCS and PLC systems), and the rigor of the calculations can be avoided.

If energy is not free and only one product composition needs to be controlled, then a linear relationship can be assumed. In this case, product flows will be directly proportional to feed rate when separation is fixed.

$$D = m_1(K_3F) \tag{8.21(48)}$$

or

$$B = m_2(K_4F) \tag{8.21(49)}$$

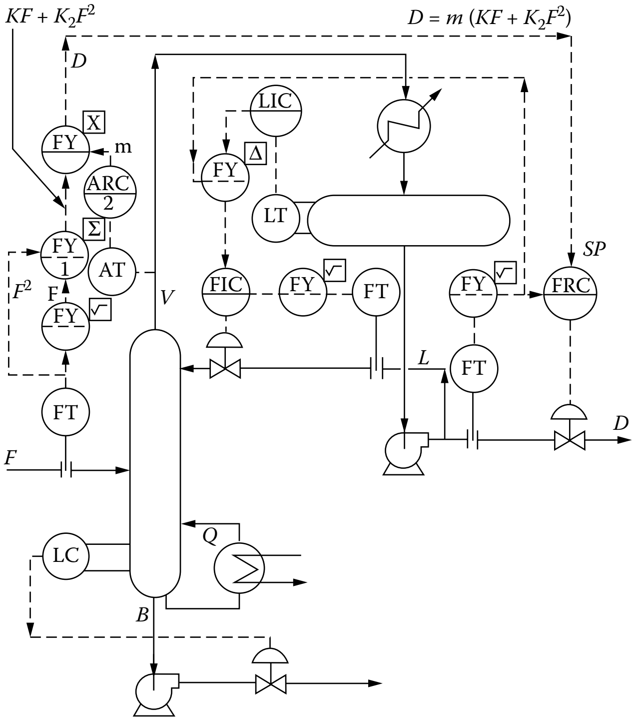


FIG. 8.21n

Maximum recovery system: instrumentation solves quadratic equation for distillate rate.

and

$$Q = m_3(K_5F) \tag{8.21(50)}$$

or

$$L = m_4(K_6F) \tag{8.21(51)}$$

where

D = distillate rate
B = bottoms flow rate
F = feed rate
Q = heat input rate
L = reflux rate
K_3, K_4, K_5, K_6 = adjustable coefficients
m_1, m_2, m_3, m_4 = feedback trim signals

Composition Control of Two Products

Because of the many variables that affect product composition, which are difficult to anticipate or control (e.g., feed composition), and because composition specifications for both products may be tight, some columns require better control than can be achieved by the previous constant separation strategy.

One method that can be used on some columns for achieving the required product specifications is to directly control the compositions of both products. One benefit of dual composition control is minimized energy consumption. However, it is difficult to implement dual composition control on many columns because of the severe interaction problems that may exist.

Also, with a given feed and tower design, it may not be possible to achieve two arbitrarily chosen product compositions.

An example of a feedforward dual-composition control model will be described here, after which a method for determining the degree of interaction, based on actual process data, will be discussed.

The control of distillate composition can still be done by manipulating distillate flow as required by

$$D = F\left(\frac{z-x}{y-x}\right) \tag{8.22(52)}$$

However, in order to also enforce composition control of the bottom product, an additional manipulated variable is needed. Another product stream cannot be independently manipulated without changing the accumulation in the column, which is not practical. The energy balance must, therefore, be adjusted to control bottoms composition x.

The relationship between x and the energy balance was developed by Shinskey,[5] for binary mixtures, as a function of separation S:

$$S = \frac{y(1-x)}{x(1-y)} \tag{8.21(53)}$$

For multicomponent mixtures separation is defined as:

$$S = \frac{y_L/x_L}{x_H/y_H} = \frac{y_L/x_H}{x_L/\,y_H} \tag{8.21(54)}$$

where the separation factor is the ratio of light to heavy key in the distillate divided by the same ratio in the bottoms product.

The relationship between separation (S) and the ratio of boil-up to feed (V/F) over a reasonable operating range is

$$V/F = a + bS \tag{8.21(55)}$$

where a and b are functions of the relative volatility, the number of trays, the feed composition, and the minimum V/F. The control system therefore computes V based on the equation for a binary mixture as

$$V = F\left(a + b\frac{y(1-x)}{x(1-y)}\right) \tag{8.21(56)}$$

Because y is held constant, the bottom composition controller adjusts the value of the parenthetical expression if an

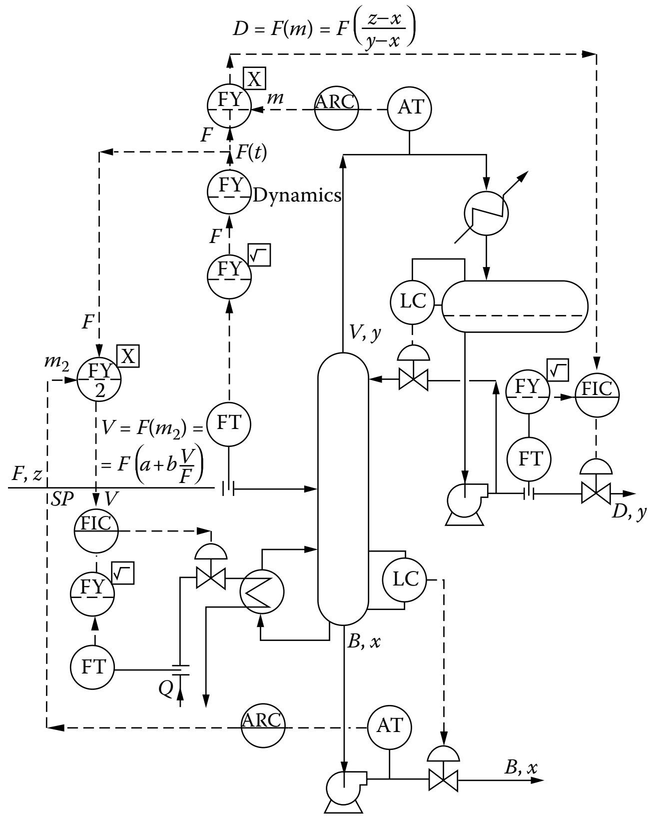


FIG. 8.21o
Feedforward control system provides closed-loop composition control of two product streams.

error should appear in x. Let $V/F = y(1 - x)/(1 - y)$, and the control equation becomes

$$V = F(a + b[V/F]) \qquad \textbf{8.21(57)}$$

where $[V/F]$ equals the desired ratio of boil-up to feed.

The system implementing Equation 8.21(52) is shown in Figure 8.21o. FY-1 and FY-2 are multipliers. The FY block labeled "dynamics" is a special block for dynamic compensation similar to the one described in Figure 8.21l and 8.21m.

Included in a and b are the relationship between boil-up (vapor rate), and energy flow (Q), and the minimum ratio of boil-up to feed. Equation 8.21(57) can therefore be written

$$Q = kF([V/F]_{\min} + [V/F]) \qquad \textbf{8.21(58)}$$

where k represents the proportionally constant.

Two Products with Interaction

Interaction always exists between the material and energy balances in a distillation column. In some columns, this interaction is not severe enough to impede closed-loop composition position control of the two product streams, but in others it is. The severity is a function of feed composition, product specification, and the pairing of manipulated and controlled variables.

Severe interactions frequency occur when the energy balance is manipulated by two independent composition controllers. A column in which reflux flow and steam flow are the manipulated variables is an example of a severely interacting column. The control system equations are

$$Q = kF([V/F]_{\min} + [V/F]) \qquad \textbf{8.21(59)}$$

$$L = F([L/F]) \qquad \textbf{8.21(60)}$$

where L is the reflux rate and $[L/F]$ is the desired reflux-to-feed ratio.

Note that in the control system described by these two equations, the rates of products leaving the column are dependent on two energy balance terms. Increasing heat input at the reboiler forces the composition controller that is resetting reflux flow to increase heat withdrawal, and the top and bottom composition controllers, therefore, "fight" each other. The only way to avoid this fighting is by preventing a change at one end of the column from upsetting the other end.

The heat input is changed when the bottom composition controller is upset. If the upset is because of a high concentration of light ends in the bottom product, heat is increased to adjust the separation being performed and to drive the extra light ends up and out the top. The top composition controller does not know how to split the increased vapor load, but it sees a measurement indicating an upset and responds to an increase in heat input by increasing the reflux flow. Theoretically, if the reflux rate is compensated for the change in heat input, the top composition controller upset can be avoided.

One can find the relationship between reflux L and heat input Q by solving Equations 8.21(59) and 8.21(60) for L in terms of Q. The resultant equation is of the form

$$L = k_1Q - k_2F \qquad \textbf{8.21(61)}$$

The values of k_1 and k_2 are found by deriving Equation 8.21(61) using actual process values of $[L/F]$, $[V/F]_{\min}$, and $[V/F]$.

The decoupling equation, Equation 8.21(61), replaces Equation 8.21(60) in the control model. The resulting system is shown in Figure 8.21p.

The system is now half-decoupled: A change in heat input will not upset the top temperature, because the decoupling loop adjusts the reflux independently of the top temperature (analysis) controller. However, the heat input is still coupled to reflux, because a change in reflux will still cause the bottom temperature controller to adjust steam flow. This degree of decoupling is enough to reduce the interaction approximately 20-fold. The two multipliers are scaled as described previously (under the paragraph Scaling), and the adder is tuned on-line.

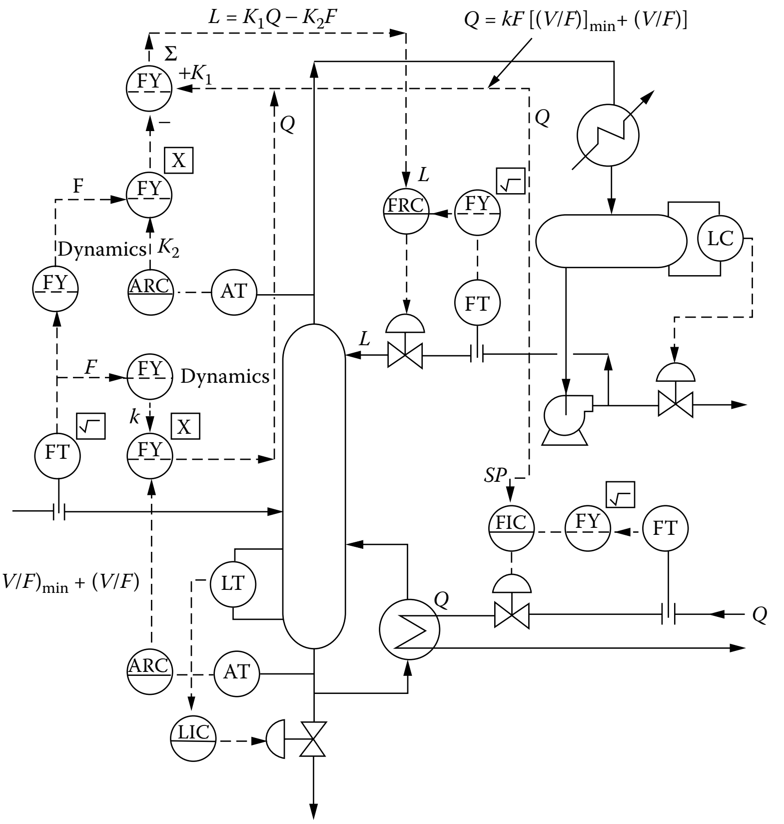


FIG. 8.21p
The decoupling of a distillation tower when the compositions of both products are being controlled.

Classical decoupling schemes,[6] however, often do not provide a solution to the problem of interaction because of practical problems encountered on real columns.[7,8] Decoupling systems that include overrides can drive to saturation when constraints are encountered. Most seriously, decouplers applied to systems with negative interaction (defined later) may have very little tolerance for errors in decoupler gains.

For this group, which always includes the interaction encountered in reflux and boil-up controls, small errors can transform a system that provides complete decoupling into one that provides no control at all. Because the proper decoupler gains depend on the process gains, which inevitably change with variations in feed rate, product specifications, and column characteristics, these systems require constant attention and adjustment beyond the ordinary capability of plant operating personnel.

The difficulties associated with the application of decoupling systems have prompted a re-examination of interaction itself. The problem may be postulated in two ways:

1. For a given column, is the interaction equally strong in each of the control structures available to the designer?
2. For a given control structure, will the interaction be equally strong in every column in which it is applied?

The stumbling block of loop assignment may in this way be converted into a stepping stone by providing the opportunity to select a control structure that will exhibit minimum interaction in any particular application.

Shinskey and Ryskamp have given consistent guidelines for assigning loops to minimize loop interaction. Shinskey[9] suggests that the controller assigned to the more pure product should manipulate separation. Ryskamp[10] suggests that the controller for the component with the shorter residence time should adjust vapor flow, and the controller for the component with the longer residence time should adjust the liquid/vapor ratio.

Feed Composition Compensation

Occasionally, changes in feed composition occur too fast to be handled by feedback control, and feedforward compensation for these changes is necessary (Figure 8.21q).

The basic material balance equation, Equation 8.21(62), already has a term, z, representing concentration of the key component in the feed:

$$D = F\left(\frac{z-x}{y-x}\right) \qquad \mathbf{8.21(62)}$$

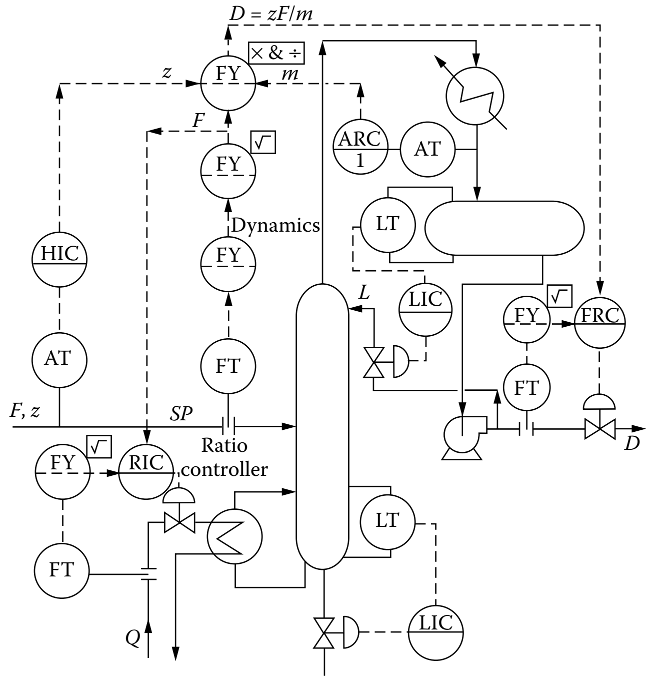


FIG. 8.21q
Feed composition measured and used to compute distillate flow.

When z is measured, the equation for distillate can be simplified to

$$D = zF/m \quad \textbf{8.21(63)}$$

where m is the output of the overhead analyzer feedback trim controller (ARC-1). The auto/manual station (HIC) is used in the event of analyzer failure. Dynamic compensation is placed on the flow feed signal only.

The control of the bottoms flow in Figure 8.21q is indirectly provided by the feedforward control of the reboiler heat input based on the (dynamically compensated) feed flow rate. If, instead of this approach, feedforward analyzer control of the bottoms flow is desired, Equation 8.21(64) can be utilized.

$$V = m - K(B/F)(F_L) \quad \textbf{8.21(64)}$$

where F_L is the dynamically compensated feed rate.

Substituting the B/F ratio:

$$B/F = \frac{(100 - \%HK_D - \%LLK_F - \%LK_F)}{(100 - \%HK_D - \%LK_B)} \quad \textbf{8.21(65)}$$

gives the feedforward expression for the vapor rate up the tower:

$$V = m - K\left[\frac{(100 - \%HK_D - \%LLK_F - \%LK_F)}{(100 - \%HK_D - \%LK_B)}\right]F_L \quad \textbf{8.21(66)}$$

SUPERVISORY CONTROL

On-line computer control can greatly enhance the profitability of the distillation process and data collection improvements, and increased flexibility can often justify computer control even if rigorous on-line optimization is not implemented. The optimization strategy can be implemented in the supervisory mode (recommendations to the operator) or in the automatic mode and can involve the whole plant or only particular subsystems of the total process.

The main computer control functions applied to distillation include engineering calculations, operating assistance, quality controls, and heat balance controls.[11] The primary engineering calculations are made from material and heat balances around column sections and include tray loadings, internal vapor flows, internal liquid flows, and heat duties. These calculations are helpful as operating guidelines and as inputs for on-line control. However, they are usually based on steady-state conditions, and therefore the input signals must be averaged to make the calculations.

Response, although normally fast enough for on-line control, may not be adequate if frequent, short-term disturbances must be handled. However, information gained from these types of calculations can often justify the computer system by providing better operating guidelines, even if it is not used for on-line control.

An example of a computerized control system that makes the operator's job easier is the balancing of the heater coil outlet temperature on furnaces (Figure 8.21r).

Other examples of computer controls include (1) the feedforward adjustments of products and pump-arounds on the basis of feed rates, (2) the control of the column's bottom level by throttling the feed preheater bypass flows, and (3) the control of overhead receiver level. Such control systems will be described in more detail later. Because many of these calculations result in the need for changing the set points of several controllers simultaneously, supervisory control can reduce the workload and the potential for human errors by the operator.

Product quality controls are enhanced if the computer adjusts the column temperature and side-draw flow rates to control product specification. Often computers are used to infer product specifications from local flows, temperatures, and pressures. Examples of these inferred calculations include true boiling point (TBP) cut points, ASTM 95% boiling points, Reid vapor pressure (RVP), octane, viscosity, freeze points, cloud points, and pour points.

These calculated measurements can be used in feedback controllers themselves or as a fast inner loop with an analyzer trim. The advantage of such model-based controls is that they can anticipate future events, because their outputs are not delayed by the process dead time and time lag. This results in tighter control.

Adjustment of pump-around reflux flows, as shown in Figure 8.21s, is an application example where the computer

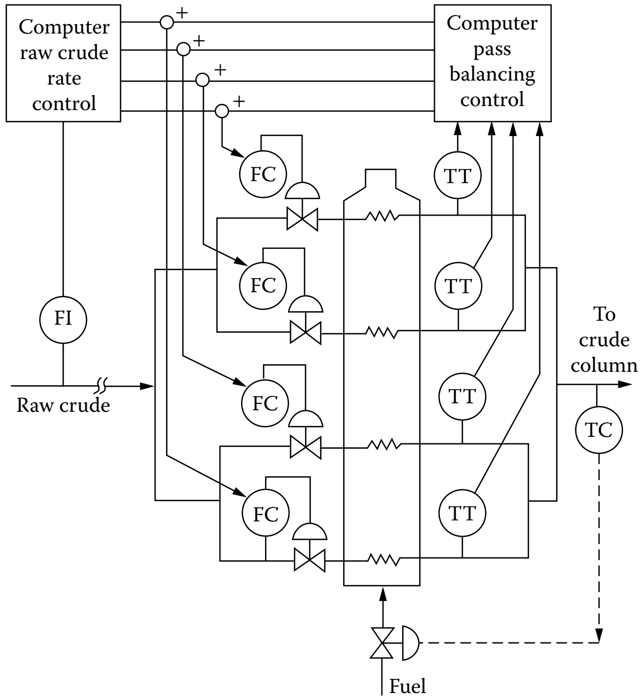


FIG. 8.21r
The balancing of heater coil charge rates in crude oil furnaces can be under computer control or supervision.

assists in heat balance control. The goal of such systems usually is to maximize the exchange of heat to feed, subject to certain limits,[19] which will be discussed in the subsection "Optimization."

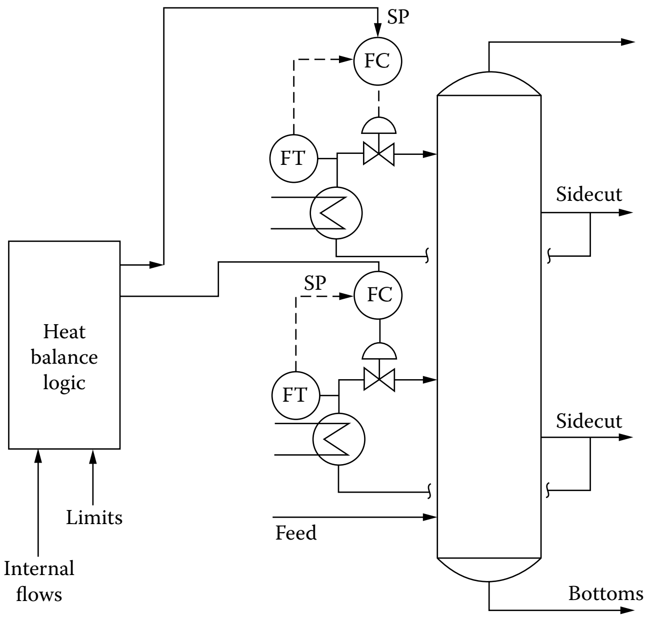


FIG. 8.21s
Computer adjustment of pumparound refluxes.

A careful analysis of the limits and operating constraints is essential, because if the system is not designed to provide limit checks and overrides to handle operating limits, frequent operator intervention will be required during upsets. This can cause a lack of confidence in the computer system, which can result in the column's being off computer control more than necessary.

The Total Model

It is possible to design a system to compensate for all load variables: feed rate, composition, enthalpy, reflux, and bottoms enthalpy. The goal of these systems is to overcome the problems associated with unfavorable interactions and to isolate the column from changes in ambient conditions. These problems can usually be solved by careful system analysis and variable pairing, thus avoiding complicated total energy and material balance control systems. The complexity of the total material and energy balance systems is made apparent by the list of equations required in the model:

Feed enthalpy balance
Bottoms enthalpy balance
Internal reflux computation
Reboiler heat balance
Overall material balance

Suboptimization

The following derivations will provide insight into the derivation and modeling of optimization equations and the handling of constraints. However, the reader is advised that every distillation column is unique; the examples given here are for illustrative purposes only and should not be considered to be suboptimal or optimal solutions for every column.

Optimization of a single distillation column normally implies a maximum profit operation, but to achieve maximum profit, the price of the column's products must be known. It is impossible to control every column in a system based on this criteria, because the prices of products for many columns are unknown. Product prices are often unknown because the products are feed streams to other units, whose operations would need to be taken into account to establish the column's product prices.

When product prices are unknown, it is possible to carry optimization only to the stage at which specified products can be produced for the least operating cost. This can be called an optimum with respect to the column involved, but only a suboptimum with respect to the system of which the column is a part.

When column product prices are known, complete economic optimization can be achieved. However, a number of different situations may still exist. If there is a limited market for the products, then the control problem is to establish the separation that results in a maximum profit rate. Such an optimum separation will be a function of all independent inputs to the column involved.

When an unlimited market exists for the products, and sufficient feedstock is available, the optimization problem becomes more difficult. Not only must the optimum separation be established, but also the value of feed must be determined. Optimization for this case results in operating the column at maximum loading or at maximum energy efficiency.

One of three possible constraints will be involved: Throughput will be limited by the overhead vapor condenser, the reboiler, or the column itself. In some cases, the constraint will change from time to time, depending upon product prices and other independent variables of the system. The design of optimal automatic control systems to single columns should follow three logical steps:

1. Design the basic controls to regulate basic functions, such as pressures, temperatures, levels, and flows
2. Configure the controls to regulate the main sources of heat inputs, including regulation of internal reflux flow rate, feed enthalpy, and reboiler heat flow rate
3. Apply controls to regulate the specified separation

A single column that is automated in this manner is called a suboptimized system. This suboptimum is defined as an operation that will produce close to the specified separation, whether or not that separation is ideal with regard to the total system of which the column is a part. If product purities are higher than specified, the operation cannot be considered suboptimum.

When a single column is automated through the suboptimum operation stage, it will still exhibit up to five degrees of freedom. As a basis for proceeding into the optimization phase, Figure 8.21t is presented as one example of a column automated through the suboptimization stage.

As shown in Figure 8.21t, the system used to regulate the separation is a predictive control system, similar to that described earlier. The function of the predictive control system is to manipulate the energy balance (reflux flow rate) and the material balance (bottom product flow rate) to give the specified separation. The equations derived for these manipulations are called the operating control equations.

Figure 8.21u is another example of suboptimization. Both figures achieve their goal of operating at suboptimum; the difference is mainly in their the basic controls. In the paragraphs below, the steady-state operating equations and the dynamics for bottom product and reflux will be described.

Bottoms Product Operating Equation The equation for predicting bottom product flow rate was derived in Section 8.19 as

$$B/F = \frac{(100 - \%HK_D - \%LLK_F - \%LK_F)}{(100 - \%HK_D - \%LK_B)} \qquad 8.21(67)$$

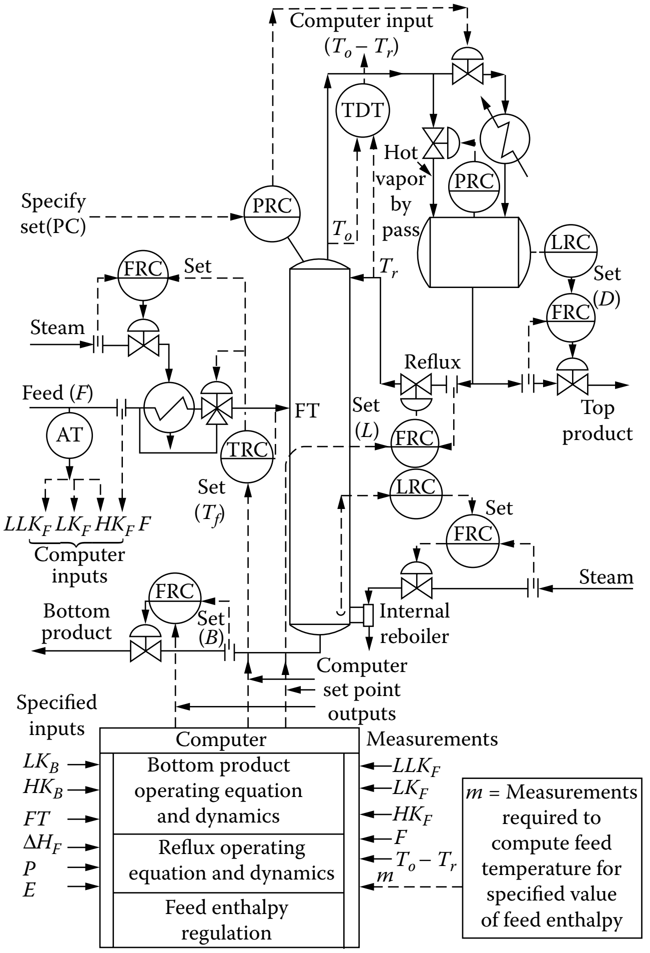


FIG. 8.21t
Distillation column automated through the suboptimization stage (to produce close to the specified separation.)

where

$\%LLK_F$ = lighter than light key in the feed (mol%)
$\%LK_F$ = light key in the feed (mol%) = z
$\%LLK_D$ = lighter than light key in the distillate product (mol%)
$\%LK_D$ = light key in the distillate product (mol%) = y
$\%HK_D$ = heavy key in the distillate product (mol%)
$\%LK_B$ = light key in the bottoms product (mol%) = x

Assuming that the composition in both ends of the tower are to be held at specification:

$$B/F = \frac{(100 - \%\overline{HK}_D - \%LLK_F - \%LK_F)}{(100 - \%\overline{HK}_D - \%\overline{LK}_B)} \qquad 8.21(68)$$

where

$\%\overline{HK}_D$ = specification of the heavy key in the distillate product (mol%)
$\%\overline{LK}_B$ = specification of the light key in the bottoms product (mol%) = x

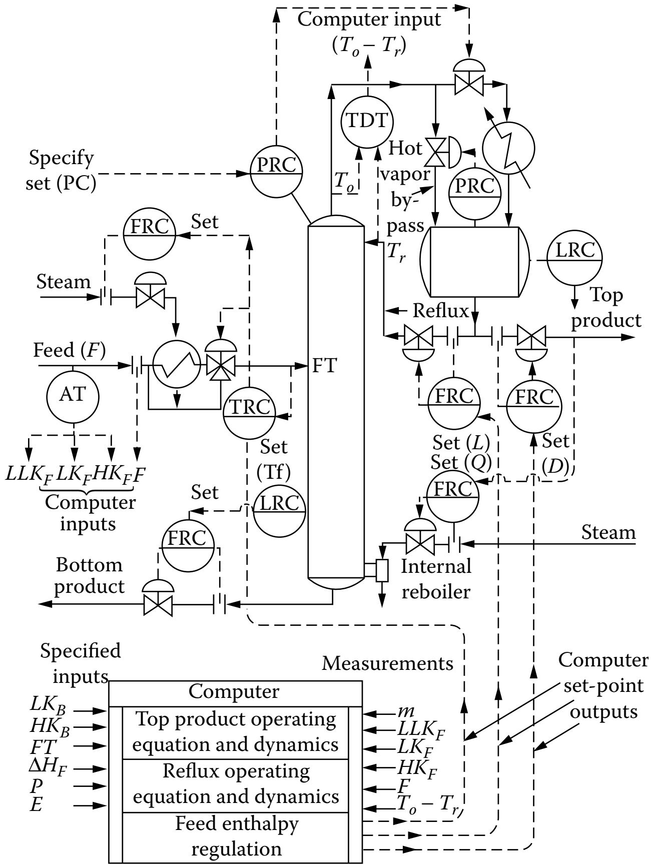


FIG. 8.21u

Alternative method of column automation through the suboptimization control stage.

Derivation of the internal reflux operating equation is more difficult. Typically, this equation is developed in two parts:

$$(L_i/F) = (L_i/F)t + (L_i/F)e \qquad \textbf{8.21(69)}$$

where

(L_i/F) = internal reflux to feed flow rate ratio required to give a specified separation
$(L_i/F)t$ = theoretical part of reflux operating equation
$(L_i/F)e$ = experimental part of reflux operating equation

The experimental part of this equation is necessary because the effect of loading on the overall separating efficiency (E) is normally unpredictable. Both parts of the reflux operating equation are functions of all independent inputs to the system. However, simplifications are normally considered for the experimental part, as follows:

$$(L_i/F)t = f_1[(LLK_F),(LK_F),(HK_F),(\bar{E}),(\overline{FT}),$$
$$(\overline{\Delta H}_F),(\bar{P}),(\%\overline{LK}_B),(\%\overline{HK}_D)] \qquad \textbf{8.21(70)}$$

$$(L_i/F)e = f_2(L_i) \qquad \textbf{8.21(71)}$$

where

$\bar{E}$ = specified constant average efficiency
$\overline{FT}$ = specified value of feed tray location
$\overline{\Delta H}_F$ = specified value of feed enthalpy
$\bar{P}$ = specified value of column pressure

The theoretical part of Equation 8.21(69) is normally developed by tray-to-tray runs of calculations performed by an off-line digital computer. A statistically designed set of runs is made, and the information thus obtained is curve-fitted to an assumed equation form.

Once the steady-state theoretical equation is developed and placed in service, the experimental part is determined by on-line tests. These tests involve operating the column at different loads to determine the correction required to $(L_i/F)t$ for the separation to be equal to that specified. Average overall efficiency $\bar{E}$ is set to make (L_i/F)t required to equal the actual L_i/F that exists. The loading tests are carried out under this condition.

Often plant tests are performed to determine L_i/F without consideration to the theoretical term. For a given distillate composition, the calculated internal reflux in Equation 8.21(72) is found for several different feed rates.

$$\frac{L_i}{L} = \frac{\Delta H_L}{\Delta H_{L_i}} \cdot \left[\left(1.0 + \frac{C_{p_L}}{\Delta H_L} \cdot (T_o - T_r)\right)\right] \qquad \textbf{8.21(72)}$$

The result often gives the relationship as shown in Figure 8.21v. Implementation of these curves is generally via a polynomial equation or segmented function curve.

Reflux Operating Equation Internal reflux controls were described in Figure 8.21h. One can approximate internal reflux flow rate of a distillation column by making a heat balance around the top tray. If that is done, the following equation is obtained:

$$(L_i) = LK_2[1 + K_1(T_o - T_r)] \qquad \textbf{8.21(73)}$$

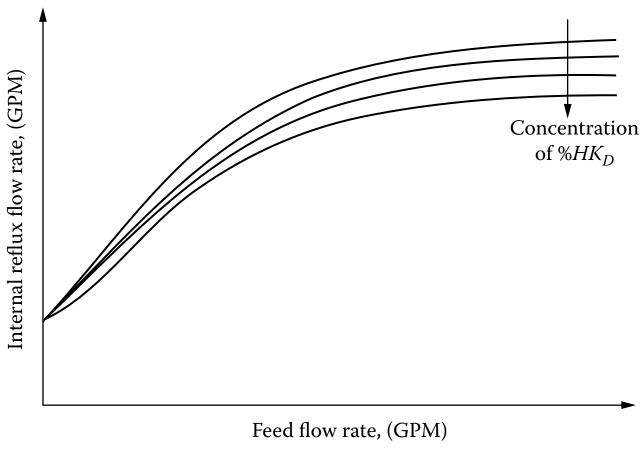


FIG. 8.21v

Relationship of internal reflux to feed at several distillate compositions.

Substitute this equation into Equation 8.21(69) to eliminate L_i:

$$(L/F) = \frac{[(L_i/F)_t + (L_i/F)_e]}{K_2[1 + K_1(T_o - T_r)]} \qquad \mathbf{8.21(74)}$$

where

L = external reflux flow rate

K_1 = ratio of specific heat to heat of vaporization of the external reflux

K_2 = ratio of heat of vaporization of external reflux to heat of vaporization of internal reflux

T_o = overhead vapor temperature

T_r = external reflux temperature

Dynamics Equations Equations 8.21(68) and 8.21(69) are steady-state equations. Applied without alteration, undesirable column response will result, especially for sudden feed flow rate changes. Feed composition changes are less severe than are feed flow rate changes and seldom require dynamic compensation.

Dynamic elements should compensate for feed flow rate changes in such a way that when the feed flow rate changes, the column's terminal stream flows should respond at the proper time, in the correct direction, and without overshoot. The simplest form of dynamics to meet these criteria involves dead time plus a second-order exponential lag response. The feed flow rate signal is passed through this dynamic element before being used in the operating equations to obtain the bottom product flow rate (B) and the reflux flow rate (L) set points. The transfer function for the dynamic element is

$$\frac{(F_L)}{(F)} = \frac{Ke^{-ts}}{(T_1 s + 1)(T_2 s + 1)} \qquad \mathbf{8.21(75)}$$

where

(F_L) = feed flow rate lagged

(F) = feed flow rate measured

t = dead time

T_1, T_2 = time constant

Using Equation 8.21(75), F is eliminated from the left side of Equations 8.21(68) and 8.21(69) to obtain the complete set of operating equations as used in Figure 8.21v. These equations become

$$B = F\left[\frac{Ke^{-ts}}{(T_1 s + 1)\,(T_2 s + 1)}\right]$$

$$\left[\frac{100 - (\%\overline{HK}_D) - (\%LLK_F) - (\%LK_F)}{100 - (\%\overline{HK}_D) - (\%\overline{LK}_B)}\right] \qquad \mathbf{8.21(76)}$$

$$L = F\left[\frac{Ke^{-ts}}{(T_1 s + 1)\,(T_2 s + 1)}\right]\left[\frac{(L_i/F)_t + (L_i/F)e}{K_2[1 + K_1(T_o - T_t)]}\right] \qquad \mathbf{8.21(77)}$$

where in functional form:

$$(L_i/F)_t + (L_i/F)_e = f_1[(LLK_F),(LK_F),(HK_F),(\bar{E}),(\overline{FT}),$$

$$(\overline{\Delta H}_F),\ (\bar{P}),(\%\overline{LK}_B),(\%\overline{HK}_D)] + f_2(L_i) \qquad \mathbf{8.21(78)}$$

Note that the bottoms and overhead dynamic constants, t, T_1, and T_2 in Equations 8.21(76) and 8.21(77), are not necessarily the same values.

Application of Equations 8.21(76), 8.21(77), and 8.21(78) will result in a suboptimized operation. This is an operation producing a performance close to the specified one. The control block diagram of Figure 8.21w illustrates the application of these equations.

Inspection of Equations 8.21(76) and 8.21(78) shows that the system still has five degrees of freedom. Therefore, feed tray location ($\overline{FT}$), feed enthalpy ($\overline{\Delta H}_F$), column pressure ($\bar{P}$), concentration of heavy key component in the top product ($\%\overline{HK}_D$), and concentration of light key component in the bottom product ($\%\overline{LK}_B$), must all be specified. Although tray efficiency is also included, it remains a fixed value, as explained earlier.

Local Optimum Variables

Feed enthalpy and column pressure are local optimization variables that can be manipulated to achieve two different objectives to increase profitability: minimizing utilities costs or maximizing throughput. Often, this type of optimization is implemented via valve position controllers. The purpose of valve position controllers is to drive the column to a constraint condition on either reboiler heat, condenser duty, or column loading.

Strategies to reduce utility costs include a valve position controller cascaded to pressure control. This is commonly referred to as "floating-pressure control." Strategies to increase throughput include valve position controller cascaded to feed flow rate. In both cases, the valve position controller will drive the manipulated variable to an equipment constraint.

Minimum Pressure Control Floating-pressure operation can often reduce energy consumption by providing minimum pressure operation within the constraints of the system. It is possible to operate a total condensing distillation column with no pressure control.[8] Although this provides for optimum operation at steady state, major problems could occur (e.g., flooding the column) during transient upsets. Flooding is caused by high vapor rates and can result in entrainment or foaming, or in preventing the liquid from flowing down the column. To prevent this, pressure control should be provided.

Most distillation columns are operated with constant pressure control. However, several advantages can be achieved

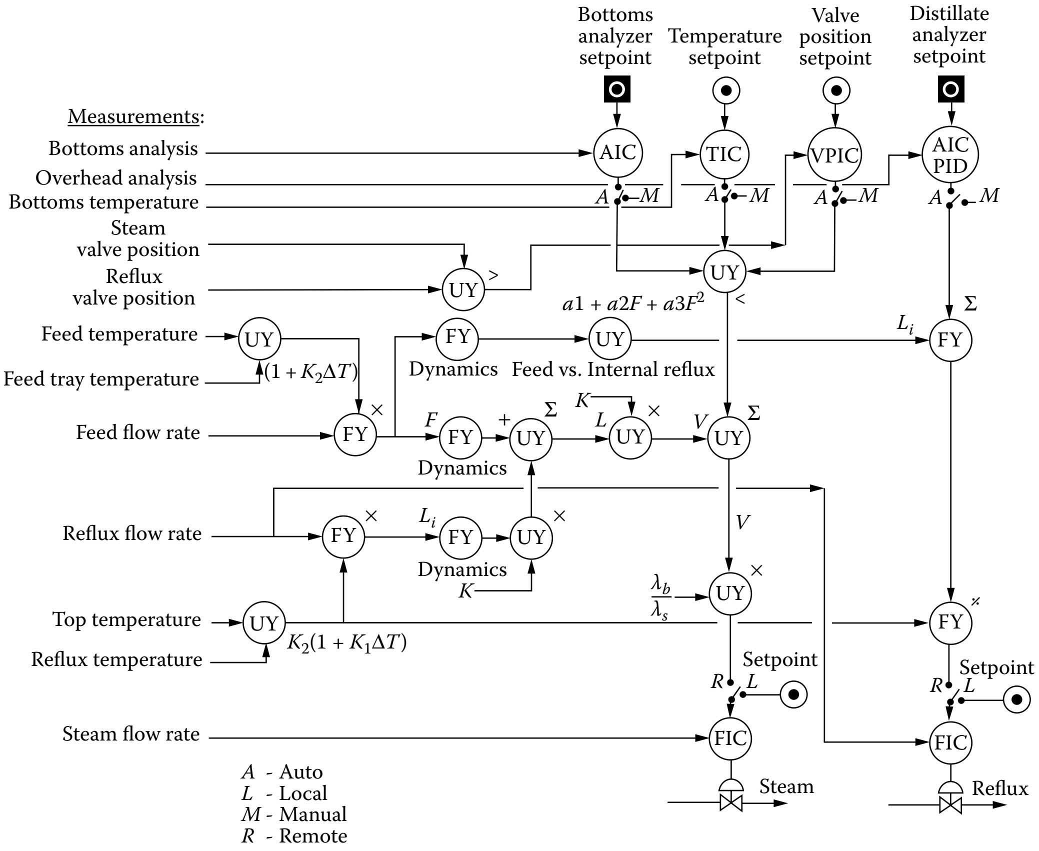


FIG. 8.21w

Fractionator control block diagram.

through minimizing operating pressure in most columns. Resistance to floating the pressure is largely due to the desire to use temperature measurements to indicate and control compositions. As analyzers are replacing temperature measurements, the traditional desire for constant pressure is also being removed. However, even when analyzers are not used, temperature measurements can be compensated for pressure variations.

Floating-pressure control allows the tower to operate with minimum pressure within the constraints of the system. Lower pressure increases the relative volatility of distillation components, thereby reducing the heat input required to effect a given separation. Other advantages include increased reboiler capacity and reduced reboiler fouling due to lower temperatures. Reducing pressure also affects other parameters, such as tray efficiencies and latent heats of vaporization. Some of these effects occur in opposite directions; therefore, floating-pressure control is desirable for some columns and not for others. The components being separated are the major determining factor.

Minimum pressure operation can be achieved by manual or automatic adjustment of the set point of the pressure controller to keep the condenser fully loaded at all times. However, to prevent upsets caused by rapid set point changes, the valve position control (VPC) scheme shown in Figure 8.21x is used.[8] The VPC adjusts the set point of the pressure controller. This maximizes cooling by holding the condenser control valve in the fully open or fully closed position, depending on whether the valve bypasses, throttles, or floods the condenser.

The pressure controller should incorporate proportional plus integral action to provide rapid response to upsets. The VPC should be an integral-only controller, so a rapid change in valve position will not produce a proportional change in the pressure set point. The integral time setting of the VPC should be approximately 10 times that of the overhead composition controller. In addition, it is common practice to limit the range within which the VPC can adjust the set point and to provide the external feedback shown in Figure 8.21x to eliminate reset windup when the VPC output reaches one of these limits. The setpoint for the VPC is given a value, such as 90%, so that the controller has a range in which to operate.

Partial condensers require a different approach to floating-pressure control because they are sensitive to cooling and accumulation of noncondensables. Typically, pressure in

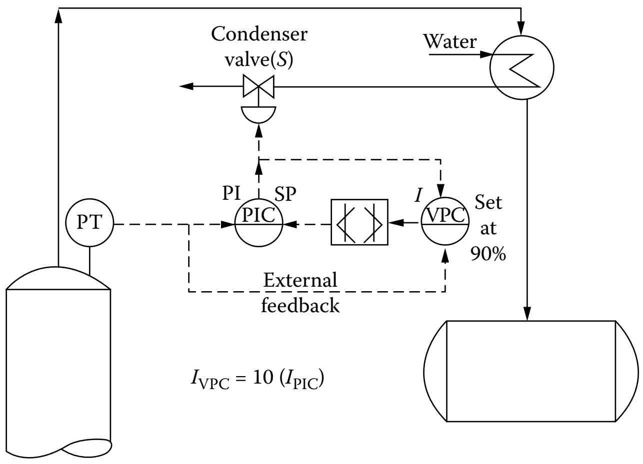


FIG. 8.21x
VPC provides floating-pressure control of total condenser.

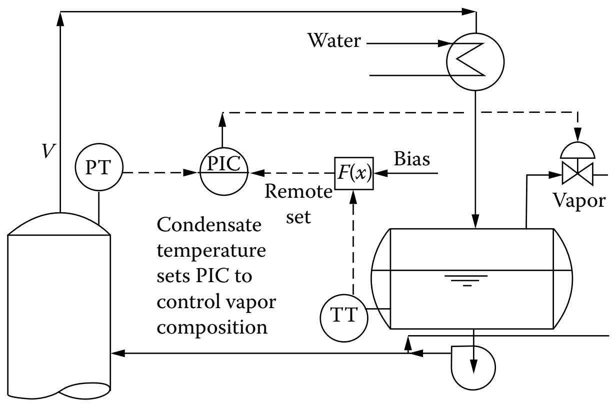


FIG. 8.21z
Floating-pressure control for a partial condenser with both liquid and vapor products.

these systems is controlled by addition or venting of inert gases. Though simple, this practice requires a source of inert gases, does not allow for steady-state optimum pressure operation, wastes overhead products that are vented with the noncondensables, and can create problems in downstream units through the addition of noncondensables.

Figure 8.21y shows a partial condensing system with no liquid product. Here, both level and pressure controllers are used to provide floating-pressure operation.^{8} The level controller acts as the VPC in the total condensing system to provide complete flooding (on the refrigerant side) in the long term, and the pressure controller handles short-term upsets.

When both liquid and vapor products are withdrawn, an additional control loop is required to control the composition of the vapor, as shown in Figure 8.21z. Here column pressure is controlled by vapor flow, but the set point must be adjusted for changes in accumulator (condenser) temperature.^{8} The temperature measurement is characterized to an equivalent vapor pressure representing the desired composition. A bias adjustment should be incorporated to readjust the relationship between pressure and temperature for desired changes in composition. This method completely eliminates the need for throttling the cooling water to the condenser.

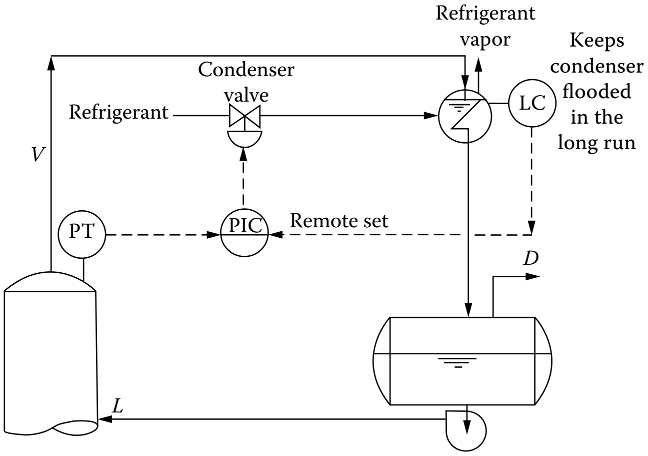


FIG. 8.21y
Floating-pressure control for a partial condenser with no liquid product.

It is important to realize that composition control must take precedence over pressure optimization. The pressure-valve position control loop response must be much slower than that of the composition control loops. Also, potential effects of pressure changes on upstream and downstream units must be considered. For example, if the pressure of an upstream tower provides the driving force to move product to a downstream tower, pressure minimization may not be practical. Fractionators using vapor recompression, such as a propylene splitter (Figure 8.19mm in Section 8.19) with a heat pump, may actually benefit from increasing pressure rather than reducing it.

Feed Maximization Control Where product demand and the availability of feedback is unlimited, increasing throughput maximizes profitability. In such installations, a valve position controller can be cascaded to the feed flow controller in order to increase the feed rate until an equipment constraint is reached. Figure 8.21aa illustrates a cascade configuration to

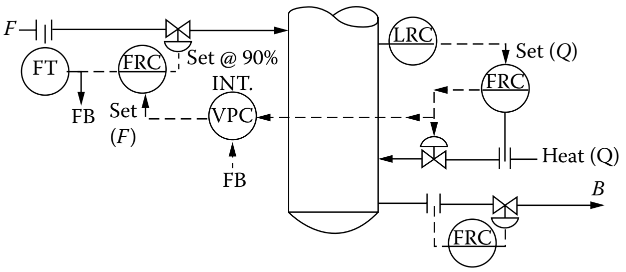


FIG. 8.21aa
Valve position control maximizes feed until a reboiler constraint is reached.

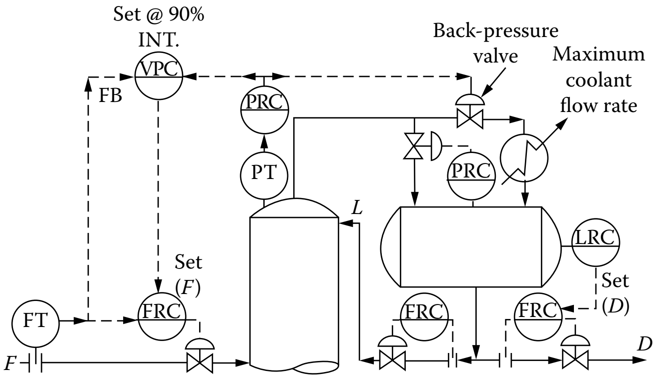


FIG. 8.21bb
Valve position control maximizing feed until a condenser constraint is reached.

keep the reboiler fully loaded. This strategy is particularly effective when the cost of reboiler heat is negligible, such as when waste steam is used that would be vented otherwise.

Likewise, if a condenser limit is involved, the opening of the column pressure control valve can be measured and the feed rate manipulated by a controller to maintain the back-pressure valve near open. Also, if a hot vapor bypass exists around the condenser, the opening of the bypass valve will also indicate the condenser load. These procedures are illustrated in Figure 8.21bb. Instead of minimizing pressure in order to reduce utility usage, here the objective is to maximize throughput.

The VPCs in Figures 8.21aa and 8.21bb are usually selected as integral-only controllers and are set at around 90% of valve stem lift, which on an equal-percentage valve corresponds to 70% of maximum flow. The integral setting of the VPC is slow, about 10 times the integral setting of the FRC, the set point of which the VPC adjusts in a cascade arrangement. In order to eliminate reset windup, the VPC is provided with external feedback from the slave transmitter.

In many cases, it is not known which single constraint will be encountered as feed is maximized. The critical constraint may vary over time. In such cases, a multiple constraint network is implemented.

Feed Enthalpy Control When an economizer is used to preheat the feed with the bottoms product (Figure 8.19rr), it is advantageous to maximize the amount of heat that is recovered from the bottoms product. In this case, the valve position controller (VPC) resetting the temperature controller as shown in Figure 8.21cc ensures that the flow in the bypass is kept at a minimum value. The tuning and feedback requirements of this VPC are similar to those of Figures 8.21x, 8.21aa, and 8.21bb.

Constant temperature feed does not necessarily mean constant feed quality. If feed composition varies, its bubble point also varies. It is common practice to set the temperature control at a point that is equivalent to the bubble point of the heaviest feed. As the feed becomes lighter, some of it will evaporate, but this variation can be handled by subsequent controls.

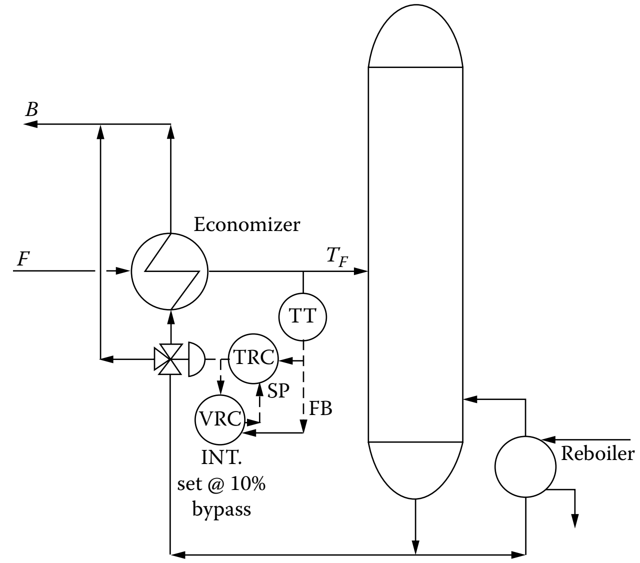


FIG. 8.21cc
Maximizing feed preheat through valve position control.

Control of feed enthalpy instead of temperature can be achieved provided the proper measurements are in place. Consider Figure 8.21dd, in which the cold feed at temperature T_f first passes through the bottoms-to-feed economizer, then through the steam preheater.

The feed may be partially vaporized as it leaves the economizer. In order to calculate the feed enthalpy as it enters

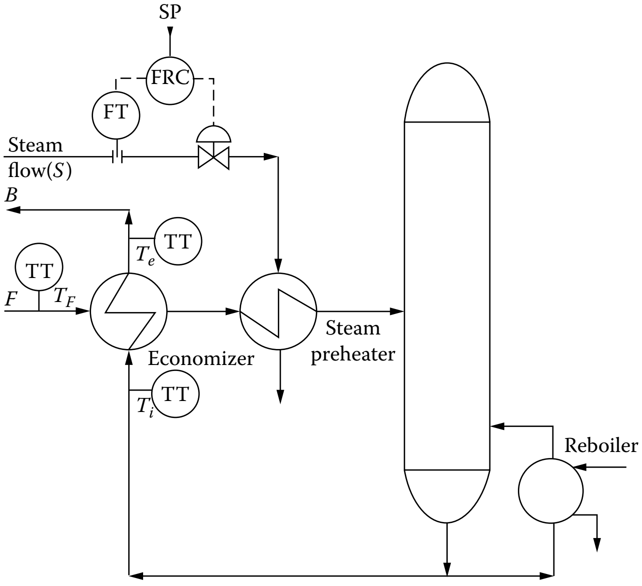


FIG. 8.21dd
Feed enthalpy control.

the column, the following heat balance can be written:

$$(F)(\Delta H_F) = (F)(C_{p_F})(T_f) + (B)(C_{p_B})(T_i - T_e) + (S)(\Delta H_{\text{stm}}) \qquad \textbf{8.21(79)}$$

where

ΔH_F = feed enthalpy as it enters the tower, BTU/lb (kcal/kg)
F = feed flow to preheater, lb/h (kg/h)
S = steam flow, lb/h (kg/h)
C_{p_F} = feed heat capacity, BTU/lb°F (kcal/kg°C)
T_f = feed temperature before preheaters,°F (°C)
B = bottoms flow to economizer, lb/h (kg/h)
C_{p_B} = bottoms heat capacity, BTU/lb (kcal/kg)
T_i = bottoms temperature to economizer,°F (°C)
T_e = bottoms temperature after economizer,°F (°C)
ΔH_{stm}= steam heat of vaporization minus condensate heat, BTU/lb°F (kcal/kg°C)

Rearranging Equation 8.21(79) gives

$$\Delta H_F = (C_{p_F})(T_f) + (B/F)(C_{p_B})(T_i - T_e) + (S)(\Delta H_{\text{stm}})/F \qquad \textbf{8.21(80)}$$

or, solving for the manipulated variable, the steam flow that will provide constant feed enthalpy is

$$S = [F(\Delta H_F - C_{p_F} T_f) - B(C_{p_B} \Delta T)]/\Delta H_{\text{stm}} \qquad \textbf{8.21(81)}$$

The primary effect of increasing feed enthalpy is to decrease vapor-liquid circulation below the feed tray relative to that above the feed tray. When feed preheat is less expensive than reboiler heat, or when the reboiler is the limiting constraint, maximum feed preheat is often optimal. When condenser capacity is limiting, or flooding is encountered above the feed tray, preheat is not desired.

OPTIMIZATION

Optimization implies maximum profit rate. An objective function is selected, and manipulated variables are chosen that will maximize or minimize that function. This is similar to the PID equation that is designed to minimize the error between the set point and measurement, but at a higher level. Optimization can be applied in several layers. Local optimization is the optimization of a single column. Normally, the goal of optimization of a single tower is to obtain minimum energy consumption or maximum throughput.

Unit optimization addresses several columns in series or parallel. It is concerned with the effective allocation of feedstocks and energy among the members of that system. Plantwide optimization involves coordinating the control of distillation units, furnaces, compressors, and so on to maximize profit from the entire operation. All lower-level control functions respond to set points received from higher-level optimizers.

If Product Prices are Unknown

Unless the prices of terminal products of a distillation column are known, it is impossible to maximize the profit rate for that column without taking into account all other aspects of the overall plant of which the column is a part. Thus, optimization for a single distillation column whose terminal product prices are unknown is a matter of producing specification products for minimum operating cost. Determining the optimum separation for the column requires optimizing the overall system of which the column is a part.

Optimization of a single column whose product prices are unknown involves determining the values for the location of the feed tray ($\overline{FT}$), for the feed enthal ($\overline{\Delta H}_F$), and for column pressure ($\bar{P}$) that result in minimum operating costs for whatever separation is specified. Any applicable mathematical approach can be used to establish values for ($\overline{FT}$), ($\overline{\Delta H}_F$), and ($\bar{P}$) that will result in minimum operating costs.

Assuming that Equations 8.21(76), 8.21(77), and 8.21(78) are available, it is a relatively simple matter to establish optimum values for these three variables that result in minimum operating costs. Because these variables have specific constraint values, one method involves a search technique. It is usually difficult to justify the search technique for on-line computer control. Therefore, a statistical design study can be made off-line on another computer that allows correlation of the variables with each of the three optimizing variables.

Theoretically, three equations in functional form describe the optimum for ($\overline{FT}$), ($\overline{\Delta H}_F$), and ($\bar{P}$)

$$(\overline{FT})_o = f_3[(\%LLK_F), (\%LK_F), (\%HK_F), (F), (\%\overline{HK}_D), (\%\overline{LK}_B)] \qquad \textbf{8.21(82)}$$

$$(\overline{\Delta H}_F)_o = f_4[(\%LLK_F), (\%LK_F), (\%HK_F), (F), (\%\overline{HK}_D), (\%\overline{LK}_B)] \qquad \textbf{8.21(83)}$$

$$(\bar{P})_o = f_5[(\%LLK_F), (\%LK_F), (\%HK_F), (F), (\%\overline{HK}_D), (\%\overline{LK}_B)] \qquad \textbf{8.21(84)}$$

In a majority of cases, the optimum for column pressure, ($\bar{P}$), will be the minimum value with constraints of the system. Normally, ($\bar{P}$) can be lowered until the condenser capacity is reached, or until liquid entrainment in the vapor on the trays is initiated.

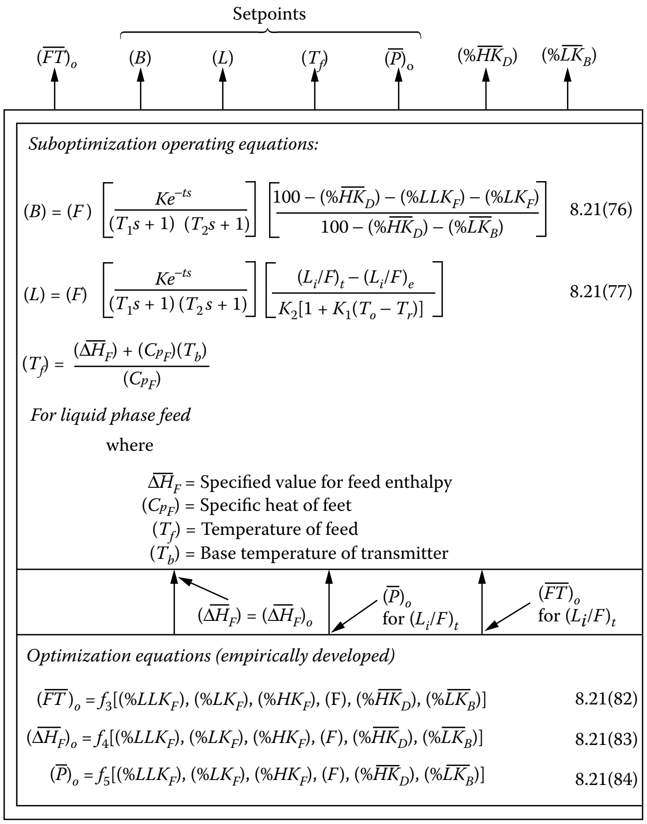


FIG. 8.21ee

Optimizing control for single column when product prices are unknown (criterion is to produce specification products for the least operating cost).

Figure 8.21ee illustrates all the control equations involved to optimize the operation of a column when product prices are unknown. Determination of column pressure is handled by the predictive control equation, Equation 8.21(84).

If Products Prices are Known

When terminal product prices for a single column are known, the column is optimized to obtain the specified separation for the least operating cost. The problem now is to determine values for the separation that will maximize profit rate. There are, however, a number of different situations that may exist, which will be covered below. Several general optimizing policies can be stated first. The purpose of these general policies is to reduce the number of variables involved in design of the optimizing control system.

Optimizing Policies

1. The optimum separation for a single distillation column can be determined independently of feed cost.
2. One condition resulting in optimized operation for a single distillation column is production of that product with the highest unit price at minimum specified purity.
3. The optimum separation to specify for a single distillation column is not a function of each terminal product price, but a function of the price difference between products.
4. When the individual components in the products of a single column have separate assigned prices, the optimum separation is not a function of all the component prices but it is a function of the price difference between the heavy key component in the top and bottom ($PHK_D - PHK_B$) and of the price difference between the light key component in the top and in the bottom products ($PLK_D - PLK_B$).

These policies are derived by the evaluation of the partial differential equations that describe the profit rate for a single column with respect to the specified separation, noted as $(\overline{LK}_D), (\overline{HK}_B)$.

Operating Constraints

When product prices for a distillation column are known, complete economic optimization almost always requires the operation to be against a constraint. If not against an operating constraint, the optimum will occur when the specified separation $(\overline{LK}_B), (\overline{HK}_D)$ is of such a value that incremental gain in product worth is equal to incremental gain in operating cost. Because the majority of cases will involve operating constraints, it is important to understand the principles involved.

Loading a distillation column is affected by the specified separation and by the existing feed rate. Loading is increased by specification of a better separation or by an increase in the feed rate at a constant separation. In general, both the feed rate and separation are involved as optimizing variables when an unlimited market exists for the products and when sufficient feedstock is available.

The operating constraints normally involve the capacities of (1) the condenser, (2) the reboiler, and (3) the column. As feed rate is increased, or an increased separation specified, one of these three constraints will be approached. The constraint encountered may also depend on prevailing external conditions, e.g., ambient conditions, steam pressure, and feed composition. The optimum may then be different depending upon these other conditions and the constraint first reached.

Condenser Constraint Capacity of a given condenser at maximum coolant flow rate is a function of the differential temperature between the overhead vapor and the coolant media. One useful approach to operating a column against the condenser constraint requires correlation of maximum vapor flow rate with this temperature difference. Such a correlation can be obtained by column testing. The information obtained by on-line tests is curve-fitted to some general form, such as in Equation 8.21(85):

$$(V_o)_{\max} = a_1 + a_2(\Delta T) + a_3(\Delta T)^2 \qquad \mathbf{8.21(85)}$$

where

$(V_o)_{\max}$ = maximum overhead vapor flow rate that will load the condenser maximally
ΔT = temperature difference between overhead vapor and coolant to the condenser
a_1, a_2, a_3 = coefficients

Values for feed rate and for the separation can be determined that will result in $(V_o)_{\max}$ to load the condenser. Column pressure and condenser fouling are also major variables that will affect overhead vapor temperature. For improved accuracy, the temperature of the overhead vapor can be modeled as a function of all independent variables in the system for use in the ΔT determination. This would result in a completely predictive system for loading the condenser. As the condenser becomes fouled, new coefficients must be established for Equation 8.21(85).

Reboiler Constraint Capacity of the reboiler at maximum flow rate of the heating media is a function of the temperature difference between the heating media in the reboiler tubes and the liquid being reboiled. Just as with the condenser, tests can be conducted to correlate maximum vapor flow rate out of the reboiler with temperature difference across the reboiler tubes.

Also, temperature of the reboiler can be expressed as a function of the column's independent inputs. Column pressure and reboiler fouling are major variables that will affect the temperature of the reboiler liquid. Generally, when using waste streams such as low pressure steam that would alternatively be vented, it is optimal to operate against a reboiler constraint.

Column Constraint The capacity of a given column is a function of liquid and vapor flow rates within the column as well as of the column pressure. Often, capacity is limited by entrainment of liquid by the vapor. At low internal liquid flow rates, a higher vapor flow rate can be used. Also, column capacity will be greater at higher pressures. If capacity is limited by entrainment, then loading can be increased at higher pressures. However, if column capacity is limited by the tray downcomers, internal liquid flow rate can be increased by lowering pressures. Therefore, the capacity-limiting parameter must be known.

Over a limited range, a linear relationship can be assumed between column pressure (P), liquid flow rate (L), and the maximum vapor rate ($V_{\max}$) that will initiate entrainment. Figure 8.21ff shows data that were obtained by column testing, because it is difficult to predict these effects.

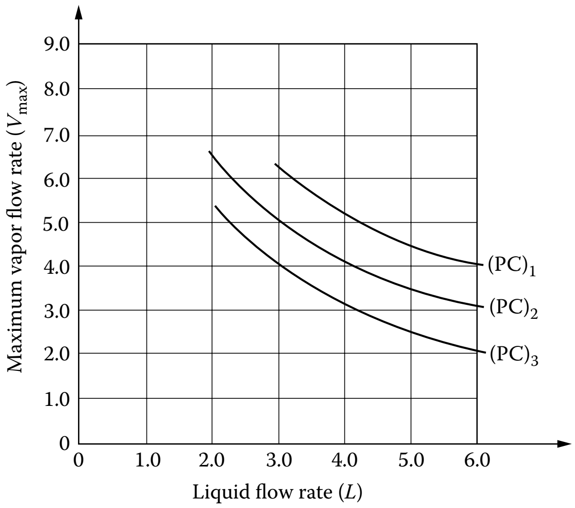


FIG. 8.21ff
Typical relationship between liquid flow rate (L) and the vapor flow rate ($V_{\max}$) that will initiate entrainment.

An equation can be developed from the test data to cover a limited range of liquid and vapor flow rates and pressure: Equation 8.21(86). This relationship can usually be considered linear.

$$V_{\max} = a_1 + a_2(L) + a_3(P) \qquad \mathbf{8.21(86)}$$

This equation is useful in predicting the values of feed flow rate and of the separation that will cause maximum vapor flow rate to exist. The use of these loading functions to optimize the operation of a column will be covered later in this section.

Limited Market and Feedstock

Assuming that the column is already equipped with the operating control functions given in Figure 8.21ee, the overall control system is illustrated in Figure 8.21gg. In order to keep the drawing clear, the basic controls are not shown here, but can be seen in Figure 8.21t.

This column is assumed to have four feed components having concentrations of LLK_F, LK_F, HK_F, and HHK_F. Three of these components appear in each product. The separation for the column is fixed by specification of the concentration of heavy key component in the top product ($\overline{HK}_D$) and the concentration of light key component in the bottom product ($\overline{LK}_B$).

The top and bottom products have unit prices of PD and PB, respectively. All the components in the feed that are

Inputs and outputs for computer control

Measured inputs	*Specified inputs*	*Constants*	*Outputs*
(LLK_F)	$(\overline{HK_D})$	$(LK_D)_{ss}$	$(\overline{FT})_o$
(LK_F)	$(\overline{LK_B})$	$(HK_B)_{ss}$	$(\overline{P})_o$
(HK_F)		K	(T_f) from $(\mathrm{FT})_o$
(F)		Cp_F	(L)
$(T_o{-}T_r)$		T_b	(B)
		T_1	
		PF	
		PD	
		PB	
		⊝	

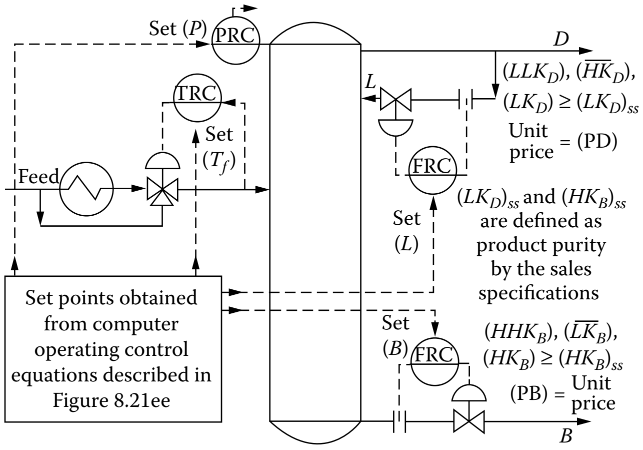


FIG. 8.21gg
With prices of the products (PD, PB) known, the problem is to find the separation (HK_D, LK_B) that will maximize profit rate.

lighter than the light key component, i.e., LLK_F components, will go to the top product, whereas all the components that are heavier than the heavy key component, i.e., HHK_F components, will go to the bottom product. Therefore, the operation of the column can do nothing about the distribution of these two components between the top and bottom products. $\overline{HK_D}$ will be the impurity component in the top, and $\overline{LK_B}$ the impurity component in the bottom.

The concentrations of these two components in the products can be fixed by specification of their values as inputs to the predictive controller. LK_D is the purity component in the top product, and HK_B is the purity component in the bottom. Sales specification for the two products usually state the minimum acceptable purities.

$$(HK_B) \geq (HK_B)_{ss} \qquad \textbf{8.21(87)}$$

$$(LK_D) \geq (LK_D)_{ss} \qquad \textbf{8.21(88)}$$

where $(HK_B)_{ss}$ and $(LK_D)_{ss}$ are minimum purity sales specifications.

Control Equations

Control equations for the heavy key in the top product and for the light key in the bottoms product will be developed first for the case in which the top product is considered to be more valuable and second for the case in which the bottom product is considered to be more valuable for a two-product fractionator.

Top Product Having Higher Value

As stated earlier, the general optimizing policy is that the product with the highest unit price must be produced at minimum specified purity. The heavy key component in the top product $(\overline{HK_D})$ is the control component. Its set point can be calculated as a function of feed component concentrations, minimum sales purity specification, $(LK_D)_{ss}$, and concentration of light key component in the bottom product $(\overline{LK_B})$.

Heavy Key in Top Product The following material balance equations are involved in determining the optimum value for the concentration of the heavy key in the top product, $(\overline{HK_D})_o$.

$$(\overline{HK_D})_o = 1.00 - (LK_D)_{ss} - LLK_D \qquad \textbf{8.21(89)}$$

$$(LLK_F)(F) = (LLK_D)(D) \qquad \textbf{8.21(90)}$$

$$D = (F - B) \qquad \textbf{8.21(91)}$$

$$B = F\left[\frac{(1-(\overline{HK_D})_o = [(LLK_F)+(LK_F)]}{1-(\overline{HK_D})_o-(\overline{LK_B})}\right] \qquad \textbf{8.21(92)}$$

In Equation 8.21(89), LLK_D is eliminated by use of Equation 8.21(90). D is eliminated by Equation 8.21(91), and B is eliminated by Equation 8.21(92). The following is obtained:

$$(\overline{HD_D})_o = \frac{\begin{array}{c}(LK_F)+(\overline{LK_B})[(LLK_F)-1]\\ -(LK_D)_{ss}[(LLK_F)+(LK_F)-(\overline{LK_B})]\end{array}}{(LK_F)-(\overline{LK_B})} \qquad \textbf{8.12(93)}$$

Light Key in Bottom Product Having obtained the value for optimum concentration of heavy key in the top product, the problem reduces to finding the value for optimum concentration of light key component in the bottom product, $(\overline{LK_B})_o$.

The maximum value that can be specified for LK_B to meet the minimum sales specification $(HK_B)_{ss}$ will be as follows:

$$(\overline{LK_B})_{\max} = \frac{\begin{array}{c}[1-(\overline{HK_D})_o][(LLK_F)+(LK_F)+(HK_F)+(HK_B)_{ss}]\\ -[(LLK_F)+(LK_F)][1-(\overline{HK_B})_{ss}]\end{array}}{(LK_F)-(\overline{HK_D})_o}$$

8.21(94)

This equation is derived by the same procedure that was used for Equation 8.21(94). As $\overline{LK_B}$ is lowered from its maximum value, the flow rate of the bottom product will decrease and the top product flow rate will increase. Because the top product is the highest unit price, profit will increase as is $\overline{LK_B}$ lowered. Also, operating costs will increase, because a better separation is being specified.

As $\overline{LK_B}$ is lowered further, one of two things will occur to establish $(\overline{LK_B})_o$ (the lowest value for LK_B). Either the change in operating cost will approach the change in profit rate, or the column will be loaded against an operating constraint. This constraint will be a condenser limit, a reboiler limit, or a column limit. Whichever occurs first will establish $(\overline{LK_B})_o$.

Assume, for example, that no constraints are involved. In that case, the value of the products will be given by

$$PW = (PD)(D) + (PB)(B) \qquad \mathbf{8.21(95)}$$

where PW is the total product worth rate, dollars/unit time. Eliminate B by

$$B = F - D \qquad \mathbf{8.21(96)}$$

to obtain

$$PW + (PD - PB)(D) + (PB)(F) \qquad \mathbf{8.21(97)}$$

Taking the partial derivative of this equation with respect to $\overline{LK_B}$, the following is obtained:

$$\frac{\partial(PW)}{\partial(\overline{LK_B})} = (PD - PB)\frac{\partial(D)}{\partial(\overline{LK_B})} \qquad \mathbf{8.21(98)}$$

Operating costs can be approximated closely by

$$(OC) = (CL_i)(L_i) \qquad \mathbf{8.21(99)}$$

where CL_i is the unit cost of operation per unit of internal reflux. Taking the partial derivative of this equation with respect to $\overline{LK_B}$, the following is obtained:

$$\frac{\partial(OC)}{\partial(\overline{LK_B})} = CL_i\frac{\partial(Li)}{\partial(\overline{LK_B})} \qquad \mathbf{8.21(100)}$$

When the change in operating cost is equal to the change in product worth, the value for $(\overline{LK_B})_o$ is determined.

$$(PD - PB)\left[\frac{\partial(D)}{\partial(\overline{LK_B})}\right] = CL_i\left[\frac{\partial(L_i)}{\partial(\overline{LK_B})}\right] \qquad \mathbf{8.21(101)}$$

Evaluation of this equation will yield a value for $(\overline{LK_B})_o$. D must be expressed in terms of the specified separation, the feed composition, and flow rate. L_i must be expressed in terms of the independent inputs to the system as used in the internal reflux control equation. The simultaneous solution of Equations 8.21(93) and 8.21(101) will yield optimum values for LK_B and HK_D: $(\overline{LK_B})_o, (\overline{HK_D})_o$. The total solution is indicated in Figure 8.21hh.

Bottom Product More Valuable

The general optimizing policy requires that the product with the highest unit price be produced at minimum specified purity.

Light Key in Bottom Product Because the bottom product must be produced at minimum purity, Equation 8.21(94) gives the optimum value for $\overline{LK_B}$. The optimum $\overline{LK_B}$ equals $(\overline{LK_B})_{\max}$. This, in turn, gives the minimum specified sales purity $(HK_B)_{ss}$.

Heavy Key in Top Product The maximum value for heavy key in the top product is given by Equation 8.21(93). As $\overline{HK_D}$ is reduced from its maximum allowable value, flow rate of the top product will decrease and flow rate of the bottom product will increase. Therefore, total worth of the products will increase together but with increased operating costs due to increased reflux required. The optimum value for $\overline{HK_D}$ will occur when an operating constraint is encountered or when the incremental gain in product worth is equal to the incremental gain in operating cost.

For this case, assume that the column approaches an operating constraint as $\overline{HK_D}$ is lowered. Let this constraint be flooding above the feed tray because of excessive entrainment of liquid in the vapor.

First, an equation needs to be developed for vapor flow rate above the feed tray in terms of variables that are contained in the bottom product and reflux operating equations. The loading equation is developed as follows:

$$L_i + D = V_t \qquad \mathbf{8.21(102)}$$

where

L_i = internal reflux flow rate
D = top product flow rate (distillate)
V_t = vapor flow rate above feed tray

V_t can have a maximum value as given by Equation 8.21(86), which is developed specifically for the column involved. Therefore, equate the right side of Equation 8.21(86) with the left side of Equation 8.21(102) to obtain

$$L_i + D = a_1 + a_2(L) + a_3(\bar{P}) \qquad \mathbf{8.21(103)}$$

where

a_1, a_2, a_3 = coefficients of experimental loading equation
L = liquid flow rate in column at the point at which flooding occurs
P = column pressure

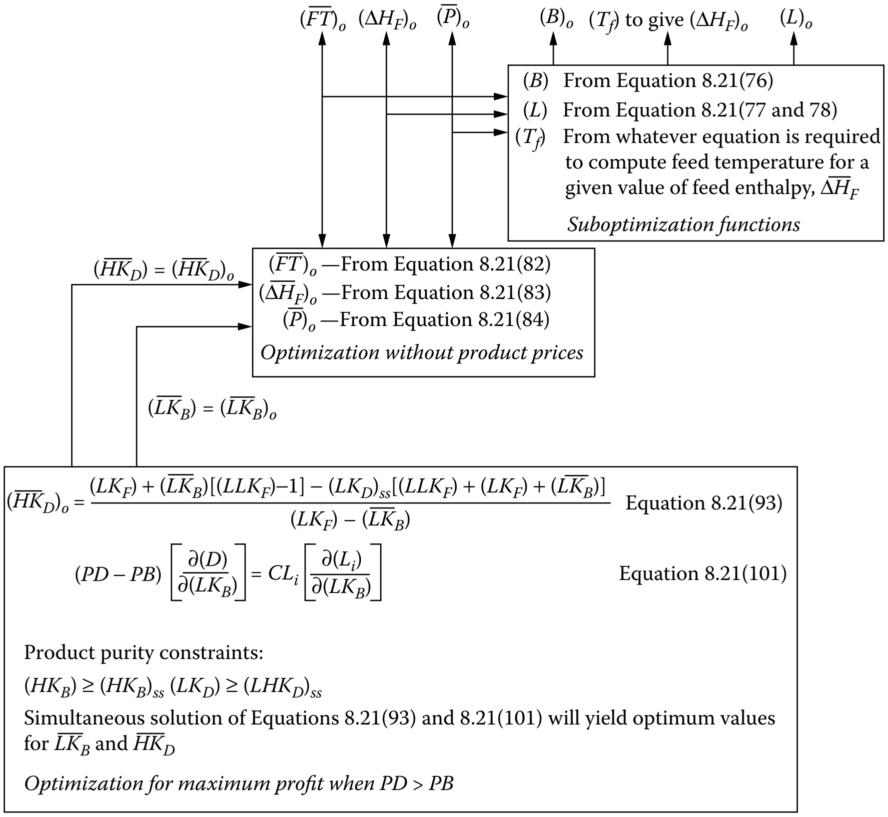


FIG. 8.21hh
Control hierarchy to maximize profit rate when operating constraints are involved.

For this case, L will equal L_i. Also, eliminate D by substituting $(F - B)$. Therefore, by substituting

$$L = L_i = \left[\frac{L_i}{F}\right]\mathrm{F} \qquad \textbf{8.21(104)}$$

and

$$B = \left[\frac{B}{F}\right]F \qquad \textbf{8.21(105)}$$

the following is obtained:

$$a_1 + a_3(\bar{P}) + \left[(a_2 - 1)\frac{L_i}{F} + \frac{B}{F} - 1\right]F = 0 \qquad \textbf{8.21(106)}$$

In this equation a_1, a_2, and a_3 are known from the experimental loading equation. $\bar{P}$ is specified, F is measured, and L_i/F and B/F are obtained from the operating control equations for reflux and bottom product flow rate.

It is now possible to find the optimum value for $\overline{HK}_D$. As $\overline{HK}_D$ is lowered from its maximum allowable value as given by Equation 8.21(93), the values for L_i/F and B/F will change. $\overline{HK}_D$ can be lowered until loading Equation 8.21(102) is equal to zero. Also, column pressure $\bar{P}$ can be raised to allow a greater loading that will result from a lower $\overline{HK}_D$. A point will be found at which maximum profit rate will exist.

The maximum limit for $\bar{P}$ will be determined by several factors. As $\bar{P}$ is increased, operating costs will also increase, because of the resulting smaller differential temperature at the reboiler. Therefore, one possible limit would be a reboiler (or operating cost) limit.

Another pressure limit will be set by the column pressure rating. Another limit for maximum $\bar{P}$ may be determined by requirements of upstream processing equipment. Yet another limit for maximum $\bar{P}$ could be loading of the downcomers between each tray. This comes about because at higher vapor densities, disengagements of vapor from the liquid becomes more difficult. Therefore, at some maximum pressure, density of the vapor can approach the point at which vapor will not have sufficient time to disengage from the liquid in the downcomers, and a condition known as downcomer flooding will occur.

Assume for the purpose of illustration that the maximum for $\bar{P}$ is set at the pressure rating of the column. Therefore, $\bar{P}$ will be set and left at this value.

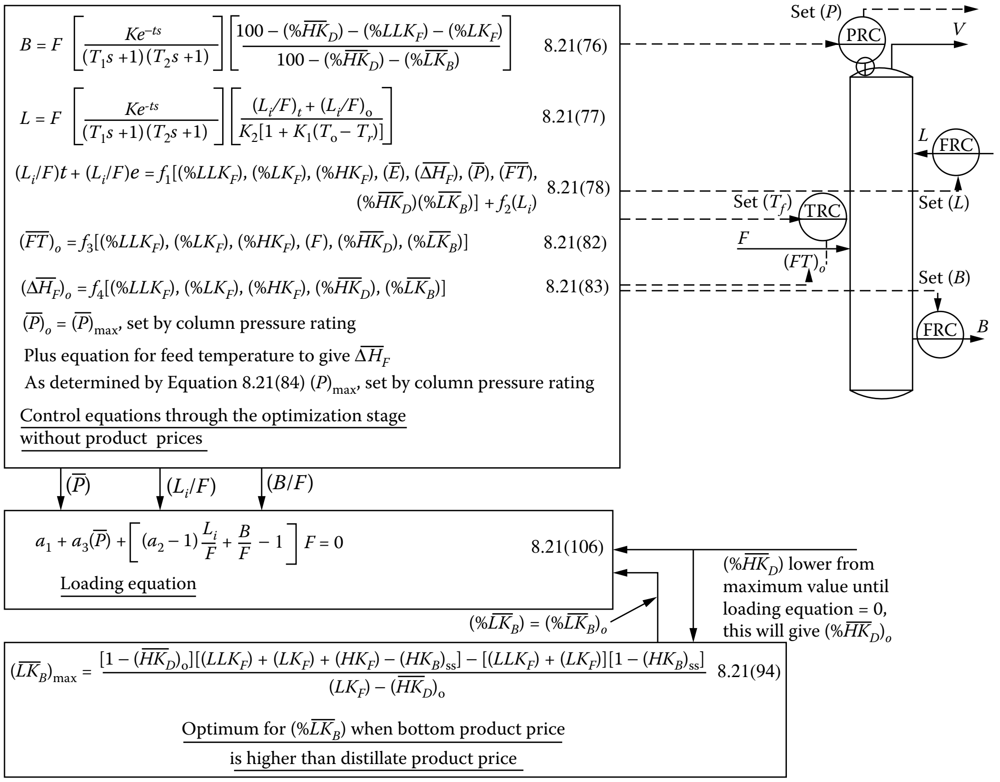


FIG. 8.21ii

System for maximizing profit rate when bottom product price is higher than top product price and when entrainment above the feed tray limits loading.

Figure 8.21ii shows the optimizing control system connected to a typical distillation column. Only part of the basic controls are shown for purpose of clarity. The basic controls are not show in this figure and can be assumed to be the same as shown in Figure 8.21t.

Unlimited Market and Feedstock

When an unlimited market exists for the products, the feed flow rate and separation resulting in maximum profit rate must be determined. Operation for a column under this condition will always be against an operating constraint. The following example assumes that the operating constraint is the overhead vapor condenser capacity.

In general, the overall optimization problem for this case is illustrated in Figure 8.21jj. Values must be determined for the optimum separation, $(\overline{HK}_D)_o, (\overline{LK}_B)_o$; feed flow rate, $(F)_o$; and column pressure, $(\overline{P})_o$, that will give maximum profit. One of the key component specifications, $\overline{HK}_D$ or $\overline{LK}_B$, can be easily determined from the general optimizing policies. Often, the incremental gain in recovery of the most valuable product will not exceed the incremental gain resulting from increasing feed flow rate. This means that both products operating at minimum purity will allow the largest quantity of feed to be charged for maximum profit rate.

Optimum Concentrations Both products must be produced at minimum purity to achieve the most profitable operation for this case. Therefore,

$$\overline{LK}_D = (\overline{LK}_D)_{ss} \qquad \textbf{8.21(107)}$$

$$\overline{HK}_B = (\overline{HK}_B)_{ss} \qquad \textbf{8.21(108)}$$

The concentration for each control component ($\overline{LK}_B$ and $\overline{HK}_D$) must be as follows to satisfy Equations 8.21(107) and 8.21(108):

$$(\overline{LK}_B)_o = (\overline{LK}_B)_{max} \qquad \textbf{8.21(109)}$$

$$(\overline{HK}_D)_o = (\overline{HK}_D)_{max} \qquad \textbf{8.21(110)}$$

$(\overline{HK}_D)_{max}$ and $(\overline{LK}_B)_{max}$ are given by Equations 8.21(93) and 8.21(94). The separation is optimized independently by column pressure $(\overline{P})$ and feed rate (F).

Loading Constraint As was previously explained, the overhead vapor condenser limits loading for this example. Column

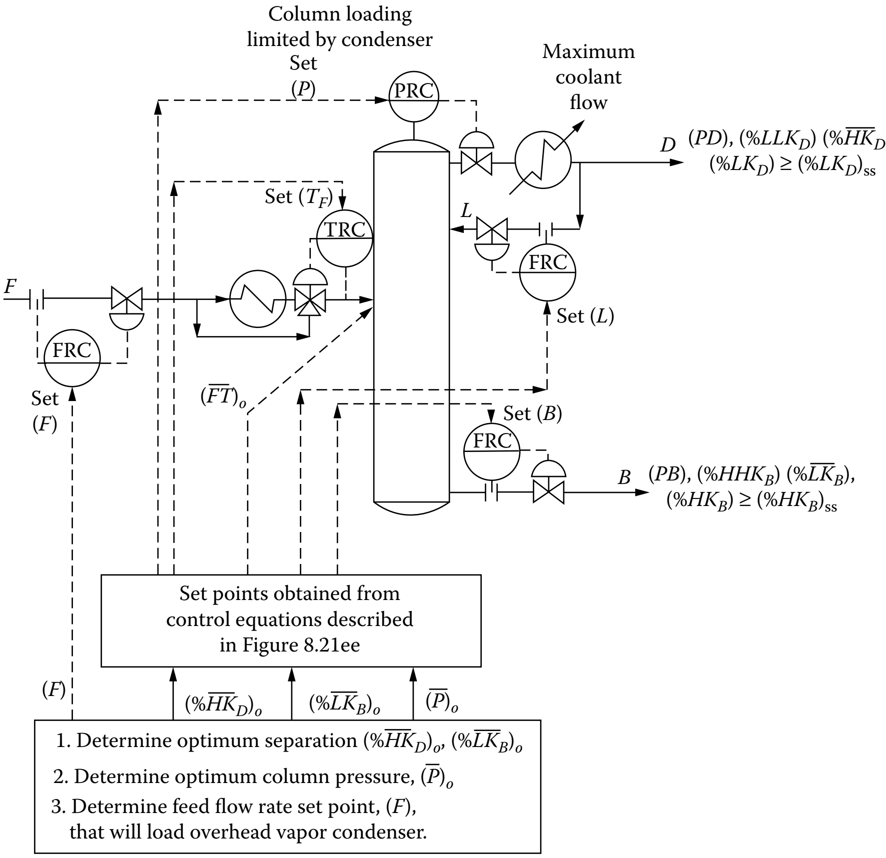


FIG. 8.21jj
When an unlimited market exists for the products and the product prices (PD, PB) are known, the optimum separation, the optimum column pressure, and the value for feed flow rate that will give maximum loading must be found.

pressure must, therefore, be operated at maximum in order to obtain maximum condensing capacity. (ΔT across the condenser tubes will be the largest at maximum column pressure, and therefore maximum condenser capacity will result.) The maximum pressure that can be used will be determined by one of the following five constraints, assuming operating costs do not become prohibitive before a physical constraint is reached:

1. Column downcomer capacity
2. Reboiler capacity
3. Upstream equipment pressure specifications
4. Pressure rating of the column shell
5. Fouling of the reboiler or condenser tubes

As column pressure is increased, the capacities of the reboiler and tower trays will be approached. Also, the pressure rating of the shell and of other processing equipment will be approached. The one of the five constraints that is approached first, as column pressure is raised, will set the operating pressure. Assume for the purpose of illustration that the shell pressure rating is the limit on column pressure. Pressure can, therefore, be set constant at the pressure rating of the column shell.

Optimum Feed Flow Rate After the optimum separation $(\overline{LK}_B)_o$ and $(\overline{HK}_D)_o$ and the optimum column operating pressure $(\overline{P})_o$ have been determined, the feed rate can be increased until the condenser capacity is approached. There are several ways that feed rate can be manipulated to maintain condenser loading. One very useful method involves a predictive control technique.

Overhead vapor flow rate can be expressed in terms of various independent inputs and terms obtained from the operating control equations. Development of the predictive control equation proceeds as follows:

$$V_o = D + L \qquad \textbf{8.21(111)}$$

where
V_o = vapor flow rate overhead
D = top product flow rate
L = external reflux flow rate

First, eliminate D by $D = F - B$. Now, the maximum overhead vapor flow rate will be given by Equation 8.21(85) and Equation 8.21(111) can be set equal.

$$a_1 + a_2(\Delta T) + a_3(\Delta T)^2 = L + F - B \qquad \textbf{8.21(112)}$$

or

$$F = a_1 + a_2(T_o - T_c) + a_3(T_o - T_c)^2 + B - L \qquad \textbf{8.21(113)}$$

where

a_1, a_2, a_3 = coefficients for condenser-loading equation that are determined by column tests
F = feed flow rate
T_o = overhead vapor temperature
T_c = temperature of coolant to overhead vapor condenser
B = bottom product flow rates from the output of operating control equation
L = reflux flow rates from the outputs of operating control equation

Temperature of the overhead vapor will be a function of all independent inputs to the system. However, column pressure is usually the main variable of concern. For this example let

$$T_o = f(\bar{P}) \qquad \textbf{8.21(114)}$$

where $f(\bar{P})$ is some function of column pressure.

In many cases $f(\bar{P})$ can be considered a linear function such as

$$f(\bar{p}) = d_1 + d_2(\bar{p}) \qquad \textbf{8.21(115)}$$

This equation can be determined off-line from correlation of data obtained from flash calculations at the average composition of the existing overhead vapor. If changes in composition of the overhead vapor affect temperature of the overhead vapor by a significant amount, then composition also has to be taken into account. Composition for the overhead vapor can be easily approximated from feed composition analysis. If Equation 8.21(113) is carried to this extent, then the feed flow rate can be predicted to keep the condenser against its maximum capacity. For the purpose of illustration here, T_o is assumed to be a function of column pressure only.

Eliminate T_o from Equation 8.21(113) by Equations 8.21(114) and 8.21(115) to obtain

$$a_1 + a_2[d_1 + d_2(\bar{P}) - T_c] + a_3[d_1 + d_2(\bar{P}) - T_c]^2\ B - L = F_{\max} \qquad \textbf{8.21(116)}$$

$\bar{P}$ and T_c are measured, and B and L are obtained from the operating equations' set point calculations. $F_{\max}$ will be the feed rate required to load the condenser for the particular values of $(\overline{HK_D})_o$ and $(\overline{LK_B})_o$.

Figure 8.21kk shows the overall optimizing control system. Only the necessary basic controls are shown. The other controls can be assumed to be as shown in Figure 8.21t.

Reboiler Limiting Let us now assume that loading is limited by the reboiler instead of the condenser. Optimum separation remains the same. However, now column pressure must be operated at a minimum value in order to gain maximum reboiler capacity. For this example, assume that minimum column pressure is set by the pressure requirements of downstream equipment. Therefore, column pressure is set at a constant value and will not be changed unless the pressure requirements of downstream equipment are changed.

Having achieved the optimum separation (minimum purity of products) and optimum column pressure, the feed rate can now be increased up to the maximum capacity of the reboiler. This, then, will represent the most profitable operation.

Again, manipulation of the feed flow rate can be handled by a predictive control technique. Liquid flow rate below the feed tray (L_f) is given by

$$L_f = L_i + F_i \qquad \textbf{8.21(117)}$$

where F_i is the internal feed flow rate.

$$F_i = F[1 + (K_F)(T_v - T_t)] \qquad \textbf{8.21(118)}$$

where

K_F = a constant equal to the specific heat of the feed divided by the heat of vaporization
T_v = temperature of vapor above the feed tray
T_f = temperature of feed at column entry

The vapor flow rate out of the reboiler is given by

$$V_B = L_f - B \qquad \textbf{8.21(119)}$$

Now, substitute Equation 8.21(118) into Equation 8.21(117) to eliminate F_i. Then, substitute Equation 8.21(117) into Equation 8.21(119) to eliminate L_f. The following is obtained:

$$V_B = L_i + F[1 + (K_F)(T_v - T_f)] - B \qquad \textbf{8.21(120)}$$

Next, substitute $(L_i/F)F$ for L_i, and $(B/F)F$ for B. Then solve for F to obtain

$$F = \frac{(V_B)_{\max}}{(L_i/F) + 1 + (K_F)(T_v - T_f) - (B/F)} \qquad \textbf{8.21(121)}$$

where

F = set point of feed flow controller
$(V_B)_{\max}$ = maximum reboiler heat input rate

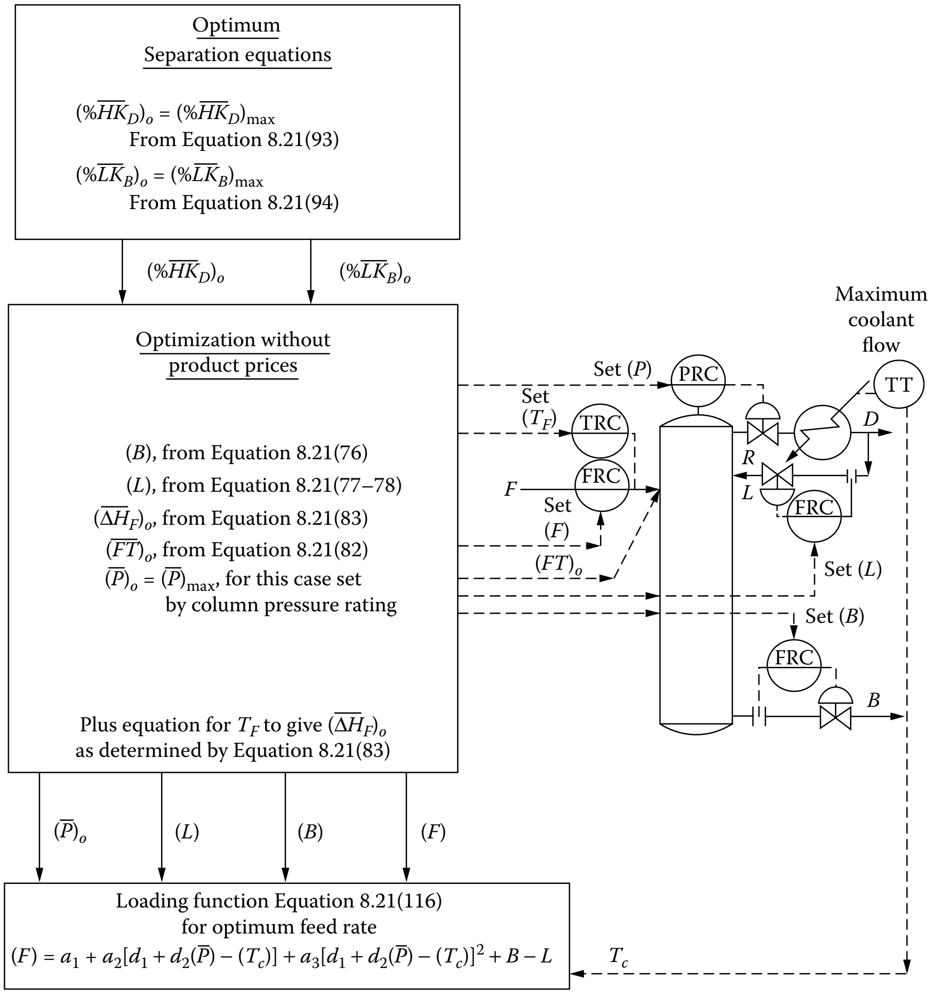


FIG. 8.21kk
Optimizing for maximum profit rate when unlimited market exists for the products and loading is limited by condenser capacity.

Equation 8.21(121) calculated that feed flow rate which is required to cause the vapor rate $(V_B)_{\max}$ to exist for all separations specified. L_i/F and B/F are obtained from the operating control equations used to achieve a suboptimum operation. T_v and T_f are measured.

Equation 8.21(121) is used by specifying $(V_B)_{\max}$ and then evaluating the column operation. After sufficient time for the column to stabilize, the reboiler valve position (output of reboiler heat flow controller) is observed. If, for example, the reboiler valve is 85% open, $(V_B)_{\max}$ can then be increased until the reboiler heat control valve is near its maximum opening, say, 95% open.

Enough room must be left to maintain control. $(V_B)_{\max}$ can be adjusted by the plant operator to maintain the reboiler valve near open or can be handled automatically by a valve-position-based feedback controller. Once $(V_B)_{\max}$ is established by experience, few adjustments will be required to maintain the column in a fully loaded condition. Adjustments to $(V_B)_{\max}$ will be required only as the heat transfer capability of the reboiler varies. The control scheme is illustrated in Figure 8.21ll.

CONCLUSIONS

Example solutions to some of the common distillation column optimizing problems have been given. Although many different situations can exist, they usually are combinations of those presented.

Optimization by feedback control methods cannot approach the quality of control obtained by predictive (feedforward) techniques. This is true even though the predictive control equations may require updating by feedback. In effect, predictive optimization control greatly attenuates any error that must be handled by feedback (updating).

The application of feedforward optimizing control forces development of mathematical models of the component parts

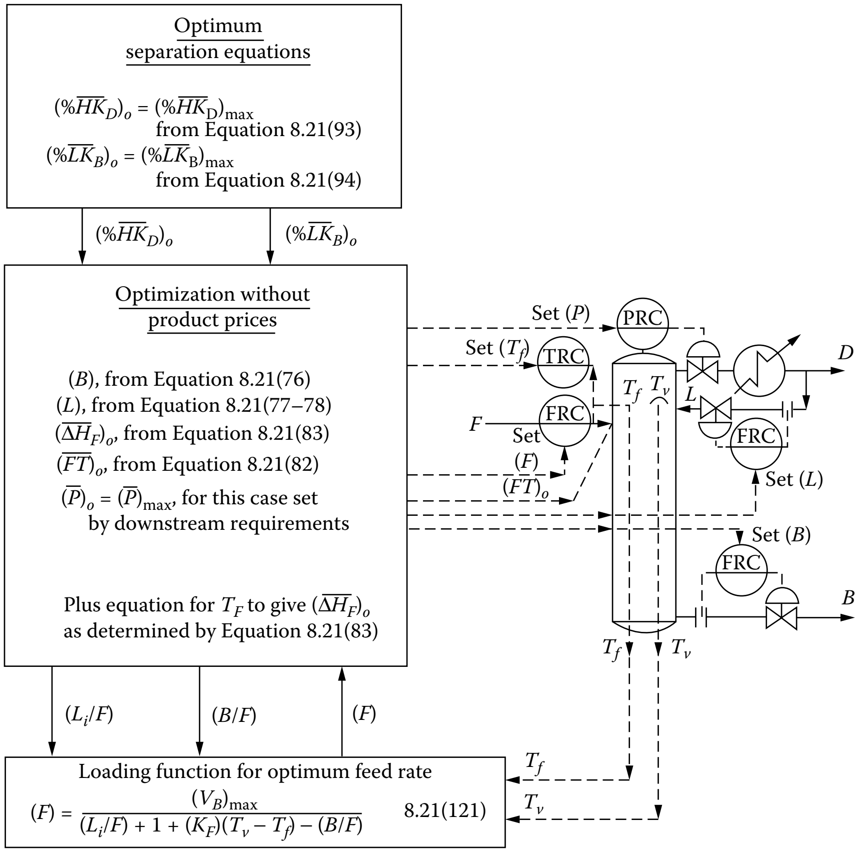


FIG. 8.21ll

Optimization of a distillation column when unlimited market exists for the products, prices of the products are known, and loading is limited by the reboiler.

of a process. The mathematical models developed for optimizing unit operations will eventually be required to extend optimization to include an entire plant complex.

The APC products most applicable to distillation modeling are the white box models, where the theoretical dynamic models are derived on the basis of the mass, energy, and momentum balances of the process. Fuzzy logic and black box models are used less often, as they are more applicable to processes that are poorly understood or when it is acceptable to use a complete mechanistic empirical model constructed solely from a priori knowledge.

References

1. Lipták, B. G., "Process Control Trends," *Control,* January, March, and April 2004.
2. Jensen, B. A., and Collins, P. L., "Incentives for Tighter Fractionator Control," *Control,* November 1990.
3. Smith, D. E., Stewart, W. S., and Griffin, D. E., "Distill with Composition Control," *Hydrocarbon Processing,* February 1978.
4. Van Kampen, J. A., "Automatic Control by Chromatograph of a Distillation Column," Convention on Advances in Automatic Control, Nottingham, England, April 1961.
5. Shinskey, F. G., *Process-Control Systems, Application, Design, Adjustment,* 3rd edition, New York: McGraw-Hill Book Company, 1988.
6. Luyben, W. L., "Distillation Decoupling," *AIChE Journal,* Vol. 16, No. 2, pp. 198–203, March 1970.
7. Shinsky, F. G., "The Stability of Interacting Loops with and without Decoupling," presented at the IFAC Symposium on Multivariable Control, Fredericton, New Brunswick, July 4–8, 1977.
8. Gordon, L. M., "Practical Evaluation of Relative Gains: The Key to Designing Dual Composition Controls," *Hydrocarbon Processing,* December 1982.
9. Shinskey, F. G., *Distillation Control for Productivity and Energy Conservation,* 2nd edition, New York: McGraw-Hill Book Company, 1984.
10. Ryskamp, C. J., "New Control Strategy Improves Dual Composition Control," *Hydrocarbon Processing,* June 1980.
11. Van Horn, L. D., "Crude Unit Computer Control: How Good Is It?" *Hydrocarbon Processing,* April 1980.
12. Garcia, C. E., and Morari, M., "Internal Model Control: 1. A Unifying Review and Some New Results," *Industrial & Engineering Chemistry Process Design & Development,* Vol. 21, pp. 308–323, 1982.
13. Richalet, J. A., Rault, A., Testud, J. L., and Papon, J., "Model Predictive Heuristic Control: Applications to an Industrial Process," *Automatica,* Vol. 14, pp. 413–428, 1978.
14. Cutler, C. R., and Ramaker, B. L., "Dynamic Matrix Control: A Computer Control Algorithm," presented at the AIChE 86th National Meeting, Houston, TX, 1979.

15. Singh, D. K., "Neural Network Model-Based Control of Chemical Processes," *Proceedings of the ISA/90 International Conference and Exhibit,* Part 2, New Orleans, LA, October 1990, pp. 965–968.
16. Douglas, J. M., Jafarey, A., and McAvoy, T. J., "Short-Cut Techniques for Distillation Column Design and Control, Part 1: Column Design," *I&EC Process Design and Development,* Vol. 18, pp. 197–202, April 1979.
17. Lee, P. L., and Sullivan, G. R., "Generic Model Control," *Computers and Chemical Engineering,* Vol. 12, p. 573, 1988.
18. Tai, H. M., Wang, J., and Ashenayi, K., "A Neural Network Tracking Controller for High-Performance Applications," *Proceedings of the Oklahoma Symposium on Artificial Intelligence,* November 1991, p. 225.
19. Bannon, R. et al., "Heat Recovery in Hydrocarbon Distillation," *Chemical Engineering Progress,* July 1978.

Bibliography

For pre-1990 literature, refer to Section 8.19

1990

Cingara, A., Jovanovic, M., and Mitrovic, M., "Analytical First-Order Dynamic Model of Binary Distillation Column," *Chemical Engineering Science,* Vol. 45, No. 12, pp. 3585–3592, 1990.

Ding, S. S. and Luyben, W. L., "Control of a Heat-Integrated Complex Distillation Configuration," *Industrial & Engineering Chemistry Research,* No. 29, p. 1240, 1990.

Farrel, R. J. and Polli, A., "Comparison of Unconstrained Dynamic Matrix Control to Conventional Feedback Control for a First Order System," *Proceedings of the ISA/90 International Conference and Exhibit,* Part 2, New Orleans, LA, October 1990, pp. 1037–1046.

Gerstle, J. G., Hokanson, D. A., and Anderson, B. O., "Multivariable Control of a C2 Splitter," *Proceedings of the AIChE Annual Meeting,* Chicago, IL, November 1990.

Kister, H. Z., *Distillation Operation,* New York: McGraw-Hill Book Company, 1990.

Li, R., Olson, J. H., and Chester, D. L., "Dynamic Fault Detection and Diagnosis Using Neural Networks," *Proceedings of the Fifth IEEE International Symposium on Intelligent Control,* Philadelphia, PA, September 1990, pp. 1169–1174.

Papastathopoulou, H. S. and Luyben, W. L., "Potential Pitfalls in Ratio Control Schemes," *Industrial & Engineering Chemistry Research,* October 1990, pp. 2044–2053.

Pitt, M. J., *Instrumentation and Automation in Process Control,* New York: E. Horwood, 1990.

Riggs, J. B., "Advanced Model-Based Control of a Sidestream Draw Column," *Proceedings of the ISA/90 International Conference and Exhibit,* Part 2, New Orleans, LA, October 1990, pp. 1023–1032.

Riggs, J. B., Watts, J., and Beauford, M., "Industrial Experience with Applying Nonlinear Process Model-Based Control to Distillation Columns," *Proceedings of the ISA/90 International Conference and Exhibit,* Part 2, New Orleans, LA, October 1990, pp. 1047–1054.

Skogestad, S., Jacobsen, E. W., and Morari, M., "Inadequacy of Steady-State Analysis for Feedback Control, Distillate, Bottom Control of Distillation Columns," *Industrial & Engineering Chemistry Research,* December 1990, pp. 2339–2346.

1991

Coughanowr, D. R., *Process Systems Analysis and Control,* 2nd edition, New York: McGraw-Hill Book Company, 1991.

Farhat, S., Piouleau, L., Domenech, S., and Czernicki, M., "Optimal Control of Batch Distillation via Nonlinear Programming," *Chemical Engineering and Processing,* Vol. 29, No. 1, pp. 33–38, January 1991.

Figuerosa, J. L., Desages, A. C., Romagnoli, J. A., and Palazoglu, "Highly Structured Stability Margins for Process Control Systems. A Case Study of Decoupling Control in Distillation," *Computers & Chemical Engineering,* Vol. 15, No. 7, pp. 493–502, July 1991.

Gerstle, J. G. and Hokanson, D. A., "Experiences with Applying Dynamic Matrix Control in Olefins Plants," *Proceedings of the ISA/91 International Conference and Exhibit,* Part 1, Anaheim, CA, October 1991.

Houk, B. G., Snowden, D. L., and Stevenson, T. E., "Improved Control of Ethylene Recovery Train Using Dynamic Matrix Control," *Proceedings of the ISA/91 International Conference and Exhibit,* Part 1, Anaheim, CA, October 1991.

Jacobsen, E. W., and Skogestad, S., "Control of Unstable Distillation Columns," *Proceedings of the American Control Conference,* Green Valley, AZ, Vol. 1, pp. 773–778, 1991.

Jensen, B. A., "Improve Control of Cryogenic Gas Plants," *Hydrocarbon Processing,* May 1991.

Kettinger, J. G., Gaines, L. D., and McGee, N. F., "Multivariable Control a Depropanizer," *Proceedings of the ISA/91 International Conference and Exhibit,* Part 1, Anaheim, CA, October 1991.

O'Connor, D. L., Grimstad, K., and McKay, J., "Application of a Single Multivariable Controller to Two Hydrocracker Distillation Columns in Series," *Proceedings of the ISA/91 International Conference and Exhibit,* Part 1, Anaheim, CA, October 1991.

Papadopoulos, M. N. and Berkowitz, P. N., "Multivariable (MVC) Process Control. A User-Based On-Line Optimal Control System," *Proceedings of the Industrial Computing Conference,* Anaheim, CA, October 1991.

Papastathopoulou, H. S. and Luyben, W. L., "Control of a Binary Side-Stream Distillation Column," *Industrial & Engineering Chemistry Research,* April 1991, pp. 705–713.

Patwardhan, A. A. and Edgar, T. F., "Nonlinear Model-Predictive Control of a Packed Distillation Column," *Proceedings of the American Control Conference,* Green Valley, AZ, Vol. 1, 1991, pp. 767–772.

Radhakrishnan, T. K., and Gangiah, K., "Critical Study of Multivariable Self-Tuning Algorithms for Distillation Control," *Chemical Engineering & Technology,* Vol. 14, No. 6, pp. 399–405, December 1991.

Sandelin, P. M., Haeggblom, K. E., and Waller, K. V., "Disturbance Rejection Properties of Control Structures at One-Point Control of a Two-Product Distillation Column," *Industrial & Engineering Chemistry Research,* June 1991, pp. 1182–1186.

Sandelin, P. M., Haeggblom, K. E., and Waller, K. V., "Disturbance Sensitivity Parameter and Its Application to Distillation Control," *Industrial & Engineering Chemistry Research,* June 1991, pp. 1187–1993.

Tinetti, G., Ghassan, A., and Cheung, J. Y., "Comparing Intelligent and Classical Control Systems in a Nonlinear Environment," *Proceedings of the Oklahoma Symposium on Artificial Intelligence,* November 1991, p. 199.

Yang, D. R., Seborg, D. E., and Mellichamp, D. A., "Combined Balance Control Structure for Distillation Columns," *Industrial & Engineering Chemistry Research,* September 1991, pp. 2159–2168.

Yang, D. R., Waller, K. V., Seborg, D. E., and Mellichamp, D. A., "Dynamic Structural Transformations for Distillation Control Configurations," *AIChE Journal,* Vol. 36, No. 9, pp. 1391–1402, September 1991.

1992

Brambilla, A. and D'Elia, L., "Multivariable Controller for Distillation Columns in the Presence of Strong Directionality and Model Errors," *Industrial & Engineering Chemistry Research,* February 1992, pp. 537–543.

Luyben, W. L. (ed.), *Practical Distillation Control*, New York: Van Nostrand Reinhold, 1992.

Rovaglio, M., Raravelli, T., Biardi, G., Gaffuri, P., and Soccol, S., "Precise Composition Control of Heterogeneous Azeotrophic Distillation Towers," *Computers & Chemical Engineering,* Vol. 16, 1992.

Yiu, Y., and Papadopoulos, M. N., "On-Line Optimal Advanced Control of a Ryan Holmes Gas Processing Process," *Advances in Instrumentation and Control,* Vol. 47, Part 1, Research Triangle Park, NC: Instrument Society of America, 1992, pp. 967–975.

1993

Anderson, B., and Mejdell, T., "Using Temperature Profile for Product Quality Estimation on a Distillation Column," *Advances in Instrumentation and Control,* Vol. 48, Part 1, Research Triangle Park, NC: Instrument Society of America, 1993, pp. 39–46.

Freitas, M. S., Campos, M. C. M. M., and Lima, E. L., "Dual Composition Control of a Debutanizer Column," *Advances in Instrumentation and Control,* Vol. 48, Part 1, Research Triangle Park, NC: Instrument Society of America, 1993, pp. 501–509.

Ganguly, S., "Model Predictive Control of Distillation," ISA/93 Technical Conference, Chicago, IL, September 19–24, 1993.

McKetta, J. J. (ed.), *Unit Operations Handbook, Vol. 1: Mass Transfer*, New York: Marcel Dekker, 1993.

1994

Gokhale, V., Shukla, N., and Munsif, H., "Analysis of Advanced Distillation Control on a C_3 Splitter and a Depropanizer," 1994 AIChE National Annual Meeting, San Francisco, CA, November 1994.

1995

Banerjee, A., and Arkun, Y., "Control Configuration Design Applied to the Tennessee Eastman Plantwide Control Problem," *Computers. Chem. Engng.,* 19(4), 453–480, 1995.

Diwekar, U. M., *Batch Distillation: Simulation, Optimal Design and Control (Series in Chemical and Mechanical Engineering),* Taylor & Francis, September 1995.

Fleming, B., and Sloley, A.W., "Feeding and Drawing Products: The Forgotten Part of Distillation," *Proceedings of the ChemShow and Exposition,* New York, December 1995.

Hurowitz, S. E., and Gokhale, V., "A Dynamic Model of a Superfractionator: A Test Case for Comparing Distillation Control Techniques," DYCORD '95, 4th IFAC Symposium, Helsingor, Denmark, June 1995.

Lundstrom, P., and Skogestad, S., "Opportunities and Difficulties with 5 × 5 Distillation Control," *J. Process Control,* 1995, 5, 249–261.

Musch, H. E., and Steiner, M., "Robust PID Control for an Industrial Distillation Column," *Control System Magazine,* 1995, 15, 4, 46–55.

Rawlings, J. B., "Dynamics and Control of Chemical Reactors, Distillation Columns, and Batch Processes (Dycord '95)," a postprint volume from the 4th IFAC Symposium on Dynamics and Control of Chemical Reactors, Distillation Columns, and Batch Processes (DYCORD '95).

1996

Koggersbøl, A., Andersen, B. R., Nielsen, J. S., and Jørgensen, S. B.,"Control Configuration for Energy Integrated Distillation," *Computers & Chem Eng.,* 20 (supplement), pp. S853–S858, 1996.

Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, pp. 340–347.

1997

Anderson, N. A., *Instrumentation for Process Measurement and Control,* 3rd edition, Boca Raton, FL: CRC Press, October 1997.

Hurowitz, S. E., and Anderson, J. J., "Distillation Configuration Selection for Dual Composition Control," AIChE Spring National Meeting, Houston, TX, April 1997.

Hurowitz, S. E., and Anderson, J. J., "Control of High-Purity Distillation Columns," Control 97 Conference, Sydney, Australia, October 1997.

Linsley, J., "New, Simpler Equations Calculate Pressure-Compensated Temperatures," *Oil & Gas Journal,* May 24, 1997, pp. 58–64.

Mahoney, D. P., and Fruehauf, P. S., "An Integrated Approach for Distillation Column Control Design Using Steady State and Dynamic Simulation," Aspentech technical articles, March 1997.

Tham, M. T., "Distillation," Base Document URL: http://lorien.ncl.ac.uk/ming/distil/distil0.htm, Date: October 1997.

Skogestad, S., "Dynamics and Control of Distillation Columns: A Tutorial Introduction," *Trans. IChemE,* Vol. 75, Part A, pp. 539–562, 1997.

1998

Betlem, B. H. L., Krijnsen, H.C., and Huijnen, H., "Optimal Batch Distillation Control Band on Specific Measures," *Chemical Engineering Journal,* 71, pp. 111–126, 1998.

Luyben, W. L., Tyreus, B. D., and Luyben, M. L., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Ochiai, S., "Calculating Process Control Parameters from Steady State Operating Data," *ISA Transactions,* Vol. 36, No. 4, pp. 313–320, 1998.

Riggs, J. R., "Improve Distillation Column Control," *Chemical Engineering Progress,* October 1998, pp. 31–47.

Stichlmair, J. G., and Fair, J. R., *Distillation: Principles and Practices,* New York: John Wiley & Sons, 1998.

1999

Hurowitz, S. E., Anderson, J., Duvall, M., and Riggs, J. B., "An Analysis of Controllability Statistics for Distillation Configuration Selection," Presented at the AIChE Annual Meeting, Dallas, TX, November 1999.

Eker, I., and Sakthivel, K., "Automation & Lube Oil Additives Blending Plant Using an S88.01-Consistent Batch Software: A Case Study," *Proceedings of the World Batch Forum,* San Diego, CA, April 1999.

2000

Sloley, A. W., "Steady under Pressure: Distillation Pressure Control," presented at the American Institute of Chemical Engineers Spring Meeting, March 6–9, 2000.

Betlem, B. H. L., "Batch Distillation Column Low-Order Models for Quality Control Program," *Chemical Engineering Science,* 55, pp. 3187– 3194, 2000.

Hugo, A., "Limitations of Model Predictive Controllers," *Hydrocarbon Proc*., January 2000, p. 86.

Roffel, B., Betlem, B. H. L., and De Ruijter, J. A., "Modeling and Control of a Cryogenic Distillation Column," *Computers and Chemical Engineering,* 24, pp. 111–123, 2000.

Roffel, B., "Distillation: Instrumentation and Control Systems," in *Encyclopedia of Separation Science,* Academic Press, 2000.

Willis, M. J., "Selecting a Distillation Column Control Strategy (A Basic Guide)," University of Newcastle, UK: 2000.

2002

Florez, M., "Batch Distillation: Practical Aspects of Design and Control," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2002.

Cook, B., Engel, M., Landis, C., Tedeschi, S., and Zehnder, A., "Synthesis of Optimal Batch Distillation Sequences," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2002.

2003

Blevins, T. L., McMillan, G. K., Wojsznis, W. K., and Brown, M. W., *Advanced Control Unleashed,* Research Triangle Park, NC: ISA, 2003.

Kralj, F., "Application of the S88 Model in the Control of Continuous Distillation Facilities," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

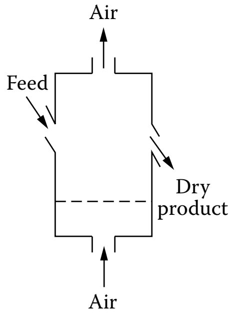
2004

Hurowitz, S., Anderson, J., Duvall, M., and Riggs, J. B., "Distillation Control Configuration Selection," submitted to *J. Process Control,* March 2004.

Lipták, B. G., "Process Control Trends," *Control,* January, March, and May, 2004.

8.22 Dryer Controls

B. BLOCK (1970, 1985) **B. G. LIPTÁK** (1995) **F. G. SHINSKEY** (2005)



Flow sheet symbol

INTRODUCTION

The equipment discussed in this section serves to dry solids. The process involves the removal of liquids, such as water or other solvents, by adding heat to vaporize them. The types of dryer designs discussed can be grouped into batch and continuous. They can be further subdivided on the basis of whether the dryer itself is heated (*nonadiabatic*) or unheated and the heat of vaporization is provided by preheated air (*adiabatic*).

Adiabatic dryers can be *co-current* (the solids and the drying air move in the same direction), *countercurrent* (hot air and solids move through the dryer in opposite directions), or *fluidized bed*. The control of each of these designs will be briefly discussed in this section.

The subject of dryers is so extensive that some limitations must be imposed on the discussion. Therefore:

1. The dryers and drying principles discussed will be limited to those related to the removal of a volatile solvent from a solid material. Such processes as the removal of a solvent from an air stream (air drying) or the removal of one solvent from another (drying of an organic solvent) will not be covered.
2. The solvent to be removed will be taken as water.
3. The heating medium used may be steam or the firing of a fuel.

ADIABATIC DRYING

The energy source for adiabatic drying is the sensible heat of a stream of air directed toward a bed or a stream of moist solids. Heat flows from the air to the solids, evaporating some of their moisture — in this respect, the governing force for drying is that of heat transfer. In this process the air is cooled, but its energy content (enthalpy) does not change, because the sensible heat loss equals the latent heat associated with its increased moisture content.

The Psychrometric Chart

Figure 8.22a is a psychrometric chart relating the absolute and relative humidity of air to its temperature and enthalpy. The vertical scale is absolute humidity in grains of moisture carried by a pound of dry air (1 lb is 7000 grains). Dry-bulb temperature is the horizontal scale.

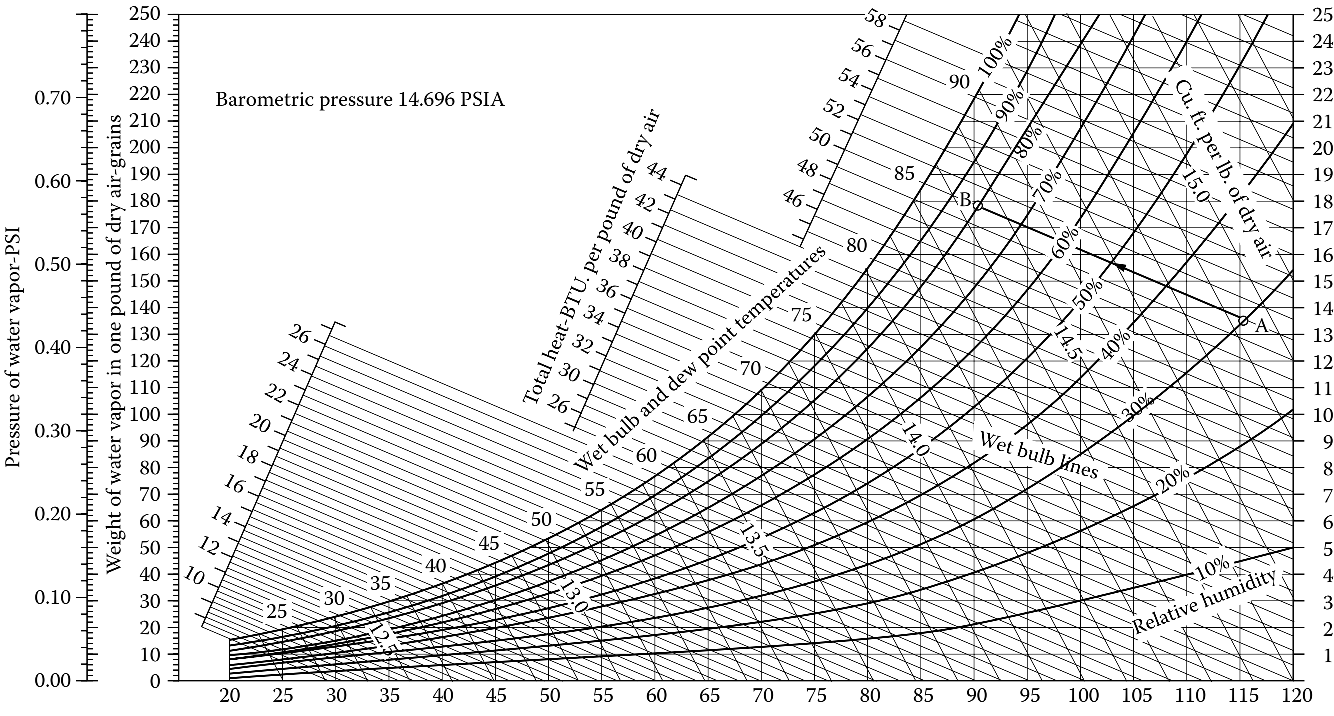
The uppermost curve is the saturation limit — 100% relative humidity, the remaining curves representing lower levels of RH. The temperature of the air at 100% RH is also its dew point. Consider, for example, a hot summer day with a (dry-bulb) temperature of 85°F and a RH of 75%. The absolute humidity of the air would be about 137 grains/lb dry air, and its dew point would be about 76°F.

Such air might be used to dry lumber, for example, by heating it to 115°F. Point A on the psychrometric chart would then represent the condition of the heated air that is sent to the dryer — notice that its RH at that point is only 30%.

The scale in the center of the chart is marked in units of BTU/lb dry air total heat, or enthalpy. As the air passes through the adiabatic dryer, its enthalpy will remain at 50 BTU/lb, while it cools and evaporates moisture from the solids. The path it takes on the chart then follows a line of constant enthalpy sloping upward to the left toward point B, where it is exhausted from the dryer at 80%RH. These lines of constant enthalpy are also lines of constant wet-bulb temperature.

A wet-bulb thermometer's bulb is covered by a wick that is saturated with water. It is whirled in the air, or has a stream of air blown across it. The equilibrium temperature reached during this process is known as the wet-bulb temperature and can be used along with dry-bulb temperature to estimate RH, absolute humidity, and dew point from the chart. It is also a direct indication of enthalpy — note that the wet-bulb lines on the chart are also lines of constant enthalpy.

Wet-bulb temperature is particularly important in a drying operation, because the wet bulb is an analog of the solid material being dried. As moisture evaporates from the solid in an adiabatic dryer, its temperature approaches that of a wet-bulb thermometer in the same air, and it remains at that



Copyright, 1942, General Electric Company

FIG. 8.22a
Psychrometric chart.

temperature through most of the drying operation. Measuring the wet-bulb temperature in a dryer or even in the air entering a dryer is difficult, however, because the wick must be continuously supplied with just the right flow of water, and the wick is subject to plugging from dust and any dissolved solids in the water (Figure 8.22b).

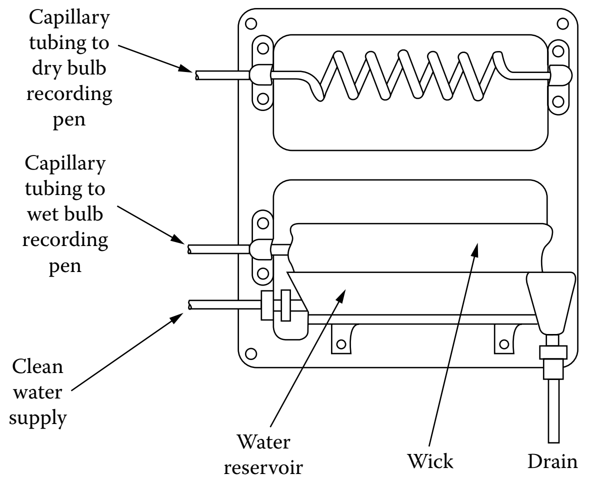


FIG. 8.22b
Wet-dry bulb hygrometer using filled system temperature sensors arranged for duct mounting.

As a result, wet-bulb temperature is rarely measured in situ, but can be calculated from other measurements such as dry-bulb and dew-point temperatures. The exception is in a lumber kiln, where conditions are relatively mild and dust-free.

Batch Drying

Many materials are dried in batches; for example, fine powders, which are later encapsulated into aspirin and other tablets. The drying is typically conducted in a fluid bed such as shown in Figure 8.22c.

A basket fitted with fine-mesh screens on bottom and top is charged with wet powder and wheeled into the dryer, where it is sealed in place. The exhaust fan is started, drawing air upward through the bed, fluidizing the wet powder. The air is heated by a steam coil to a controlled temperature before it contacts the solids.

Initially, the solids will be below the wet-bulb temperature of the heated air and will warm to that value within minutes, where evaporation will begin. The first water to be removed will be surface moisture, and it will be removed at a constant rate until the surface is partially dry. Migration of moisture from within the solids then progresses at a reduced rate. If drying is continued indefinitely, an equilibrium temperature will eventually be reached where no more moisture is removed.

When the product is later exposed to atmospheric conditions, the equilibrium will shift, however, and some moisture

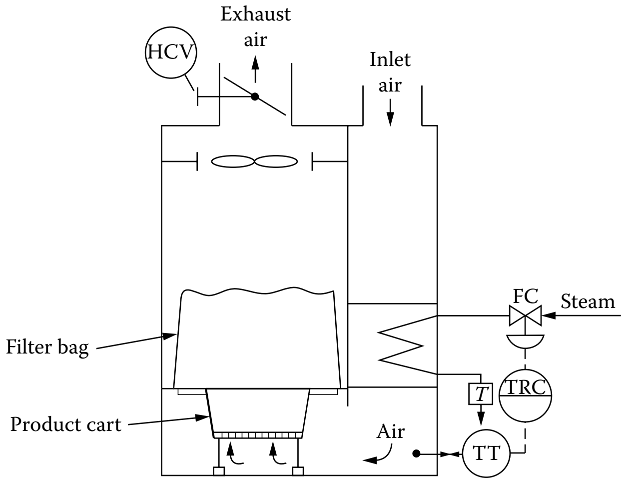


FIG. 8.22c
Batch fluid-bed dryer.

may be readsorbed from the air. A profile of moisture content against time for a typical batch dryer is given in Figure 8.22d.

The curve consists of four zones. Section A-B represents the period of preheating of the solid to the initial drying temperature; evaporation during this phase is slow. The next section, B-C, is a period of evaporation of surface moisture or moisture that migrates readily. The temperature near the surface of the solids during this time is the wet-bulb temperature of the air in contact with the product.

After the surface moisture has evaporated, the rate of evaporation drops off (section C-D). This reduction may be due in part to a case-hardening of the surface and in part to the long path necessary for the water to migrate to the surface. If the solid has water-of-crystallization, or bound water, the rate will then drop off even more, as shown in section D-E of the curve. At point E, equilibrium is established and the drying rate is zero.

Other curves are also useful for an understanding of dryer control requirements. The first of these, Figure 8.22f, represents the derivative of Figure 8.22e, that is, the variation in rate of drying as a function of time.

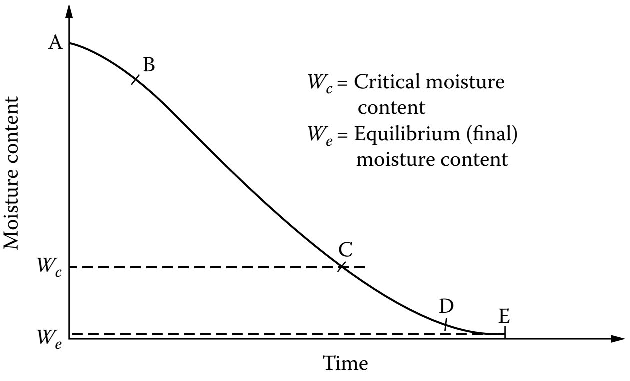


FIG. 8.22d
Typical batch drying curve.

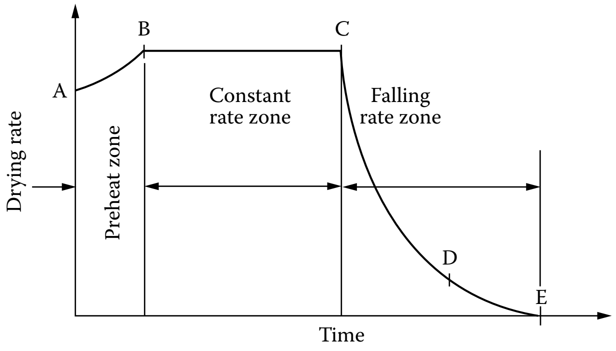


FIG. 8.22e
Rate of drying vs. time.

Another curve, Figure 8.22g, combines these two to relate the rate of drying to the moisture content of the solid.

In virtually all dryers, the desired moisture content of the product lies in the falling-rate zone, i.e., below its critical moisture content. In this zone, the moisture content of the product is proportional to the rate of drying, in fact, approximately linear with it, which gives the dryer a degree of self-regulation.

The lower the moisture content, the harder it is to remove. The relationship between rate of drying and driving force allows for moisture control without an on-line measurement.

It is common for operators of batch dryers to stop the operation when it is estimated that the product is close to its desired moisture content—on the basis of time or intuition—and sample it. The sample is typically placed in a loss-in-weight analyzer where it is weighed, dried thoroughly by

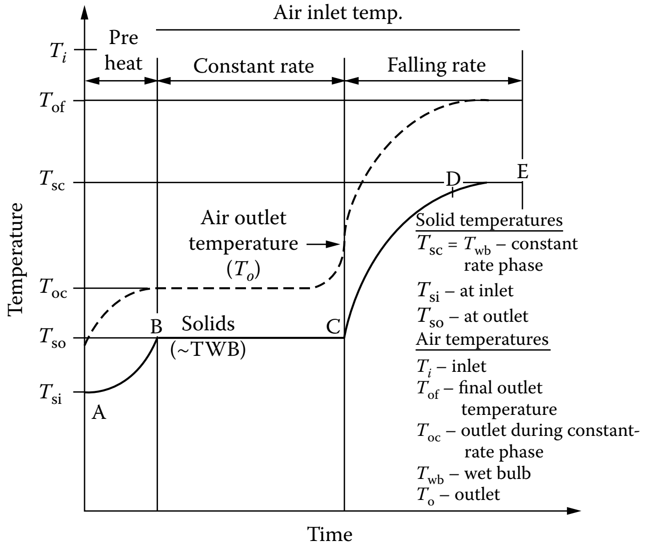


FIG. 8.22f
Product temperature rises in a batch dryer from T_{si} to T_{so} as it is heated by air, which enters at T_i and leaves at T_o.

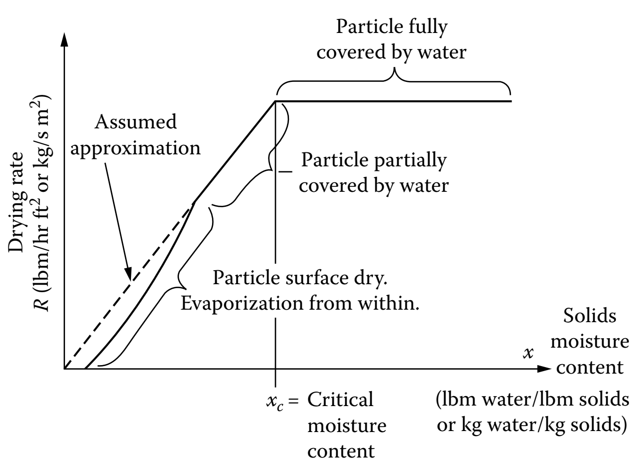


FIG. 8.22g
Drying rate versus moisture content.

infrared radiation, and weighed again, which takes a few minutes.

If the product moisture is still too high, the drying will be resumed for a time and the product sampled again. If the moisture is too low, the product may have to be sprayed with water and drying repeated. However, the conditions corresponding to the desired moisture can be estimated on-line, and the dryer can be automatically shut down when those conditions are met, avoiding unnecessary sampling and thereby saving time. To do this, a model of the drying process is needed.

Dryer Model

The rate of evaporation from a solid is proportional to its moisture content x (following the linear approximation) up to its critical moisture x_c. It is also proportional to the driving force, which is the temperature difference between the air and the solid, which is also the difference between the dry- and wet-bulb temperatures of the air, $T - T_w$:

$$dW = dAk(T - T_w)x/x_c \qquad \textbf{8.22(1)}$$

where dW is the water transferred from a solid particle of surface area dA to the air, and k is the mass-transfer coefficient. As the air passes through the bed, its temperature falls by evaporation:

$$-dT = dW\frac{H_v}{GC_p} \qquad \textbf{8.22(2)}$$

where H_v is the latent heat of evaporation, G is the mass flow of air, and C_p is its specific heat. Combining these two equations by eliminating dW, and integrating across the total area A of solids in the dryer, produces a relationship[1] between product moisture x and air temperatures:

$$x = \frac{x_c G C_p}{H_v k A}\ln\frac{T_i - T_w}{T_o - T_w} \qquad \textbf{8.22(3)}$$

where T_i is the inlet air temperature and T_o is its outlet temperature. Given a constant airflow and surface area, product moisture in the fluid bed is proportional to the ratio of the dry-to-wet-bulb temperature difference at the ends of the dryer. This relationship will be used to control product moisture in both batch and continuous adiabatic dryers.

MOISTURE CONTROL

Moisture Analyzers

On-line moisture analyzers are becoming more common in industry, and more reliable as well (Figure 8.22h). However, they are subject to a difficult environment—hot, humid, and dusty—and the material they are analyzing tends to have nonuniform properties within a product and quite-variable properties from product to product. Variable particle size and bulk density are particularly a hindrance to optical and electrical measurements.

Optical devices can only read surface moisture, while the internal amount in a lump of solids might be quite different; and a rough or irregular surface adds a high level of noise to the measurement. Devices based on loss in weight avoid the problem of variable density, but add sample-handling difficulties and are slow.

Even when an on-line measurement is found to be accurate and reliable, it is applied to a product that has already left the dryer. Most dryers are dominated by dead time, and so the analysis reflects a condition that is several minutes—perhaps even an hour—old.

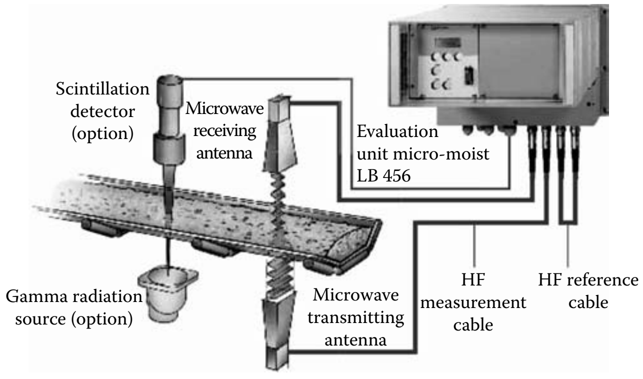


FIG. 8.22h
Noncontacting moisture analyzer can be provided with gamma radiation densitometer to compensate for variations in process density. (Courtesy of Berthold Technologies.)

By contrast, air temperatures respond readily to changing heat input and feed conditions, and if controlled, they can keep a product on specification before an analyzer can detect a significant variation. So, when an analyzer is used for control, it should set temperature in cascade and preferably be used to update an inferential model

Controlling Batch Dryers

The outlet-air temperature of a batch dryer is quite revealing of the process within. Figure 8.22i gives a typical time plot, relative to a constant inlet-air temperature at the top and a corresponding constant wet-bulb temperature at bottom. After a short preheat of the dryer and its contents, evaporation begins at a constant rate, with the outlet-air temperature settling at T_{oc}, not far above the wet-bulb temperature. After the product moisture has reached the critical value, falling-rate drying begins. This decrease in the rate of drying is reflected in a rise in outlet-air temperature, simultaneously increasing the driving force. Eventually, a temperature will be reached where the product moisture is at specification and the drying should be terminated—this is designated at T_{of} in Figure 8.22i.

To apply the fluid-bed dryer model, lump the constants of Equation 8.21(3) into a single constant K:

$$x = Kx_c \ln \frac{T_i - T_w}{T_o - T_w} \qquad \textbf{8.22(4)}$$

Wet-bulb temperature can be estimated from the measured value of T_{oc} by applying the model to the constant-rate period, setting $x = x_c$ and solving for T_w:

$$T_w = \frac{T_i - T_{\text{oc}}e^{1/K}}{1 - e^{1/K}} \qquad \textbf{8.22(5)}$$

where e is 2.718, the base of the natural logarithms. Once having estimated T_w, it can then be subsequently used in the same model to calculate T_{of}, which corresponds to the desired final moisture content x:

$$T_{\text{of}} = T_{\text{oc}} + C(T_i - T_{\text{oc}}) = CT_{\text{oc}} + (1 - C)T_i \qquad \textbf{8.22(6)}$$

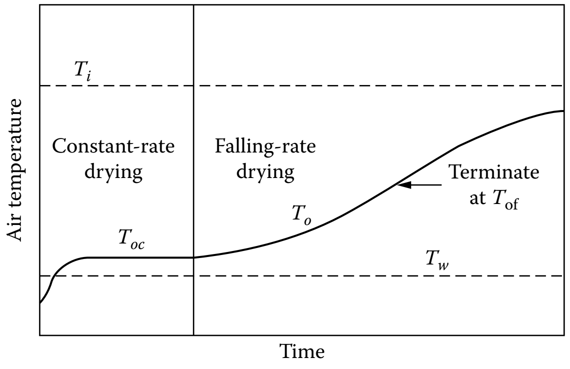
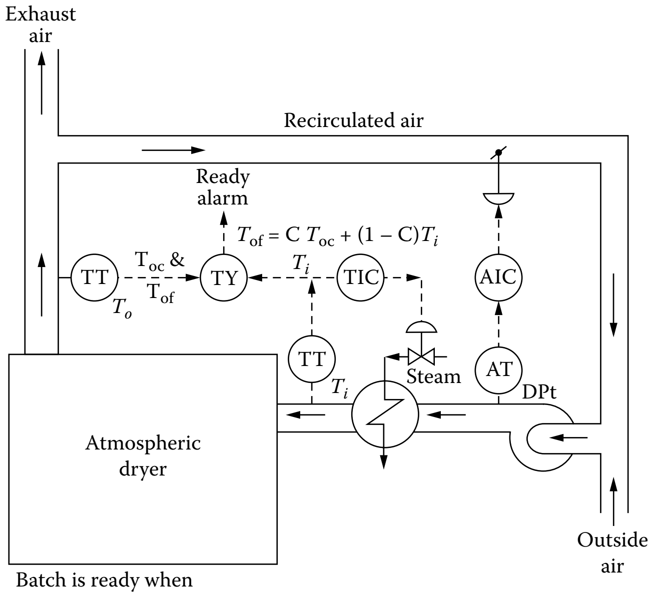


Fig. 8.22i
Product moisture should be on-specification at T_{of}.



Batch is ready when
$T_o > \text{C}\ T_{\text{oc}} + (1 - \text{C})T_i$
The smaller the constant "*C*" the drier the product will be.

FIG. 8.22j
Inferential control of a batch dryer with recirculated air.

The value of T_{oc} measured during constant-rate drying must be held in memory and used later to calculate T_{of}. When the actual measurement of T_o reaches the calculated value of T_{of}, drying should be terminated. Coefficient C is an adjustable constant, which includes the desired product moisture, batch size, and airflow. It must be determined for each product on-line. Once calibrated to a particular product, it should deliver consistent batch-to-batch moisture regardless of initial atmospheric humidity and inlet-air temperature; it is somewhat sensitive to batch size, however.

To keep inlet-air humidity from changing during the course of a run, some of the exhaust air can be recirculated under dew-point control, as shown in Figure 8.22j. At the beginning of a run, during constant-rate drying, the exhaust air will be cool and wet, and not useful for recirculation. But during falling-rate drying it becomes progressively warmer and dryer, and its recirculation under dew-point control can reduce steam consumption.

Lumber Kilns

A lumber kiln is a batch dryer where lumber is stacked in a manner that allows uniform airflow through it. Heated air circulates from one end of the kiln to the other, with the direction of flow reversed periodically for more uniform drying.

The wet-bulb temperature of the circulated air is controlled by opening an exhaust damper, which releases humid air and induces an equal flow of fresh air into the kiln. The rate of drying must be strictly limited to prevent cracking of the wood from stress caused by unequal moisture distribution.

At every reversal of flow, the temperature gradient in the kiln reverses. Temperatures at both ends of the kiln are compared in high and low selectors, with the higher selected as the inlet temperature T_i and the lower as the outlet temperature T_o. Rate of drying is limited by controlling $T_i - T_o$ with a steam valve on the air heater. T_i is also limited by a controller that is overriding the rate-of-drying controller through a low selector. These controllers are transferred to manual briefly after each flow reversal, until the temperatures settle at their new values.

Reference 2 describes the successful calculation of wood moisture using Equation 8.22(4) directly from measurements of inlet, outlet, and wet-bulb temperatures. Kx_c was adjusted to calibrate the model for each species and grade of wood, and drying was terminated when the calculated value of x reached the specification for that product. The use of the model resulted in a higher quality product, more batch-to-batch uniformity, and shorter batch times.

Continuous Fluid-Bed Dryers

The continuous fluid-bed dryer shown in Figure 8.22k uses a temperature controller on the air leaving the bed to manipulate the flow of steam to the air heater. A second controller maintains bed density by holding a constant differential pressure across it. In this dryer, rapid circulation of the solids means that the average moisture content in the bed is approximately the same as that of the product being discharged. As a consequence, the rate of drying is essentially that of the product.

The relationship between product moisture and air temperatures is then governed by Equation 8.22(4). An increase in either feed rate or moisture will lower the outlet-air temperature, causing the controller to increase steam flow to return it to set point. However, the addition of more heat to the air also raises its wet-bulb temperature, which narrows the temperature difference in the denominator of Equation 8.22(4), while raising the temperature difference in the numerator, thereby raising the level of moisture in the product. Therefore, controlling at a fixed temperature set point gives a product that is sensitive to variations in load as well as atmospheric humidity.

To provide load regulation and protection from humidity variations, the inferential model of Equation 8.22(4) can be applied. It requires a measurement of inlet-air temperature, which should be controlled in cascade from the outlet-air temperature controller. It also requires wet-bulb temperature, but this is not measurable under the conditions prevailing in the dryer—it must be calculated as a function of inlet-air temperature and humidity. The model is then inverted so that outlet-air temperature appears in the numerator, as that is the controlled variable. Outlet-air temperature is plotted as a function of inlet-air temperature for conditions of constant product moisture in Figure 8.22l, using the software in Reference 3.

Ambient temperature and dew point are entered into the program, along with fuel type in case of direct firing—this figure has no fuel input. The origin of the plot is the point where both inlet and outlet temperatures are equal to the ambient—here, 50°F.

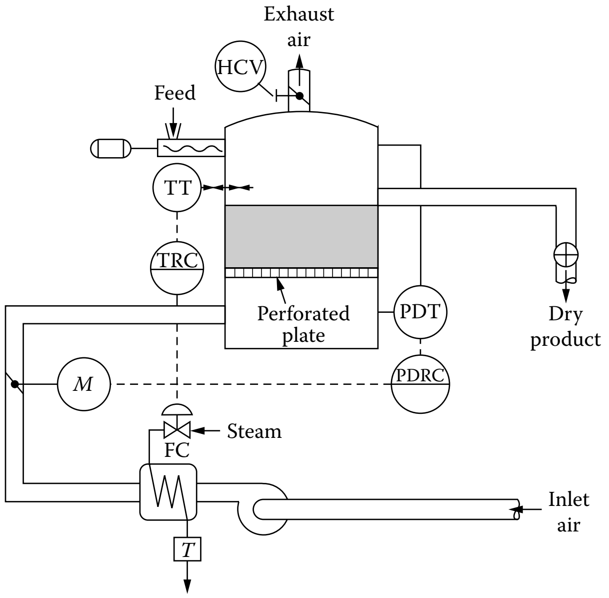


FIG. 8.22k
A continuous fluid-bed dryer with single-loop controls.

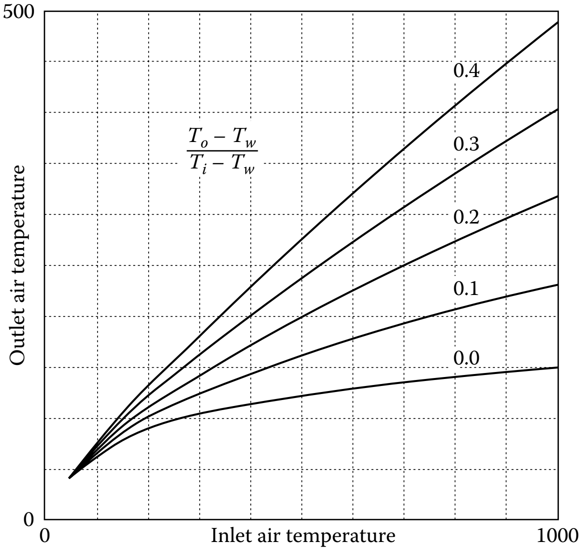


FIG. 8.22l
Each curve represents a contour of constant dryness; temperatures are in °F.

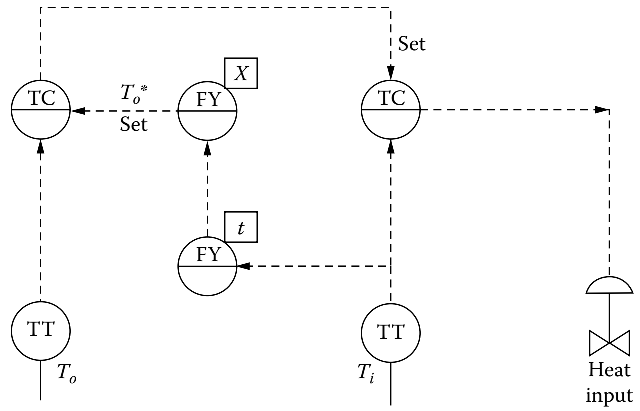


FIG. 8.22m
The inferential moisture control system has positive feedback.

The program then calculates the wet-bulb temperature for that air heated to each inlet temperature across the scale. (If the temperature is reached by combustion, then the water thus formed is added to the ambient humidity.) The wet-bulb temperature is plotted as the lowest contour, where the temperature ratio is 0. Each of the other curves is plotted by applying a different ratio, from 0.1 to 0.4, covering the conditions found in most dryers.

At a certain set of operating conditions, a fluid-bed dryer will be making on-specification product. A set of curves is plotted for that set of ambient temperature and dew point, and the inlet and outlet temperature coordinates corresponding to the desired product moisture are marked on the chart. A curve drawn through that point parallel to its neighbors then represents the locus of temperatures also producing that moisture content.

At higher temperatures and higher ratios, the curves are almost linear. Then, the model reduces to a simple matter of setting the outlet-air temperature set point T_o^* in ratio to inlet-air temperature T_i:

$$T_o^* = a + bT_i \qquad \textbf{8.22(7)}$$

where a and b are constants that give the best linear fit. However, a more accurate fit can be made by calculating wet-bulb temperature on-line and solving for T_o^* directly. A control system based on Equation 8.22(7) is shown in Figure 8.22m.

The indexing of the set point of the T_o controller forms a positive-feedback loop — a higher inlet temperature raises the outlet-temperature set point, which causes its controller to increase inlet temperature further. At the same time, the outlet-temperature measurement will also start to increase, and this is negative feedback. For a system having positive feedback to be stable, its gain must be lower and its dynamics slower than the negative feedback loop in the system.

The gains of all the curves in Figure 8.22l are 0.5 or less, compared to a negative-feedback gain of 1, so the system will not run away. But the loop will cycle if the positive feedback is not also slower. Therefore, a lag [*t*] is inserted in Figure 8.22m as a model of the dryer dynamics. This loop is very similar to the reset feedback loop in a PI or PID controller that has a positive feedback gain of 1.0 and a lag that represents the integral time. Therefore, the lag [t] in Figure 8.22m can be safely set at the same value as the integral time of the outlet-temperature controller.

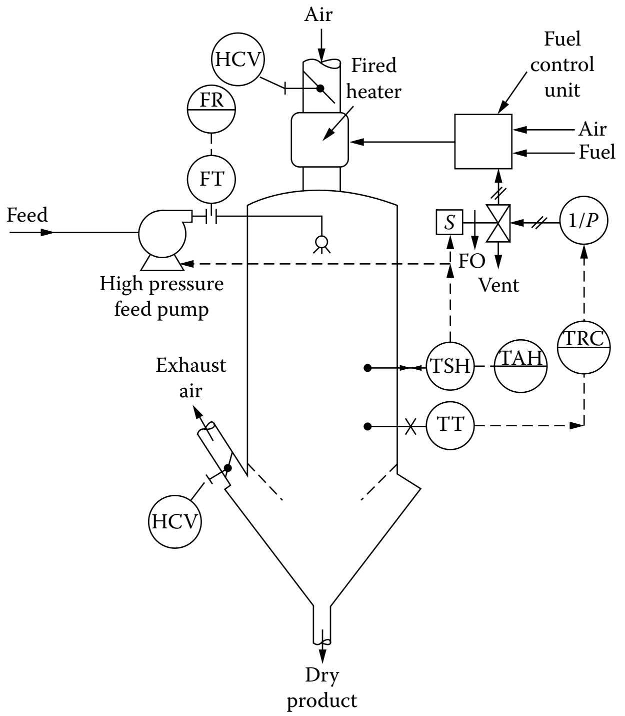


FIG. 8.22n
Spray dryer with single-loop controls.

Direct-Fired Dryers

Figure 8.22n shows a spray dryer whose air supply is heated by direct firing of a fuel. In an arrangement such as this, it is usually difficult to obtain a measurement of inlet-air temperature due to the radiation of the flame. However, given a constant airflow, air temperature should be proportional to the fuel flow. So, fuel flow is then used to index the outlet-temperature set point in the place of inlet-air temperature in Figure 8.22m. The relationship to fuel flow is not quite as accurate, but still preferable to operating at a constant set point.

The feed to a spray dryer is typically a thin slurry sprayed at high pressure through a bank of nozzles into a co-current stream of hot air. Residence time is very short, and granular product is withdrawn at the bottom of the dryer. When the model described in Figure 8.22l was originally applied to a spray dryer producing milk solids, it overcorrected—product moisture decreased following a load increase.[4] The gain factor b of about 0.23 in Equation 8.22(7) was subsequently decreased until the product moisture no longer varied with load—the final gain was 0.13.

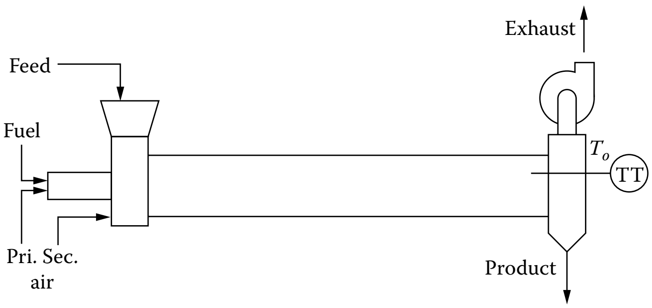


FIG. 8.22o
Ores are dried in a rotating kiln by direct firing of a fuel.

The reason for this discrepancy was discovered much later. The model used to develop the contours of Figure 8.22l was based on a constant feed moisture and variable feed rate. Within a fluid-bed dryer, the two load components produce the same effect on product moisture because of the complete back-mixing. But in a spray dryer, there is no back-mixing, so that the rate of drying varies through its length. In the particular dryer being studied, feed rate was maintained constant by pressure regulated at the nozzles, but its composition varied through a run. Wetter feed dries easier, and therefore requires less of a temperature rise to control product moisture than does an equivalent increase in feed rate.

Most common are rotary dryers, used for ores such as bauxite and gypsum, one of which is illustrated in Figure 8.22o. The center cylinder rotates slowly, conveying the solids through its length, which can be 100 ft or more. Airflow is co-current with the feed, induced by the exhaust fan. Primary air is mixed with the fuel for combustion, and a larger flow of secondary air is induced at the feed end. Inlet-air temperature cannot be measured and must be inferred from fuel flow.

There is also a problem measuring outlet-air temperature. Measurements made at an easy access point such as the suction of the exhaust fan can give temperatures below the wet-bulb—an impossibility in an adiabatic dryer. The low measurement is caused by the infiltration of ambient air through the rotary seals at the end of the dryer. To obtain an accurate measurement of outlet-air temperature requires the insertion of a resistance bulb into the rotating kiln upstream of the seal, as shown in the figure. It is best fastened under an angle-iron to protect it from falling rocks.

If a rotary dryer is exposed to feed-moisture changes, it will behave much like the spray dryer, and for the same reason. To discriminate between feed-rate and feed-moisture changes, a measurement of feed rate is required. This can be obtained by a mass-flow measurement on a conveyor feeding the dryer. On ore dryers fed by a bulldozer, feed rate may be estimated by the power load on the motor that rotates the dryer. The set point for outlet-air temperature then is calculated as

$$T_o^* = a + bF + cQ \qquad \mathbf{8.22(8)}$$

where F is feed rate and Q is the firing rate. If the feed rate is steady and the firing rate increases, the cause is an increase in feed moisture—hence coefficient c is the lower value found for the spray dryer above. If both F and Q increase, the load change is due to a rate increase—therefore, coefficient b in Equation 8.22(7) is equivalent to $b + c$ in Equation 8.22(8).

Countercurrent Dryers

Materials that have little heat sensitivity can be dried more efficiently by a countercurrent flow of air. Examples of these are sugar granulators, where the air is steam-heated, and carbon-black dryers, which are direct-fired. In these dryers, the temperature of the air leaving is unrelated to product moisture, because it is measured at the feed end. In the sugar granulator particularly, exit-air temperature reflects the temperature of the feed that is discharged hot from centrifuges. As a result, the models used for inferential control of fluid-bed and co-current adiabatic dryers do not apply when the flow pattern is countercurrent.

The best temperature to control in a countercurrent dryer is that of the product as it leaves. This can be difficult to measure, even in a sugar granulator where the product is clean, uniform, and freely flowing. Voids and dams easily form around any obstruction such as a temperature bulb. And the product temperature has much more dead time than air temperature. A reliable on-line analyzer is a real asset in controlling these dryers.

NONADIABATIC DRYERS

When heat is transferred to the solid through the walls of a dryer, enthalpy is added to the product, which raises the wet-bulb temperature of the surrounding air. An attempt to model this type of drying by estimating the internal wet-bulb temperature is difficult, because the airflow is often induced by natural draft and is neither measured nor controlled. A similar problem develops when air is recirculated within a dryer without control.

An example is a steam-tube dryer used for drying spent grain in distilleries. It is a rotating shell fitted with a bank of tubes containing condensing steam. The solids contact the tubes as the dryer rotates, and they flow at a downward slope against a natural draft of air. A rotating seal at the product end connects the steam to the tubes and draws the condensate out.

These dryers are conventionally controlled by regulating the steam pressure in the tubes. An increase in feed rate or moisture will wet more of the tube surface, thereby increasing the rate of heat transfer and condensing more steam, which the pressure regulator will then supply. However, a higher evaporative load will result in more wet-tube surface, thereby increasing the moisture content of the product.

An inferential control system was developed[5] that is similar to what was used successfully on adiabatic dryers—the

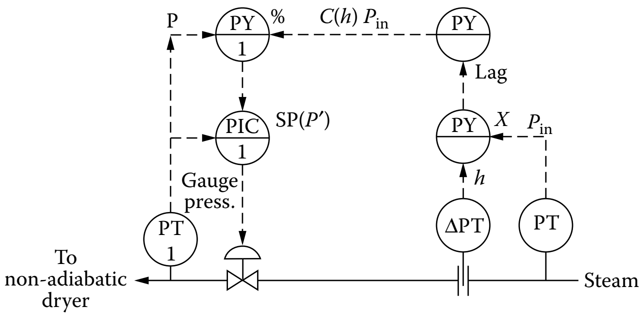


FIG. 8.22p
The ratio of the gauge pressure of the steam into the dryer (P) and the orifice differential corrected for supply pressure variations (ChP_{in}) can be adjusted to obtain the desired product dryness.

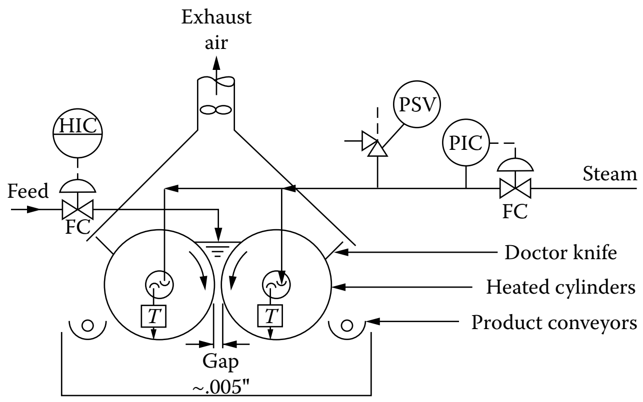


FIG. 8.22q
Double-drum dryer.

driving force for drying is increased in proportion to the evaporative load. Assuming that the moisture profile in the dryer is related to the overall heat-transfer coefficient, an attempt is made to keep that coefficient constant. Heat flow is linear with temperature difference:

$$Q = UA\Delta T \qquad \textbf{8.22(9)}$$

where U is the coefficient to be controlled, A the area of the tube surface, and ΔT the temperature difference between the solid and the steam. If steam flow increases due to a higher evaporative load, then ΔT should increase proportionately to keep U constant. This is accomplished by increasing steam pressure, as shown in Figure 8.22p.

The relationship needed between steam pressure and flow was found to be parabolic rather than linear. As a result, the differential pressure h measured across the steam orifice does not have its square root extracted before being used in the calculation. Compensation of the flow measurement for variations in supply pressure P_{in} was found to be necessary for accuracy. The pressure set point is calculated as

$$P^* = ChP_{\text{in}} \qquad \textbf{8.22(10)}$$

where C is a coefficient adjusted to make on-specification product.

Another nonadiabatic dryer that can be controlled in a similar way is the double-drum dryer shown in Figure 8.22q. Liquid is fed into the "valley" between the heated cylinders. The drums, rotating downward at the center, receive a coating of the liquid with the thickness depending upon the spacing between the rolls. The material must be dry by the time it rotates to the doctor knife, where it is cut off the roll.

The variables available for control are the speed of the cylinders, the spacing between them, the liquid level in the valley, and the steam pressure in the cylinders. The first two (speed and spacing) are usually adjusted manually. The liquid level is maintained by throttling the feed stream. Attempts to automatically control this level been thwarted by three difficulties: (1) The height of the level is only 6–9 in. (150–225 mm). (2) The liquid is constantly in a state of extreme agitation, bubbling, and boiling; and the liquid is highly concentrated and tends to plug conventional level sensors that depend on physical contact for measurement. (3) Frequently the control of feed is manual, indicated in the diagram by a manual loading station (HIC). In traditional applications, all of the other controls were on manual, and only the steam pressure was controlled automatically.

References

1. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, pp. 375–377.
2. Kinney, T. B., "Advanced Control with Optimization of Kiln-Dried Lumber with a Microprocessor-Based Unit Controller," ISA Conference, October 1985.
3. Shinskey, F. G., *Problem Solving Software,* Ver. 3.2, Foxboro, MA: The Foxboro Company, 1997, Chap. 11.
4. Myron, T. J., Shinskey, F. G., and Baker, R., "Inferential Moisture Control of a Spray Dryer," ISA Food Industry Symposium, Montreal, June 1973.
5. Myron, T. J., and Shinskey, F. G., "Product Moisture Control for Steam-Tube and Direct Fired Dryers," *MBBA Tech. Quarterly,* Vol. 12, No. 4, 1975, pp. 235–242.

Bibliography

Cook, E. M., and DuMont, H. D., "New Ideas to Improve Dryer Performance," *Chem. Eng.,* May 9, 1988.

DeThomas, F. A., "Control of Moisture in Multi-Stage Dryer," ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Fadum, O., and Shinskey, F. G., "Saving Energy through Better Control of Continuous and Batch Dryers," *Control. Eng.,* March 1980.

Hawkins, J. C., "Advances in Mineral Dryer Control Systems," 1970 ISA Conference, Reprint No. 850–70. Research Triangle Park, NC: ISA, 1970.

Krigman, A., "Moisture and Humidity," *InTech,* March 1985.

Liptak, B. G., "Optimizing Dryer Performance through Better Control," *Chem. Eng.,* February, March 1998.

McKetta, J. J., "Mass Transfer," *Unit Operations Handbook,* Vol. 1, New York: Marcel Dekker.

"Moisture/Humidity," *Measurements & Control,* February 1992.

Perry, R. H. and Green, D. W., *Chemical Engineers' Handbook,* 7th edition, New York: McGraw-Hill, 1997, Chap. 12.

Schultz, G., "Relative Humidity Measurement with an Analog Transmitter," *Measurements & Control,* December 1992.

Shinskey, F. G., "Humidity, Dew Point, and Wet-Bulb Temperature," *Instruments and Control Systems,* January, February 1975.

Shinskey, F. G., "How to Control Product Dryness," *Instrumentation Technology,* October 1968.

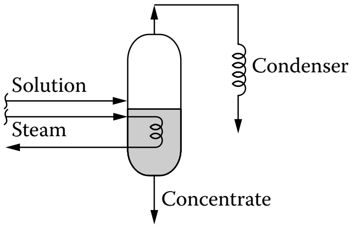
Shinskey, F. G., "Process Control Systems with Variable Structure," *Control Engineering,* August 1974.

Shinskey, F. G., *Energy Conservation through Control*, New York: Academic Press, 1978, Chap. 7.

Young, G. C., "Simple Method Eases Dryer Tests," *Chem. Eng.,* June 20, 1988.

8.23 Evaporator Controls

T. J. MYRON (1972, 1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

INTRODUCTION

Evaporation is used in a number of unit operations. For example, one of the important steps in the operation of a chiller is the evaporation of the refrigerant, which usually occurs at low pressures. The controls of evaporators serving such heat pumps are described in Sections 8.12 and 8.13. If evaporation occurs at higher pressures, we refer to that unit process as vaporization or reboiling and the controls of those processes are described in Sections 8.19, 8.20, 8.21, and 8.27. The evaporation controls that are discussed in this section are for water-based near-atmospheric processes.

Evaporation is one of the oldest unit operations, dating back to the Middle Ages, when any available energy source was used to concentrate thin brine solutions in open tanks. As archaic as this technique is, it is still widely used during the early spring in northern New England when maple sugar sap is tapped and concentrated in open-fired pans.

Solar evaporation and bubbling hot gases through a solution are other examples of concentrating solutions. For our purposes, evaporation will be limited to concentrating aqueous solutions in a closed vessel or group of vessels in which the concentrated solution is the desired product and indirect heating (usually steam) is the energy source. Occasionally, the water vapor generated in the evaporator is the product of interest, such as in desalinization or in the production of boiler feed water. In other cases neither vapor nor concentrated discharge has any market value, as in nuclear wastes.

Evaporators can be arranged in forward-feed, reverse-feed, or parallel-feed configurations, with each stage being heated by the vapors of the previous stage. Evaporation is an energy-intensive process. Its efficiency can be improved by increasing the number of evaporators in series (effects), and some of the energy can be recovered by vapor recompression.

The product concentration can be measured by a variety of analyzers, including density, conductivity, refractive index, percent solids, turbidity, and boiling or freezing point analyzers, which are all discussed in Chapter 8 of the first volume of this handbook, *Process Measurement and Analysis*.

Evaporator Terminology

SINGLE-EFFECT EVAPORATION Single-effect evaporation occurs when a dilute solution is contacted only once with a heat source to produce a concentrated solution and an essentially pure water vapor discharge. The operation is shown schematically in Figure 8.23a.

MULTIPLE-EFFECT EVAPORATION Multiple-effect evaporations use the vapor generated in one effect as the energy source to an adjacent effect (Figure 8.23b). Double- and triple-effect evaporators are the most common; however, six-effect evaporation can be found in the paper industry, where kraft liquor is concentrated, and as many as 20 effects can be found in desalinization plants.

BOILING-POINT RISE This term expresses the difference (usually in °F) between the boiling point of a constant composition solution and the boiling point of pure water at the same pressure. For example, pure water boils at 212°F (100°C) at 1 atmosphere, and a 35% sodium hydroxide solution boils at about 250°F (121°C) at 1 atmosphere. The boiling-point

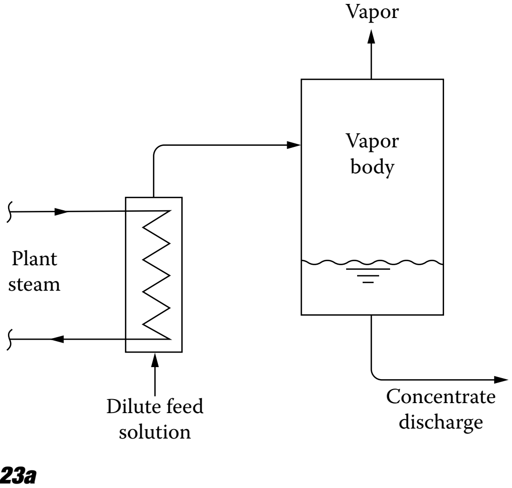


FIG. 8.23a
Single-effect evaporator.

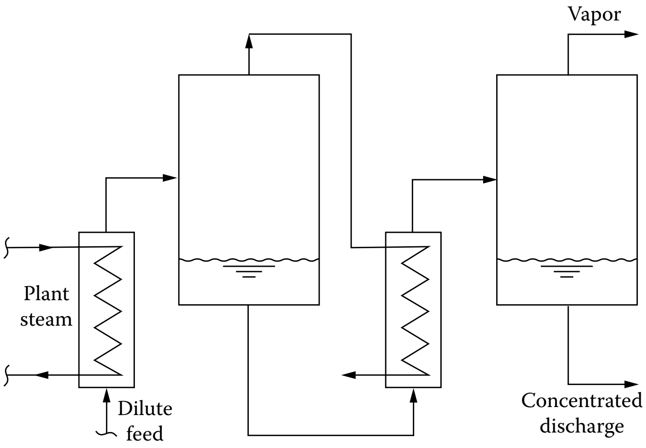


FIG. 8.23b
Multiple-effect evaporator.

rise is therefore 38°F (21°C). Figure 8.23c illustrates the features of a Dühring plot in which the boiling point of a given composition solution is plotted as a function of the boiling point of pure water.

ECONOMY This term is a measure of steam use and is expressed in pounds of vapor produced per pound of steam supplied to the evaporator train. For a well-designed evaporator system the economy will be about 10% less than the number of effects; thus, for a triple-effect evaporator the economy will be roughly 2.7.

CAPACITY The capacity for an evaporator is measured in terms of its evaporating capability, viz., pounds of vapor produced per unit time. The steam requirements for an evaporating train may be determined by dividing the capacity by the economy.

CO-CURRENT OPERATION The feed and steam follow parallel paths through the evaporator train.

COUNTERCURRENT OPERATION The feed and steam enter the evaporator train at opposite ends.

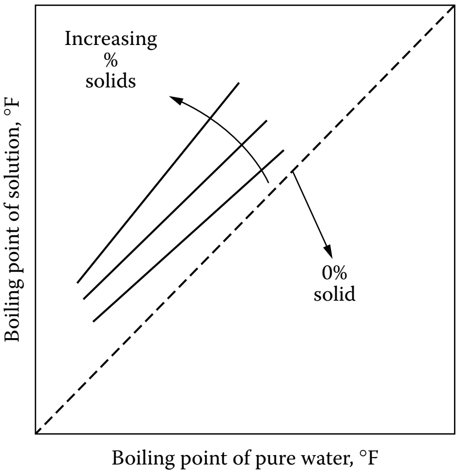


FIG. 8.23c
Dühring plot of boiling point rise.

EVAPORATOR MODELING

The Steady-State Model

Development of the steady-state model for an evaporator involves material and energy balances. A relationship between the feed density and percentage of solids is also required and is specific for a given process, whereas the material and energy balances are applicable to all evaporator processes. Figure 8.23d illustrates the double-effect evaporator from the standpoint of a material balance.

where

W_o = feed rate in lb (kg) per unit time
V_o = feed flow in gpm or lpm
V_1 = vapor flow from Effect I in lb (kg) per unit time
X_o = weight fraction solids in feed
W_1 = liquid flow rate leaving Effect I in lb (kg) per unit time
X_1 = weight fraction solids in W_1
V_2 = vapor flow from Effect II in lb (kg) per unit time
W_p = concentrated liquid product flow in lb (kg) per unit time
X_p = weight fraction of solids in product (the controlled variable)

Overall balance in Effect I:

$$W_o = V_1 + W_1 \quad \mathbf{8.23(1)}$$

Overall balance in Effect II:

$$W_1 = V_2 + W_p \quad \mathbf{8.23(2)}$$

Solid balance in Effect I:

$$W_oX_o = W_1X_1 \quad \mathbf{8.23(3)}$$

Solid balance in Effect II:

$$W_1X_1 = W_pX_p \quad \mathbf{8.23(4)}$$

Substituting Equation 8.23(2) in Equation 8.23(1):

$$W_o = V_1 + V_2 + W_p \quad \mathbf{8.23(5)}$$

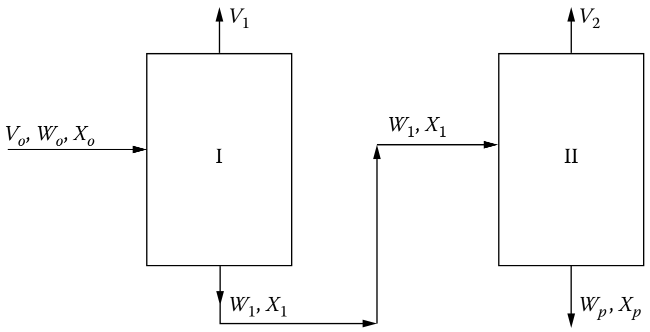


FIG. 8.23d
The material balance of a double-effect evaporator.

Combining Equations 8.23(3) and 8.23(4):

$$W_o X_o = W_p X_p \qquad \textbf{8.23(6)}$$

Solving Equation 8.23(6) for W_p and substituting in Equation 8.23(5) gives

$$W_o = V_1 + V_2 + \frac{W_o X_o}{X_p} \qquad \textbf{8.23(7)}$$

The W_o term of Equation 8.23(7) can be written in terms of volumetric flow (gallons per unit time) times density or specific gravity as follows:

$$W_o = V_o D_o = V_o D_w S_o \qquad \textbf{8.23(8)}$$

where

V_o = volumetric feed rate in gallons per unit time
D_o = fed density in lb per gallon
D_w = nominal density of water, 8.33 lb per gallon (1 kg/l)
S_o = specific gravity of feed

Substituting for W_o from Equation 8.23(8) in Equation 8.23(7) and combining terms,

$$V_o D_w S_o \left(1 - \frac{X_o}{X_p}\right) = V_1 + V_2 = V_t \qquad \textbf{8.23(9)}$$

where V_t = total vapor flow in lb per unit time.

The total vapor flow (V_t) is proportional to the energy supplied to the train (plant steam) and the proportionality constant is the economy (E) of the system, i.e.,

$$V_t = W_s E \qquad \textbf{8.23(10)}$$

where

W_s = steam flow in lbs steam per unit time
E = economy in lbs vapor per lb steam

Substituting for V_t in Equation 8.23(9):

$$V_o D_w S_o \left(1 - \frac{X_o}{X_p}\right) = W_s E \qquad \textbf{8.23(11)}$$

Equation 8.23(11) is the steady-state model of the process and includes all of the load variables (V_o and X_o), the manipulated variable (W_s), and the controlled variable (X_p). At this point the $W_s E$ portion of Equation 8.23(11) may be modified to include heat losses from the system and to include the fact that the feed may be subcooled. These are two forms of heat losses, because in either case a portion of the steam supplied is for purposes other than producing vapor. Typical values of effective heat losses vary from 3 to 5%. If, for example, a 5% heat loss is assumed, Equation 8.23(11) becomes

$$V_o D_w S_o \left(1 - \frac{X_o}{X_p}\right) = 0.95 W_s E \qquad \textbf{8.23(12)}$$

TABLE 8.23e
Weight Fraction and Specific Gravity Relationship

Solids Weight Fraction (X_o)	*Specific Gravity (S_o)*
0.08	1.0297
0.16	1.0633
0.24	1.0982

For an in-depth discussion relating to methods of computing the economy (E) of a particular evaporator system, see References 2 and 3.

The $S_o(1 - X_o/X_p)$ portion of Equation 8.23(11) is a function of the feed density, $f(D_o)$, i.e.,

$$f(D_o) = S_o \left(1 - \frac{X_o}{X_p}\right) \qquad \textbf{8.23(13)}$$

For each feed material a relationship between the density of the feed material and its solids weight fraction has usually been empirically determined by the plant or is available in the literature. See Reference 4 where the density equals percentage of solids relationship of 70 inorganic compounds is available.

Assume, for example, that a feed material (to be concentrated) has the solids–specific-gravity relationship shown in Table 8.23e.

If this feed material were to be concentrated so as to produce a product having a weight fraction of 50% ($X_p = 0.50$), the $f(D_o)$ relationship of Equation 8.23(13) could be generated as shown in Table 8.23f.

This body of data is plotted in Figure 8.23g. In all the cases investigated, the $f(D_o) = S_o$ relationship is a straight line having an intercept of 1.0, 1.0.

The $f(D_o)$ relationship can then be written in terms of the equation of a straight line: $y = mx + b$, i.e.,

$$f(D_o) = 1.0 + m(S_o - 1.0) \qquad \textbf{8.23(14)}$$

where m = slope of line.

TABLE 8.23f
Density to Weight Fraction Relationship

Solids Weight Fraction (X_o)	*Specific Gravity (S_o)*	$\left(1-\frac{X_o}{X_p}\right)$	$S_o\left(1-\frac{X_o}{X_p}\right) = \int(D_0)$
0	1.0000	1.0	1.0000
0.08	1.0297	0.840	0.865
0.16	1.0633	0.680	0.723
0.24	1.0982	0.520	0.571

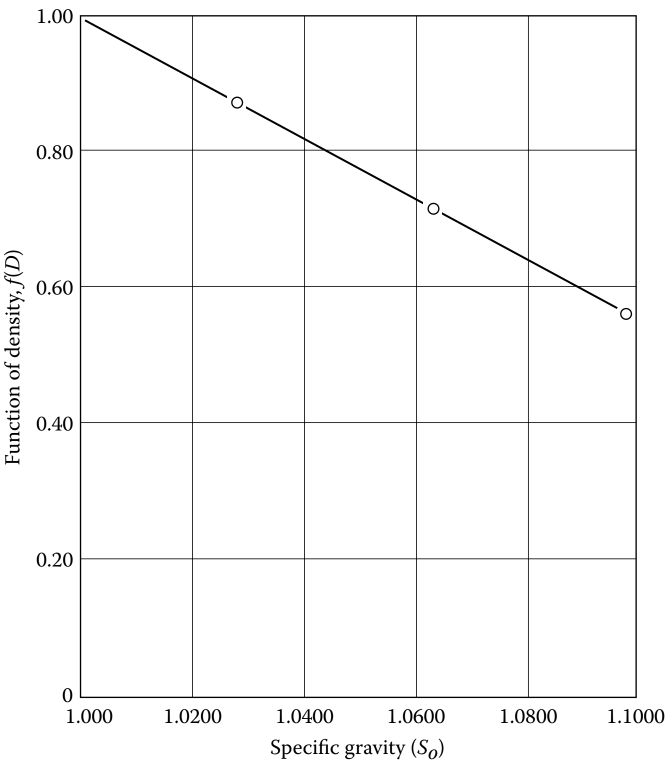


FIG. 8.23g
The straight-line relationship between density and specific gravity.

Using the data of Table 8.23f the value of m is determined as

$$m = \frac{1.000 - 0.571}{1.0000 - 1.0982} = -4.37 \qquad \textbf{8.23(15)}$$

therefore,

$$f(D_o) = 1.0 - 4.37(S_o - 1.0) \qquad \textbf{8.23(16)}$$

Substituting Equation 8.23(16) into Equation 8.23(11) and solving for the manipulated variable, steam flow, W_s:

$$W_s = \frac{V_o D_w f(D)}{E} \qquad \textbf{8.23(17)}$$

Therefore, W_s can be expressed as

$$W_s = \frac{V_o[1 - m(D-1)}{E} \qquad \textbf{8.23(18)}$$

If feed rate is manipulated in response to a variable steam flow and feed density, Equation 8.23(18) becomes

$$V_o = \frac{W_s E}{1 - m(D-1)} \qquad \textbf{8.23(19)}$$

Scaling and Normalizing

With the equation for the steady-state model defined, it can now be scaled and the digital or analog instrumentation specified. Scaling of computing instruments or software is necessary to ensure compatibility with input and output signals and is accomplished most effectively by normalizing, i.e., assigning values from 0 to 1.0 to all inputs and outputs. The procedure involves (1) writing the engineering equation to be solved, (2) writing a normalized equation for each variable in the engineering equation, and (3) substituting the normalized equivalent of each term in (2) into (1).

The first step of the procedure has already been done, because Equation 8.23(17) has already been written. To illustrate, let $V_o = 0$ to 600 gph (0 to 2.3 m^3/h), $W_s = 0$ to 2500 lb/hr (0 to 1125 kg/h), S_o = 1.000 to 1.1000, E = 1.8 lb vapor per lb of steam (1.8 kg vapor per kg of steam), and D_w = 8.33 lb/gallon (1 kg/l). The scaled equations for each input are

$$V_o = 600\,V_o' \qquad \textbf{8.23(20)}$$

$$W_s = 2500\,W_s' \qquad \textbf{8.23(21)}$$

$$S_o = 1.0000 + 0.10000 S_o' \qquad \textbf{8.23(22)}$$

where

V_o' = volumetric flow transmitter output, 0–1.0 or 0–100%
W_s'= steam flow transmitter output, 0–1.0 or 0–100%
S_o' = specific gravity transmitter output, 0–1.0 or 0–100%

The values of D_w and E need not be scaled, because they are constants. Because $f(D_o)$ is already on a 0 to 1.0 basis, the $f(D_o)$ term for the sake of completeness can be written

$$f(D_o) = 1.0 f(D_o)' \qquad \textbf{8.23(23)}$$

Operating on the $f(D_o)$ equation first, Equations 8.23(23) and 8.23(22) are substituted into Equation 8.23(16):

$$(D_o)' = 1.0 - 4.37 \times (1.0000 + 0.1000S_o - 1.0) \qquad \textbf{8.23(24)}$$

$$(D_o)' = 1.0 - 4.37S_o' \qquad \textbf{8.23(25)}$$

Substituting Equations 8.23(20), 8.23(21), and 8.23(25) into Equation 8.23(17) as well as the values of E and D_w,

$$2500W_s' = \frac{600(V_o')8.33(1.0 - 0.437S_o')}{1.8} \qquad \textbf{8.23(26)}$$

$$\begin{aligned} W_s' &= 1.11V_o'(1.0 - 0.437S_o') \\ &= 1.11V_o' f(D_o)' \end{aligned} \qquad \textbf{8.23(27)}$$

or if feed flow is the manipulated variable:

$$V_o' = \frac{W_s'}{1.11(1 - 0.437S_o')} \qquad \textbf{8.23(28)}$$

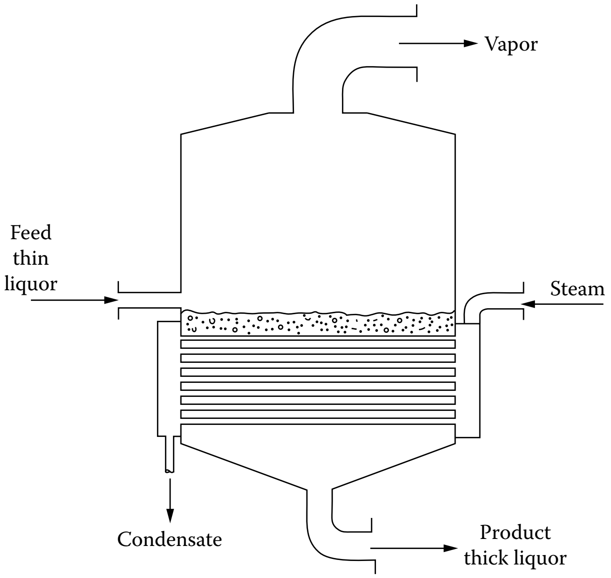


FIG. 8.23h
Horizontal-tube evaporator.

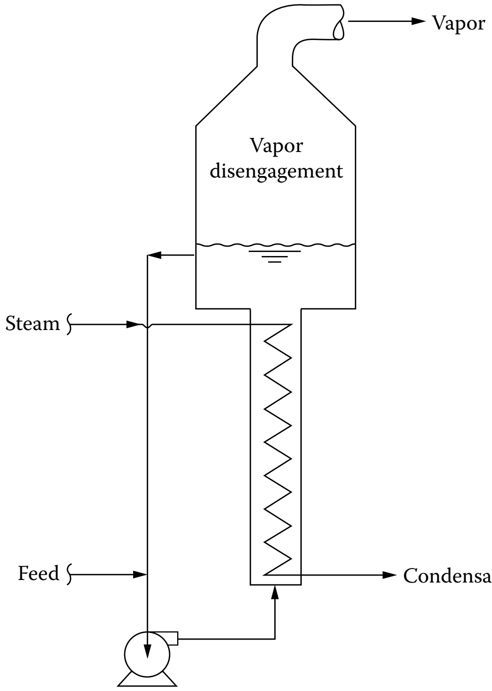


FIG. 8.23i
Forced-circulation evaporator.

EVAPORATOR DESIGNS

Six types of evaporators are used for most applications, and for the most part the length and orientation of the heating surfaces determine the name of the evaporator.

Horizontal-tube evaporators (Figure 8.23h) were among the earliest types. Today, they are limited to preparation of boiler feed water and, in special construction (at high cost), for small-volume evaporation of severely scaling liquids, such as hard water. In their standard form they are not suited to scaling or salting liquids and are best used in applications requiring low throughputs.

Forced-circulation evaporators (Figure 8.23i) are the most popular and have the widest applicability. Circulation of the liquor past the heating surfaces is assured by a pump, and consequently these evaporators are frequently external to the flash chamber so that actual boiling does not occur in the tubes, thus preventing salting and erosion. The external tube bundle also lends itself to easier cleaning and repair than the integral heater shown in the figure. Disadvantages include high cost, high residence time, and high operating costs due to the power requirements of the pump.

A short-tube vertical evaporator (Figure 8.23j) is common in the sugar industry for concentrating cane sugar juice. Liquor circulation through the heating element (tube bundle) is by natural circulation (thermal convection).

Because in the short-tube vertical evaporator the mother liquor flows through the tubes, they are much easier to clean than those shown in Figure 8.23h, in which the liquor is outside the tubes. Thus, this evaporator is suitable for mildly scaling applications in which low cost is important and cleaning or descaling must be conveniently handled. Level control is important — if the level drops below the tube ends, excessive scaling results. Ordinarily, the feed rate is controlled by evaporator level to keep the tubes full. The disadvantage of high residence time in the evaporator is compensated for by the low cost of the unit for a given evaporator load.

A long-tube vertical evaporator, or rising film concentrator (RFC), shown in Figure 8.23k, is in common use today,

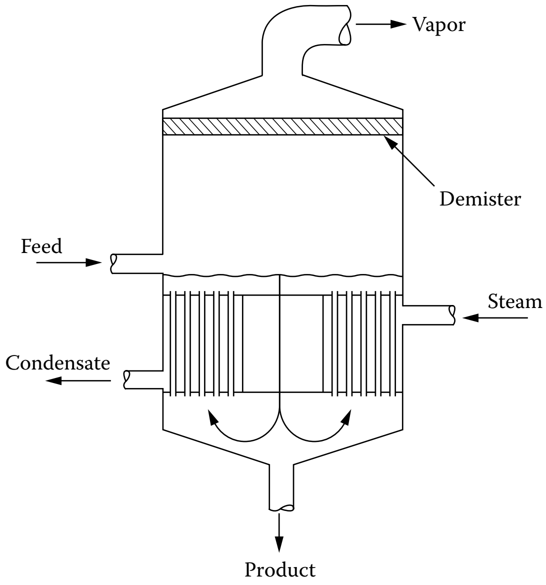


FIG. 8.23j
Short-tube vertical evaporator.

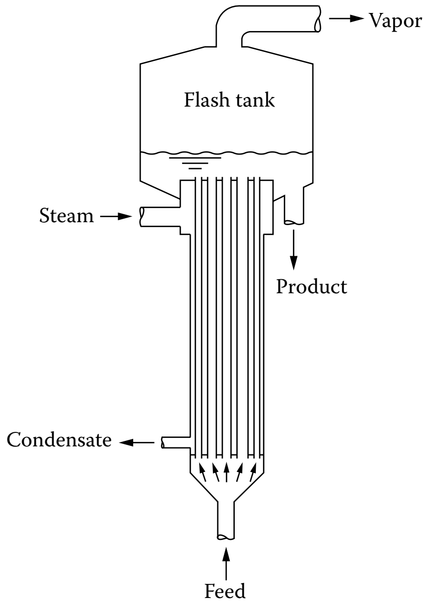


FIG. 8.23k
Long-tube vertical evaporator.

because its cost per unit of capacity is low. Typical applications include concentrating black liquor in the pulp and paper industry and corn syrup in the food industry.

Most of these evaporators are of the single-pass variety, with little or no internal recirculation. Thus, residence time is minimized. Level control is important in maintaining the liquid seal in the flash tank. The units are sensitive to changes in operating conditions, which is why many of them are difficult to control. They offer low cost per pound of water evaporated and have low holdup times but tend to be tall (20 to 50 ft, or 6 to 15 m), requiring more head room than other types.

A falling-film evaporator (Figure 8.23l) is commonly used with heat-sensitive materials. Physically, the evaporator looks like a long-tube vertical evaporator, except that the feed material descends by gravity along the inside of the heated tubes, which have large inside diameters (2 to 10 in., or 50 to 250 mm).

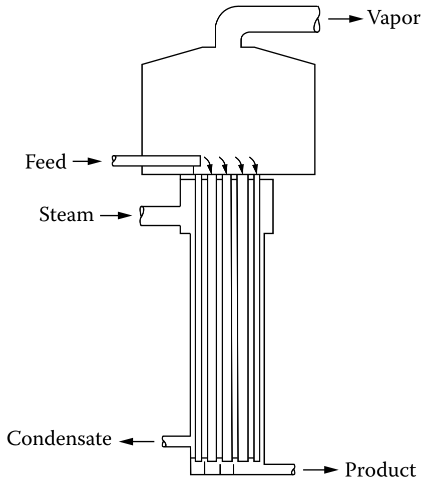


FIG. 8.23l
Falling-film vertical evaporator.

An agitated-film evaporator, like the falling-film evaporator, is commonly used for heat-sensitive and highly viscous materials. It consists of a single large-diameter tube with the material to be concentrated falling in a film down the inside, where a mechanical wiper spreads the film over the inside surface of the tube. Thus, a large heat-transfer coefficient can be obtained, particularly with highly viscous materials.

EVAPORATOR CONTROLS

In the following paragraphs the load variable will be assumed to be the flow rate and concentration of the feed stream. A later paragraph, Other Control Loops, will discuss the control of other variables, such as steam enthalpy, material balance, and absolute pressure controls.

The control systems to be considered in achieving final product concentration include (1) feedback, (2) cascade, and (3) feedforward, (4) auto-select, and (5) advanced controls. For ease of illustration, a double-effect, co-current flow evaporator will be used. Extension to more or fewer effects will not change the basic control system configuration.

The choice of the control system should be based on the needs and characteristics of the process. Evaporators as a process class tend to be capacious (mass and energy storage capability) and have significant dead time (30 sec or greater). If the major process loads (feed rate and feed density) are reasonably constant and the only corrections required are for variations in heat losses or tube fouling, feedback control will suffice.

If steam flow varies because of demands elsewhere in the plant, a cascade configuration will probably be the proper choice. If, however, the major load variables change rapidly and frequently, it is strongly suggested that feedforward in conjunction with feedback be considered.

Selective control can be superimposed on either control configuration, which will stay inactive until a limit is reached (such as running out of steam), at which point control is transferred to keep the operation (the demand for steam) under that limit. Advanced controls usually combine all the previous features and add to it a multivariable model-based predictive capability.

Feedback Control

A typical feedback control system (Figure 8.23m) consists of measuring the product concentration with a density sensor and controlling the amount of steam to the first effect by a three-mode controller. The internal material balance is maintained by level control on each effect. (A brief description of the various methods of measuring product density will be found at the end of this section; additional discussion regarding

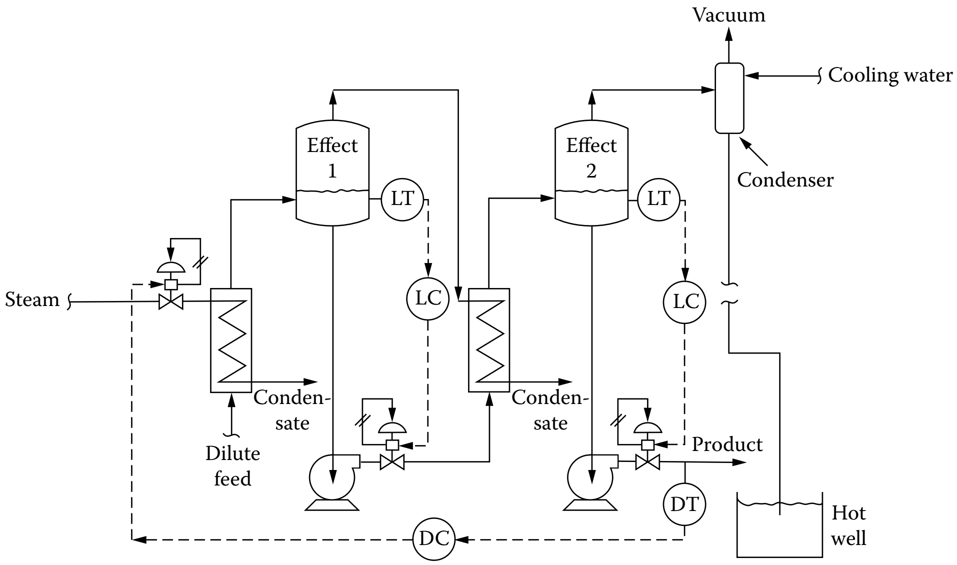


FIG. 8.23m
In a simple feedback control system the steam flow is directly throttled to keep the density of the product constant.

density measurement can be found in Chapter 6 of Volume 1 of this handbook, *Process Measurement and Analysis*.)

Ritter and associates[1] have modeled the control system configuration shown in Figure 8.23m and have found it to be very stable. They investigated other combinations of controlled and manipulated variables, which were less effective.

Cascade Control A typical cascade control system is illustrated in Figure 8.23n. This control system, like the feedback loop in Figure 8.23m, measures the product density and adjusts the heat input. The adjustment in this instance, however, is through a flow loop that is being set in cascade from the final density controller, an arrangement that is particularly effective when steam flow variations (outside of the evaporator) are frequent. It should be noted that with this arrangement the valve positioner is not required and can actually degrade the performance of the flow control loop. (For more information on cascade control see Section 2.6 in Chapter 2.)

Selective Control The selective control scheme shown in Figure 8.23o uses the previously described cascade configuration and adds to it the feature of protection against running

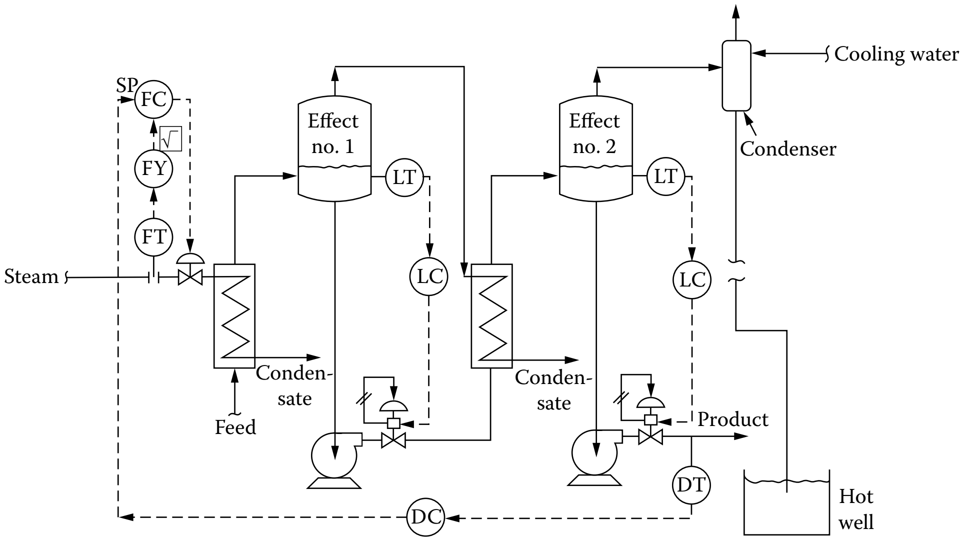


FIG. 8.23n
In the cascade version of a feedback control system the steam flow is controlled and the flow controller set point is adjusted to keep the density of the product constant.

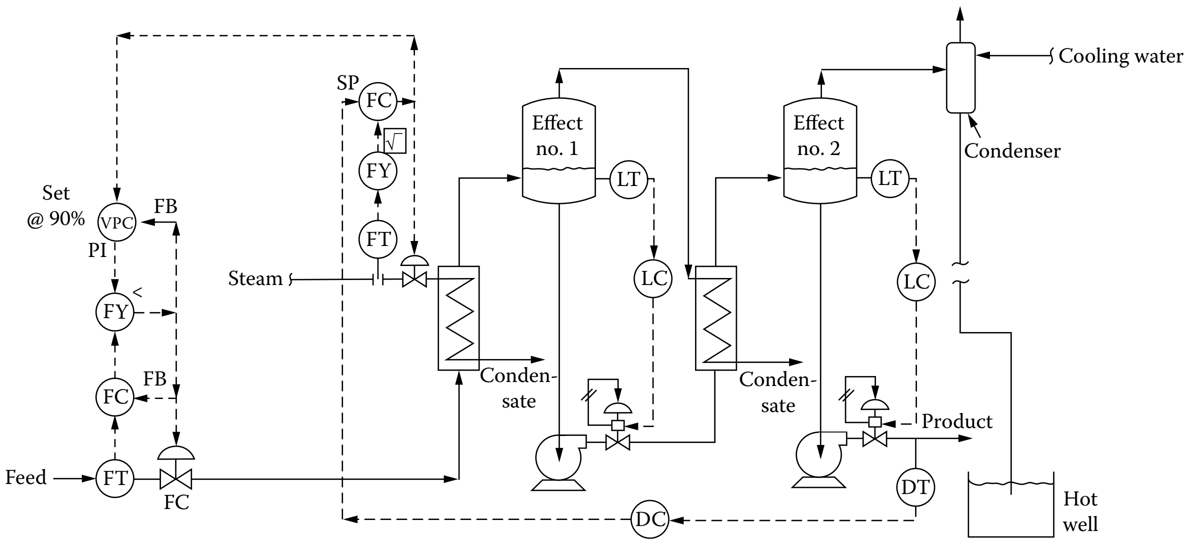


FIG. 8.23o
Valve position controller (VPC) cuts back the feed flow and thereby protects the system from running out of steam. The selective control scheme is protected from reset windup by external feedbacks.

out of steam. This feature is provided by the valve position controller (VPC), whose measurement is the steam valve opening and whose set point is about 90%. Therefore, as the feed gets too diluted and therefore requires more steam to concentrate, the steam valve opens. When it reaches 90% lift (which on an equal-percentage valve corresponds to 70% flow), the VPC output signal drops below that of the feed flow controller (FC), and therefore the low signal selector (FY) transfers the control of the feed valve to the VPC. During the period while the feed is diluted due to some upset, the VPC sets the feed flow rate to a value that corresponds to the allowable maximum opening of the steam valve.

The VPC is a PI controller with the integral mode dominating (as is the case in most valve position controllers). This allows it to operate smoothly even if the measurement signal (the steam valve opening) is noisy, as its output signal responds not so much to the error, but to the total area under the past error curve. Both controllers (FC and VPC) are provided with external feedback taken from the output of the low selector FY. This way the controller that is in control will have its output signal and its feedback signal at the same values and therefore will operate as a normal PI controller.

The controller that is not selected will operate as a proportional-only controller with a bias, where the bias is the external feedback signal. This guarantees that at the time of switchover the two outputs will be identical, and therefore the switchover will occur bumplessly (at the time of switchover the error in the idle controller has just reached zero, and therefore its output equals its external feedback signal).

Feedforward Control

In most evaporator applications the control of product density is constantly affected by variations in feed rate and feed density to the evaporator. In order to counter these load variations, the manipulated variable (steam flow) must attain a new operating level. In the pure feedback or cascade arrangements this new level was achieved by trial and error as performed by the feedback (final density) controller.

A control system able to react to these load variations when they occur (feed rate and feed density) rather than wait for them to pass through the process before initiating a corrective action would be ideal. This technique is termed *feedforward control*. (For an in-depth discussion of feedforward control, refer to Section 2.9 in Chapter 2.)

Figure 8.23p illustrates in block diagram form the features of a feedforward system. Feedforward controls utilize

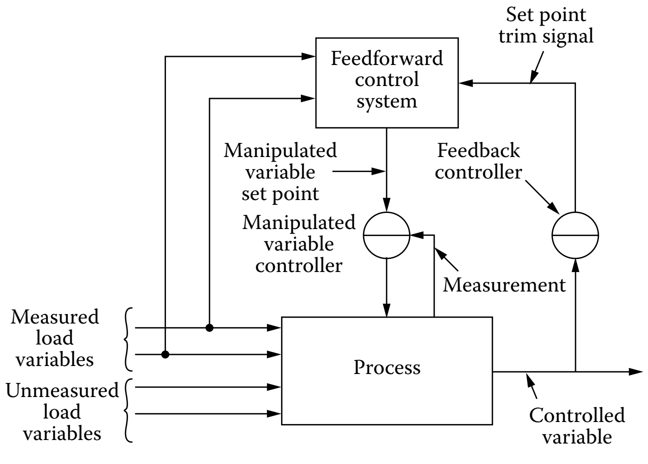


FIG. 8.23p
Feedforward system.

a simplified model of the controlled process. There are two types or classes of load variations — measured and unmeasured. The measured load signals are inputs to the feedforward control system, where they serve to compute the set point of the manipulated variable control loop as a function of the measured load variables.

The unmeasured load variables pass through the process, undetected by the feedforward system (disregarded in this simplified model), and cause an upset in the controlled variable. The output of the feedback loop then trims the calculated value of the set point to the correct operating level. In the limit, feedforward control would be capable of perfect control if all load variables could be defined, measured, and incorporated in the forward set point computation. This is what multivariable model predictive control provides, as will be discussed later.

At a practical level, then, the load variables are classified as either major or minor, and the effort is directed at developing a feedforward model that incorporates the major load variables, the manipulated variables, and the controlled variable. Such a relationship is termed the steady-state model of the process. Minor load variables are usually very slow to materialize and are hard to measure.

In terms of evaporators, minor load variations might be heat losses and tube fouling. Load variables such as these are usually easily handled by a feedback loop. The purpose of the feedback loop is to trim the forward calculation to compensate for the minor or unmeasured load variations. Without this feature the controlled variable would go off set point.

In addition to the two ingredients of a feedforward control system that have already been discussed, the steady-state model and feedback trim, there is a third ingredient. This third ingredient of a successful feedforward system application is dynamic compensation.

Dynamic compensation is required when a change in one of the major process loads also requires a change in the operating level of the manipulated variable. If the load and the manipulated variables enter the process at different locations, there usually will be an imbalance or inequality between the effects of the load and the manipulated variables on the controlled variable; i.e.,

$$\frac{\Delta \text{ controlled variable}}{\Delta \text{ load variable}} \neq \frac{\Delta \text{ controlled variable}}{\Delta \text{ manipulated variable}} \qquad \textbf{8.23(29)}$$

This imbalance manifests itself as a *transient* excursion of the controlled variable from set point. If the forward calculation is accurate, the controlled variable returns to set point once the new steady-state operating level is reached.

In terms of a co-current flow evaporator, an increase in feed rate will call for an increase in steam flow. Assuming that the level controls on each effect are properly tuned, the increased feed rate will rapidly appear at the end of the train, while the effect of the increased steam flow will take longer, because it has to overcome the thermal inertia of the process. Therefore, an increase in feed flow results in a transient decrease of the controlled variable (density), because the load variable passes through the process faster than the manipulated variable. This behavior is shown in Figure 8.23q.

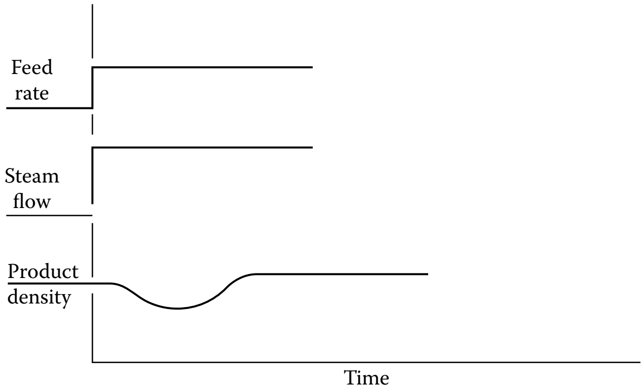


FIG. 8.23q
Load variable faster than manipulated variable.

The same sequence is seen in Figure 8.23r, except that this figure illustrates the case when the manipulated variable passes through the process faster than the load variable. Such behavior may occur in a countercurrent evaporator operation. This dynamic imbalance is normally corrected by inserting a dynamic element (lag, lead-lag, or a combination thereof) in at least one of the load measurements to the feedforward control system. Usually, dynamic compensation of that major load variable that can change in the severest manner (a step change) is all that is required. For evaporators this is usually the feed flow rate to the evaporator.

Feed density changes, although frequent, are usually more gradual, and the inclusion of a dynamic element for this variable is usually not warranted. In summary, the three ingredients of a well-designed feedforward system are (1) an accurate steady-state model, (2) properly set dynamic compensation, and (3) a correctly designed feedback trim, which corrects for the inaccuracy of the steady-state model.

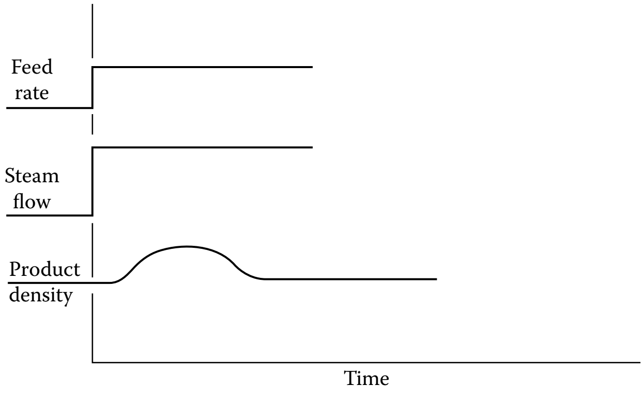


FIG. 8.23r
Manipulated variable faster than load variable.

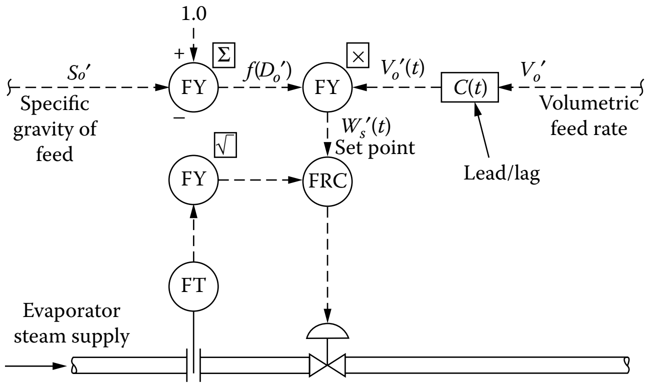


FIG. 8.23s
Implementation of the feedforward control loop for the case when the manipulated variable is steam flow, which is described in Equation 8.23(27), with first-order lag added for dynamic compensation.

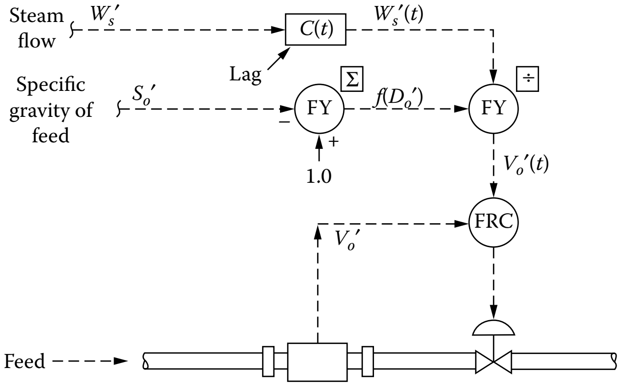


FIG. 8.23t
Implementation of the feedforward control loop for the case when the manipulated variable is feed flow, which is described in Equation 8.23(28), with first-order lag added for dynamic compensation.

Dynamic Compensation The dynamics of the co-current evaporator, in which steam is the manipulated variable, requires that a lead-lag dynamic element be incorporated in the system to compensate for the dynamic imbalance between feed rate and steam flow.

In the example, it was arbitrarily assumed that steam flow is the manipulated variable resulting in Equation 8.23(17). In some applications evaporators are run on waste steam, in which case the feed rate is proportionally adjusted to the available steam (Figure 8.23o), which makes feed the manipulated variable and steam the load variable.

Solving Equation 8.23(17) for feed rate,

$$V_o' = \frac{W_s E}{D_w(D)} \qquad \textbf{8.23(30)}$$

In this arrangement the dynamics do not change, but the manipulated variable advances through the process faster than the load variable, which requires a dynamic element having first-order lag characteristics. The instrument arrangement for the case where the manipulated variable is the steam supply was shown in Figure 8.23s.

The configuration where the manipulated variable is the feed flow is shown in Figure 8.23t and is based on Equation 8.23(28).

Feedback Trim of the Feedforward Loop As a general rule, feedback trim is incorporated into the control system *at the point at which the set point of the controlled variable appears*. For the evaporator the set point is the slope of the $f(D_o)$ relationship (Figure 8.23g). If the weight fraction of solids in the product (X_p) changes, the slope of the line changes too.

To this point the slope of the line (value of *m*) was assumed to be a constant (0.437), which value is incorporated into the summing relay or amplifier in Figures 8.23s and 8.23t. The instrumentation was scaled to make one grade of product, e.g., 50% solids. If a more or less concentrated product were desired, the gain term would have to be changed manually. In order to increase the flexibility of the control system, in Figure 8.23u a multiplier and a final product density control loop are added.

The controller output is now variable not only to permit changing the concentration of the product (slope adjust) but also to adjust the steam flow set point to compensate for the minor load variations, which up to this point were not considered.

The feedback trim for the case when steam is the manipulated variable is shown in Figure 8.23u. For the configuration where the manipulated variable is the feed flow, the product density controller will trim the value of the slope m, but in most other respects the feedforward-feedback configuration

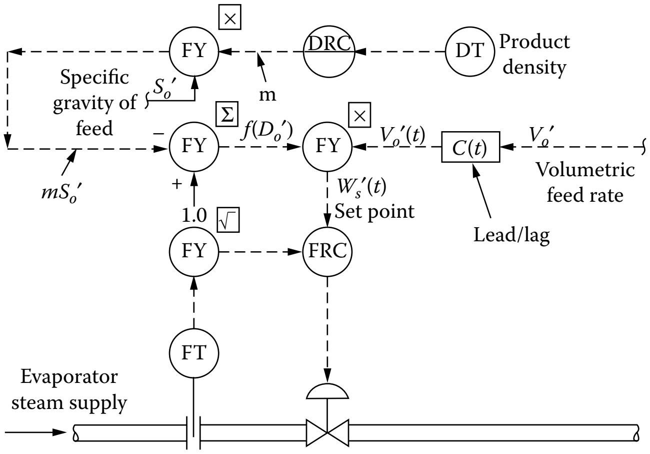


FIG. 8.23u
Feedforward control loop with feedback trimming of the slope m is shown for the case where steam is the manipulated variable.

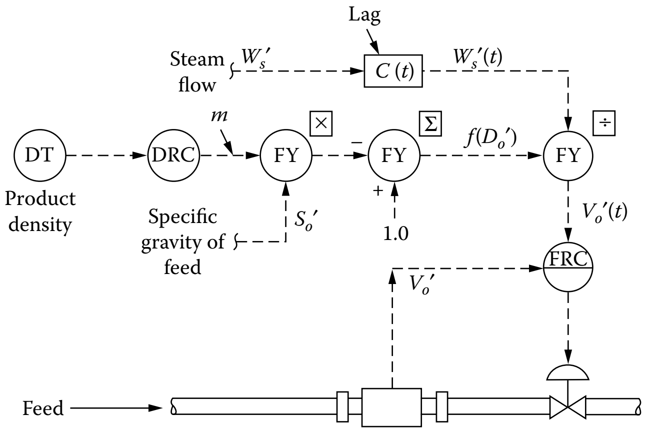


FIG. 8.23v
When feed flow is the manipulated variable, the feedforward signal based on steam flow is lagged while the product density controller trims the slope m in a feedback manner.

will be shown as in Figure 8.23v. The one exception is that the lead-lag dynamics is replaced with a simple lag.

Other Load Variables

Until this point it was assumed that the only load variables in an evaporation process are the feed flow rate and density. This is not the case, and in the next paragraphs, some of the other load variables and the controls they require will be considered.

Steam Enthalpy So far we have discussed only load changes related to feed density and feed flow rate and configured control loops to control the evaporator in response to these variations. One additional load variable that, if allowed to pass through the process, would upset the controlled variable (product density) is steam enthalpy.

In some applications the steam supply may be carefully controlled so that its energy content is uniform. In other applications substantial variations in steam enthalpy may occur. In order to correct for this, one must consider the factors that influence the energy content of the steam and design a control system that will protect the process from the changes in this load variable.

For saturated steam the energy content per unit weight is a function of the absolute pressure of the steam. If the flow of steam to the process is measured with an orifice meter, the mass flow of the steam is

$$W_s = K_1\sqrt{\frac{h}{v}} \qquad \textbf{8.23(31)}$$

where
W_s = steam flow in lb/hr (kg/hr)
h = differential head measurement in ft (m)
v = specific volume in cu. ft per lbm (m^3/kg)
K_1 = orifice coefficient dependent on the physical characteristics of the orifice

Therefore, the total energy to the system is

$$Q = W_s H_s \qquad \textbf{8.23(32)}$$

where
Q = energy to system per BTU per hour (J/hr)
W_s = steam flow in lb/hr (kg/hr)
H_s = heat of condensation in BTU per lb, or J/kg (enthalpy of saturated vapor minus enthalpy of saturated liquid)

Substituting Equation 8.23(31) into Equation 8.23(32) gives

$$Q = K_1\sqrt{\frac{(H_s)^2}{v}h} \qquad \textbf{8.23(33)}$$

For any particular application the steam pressure will vary around a normal operating pressure. To demonstrate the design of a control system to compensate for variations in the energy input into the process, assume that the steam pressure varies between 18 and 22 PSIA (124 and 152 kPa), with a normal operating value of 19.7 PSIA (135.9 kPa). The pressure transmitter has a range of 0–25 PSIA (0–172.5 kPa). These values of pressure variation and operating pressure are typical for a number of evaporator operations.

The value of the $(H_s)^2/v$ term appearing in Equation 8.23(33) will vary, depending on the steam pressure. Over a reasonably narrow range of pressures, the value of $(H_s)^2/v$ can be approximated by a straight line with the general form of

$$(H_s)^2/v = bP + a \qquad \textbf{8.23(34)}$$

where
P = absolute pressure in PSIA (Pa)
a and b = constants

Table 8.23w shows the typical values of H_s, v, $[(H_s)^2/v]'$, and P and P' selected for the specified range of pressures and for a case in which the steam is condensed at 1.5 PSIG.

$$\left(H_s^2/v\right)\text{ normal} = \frac{(970.7)^2}{20.4} = 46{,}208$$

Rewriting Equation 8.23(34) in scaled form:

$$\left(\frac{(H_s)^2}{v}\right)' = bP' + a \qquad \textbf{8.23(35)}$$

The designer can either linearize the data—using any two points from Table 8.23w—or can use a least squares computation to find the best straight line.

TABLE 8.23w
Specific Volume-Enthalpy Data

(P) Steam Supply Pressure PSIA (kPa)	*(H_s) Heat of Condensation BTU/lbm*[†] *(MJ/ka)*	*(v) Specific Volume ft^3/lbm (m^3/kg)*	$((H_s)^2/\upsilon) = \dfrac{(H_s)^2/\upsilon}{((H_s)^2/\upsilon)_{\text{normal}}}$	$P' = \dfrac{P}{P_{\max}} = \dfrac{P}{25}$
18 (124.2)	969.1 (225.4)	22.2 (1.38)	0.916	0.720
19 (131.1)	970.2 (225.7)	22.1 (1.37)	0.965	0.760
19.7* (135.9)	970.7 (225.8)	20.4 (1.26)	1.00	0.788
20 (138)	971.2 (225.9)	20.1 (1.25)	1.02	0.800
21 (144.9)	972.1 (226.1)	19.2 (1.19)	1.07	0.840
22 (151.8)	973.0 (226.3)	18.4 (1.14)	1.11	0.880

* Normal operation.

[†] Assuming that the steam is condensed at 1.5 PSIG.

In a least square computation the following values of *a* and *b* constants of Equation 8.23(35) were obtained:

$$\left(\frac{(H_s)^2}{v}\right)' = bP' + a \qquad \textbf{8.23(36)}$$

where

$a = 0.085$

$b = 1.161$

Squaring Equation 8.23(33) and substituting in Equation 8.23(36):

$$(Q^2)' = (1.161\ P' + 0.085)hk_1 \qquad \textbf{8.23(37)}$$

The parenthetical portion of Equation 8.23(37) can be rewritten so as to make the sum of the two coefficients equal 1.0, which simplifies its implementation using conventional analog hardware. This is done by multiplying and dividing each term in the parentheses by the sum of the two coefficients.

$$(Q^2)' = 1.246\left(\frac{1.161P'}{1.246} + \frac{0.085}{1.246}\right)hk_1 \qquad \textbf{8.23(38)}$$

$$(Q^2)' = 1.246(0.932P' + 0.068)hk_1 \qquad \textbf{8.23(39)}$$

The instrumentation to implement Equation 8.23(39) is shown in Figure 8.23x.

Internal Material Balance The feedforward system described earlier imposes an external material balance as well as an internal material balance on the process. The internal balance is maintained by liquid level control on the discharge of each effect.

Analysis of the performance of level loops indicates that a narrow proportional band (<10%) can achieve stable control. However, because of the resonant nature of the level loop it can cause the process to oscillate at its natural frequency. Therefore, in most installations a much lower controller gain must be used (proportional bands 50 to 100%).[5]

Because of such wider proportional bands, the addition of the integral mode is required to help maintain the set point. A valve positioner and booster relay are also recommended to overcome the usual limit cycle characteristics of an integrating process and the nonlinear nature of valve hysteresis (Figure 8.23y).

Absolute Pressure The heat to evaporate water from the feed material is directly related to the boiling pressure of the material. In most multiple-effect evaporation each effect is held at a pressure less than atmospheric in order to keep boiling points below 212°F (100°C). The lowest pressure is in the effect closest to the condenser, with pressures increasing slightly in each effect away from the condenser.

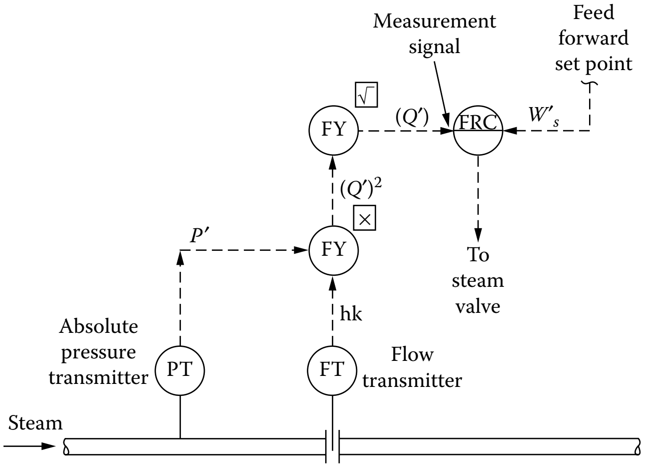


FIG. 8.23x
This control configuration compensates for variations in steam enthalpy by determining the heat input in terms of BTU/hr, based on the calculation required by Equation 8.23(39).

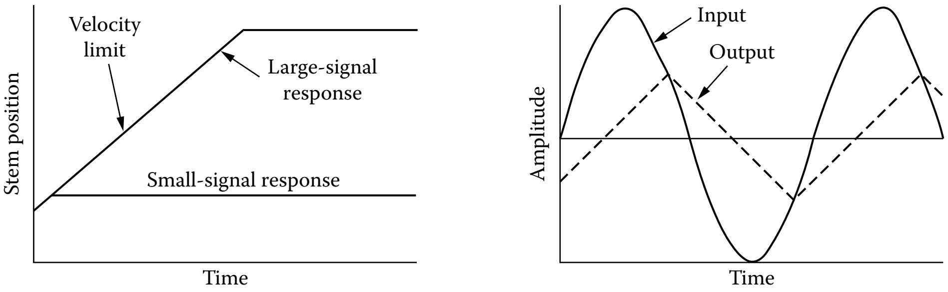


FIG. 8.23y
If a valve actuator cannot move as fast as the speed at which the control signal changes (velocity limited), a step change in the control signal (left) will result in a delayed straight-line response, while a sine wave in the control signal (right) will result in cycling.

The three possible methods of controlling the absolute pressure are (1) controlling the flow of water to the condenser, (2) bleeding air into the system with the water valve wide open, and (3) locating the control valve in the vapor draw-off line, manually setting the water flow rate and air bleeding as necessary.

Method 3 requires an extraordinarily large valve, because the vapor line may be 24 or 30 in. (600 or 750 mm) in diameter. Method 2 is uneconomical, because the expense of pumping the water offsets the savings realized by using a smaller valve on the air line. Method 1 represents the best compromise between cost and controllability, and therefore it is preferred (Figure 8.23z). (For more on condenser pressure control see Section 8.29.)

Auto-Select Controls

In many processes the final product is the result of a two-step operation. The first step produces an intermediate product that serves as the feed to a final concentrator. The aim is to ensure that the process is run at the maximum throughput consistent with the process limitations, an example of which is shown in Figure 8.23aa.

In this two-step evaporation process, three limitations are considered: (1) the steam availability to the intermediate concentrator can be reduced as a result of demands in other parts of the plant; (2) the steam availability to the final concentrator can be reduced as a result of demands in other parts of the plant; and (3) the final concentrator can accept feed only at or below a certain rate, and it is desired to run this part of the process at the rate set by its limits.

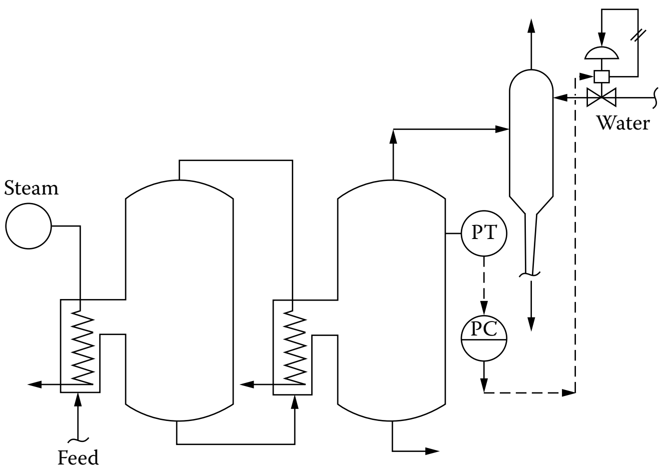


FIG. 8.23z
The recommended method of controlling the absolute pressure of evaporation is to throttle the water flow to the condenser.

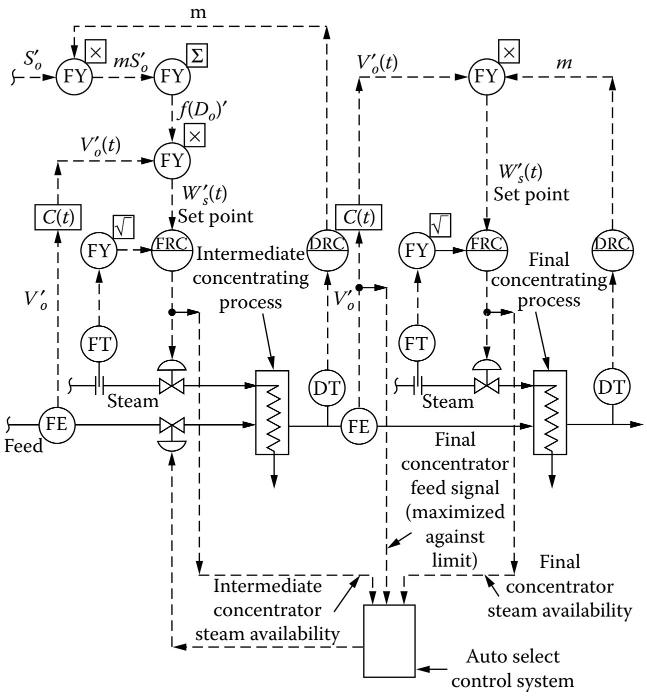


FIG. 8.23aa
This control system includes auto-select limits that consider the maximum feed rate to the final concentrator and the availability of steam in the plant.

Each part of the process has its own feedforward-feedback system. The intermediate concentrating process has the feedforward system shown in Figure 8.23u. The final concentrating process does not require feed density compensation, so that it is necessary only to establish the ratio of steam to feed and to adjust the ratio by feedback. The feed rate to the final concentrator is sent to the auto-select system that manipulates the feed to the intermediate process so as to maintain the final concentrator feed maximized at its set point limit.

Should either steam supply become deficient, the associated FRC output is selected to adjust the feed to the intermediate process. Thus, the process is always run at the maximum permissible rate consistent with the process limitation(s). (For more on auto-select control systems, see the explanation given in connection with Figure 8.23o.)

Trimming Controls

The system illustrated in Figure 8.23bb combines a number of advanced control features. These features include the selective control loop that was described in Figure 8.23o. This loop utilizes the valve position controller (VPC-4) to overrule the feed flow controller (FC-1) if the feed rate exceeds steam availability and requires the steam valve to open beyond 90%.

The feed flow rate is set by a nonlinear level controller (LC-12) on the feed tank. This tank provides surge capacity between the upstream equipment and the evaporator system. LC-12 is nonlinear, meaning that its output signal, which is the set point of FC-1, remains constant as long as the level in the tank is between 40 and 60%. This allows the evaporator loading to be stable, while production surges are averaged out in the feed tank. If the level gets above 60% or below 40%, the feed flow set point is slowly changed to match the new production rate.

The steam flow is measured on the basis of its heat content (enthalpy) as was explained in connection with Figure 8.23x. The set point of the steam flow controller (FRC-2) is obtained on a feedforward basis from the feed flow (FT-1).

Co-current evaporators usually respond faster to changes in feed flow than to steam flow. Therefore, the dynamic compensator in the feedforward loop (FY-11) is a dominant lead in Figure 8.23bb, just as it was in Figures 8.23s and 8.23u. If feed flow was manipulated in response to steam flow variations, the loop would require a dominant lag. (Countercurrent evaporators might respond faster to steam than to feed flow changes and, therefore, require different dynamic compensators.) A lead-lag compensator (FY-11) is usually provided with all feedforward systems and is adjusted as needed in the field.

The feedback trim in Figure 8.23bb is based on product density (DRC-6) and is configured similarly to the loop in Figure 8.23u. The steam flow set point is obtained in accordance with Equation 8.23(18), where E is the economy of the evaporator (the mass of vapor produced per unit mass of steam used).

If the product quality response to the changes in steam flow to the first effect (FRC-2) is too slow, the design engineer might speed up that response by adding a small trim heater to the last effect. Product quality responds faster to this trim heater, but it is more costly to add steam at this point, because the steam is used in only one effect, instead of passing through all.

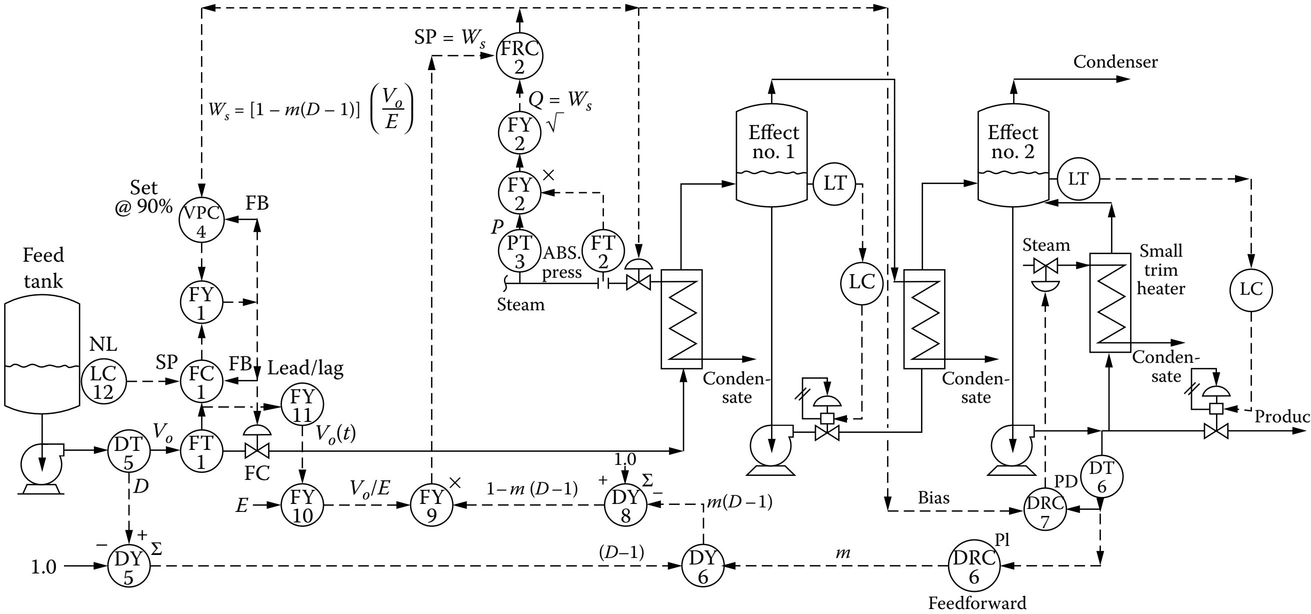


FIG. 8.23bb
Combined control system showing the discussed advanced control features plus controls of a trim heater on the last effect.

When a trim heater is used, as shown in Figure 8.23bb, the product density control is split between a proportional and integral controller (DRC-6), which controls the main steam flow, and a faster, proportional and derivative controller (DRC-7), which throttles the steam to the trim heater.

With feedforward control systems like the one shown in Figure 8.23bb, the control response is fast enough without a trim heater. Yet, if its better response can be economically justified, its controls should be coordinated with the primary steam controls. For this reason DRC-7 is biased by the main steam valve opening, so that they both move in the same direction.

Similar to the use of trim heaters, it is also possible to introduce a small stream of feed directly into the last effect. The use of this technique also improves the response of the system to variations in product quality but is also inefficient, as that stream of feed is exposed only to that one effect instead of to all. Trim controls were used prior to the introduction of feedforward controls, and in many cases the use of trim controls has been discontinued after the installation of feedforward controls.

PRODUCT DENSITY MEASUREMENT

Perhaps one of the most controversial issues in any evaporator control scheme is the method used to measure the product density. Common methods include (1) temperature difference, boiling-point rise; (2) conductivity; (3) differential pressure; (4) gamma gauge; (5) U-tube densitometer; (6) buoyancy float; (7) refractive index; and (8) oscillating Coriolis (see Chapter 6 in the first volume of this handbook).

Each method has its strengths and weaknesses (Table 8.23cc). In all cases, however, care must be taken to select a representative measurement location to eliminate

TABLE 8.23cc
Orientation Table for Density Sensors

LIQUID Density Sensor Design	*Applicable to*			*Minimum Span Based on Water SG = 1.0*	*Inaccuracy in % of Spain or SG Units*	*Design Pressure and Temperature Limitations*		*Temperature Compensation Available*	*Direct Local Indicator*	*Transmitter*
	Clean Process Streams	*Slurry Service*	*Viscous or Polymer Streams*			*PSIG/°F*	*Bars/°C*			
Angular Position Type	✓			0.1	0.5%	1000/500	69/260	N.S.		✓
Ball Type	✓			Digital	0.01 SG	600/160	41/71	N.S.	✓	✓
Capacitance Type	✓	✓	✓	0.1	1%	500/160	34.5/71	✓		✓
Displacement Type Buoyant Force Displacer	✓			0.005	1%	1500/850	130/472	N.S.		✓
Chain Balance Float	✓			0.005	1–3%	500/450	34/232	✓	✓	✓
Electromagnetic Suspension	✓			0.01	0.5–1%	200/350	14/177	✓		✓
Hydrometers	✓			0.05	1%	100/200	7/93	✓	✓	✓
Hydrostatic Head Type	✓	✓	✓	0.05	0.2–1%	5000/350	345/177	N.S.	✓	✓
Oscillating Coriolis	✓	✓	✓	0.1	0.02 SG or better	5000/800	345/426	✓		✓
Radiation Type	✓	✓	✓	0.05	1%	Unlimited	Unlimited	✓		✓
Sonic/Ultrasonic	✓	✓	✓	0.2	1–5%	1000/390	69/199	✓		✓
Twin Tube	✓	✓		Digital	0.0001	1440/356	100/180	✓	✓	✓
Vibrating Fork Type	✓	L	L	0.02	0.001 SG	3000/392	207/200	✓		✓
Vibrating Plate Type (also for gases) (currently not manufactured)	✓	L	L	0.1	0.2%	1440/200	100/95			✓
Vibrating Spool Type (also for gases)	✓	L	L	0.3	0.001 SG	725/300	50/149	✓		✓
Vibrating U.Tube Type	✓			0.05	0.00005–0.005 SG	2900/500	200/260	✓		✓
Weight of Fixed Volume Type	✓	✓	✓	0.05	1%	2400/500	165/260	✓	✓	✓

N.S.: Nonstandard
L: Limited

entrained air bubbles or excessive vibration, and the instrument must be mounted in an accessible location for cleaning and calibration. The relative location of the product density transmitter with respect to the final effect should also be considered. Long runs of process piping for transporting the product from the last effect to the density transmitter increase dead time, which in turn reduces the effectiveness of the control loop.

Boiling-Point Rise

Perhaps the most difficult and controversial method of product density measurement is by temperature difference or boiling-point rise. Dühring's rule states that a linear relationship exists between the boiling point of a solution and the boiling point of pure water at the same pressure. Thus, the temperature difference between the boiling point of the solution in an evaporator and the boiling point of water at the same pressure is a direct measurement of the concentration of the solution. Two problems in making this measurement are location of the temperature bulbs and controls of absolute pressure.

The temperature bulbs must be located so that the measured values are truly representative of the actual conditions. Ideally, the bulb measuring liquor temperature should be just at the surface of the boiling liquid. This location can change, unfortunately, if the operator decides to use more or less liquor in a particular effect. Many operators install the liquor bulb near the bottom of the pan, where it will always be covered, thus creating an error due to *head effects*, which must be compensated for in the calibration.

The vapor temperature bulb is installed in a condensing chamber in the vapor line. Hot condensate flashes over the bulb at an equilibrium temperature dictated by the pressure in the system. This temperature minus the liquid boiling temperature (compensated for head effects) is the temperature difference reflecting product concentration.

Changes in absolute pressure of the system alter not only the boiling point of the liquor but also the flashing temperature of the condensate in the condensing chamber. Unfortunately, the latter effect occurs much more rapidly than the former, resulting in transient errors in the system that may take a long time to resolve. Therefore, it is imperative that absolute pressure be controlled closely if temperature difference is to be a successful measure of product density. These systems are more effective when control of water rate to the condenser rather than an air-bleed system is used.

Conductivity

Electrolytic conductivity is a convenient measurement to use in relationships between specific conductance and product quality (concentration), such as in a caustic evaporator. For the conductivity of some aqueous liquids, refer to Figure 8.23dd. Problem areas include location of the conductivity cell so that product is not stagnant but is flowing past the electrodes; temperature limitations on the cell; cell plugging; and temperature compensation for variations in product temperature.

Differential Pressure

Measuring density by differential pressure is a frequently used technique. The flanged, extended diaphragm differential pressure transmitters are preferred for direct connection to the process (Figure 8.23ee); otherwise, lead lines to the transmitter could become plugged by process material solidifying in the lines. Differential pressure transmitters are more

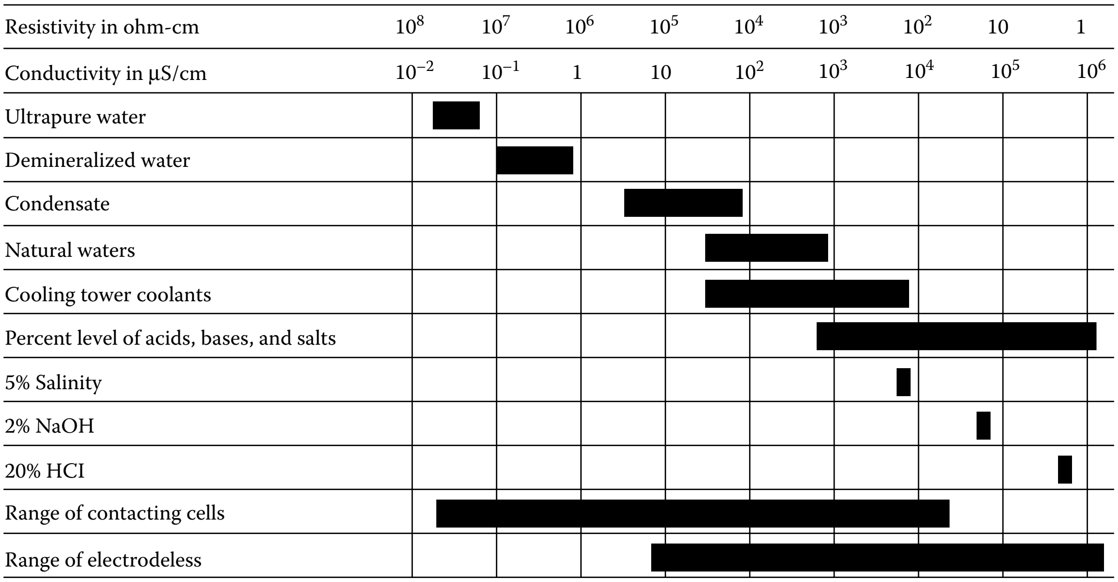


FIG. 8.23dd

Resistivity/conductivity spectrum of aqueous electrolytes. (From Light, T. S., Chemtech, August 1990, pp. 4960–4501.)

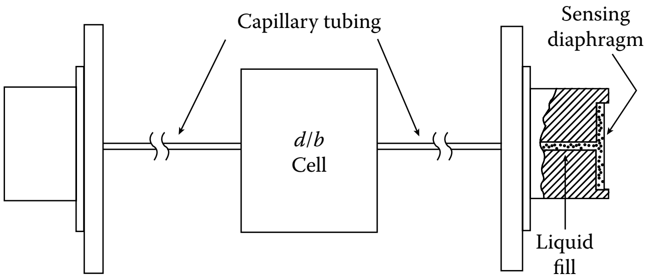


FIG. 8.23ee
d/p cell with extended chemical seal elements.

frequently used on feed density than on final product density measurements.

Gamma Gauge

This measurement is popular in the food industry because the measuring and sensing elements are not in contact with the process. It is very sensitive and not subject to plugging. Periodic calibration may be required because of decay due to the half-life of the source materials. Occasionally, air is entrained, especially in extremely viscous solutions. Therefore, the best sensor location is a flooded low point in the process piping.

U-Tube Densitometer

The U-tube densitometer, a beam-balance device, is also a final product density sensor. Solids can settle out in the measuring tube, causing calibration shifts or plugging.

Buoyancy Float

Primarily used for feed density detection, the buoyancy float can also be applied to product density if a suitable mounting location near the evaporator can be found. Because flow will affect the measurement, the float must be located where the fluid is almost stagnant or where flow can be controlled (by recycle) and its effects zeroed out. A Teflon-coated float helps reduce drag effects.

Oscillating Coriolis

When a Coriolis mass flowmeter (see Section 2.12 in Chapter 2 in the first volume of this handbook) is used for the measurement of feed flow, that same instrument can also provide a density reading for the feed. The error in these measurements is typically between 0.2 and 0.5% SG (0.002 to 0.005 g/cc).

CONCLUSIONS

The process of evaporation is well understood and easily modeled on the basis of mass and energy balance equations. For this reason, *white box*-type (see Section 2.10 in Chapter 2) multivariable models can be used to implement their advanced process control[3] (APC). When using APC-based unit operations controllers, the evaporators can be optimized to reach the criteria of maximized production, minimized energy consumption, or other goals.

References

1. Ritter, R. A. and Andre, H., “Evaporator Control System Design,” *Canadian Journal of Chemical Engineering,* 48: 696, December 1970.
2. McCabe, W. L. and Smith, J. C., *Unit Operations of Chemical Engineering,* New York: McGraw-Hill, 1956.
3. Lipták, B. G., “Process Control Trends,” *Control,* March, May, and July, 2004.
4. Perry, J. H. et al., *Chemical Engineers’ Handbook,* 4th edition, New York: McGraw-Hill, pp. 3-72 to 3-80.
5. Shinskey, F. G., *Process Control System,* 2nd edition, New York: McGraw-Hill, 1979, p. 74.

Bibliography

Babb, M., “New Coriolis Meter Cuts Pressure Drop in Half, *Control Engineering,* October 1991.

Blevins, T. L. et al, *Advanced Control Unleashed,* Research Triangle Park, NC: ISA 2003.

Blickley, G. J., “Tank Gaging Transmitter Performs More Functions,” *Control Engineering,* August 1991.

Boyes, W., “Non-Contacting Density Instrumentation,” ISA/93 Technical Conference, Chicago, September 9–24, 1993.

Capano, D., “The Ways and Means of Density,” *InTech,* November 2000.

Chen, C. S., “Computer-Controlled Evaporation Process,” *Instruments and Control Systems,* June 1982.

“Density and Specific Gravity,” *Measurements and Control,* October 1991.

DeThomas, F. A., “Control of Moisture in Multi-Stage Dryer,” ISA/93 Technical Conference, Chicago, September 19–24, 1993.

Economical Evaporator Designs by Unitech, Bulletin UT-110, Union, NJ: Ecodyne United Division, 1976.

Ewing, G., *Analytical Instrumentation Handbook,* New York: Marcel Dekker, 1990.

Extance, P. et al., “Intelligent Turbidity Monitoring Using Fiber Optics,” First International Conference on Optical Fibre Sensors, IEE, London, 1983.

Farin, W. G., “Control of Triple-Effect Evaporators,” *Instrumentation Technology,* April 1976.

Hashimoto, M. et al., “An Automated Method for Bacterial Test by Simultaneous Measurement of Electrical Impedance and Turbidity,” *Japanese Journal of Med. Engineering,* February 1981.

Hilliard, C. L., “Design Solutions for Typical Problems of Colorimetric Process Analyzers,” October 1992.

Hoyle, D. L. and Nisenfeld, A.E., “Dynamic Feedforward Control of Multieffect Evaporators,” *Instrumentation Technology,* February 1980.

Hugo, A., “Limitations of Model Predictive Controllers,” *Hydrocarbon Processing,* January 2000.

Lipták, B. G., “Process Control Trends,” *Control,* March, May, and July, 2004.

Livingston, A. J., “Density Measurement, Nuclear Gauge Measurement of Density,” *Measurements & Control,* December 1990.

Luyben, W. L. and Tyreus, B. D., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Magaris, P., “On-Line Density Measurement Is Fast and Accurate,” *Control Engineering,* June 1981.

McKetta, J. J., *Unit Operations Handbook,* Volume 1, Mass Transfer, New York: Marcel Dekker.

Ociai, S., “Creating Process Control Parameters from Steady State Data,” *ISA Transactions,* Volume 36, 1998.

Perry, R. H., *Chemical Engineers Handbook,* 7th edition, New York: McGraw-Hill, 1997.

Shinskey, F. G., *Process Control Systems,* 4th edition, New York: McGraw-Hill, 1996.

Shinskey, F. G., *Controlling Multivariable Processes,* Research Triangle Park, NC: Instrument Society of America, 1981.

Spencer, G. L. and Meade, G. P., *Cane Sugar Handbook,* 9th edition, New York: McGraw-Hill, 1963.

Stickney, W. W. and Fosbery, T. M., “Treating Chemical Wastes by Evaporation,” *Chemical Engineering Progress,* April 1976.

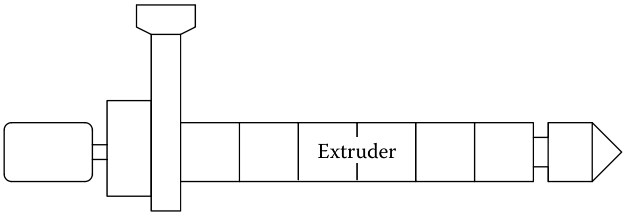
Waller, M. H., “Measurement and Control of Paper Stock Consistency,” ISA Conference in Houston, October 1992.

8.24 Extruder Controls

G. A. PETTIT (1972, 1985),

B. A. JENSEN (1995),

M. L. BERNS (2005)



Flow sheet symbol

INTRODUCTION

Integrated extrusion control requires more than controlling the extruder and associated equipment variables at appropriate set points. Extrusion control integrates all the functions an operator normally performs in the correct order at the proper time. This involves the complete integration of controlling, interlocking, sequencing, and scheduling of the extruder system with the other parts of the plant.

An equally important consideration in the automation effort is the interfacing to allow the operator or the higher level plantwide control system to retrieve the necessary data and to provide the means to control the operations of the equipment up- and downstream of the extruder.

Implementation of the various forms and levels of extruder automation is a continuing stepwise process. Achieving the benefits of extrusion automation is a matter of layering information functions and advanced supervisory strategies upon the basic regulatory and interlocking control strategies.

Extruder Production Rate

The extruder output flow is a function of the screw speed, plastic viscosity, screw leakage, pressure build-up, and so on. Assuming an uninterrupted resin supply and negligible screw overflight leakage, the relationship between the production rate of the extruder and the above listed factors is as follows:

$$Q = \frac{k_1 N - k_2 P}{\mu} \qquad \textbf{8.24(1)}$$

where

k_1 = constant, based on geometry of the last screw flight
N = screw speed
k_2 = constant, based on geometry of melting zone
P = pressure of the melt at the nose of the screw
μ = viscosity of plastic

Equation 8.24(1) shows that the output of the extruder is proportional to the speed of the screw and that an increase in output pressure reduces output flow rate.

Polymer Types and Characteristics

All lists of commercial polymers are incomplete owing to the rapid progress of polymer science and production. The common extruded products, such as polyethylene and polystyrene, extrude rather simply and are relatively stable, and their behavior is reasonably predictable.

Polymers in the vinyl chloride group are not so stable at extrusion temperatures. If not promptly removed from the process, they decompose, emitting hydrogen chloride gas, which is very corrosive to metal. Consequently, dies for vinyl chloride are commonly chrome-plated and designed to eliminate pockets or crevices where the material may become entrained and decompose.

An important characteristic of synthetic polymers that complicates their extrusion is sensitivity to shear rate, a phenomenon known as non-Newtonian behavior (Equation 8.24[1]). The relationship between apparent viscosity and shear rate is also dependent on temperature and must be considered when extruder controls are considered.[4] Certain polymers, such as polyvinyl chloride (PVC) and acrylonitrile-butadiene-styrene (ABS), are much more sensitive to shear rate than are more crystalline materials, such as nylon and acrylics. This is especially true if color pigment additives are used.

A wide variety of additives or fillers are available to reduce or eliminate various types of degradation (thermal, mechanical, chemical, environmental). Additives such as UV stabilizers, antioxidant, processing aids, flame retardants, lubricant concentrates, and impact modifiers are normally used in very small quantities. Other additives include reinforcing materials such as fiberglass and, of course, pigment additives for color. Fillers include mineral fillers, such as talc and $CaCO_3$ plus mica, which are added to improve physical properties and function as low-cost extenders.

COMPONENTS OF THE EXTRUDER SYSTEM

The extruder itself is the main component in an extrusion process. It is composed of a drive motor, gear box, screw, barrel zones, and die. On older systems, the screw drive motors are generally DC drives, while on newer systems

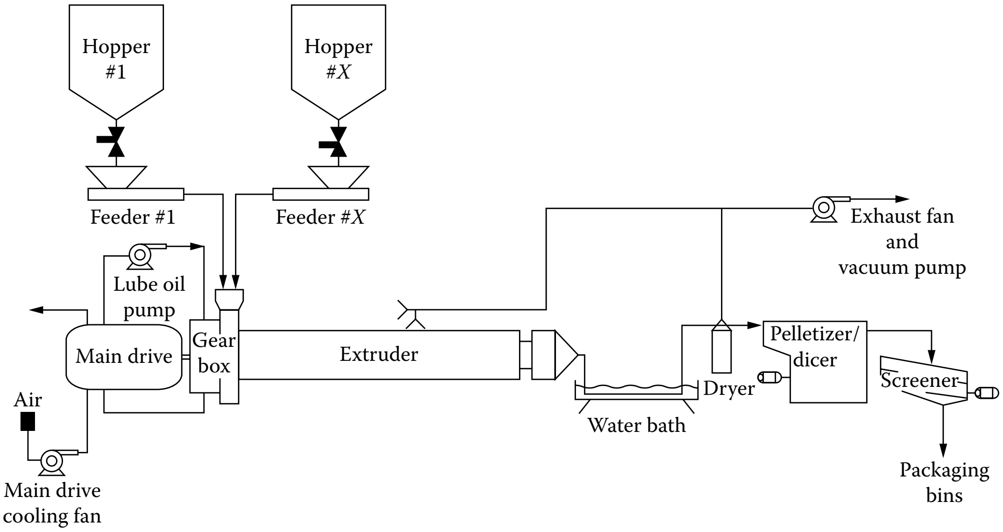


FIG. 8.24a
Extrusion flow diagram.

variable-frequency AC drives or open-loop vector-controlled AC drives are more often used (see Section 7.10 for details on variable-speed drives). The most important component in the compounding process is the screw-type extruder. In addition, an extruder system consists of several other pieces of equipment, including: blenders and feeders on the inlet of the extruder and coolers (water baths), pelletizers, tube cutter-pullers, shears, and stackers on its outlet side. In addition, the system includes such auxiliary equipment as fans and vacuum pumps.

An example of an equipment layout is shown in Figure 8.24a.

Blenders

The resin feed usually is blended with different additives such as pigments and fillers before it is fed into the extruder. The blenders can be volumetric and gravimetric. Both types consist of a resin station, one or more dosing stations, the mixer, and a controller.

The feed rate of a volumetric feeder typically ranges from 3 to 1500 ft^3/hr (0.085 to 42.5 m^3/h) and is determined by the gear ratio and the diameter of the metering screw. The associated microprocessor-based controller performs the following functions: 1) allows the operator to create, store, and recall recipes, 2) controls the running time of the dosing stations, 3) controls the running time of the agitator, 4) maintains the level of the material in each dosing station, 5) displays the status of the operation, and 6) warns the operator if abnormal conditions occur. Figure 8.24b provides a block diagram of the volumetric blender.

A gravimetric feeder measures the actual weight of each additive before loading it into the mixer. The rate of a gravimetric feeder typically ranges from 0.55 lb/hr to 16,500 lb/hr (0.25 kg/h to 7500 kg/h). A microprocessor-based controller performs the same control functions on a gravimetric blender as was described for the volumetric blender. Blender manufacturers use the same controllers for both types of feeders. Figure 8.24c provides a block diagram of the gravimetric blender.

Cooling System

When the molten extrudate exits the die, cooling is required to solidify the product. Different types of cooling systems are used, including: 1) water baths with vacuum sizing, which are used in pipe and tube manufacturing, 2) cooling rollers, which are used in sheet manufacturing, and 3) water baths, which are used in coating processes, including wire and cable covering.

The vacuum sizing tank system consists of the tank, which is under vacuum, a water reservoir, water, and vacuum pumps. The water pump circulates the water between the water reservoir and the vacuum tank through entrance and exit rings. The continuous circulation of the water ensures adequate cooling of the extrudate.

The molten extrudate enters the vacuum sizing tank through an entrance ring. The diameter of the entrance ring determines the size of the pipe or tube. The difference between the air pressure inside the tube or pipe and the vacuum outside it determines the inside diameter of the tube or pipe. The vacuum pressure in the tank can be controlled manually or automatically. (For a description of vacuum gauges and vacuum transducers, refer to the first volume of this handbook.)

When the extruded sheet leaves the die it passes through an air space of 3–10 ft (1–3 m). Here, additional cooling is

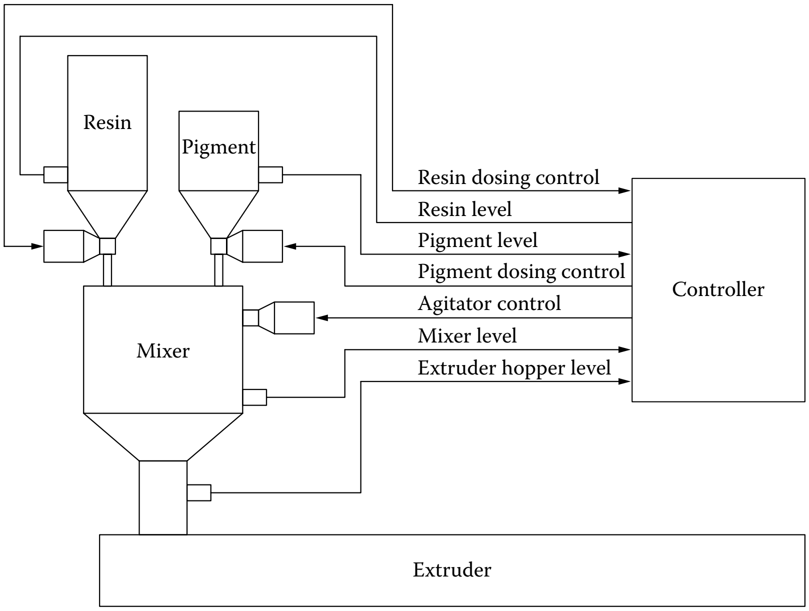


FIG. 8.24b
Block diagram of a volumetric blender and its controls.

provided by a set of cooling fans, which are placed above and below the sheet. In some applications, water-cooled rollers are also used to provide adequate cooling of the material.

When covered wires and cables are manufactured, they are cooled by passing through a cooling trough. To reduce the length of the trough and to maintain the high speed of production, it might be necessary to have several passages inside the trough.

Cutters

Depending upon the end product of the extrusion process, a cutting machine can be used to cut the extrudate strands into shorter lengths.

Pelletizers and Dicers Depending upon the type of the extrudate, a pelletizer, dicer, or some other kind of cutting machine is used to cut the strands into shorter lengths. Pelletizers

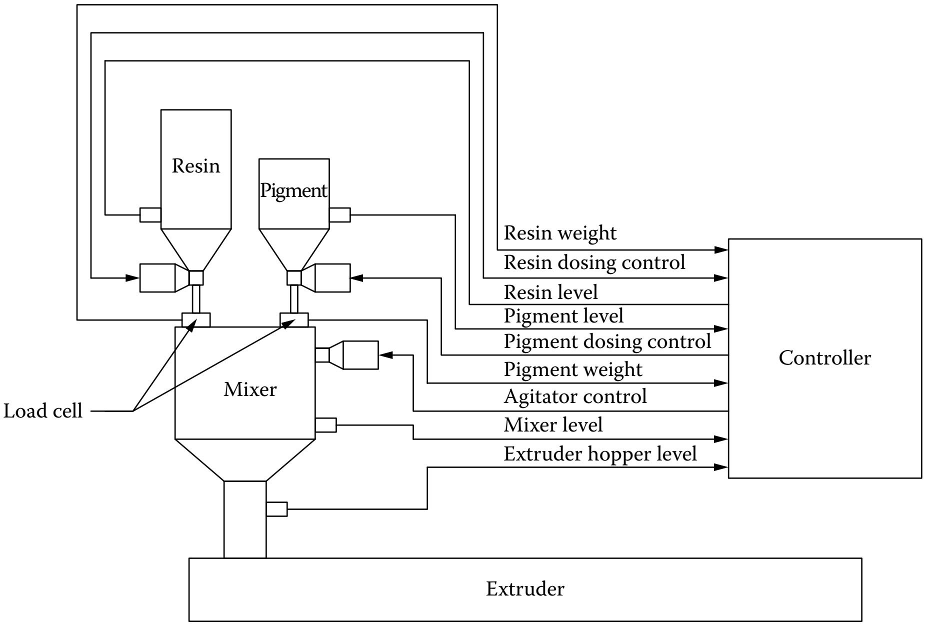


FIG. 8.24c
Block diagram of a gravimetric blender and its controls.

generally cut spaghetti-like strands into small pellets suitable for packaging in bags or boxes. Dicers are used to cut sheets into strips suitable to be wound up into rolls, or they can also be left flat.

Control of the pelletizers and dicers primarily involves the control of their speed so as to maintain the size of the pellets and the ratio of the pelletizer and feed speeds constant. The speed ratio is achieved by feedforward control of the pelletizer/dicer speed based upon the dynamically compensated master feed rate to the extruder.

Pullers-Cutters Cutter-pullers are used in the production of pipes and thin-wall tubes. As their name implies, they pull the extrudate through a vacuum sizer/cooler and cut the tube or pipe to the desired length. The speed of the puller is set by a slave controller to correspond to the speed of the extruder. Reciprocal movement of the cutting head is synchronized with the linear speed of the sleeve or pipe, and at a preset length, a blade cuts the sleeve and retracts. Then the cutting head moves back by the distance equal to the required tube or pipe length.

Auxiliary Equipment

Exhaust fans and vacuum pumps are some of the auxiliary equipment required to remove fumes from the extruder. Because during compounding a variety of gaseous by-products can form, special vent sections are provided in the extruder to allow the removal of these gases. Venting is also required to remove the vapors generated by moisture removal and devolatilizing. A decompression chamber is incorporated at the point of venting so that the process material does not extrude out through the vent port.

When the processing involves the sending of strands through a water bath, often these strands are sent through a vacuum system to dry them.

Screeners are used to sift pellets, ensuring common diameters and lengths.

EXTRUDER TYPES AND SUBSYSTEMS

Single-Screw Extruders

Single-screw extruders usually convert granular resin feeds into sheets, films, pellets, and shapes such as pipe. These extruders are described by their screw diameters (in inches or millimeters) and by their L/D ratio, L being the screw length and D the screw diameter.

Single-screw extruders are available in almost any size imaginable. Common sizes are $^5/_8$, $^3/_4$, 1, $1^1/_2$, $2^1/_2$, $3^1/_2$, $4^1/_2$, $5^1/_2$, 6, 8, 12, 15, and 20 in. (16, 19, 25, 37.5, 62.5, 87.5, 112.5, 132, 150, 200, 300, 380, and 500 mm). L/D ratios range from 5:1 to 48:1, with 20:1 to 30:1 being some of the common choices. The single-screw extruder is by far the most frequently used design; however, the twin-screw extruder is becoming even more popular.[1]

Twin-Screw Extruders

Machines using twin screws are generally large-volume production units used for resin pelletizing in petrochemical plants. They are equipped with various combinations of intermeshing and nonmeshing screws that can be either the co-rotating or the counter-rotating variety. The following types of processing can be performed in a single machine: 1) melting, 2) mixing and blending, 3) homogenizing, gelling, and dispersing, 4) reacting, 5) pumping, 6) compounding and formulation, 7) devolatilizing and degassing, and 8) drying.

Twin-screw machines are often melt-fed directly from polymerization reactors and perform multiple functions on the polymer prior to pelletizing and packaging it as a finished product. The twin-screw extruder is normally selected as the solution to many compounding and reactive extrusion tasks. Twin-screw extruders can be either intermeshing or nonintermeshing. Nonintermeshing extruders behave like two single-screw extruders with only minor interactions between the two screws.

A further subdivision of twin-screw types is the direction of rotation. Co-rotating extruders have both screws rotating in the same direction, and therefore the material is exchanged between the screws, while counter-rotating screws transport the process material through the extruder in a figure eight channel.

Compounding requires that the resin be melted and homogenized while incorporating additives or fillers at a given shear level. The key is the isolation of the high, medium, and low shear sections along the screw length, as well as the feeding of additives at the appropriate points. The advantages of twin screws include the capability for mixing, dispersion, and heat control in addition to efficient conveying.

Extruder Dies and Barrels

Die Types The shape and the ultimate use of the extruded product are defined by the die shape. Dies are broadly classified as follows:

1. Sheet dies for extruding flat sheets, up to 120 in. (3 m) wide and $^1/_2$ in. (12.5 mm) thick.
2. Shape dies for making pipe, gaskets, tubular products, and many other designs.
3. Blown film dies, using an annular orifice to form a thin-walled envelope. The diameter of the envelope is expanded with low-pressure air to roughly three times the annular orifice diameter to form a thin film. The process is used for films up to 5 mils thick (0.005 in., or 0.125 mm) at the upper limit.
4. Spinnerette dies for extrusion of single or multiple strands of polymer for textile products, rope, tire cord, or webbing.

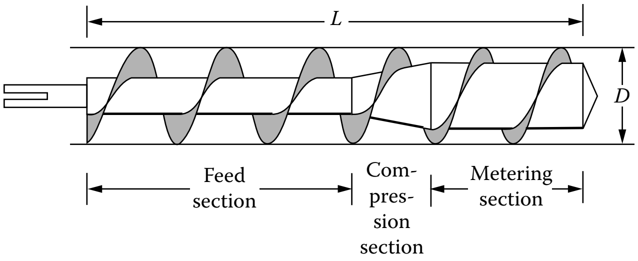


FIG. 8.24d
Typical extruder screw.

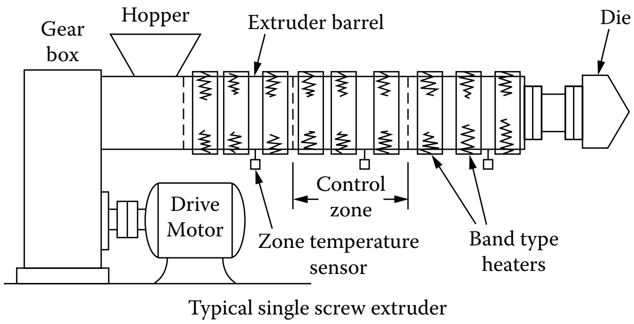


FIG. 8.24e
Single-screw extruder with band heaters.

5. Pelletizing dies for granular products in resin production, synthetic rubbers, and scrap reclaiming. These dies form multiple strands roughly $^1/_8$ in. (3.125 mm) in diameter. Rotating knives continuously cut the strands in short lengths, after which the pellets drop into water for cooling.
6. Cross-head dies for wire coating, in which the bare wire or cable enters the die and emerges coated with semimolten polymer. The wire enters and leaves the die at an angle of 90° to the extruder axis.

Many special configurations of extruder dies are used to produce composite films. In such designs, two polymers enter the die from two extruders and exit as a sheet. The top and bottom polymer layers are of different chemical composition so that one might obtain a film or sheet of two colors or to utilize the other desirable characteristics of both materials.

Dies for rigid foam production are similar to blown film dies. Special dies with moving parts can continuously extrude netting. Most dies require a short connecting pipelike piece, commonly called the adaptor, which connects to the extruder head. Also, twin-screw extruders require an additional special adaptor piece called an "eight-to-oh" (8/O) zone to convert the twin configuration (which physically looks like a figure eight lying on its side) to an output configuration of the adaptor (zero or the letter "O").

A typical extruder screw is shown in Figure 8.24d.

Barrels and Heater Rings The barrel of an extruder is usually divided into roughly 15–18 in. (375–450 mm) long temperature control zones. A $4^1/_2$ in. (112.5 mm) extruder, for example, may have four to six barrel zones. The number of zones depends upon the L/D ratio, the number of feed points for additives, and on the type of material being extruded; 4–6 zones is most common, but it is not unheard of to have 12–15 zones (Figure 8.24e).

Conventional temperature control loops include a power control device, temperature sensor, and heater for each zone. Adapter, 8/O, and die zones may all be provided with temperature controllers.

Extruders require large heater ratings to decrease heating time. A typical barrel zone electrical heater can be rated anywhere from 1.2 to 11 kW and greater. The size generally is determined by the type of feed, the screw type, and the resin feed location. External heating of the barrels can be provided by electric resistance heaters and by hot oil systems.

Band-type silicon control rectifier (SCR) resistance heaters are of two-piece construction to facilitate removal. Another type of electric heating is inductance heating, which is accomplished by coiling a copper wire about the barrel to induce an electromagnetic field. This technique is more responsive than resistance-type heating. The coil is energized with 60 cycle current and is usually controlled by contactors.

After extrusion has begun, heat is internally generated from the friction and shear of the rotating screw. This heat is a function of the square of the screw speed. The amount of heat generated is also influenced by screw design, head pressure, and resin viscosity. In some zones, the melt temperature may rise above accepted maximum, and barrel cooling is required.

Cooling Systems

The barrels are commonly cooled by fans (Figure 8.24f) or by a water-cooled jacket. Many extruders use aluminum heating shells with casting heaters of encased nichrome coiled

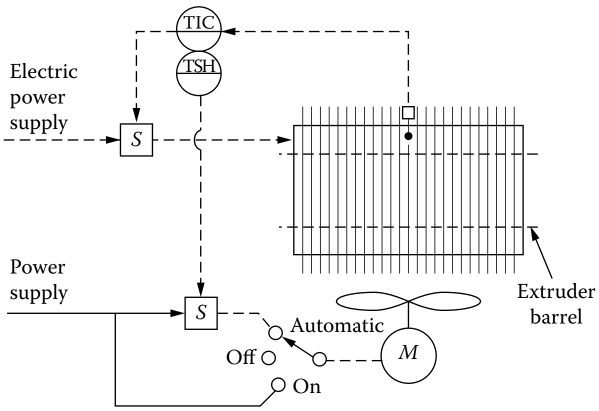


FIG. 8.24f
The extruder barrel is provided with fins and with temperature controls using electric heating and cooling by air fans.

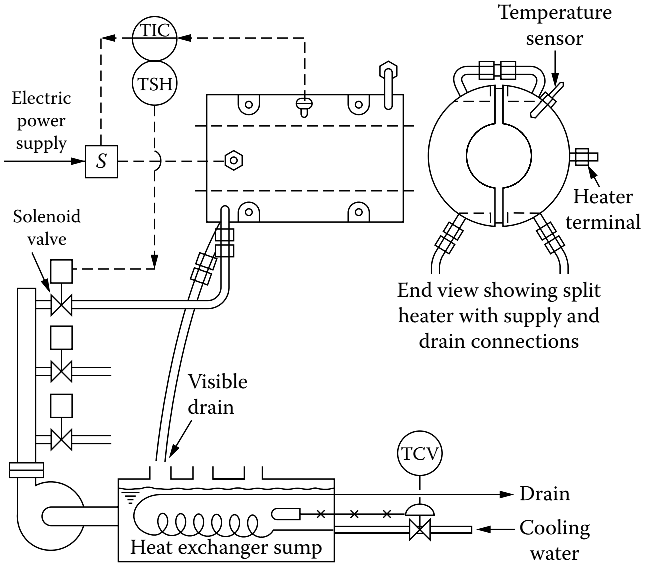


FIG. 8.24g
The extruder barrel is provided with fins and with temperature controls using electric heating and cooling by water.

throughout the aluminum jacket. The shells are constructed in two segments clamped together to surround the extruder barrel. Shells for fan cooling have fins cast to the outside surface to increase their surface area. Motor-driven fans or blowers are positioned directly below the zone, and when switched on by a high-temperature switch, they cool the barrel by forced convection. A zone fan on a $2^1/_2$ in. (62.5 mm) extruder can typically remove the equivalent of 5 kW per hour.

Water-cooled extruders are provided with aluminum shells with cast-in tubing as well as heaters. Treated water is continuously circulated through the coils by a common pump and heat exchanger (Figure 8.24g). Solenoid valves, regulated by the zone temperature controller, allow circulation to each zone in order to maintain the proper barrel zone temperature. The solenoid valves can be operated by an auxiliary high-temperature switch, which is part of the zone temperature controller. The switch is designed to operate in a time-proportioning manner with extra slow cycle rate and capability for very short pulses.

Water-cooling can be too effective compared with fan cooling. The cooling water is usually flashed into steam at most barrel temperatures used for processing thermoplastic materials. Some machine builders use compressed air to clear the water from the passages and to eliminate trapped fluid, which could cause erratic cooling.

Cooling water is often injected in very short pulses, which are followed by the immediate removal of the water to reduce cooling. Running the exchanger sump at higher temperatures also reduces the severity of water-cooling.

SENSORS, VARIABLES, AND THEIR CONTROL

Temperature Measurement

Accurate temperature measurement is very important for efficient extrusion and overall quality of the end product. The most common extruder temperature sensors are thermocouples (TCs) and resistance temperature detectors (RTDs). RTDs detect the change in their electrical resistance, which is proportional to their temperature, while thermocouples produce a millivoltage output, which is related to the temperature of the TC's junction.

A detailed discussion of all temperature sensors is provided in Chapter 4 of the first volume of this handbook.

Temperature Control

The zone temperature of extruders is usually controlled by time proportioning PID controllers. In such a control system, the on-time of the electric heater is modulated by the PID controller, while the total cycle period (usually 8–10 sec) remains fixed. In some systems each zone is provided with its own single-loop controller, while in others a programmable logic controller (PLC) executes the PID algorithms in sequence.

Time-proportioning control is less expensive than continuous modulation, because it allows the use of electromechanical or solid-state contactors. A block diagram of the time-proportioning controller is shown on Figure 8.24h.

Duration of on-time period of the final control element is a function of the continuous output of the PID controller. Therefore, if the output of the controller is 40% and the cycle period is 10 sec, then the on-time period of the solenoid or contactor is 4 sec. The output signal generated by the time-proportioning control system in response to a continuous PID controller output is illustrated in Figure 8.24i.

Temperature control is complicated by several factors. These include heating caused by the shearing action of the screw, changes in the feed rate, conduction of the heat along

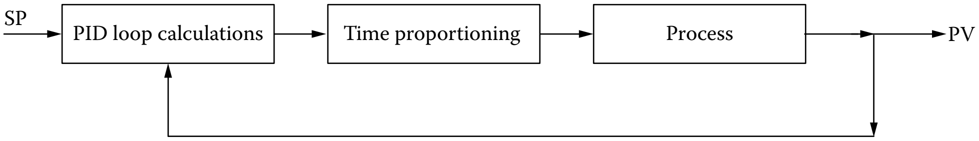


FIG. 8.24h
Block diagram of a time-proportioning controller.

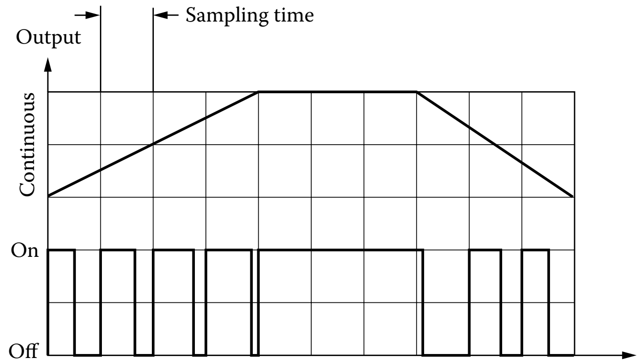


FIG. 8.24i
Illustration of the continuous output signal from a PID controller (top) and the corresponding output generated by a time-proportioning controller (bottom).

the barrel, and the difference between the process gains that exist during heating and during cooling. An increase in screw speed increases the shear in the screw channel and the mechanical work, which raises the resin temperature. At the same time an increase in screw speed and output flow rate reduces the amount of cooling, because the resin spends less time in contact with the same heat-transfer surface.

If the operating pressure is low, due to a low-resistance die size, the amount of frictional heat developed may not be sufficient to compensate for the decrease in conducted heat, and therefore, the resin temperature will drop as the screw speed increases. At higher pressures, the temperature may show an initial increase before dropping off. At still higher pressures (heavy screen packs or small dies), the same screw and operating conditions can result in an increased temperature with increasing screw speed, because the mechanical heating effect exceeds the decrease in conducted cooling.[2]

In some cases, as much as 80–100% of the heat produced throughout the extruder can be generated by screw shear alone. Effective temperature control requires accurate removal of heat as well as accurate application of the heat from the barrel heaters.

Zone temperature control is designed to maintain a constant temperature in each barrel, adaptor, and die zone to achieve a desired profile and thus guarantee the consistent melting of the extrudate. Feeding and compression zones are most effectively controlled at relatively constant temperatures, because under such conditions they will provide the optimum wall friction for efficient feeding.

Temperature can have a dramatic effect on both throughput and product quality. Good temperature control in the rear zones can increase the hourly output for some materials. On the other hand, some crystalline thermoplastics are relatively impervious to temperature variations of as much as 50°F.

Zone temperature is usually controlled either by single-loop PID control loops or by PLC-executed PID controls. However, the installation and maintenance of the RTD-, thermocouple-, or thermistor-type sensors is critical. (See Chapter 4 in the 4th edition of Volume 1 of this handbook.) The major problem is to maintain an optimal temperature profile, which is a function of the dynamics of heating, cooling, shear heat, and conduction. Adaptive tuning and the use of dead time PID control can improve performance.

Pressure Measurement

The accurate knowledge of extruder pressure is very important for efficient extrusion. Chapter 5 of the first volume of this handbook provides an in-depth description of all the available pressure sensors.

The commonly used extruder pressure sensors include grease-sealed gauges. These are direct-reading Bourdon tube gauges with a tube tip capillary for complete filling with high-temperature grease, usually silicone. The grease remains viscous at high temperatures and prevents the molten polymer from entering and solidifying in the gauge or piping. Disadvantages of this sensor include the need for periodic greasing and the occasional contamination of the product.

Force-balance transmitters provide a linear output signal that can be sent to digital or analog pressure indicators, recorders, and controllers. Their force-balance operating principle makes them less sensitive to process temperature variations.

Strain gauge-type pressure transducers are higher accuracy pressure sensors, and for that reason, they are widely used in extruder applications.

Pressure Control

Synthetic fiber processes frequently use an extruder to melt and transport nylon and polyesters to a bank of gear pumps feeding individual spinning die heads. The shear characteristics of these polymers are near-Newtonian and their viscosity is relatively constant at wide variations in shear rate. In such processes, the pressure is nearly proportional to screw speed, and therefore pressure can be controlled by the manipulation of the screw speed.

In contrast, PVC can be pumped through a restriction orifice at increasing rates without an increase in the pressure drop across the orifice, because its apparent viscosity decreases with increased pumping shear rate. In this case, the pressure cannot be controlled by the manipulation of the screw speed.

Figure 8.24j illustrates the control scheme used in fiber processes. Because some of the fiber spinning pumps frequently fail or are stopped intentionally, it is necessary that the pressure controller quickly respond to such events by immediately slowing the extruder screw to a new output rate determined by the number of constant volume pumps remaining in operation.

In the past, screw drive motors were generally DC drives or eddy current clutches. Today, these drives are being substituted by variable-frequency AC drives. These drives can

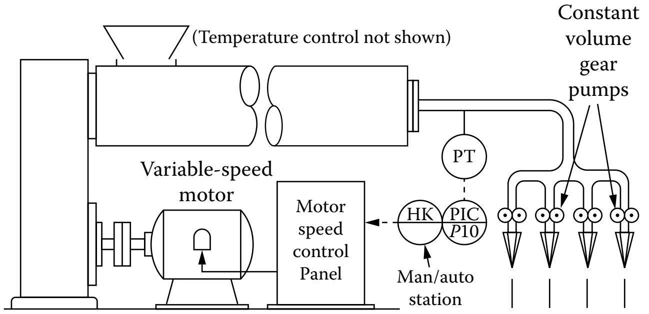


FIG. 8.24j
Extruder pressure control used in the production of synthetic fibers.

be modulated by either digital or analog control signals. The variable-frequency drives provide virtually instantaneous response, which is important to guarantee stable control.

Depending upon shear characteristics, die head pressure can remain uncontrolled or can be controlled by either feed rate or screw speed. Synthetic fiber processes frequently use an extruder to melt and transport nylon and polyesters to a bank of gear pumps feeding individual spinning die heads. The shear characteristics of these polyesters are near-Newtonian and, thus, have relatively constant viscosity at wide variations in shear rate. This characteristic is responsible for the discharge pressure being nearly proportional to screw speed, which is necessary for pressure control by the regulation of the screw speed.

In contrast, PVC can be pumped through a restricted orifice at increasing rates without an increase in back-pressure because its apparent viscosity drops with increased pumping shear rate. The pressure of such a process is nearly impossible to control by manipulating screw speed.

Film Thickness Control

Blown film, which is used in the production of tubular and flat film to 5 mils (0.125 mm) thickness, requires thickness measurement and control. Thickness can be adjusted by the speed of the take-up rolls if the thickness uniformly varies (in machine direction) across the film. Increasing the take-up speed reduces the film-gauge, while decreasing the speed increases its thickness (Figure 8.24k).

Thickness variations in die direction (across the width) can be measured, but most attempts to automate sheet die nip adjustments have been unsuccessful. Sheet dies extrude materials up to $^1/_2$ in. (12.5 mm) thick and 120 in. (3 m) wide.

Measurement of film thickness to 100 mils (2.5 mm) is made with radiation instruments (Figure 8.24l), using beta rays. Among design variations are scanning heads that measure and record thickness over the entire width of materials. Infrared, LVDT, laser, capacitance, ultrasonic, and mechanical devices have also been used. For details on these sensors, refer to Section 7.20 in Chapter 7 in the first volume of this handbook.

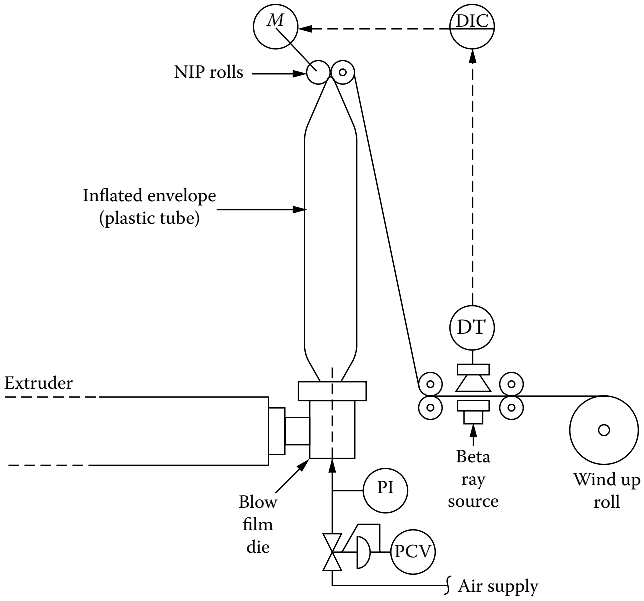


FIG. 8.24k
Film thickness control through nip roll speed manipulation.

Polymer viscosity variations of the feed constitute the most difficult aspect of extruder design, simulation, and control. The feed viscosity tends to have more influence on the extrusion process and on the dimensional quality of the pelletizing extruder output than all other variables. For the description of viscosity detectors, see Sections 8.62 to 8.64 in Chapter 8 in Volume 1 of this handbook.

In tube and pipe production lines, the outer diameter, wall thickness, and concentricity are determining factors of produce quality. Outer diameter is set by the entrance ring of

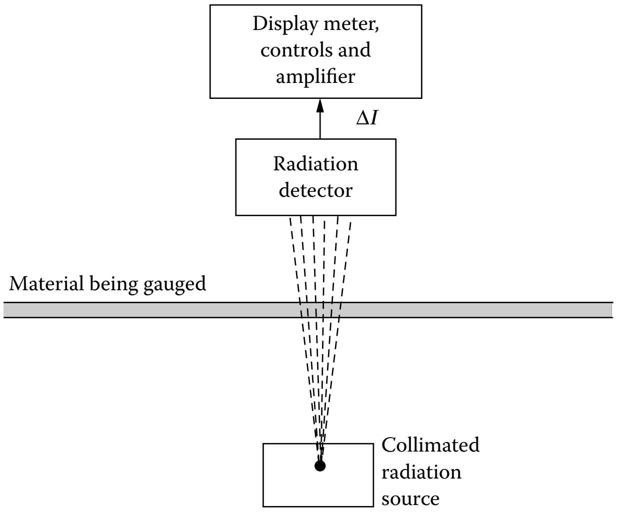


FIG. 8.24l
Radiation absorption gauge.

the water tank and can be slightly adjusted by the speed of the puller. Concentricity can be adjusted by controlling the vacuum inside the vacuum tank. Wall thickness can be adjusted by controlling extrusion speed.

EXTRUDER CONTROL SYSTEMS

The overall extrusion control system serves more than controlling the extruder and associated equipment at appropriate set points. Extrusion control should integrate all the functions that an operator normally performs into a control system that keeps the steps of the extrusion process in the correct order at the proper time. This includes the complete integration of controlling, interlocking, sequencing, and scheduling the extruder system with the other parts of the plant.[5]

An equally important consideration in the automation effort is the interfacing provided for the operator to retrieve data and to provide the means for control. Modern extruder control systems are based on programmable logic controllers or industrial computers to control temperatures and screw speeds. Using programmable logic controllers and networking offers systems that not only govern machine functions and parameters but also integrate them with upstream equipment, such as the feeders and mixers, and with downstream equipment, such as pelletizers, dryers, and conveyors.

Basic Control

Regulatory, discrete, and interlock controls are needed to operate an extrusion system. The typical basic control system consists of the several subsystems, including: 1) feeder mass flow rate control, 2) individual zone temperature control along the extruder barrel, 3) motor speed control, 4) die head pressure controls, and 5) auxiliary discrete devices for on/off operation of fans, pumps, pelletizers, and so on.

Mass flow rate control is designed to allow the operator to set the overall mass flow rate to the extruder. Feed rate control for blending is implemented by setting mass flow rate set points for each feeder, based upon the individual percentages determined by the recipe, and multiplied by the overall mass flow rate. For a detailed discussion of mass flow meter designs, see Sections 2.11 to 2.13 in Chapter 2 in the first volume of this handbook.

Main screw speed control is designed to maintain a reasonably consistent torque and specific energy of the extrudate at various feed rates. Implementation is based upon feedforward control from the master feed rate to the extruder screw speed, which is adjusted in a simple ratio. Extruders used in compounding tend to be run at top speed for extended periods and require only minor variations in RPM.

Interlocks build upon regulatory and discrete control functions in order to provide additional equipment and personnel protection in the case of hardware failure. The most common interlocks control the main drive motor, lube oil pump, feeders, feed hoppers, and pelletizers/dicers.

Advanced Control

Advanced controls include the cascade control of melt temperature, zone flux control, auto-tuning of temperature controllers, maximizing production, and extrudate quality control.

Melt Temperature Control The output melt temperature is a function of the internal shear energy (converted to heat energy) plus or minus the conducted heat (barrel cooling), depending on the operation. The temperature of the melt is as important, because it influences the output rate for quality extrusion. The cascade control to reset the zone controller set point achieves continuous temperature control of the melt. Figure 8.24m shows the arrangement for cascade feedback.[6]

One special feature is that the system allows only depression of the zone temperature controller set points, which is a safety consideration because of heat degradation and pressure build-up in polymer systems. A safety interlock with the extruder screw drive is usually incorporated in order to prevent both polymer freezing during shutdowns and drive damage at start-up.

Zone controllers provide both heating and cooling, and their set points are regulated by the melt controller. Each

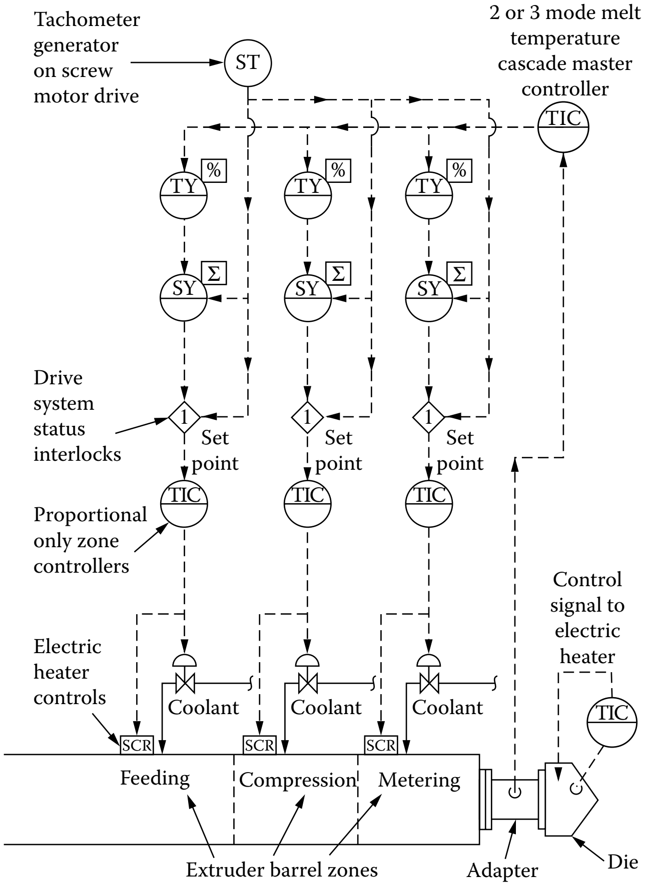


FIG. 8.24m
Melt temperature cascade system.

zone temperature set point can be adjusted individually to follow a certain percentage of the feedback cascade signal so that a preset program of zone depression will follow a definite barrel temperature zone profile curve. The extruder metering zones are capable of the greatest heat transfer and, therefore, receive the greatest percentage of feedback and most set point depression.

Feeding and compression zones are most effective at a relatively constant temperature to provide optimum wall friction for most efficient feedings; therefore, they usually receive a small percentage of the feedback signal. The melt controller must have proportional and integral control actions[7] with rate action being beneficial. Some cascade systems also incorporate a tachometer feedforward signal to supply set point depression proportional to screw speed. This function reduces temperature departures from the control point caused by screw speed drift or by intentional speed changes.

Flux Control Zone flux control is designed to maintain the desired temperature at all points throughout the extruder so that a consistent melt of the extrudate is achieved, which is the same objective as that of basic zone temperature control. Flux controllers are used to include feedforward control from feed rate and the drive motor screw.

Flux is defined as the heat per unit area in the extruder zone. It is the energy due to the heating elements, plus the energy due to the screw shear, minus the energy removed by cooling, and minus the energy required to heat and melt the resin. These factors must be dynamically compensated and applied to the measurement of the flux controller.

Both the motor and the feed energy terms provide feedforward responses to "load" disturbances. These disturbances are generally not significant during normal operation but become important during start-up and shutdown. This is because as the screw starts, the motor power begins to rise, which increases the flux controller's measurement. In response to this, the flux controller calls for a decrease in heat, resulting in a decrease in output to the heating element or an increase in the rate of cooling.

As feed is added to the extruder, the heat required to warm and melt the feed results in a decrease in the flux controller's measurement. In response to this, the flux controller's output increases, calling for more energy to be put into the zone. Therefore, the control algorithm must include dynamic terms, which have to be field-tuned. Figure 8.24n illustrates the control block diagram for zone flux control.

Auto-Tuning Maintaining the extruder barrel temperature profile is the most important factor in controlling the quality of the final product. Therefore, the temperature control loops must be tuned to ensure their optimal performance. Loop tuning is performed by initiating a small change of the set point (usually 10%) and adjusting the gain of the control loop to achieve the fastest response (see Figure 8.24o). For an in-depth discussion of PID controller tuning, refer to Section 2.35 in Chapter 2.

Most PLCs and microprocessor-based single-loop controllers have an auto-tuning feature that implement the Ziegler-Nichols method of tuning. Adaptive auto-tuning is designed to continually update the PID-type zone temperature controller's tuning constants in response to process gain variations caused by load variations. There are three methods of adaptive auto-tuning:

Programmed adaptive auto-tuning is based on automatic adjustment of the proportional gain as a function of changes in process dynamics. For example, the proportional gain of the controller can be affected by changes in the controlled variable, in the set point, and so on. The programmed adaptive algorithm can be set for independent gain adjustment using any combination of the process-related variables.

Model-based auto-tuning uses an internal model of the process to determine the optimal PID settings. In this case, the controller introduces step changes above and below the set point and observes reaction to these changes. The PID settings are then calculated according to predetermined criteria.

The *pattern recognition* method continuously examines the response of the process to naturally occurring disturbances such as set point or load changes. When the resulting error exceeds some threshold value (controller looks for peak values) and after several peaks (usually two or three), the controller determines the period of oscillation that is related to the dead time of the loop and, based on it, calculates the new PID tuning constants. See Chapter 2 for a general discussion of control theory and specifically for process modeling and controller tuning.

Maximized Feed Rate Control Figure 8.24p illustrates a master feed control system that is designed to maximize the throughput of resin to the extruder by increasing the feed rate until a constraint is reached. The master feed rate controller set point is constrained by the outputs of a number of limit controllers, which are sent to a select network. The output of the selector is the lowest of any of the constraint controllers.

The output of the low select then becomes one of two inputs into a high selector limit. The other input is the minimum feed rate setting of the extruder. The resulting output is the cascade master's feed rate set point for the feed hopper flow controllers. A set point for minimum feed rate is included in order to have a low limit clamp.

The constraint variables normally include the die head pressure, drive motor amps (or motor torque), extruder throat or chute level, maximum feed rate, and minimum feed rate. Many compounding extrusion processes can take advantage of a maximum feed rate control because of excess upstream storage capabilities. However, when feed to the extrusion system is from some other continuous process, this type of control may not be possible because of limited feedstock availability.

The ultimate test of the quality of any control system is the quality of the extrudate. Quality is affected by many

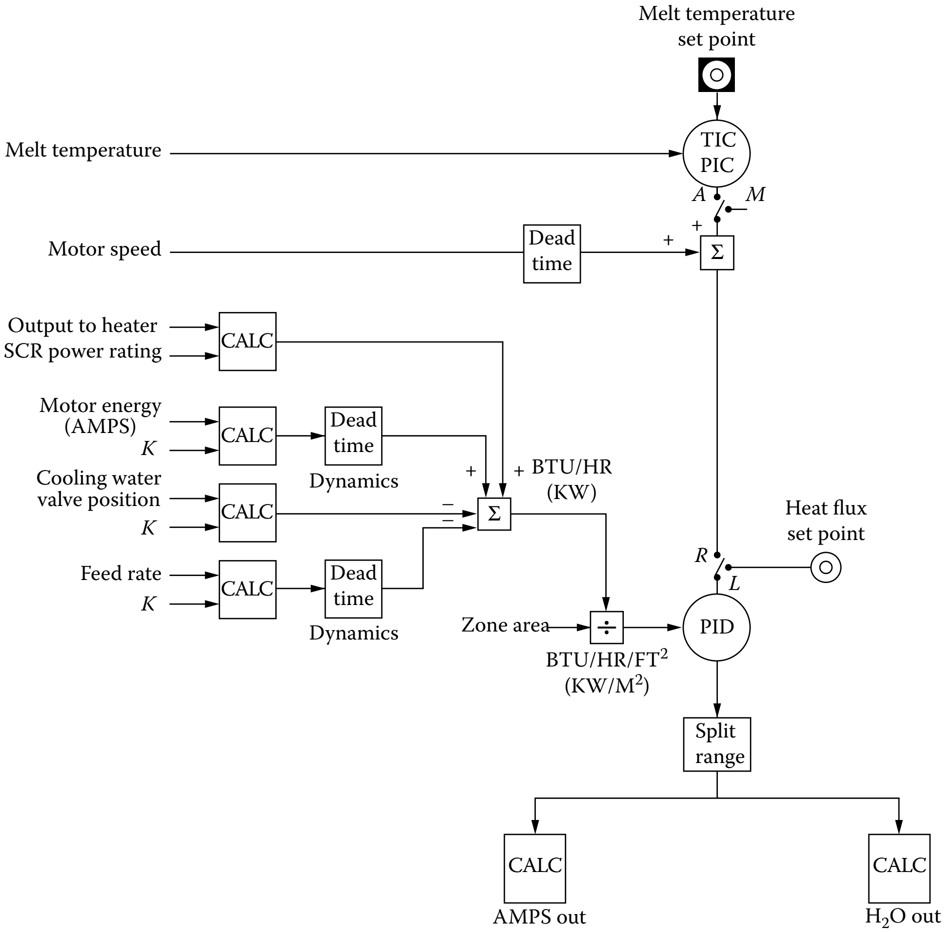


FIG. 8.24n
Heat flux control.

parameters, including feed mixing, barrel temperatures, melt temperatures, screw speed (shear rate), enthalpy (from external sources), pressure, and rheological conditions. On-line rheometers, machine vision analyzers, or other specialty instruments can be used to measure certain extrudate properties. They are becoming more numerous and more reliable.

Quality indicators measured by these instruments include such properties as melt flow index, viscosity, gels, color,

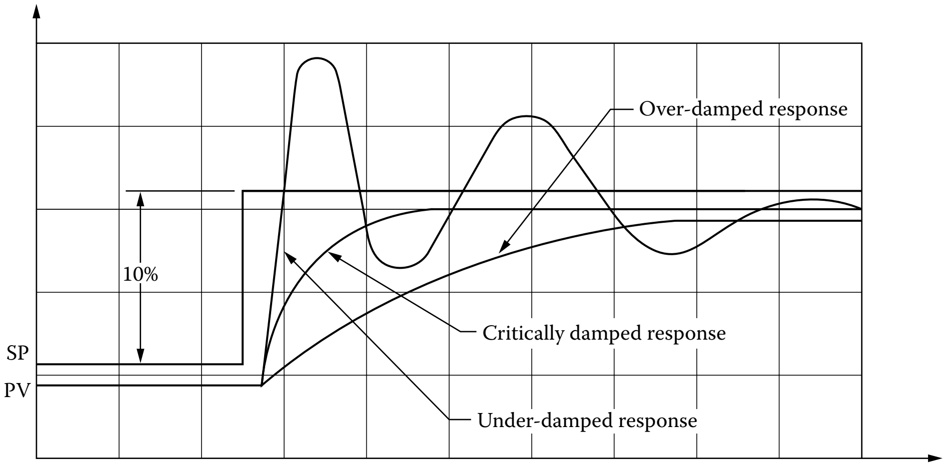


FIG. 8.24o
The response of a PID loop as a function of narrowing the proportional band (increasing the controller gain). The cycling curve corresponds to a high gain setting, the critically damped curve to the correct setting, while the overdamped curve to a low gain setting.

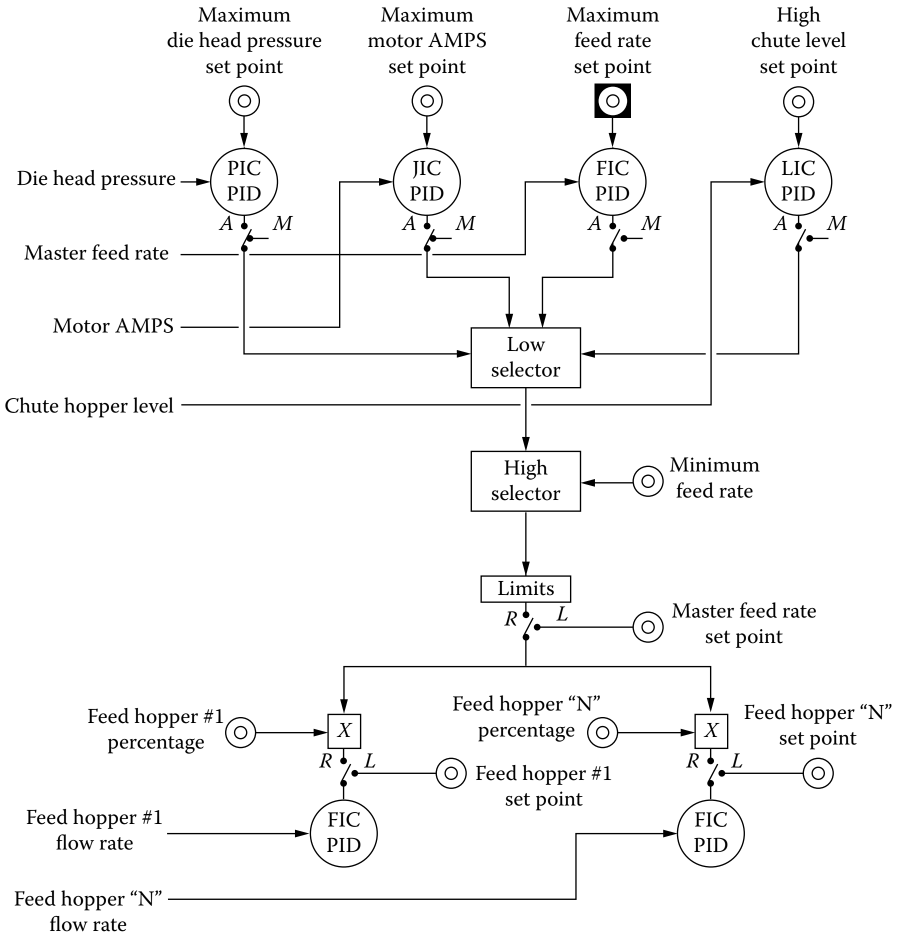


FIG. 8.24p
Optimized control system, serving to maximize the feed rate to the extruder while obeying limit constraints.

agglomerates, and excess ingredients, which are all described in Chapter 8 of Volume 1 of this handbook. A combination mass flowmeter and viscosity detector is illustrated in Figure 8.24q.

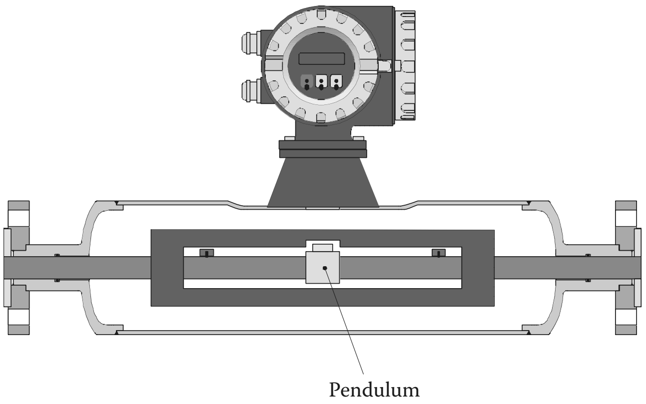


FIG. 8.24q
A pendulum mounted on a single Coriolis tube generates both torsional and lateral oscillations, which allow for the simultaneous measurements of mass flow and viscosity.

A measurable, repeatable value can be used as feedback to one of the above-listed quality-related parameters. Calculations determining extrudate properties can also be used. Choosing a manipulated variable from the above list is challenging because of the lack of sensitivity and the limited number of degrees of freedom available.

One method of controlling the quality of the extrudate is the manipulation of the feed ratio, i.e., the mixing of a particular feed resin with another resin of different properties while all other parameters are held constant. Model-based control or multivariable control can also be used to control extrudate quality, but in most polymer processes the majority of quality parameters of the feed can only be corrected at the reactor and not at the extruder.

The large dead time of the process can also render feedback control impractical.

Engineering Calculations Increased capabilities of modern digital controls allow engineering calculations to be displayed on the human-machine interfaces (HMIs) for the operator. The displayed variable trends include specific energy, extruder motor speed, torque, and current. Additional parameters such as actual zone temperatures and deviations, equipment run times, and other values are also displayed on auxiliary screens.

One such calculation is the physical property known as melt factor or melt flow index (MFI) of the extrudate. Another calculation could include viscosity or other indexing properties. The quality calculation is normally computed as a continuous on-line value. Because the laboratory result of the sample is not available until some time after the sample was taken, recording the current conditions allows a time-synchronized comparison of the laboratory result with the calculated value.

The process conditions are recorded at the time that the sample is taken. The conditions from which MFI is normally calculated are die pressure, melt temperature, and mass flow rate. However, feed mixing, screw speed, and rheological conditions are other nonmeasured parameters affecting extrudate properties.

Another useful calculation is specific energy of the extruder at the die head. The power input, whose units are horsepower (watts), is multiplied by an efficiency constant and by a constant conversion factor. This value is divided by the total feed rate to give a result in the units of hp/lb (kw/kg). This property relates to the shear rate and, thus, to the viscosity of the resin.

Torque and percent maximum torque are important operational parameters for the extruder. A linear relationship can be assumed to exist between the motor current (in amps) and screw speed (in RPM). A constant voltage is assumed, allowing power to be calculated from motor amps. Percent maximum torque is also calculated from the current and screw speed. The current torque measurement is divided by the maximum current rating and maximum drive speed, and is expressed in percent.

Equipment run-time numbers are generally of interest to maintenance personnel. The number of times the main drive motor has been started, the percentage up-time since initialization, total run time, and percentage up-time can be calculated for each extruder. The cutting time and the number of times each pelletizer/dicer has been started gives maintenance personnel information on when cutting blades should be sharpened or when the equipment should be overhauled.

Sequences Sequences are automated so that step-by-step procedures control the routine activities. Included in this category are heater/contactor check, heat-up, start-up, shutdown, and alarm disabling/enabling.

Heater/contactor checks provide the ability to test proper operation of the heater/contactors and extruder thermocouples, for those heating systems that are using electrical heating elements. A heater/contactor is suspected to be "bad" whenever the maximum output is sent to the zone and the expected amperage read back is not within a certain percentage of the expected value. The problem is either a bad contactor or bad heater legs.

Extruder heat-up provides the ability to heat the extruder to preset values in a series of steps and to display the duration of the heat-up. This results in a heat-up profile in accordance with the manufacturer's specifications in order to minimize thermal stress on the equipment.

Extruder start-up provides the ability to bring the extruder equipment on-line from a heat-up condition to preset values in a series of steps. The procedure checks system interlocks, starts auxiliary equipment, and ramps the extruder speed and feed rate to their preset recipe values. In addition, the appropriate alarms are enabled as the sequence progresses. Once the feed material is ready and a heat-up of the extruder has been performed, the extruder can be started up in an orderly stepwise manner.

Extruder shutdown sequencing provides the ability to bring the extruder equipment off-line from a running condition in a series of steps. The procedure turns off the feeders, decreases screw speed in a preset manner, ramps the pelletizer speed to a minimum value, and shuts down the main drive and auxiliary equipment. Appropriate alarms are inhibited as the sequence progresses. In addition, a "panic button" is normally provided for emergency shutdown. This button cuts all power to the line, resulting in immediate shutdown of the extrusion process in case of emergency.

The enabling and disabling of alarms is automated to reduce the frequency of unnecessary signals. Normally this function can be performed by the start-up and shutdown procedures. Whenever a unit is being started up, alarms need to be enabled to warn of abnormal occurrences during normal operating conditions. Similarly, when a unit is being shut down, certain alarms need to be disabled or inhibited to prevent alarms from actuating, because it is normal that during shutdown a number of alarm thresholds will be violated as equipment is being shut down.

Integrated Control

Management information control is a higher level automation activity. Timely information provided to engineers and management allows them to better evaluate performance and maintain consistency. Integrated control includes those additional steps that serve to optimize the overall utilization and performance of the extruders. Included in this category are line scheduling, lot history, recipe management, and statistical process control and statistical quality control (SPC and SQC).

Line scheduling provides the ability to optimize extruder availability. Optimizing extruder availability maximizes overall throughput. An individual extruder, chosen from a set of many extruders, may need to be made available to run a particular type of raw material to produce a certain grade of product. Factors affecting the scheduling of a single extruder line includes the amount of material to be run, availability of raw material, grade of material, color of material, size of extruder (screw diameter and number of barrel zones), extruder configuration (single or twin screw), and the availability of auxiliary equipment.

Lot history is a function that captures all the information associated with the process of extruding a resin product. This is analogous to batch tracking in a chemical plant. The information associated with extrusion of a resin product is normally categorized according to a lot number. Select process alarm information and process conditions are captured and stored as part of the production history on file. Quality standards within the batch industry mandate the reporting of all activity associated with a batch run.

Recipe management maintains a database of master recipes for various products, formulas, and procedures. Specific information for the extrusion equipment and extrusion lines that can perform the required operations are contained within the recipe. The recipe is selected and accessed according to the relationship of the material to be processed with a specific extruder line.

The recipe management function provides the ability to automate the start-up and normal operation of the extruder. Set points stored from previously run recipes and formulas are useful, because they store the relationship to the grade of materials that were processed in the past and can be inserted into the controllers for new runs on the same extruder and associated equipment. These set points include barrel zone temperatures, feed rates, and screw speed.

Statistical quality control provides a history of quality data for the extrudate. This process flags statistically improbable events in real time, evaluates process performance, and validates newly installed operating conditions. Various calculations and statistical tests can be made for each run. Real-time and historical graphs can then display these data to the operators and engineers as required.

CONCLUSIONS

During the last ten years digital control systems including programmable logic controllers have become more powerful. Today, even mid-size PLCs can perform complex calculations, including some modeling and floating-point mathematics. This allows extruder manufacturers to implement control systems with complex algorithms that include auto-tuning of temperature control and pressure control loops.

In addition, the human-machine interfaces allows storage of recipes and can display historical trends of many operating parameters. Because most mid-size PLCs have extensive networking capabilities, extruder manufacturers are able to offer integration with upstream equipment, such as feeders and mixers, and with downstream equipment, such as pelletizers, dryers, and conveyors.

System control and data acquisition (SCADA) packages provide plant management with production information that allows them to improve quality control, planning, and allocation of resources.

References

1. Collins, S. H., "Update on Continuous Compounding Equipment, Part 2: Twin-Screw Extruders and Specialized Compounders," *Plastics Compounding,* November/December 1982, pp. 29–44.
2. Maddock, B. H., "Factors Affecting Quality in Polyethylene Extrusion," Thirteenth Annual National Conference of the Society of Plastic Engineers, 1967.
3. Mielcarek, D. F., "Twin-Screw Compounding," *Chemical Engineering Progress,* June 1987, pp. 59–67.
4. Bernhart, E. C., *Processing of Thermoplastic,* New York: Van Nostrand Reinhold, 1959.
5. Jensen, B. A. and George, J., "Extruder Automation," *Proceedings of the ISA/90*, Vol. 45, Part 2, New Orleans, LA: Instrument Society of America, October 1990.
6. Pettit and Ahlers, "Extruder Controls," *SPE Journal*, Vol. 24, November 1968.
7. Meissner, S. C. and Meneges, E. P., *Technical Papers,* Vol. XVII, Society of Plastic Engineers, May 1971.
8. Bateson, R. N., *Introduction to Control System Technology,* 5th edition, Engelwood Cliffs, NJ: Prentice Hall, 1996.
9. Johnson, C. D., *Process Control and Instrumentation Technology,* 6th edition, Upper Saddle River, NJ: Prentice Hall, 1996.
10. Petruzella, F. D., *Programmable Logic Controllers,* New York: McGraw Hill Companies, 1990.
11. Hensen, F., *Plastic Extrusion Technology*, 2nd edition, Cincinnati, OH: Hanser-Gardner, 1997.

Bibliography

Bouilloux, A., Macosko, C. W., and Kotnour, T., "Urethane Polymerization in a Counter-Rotating Twin-Screw Extruder," *Industrial and Engineering Chemistry Research,* Vol. 30, pp. 2431–2436, November 1991.

Chui, Sh.-H. and Pong, Sh.-H., "Fuzzy Supervisory Predictive PID Control of Plastic Extruder Barrel," *Journal of Applied Polymer Science,* September 1999, pp. 541–555.

Colbert, J. A., "Optimization of Compounding Extruders," *Plastics Compounding,* December 1989, pp. 25–32.

Collins, S. H., "Update on Continuous Compounding Equipment, Part 1: Single-Screw Extruders," *Plastics Compounding,* September/October 1982, pp. 69–82.

Fisher, H., "Electric Drives in Extrusion," *Kunststoffe Plast Europe,* September 1998, pp. 26–29.

Hopkins, B., "Temperature Control of Polymerization Reactors," *Instrumentation Technology,* May 1973.

Janssen, L. P., *Twin-Screw Extrusion,* New York: Elsevier, 1978.

Kepper, D. and Haney, J., "An Overview of Modern Extrusion Technology," *Rubber World,* June 1987.

Kilham, L. B., "A Systems Approach to Quality Control in the Production of Polymer Products," *Tappi Journal,* April 1989, pp. 113–116.

Klein, I. and Klein, R. J., "The Effect of Resin Properties on the Extrusion Process," *Tappi Journal,* August 1989, pp. 99–103.

Kramer, W. A., "Extruder Temperature Control with Adaptive Reset," *Engineering Plastics,* September 1996, pp. 341–350.

Marschall, U. and Reyer, K.-H., "Automating Extruder Control," *Kunststoffe Plast Europe,* February 1998, pp. 10–12.

Martin, T. L., "Temperature Control in Plastic Processing," *Plastics,* April 1967.

Mercure, P. K. and Trainor, R., "Modeling and Optimization of Extruder Temperature Control," *Polymeric Materials Science and Engineering,* Vol. 58, pp. 692–696, 1988.

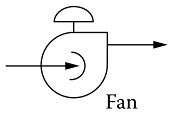
Mober, C., "Control Schemes for Plastic Molding Machines Conserve Energy," *Instrument and Control Systems,* April 1982.

Tsai, Ch.-Ch. and Lu, Ch-H., "Fuzzy Supervisory Predictive PID Control of Plastic Extruder Barrel," *Journal of the Chinese Institute of Engineers,* September 1998, pp. 619–624.

Yang, B. and Lee, L. J., "Process Control of Profile Extrusion Using Thermal Method, Part 1. Mathematical Modeling and System Analysis," *Polymer Engineering Science,* November 1988, pp. 697–707.

Yang, B. and Lee, L.J., "Process Control of Profile Extrusion Using Thermal Method, Part 2. Closed Loop Control," *Polymer Engineering Science,* November 1988, pp. 708–717.

8.25 Fan Controls



Flow sheet symbol

B. G. LIPTÁK (1995, 2005)

INTRODUCTION

The transportation of gases and vapors is an important unit operation in all plants, particularly in the unit operations that involve the moving of large volumes of air, such as dryers, boilers, cooling towers, and HVAC systems. The reduction in the cost of air or gas transportation contributes to plant optimization. This section describes the automatic control of fans and provides some advice on how to operate them at minimum cost.

Fans transport large volumes (up to 1 million CFM) of gases at low pressures, usually at discharge pressures of a few inches of water to up to a couple of PSIG. For a detailed discussion of fan controls for HVAC, boiler, and cooling tower applications, refer to Sections 8.2, 8.6, 8.16, and 8.17. The emphasis of this section will be on the industrial application of fans.

FAN TYPES

Fan designs are classified into axial flow and radial (also called centrifugal) flow types, because of the difference between the nature of the flow through the blade passages. They move large volumes of air at relatively low pressures (in. H_2O), with the pressure-flow characteristics as shown in Figure 8.25a.

If the operating point is to the left of maximum pressure on the fan curve, and particularly if the fan pressure exceeds 10 in. H_2O, it is likely that pulsation and unstable operation will occur. This maximum point on the fan curve is referred to as the *surge point* or *pumping limit*. All fans must always operate to the right of the surge point.

Under low load conditions, the fan can be kept out of surge by artificially increasing the load by venting the gas that is not required by the process. The other possible way of eliminating surging is to substitute the use of dampers on the fan discharge as final control elements with blade pitch, speed, or vane control.

FAN CONTROLS

Fan Throttling

Figure 8.25b shows both fan curves and system curves. The system curves, the relationship between flow (velocity) and fan discharge pressure, in mostly friction transportation systems are parabolic. The operating point for the process is where the fan curve crosses the system curve (point *A* on Figure 8.25b).

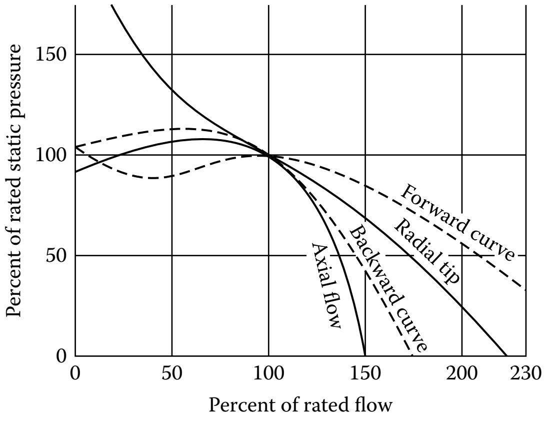
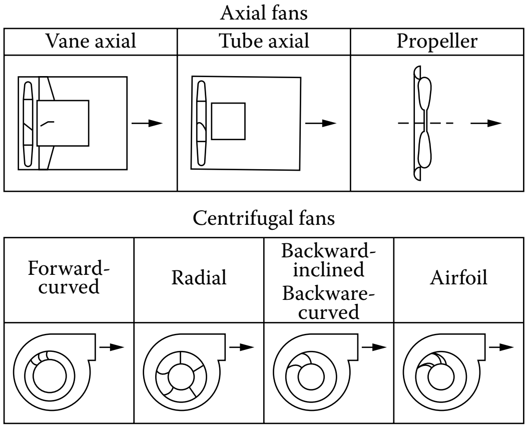


FIG. 8.25a
Different fan types exhibit different pressure-flow characteristics.[1]



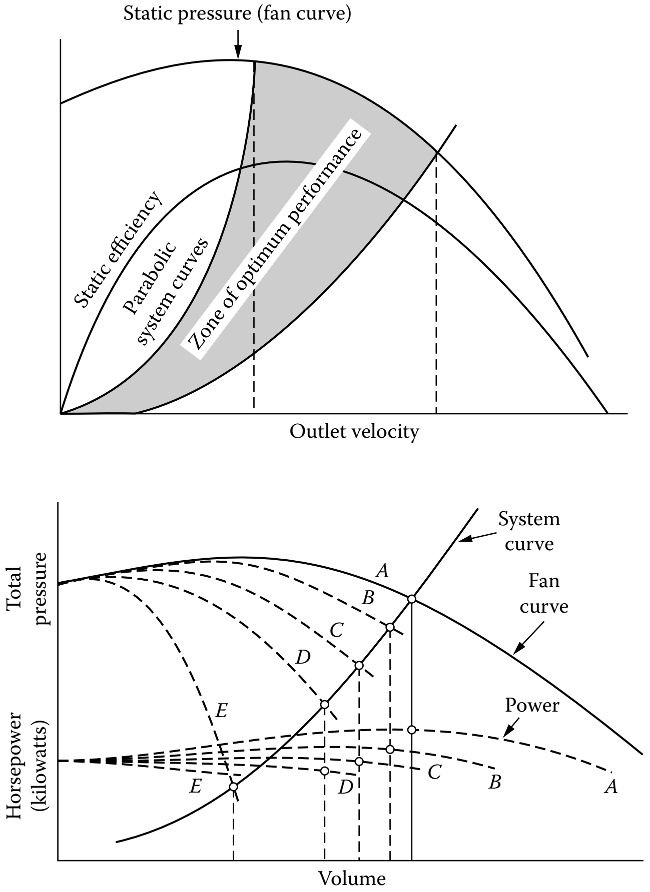


FIG. 8.25b
Upper curve shows optimum operating zone for fans; lower curves describe the effect of suction vane throttling on the fan curve.[1]

When the flow and pressure requirements of the process correspond to a point that is below the curve of the constant speed fan (points *B* to *E* in Figure 8.25b), one way to bring them together is to introduce an artificial source of pressure drop. In this figure, that source of pressure drop is a suction damper (a vane), which, as it is throttled, modifies the fan curve (from point *A* to *E*).

While the introduction of this artificial pressure drop does result in the wasting of transportation energy, its advantage is that this throttling shifts the surge point to the left and thereby, at lower flows, allows for stable operation.

Tube-axial fans are provided with adjustable pitch blades that permit the balancing of the fan to match the varying process load either automatically or by infrequent manual adjustments. Vane-axial fans are also available with controllable pitch blades (that is, pitch that can be varied while the fan is in operation) for use when frequent or continuous flow adjustment is needed. Throttling by varying the pitch angle retains high efficiencies over a wide range of conditions.

Figure 8.25c illustrates both the variable-blade pitch design and its performance curves. The efficiency is near maximum at zero blade pitch angle, and it drops off as the pitch angle is increased.

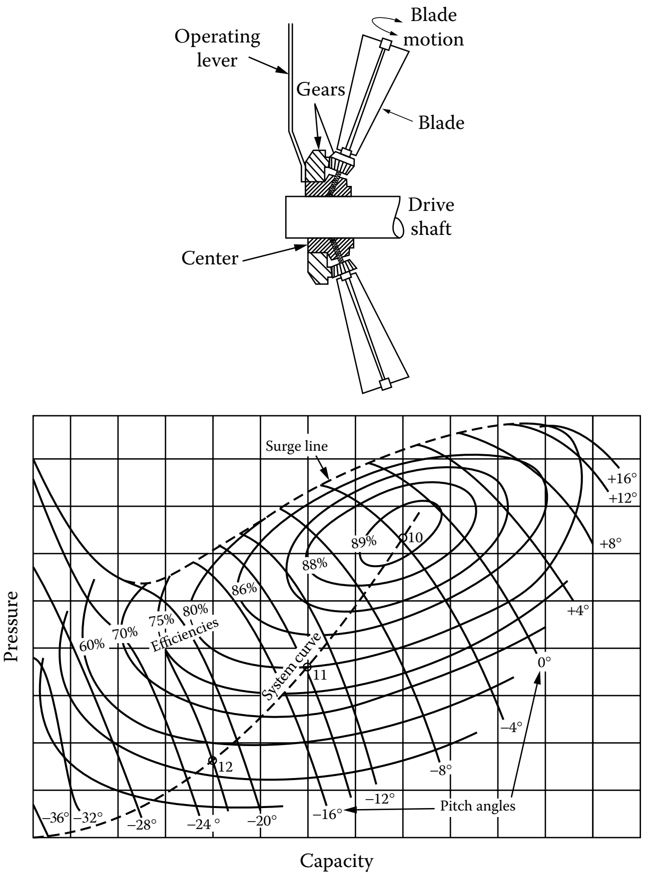


FIG. 8.25c
The design and characteristics of axial-flow fans with variable pitch control are described here. (Lower part is from Reference 2.)

From the standpoint of power consumption, the most desirable method of control is to vary the fan speed to match reduced process loads. If the load does not change too frequently, belt-driven fan drives can be considered. The speed in such designs is adjusted by changing the pulley on the drive motor of the fan. When the process load varies often or when continuous fan flow modulation is desired, electrical or hydraulic variable-speed motors are required.

Figure 8.25d illustrates both the fan curves and the power consumption of a variety of load controls on fans at partial loads. From the standpoint of noise, variable-speed is preferred to the variable-blade pitch design. On the other hand, both the variable-speed and the variable-pitch throttling designs are much quieter and more efficient than the discharge damper or suction vane throttling-type systems.

Safety Interlocks

All fans should be provided with safety interlocks, such as the ones illustrated in Figure 8.25e. In the illustrated design, interlock #1 will stop the fan if either excessively high pressure develops on its discharge side (PSH-02) or excessively high vacuums are detected on its suction side (PSL-03). Both of these pressure switches are protecting the ductwork from

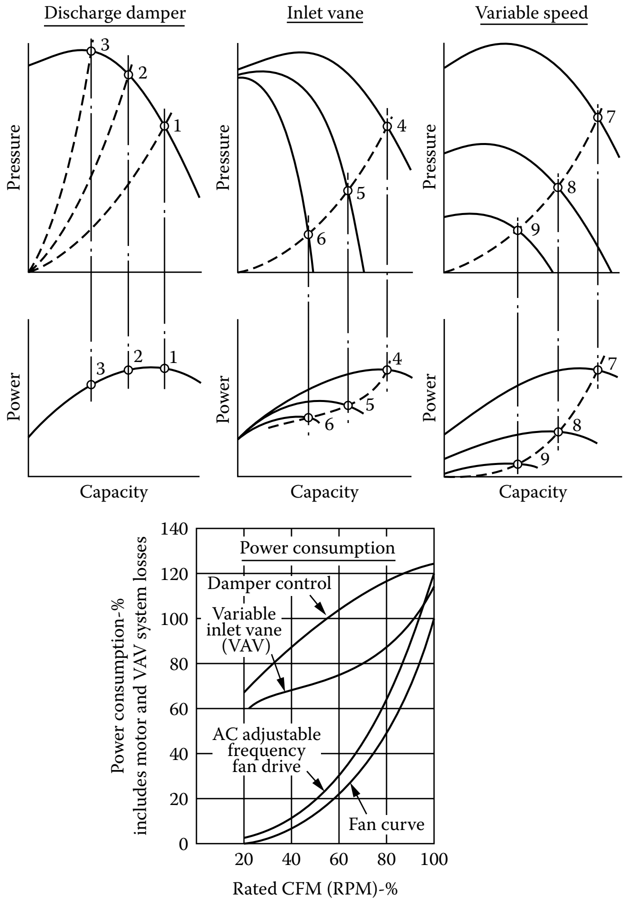


FIG. 8.25d
Turndown efficiency and energy conservation are directly related. (Upper part from Reference 2.)

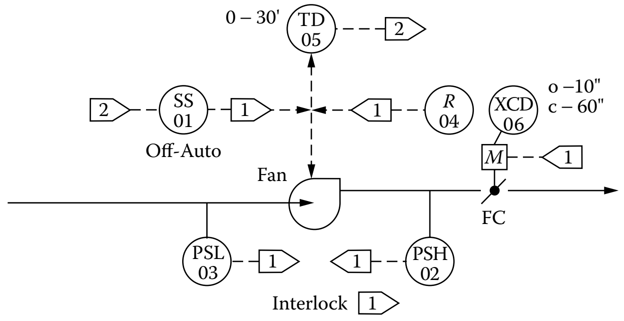
bursting (PSH) or from collapsing (PSL) due to the extreme pressure conditions. Both of these conditions can be caused by accidental blockage of the airflow.

Once the fan is shut down, it cannot be restarted until the reset button (R-04) is pressed. This gives the operator an opportunity to eliminate the cause of the abnormal pressure condition before restarting the fan. Whenever the fan is stopped, its discharge damper (XCD-06) is automatically closed to protect it from flow reversal, which can occur if several fans are connected in parallel. This discharge damper is designed for fast opening and slow closure, to make sure that it is always open when the fan is running.

The time delay (TD-05) guarantees that once the fan is started, it will run for a preset period, unless the safety interlocks turn it off. This protects the motor from overheating as a result of excessive on/off cycling. The fan cycling interlock (#2) is described below, in the paragraph titled Optimizing Multiple Fans.

Fan Controls in HVAC Applications

Air-handler controls are discussed in detail in Section 8.2. Therefore, here only the highlights will be mentioned of a fan configuration where supply fan(s) is/are transporting heat or cooling by sending conditioning air into a number of spaces (users), while return air fan(s) bring most of that air back for reheating or recooling. Figure 8.25f shows the fan controls of a variable-volume HVAC system.



Conditions				*Actions*		
SS-01	*PSH-02*	*PSL-03*	*R-04 (Reset)*	*Fan*	*XCD-06 (Damper)*	*TD-05 (Note #1)*
Off	—	—	—	Stop	Close	Time
Auto-On	High	—	—	Stop	Close	Time
		Low	—	Stop	Close	Time
	Not high	Not low	Not reset	Stop	Close	Time
			Reset	Run	Open	Reset to start

— = Any state or condition (don't care)
Note #1: Shut-down delay used in interlock 2

FIG. 8.25e
Illustration of the functioning of safety interlocks, which evaluate the state of suction and discharge pressures plus reset button status and, based on them, determine the fan, damper, and time delay status.

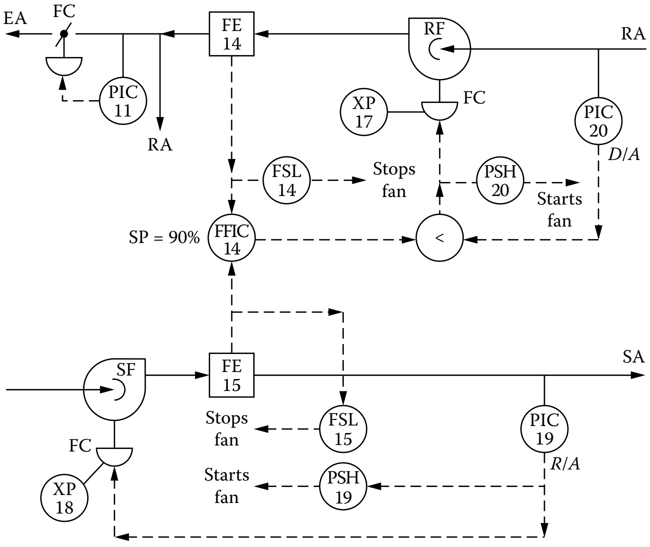


FIG. 8.25f
Variable-volume fan controls operate as shown here.

This configuration minimizes the operating cost by automatically maintaining the pressurization loss in the building at 10%, which is the usual minimum fresh air requirement in occupied spaces. This is achieved by setting the flow ratio controller FFIC-14 at 90%, meaning that the return fan station is modulated to return 90% of the air supplied to the zones.

PIC-20 is provided for dynamic balancing of the system. This is needed because the conditioned zones represent a rather large capacity, and therefore a change in supply airflow does not immediately result in a change in return flow. By overriding FFIC-14, PIC-20 prevents this ratio controller from increasing the return airflow faster than required and thereby prevents cycling or collapsing the suction ductwork by developing excessive levels of vacuum.

Parallel Fan Balancing

When two fans are operated in parallel, serving the same load, and one of them is idle, its damper should be closed in order to prevent recirculation from the operating fan. When one of the two fans is shut down or started up, it is important to simultaneously correct the damper position on the operating fan so that the supply pressure to the process will not be upset.

This goal can be achieved by installing a flow cascade loop as shown in Figure 8.25g. Here, the cascade master (FRC-104) compares the actual airflow being sent to the process with the airflow required, and if there is a difference, it changes its output signal, which is the set point to the cascade slave (FRC-103).

The measurement of FRC-103 is the sum of the damper position signals. Therefore, when one of the fans is shut down (and its damper is closed), the sum of the positions signal drops, while the set point of FRC-103 stays unaltered. Consequently, FRC-103 will quickly open the damper on the fan that remains in operation in order to keep the system balanced.

Optimizing Multiple Fans

The fan station is optimized when it is meeting the process demand for air or other gases at the lowest possible cost. The minimum cost of operation is achieved by first minimizing the number of fans that remain in operation and then by reducing the power consumption of these operating fans to the minimum. Figure 8.25h illustrates a control system that fulfills both of these goals. The lower part of the figure describes the control loops, while the operating fan curves are shown in the upper left and the interlock table in the upper right.

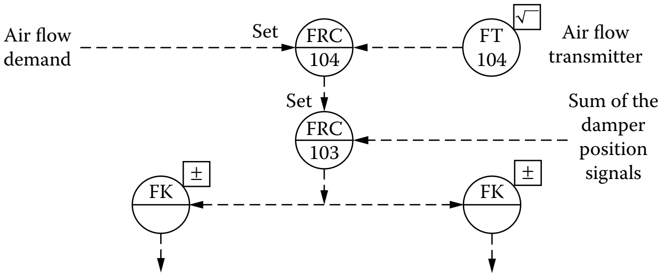


FIG. 8.25g
Flow to valve position cascade loop provides balancing controls for parallel fans.

The second fan is started and stopped in response to load variations according to the logic in the interlock table in the upper right of the figure. As the process demand increases, the speed of the operating fan (fan #1) is gradually increased. When its maximum speed is reached (point *A* on the fan curve) PSH-07 activates interlock #2 to start the second fan.

If the speed control signal were left unchanged when the second fan was started, an upset would occur, because the operating point instantaneously jumps from point *A* to *C*. In order to eliminate this temporary surge in discharge pressure (which otherwise could shut down the station), PY-07 is introduced. This is a signal generator, which, when actuated by interlock #2, drops its output to a value "*x*." The value of "*x*" corresponds to the required speed for the two-fan operation at point *A*.

The low-signal selector PY-09 immediately selects this signal *x* for control, and thereby, the upset is avoided. After actuation, the output signal of PY-07 slowly rises to full scale. As soon as it rises above the output of PIC-10, the low-signal selector (PY-09) disregards it and control is returned to PIC-10.

Once both fans are smoothly operating, the next control task is to stop the second fan when the load drops to the point where a single fan can meet it. This sequence is controlled by the low-flow switch FSL-08, which is set at 90% of the capacity of one fan (point *B* on the system curve). When the flow drops below the setting of FSL-08, interlock #2 is actuated to stop the second fan.

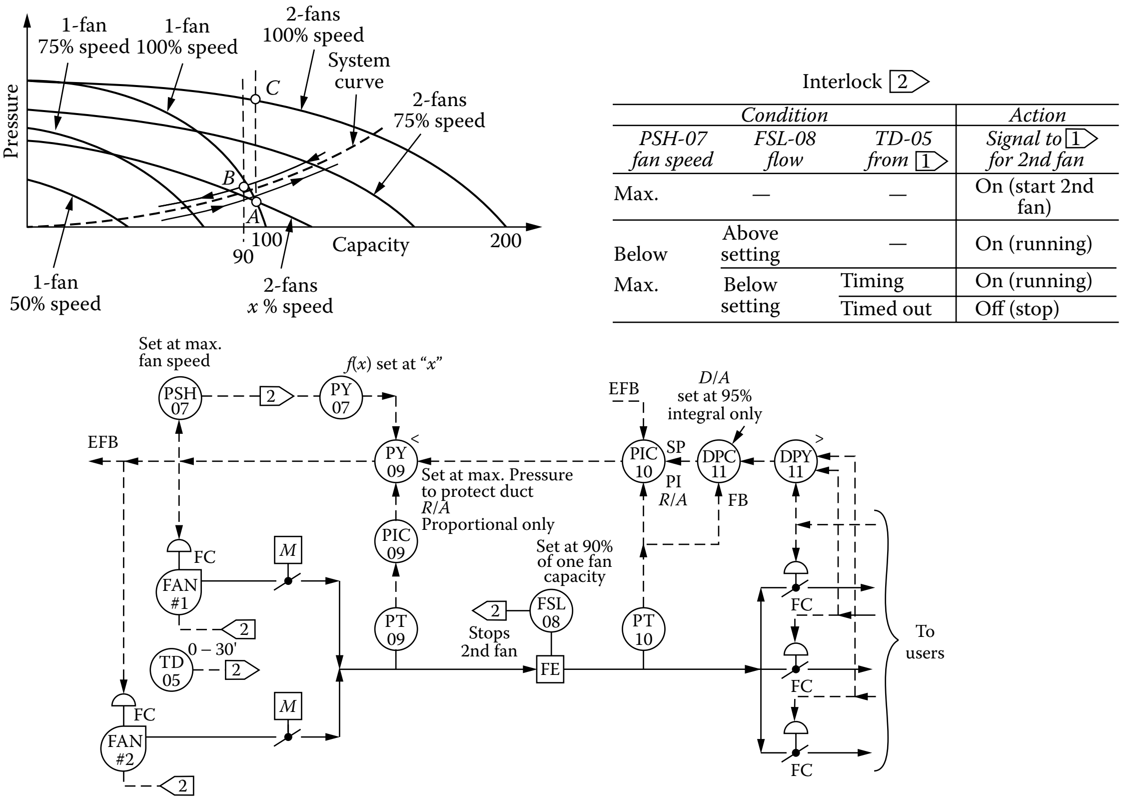
The stopping of the second fan is delayed until time delay TD-05 times out. This shutdown delay protects the fan from overheating due to excessive on/off cycling. Therefore, if TD-05 is set for, say, 20 min, the fan cannot be started more than three times an hour.

The fan cycling controls described here can be used to cycle any number of parallel fans. For each additional fan, another FSL-08 and PY-07 needs to be added. If n is the number of fans in operation, then FSL-08 is to be set for 90% of the capacity of $(n - 1)$ fans and PY-07 is to be set for "*x*" corresponding to the required speed of $(n + 1)$ fans at point *A*.

Optimizing the Discharge Pressure

The optimum discharge pressure is the minimum pressure that is still sufficient to satisfy all the users. This minimum pressure is found by observing the opening of the most-open damper. If even the most-open damper is not fully open, the supply pressure can be safely lowered, whereas if the most-open damper is fully open, the supply pressure must be raised. This supply-demand matching strategy not only minimizes the use of fan power but also protects the users from being undersupplied.

The optimization loop functions as follows: DPY-11 selects the opening of the most-open damper and sends its



Interlock [2>

Condition			Action
PSH-07 fan speed	*FSL-08 flow*	*TD-05 from* [1>	*Signal to* [1> *for 2nd fan*
Max.	—	—	On (start 2nd fan)
Below Max.	Above setting	—	On (running)
	Below setting	Timing	On (running)
		Timed out	Off (stop)

FIG. 8.25h

Optimized fan station controls include automatic fan cycling and a damper position to discharge pressure cascade loop to keep the most-open user supply damper at 90% opening.

opening as the measurement signal to the damper position controller DPC-11. If that signal is under 95%, this controller lowers the supply pressure set point, which in turn reduces the fan speed and opens all the load supply dampers until one has opened to 95%. Opening up all the dampers lowers the resistance to flow and saves fan power.

DPC-11 is an integral-only controller with an integral time (minutes/repeat), which is set to be 10 times that of the integral setting of PIC-10. This guarantees smooth, stable control, even if the dampers are unstable or cycling. The external feedback signal (EFB) protects the damper optimizer controller (DPC-11) from reset windup when PIC-10 is switched from the cascade configuration to manual set point.

PIC-10 is the load-following controller. It compares the optimized set point with the actual header pressure and adjusts the fan speed(s) or the blade pitch angle(s) of the fan(s). When its output signal reaches the maximum set on PSH-07, it starts another fan.

The role of PIC-09 is to provide overpressure protection at the fan discharge. Under normal conditions the fan discharge pressure is much below the set point of PIC-09, and its output is saturated at its maximum value. Therefore, under normal conditions PY-09 will select the output signal from PIC-10 for control. When the pressure limit set on PIC-09 is reached, it takes over control from PIC-10 and protects the ductwork from being damaged.

CONCLUSIONS

Fan optimization is an effective means to lowering the operating costs of plant operation. As can be seen from Figure 8.25d, if the average load is 60% of full capacity, the above-described optimization strategies can reduce the yearly operating cost to less than 50%.

References

1. American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE), *ASHRAE Handbook,* Equipment Volume, Atlanta, GA: ASHRAE.
2. Buameister, T. (ed.), *Mark's Standard Handbook for Mechanical Engineers,* New York: McGraw-Hill, 2004.

Bibliography

ASHRAE Standard 62-2001, "Ventilation for Acceptable IAQ," Atlanta, GA: ASHRAE, 2001.

Avery, G., “VAV Economizer Cycle,” *Heating, Piping, Air-Conditioning,* August 1984.

Bentley, D. E., *Machinery Protection Systems for Various Types of Rotating Equipment,* Minden, NV: Bentley Nevada Corp., 1980.

Cooper, F. G., “Low Pressure Sensing and Control,” *InTech,* September 1992.

Haines, R. W. and Wilson, C. L., *“HVAC Systems Design Handbook,* New York: McGraw-Hill, 2003.

Kovach, E. G., “Technology of Efficient Energy Utilization: The Report of a NATO Science Committee Conference Held at Les Ares, France, 8–12 October, 1973,” Brussels, Belgium: Scientific Affairs Division, NATO, 1973.

Kusuda, T., “Intermittent Ventilation for Energy Conservation,” NRS report of ASHRAE Symposium in Dallas, TX, February 1976.

Luciano, J. R., “Energy Conservation Techniques for Hospital Operating Rooms,” *ASHRAE Journal,* May 1983.

Nordeen, H., “Control of Ventilation Air in Energy Efficient Systems,” ASHRAE Symposium #20, Paper #3.

Roberts, W. B. and Rogers, W., “Turbine Engine Fuel Conservation by Fan and Compressor Profile Control,” Symposium on Commercial Aviation Energy Conservation Strategies, National Technical Information Service Document PC A 16/MF A01, April 2, 1981, pp. 231–256.

Romero, C., Sarunac, N., and Levy, E., “Sootblowing Optimization in Coal-Fired Boilers,” Lehigh University Energy Research Center Energy Liaison Program Annual Meeting, 2001.

Rosaler, R. C., *HVAC Maintenance and Operations Handbook,* New York: McGraw-Hill, 1997.

Shih, J. Y., “Energy Conservation and Building Automation,” ASHRAE Paper #2351.

Spielvogel, L. G., “Exploding some Myths about Building Energy Use,” *Architectural Record,* February 1976.

8.26 Fuel Cell Controls

A. C. GIBSON (2005)

Types of Fuel Cells:	A. PAFC: Phosphoric Acid Fuel Cell B. PEMFC: Proton Exchange Membrane Fuel Cell/Polymer Electrolyte Membrane Fuel Cell C. DMFC: Direct Methanol Conversion Fuel Cell D. AFC: Alkaline Fuel Cell E. MCFC: Molten Carbonate Fuel Cell F. SOFC: Solid Oxide Fuel Cell G. ZAFC: Zinc Air Fuel Cells H. MFC: Microbiological Fuel Cell
Rangeability:	4 to 1 turndown is common in reforming fuel cells, 10 to 1 is achievable in low-temperature hydrogen fueled cells.
Efficiencies:	Fuel cell peak efficiencies range from 85% for high-temperature solid oxide fuel cells to below 20% for direct conversion methanol fuel cells. At low-power outputs efficiencies drop dramatically in all fuel cells.
Materials of Construction:	Materials of construction vary with the cell design
Costs:	Costs for fuel cell systems range around $100 per watt at this time (2004). Cost reductions are critical to the transition from laboratory curiosity to commercially viable systems.
Partial List of Suppliers:	There are a multitude of fuel cell manufacturers throughout the world, each of which tends to specialize in a specific design type; the volatility in the market leads to a rapid turnover of Web site links. For a survey of fuel cell manufacturers, with access to catalogs and mechanical details the Web offers several excellent sources, such as Ballard Power Systems Inc., www.ballard.com CFCL Ltd., www.cfcl.com.au E-TEK, www.etek-inc.com, www.denora.it Fuel Cell Today, www.fuelcelltoday.com Fuel Cells 2000, www.fuelcells.org FuelCell Info.Com, www.fuelcell-info.com FuelCell Energy, www.ercc.com Forschungszentrum Julich, www.fuelcells.de Honeywell, www.honeywell.com Plug Power, www.plugpower.com U.S. Fuel Cell Council, www.usfcc.com UTC Fuel Cells, www.utcfuelcells.com Westinghouse Electric Company, www.westinghouse.com

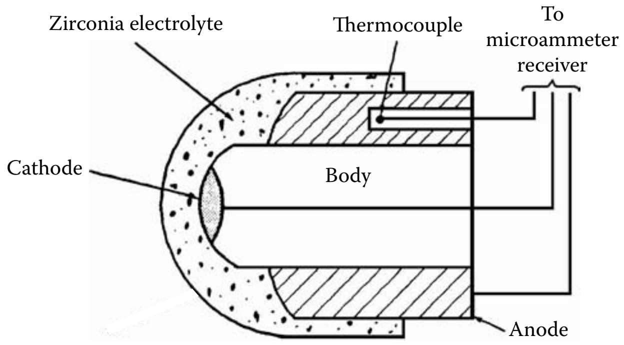


FIG. 8.26a
Galvanic cell design. Note: External heater not shown.

INTRODUCTION

Fuel cells are a means of converting chemical energy into electrical energy and, hence, to work. The term *fuel cell* describes a type of primary galvanic cell[1] using replaceable anode and cathode reagents. Conventional primary cells are limited by the ability to store reagents on the current-collecting plates. Fuel cells bypass this limitation by providing a continuous feed from an external reservoir.

In theory, almost 100% conversion efficiency is possible, as the energy conversion is via electrochemical reactions, bypassing the losses inherent in thermal conversion. In practice, however, other limitations, due to electrical losses and chemical concentration gradients, reduce the achievable performance to between 30 and 85% efficiency.

Any gaseous or liquid oxidizer and fuel can be used in a fuel cell system, but the preferred oxidizer is oxygen, as the exhaust in that case is relatively innocuous water and carbon dioxide.

The justifications for the use and development of fuel cell systems include a number of grounds. The most common is high efficiency, but low-temperature fuel cells are no more efficient than conventional steam systems and less efficient than some combined cycle systems that are now in service. However, fuel cell systems achieve this performance on a much smaller scale.

Efficiency and Pollution

An even more important benefit is the low emission achievable by the positive control of the reduction oxidation reactions, which eliminates some of the more undesirable pollutants from the exhaust streams. Fuel cells are also suitable for special-purpose applications where closed cycle operation is desired.

As the low-temperature fuel cell systems operate at temperatures too low for nitrogen oxides to form and because high-temperature fuel cells prevent the contact with nitrogen in atmospheric air (at least as far as the exhaust handling is considered), nitrogen oxide pollution is not a problem, and consequently ozone pollution from these units is also reduced.

Sulfur oxide emissions are also low, but it is still preferred to remove the sulfur at the source rather than from the exhaust. This is not an inherent advantage of fuel cells but rather an operational requirement and restriction for reforming fuel cells, as sulfur will poison the catalyst used in the process.

In the case of hydrogen-powered fuel cells, the above-mentioned advantages must be balanced with the consequences of the hydrogen production method used. If it is made by reforming of natural gas, that is a rather wasteful process if the carbon dioxide is vented and, therefore, the energy content of the carbon is not utilized.

Electrolytic[2] systems are cleaner but their power consumption and power generation are essentially the same (somewhat less), so unless renewable energy sources are used, their net efficiency and pollution reduction are very poor. A further problem with hydrogen fuel cells is losses while the hydrogen is in storage. Losses of up to 30% per month can be expected in automotive applications, because of leakage.

The alternative mode of operation is that of the fuel cell-operated storage battery. In this case, the fuel cell is run in reverse to generate hydrogen from the electrolysis of water. In this mode of operation, the fuel cell functions as a closed system and provides a battery with greatly increased storage capacity compared to batteries that store energy in solution or as a solid on the electrodes.

Low-temperature polymer electrolyte fuel cells are reliable in the regenerative mode. Solid oxide fuel cells have also shown promise of high efficiency due to a combination of their superior performance of the basic technology and their reduced cell voltage compared to the low-temperature fuel cells.

High-temperature solid oxide cells do not require the catalytic coatings used for the reforming fuel cell designs. Low-temperature polymer fuel cells are in service as oxygen generators in submarines around the world.

Fuel cell-based energy storage is not particularly efficient, and other options may be more effective on an industrial scale. For closed cycle applications, such as spacecraft and submarine vehicles, however, the combination of light weight, reasonable power density, and compact size of fuel

---

[1] In a galvanic cell (Figure 8.26a), no external voltage is applied to the cell, which consists of an electrolyte and two electrodes. For example, if the cathode is a noble metal, the anode is a base metal, and if the electrolyte is a zirconia oxide solution, which contains dissolved oxygen.

For the reactions that take place on the solid oxide fuel cell (SOFC) electrodes, refer to Equations 8.26(15) to 8.26(19).

As the cathode reduces an oxygen molecule into hydroxide, it releases four electrons, which causes a current flow. This flow is in proportion to the oxygen concentration in the electrolyte. As the current flows, the analyte diffuses and the cell is depolarized. The electrodes are consumed during this operation and require periodic replacement.

The minimum operating temperature is 752°F (400°C), the typical operating temperature is 1112°F (600°C) and the SOFC operating temperature is 1472°F to 1742°F (800°C to 950°C).

[2] An electrolytic cell is similar to a galvanic one, but it operates with an external voltage connected to its electrodes, and its electrodes are not consumed.

cells compares favorably to the equivalent battery-based systems with similar storage capacity.

An unfortunate aspect of the presently existing fuel cell systems is the difficulty in scaling up their designs. Often, the doubling of their power output also tends to double their complexity.

Historical Perspective

The oldest form of a fuel cell is the low-temperature Groves "gas battery" cell.[3] This cell was developed by Sir William Groves in 1839 and was based on the research work on water electrolysis by William Nicholson and Anthony Carlisle in 1800. It used fine-wire platinum electrodes encapsulated in glass tubes to demonstrate a reversible reaction. When a power supply was connected to the electrodes, this cell generated hydrogen and oxygen and then it became able to generate a current. The fact that it worked at all was an accident, resulting from the fact that Groves used the catalytic metal, platinum, for current collectors. Yet it formed the basis for what was to come.

The first attempt at a commercial fuel cell was a coal gas-driven fuel cell developed by Ludwig Mond and Charles Langer in 1889. This unit used a sulfuric acid electrolyte impregnated into a ceramic separator plate with perforated platinum electrodes. Coal gas is a mixture of hydrogen, methane, and carbon monoxide, and it is produced by injecting steam into a bed of coke at high temperatures and then scrubbing before being used. This 1889 coal gas-based fuel cell demonstrated that much more research was required before a viable product could be produced. This was the case because carbon monoxide rapidly poisons the platinum catalysts at low temperatures by binding to their surfaces. The other limitation was the need to remove the water generated by the oxidation reaction.

A theoretical basis for the fuel cells was provided in 1893 by Friedrich Wilhem Ostwald, who was a pioneer of electrochemistry. Based on experiments, he provided a solid basis for the further development in this field, because he determined the relationships between the electrodes, electrolytes, reducing and oxidizing agents, and anions and cations.

This was further investigated by Walter Herrmann Nernst (who in 1897 demonstrated an early yttrium-doped zirconia device in the form of the Nernst Lamp) and provided the theoretical basis for current research.

High-Temperature Designs

High-temperature fuel cells were first investigated by Emil Baur in Switzerland at Braunschweig and Zurich, up to the World War II, using metal oxide fuel cells with liquid silver electrodes. Their experiments with zirconium, yttria, cerium, lanthanum, and tungsten oxides laid the foundations for the current work on solid oxide fuel cells but were frustrated by problems caused by unwanted reactions that occurred with the calcium dopants. This led to a redirection of the development efforts towards molten salt fuel cells.

In parallel with Baur's experiments, Francis Thomas Bacon in 1939 was experimenting with nickel electrodes in high-pressure fuel cells. He used pressures above 200 atmospheres in demonstration fuel cells intended for submarine use. This led, in 1958, to the development of the alkali fuel cell stacks, which are capable of generating significant power levels.

This technology was the basis of the molten potassium hydroxide fuel cells built by Pratt and Whitney (now UTC Fuel Cells) in the mid-1960s for the Deep Submergence Rescue Vehicles of the military. Their "black" project is a counterpart of the Deep Submergence Search Vehicles project and the later of the Apollo spacecraft fuel cells. This in turn led to Lockheed's Deep Quest and to the current space shuttle systems.

The required high operating temperatures of phosphoric acid and of potassium hydroxide fuel cells made black start operation difficult. Another limitation was that in order to function reliably, these cells require exceptionally pure, carbon dioxide-free hydrogen fuel.

Polymer Technology

Polymer technology permitted General Electric to develop the next step in the development of the proton exchange membrane cells for the Gemini program. The benefits of high power-to-weight ratios outweighed the high costs of developing an essentially experimental system to an acceptable level of reliability. The project was helped by NASA's "cost is no object" philosophy of the time, and by the short operational durations that were dictated by the need to complete the dictated missions.

Even so, failures of the fuel cells curtailed all of the first flights in which they were used, from Gemini GT-2 to Gemini 7 (or over half the flights). The problems were finally ironed out, only by then the Apollo program had decided to switch to the older alkaline technology.

Therefore, anyone who is moving into the field of fuel cell development should not underestimate the difficulties, as a similar number of unexpected problems seem to occur in connection with each new design. New developments are in progress in a number of commercial organizations, with Ballard being the most commercially significant.

High cost and low reliability have dogged the implementation of fuel cells and still limit the existing projects from producing a commercially viable power generation system. Unjustified optimism has brought down more fuel cell devel-

---

[3] A battery is a device that converts chemical energy into electrical energy. A battery consists of two dissimilar substances (the two electrodes) and an electrolyte that acts chemically on the electrodes and functions as an ionic conductor for the transfer of electrons between the electrodes. The dry-cell battery of a flashlight consists of carbon and zinc. Wet-cell batteries are rechargeable and use liquid electrolytes such as a sulfuric acid solution, with the negative electrode being lead and the positive being lead coated by lead oxide.

The standard dry cell is manganese dioxide and zinc with a carbon electrode current collector and ammonium chloride electrolyte.

opment programs than any other cause. Adequate design and development budgets are also critical, and the costs are high.

PERFORMANCE AND DESIGNS

Electrical Performance

The potential voltage delivered by a fuel cell can be predicted by the Nernst equation:

$$E^N = E^0 - \frac{RT}{zF}\ln\left(\frac{a(\text{Reduction})}{a(\text{Oxidation})}\right) \qquad \textbf{8.26(1)}$$

E^N = normalized voltage (cell open circuit voltage)
E^0 = theoretical reversible voltage at reference conditions (°K)
R = universal gas constant = 8.314 472 (±0.000 015) Joules/mole/Kelvin = 8.314 472 · 10^7 erg/mole/Kelvin
T = absolute temperature in degrees Kelvin
F = Faraday constant = 96 485.3383(±0.0083) Coulomb mol^{-1}
Z = charge number of the electrode reaction = the number of moles of electrons involved in the reaction as written
$a(Reduction)$ = chemical activities on the reduction side of the equation (for gases, the partial pressure)
$a(Oxidation)$ = chemical activities on the oxidation side of the equation (for gases, the partial pressure)

Or as applied to common fuel cell applications:

$$E^N = E^0 - \frac{RT}{2F}\ln\left(\frac{[H_2O]}{[H_2][O_2]^{1/2}}\right) \text{ with } E^0 = 1.229\text{ V} \qquad \textbf{8.26(2)}$$

As can be seen from Equation 8.26(2) the cell voltage will be decreased by an increase in temperature and will be reduced as the reagents are consumed (if not replaced). Maintaining a high partial pressure of reagents will also improve the cell's stability as the load increases. Contamination with waste products will reduce the cell voltage, as will inert diluents that are not purged.

The cell voltage achieved at the load inputs is further reduced by the resistance of the electrolyte and by the resistance of the electrodes and connecting cabling. At high current rates and high temperatures, these losses caused by a high-temperature nickel cable can be significant.

The actual delivered voltages typically range from 1 to 0.6 V per cell. This voltage is a very good measure of cell efficiency, if the actual achieved output is compared with the one predicted by the Nernst Voltage.

In theory, the fuel cell fuel consumption can be predicted by the relationship that 1 mole of hydrogen gas yields 2 moles of electrons when ionized. This is equivalent to a charge of 2 × 96,485.3383 (±0.0083) Coulomb/mol or amp/sec/mol.

At 0°C 1 mole will fill 22.72 liters of space at standard atmospheric pressure. So, for a current draw of 1 amp the fuel cell will need to be receiving 0.1177 Ncm^3 per sec (7.062 Ncm^3 per min) at 100% efficiency. Considering the actual efficiencies of fuel cells, at least double this value (0.2355 cm^3) would, however, be more typical. This gas requirement does not seem to be much, but because each cell in a stack is consuming this much gas, the total volumetric feed requirement of the system rises rapidly.

Functional Requirements

Fuel cells are a means of converting chemical energy into electrical energy and, hence, to work. In theory, because fuel cells convert the energy in chemical bonds to electrical energy by means of a direct redox reaction, efficiencies should be able to approach 100% with minimal generation of unwanted reaction products.

Noise Similarly, because the basic fuel cell needs no mechanical drive, its operation is quiet and involves no frictional losses. These characteristics should make it acceptable to locate the generator near the final user, producing a more even distribution of the generation capacity.

These expectations necessitate that quiet-running auxiliaries be used to transport the reagents, particularly the fans and blowers that are used to transport the oxidizer, usually air (and for transporting and compressing the fuel in case of low pressure supplies, such as waste gas). In order to minimize the noise generated by the electrical systems, they should be well-potted to prevent their motion, which can be induced electrostatically or electromagnetically.

Blower noise can be reduced by providing variable-speed drives for the blowers and by careful duct design. In addition, the feathering of the blade edges and the use of noise-reducing enclosures are recommended, but should be so designed that it will not result in heat problems by overheating the blower drive.

Variable-speed control is also desirable, because it helps reduce the parasitic losses, if the drive is powered directly from the DC bus of the fuel cell, by avoiding conversion losses, but this requires a cell bus voltage consistent with the VSD AC bus voltage.

To reduce the electrically generated mechanical noise, the inductor and transformer windings should be carefully designed and the AC bus bars should be carefully secured. Some inverters that are used in conventional variable-speed drives can be exceptionally loud, as can the output filters needed for hazardous area certification.

Size and Reliability Fuel cell system size is important for mobile applications, and volumetric efficiency is even more critical. The volume of a 40 kW system today is about 10 m^3 with all its auxiliaries. This compares with 2 m^3 for a similar capacity diesel generator. Therefore, one goal of fuel cell research is to increase the power densities both in terms of volume and area of the unit.

Current state-of-the-art fuel cells are finally approaching the reliability levels that are acceptable for some applications. On the other hand, they are yet to live up to the expectation of a financially viable power generation, because of their low power efficiencies.

The cost per installed watt is still one to two orders of magnitude higher than that of the conventional technologies, and significant generation benefits are required to justify this added capital cost. The theoretical benefits include lower emissions, lower noise levels, and lower operating costs.

In reality, these benefits are undermined by the remaining unsolved problems. The advantage of lower emissions is undermined by the problems involved in generating and storing the hydrogen gas needed for the simplest (and so far most efficient) fuel cells. The advantage of low noise levels is limited by the limited availability and higher cost of the blowers. The advantages of higher power densities and lower operating costs are limited by the life expectations of both the control hardware and of fuel cell stacks, which today still result in low availability.

Recent U.S. military experience with phosphoric acid fuel cells resulted in under 1800 hours mean time between failure (MTBF) with an availability of 67%. This is comparable to the overall service intervals for diesel generators, and one fifth of the service interval for a typical gas turbine generation set.

All of these areas require further efforts by the developers to achieve a commercially viable product. Today, the typical system still requires servicing for the replacement of its scrubber packs every 3 to 4 days.

FUEL CELL TYPES AND FEATURES

Fuel cells fall into a number of distinct strains, each with its own advantages and disadvantages. Each cell type has a preferred feed stock (fuel) and operating regime, and in each of them a specific set of reactions take place, which affects the optimum operating conditions.

Fuel Cells Reactions

Low Temperature (Proton Exchange Membrane and Phosphoric Acid)

Anode Reaction: $2H_2 => 4H^+ + 4e^-$ **8.26(3)**

Cathode Reaction: $O_2 + 4H^+ + 4e^- => 2H_2O$ **8.26(4)**

Overall Cell Reaction: $2H_2 + O_2 => 2H_2O$ **8.26(5)**

(Note: These PEMFC cells are easily poisoned by carbon monoxide, which can form a complex on the surface of the platinum electrodes, because the fuel cell operating temperature is not high enough to oxidize it into carbon dioxide.)

Direct Methanol Conversion

Anode Reaction:

$2CH_3OH + 2H_2O => 2CO_2 + 12H^+ + 12e^-$ **8.26(6)**

Cathode Reaction: $3O_2 + 12H^+ + 12e^- => 6H_2O$ **8.26(7)**

Overall Cell Reaction:

$2CH_3OH + 3O_2 => 2CO_2 + 4H_2O$ **8.26(8)**

(Note: These cells have the same problems as the PEMFCs as far as carbon monoxide poisoning is concerned. They also have a limited ability to handle their own waste products, which leads to low efficiency and poor utilization of their fuel.)

Alkaline

Anode Reaction: $2H_2 + 4OH^- => 4H_2O + 4e^-$ **8.26(9)**

Cathode Reaction: $O_2 + 2H_2O + 4e^- => 4OH^-$ **8.26(10)**

Overall Cell Reaction: $2H_2 + O_2 => 2H_2O$ **8.26(11)**

(Note: This fuel cell design can be poisoned by carbon dioxide and carbon monoxide. Therefore, it is limited to fuels such as high purity gas or electrolytic hydrogen. It cannot use reformate from natural gas unless highly refined. If ambient air is used as oxidizer, it must be scrubbed.)

Molten Carbonate

Anode Reaction:

$2CO_3^{2-} + 2H_2 => 2H_2O + 2CO_2 + 4e^-$ **8.26(12)**

Cathode Reaction: $2CO_2 + O_2 + 4e^- => 2CO_3^{2-}$ **8.26(13)**

Overall Cell Reaction:

$4H_2(g) + O_2(g) + 2CO_2 => 2H_2O(g) + 2CO_2$ **8.26(14)**

(Note: This fuel cell requires a feed of highly concentrated carbon dioxide. This carbon dioxide can be recovered from the exhaust if the effort is felt worthwhile.)

Solid Oxide

Anode Reaction: $4H_2 + 2O^{2-} => 2H_2O + 4e^-$ **8.26(15)**

$2CO + 2O^{2-} => 2CO_2 + 4e^-$ **8.26(16)**

Cathode Reaction: $2O_2 + 4e^- => 2O^{2-}$ **8.26(17)**

Overall Cell Reaction: $2H_2 + O_2 => 2H_2O$ **8.26(18)**

$2CO + O_2 => 2CO_2$ **8.26(19)**

(Note: These solid oxide cells have some direct reforming ability when operating at their design conditions.)

TABLE 8.26b
Features and Characteristics of a Number of Fuel Cell Designs

	Fuel Cell Type	*PAFC*	*PEMFC*	*MBFC*	*DMFC*	*AFC*	*MCFC*	*SOFC*	*ZAFC*
Feed Component	H_2	Fuel	Fuel	Diluent	Fuel	Fuel	Fuel	Fuel	Fuel
	CO	Poison (0.5–1.5%)	Poison	Poison	Poison	Poison	Fuel (a) (d)	Fuel	Fuel
	CH_4	Diluent	Diluent	Fuel	Diluent	Poison	Diluent (b)	Fuel (a)	Fuel
	CH_3OH	Diluent	Diluent	Fuel	Fuel	Poison	Fuel	Fuel	Fuel
	C_xH_x	Diluent	Diluent	Fuel	Diluent	Poison	Diluent	Fuel (a)	Fuel
	CO_2	Diluent	Diluent		Diluent	Poison	Reagent	Diluent	Diluent
	H_2O	Diluent	Reagent (membrane needs water)	Reagent (membrane needs water)	Reagent	Reagent	Reagent	Reagent	Diluent
	Sulfur (H_2S, COS)	Poison (@ 50 ppm)	No information	Fuel*	No information	Poison (@ 50 ppm)	Poison (@ 0.5 ppm)	Poison (@ 0.5 ppm)	Poison (@ 1 ppm)
Charge Carrier		H^+	H^+	H^+	H^+	OH^-	CO_3^-	O^{2-}	OH^-
Efficiency		40–50%	40–50%	1–2%	20–40%	70%	60%	60–85%	
Operating Temp.		150–220°C	< 100°C	15–40°C	65–220°C	50–120°C	650°C	650–1000°C	700°C
Electrolyte		H_3PO_4	Teflon (inert)	Teflon (inert)	Teflon (inert)	Potassium hydroxide (KOH)	Lithium carbonate/ potassium carbonate or sodium carbonate	Zirconia/yttria	Zinc oxide
Electrodes	Anode	Platinum	Platinum/ carbon	Platinum/ carbon	Platinum/ ruthenium	Platinum	Nickel	Nickel/nickel oxide	Zinc anode
	Cathode	Platinum	Platinum/ carbon	Platinum/ carbon	Platinum	Platinum	Nickel	Manganese oxide/ vanadium oxide lanthanum/ strontium/tin doped	
Comments									

Notes:

(a) In reality, CO, with H_2O, shifts to H_2 and CO_2, and CH_4, with H_2O, reforms to H_2 and CO faster than reacting as a fuel at the electrode.

(b) A fuel in the internal reforming MCFC.

(c) Molten carbonate fuel cells, solid oxide fuel cells, and zinc air fuel cells all can consume complex hydrocarbons (C_xH_x).

(d) CO can strip nickel at high temperatures by forming $Ni(CO)_4$ gas.

(e) Microbiological fuel cells use bacteria to catalyze the conversion of carbohydrates to fuel, low efficiencies can be balanced against low fuel cost, including domestic sewage.

PAFC: phosphoric acid fuel cell
PEMFC: proton exchange membrane fuel cell/polymer electrolyte membrane fuel cell
MBFC: microbiological fuel cell
DMFC: direct methanol conversion fuel cell
AFC: alkaline fuel cell
MCFC: molten carbonate fuel cell
SOFC: solid oxide fuel cell
ZAFC: zinc air fuel cell

* The use of sulfur compounds was subject fo further testing.
Results are yet to be published.

Zinc Air

Anode Reactions:

$$CH_4 + H_2O \Rightarrow CO + 6H^+ + 6e^- \quad \textbf{8.26(20)}$$

$$2Zn + 2OH^- \Rightarrow 2ZnO + 2H^+ + 2e^- \quad \textbf{8.26(21)}$$

$$CO + ZnO \Rightarrow CO_2 + Zn \quad \textbf{8.26(22)}$$

$$\text{Cathode Reaction: } O_2 + 2H^+ + 2e^- \Rightarrow 2OH^- \quad \textbf{8.26(23)}$$

$$O_2 + 4H^+ + 4e^- \Rightarrow 2H_2O \quad \textbf{8.26(24)}$$

$$\text{Overall Cell Reaction: } CH_4 + 2O_2 \Rightarrow CO_2 + 2H_2O \quad \textbf{8.26(25)}$$

Fuel Cell Design Features

Fuel cells share a number of characteristics with heat exchangers. These include the fact that similar forces limit the transfer of heat as the ones that limit the transfer of ions. Also, the exchange of ions is dependent upon the available surface area in contact with the process fluids and by the resistance to transfer of energy through the separator material. The transfer of energy also depends upon the movement and mixing of the molecular components of the fluids on either side of the exchange surfaces.

A simple tubular design has been popular for solid oxide fuel cells, because it is easy to seal. On the other hand, because the surface area is a function of tube diameter, deducing the diameter of the tubes increases both the transfer area and the pressure drop through the smaller tubes. Also, it is hard to apply the current-collecting coatings to small-diameter tubes. For these reasons, tubular designs are very limited in terms of the ultimate achievable power densities.

Flat plate designs have, therefore, become the default standards for low-pressure services, as films are easy to fabricate, particularly if liquid or film electrolyte systems are involved.

In either design there are design limitations to the task of bringing turbulent reagent streams to the active surfaces. Reagent mixing can be enhanced by narrowing the passages between the electrodes or by placing obstructions into these streams to induce turbulence in the fluid flows. The downside of these solutions is that they both interfere with the fluid flow, and therefore they both increase pressure drop. In addition, if the dimensional tolerances cannot be accurately maintained or if the system model is not accurate, such obstructions can potentially interfere with the gas distribution between cells.

Sealing the flat-panel systems is not easy. Low-temperature systems can be sealed by elastomers or can have welded joints in the polymer membranes.

In high-temperature systems, a combination of thermal expansion and the poor adhesion of the conventional sealing materials must be considered. Solid oxide fuel cells have used metallic carriers to serve as both supports and current collectors, but thermal expansion differences between the ceramic electrolyte and metal carrier leads to leaks that are impossible to seal and can also shatter the cells. A newer solution is to use monolithic ceramic structures with ceramic glass seals. In such designs, the conductive glass links interconnect the cells into functioning stacks.

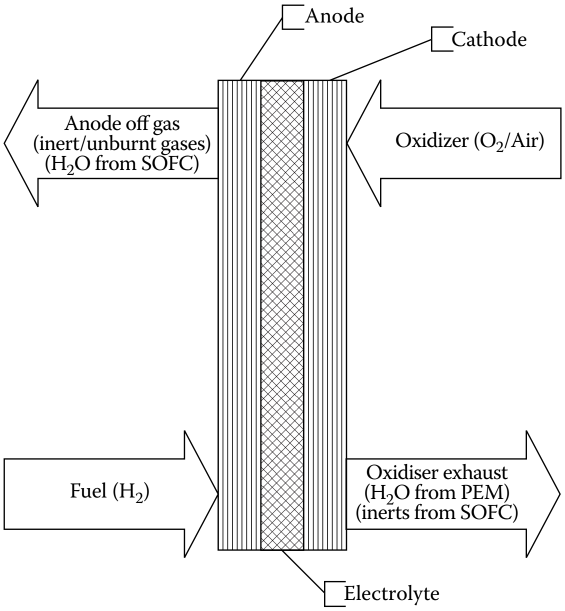


FIG. 8.26c
Illustration of one of the cells in a stack of solid oxide fuel cells (SOFCs) where the hydrogen fuel contacts the anode and the oxidizer acts on the anode surface. The overall reactions in this cell are described by Equations 8.28(18) and 8.26(19).

Single cells are rarely able to produce enough power as is required by commercial applications. The cells are therefore connected into batteries or stacks. Commercial stacks frequently have more than 100 and sometimes as many as 400 cells.

If the dimensional tolerances are not tightly controlled, a major limitation to balancing the fluid flows between the cells in the stack will result. Therefore, providing good control of the fuel and oxidizer flows in every cell (Figure 8.26c) increases both cost and size, which conflicts with the space constraints in commercial systems. If tight intercellular spacing is maintained to improve gas distribution and volumetric efficiency, small errors in the gaps, which can be due to the limitations in manufacturing tolerances, can produce large variations in the actual gas flows between the cells in a stack, which limits the effective stack efficiency as the most starved cell sets the current capacity.

FUEL CELL CONTROLS

Fuel cell control is conceptually simple. The fuel and oxidizer flows are regulated to ensure that the reagent flows match the required current generation, taking into account the cell efficiency factor.

At low levels of current generation and low volumetric efficiencies, simple diffusion is adequate to ensure that some current output is generated. This is the level of operation, achieved by the Groves gas battery, which is used by simple methanol cells. When higher power outputs are required, some type of control system is required to match the power generation with the power requirement of the process.

Oxidant and Fuel Flow Metering

In air breathing cells, the air flow can be maximized to ensure the availability of oxygen by maximizing its partial pressure. The partial pressure achieved is subject to maximum capacity limitations and to limitations imposed by losses due to heating in high-temperature cells and due to pumping losses in other cell types.

In the proton exchange-type cells this maximized air flow also sweeps the waste gases, principally water, away from the cathodes. In fuel cells running at high volumetric power levels, cold air bypass streams are provided to cool the cells if needed and to balance the temperatures between the cell stacks that are connected to a common manifold.

The fuel flows are generally metered more precisely, because of their generally higher costs, and are throttled to meet the current demand of the process being served. Fuel flow control in low-temperature cells is relatively simple. The control valves used are servomotor-driven quarter-turn valves (Chapter 6) and the flow sensors can be the type used on gas furnaces (Section 8.27).

In high-pressure systems, more expensive globe-type control valves (Section 6.19 in Chapter 6) are usually used with pneumatic or electric actuators.

Gas flow metering is relatively simple using thermal dispersion-type mass flow sensors (Figure 8.26d). These sensors provide reasonable accuracy (1–2% of full scale) and are relatively affordable (for details see Section 2.13 in Chapter 2 in the 4th edition of the first volume of this handbook). In the author's test, a bypass-type thermal dispersion meter with gas composition compensation was used (Yamatake CMS series).

The previously used flow sensor option in research fuel cells was the integrated circuitry-type differential sensors, which was limited by its low tolerance for water contamination. This was a serious limitation in reforming-type fuel cell applications, because there the presence of water is essential for the operation of the system, and no protection against water contamination was provided in the smaller and cheaper IC sensors.

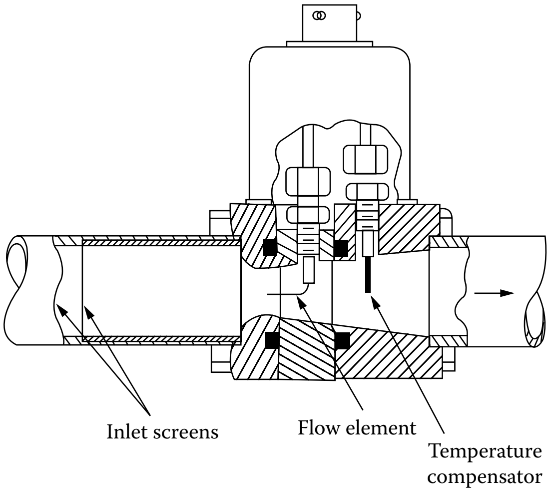


FIG. 8.26d
Venturi-type thermal mass flowmeter. (Courtesy TSI Inc.)

Spacecraft and High-Temperature Cell Applications The flow measurement tasks in spacecraft and similar closed circuit fuel cell applications are slightly different. This is because the oxygen content needs to be conserved with the same vigor as the hydrogen, and blowdowns are restricted to what is needed to maintain power output in the face of water fouling of the anode. As pure gases are used, partial pressure reductions are not a problem, other than for the production of wastewater. Therefore, constant supply pressures can be used, which greatly simplifies the control requirements, as the system becomes self-regulating on the demand side. In such applications, the waste can simply be blown down to a reservoir, based on time and current draw.

In high-temperature cells, special valve designs are required to handle the fuel gas, which is at up to 1000°C. These valves are used to balance the gas flows between stacks and for stack isolation to permit continued operation, when some of the stacks fail. In the automotive industry, for exhaust gas recirculation, such high-temperature valves are successfully operated with high-temperature servomotor drives. If these were available to fuel cell system integrators, they would solve many of the problems in high-temperature fuel controls.

Unfortunately, the major manufacturers of these devices (Mitsubishi and Hitachi) will not sell them to competing fuel cell manufacturers. This means that each fuel cell system designer has to reinvent the wheel each time a new system is designed, which is significantly raising the cost of complete systems.

Indirect Flow Detection In most fuel cells a measurement of the cell voltages (based on the stack voltages) will indicate if the fuel or air flow is inadequate, as the cell voltage will drop substantially. In large stacks (20 or more cells), this stack voltage-based measurement can be too insensitive to be of much use. On the other hand, the measurement of the voltages generated by the individual cells can be too complex to be economically viable. In case of low-voltage systems, cost is not a serious problem, and individual monitoring is possible for up to 40 cells using commercially available products.

In high-temperature SOFC cells, fuel utilization can also be measured by the oxygen content in the exhaust fuel gas. Using a zirconium oxygen probe (Figure 8.26e), the unused oxygen is drawn through the partially stabilized zirconia separator (partially stabilized zirconia is a mix of zirconia and 10–15% yttrium, which improves its thermal shock and impact resistance), which is balanced against the oxygen present in the fuel, which is measured by a second sensor. An overloaded individual cell will release excess oxygen into the fuel gas exhaust stream.

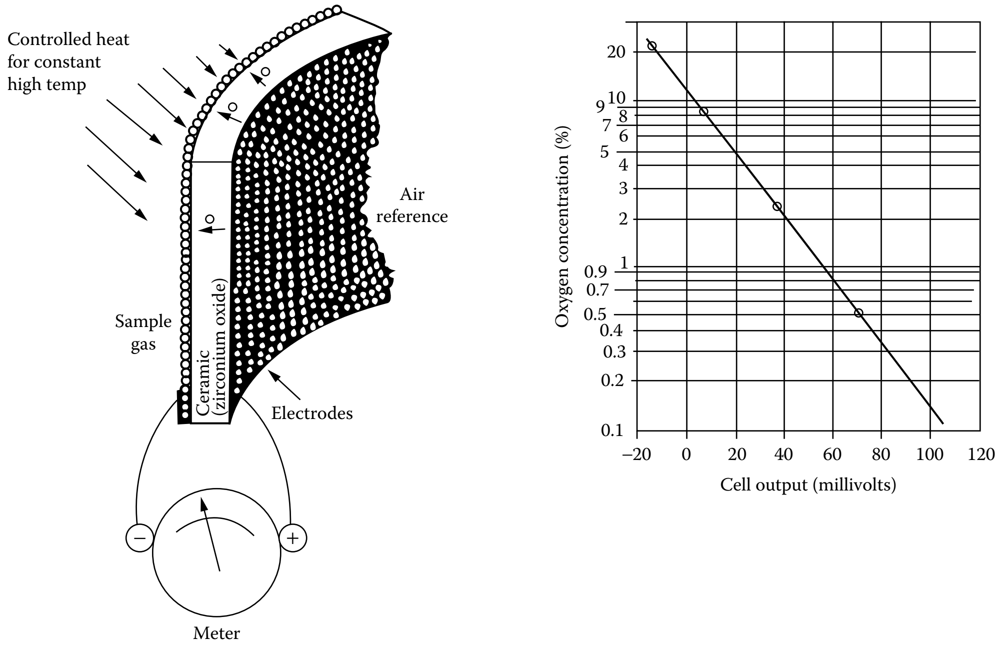


FIG. 8.26e

The flow of oxygen ions through the hot zirconium oxide electrolyte causes a voltage difference across the thickness of the element. (Courtesy of Ametek Inc.)

Both cell voltage and fuel oxygen content measurement respond to the worst cell in the stack. Erroneous readings can also occur if a cell is cracked, as this will result in the same reading as will an overloaded cell. In the case of overloading the fuel needs to be increased, while in the case of cracking the cell has to be isolated, so this reading is insufficient for accurate diagnosis.

Temperature monitoring of the stack exhaust is needed to resolve if the temperature rise is a consequence of cell overloading or of stack fire.

Both the cell voltage and the oxygen monitoring are negative correlations, because these measurements give an indication only when the utilization of oxygen drops. While this is useful information when the load is rising, the detection of falling load conditions require another mechanism to ensure that fuel is not being wasted. Therefore, as a means of checking, it is desirable to also measure the cell performance on the basis of a fuel cell model.

Feedback Control of Feed Flow In a load-following control system, as the demand for current falls, the fuel feed flow is reduced until a predetermined minimum value is reached or until a fuel deficit is detected. If the fuel flow cannot be raised when a fuel deficit is detected, the inverter drawing power from the cell stack will reduce the current draw. If the fuel cell system is connected to a grid that is also connected to a booster, this is not a problem, because more power can be drawn from the grid to make up the shortfall.

If the grid is isolated, it is necessary to shed some of the load on the cells by making up the shortfall from battery or other storage systems, such as capacitor banks, in the short term.

In multiple-stack installations it is important to control the performance of each stack separately to ensure that one stack cannot discharge into another. This is necessary because the manufacturing of identical stacks is just about impossible with the current means of manufacturing in the industry. This is particularly a problem for active anode SOFC and molten carbonate cell designs, because the oxygen drawn though the cell electrolyte can oxidize and destroy the catalytic ability of the cell.

To ensure that back feeds cannot occur, it is usual to design the stack electronics to act as a variable step-up DC/DC inverter with diode protection. In fuel cells, current flow always equals gas flow. In other words, whenever current is drawn, gas must also move from one place to another. With PEM cells, the current drawn results in moving of hydrogen to the oxidizer side, while in high-temperature cells, it results in moving oxygen into the fuel side.

If the reagents are not available to remove these flows, reagents can appear on the wrong side of the cell, which can cause adverse reactions or, in extreme cases, can cause cell fires and explosions. In isolated stack situations this is not a problem, but for installations connected to a grid or in multiple-stack installations, it can cause serious problems if reverse currents are permitted to flow.

Fuel cell control is relatively simple for one cell; the difficulties come when multiple cells need to be controlled in parallel. The control challenge is similar to controlling a single cylinder in a conventional engine, to controlling 24 cylinders in parallel; of course, no one has yet built a functioning 400 cylinder engine. Added to this challenge is the fact that because all cells are connected in series in a stack, the performance of the worst cell also determines the performance of the best cell.

Pressure Balancing and SOFC Problems Further concerns are the balancing of the inlet pressures on both sides of the cells to ensure that the maximum differential pressures are not violated. This is particularly a problem when thin film electrolytes are used in PEM cells or with some SOFC systems. The absolute pressures can, however, be raised quite high, as the strength of the outer casing is not limited by the requirement to enhance diffusion across a membrane, thereby potentially improving the power density of some fuel cell systems. This improvement has to be balanced against the energy cost of pressurizing the cell system. If fed from a high-pressure source like a gas bottle this cost is negligible at the consumer's end but can be significant at the source.

Solid oxide fuel cells have their own peculiar problems. Because of the elevated inlet gas temperatures, special materials are required to permit safe operation. This is because the chromium oxide films that are usually used are removed by the action of the high-temperature fuel gases and the steam, which cause accelerated corrosion and pitting.

Some of the protection techniques applied have been borrowed from the gas turbine industry, where the gas ducts had to be protected from the fuel supply. Commonly used materials are self-aluminizing steels, although alumina coatings have also been applied to Hastelloy and stainless steel to good effect, in a manner commonly used during the Second World War by the Junkers company in the production of the early jet engines. This technique utilizes the high-temperature strength of the material while bypassing its Achilles' heel.

Thermal Control

Temperature controls are used to override the power generation controls in fuel cells. The operating temperature of low-temperature cells can rise due to a combination of increased current draw, cell impedance, and elevated gas inlet temperatures from the external reformers to raise the temperature of the cell system. This can degrade electrolyte and can destroy the cell.

In high-temperature cells, this heating effect can be useful, because at low temperatures it will increase the load on a cold cell to raise its temperature and will reduce the load on hot stacks. Temperature controls are usually successful when a combination of heat exchanger bypasses and auxiliary heater controls are used to maintain the temperature of loaded cells.

Idling fuel cells will, however, require external supplements to maintain normal operating conditions. These supplements can be provided by either fired heaters or by external electrical heating. Efficient insulation is essential for all high-temperature cells.

A further issue for high-temperature fuel cells is the melting temperature and related vapor pressure of the current-collecting cabling. While most commercial high-temperature cable is nickel, its high resistance makes its use in fuel cell systems unwise and inefficient. Instead, solid silver cabling is used; however, the silver's vapor pressure limits its service to below 900°C, with an 850°C limit being more common. Recent developments in inexpensive cable insulation including ready availability of commercially produced silica and Nextel insulated thermocouple wiring simplifies the solving of the problems associated with sensor wiring. Therefore, this no longer is a serious problem.

CONTROL OF THE AUXILIARY SYSTEMS

Figure 8.26f describes a high-temperature fuel cell system that includes a number of auxiliary unit operations. The fuel treatment controls described in the following paragraphs include the scrubbers, humidifiers, and reformers. In addition, the controls of the exhaust systems, the inverters, and the required shutdown controls will also be covered.

Fuel Treatment Control

For all fuel cells, except those running on high purity hydrogen and similarly prepared fuels like methanol, some form of fuel treatment is required. The main problem with fuel supplies intended for conventional combustion systems is the presence of minor contaminants containing ash-making chemicals and sulfur compounds. In fuel cell applications, the sulfur compounds form corrosive substances that poison the catalysts in the reformer stages and the fuel cell itself.

Scrubber Controls Natural gas supplies can be a problem, as the hydrogen sulfide concentration in natural gas can reach 5.7 mg/m^3, and other sulfur compounds including mercaptan odorant (which typically contains 5 mg/m^3 sulfur) can total 50 mg/m^3 without violating industry standards. While these concentrations are low, they are sufficient to poison a fuel cell and the reformer stage within a short period if not scrubbed out.

Heavier oil supplies such as diesel (which can be used in solid oxide fuel cells) can also have a substantial sulfur content. Diesel oil received from third world refineries can contain as much as 10% of sulfur compounds due to lax regulation and cost pressures.

The most proven scrubbing technique to absorb sulfur compounds is to use a bed of sintered zinc oxide pellets that are heated to 300°C, but this requires a slipstream of hydrogen-rich gas to crack the more complex compounds to hydrogen sulfide. Iron oxides have also been used as absorption materials in similar installations. The saturated absorbent can then be either dumped (the usual option) or regenerated and recycled at a suitable facility.

This can result in a chicken-and-the-egg scenario at start-up, because hydrogen is needed for start-up but is not available until after start-up. This can be solved by either providing a bottle of hydrogen for start-up or a cold start-up system,

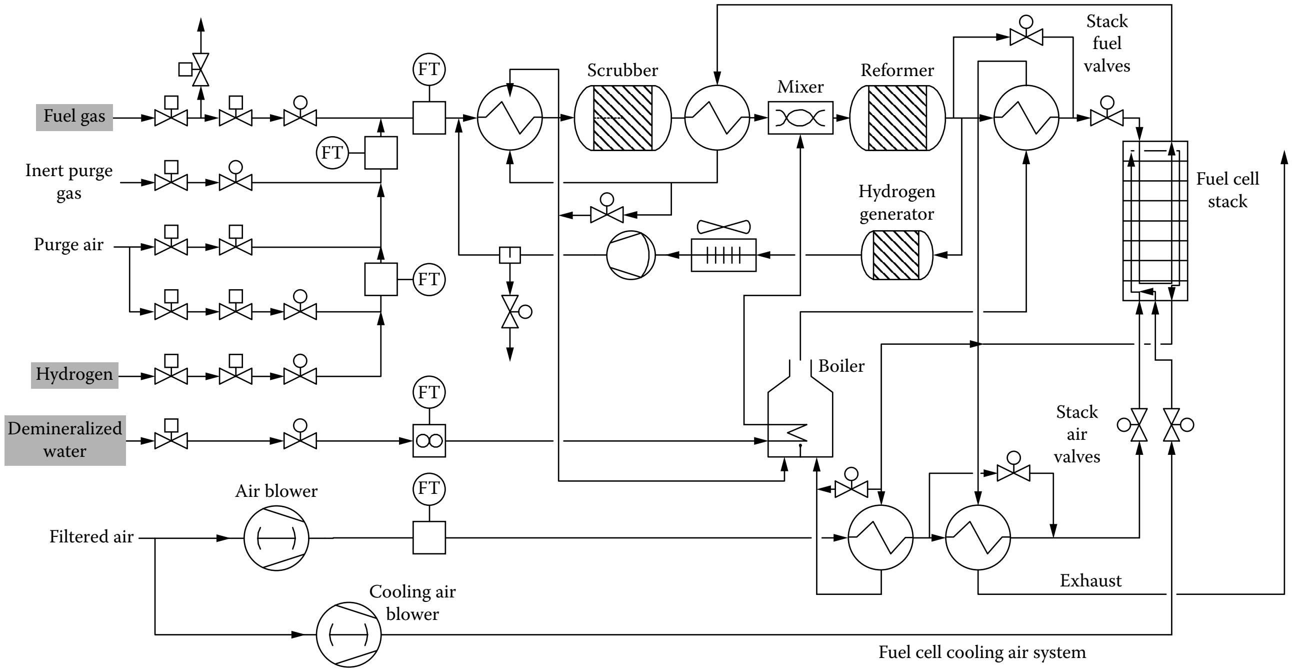


FIG. 8.26f
A simplified flow diagram of a high-temperature fuel cell system.

which is storing the hydrogen produced during the previous run.

The scrubber operation requires tight temperature control, because if the absorber is too hot, the sulfur is released, and if it is too cold the absorption process will not take place. Usually, electric heating is used to raise the bed temperature, and the heating is modulated on a feedforward basis as a function of the gas flow rate to prevent overheating or overcooling on load changes. The scrubbing process is not tolerant of water contamination, because it can cause the bed to collapse, if exposed for prolonged periods to high water vapor pressures or to condensation.

Alkaline fuel cells also require the removal of carbon oxides including carbon dioxide. The current carbon dioxide removal technology is, however, not suitable for use in mobile (or even compact) systems in long-term operation.

Humidification Controls After the scrubber (see Figure 8.26f), water vapor must be added to the scrubbed mix, in order for the shift reaction to properly function in the reformer. The shift reactions in low- and high-temperature cells are as follows:

$$\text{Low-Temperature Cell: } CH_4 + 2H_2O \Rightarrow CO_2 + 4H_2 \qquad \mathbf{8.26(26)}$$

$$\text{High-Temperature Cell: } 2CH_4 + 3H_2O \Rightarrow 2CO + 5H_2 \qquad \mathbf{8.26(27)}$$

In case of complex (saturated) hydrocarbons, the shift reaction is

$$2C_xH_{2x} + 2H_2O \Rightarrow 2xCO + 4H_2 \qquad \mathbf{8.26(28)}$$

The steam flow to the mixer is controlled on a mass flow ratio basis to the gas flow, considering the gas composition and the degree of reformation required.

For research applications, humidification can be controlled by bubbling the fuel gas through a water bath, the temperature of which is adjusted by the humidity controller and corrected for the system pressure. Such controls are similar to that used by the Carrier air-conditioning systems dating back to the early 20th century. For any system larger than a test bed, however, the bath system breaks down, as the bubble size generated by the available passive diffusers is too large for effective humidification in any reasonably sized bath.

Therefore, for larger humidification systems, steam generators are required (Figure 8.26f). These being low-pressure boilers, the steam must be superheated (up to 900°C) to minimize the thermal losses in high-temperature fuel cells. The steam produced by the boiler is mixed with the fuel stream by in-stream mixers ahead of the reformer. Mixing the gas with the highly superheated steam also assists with the heating of the fuel gas before it enters the reformer. The boiler itself must be designed for dry operation, as the bulk of the system will act as a superheater coil with no control of the heat flux.

Controlling the water feed to these boilers is a challenge in itself, as the flow rates are too low for most metering systems that are available at a reasonable price. For a 40 kW fuel cell system, the required water flow range is 0.25–10 gr/sec (see Chapter 2 in the first volume of this handbook for an in-depth discussion of low flow measurement).

In a particular application, paddlewheel flow meters have been made to function with adequate turndown in this range, but special linearization was required to accommodate for the nonlinear response at low flow. Thermal dispersion flow meters would also have been an option, but the price was too high for a commercially viable system.

Conventional control valves presented inadequate turndown (for high-turndown control valves, refer to Chapter 6), so in this particular installation pulsed and pulse-modulated solenoids were used, with the flow being averaged to match the required ratio to the fuel gas.

Limiting the outlet temperature both before and after the mixer is important to prevent flooding or forming slugs of water, which are then injected into the fuel stream. The boiler feedwater flow is also limited on the basis of low steam temperature and on the basis of an interlock that responds to the boiler casing temperature.

Some consideration has been given to the recycling of the water content of the exhaust but the carbon dioxide content makes this difficult in small systems sensitive to carbon monoxide and dioxide.

Reformer Controls All fuel sources other than hydrogen require some reformation to provide a hydrogen-rich gas feed to the fuel cells and the scrubber. Some fuel cells, such as the direct conversion methanol cell systems and to a more restricted extent the solid oxide fuel cells, use reforming catalysts on the cell surfaces.

However, in order to extract the most energy out of the feed stock, it is common practice to add a reformation stage using a water gas shift reaction to separate the hydrogen gas from the carbon backbone of the base fuel by a combined use of heat, steam, and catalyst. The reformer reaction requires temperatures higher than that of the scrubber stage. They are around 300–400°C for common iron/chrome catalysts but temperatures of over 500°C have also been found to be effective.

To reach such temperatures requires that the gas stream be further heated either by a heat exchanger that is using waste heat or by electric heating, which is also needed for start-up. Low-temperature reformer units operating at 160–250°C and using copper zinc oxide-based catalysts reduce the cooling requirements of the low-temperature fuel cells, but at the expense of the high risk of fire if the system is shut down incorrectly.

During start-up, the catalyst first has to be activated with hydrogen gas. This first requires a purge to strip the atmospheric oxygen and condensation from the beads (or carrier if used), and then heating to the operating temperature. The traditional means of purging is to use an inert blanketing gas for this purpose until the minimum activation temperature is reached.

A cheaper (at least in terms of inert gas) alternative is to heat up the system to 150°C by using air, at that point purging the condensate and then switching to fuel gas to purge the air while remaining below the autoignition temperature of the gas stream. This, however, requires additional valves in the fuel train to isolate the fuel from the air during normal operation and requires an explosionproof design of the complete fuel cell system.

Common reformer catalysts include iron/chromium complexes, nickel-based catalysts, vanadium/cobalt, and the very pyrophoric (but very effective) copper/zinc oxide-based catalysts. A more expensive option is platinum-based ceramic catalysts but care has to be taken to prevent carbon fouling, which will need to be burned off.

Recent developments have shown that this catalyst can withstand oxygen contents similar to normal air at 550°C and can operate at 400°C. This is an area of intense current research, with copper oxides showing great promise. Gold-based catalysts have also shown some recent promise, but their high cost will probably restrict their use to niche applications.

With either technique, once the minimum operating temperature is reached, hydrogen is introduced into the reformer to strip the oxide layer from the catalyst that has been made passive. As this is an endothermic reaction, extra heating is required. In addition, close control of the hydrogen flow is also required to ensure that even reduction is achieved. The hydrogen is usually supplied from gas bottles, but some attempts have also been made to use internal hydrogen generators to render the systems more self-contained and eliminate a potential fire hazard.

The hydrogen generators are simply miniature versions of the reformer. They are driven to higher temperatures and are not made passive on shutdown and can also be used to improve the performance of the slipstream to the scrubber in normal operation.

The slipstream needs to be stripped of water, which is generated by the shift reaction before being recycled. Both air- and water-cooled condensers have been used for this task with some success. The condensate can be removed by modified flameproof electric steam traps.

Low-temperature fuel cells require a further scrubber stage after the reformer to remove any carbon monoxide using platinum group metals, in combination with cobalt, silver, iron, copper, and molybdenum.

Exhaust Systems Controls

As most fuel cells do not produce complete conversion (other than hydrogen types), and most do vent unburned gas, some means of handling the waste fuel is required.

In proton exchange membrane-type cells, the hydrogen ions migrate to the oxidizer or cathode side of the cell, which results in a level of water or water vapor building up in the oxygen stream. This can greatly reduce the available reaction surface area and, therefore, must be flushed before that happens.

In hydrogen fuel cells of this type, the fuel stream is not contaminated with waste products and no exhaust is required. In methane- and methanol-powered PEM cells, reaction products from the fuel side eventually can poison the cell by limiting the maximum achievable efficiency, so the fuel stream has to be periodically purged.

In high-temperature SOFC cells the oxygen moves from the cathode to the anode side and reacts with both the hydrogen and carbon monoxide (and the anode material if a reverse bias is permitted to occur), generating water vapor on the fuel side of the cell requiring a continuous purge. The air side is depleted correspondingly and the resulting nitrogen-rich gas also needs to be purged before the low oxygen levels affect the cell output voltage.

Recycling of the waste fuel gas stream is possible in pure hydrogen-powered systems, if no contamination problems are present. This recycling in hydrogen-powered systems requires the removal of the water and recompression for injection to the intake stream.

Hydrocarbon fuel cells have the additional burden of carbon dioxide and monoxide accumulation. Their extraction is uneconomic in small installations; therefore, an afterburner is used to neutralize the exhaust stream and to extract the residual energy.

Either conventional burners or catalytic converters can be used in smaller systems. Their controls are the same as those of small heating units that are provided with pilot burners to ensure light off. In larger systems, one has the option of using small gas turbines to extract the energy remaining in the gas by generating electricity (at a slightly lower efficiency). Enough flammable components must remain in the gas stream to ensure that the combustor can maintain a stable flame.

With catalytic combustors, the opposite is the case, because when fuel-rich gases are vented in response to the shedding of the loads, the control of combustor temperature can become critical, requiring water sprays in extreme cases.

High-temperature systems can use the residual energy in the exhaust gases to heat the incoming feed gas to the operating temperature of the cell or to generate steam for humidification. This leaves considerable levels of available energy, which can be used for industrial and domestic heating or cooling by the use of absorption chillers. (For the controls of chillers, refer to Sections 8.12 and 8.13.) The downside of recovering the residual energy is that if the burners are run at too high a temperature, the NO_x emissions will no longer be low.

Inverter Controls

Because fuel cells produce direct current and the power distribution grid uses alternating current (for very good reasons), inverters are required to make the fuel cell-type generators compatible with the installed base of electrical consumers. The required technology for the inverters is the same as is used by the solar power industry, and the designs and requirements are frequently shared between the two industries. The solar power industry being more developed, it tends to lead the way, and therefore most specifications tend to refer to solar activities, but where a certification framework is required for fuel cells, most authorities will use the existing specifications written for the solar power systems.

For simple grid isolated systems, a simple electronic inverter that is similar in function to the converters used in variable-speed drives is used to drive the load. Some larger fuel cell systems also use this strategy to serve their internal loads of small blowers and other drives directly off the internal DC bus. More complex systems are required for interfaces to the distribution grid.

The simplest of these complex systems is the back-up power supply configuration, where an undervolt switch trips the fuel cell system into service on grid failure. In this case, frequency and voltage control are not critical, because the fuel system is transferred smoothly to the grid, when its voltage is zero.

The next level of complexity is when an uninterruptible power supply (UPS) is required, which is an extension of the conventional battery-backed UPSs. In this configuration, the mains power grid charges a battery bank in parallel with a fuel cell system, and the grid failure results in no interruption of the supply, because the battery bank gives time for the fuel cell system to ramp up.

The most difficult requirement is to provide a grid-connected inverter. Such an inverter must maintain the same frequency as the incoming grid and must detect the loss of the grid and respond to such an event either by isolating itself from the grid or tripping off-line. The usual means of detecting a grid disconnect or island situation is to have the output frequency periodically drift up or (more usually) down, because if the grid is in operation, the phase shift will be readily detectable. At 2 Hertz away from the nominal grid frequency the inverter will trip to island mode and either shut down (pending return of the mains supply), or disconnect itself from the grid.

The alternative method of detecting grid failure is by detecting the voltage deviation (either high or low) and set the inverter to trip if the voltage deviation moves outside a +10% or –6% deviation from nominal supply values. The purpose of doing this is to protect linesmen from possible back feeds from generation systems, which they may not be aware of, when they are isolating the main supply. This system must detect a grid link failure within 2 min. With multiple inverters driving the same load, one of the inverters has to be the frequency master in order for this test to function correctly, while the others are operated in load share mode.

Grid-connected inverters can either drive into the grid as a power supply or can supply some of a large load. The third option is to drive the main's current draw to zero, which requires a more delicate balance. The choice among these options depends upon the specifics of the local electricity supply and upon the receptiveness of the supply authority to generators that are not owned by them.

Base load fuel cells are, as yet, uncommon but will most likely comply with the requirements for conventional generator sets regarding frequency stability and load sharing modes of operation. No fuel cell system has as of yet demonstrated either the reliability or the power output needed for base load operation, except for the smallest and least critical operations.

Shutdown Control Requirements

The fuel cell's control system should also provide the controls for a safe shutdown. The emergency shutdown sequence starts with the isolation of the fuel gas supply and is continued with "inerting" of the fuel cell system to snuff out any fires that could be caused by leaks and to prevent the possibility of the fuel cell becoming an ignition source.

Conventionally, this is done by applying a nitrogen blanket, which is supplied from a high-capacity liquid nitrogen tank. High-temperature cells that are located in or near hazardous areas may require more drastic design features, such as using liquid CO_2 to cool the cell to below the auto-ignition temperature of 400°C (destroying the fuel cells in the process). This occurs in under 2 min and as such it complies with requirements for a fire suppression system and removes the need for the nitrogen supply.

A more normal shutdown sequence would first flush the fuel cell, reformer, and scrubber with nitrogen or carbon dioxide (if it is safe for the cell design). This step is followed by a slow bleed of air and nitrogen to repassivate the fuel cell and reformer under temperature control. If this is not done gradually the reformer can reach temperatures high enough to violate its containment (melt down) and become unrecoverable. The reformer performance does decrease after this treatment but over 90% of its capacity can be retained. Similar problems are present when reforming in the fuel cells themselves.

High-temperature fuel cells are governed by the same gas safety regulations as are fired heaters and boilers, because they also operate at above the auto-ignition temperature of the fuel gas. Some gas reformers are also in the same category. Therefore, double block and bleed arrangements are required for commercial systems in all but the smallest installations, and rapid removal of internal ignition sources is required after a forced shutdown, in addition to meeting the system purge requirements that are set in the relevant standards.

CONCLUSIONS

Further research is required to reduce the cost of instruments for all fuel cell systems. For example, a complex fuel cell system can require upwards of a hundred flow control valves. Even if the cost is only $200 for a typical low-cost commercial valve this cost can exceed the total cost of alternative generation equipment by a sizable margin. Transition to high-temperature fuel cells pushes the valve price up as special materials are required, yet low cost is critical for commercial viability and salability of fuel cells, if they are ever to move out of the laboratory and into general use.

Another source of problems are the thermal losses induced by instrumenting a typical high-temperature system. Thermocouple wiring and metallic capillary tubing to pressure and flow transducers can easily double the thermal losses of the system. Therefore, nonmetallic and thermally insulating sensor leads are required at reasonable costs for reliable operation at 1200°C. Today's costs of tens of thousands of dollars per point for using such commercial instruments as optical strain gauges and thermometry systems are totally unacceptable.

It is essential that in the field of fuel cell controls the instrumentation pricing changes from what is acceptable for research purposes to the sensor prices that car manufacturers are paying today. This is essential to the commercialization of the current generation of fuel cell technology. The reduction in price must also be matched with several orders of magnitude improvement in reliability. Such reliability today is available only in the current generation intelligent instrumentation (but at a cost that is unacceptable for commercial fuel cell systems).

The same reduction in pricing and increase in reliability are also needed for the auxiliary fuel cell equipment controls, from heat exchangers to the inlet gas train.

Bibliography

Binder, M. J., Taylor, W. R., and Holcomb, F. H., "Experience with the DOD Fleet of 30 Fuel Cell Generators," U.S. Army Engineer Research and Development Center, 2001. International Gas Research Conference (IGRC 2001), Amsterdam, The Netherlands, November 5–8, 2001.

Hartvigsen, J., Khandkar, A., and Elangovan, S., "Development of an SOFC Stack Performance Map for Natural Gas Operation," SOFCo., Salt Lake City, UT, Sixth International Symposium on Solid Oxide Fuel Cells, October 17–22, 1999, Honolulu, HI.

Dorf, R. C. (ed.), *Electrical Engineering Handbook*, 2nd edition, Boca Raton, FL: CRC Press, 1997.

Perry, R. H. and Green, D. W., *Perry's Chemical Engineers Handbook,* 7th edition, New York: The McGraw-Hill Companies.

AS4777 — 2002, Parts 1 to 3, "Grid Connection of Energy Systems via Inverters," Standards Australia.

AS4564 — 2003, AG864 — 2003, "Specification for General-Purpose Natural Gas," Standards Australia/Australian Gas Association.

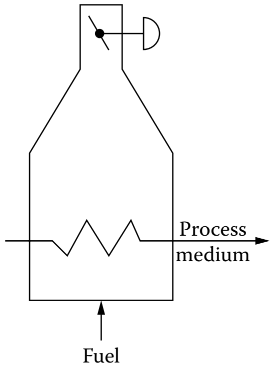
JIS C 8801:2002, "General Rules for Phosphoric Acid Fuel Cell Power Generating System," Japan Industrial Standard.

JIS C 8802:2003, "Accelerated Life Test Methods for Phosphoric Acid Fuel Cell," Japan Industrial Standard.

UL1741, "Static Inverters and Charge Controllers for Use in Photoelectric Power Systems," Underwriters Laboratories.

8.27 Furnace and Reformer Controls

A. M. CALABRESE (1970, 1985) **B. A. JENSEN** (1995)
S. E. GAERTNER AND B. G. LIPTÁK (2005)



Flow sheet symbol

INTRODUCTION

The discussion that follows covers a wide range of process furnaces, including heaters, cracking furnaces, and reformers. Each type of furnace is described from a mechanical and process standpoint, in enough detail to make the recommended control strategies understandable. Examples are presented so that extrapolation can be made to services not specifically discussed in this section. The recommended control strategies are broken down into the following categories: 1) process controls, 2) fuel firing controls, and 3) safety controls.

In the first part of the section, general information is provided concerning the furnace and fired heater processes, their sensor and analyzer requirements, and their basic control requirements, including air preheat controls. In the second part of this section, specific control systems are described for start-up heaters, process heaters, and reforming and cracking furnaces (Figure 8.27a). The last part of this section discusses some of the more advanced control strategies that can be applied to furnace controls and optimization.

In furnaces and heaters, thermal energy is transferred to a process stream or feed in a controlled manner. The typical furnace, or heater (the terms will be used interchangeably), usually takes the form of a metal housing lined with an insulating refractory. The charge can enter as a liquid, gas, or two-phase mixture and may or may not be transformed to a different state by the energy supplied. In these types of furnaces, the charge can be carried through the furnace or heater continuously through metal tubes or can be batch-heated by leaving it stationary after it enters the furnace.

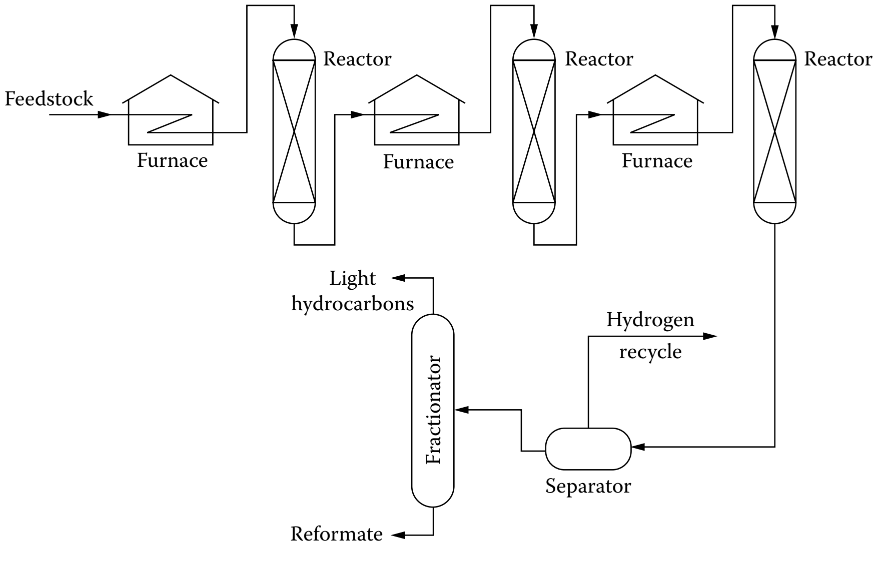


FIG. 8.27a
Illustration of a multistage reforming process.

The discussion in this section covers the controls of 1) refinery feed heaters and reboilers, 2) pyrolysis furnaces for ethylene production, and 3) steam reformers for hydrogen or synthetic gas production.

As will be seen later, the heat content of the hot flue gases leaving the furnace can be used to preheat process feed in a feed preheater or can preheat the boiler feedwater and generate process steam.

Controls for the more common types of process furnaces will first be described from a process standpoint. Next, the important process variables will be noted and the methods used to measure and control them will be outlined. The various types of firing controls applicable to the particular type of furnace will then be described. Finally, specific safety considerations and instrumentation will be presented.

As will be seen, the common control strategies include a feedback temperature/flow cascade loop, where the furnace outlet temperature adjusts the fuel flow into the fired furnace. The firing controls may use feedforward based on the feed flow rate in order to speed up the control system response to changes in feed flow rate.

Furnaces with long dead times or with highly interacting control loops tend to be unstable (Figure 8.27a). To overcome these problems, the use of mathematical analysis and process simulation may become a necessity. The control systems outlined have been empirically developed, and most of them are tried and reliable designs. In the more advanced control strategies, which are described in the last part of this section, process dynamics are also taken into full consideration.

GENERAL CONSIDERATIONS

The primary functions of furnace control systems are

1. To ensure that the charge receives the heat energy at the proper rate. In feed heater furnaces, the main controlled variable is the exit temperature of the charge.
2. To maintain efficient combustion of the fuel. Proper combustion of the fuel requires the regulation of the air-fuel ratio, and the control of the atomizing steam flow rate when fuel oil is burned. Other important controlled variables are draft and stack emissions of NO_x and of other pollutants.
3. To maintain the safety during all phases of furnace operation so as to prevent explosions or fires.

Standards Applicable to Furnace Controls

The ANSI, API, CFR, IEC, ISA, NFPA, and OSHA standards that are used to regulate furnace controls are listed at the end of this section, before the Bibliography.

In furnace operations, safety is a very important consideration. In the case of refinery furnaces, the potential sources of hazards include 1) fire or explosion caused by tube rupture. Such rupture can occur because of tube overheating due to feed flow loss or because of flame impingement; 2) explosion in the firebox caused by the loss of flame or by improper ignition or purge procedures; and 3) implosion of the furnace or refractory damage caused by high draft.

Furnace design and operation must follow OSHA's code of federal regulations: 40CFR1910.119, "Process Safety Management." The standard ANSI/ISA-84.01–1996, "Application of Safety Instrumemented Systems (SIS) for the Process Industries," must also be implemented, usually by applying independent PLCs for the control of firing shutdown sequences.

The required level of reliability must be determined by the analysis of the required Safety Integrity Level (SIL), described in the ISA Standard TR84.02-2002. The SIL level is established using process hazard analysis (PHA), which determines if dual- or triple-redundant PLCs must be used. An international companion standard that is also applicable in making this decision is IEC 61508, "Functional Safety–Safety-Related Systems." In addition, local codes and industry practices, usually set by the insurers of the plant, should also be consulted before controls are designed and installed.

The Combustion Process

To completely burn the combustible portion of any given fuel, an ideal quantity of oxygen and, therefore, an ideal quantity of air is required. Under real conditions, the inefficiencies in combustion require that some additional air be provided to ensure complete combustion. The quantity over and above the ideal (stoichiometric) amount of air is called the "excess air."

On a volumetric basis, air contains 0.21 oxygen and 0.79 nitrogen (on a mass basis 0.232 oxygen and 0.768 nitrogen). For example, in the combustion of methane (CH_4), the ideal amount of air required per pound is 4/0.232 = 17.24 lb, and in volumetric units, the volume of air required per cubic foot of methane is 2/021 = 9.52 ft^3.

The ratio of air supplied per pound of fuel (W_1) to the air theoretically required (W) is

$$(W_1/W) = \{3.02[N/(CO + CO_2)]\}/\{34.56[C/3 + H - O/8]\} \qquad \mathbf{8.27(1)}$$

Combustion Efficiency and Excess Air If we call "a" the ideal (minimum) amount of air required for complete combustion and we call "*xa*" the actual amount of air admitted, then "x – 1" is the excess coefficient. On a volumetric basis, the excess air requirement (E in %) can be calculated by

$$E = K[21/(21 - \%O_2) - 1] \times 100 \qquad \mathbf{8.27(2)}$$

where $K = 0.9$ for gas, 0.94 for oil, and 0.97 for coal.

Using the stack gas composition analyzer readings for carbon monoxide (CO), the air supply to the furnace can be adjusted to maintain the combustion efficiency at the optimum. Figure 8.6ttt in Section 8.6 shows the relationship between measured ppm concentration of CO in the stack gases and furnace efficiency.

The optimum percentage of excess oxygen is some percentage greater than what is theoretically needed for complete combustion. This margin is usually between 2 and 4% for most furnaces. For continuous emission monitoring system (CEMS) purposes, the pollutant concentrations are calculated with a reference basis of 3% O_2 on a dry basis.

The amount of carbon monoxide in the stack gases can be measured to determine the optimum set point for excess O_2 control. Elevated CO levels indicate that some of the fuel is unburned, which results in the waste of heat and the danger of potential explosion at high CO levels. In some installations, instead of continuously monitoring CO, only a one-time measurement is made to establish the crossover between the presence of excess oxygen and excess unburned fuel.

Combustion efficiency is usually defined in terms of heat input and output. In the case of furnaces, an energy balance around the unit must account for the total heat input and the total useful heat output. The difference is the total heat loss, as illustrated in Figure 8.6kkk in Section 8.6.

$$H_I = H_O + H_L \tag{8.27(3)}$$

where

H_I = heat input rate

H_O = heat output rate

H_L = heat loss rate

Furnace efficiency can be defined by one of the following three equations:

Input-output method:

$$EFF = 100 \cdot (H_O/H_I) \tag{8.27(4)}$$

Input-loss method:

$$EFF = 100 \cdot (1 - H_L/H_I) \tag{8.27(5)}$$

Output-loss method:

$$EFF = 100 \cdot [1/(1 + H_L/H_I)] \tag{8.27(6)}$$

Now, if these equations are expressed in terms of

Q = heat absorbed by fluid

W = fuel rate

I = heating value of fuel

L = total heat loss

Equations 8.27(4) to 8.27(6) become

$$EFF = Q/(W \cdot I) \cdot 100 \tag{8.27(7)}$$

$$EFF = 1 - L/I \cdot 100 \tag{8.27(8)}$$

$$EFF = Q/(Q + W \cdot L) \cdot 100 \tag{8.27(9)}$$

Although, in theory, any of these equations should produce the same results, experience has shown that the input-loss method [Equations 8.27(5) and 8.27(8)] produce the most reliable results, because the other methods are more sensitive to errors in process measurements.

Many furnaces burn gas preferentially and oil only as a standby. Emission compliance for NO_x is easier if gas fuel is being burned. However, some of the control systems described in this section are suited for combination firing (i.e., firing both gas and oil simultaneously).

Combustion Air Preheat With regard to furnaces, the overall efficiency is the ratio of the amount of heat absorbed by the process to the amount of heat liberated by the fuel. One of the major heat losses in a furnace is the heat lost up the stack in the form of hot flue gases. Therefore, much attention has been devoted to reducing the flue gas exit temperature as much as practical (to approximately 300°F, or 150°C).

A common method of reducing this waste is to add a combustion air preheater to the furnace. This device is a heat exchanger that transfers some of the heat from the hot flue gases to the cold combustion air, before it enters the burners (Figure 8.6ll in Section 8.6). Thus, the heat energy re-enters the furnace and is available for transfer to the process stream.

In addition to the combustion air preheater, it is usually necessary to add a forced draft fan to push the combustion air through the preheater (exchanger) instead of relying on natural draft. The instrumentation necessary to control this equipment is relatively simple. Figure 8.27b shows a typical combustion air preheat system.

In this control system, the forced draft fan is driven by a steam turbine drive, equipped with a variable-speed governor. A pressure controller (PIC-1) maintains the discharge pressure (and consequently the airflow) by manipulating the fan speed. A low-pressure alarm (PAL-1) is included to alert the operator of possible fan malfunction.

A trip interlock (FY-1) is provided to trip the forced draft fan to minimum speed in the event of failure of the induced draft fans. This precaution is necessary to prevent positive pressure in the hot combustion chamber, which could cause the hot gas or flames to blow out through the burner openings and other leaks in the furnace.

Temperature indicators TI-1, TI-2, TI-3, and TI-4 are provided on both streams in and out of the combustion air preheater to monitor the operation. Pressure measurements PI-1 and PI-2 are also included to allow the checking for excessive pressure drop across the unit.

Safety Considerations

As it is also discussed in the next paragraph, some 70% of furnace explosions occur during start-up or shutdown, where operator involvement is significant, while 21% of accidents were found to be caused by undocumented changes after commissioning.

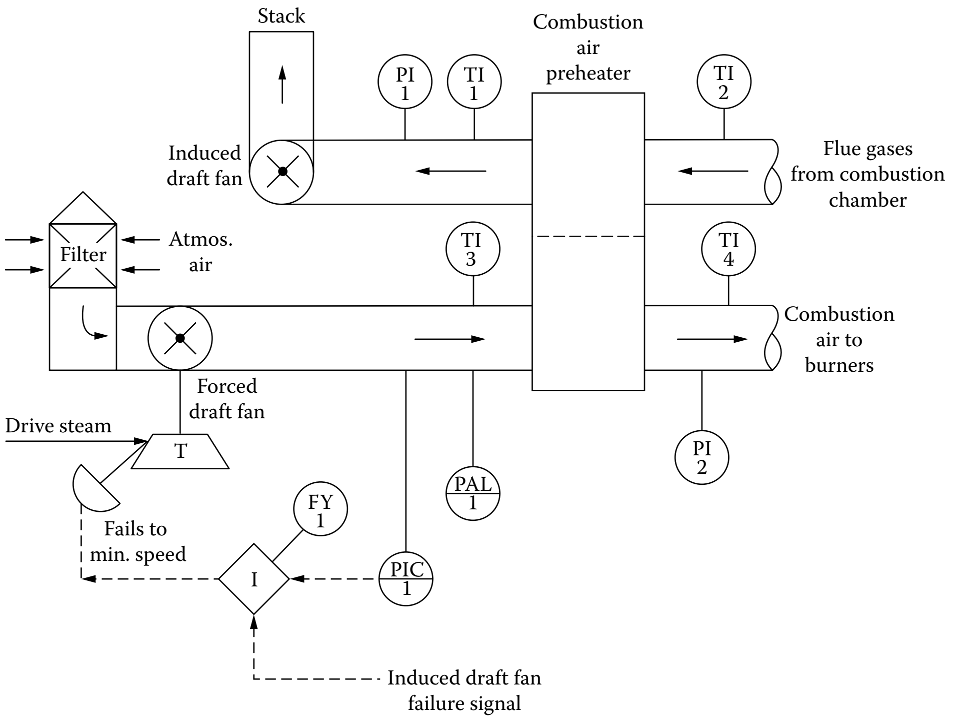


FIG. 8.27b
Combustion air preheat controls.

Fuel Gas Controls Figure 8.27c shows a typical fuel gas header, which can be branching off to several furnaces after the knockout pot. The relief valve shown should be piped to a relief header. For larger furnaces, the pressure control would be by a continuously throttling controller, rather than by a regulator.

The primary function of the furnace's safety system is to shut down the fuel supply if unsafe conditions evolve. In earlier furnace designs, the fuel control valve also served as the shut-off valve. This is no longer permissible, and separate, dedicated shut-off valves are required. Figure 8.27c illustrates a typical fuel gas regulating station with the safety shut-off valves left out and only their location being shown in the broken pipe line.

Figure 8.27d shows the safety shutdown valving usually used in the gas feed to a furnace. The top configuration is installed in the main gas line and consists of three on/off, tight shut-off, usually rotary ball valves. The dual shut-off

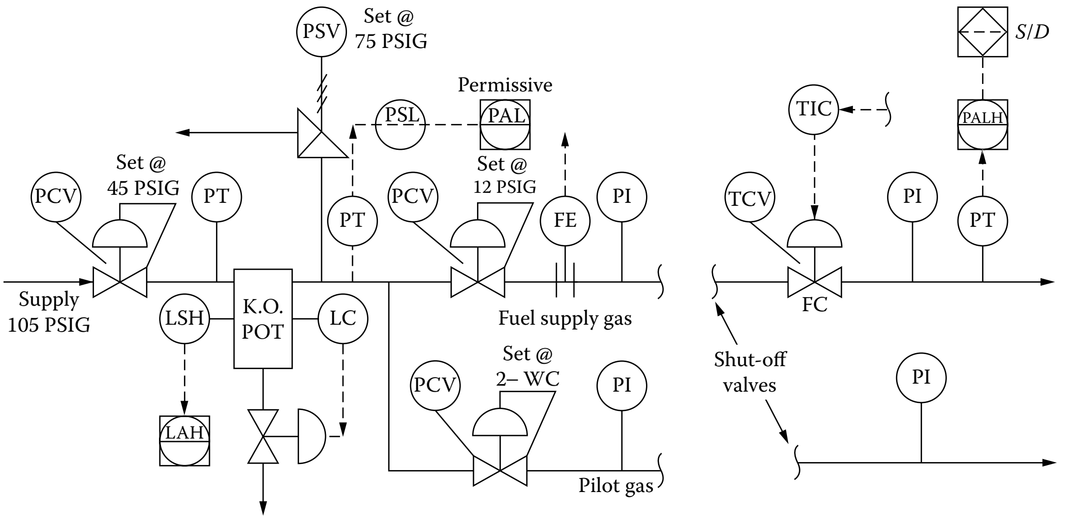


FIG. 8.27c
Fuel gas regulation station showing both the main gas supply line (top) and the pilot gas line (bottom). The locations of the safety shut-off valves are also shown in the broken pipe, but their actual configuration is separately illustrated in Figure 8.27d.

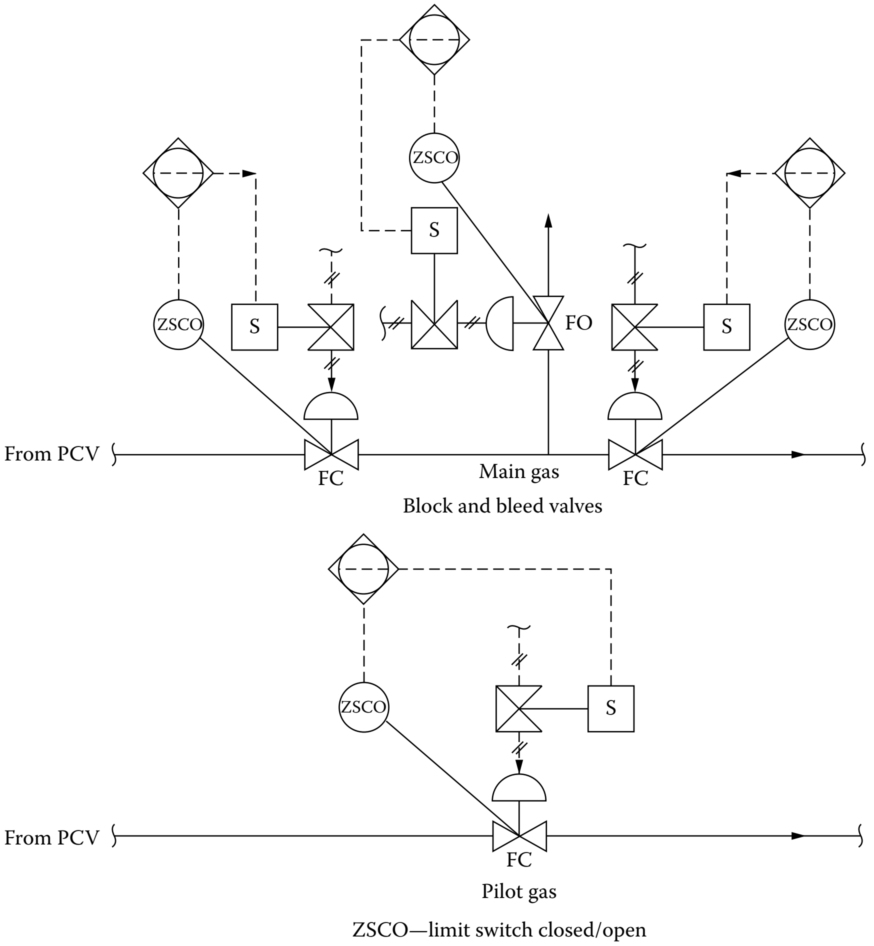


FIG. 8.27d

Safety shut-off valve configuration for the main gas (top) and the pilot gas (bottom) lines. Dual shut-off valves and a vent valve are used for the main, single shut-off valve for the pilot line.

valves are fail closed (FC), while the vent valve between them is fail open (FO). Therefore, when the shutdown interlocks are actuated, the two FC valves close, while the FO vent valve opens and vents the gas between the two FC valves, plus any leakage that might occur, to a safe location. The ZSCO (closed-open limit switches) provided on all three valves confirm to the controlling PLC that the interlock requirements have been performed.

The pilot gas line is usually provided with only a single shut-off valve with limit switch.

Safety Instrumented Systems SIS (formerly known as emergency shutdown system, or ESD) systems in general are discussed in detail in Sections 5.8 and 6.10. When designing a furnace control system, the designer should evaluate the following functional SIS requirements and clearly define the following:

- Define the safe state of the furnace (e.g., fuel shut-off, forced draft dampers open, and so on)
- Define the process variables, their normal operating range, and their limits
- Define the process outputs (conditions) and the actions required in response to them
- Considerations for manual shutdown
- Reset functions
- Actions on loss of power; UPS requirements
- Response time requirements for the SIS to bring the process to a safe state
- Operator interface criteria, including alarms and alarm responses
- Diagnostics; on-line and off-line testing
- Minimize nuisance trips while maintaining safety

As the final control elements in any SIS system are valves and dampers, their requirements should be defined by giving consideration to the following:

- Opening/closing speed
- Leakage
- Fire resistance–body and actuator
- Failure position

- Performance after staying in the same position for long periods
- Use separate shut-off and control valves
- Selection and correct sizing of solenoid valves and protecting them against plugging with bug screens
- Installing the solenoids between the valve positioner and the valve actuator
- Providing all shut-off valves with open-closed position detecting limit switches, which are wired to the SIS

The reliability of PLC-based SIS systems is increased by implementing them in redundant or triple-redundant (voting system) configurations, as far as their processors and power supplies are concerned. Important measurement and status input signals also are configured into 2 out of 3 voting systems, to avoid nuisance shutdowns caused by the failure of a single device or instrument. The shutdown output signals should also be redundant, and these redundant signals should originate on different I/O cards in the DCS or PLC systems.

The SIS safety and shutdown system is usually located on a PLC that is separate from the control equipment that directs the normal operation of the furnace. Such safety systems usually include a flame scanner. It is important to keep the air flowing through the firebox during a shutdown. In case of tube rupture, this may generate a lot of smoke, but it will protect the system from explosions that are caused by fuel-rich mixtures. Shutdowns are usually triggered by the following conditions:

- Low gas pressure: flame may go out and the resumption of gas flow can cause explosion
- High gas pressure: may blow flame away from burner
- Loss of flame, which can be detected by flame scanners
- High draft: may buckle furnace
- Low to zero draft: may push flames out
- Low feed flow: will cause tube overheating, coking, and eventual tube rupture
- Safety system failure: if PLC watchdog timer times out, this triggers hardwired fuel valve closure.

The emergency shutdown system software package within the PLC logic (also called the "problem solver") will cause the following actions as applicable:

- Fuel gas shutdown, with manual reset
- Pilot gas shutdown (not all furnaces)
- Opening of stack and fan dampers
- Ammonia injection shutdown–NH_3 valve, blowers (if the process includes them)

On furnaces with two induced air fans, dampers may be programmed to close if the fan is not running. The feed is usually shut down later manually when tubes have cooled off.

Guidelines for training can be found in OSHA Regulation 1910.119, which defines the requirements for safety systems-related training.

Procedures for on-line testing must also be established. Manual bypass switches should be available for each input that can initiate a shutdown so that the device can be tested periodically. If a bypass is open, that condition should also be indicated locally and alarmed in the central control room. In connection with testing of ESD systems, the ISA standard (S91.1-1995) should be followed, which lists the "Requirements for Functional Testing."

Purge Controls After a shutdown, the pilots must not be turned on and the furnace must not be reignited until all combustible gases are purged out. The purge cycle is initiated by the operator, who must not start the cycle timer in the safety PLC system until after the purge time has elapsed. At that point, the PLC unblocks the shut-off valves. A pilot flame detector is usually also part of the purge logic, and if no flame is detected after a few seconds, the purge will shut down and the cycle has to be restarted.

Large, PLC-controlled furnaces that are provided with emission abatement and monitoring systems are also provided with their own local control panel (Figure 8.27e), which is integrated into the total DCS- or PLC-based control system. Flow and draft indicators, alarms and shutdown bypass/test switches, and purge cycle controls are also often provided on these local panels. The purge-related pilot lights are usually designated as "purge on," "purge failed," or purge complete." After the purge is complete, the pilot burners can be ignited, and later the main gas control valve can also be opened.

Pollution Abatement

Furnace emissions must meet federal (EPA) and local air quality requirements. The applicable standards and guidelines include:

- ANSI/API 536, "NO_x Control for Fired Equipment in Refineries"
- 40CFR Part 60, "New Source Performance Standard" (section for oil refineries)
- 40CFR Part 61, "National Emissions Standards for Hazardous Pollutants"

Larger furnaces must be fitted with CEMS, which will be discussed in the next paragraph after NO_x controllers.

Gas- and oil-fired furnaces, operating at over 2800°F in their hot zones, typically produce about 100 ppm NO_x in their stack gas. EPA standards require that this concentration be reduced to 5 ppm or even 2.5 ppm. This can be achieved by several means:

- Use of cool-burning burners, with optimal air/fuel mixing nozzle design
- Use of flue gas recirculation, where part of the flue gas is returned into the firebox to cool the flame

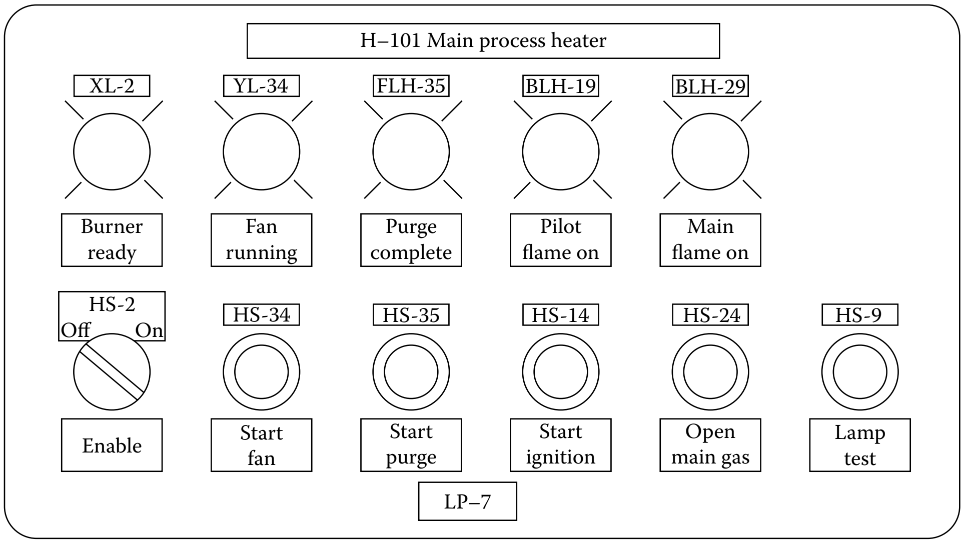


FIG. 8.27e

Local burner panel with status-indicating running lights and manual pushbuttons for purging and testing a typical furnace.

- Installation of catalytic converters between furnace and stack
- A combination of all above

Installing new burners and providing recirculation, by means of fans, can reduce NO_x to around 25 ppm. This still leaves the need for costly catalytic converters, which use vaporized ammonia to convert NO_x (mostly N_2O_3), according to Equation 8.27(10) as:

$$N_2O_3 + 2NH_3 = 2N_2 + 3H_2O \qquad \mathbf{8.27(10)}$$

To accomplish this conversion, because of the resistance of the selective catalytic reduction (SCR) catalyst bed, forced draft fans with control dampers are required. They must be installed on elevated platforms, making the installation expensive.

In the SCR process, by using a suitable catalyst, the NO_x reaction can be carried out at 300–400°C, a temperature normally available in a flue gas system. Figure 8.27f shows a schematic diagram of the SCR process. In this process at least 1% O_2 must be present in the flue gas, which is normally the case in boiler and furnace applications.

Using a movable injection lance is another effective means of reducing NO_x, CO, and volatile organic carbon (VOC), while improving thermal efficiency by optimizing the combustion process in boilers. The injection lance can be used with or without low NO_x burners.

These lances reduce NO_x emissions by 60–90% by injecting air, ammonia, or urea through retractable lances in the upper furnace. This creates turbulence and better mixing in the gas flow in addition to lowering the stoichiometric ratio of O_2. The lances are automatically retracted from the boiler on a regular basis and cleaned to remove the accumulated layers of soot and other depositions. Their costs are about 75% less than those of SCRs.

SCR Controls The controls of the SCR process are shown in Figure 8.27f. Here, from a storage tank, liquid ammonia is metered by a positive displacement pump or by Coriolis flowmeter to the ammonia vaporizer, which is heated by the dilution air. From the vaporizer, an array of small pipes takes the ammonia vapors to the distribution grid for injection at different elevations into the firebox.

Each pipe section in the distribution grid is provided with a balancing throttling valve, which serves to evenly distribute the vapors. The control valves in the major branches are controlled by the ammonia flow controller (FIC-1), which also controls the positive displacement ammonia pump. This controller measures the ammonia flow and in addition receives a feedback signal from a chemiluminescent NO_x analyzer transmitter and a feedforward signal, FT-1, the fuel gas flow transmitter, which detects the firing rate (gas flow) of the furnace or boiler.

FIC-1 also receives a trimming signal from the oxygen analyzer transmitter. It was found that increasing the amount of excess oxygen initially tends to increase the formation of NO_x, but further increases tend to decrease it due to flame cooling. The injection of ammonia is stopped by TSHL-1 (high-low temperature switch) when the converter temperature drops under 600°F, to prevent catalyst damage.

The NO_x Controller Figure 8.27g describes the ammonia controller (FIC-1 in Figure 8.27f) that is used as a cascade slave of AIC-1 to control the NO_x in a crude oil distillation heater furnace, which has a maximum gas firing rate of 8.5 MMSCFD. The figure also shows the various feedforward and trimming adjustments of the total control algorithm.

Feedforward is applied from the fuel gas flow transmitter (FT-1), which is the main determining factor in arriving at the set point of FIC-1. Oxygen content, detected by the transmitter AIT-2, serves to trim the fuel gas flow signal. Therefore,

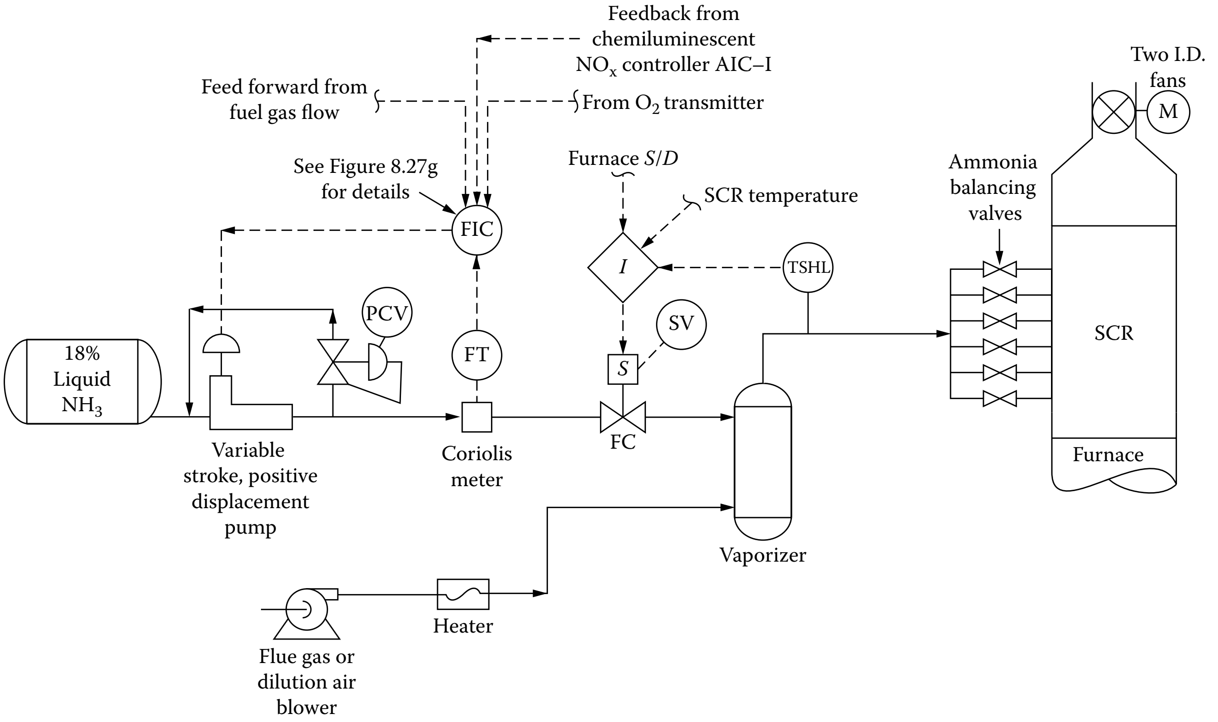


FIG. 8.27f
Control system for an SCR-type NO_x reduction system for the injection of ammonia.

changes in the output signals of FT-1 and AIT-2 will instantaneously change the set point of the ammonia flow controller, FIC-1. The basis for this relationship is that the NO_x production is proportional to both fuel flow and to excess oxygen, because the free nitrogen radicals—at over 2800°F—will seek out the extra oxygen and will combine with it.

The cascade master of the feedback portion of this control system is the NO_x controller, AIC-1. This loop is slow; its dead time can be 10 min, because the time lags from the NH_3 injection points to the stack and then through the sample line to the NO_x analyzer all add up before a measurement is obtained.

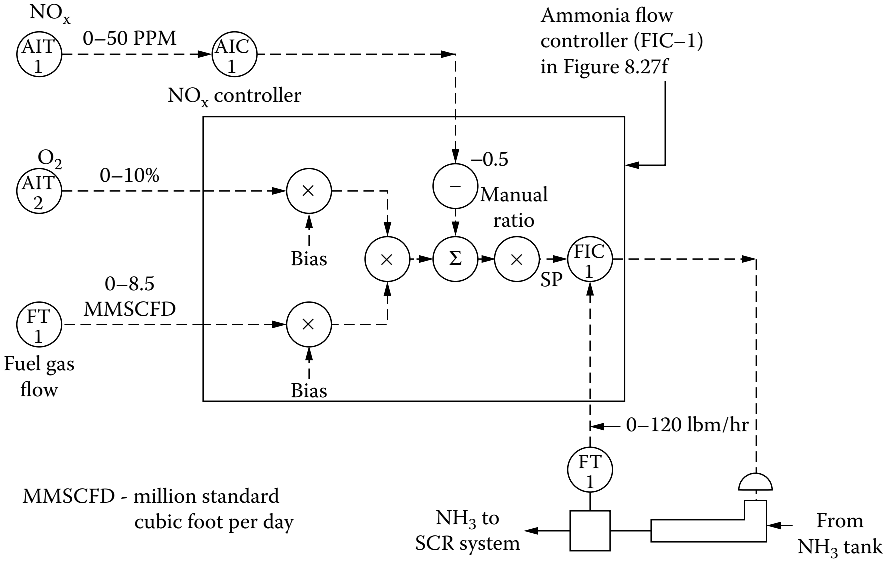


FIG. 8.27g
The configuration of the control algorithm of the ammonia controller (FIC-1 in Figure 8.27f) used to control the NO_x emissions of a furnace.

For the control algorithm shown in Figure 8.27g all measurement and control signals are scaled, so that the measurement output ranges of each transmitter or controller are converted to a 0–1 scale. For example, the output of the NO_x analyzer controller (AIC-1) is also 0–1, and when it is 50% (a value of +0.5), by subtracting −0.5, its influence on the set point of FIC-1 is eliminated. On the other hand, when the feedforward portion of this algorithm cannot correct for the changes in the SCR process, the feedback controller (AIC-1) will start moving its output away from 0.5 and will start correcting the FIC-1 set point.

Such an algorithm can be implemented either by a single-loop programmable controller or by the DCS system.

For a particular set of conditions, the following calculations can serve as an example of the operation of this control scheme:

At an average NO_x concentration of 110 ppm, the total unabated NO_x flow rate in the stack has been calculated to be 26.7 lbm/hr. The conversion of 1 mole of NO_x (Mw = 47) requires about 1 mole of NH_3 (Mw = 17). Thus, the rate at which pure ammonia should be added is 17/46 × 26.7 = 9.87 lbm/hr NH_3. Ammonia is received in 19% solution; consequently, this solution should be injected at a rate of 9.87/0.19 = 52 lbm/hr.

It was found that excess oxygen also contributes to NO_x formation and that up to 4% more NO_x is formed for each additional percentage of excess O_2 over the concentration of 2%. Because of draft limitations, the furnace in this example operates at a 5% excess oxygen level (3% over the 2% minimum). Therefore, 3 × 4% = 12% added NH_3 is required, which comes to 1.12(52) = 58.2 lbm/hr.

To measure the ammonia flow rate, a Coriolis meter with a range of 0–120 lbm/hr can be used, so that the 58.2 lbm/hr rate would approximately correspond to 50% of range.

The "manual ratio" correction factor applied to the set point of FIC-1 is available for the operator to correct for such differences as the calculated 58.2 lbm/hr and the 50% range of the flow meter.

Continuous Emission Monitoring Systems In order to comply with the requirements of the federal agencies (see "Standards and Guidelines" at the end of this section) and with local air quality district requirements, microprocessor-based, self-calibrating CEMS analyzers are used. The CEMS microprocessor calculates the pollutant concentrations reduced to 3% O_2 (as a measure of gas dilution) and provides rolling averages, peaks, their duration, and other records as required by the local agency. The CEMS is usually wired directly to the agency's computer. Emissions are referenced to the firing rate (gas flow) to provide the total mass of pollutants emitted.

For gas-fired heaters, the pollutant that is continuously monitored is only NO_x. In case of fuel oil-fired boilers and furnaces, SO_2 is also monitored. Other pollutants, such as CO, CO_2, and particles (opacity) are only checked periodically by portable instruments. EPA 40CFR60 provides specific criteria for CEMS.

Described here is a typical CEMS system used on a refinery furnace with selective catalytic reduction-based NO_x abatement controls.

The CEMS can utilize a zirconium oxide-based stack oxygen analyzer for firing controls and a chemiluminescent NO_x analyzer.

The CEMS provides NO_x readings (corrected to 3% O_2) and an O_2 signal for the ammonia injection controller (Figure 8.27g). The data collected by the CEMS is also used by the local air quality agency and by the plant monitoring system (PMS), using Equation 8.27(2). The O_2-corrected NO_x emission (E) is calculated on the basis of the actual measured concentration O_2 concentration (C).

If sampling-based analyzers are used (and not directly inserted probe-type analyzers), the components of CEMS systems can include:

- Stack sample probe, located usually $^2/_3$ up the stack (Figure 8.27h)
- Heated sample line with thermostat
- Sample conditioning cabinet with sample pump
- O_2 and NO_x analyzers
- Microprocessor with auto-calibration controller and solenoid valves
- Rack of calibration gas bottles

Sampling The stack sample probe is a stainless steel tube that extends into the stack and has a HastelloyC, 0.5-μ-size filter, a test and sample gas connection, and an electric heater with RTD. The sample line is heated to 300°F to keep moisture from condensing. An alternative method is to use a dilution extractive sample probe (Figure 8.27i). This method of sampling has the advantage of not requiring a heated sample line or a chiller/dryer.

In the standard sampling system, a sample conditioning cabinet is provided, where the sample is quickly chilled to knock out water without washing out the gases of interest. These components are usually followed by a gas dryer, then a bypass rotameter, which is used to control the sample gas flow rate to the NO_x analyzer.

A solenoid manifold is also provided, so that auto-calibration can be performed every 24 hours, by injecting zero and span gases. Sample system alarms are also provided to signal such conditions as high sample gas moisture content, low sample line temperature, and high condensate level.

NO_x Analyzer In the chemiluminescent NO_x analyzer, the reaction that takes place is described by Equation 8.27(11):

$$NO + O_3 \rightarrow NO_2 + \text{light @ } 0.6\text{–}3\mu \qquad \textbf{8.27(11)}$$

The intensity of light emission is proportional to the NO concentration in the sample. The chemiluminescence analyzer is usually provided with a catalytic converter to reduce the NO_x to the NO form and then uses its ozone generator to measure NO_x in the range of 0–50 ppm NO_x. The main components of this analyzer are shown in Figure 8.27j.

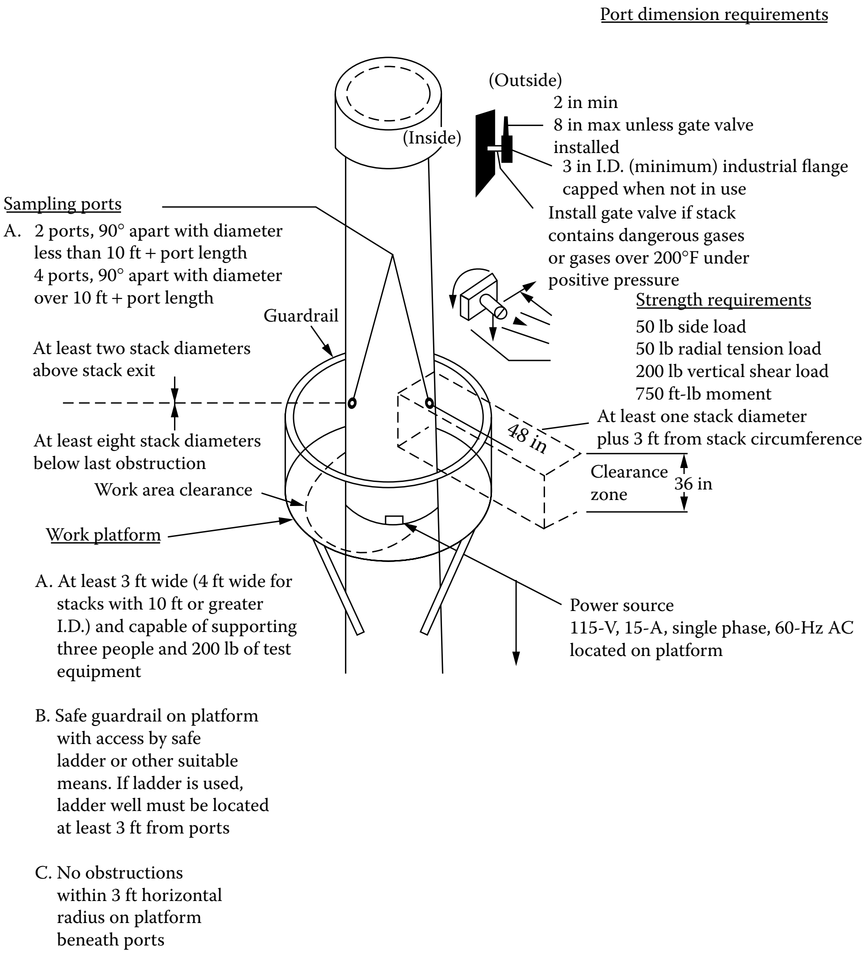


FIG. 8.27h
The location of a stack sampling probe.

Analyzers for Furnace Control

In addition to the analyzers discussed above, in connection with pollution abatement-related furnace analyzers, are the analyzers that are used in controlling the furnace's operation. These include oxygen analyzers, combustibles detectors, and fuel heating value (calorific) analyzers.

Oxygen and Excess Air Analyzers An oxygen analyzer is commonly used on furnaces to measure the oxygen content of the flue gas. From this information, the amount of excess air being delivered to the furnace can be computed. Keeping the excess air at a minimum results in conservation of heat, because the sensible heat content of the excess air is discharged into the atmosphere and is an outright heat loss.

The efficiency of combustion can be maximized by throttling the air of the furnace to maintain the oxygen content of flue gas at a desirable value. The analyzer is also tied to a low-oxygen alarm, which warns the operator if hazardous furnace atmospheres are developing.

The decision whether to install an oxygen analyzer is based on economic considerations, and energy efficiency usually justifies it. The cost of the analyzer and its upkeep must be balanced against the energy savings and increased safety. If the firing load or the type of fuel to the furnace often changes, the oxygen analyzer is a necessity from a safety standpoint alone.

The O_2 sensor can be a solid-state heated zirconium oxide probe that can be placed directly in the stack (Figure 8.27k). An instrument air purge connection is recommended for periodic blowing. At temperatures of over 600°F, these probes produce an open circuit voltage that is related to the partial pressure of oxygen in accordance with the Nernst Equation shown in Equation 8.27(12):

$$E = \frac{RT}{nF} \ln \frac{O_2\text{partial pressure in reference gas}}{O_2\text{partial pressure in sample gas}} \qquad \textbf{8.27(12)}$$

where

E = the open-circuit voltage developed
R = the universal gas constant
T = the temperature
n = the number of electrons transferred per molecule of oxygen
F = Faraday's constant

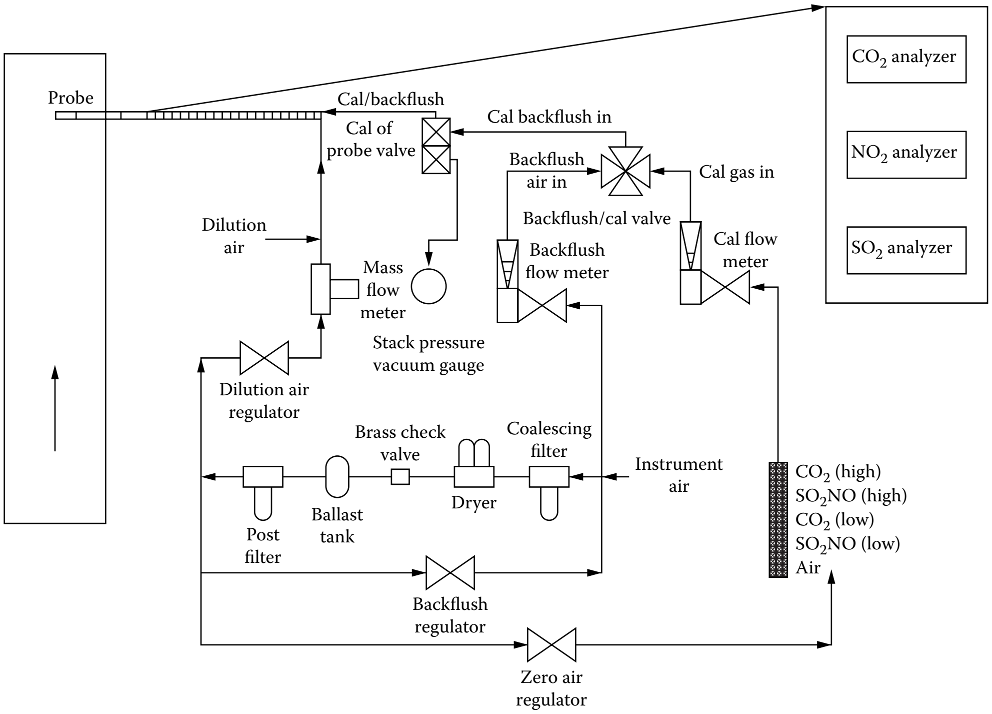
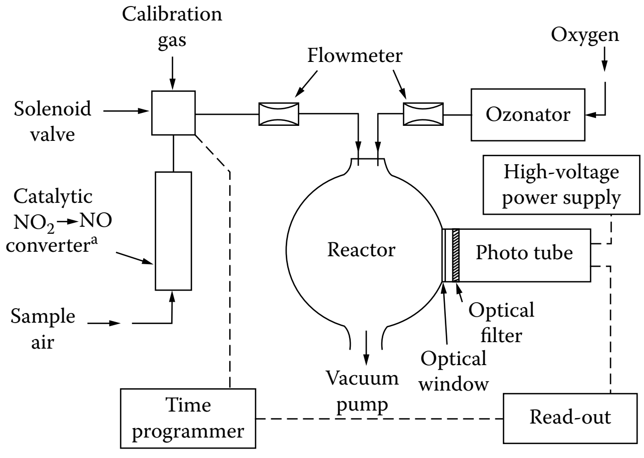


FIG. 8.27i
The components of a dilution-extractive sampling system.

Paramagnetic oxygen analyzers are less frequently used. They can be installed on the sides of fireboxes, to check for potentially explosive fuel-rich pockets. The sample gas is drawn to these analyzers by an air aspirator provided with a water seal.

Combustible Analyzers Combustible analyzers measure the amount of unburned fuel in the furnace flue gas (CH_4, CO); they are mostly used as portable instruments during emissions certification. For an in-depth discussion of combustible analyzers refer to Section 8.16 in Chapter 8 of the first volume of this handbook.



[a]Converter can be omitted if only NO is measured

FIG. 8.27j
Chemiluminescence nitric oxide analyzer.

The concentration of one combustible gas component, carbon monoxide, is sometimes used as a constraint variable in a “low select” configuration, in connection with the excess oxygen controller as described in connection with Figure 8.6bbbb in Section 8.6. The CO analyzer can be of the infrared type and can be installed to look “across the stack.”

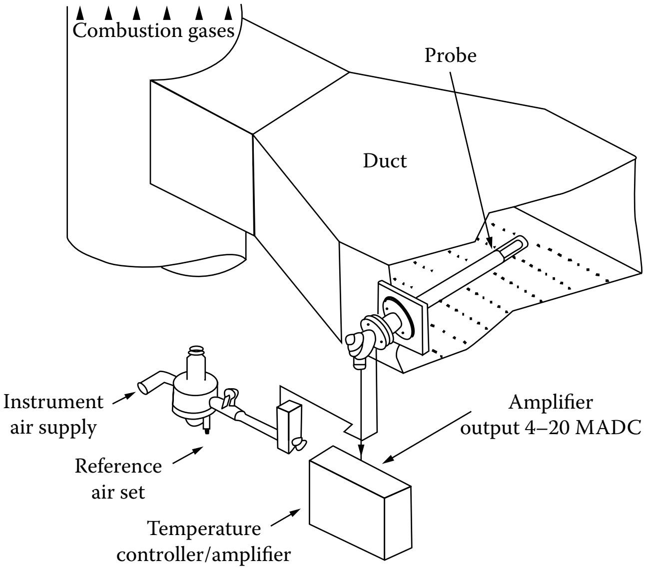


FIG. 8.27k
High-temperature, electrochemical, zirconium oxide-based oxygen analyzer probe.

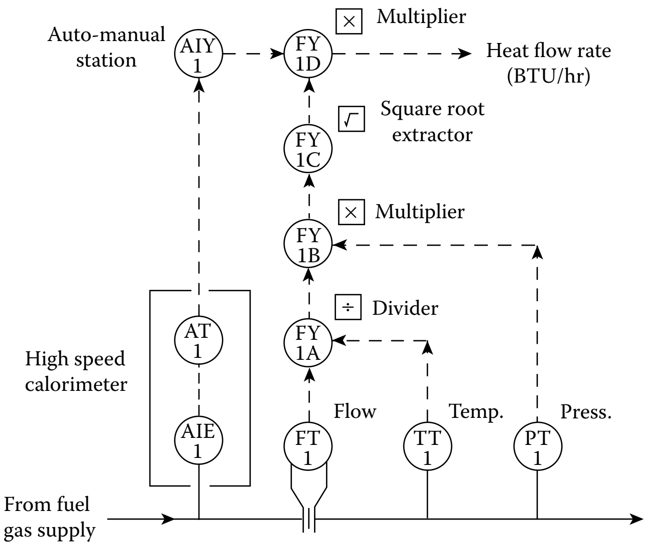


FIG. 8.27l
Heat flow rate detection loop.

NH_3 Analyzers Wet chemical NH_3 analyzers are used to detect the ammonia slip when selective catalytic reduction is applied for emission control. Excess ammonia in the flue gas is referred to as "ammonia slip" and must be avoided. NH_3 is difficult to sample because it is very easily dissolved in water. Experimental laser detectors, beaming across the stack, have been used as ammonia analyzers with some success.

Fuel Gas Heating Value When the refinery waste gas and utility gas are combined, the mixture resulting from their varying heating values will also have a varying heating value. BTU analyzers are used to measure the heat flow rate (Figure 8.27l) provided by the fuel being fed to the furnace (see Section 2.4 in Chapter 2 in Volume 1 of this handbook).

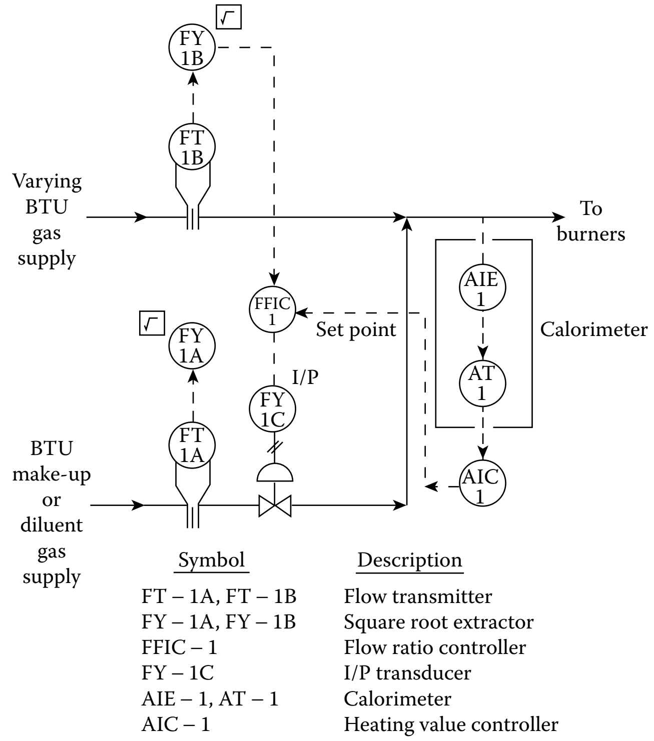
In order to stabilize furnace firing, the heating value of the fuel should be kept constant. This can be achieved by blending refinery gas with natural gas, propane, or some other supplemental fuel having a high and constant heating value. If the heating value of the total stream drops, more of the supplemental fuel is blended in.

This BTU instrument (calorimeter) can also serve as a safety device by actuating an alarm when the heating value of the fuel drops to the point where it cannot support combustion.

Both the blending of the fuels and the emergency shutdown controls can be automated (Figure 8.27m). In this control configuration, if the BTU analyzer detects a drop in the heating value, the fuel supply is shut down.

If the heating value of the two gases that are blended are known, but their ratio varies, the resulting variation in the total heating value can be compensated for by a gas blending system on the main fuel gas header.

Relatively inexpensive BTU analyzers are available for "sweet" gases. They operate by burning a sample stream of fuel and measure the oxygen concentration of the combustion products. The BTU measurement so obtained can be used in a feedforward manner, to compensate for the varying BTU content of the fuel values.



Symbol	Description
FT – 1A, FT – 1B	Flow transmitter
FY – 1A, FY – 1B	Square root extractor
FFIC – 1	Flow ratio controller
FY – 1C	I/P transducer
AIE – 1, AT – 1	Calorimeter
AIC – 1	Heating value controller

FIG. 8.27m
Control system to automatically maintain the heating value of mixed gas streams.

Figure 8.27n illustrates a fuel gas blending controller. The formula it uses to arrive at the total heating value of the blended stream is

$$HT = (F1H1 + F2H2)/ FT \qquad \textbf{8.27(13)}$$

where:

HT and FT are the total BTU content and total flow of the mixed fuel
H1 and F1 are the flow and the BTU content of fuel 1
H2 and F2 are the flow and the BTU content of fuel 2

Furnace Instrumentation

While the control hardware and software are constantly changing, today's furnaces in larger plants such as refineries are usually controlled by DCS systems, with independent PLCs used for safety shutdown. In case of some furnaces, PLCs with human-machine interfaces (HMIs) are also used in place of DCS systems, as are graphic software-supported PC displays located in the control room and in management offices. Smaller furnaces can be controlled by single-loop analog or digital PID controllers.

Both the DCS systems or the microprocessor-based unit controllers provide adaptive gain, auto-tuning, feedforward, and the capability of decoupling the interacting loops. Historical trends and temperature profiles can be graphically displayed,

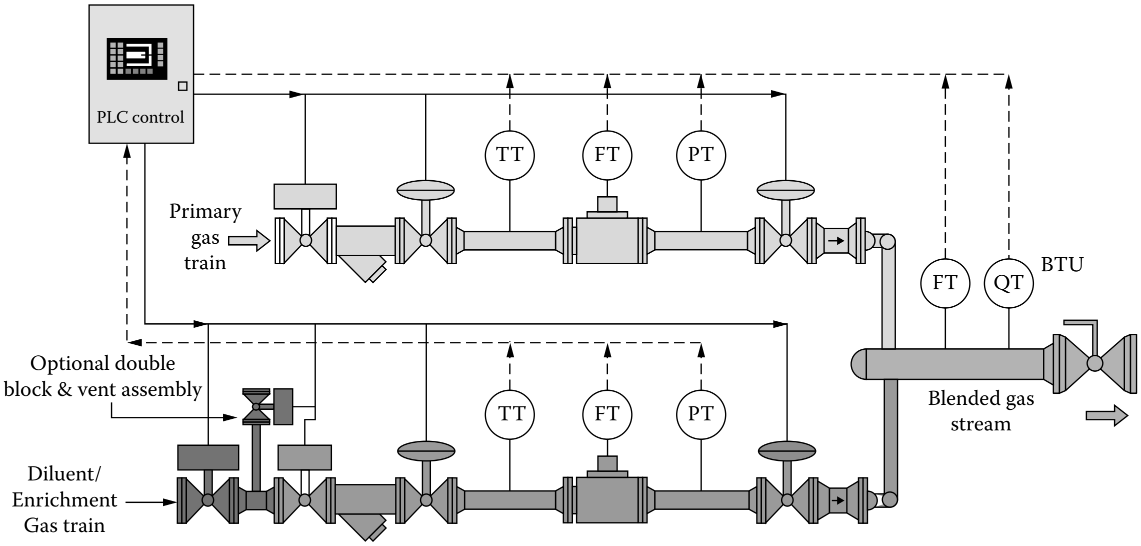


FIG. 8.27n
Pressure- and temperature-compensated fuel gas blending control system used when the BTUs of the blended streams are known, but their ratio varies. This typical controller also handles the safety shutdown interlocks.

and alarm conditions and status summaries can be shown on graphic displays with event printouts. Nuisance alarms can be eliminated by the use of smart alarm logic, and diagnostic messages can be generated for operating and maintenance information. All of these capabilities contribute to reducing the frequency of furnace shutdowns and increase its safety.

Transmitters and Valves Mostly 4–20 mA analog transmitters are used with the digital HART protocol superimposed over the analog signals to the DCS. In order to save on wiring costs, newer installations use Fieldbus networks, but their reliability for critical systems is still under evaluation.

Most field devices and their wiring need to be intrinsically safe or explosionproof, approved for Class 1 Div. 2 Gr. B, C, D. Intrinsically safe units are provided with current-limiting barriers. Most transmitters and I/P or E/P converters are provided with NEMA4X/7 explosionproof enclosures. For analyzers, purged enclosures are also used with proper safeguards (see ISA Standard S12.4).

Field instrumentation should not be located close to potential fire sources, such as on furnace walls. Their wiring should be routed from furnaces to junction boxes or from DCS I/O cabinets that are at a safe distance from potential fire sources, so that in case of a furnace fire, only the wiring to the junction boxes will need to be replaced.

Final control elements, such as control valves or dampers, should be selected so their gain variation will compensate for the variations in the process gain, with the goal being a constant loop gain over the range of the valve openings. Equal-percentage (gain increasing with load) valve characteristics should be used when the process gain drops as the load rises, in order to keep the loop gain at around 0.5.

Control valves should be provided with positioners, in order to overcome packing friction-caused hysteresis, but it is important to make sure that the positioner is faster than the controller that sets it. Newer E/P positioners also provide information on valve diagnostics such as packing problems.

Firing shutdown valves should always be separate from fuel control valves and be of the tight shut-off type with fire-resistant design. Usually, ball valves with steel seat rings are used for shut-off services. One should also consider the use of fireproof insulation of the valve actuators and of the air tanks on air-fail switch valves.

Flame Scanners Only some refinery furnaces are provided with flame scanners. Ultraviolet, visible, and infrared designs can be considered. Ultraviolet flame detectors measure the flicker of a flame in the UV band and are not blinded by the hot refractory (Figure 8.27o). Infrared sensors are able to

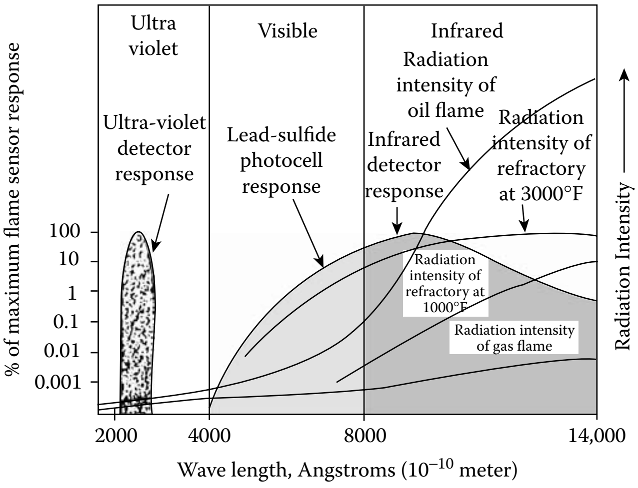


FIG. 8.27o
Applicable ranges of selected flame sensors and the range of hot refractors effect.

penetrate through the flame by-products such as smoke, atomized oil, or steam. Instrument air for purging and sometimes cooling water is also required for these sensors.

FURNACE TYPES

Start-Up Heaters

Start-up heaters are discussed here first because of their simplicity. These units are required at the start-up of a process unit, and their use lasts from a few hours to a few weeks. They are usually vertical, cylindrical units with vertical process tubes along the inner walls and with a single burner centered in the floor. Draft is normally by natural convection induced by a stack mounted on the top of the heater.

Processes that require start-up heaters include catalytic cracking units to heat up fluidized catalyst beds, ammonia units to heat up the ammonia converter catalysts, and fixed-bed gas-drying units to regenerate the dryer beds.

The start-up heater usually heats an intermediate stream, such as air or natural gas, which in turn heats another fluid or solid, such as a reactor catalyst. For example, the start-up heater for an ammonia unit is used to heat "synthesis gas" (primarily a mixture of hydrogen and nitrogen), which in turn is used to heat the catalyst bed in the ammonia converter to a temperature of 700°F (371°C).

The synthesis gas is recirculated through the catalyst bed, gradually bringing it up to operating temperature. After normal operation has begun, the exothermic nature of the ammonia synthesis reaction keeps the bed at a certain temperature, and the start-up heater can be shut down.

Process and Firing Controls The important variables are the flow and temperature of the "synthesis gas." Figure 8.27p shows the necessary controls for the operation of this unit. The flow of the cold synthesis gas is measured by the flow transmitter (FIT-1) and is indicated on a flow indicator (FI-1). The desired gas flow is manually adjusted by operating the hand valve (HV-1), while observing the local flow indicator (FIT-1). The effluent synthesis gas temperature is maintained by the temperature controller (TIC-1), which uses a thermocouple to measure the gas temperature and controls the flow of fuel gas by modulating the control valve TV-1.

The fuel gas firing rate is set by process temperature controller (TIC-1). The heater draft (i.e., negative pressure in the firebox) is produced by the stack; the operator sets it by observing the draft gauge (PI-1) and by manually adjusting the position of the stack damper. Once initially set, the damper is rarely adjusted again, unless the furnace conditions or loads change drastically.

Safety Controls Start-up heaters are usually controlled by simple PLCs or relay systems. Most start-up heaters have flame scanners. Their SIS system requirements have already been described in a previous paragraph.

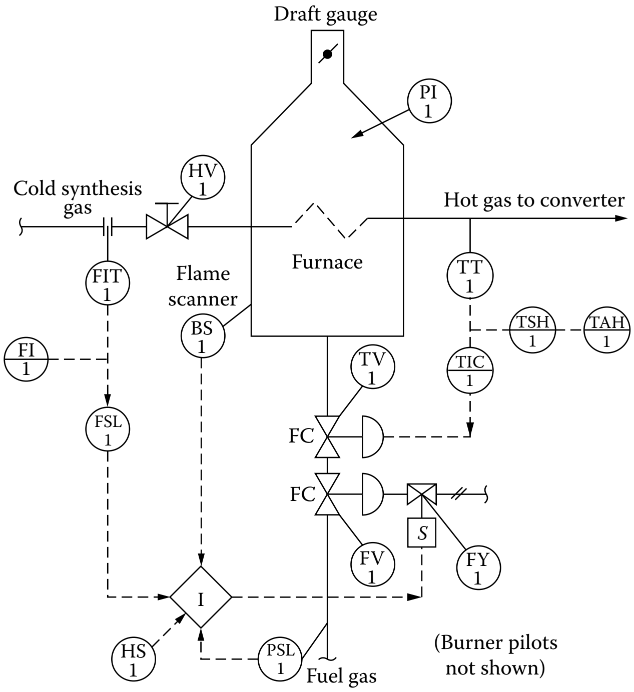


FIG. 8.27p
The controls of a start-up heater.

Fired Reboilers

The fired reboiler provides heat to a distillation tower by heating the tower bottoms and vaporizing a portion of it. Normally, a tower reboiler uses steam or another hot fluid as a source of heat, but where heat duties are great or where tower bottom temperatures must be high, a fired reboiler can be used. Depending upon its size, the reboiler may be of the vertical cylindrical type or the larger, conventional horizontal-type furnace.

The fired reboiler heats and vaporizes the tower bottoms as this liquid circulates by natural convection through the heater tubes. The coils are generously sized to ensure adequate circulation of the bottoms liquid. Temperature of the reboiler return fluid is generally used as the means of controlling the heat input to the tower, provided that the bottoms product material has a wide boiling-point range. Overheating the process fluid is a contingency that must be prevented, because most tower bottoms will coke or polymerize if they are subjected to excessive temperatures for some length of time.

Process and Firing Controls A common control scheme is shown in Figure 8.27q, which depicts the tower bottom along with the fired reboiler. It is usually not practical to measure the flow of tower bottoms to the reboiler, first, because the liquid is near equilibrium (near the flash point), and second, because it is usually of a fouling nature, tending to plug most flow elements.

Proper circulation of the fluid is provided in the careful hydraulic design of the interconnecting piping. An important variable is the reboiler return temperature, which is controlled by TRC-1 throttling the fuel gas control valve. The high-

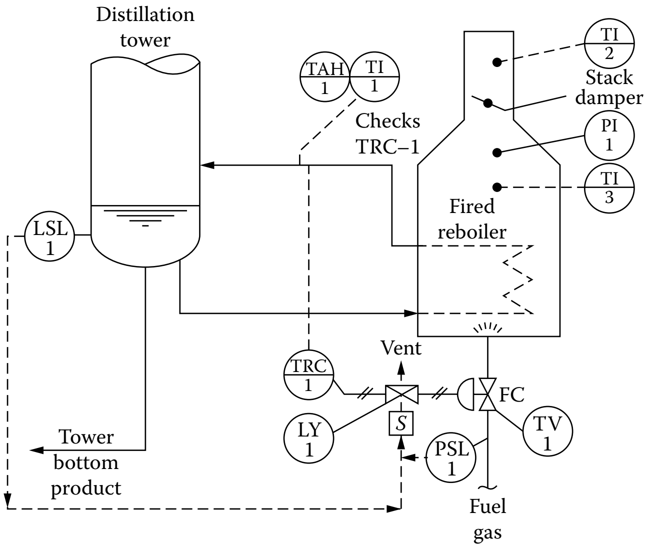


FIG. 8.27q
Fired reboiler controls.

temperature alarm (TAH-1) is provided to warn the operator if the process fluid reaches an excessive temperature.

The reboiler return temperature is actually an inferred indication of the percentage of vaporization. The sensitivity of TIC-1 depends upon the boiling-point range of tower bottoms. A wide boiling-point range provides adequate sensitivity. However, a narrow range does not control well. Often, for these cases, a differential pressure controller is used in place of the reboiler temperature to infer percentage of vaporization. A differential pressure controller (DPIC, not shown) offers increased sensitivity to firing for narrow-range boiling-point fluids.

The firing controls of fired reboilers are relatively simple and are similar to those of the start-up heater. The process temperature controller (TIC-1), or tower differential pressure controller, sets the set point of the gas pressure or flow controller. The fuel in this case is gas, but it could just as well be fuel oil. The furnace draft can be manually set by means of the operator throttling the stack damper while observing the draft gauge (PI-1). The stack temperature (TI-3) and the tube skin temperatures are monitored as checks for excessive temperatures that may develop during periods of heavy firing.

Safety Controls The major dangers in this type of furnace are caused by either the interruption of process fluid flow or by the stoppage of fuel. The loss of process fluid can occur if the liquid level in the tower bottom is lost. If this happens, flow will stop in the reboiler tubes, and a dangerous overheating of these tubes may result. To protect against this, the low-level switch (LSL-1) is wired to close the fuel gas valve (TV-1). For a more detailed description of the required shut-off safety system, refer to Figures 8.27c and 8.27d.

Process and Crude Oil Heaters, Vaporizers

The feed heater of a refinery crude unit is representative of this class of furnaces (Figure 8.27r). Crude oil, prior to distillation into the various petroleum fractions (gasoline, naphtha, gas oil, heavy fuel oil, and residual) in the "crude tower," must be heated to around 750°F and partially vaporized. The heating and vaporization are done in the crude heater furnace, which consists of a firebox with preheat coils and vaporizing coils.

Larger-duty heaters usually have multiple zones encompassing multiple passes, and they heat the oil inside the coils in the convection section of the furnace. This is the portion that does not see the flame but is exposed to the hot flue gases on their way to the stack. The vaporizing takes place at the end of each pass in the radiant section of the furnace (where

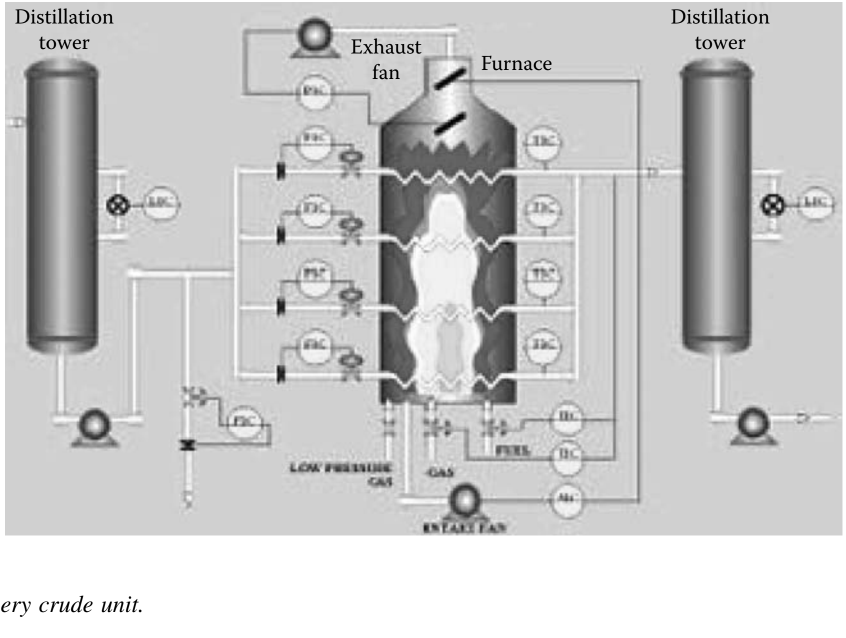


FIG. 8.27r
The feed heater of a refinery crude unit.

the coils are exposed to the flame and the luminous walls of the firebox). The partially vaporized effluent then enters the crude tower, where it is flashed and distilled into the desired "cuts."

Other process heaters that fall into this category are refinery vacuum tower preheaters, reformer heaters, hydrocracking heaters, FCCU feed heaters, and dewaxing unit furnaces. The control system presented is also applicable to these types of process heaters.

Process Controls The prime variables of this process that need to be controlled are

1. The flow of the feed flow to the unit
2. The proper splitting of the total flow into the parallel paths through the furnace, in order to prevent overheating of any one of the streams and to protect against its resultant coking
3. The correct amount of heat supplied to the process stream

Figure 8.27s shows the typical process controls for this type of furnace. The crude feed rate to the unit is set by the flow controller (FIC-4). This flow is split into the parallel paths of the furnace by adjustment of the manually set control valve openings by the hand indicating control (HIC) stations (HIC-1 through HIC-3).

The temperature indicators TI-1 through TI-3 (there may be several TIs on each pass) are periodically observed to determine if unbalanced temperatures are developing in any one of the passes. If such a trend develops, the flow through that pass is altered slightly to drive its temperature back toward the desired norm.

The outlet temperature controller (TIC-4) of the combined crude stream sets the firing rate through a cascade fuel flow or pressure controller (not shown).

However, the desired heat input into the feed stream can be more difficult to control, because the effluent of the furnace is partially vaporized and the feedstock varies in composition depending upon its source. If the feed were only heated and no vaporization took place, the control would require only that the effluent temperature be maintained. If complete vaporization and superheating occurred, this too could be handled by straight temperature control.

In the case of partial vaporization, combined with a variable-feed composition, effluent temperature control alone is not sufficient for reliable control. The composition and, hence, the boiling-point curve of the feed varies with time, and the required control temperature itself also varies. Therefore, additional information is required, and it is obtained from the distillation process, which is downstream of the furnace. By observing the product distribution from the fractionation, a need for changing the heat input can be determined.

Current practice is to achieve approximate control with a temperature controller (TIC-4) whose set point is periodically changed by the operator to account for feed variations. The operator depends both on experience and on the results in the fractionator (possibly a crude tower optimization strategy) to determine the proper temperature setting.

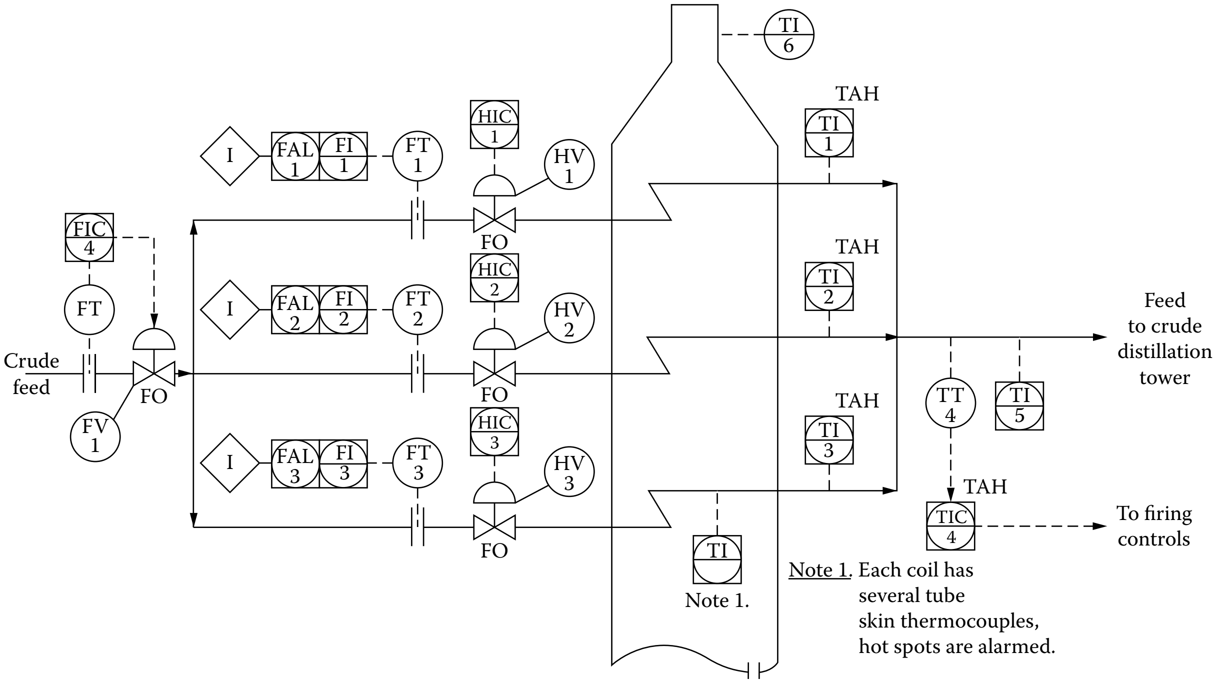


FIG. 8.27s
The controls of a crude heater-vaporizer, in which the flow distribution between the passes is manually adjusted.

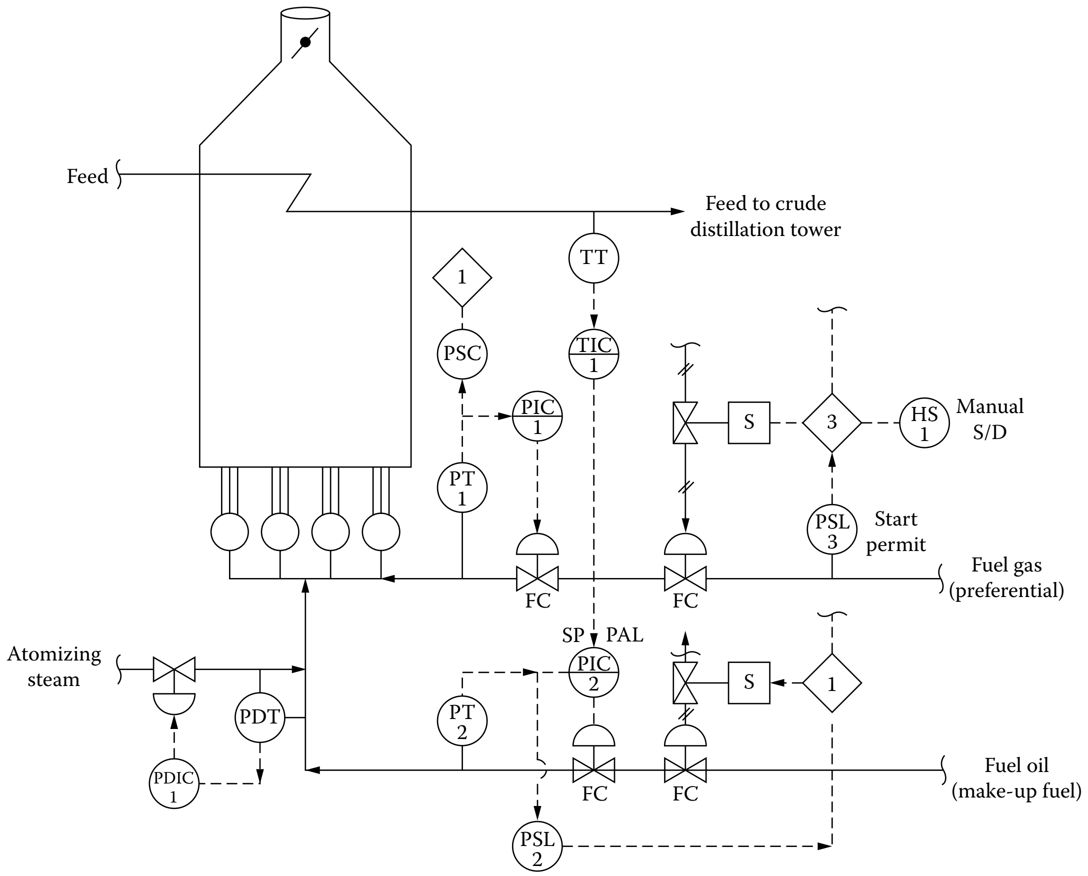


FIG.8.27t
The firing controls of a crude oil heater-vaporizer, in which fuel gas is the preferred fuel and oil is used as the make-up fuel.

Firing Controls

Figure 8.27t shows the firing controls for the furnace using fuel gas (or fuel oil as standby). The fuel gas headers serve many burners spaced equally along the floor of the firebox. The burners, being essentially fixed-diameter orifices, will pass more or less fuel depending on the header pressure. In the past, a pressure controller was often used in place of flow control. Because flow through an orifice is a function of the square root of the pressure, nonlinearity is eliminated by the use of pressure control.

Although fuel oil is the make-up fuel (trim medium), it must be available in sufficient quantity to take the whole load in the event of a fuel gas interruption. The temperature controller (TIC-1) varies the set point of the fuel oil header pressure controller (PIC-2) to satisfy process load requirements.

All burners have turndown ratio limitations. This is the ratio of minimum to maximum fuel burning capacity, which ratio is usually 3:1. Low pressure in the fuel oil header is indicative of the potential approach of a minimum firing condition. Consequently, the low-pressure switch (PSL-2 on PIC-2) is used to warn the unit operator when this situation arises. The operator then has the option of going out to the furnace and manually turning off some of the individual fuel oil burners or to reduce the fuel gas firing rate by lowering the set point of PIC-1. Either change will increase the fuel oil demand and, thus, bring the system back into a stable operating zone.

Newer installations use flow instead of pressure controllers (as shown in Figure 8.27u) on the fuel oil make-up to better control the heat balances. These controls are designed to meet a percentage of the total firing duty assigned to the hearth or floor burners to prevent furnace damage that can result from uneven firing. The percentage of the total firing duty that is assigned to the floor burners is a percentage that is recommended by the furnace manufacturer to protect metallurgy, reduce maintenance, and extend furnace life.

In Figure 8.27u, the effluent temperature controller (TIC-4 in Figure 8.27s) is the cascade master, and it adjusts the set point of FIC-5 to maintain the furnace effluent temperature. Cascading the TIC-4 to FIC-5 is beneficial in that the flow controller compensates for the disturbances (e.g., changes in fuel supply pressure) without allowing them to upset the effluent temperature, whereas the TIC-4 provides the correction required to correct the process for slower ambient changes.

Draft Control Furnace draft is normally maintained by free convection in the furnace (i.e., the stack produces a negative pressure in the firebox), and air is drawn in through louvers

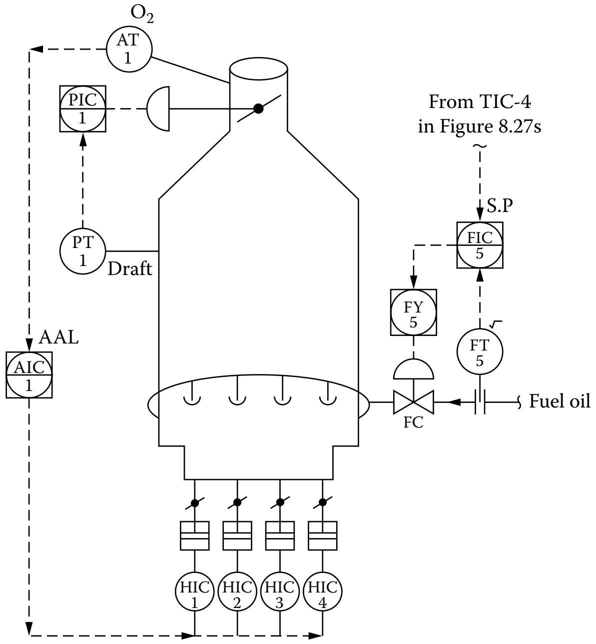


FIG. 8.27u
The firing controls of a crude oil heater-vaporizer with both stock and burner draft controls.

along the sides or bottom of the furnace. This is known as natural draft, because it is produced without mechanical means. Draft is controlled by pneumatically or motor-operated dampers in the stack.

Some furnaces use burner louvers with piston actuators for oxygen control (Figure 8.27u). Here, the overall furnace draft pressure is controlled by dampers in the stack, which are throttled by PIC-1, the draft controller, which strongly influences the O_2 content, because an increase in draft increases excess O_2. It is difficult to maintain the optimum (minimum) excess O_2 concentration in the flue gas without reaching dangerous pressure conditions at the furnace vault. Low excess O_2 and safe high arch/draft pressure limits are sometimes implemented in a duty constraint (cut-back) control configuration.

Damper characteristics are usually less than ideal, and linkages may stick. Installing powerful damper actuators will pay off in fuel savings. Newer furnaces with induced draft fans are usually provided with segmented damper blades, which have a near-linear characteristic. The nonlinearity of dampers can be compensated by using error-squared algorithms in their PID controllers.

Draft may violently oscillate, which can cause noise and furnace vibration. Therefore, measurement signal dampening is often required for PIC-1 in Figure 8.27u. The dampening filter should be so designed as to filter out fast pressure oscillations, but allow for the accurate measurement of all major pressure excursions in draft.

Changes in ambient temperatures, wind force and direction, or rain can also upset both the draft and the temperature controls. Draft controls set points should be selected with sufficient safety margins that a major storm will not cause the development of excessively low draft pressures.

If the furnace is large, or is provided with SCR NO_x abatement controls, forced draft fans that operate at positive pressures will serve the firebox. If the fans are located at the outlet of the combustion zone, they usually operate at below atmospheric pressures and are called induced draft fans. They usually are provided with variable-speed drives (VSDs).

Excess Oxygen Control The zirconium oxide stack oxygen analyzer (Figure 8.27k) that is used for combustion control can also be used to display the O_2 concentration next to the draft PIC (Figure 8.27u). This allows the operator to adjust the stack damper draft controller set point to keep the furnace pressure safely "on the negative side" and yet approach the optimum excess air concentration.

Attempts to reduce excess air by maintaining excess O_2 to 2–3% often fail, because of the draft limitations. Most furnace SIS systems are set to shut down if draft is out of limits for over 5 sec. The high-draft or low-pressure limit is usually set at any pressure below –1.0" W.C. (in order to prevent buckling of the steel or damaging the refractory lining). The high-pressure or low-draft limit is usually set at any pressure that is below 0" W.C., because overpressure could force the flames out and cause the burning of the sidings.

The sample tap (Figure 8.27h) serving the O_2 analyzer is often used to also bring a heated sample to the NO_x analyzer.

Safety Controls The primary sources of hazards in the operation of process heaters include the interruption of the charge flow, and the interruption of the fuel flow and resultant loss of flame. The reduction of crude charge rate below a minimum rate will result in overheating and possible tube rupture. Resumption of flow may cause hydrocarbon leakage into the firebox, with catastrophic results. The low-flow alarms (FAL-1, -2, or -3 in Figure 8.27s) alert the operator of this impending condition. If this happens, the operator has the option of correcting the fault or terminating the firing. If the instrument air supply fails, the control valves (FV-1 and HV-1 through HV-3) will fail open, thereby maintaining the flow through the furnace coils.

For the discussion of general safety requirements and the related PLC system design, refer to the earlier discussion on "Safety Instrumented Systems."

Reformer Furnaces

The purpose of a reformer in an ammonia plant is to produce hydrogen, which is used with nitrogen in the synthesis of ammonia (NH_3). Hydrogen is produced by reforming the hydrocarbon feed. This feed (usually methane or naphtha) is reformed with high-pressure steam, as shown in the reaction described in the equations below:

$$\underset{\text{(methane)}}{CH_4} + \underset{\text{(steam)}}{H_2O} = \underset{\text{(hydrogen)}}{3H_2} + \underset{\text{(carbon monoxide)}}{CO} \qquad \textbf{8.27(14)}$$

$$\text{General}: C_nH_{(2n+2)} + nH_2O = (2n+1)H_2 + nCO \qquad \textbf{8.27(15)}$$

The carbon monoxide is removed by further reaction later in the process. The reaction is endothermic (absorbs heat) and takes place at pressures of approximately 450 PSIG (3.10 MPa) and at temperatures of around 1500°F (815°C).

The reaction takes place as the feed gas and steam pass through tubes filled with a nickel catalyst, which is heated in the radiant section of the reformer furnace. Steam must be provided in excess of the reaction requirements to prevent the side reaction of coke formation on the catalyst. The coking of the catalyst deactivates it, requiring expensive replacement. To minimize coking, steam is usually supplied in a ratio of 3.5:1 by weight, relative to feed gas. Special precautions must be taken to maintain the excess steam at all times, because even a few seconds of interruption in the steam flow, while feed gas continues, can completely ruin the catalyst charge.

Process Controls The major controlled process variables are the feed gas flow, the reforming steam flow, and the temperature and composition of the effluent. As illustrated in Figure 8.27v, the feed gas flow is maintained by means of a pressure-compensated flow controller (FIC-1). Pressure compensation of flow corrects the measurement for fluctuations in feed gas pressure. The steam rate is maintained by means of FFC-2, and the ratio of steam to feed gas flow is continually monitored.

The ratio controller FFC-2 measures the ratio by dividing the gas flow signal from FT-1 with the steam flow signal from FT-2. If this ratio falls below the limit of approximately 3:1, a low flow ratio alarm (FAL) is sounded. If the ratio continues to fall below approximately 2.7:1, the feed gas is shut off by closing the valve (HV-1).

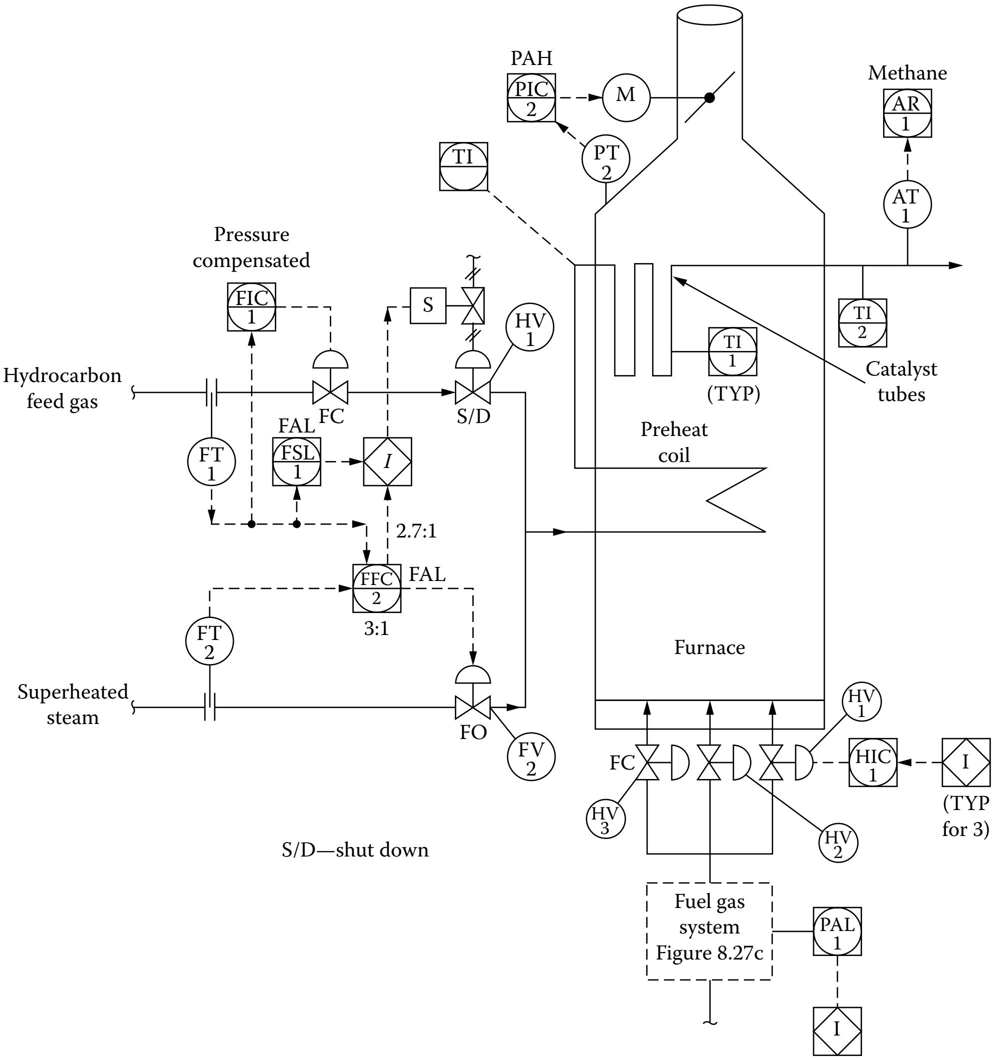


FIG. 8.27v
Reforming furnace process controls by manipulating the steam flow in ratio to the hydrocarbon feed.

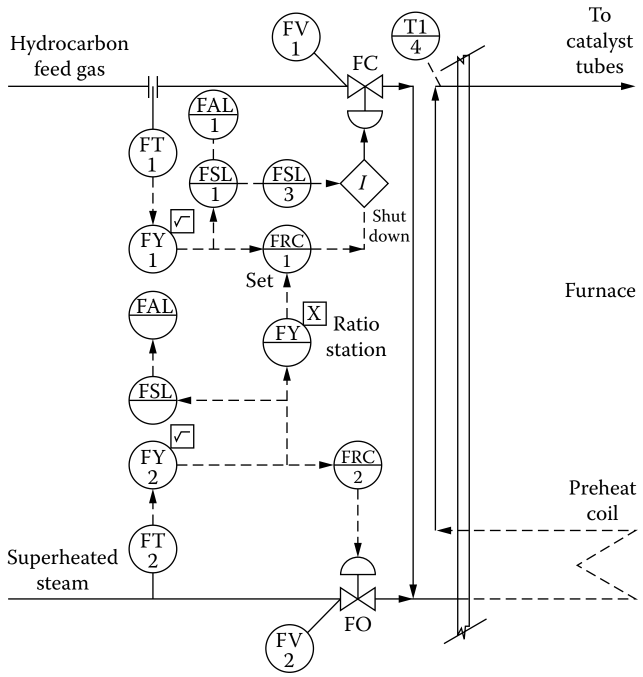


FIG. 8.27w
The reformer furnace control shown here (ratioing feed flow to the steam flow) is less stable than keeping the steam flow constant and ratioing the feed, as shown in Figure 8.27v.

If there is no separate shut-off valve (HV-1), the feed valve (FV-1) must be a quick-closing valve (4–5 sec for full closure) so that the flow of gas can be stopped almost instantly if the flow of reforming steam fails, thus protecting the reformer catalyst. If the shut-off valve HV-1 is used, but it is electric motor operated, FV-1 must also be a single-seated tight shut-off valve, to prevent leakage during the time while the electric-operated shut-off valve is closing.

An alternative to the controls shown in Figure 8.27v is illustrated in Figure 8.27w. Here, the steam flow is on straight flow control, and the hydrocarbon gas feed tracks the steam flow on ratio control. In theory, this appears to be an improvement over the previous control system, but in practice, the dynamics of the measurements and of the process make it less stable. The noise in the steam flow measurement and lags may result in cycling of the gas flow set point. Therefore, the integrated error in the gas flow obtained by the control strategy in Figure 8.27w can be considerably greater than when Figure 8.27v is implemented.

Manipulation of the steam instead of the feed gas is the most common method of alleviating this problem.

The effluent analyzer (AT-1 in Figure 8.27v) is used to determine reaction completion by measuring the methane (CH_4) content of the stream. Manipulating the profile of the furnace temperature can serve to achieve the desired degree of conversion.

This analyzer (AR-1) can be either an infrared or a chromatographic analyzer. The analysis is relatively simple, involving the measurement of 0–10% methane in a background of hydrogen and carbon monoxide. The only difficulty is the high water content in the stream, caused by the presence of excess steam used in the reaction. Therefore, water removal devices are required in the analyzer sampling system.

In newer DCS controls, the rates of competing reactions can be predicted so that the firebox and the optimum heat input profiles can be computed and set.

Firing Controls Firing of a reformer furnace, because of its massive design and great heat inertia, is often manually controlled. The furnace has approximately a dozen fuel headers with about 20 individual burners per header. Process temperature indicators TI-1 and TI-2 (on Figure 8.27v) are provided at the exit of the reaction tubes. These are constantly monitored, and periodic adjustment of firing is made by manipulation of the fuel header control valves (HV-1, HV-2, HV-3) in Figure 8.27v. The pressure upstream of the header control valves is controlled, so that once the valve stroke is set, fuel flow remains constant.

The fuel used is usually natural gas, with a small amount of purge gas from the NH_3 synthesis loop blended in.

Draft is usually maintained in these furnaces by a steam turbine-driven induced draft fan. The furnace draft (negative pressure) in the firebox is controlled by adjusting the fan speed. The pressure indicator controller (PIC-2) measures the pressure in the furnace (via PT-2) and controls the speed of the fan by adjusting the fan turbine governor to hold the furnace pressure at the desired setting. The draft points PI-1 through PI-4 are used to manually set the openings of the air inlet louvers, on the side of the furnace, to balance the drafts at various points in the furnace.

Safety Controls SIS controls are essentially the same as for the other furnaces described previously.

Upon instrument air failure, the gas feed valve (HV-1 in Figure 8.27v) closes to prevent coking the reformer catalyst. The fuel valves (HV-1, HV-2, and HV-3) close to stop firing during emergencies. The steam valve (FV-2 in Figure 8.27v) is a fail-open valve to maintain the flow of cooling steam through the furnace coils during such emergencies.

Cracking (Pyrolysis) Furnaces

An example of a cracking furnace is the ethylene pyrolysis furnace. Feedstock, which can vary from heavy gas oil, LPG, propane, to ethane, is preheated and vaporized in a preheat coil in the furnace and is then mixed with steam and cracked.

Steam is added to the feed in a fixed ratio to the hydrocarbon to reduce the partial pressure of the hydrocarbon feed. This tends to maximize the amount of olefins produced and to minimize the coke build-up in the coils.

The feed is heated to 1500°F (815°C) in the pyrolysis coils, which causes the cracking of the long chain hydrocarbons into shorter chain molecules and initiates the forming of such unsaturated (olefin) molecules as ethylene. The severity of cracking is dependent upon the temperature achieved and upon the residence time in the pyrolysis coils. Therefore, the distribution of furnace products is dependent upon the degree of firing and upon the temperature profile in the furnace. Effluent from the furnace is quickly quenched to prevent recombination of products into undesirable polymers.

Process Controls The important variables to be controlled in order to obtain the desired effluent product distribution are hydrocarbon feed flow, steam flow, coil temperatures, and firebox temperature. The firing controls discussed in connection with the previously discussed furnaces are also applicable here.

As shown in Figure 8.27x, the charge to the cracking furnace is determined by the flow controls on each of the individual passes rather than by total flow control. It is important to keep the flow through the coils constant, because coking causes a gradual build-up in the pressure drop through the coils and reduces both the flow and the heat transfer.

If the feed distribution were left purely to hydraulic splitting, the coking would start in one coil and reduce the flow through that coil, which in turn would cause overheating of that coil, producing even more coking. Eventually, the flow would be reduced to such a low rate that the overheating of the coil would cause its melting and rupture. The individual flow control valves serve to introduce variable pressure drops, which are adjusted to guarantee that all coils will coke at an almost equal rate.

The steam flow controllers (FIC-7 is shown as typical for all three) regulate the steam flow to match the total flow of hydrocarbon feed, controlled by FIC-4. The temperature controller (TIC-4) sets the firing controls by adjusting the total heat input to the process and brings the effluent temperature to the desired value.

The temperature indicator points (TI-1 through TI-6, as well as the tube skin temperature sensors) are monitored by the process operator to maintain a certain furnace temperature profile and, hence, a certain product distribution in the furnace effluent. The temperature relationships are accomplished by manually trimming burners at the required places in the firebox.

In order to determine whether the desired product specifications are being met, an analyzer is installed in the furnace effluent (AT-1). This analyzer is usually a chromatograph, though mass spectrometers could also be used, if the speed of response of this analysis was critical. This analyzer measures most of those components in the effluent stream that are lighter than butane.

This analysis is a difficult one, primarily because of the sample handling requirements. The sample has a high water content, and the water must be condensed and removed before the sample entrs the analyzer. The sample also has a

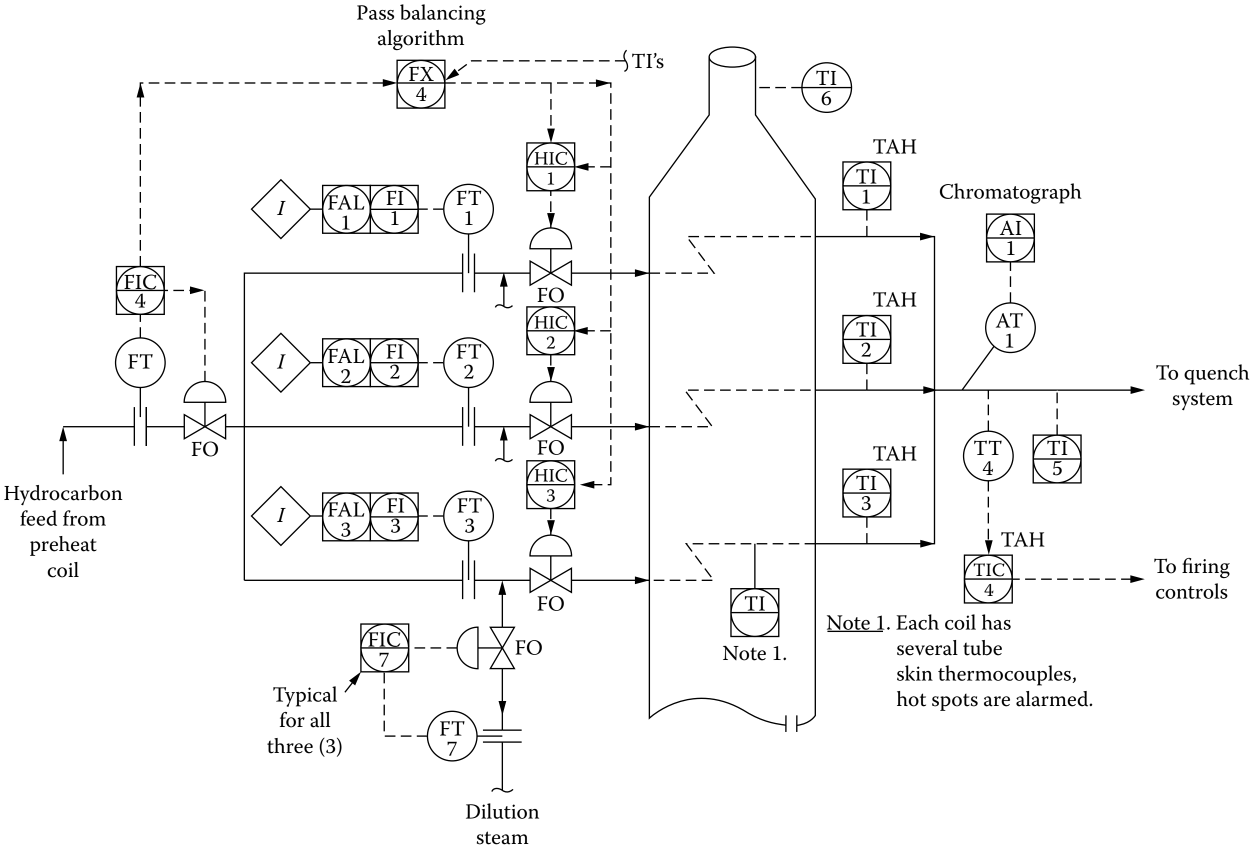


FIG. 8.27x
Pyrolysis furnace process controls.

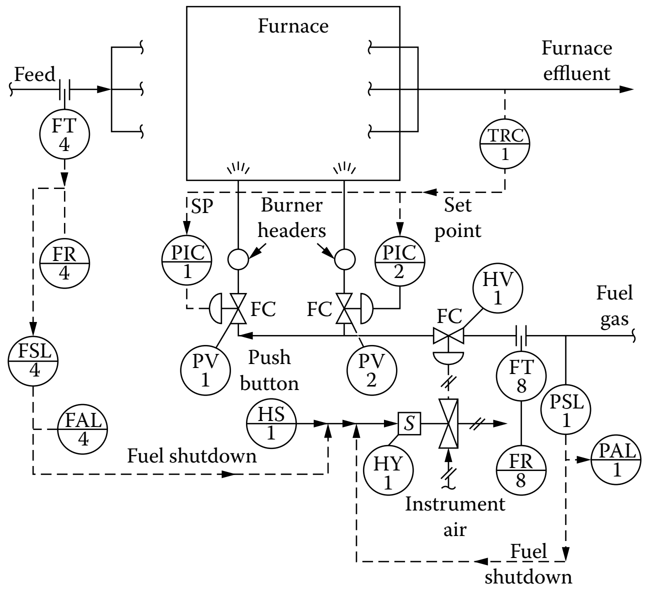


FIG. 8.27y
Pyrolysis furnace firing controls.

large amount of entrained coke and tars, which likewise must be eliminated (scrubbed out).

Safety Controls The major hazards in operating cracking furnaces are the interruption of the feed flow, the interruption of the fuel flow, coking of the individual coils, and the failure of the instrument air supply. Interruption of feed flow can result in a dangerous situation if firing is maintained at the normal rate, because the tubes are not designed for the excessive temperatures that result if charge is stopped or drastically reduced, and the danger of tube rupture is, therefore, pronounced.

The low-flow alarm (FAL-4, Figure 8.27y) is provided in the feed stream to warn of impending danger. Once it is verified that the danger is real, the vent solenoid valve (HY-1) can be tripped by the operation of the pushbutton (HS-1) that vents the diaphragm of the emergency valve (HV-1), shutting off the fuel gas. This shut-off can be automated, so that on a drop in the feed flow, a low-flow switch (FSL-4) automatically actuates the trip solenoid (HY-1), which cuts off the fuel flow.

Interruption of fuel flow will cause burner flameout, and resumption of fuel flow may result in a dangerous (fuel-rich) fuel/air mixture. To protect against such resumption of fuel flow, the low-pressure switch (PSL-1) is used in Figure 8.27y, which trips the emergency shut-off valve (HV-1) via the solenoid valve (HY-1). The solenoid valve is the manual reset type; therefore, once tripped it will reopen only if manually reset.

Excessive coking of an individual furnace coil can occur as a result of the restriction of the flow through it. This is a self-worsening effect and tends to cause dangerous overheating. The prevention of this situation is of prime concern. Should such a condition occur, it will be detected by the high-temperature alarms (TAH-1 through TAH-3 in Figure 8.27x) to warn the operator.

In response, the operator can increase the set point of the appropriate feed flow controller, thus forcing more fluid through the hot tube, thereby hoping to bring the temperature down. If this does not alleviate the condition, the operator has no alternative but to shut down the firing.

Failure of instrument air will result in failing all the control valves open, as in Figure 8.27x. The feed valves and the steam valves open on air failure to continue the flow through the furnace coils and, thereby, to prevent overheating and possible rupture of the coils.

ADVANCED CONTROLS

Advanced furnace controls include feedforward control based on feed rate, cross-limiting firing (Figure 8.6rr in Section 8.6), coil balancing (Figure 8.21r in Section 8.21), and optimization. Implementation of these control strategies is normally aided by expert system software, distributed control systems, single or dual loop programmable controllers, or dedicated microprocessor-based devices can also perform this type of control to a limited extent. An example of each type of control is presented in the next paragraphs. Control systems can include any combination of these advanced techniques.

Feedforward Control

In feedforward control, a simplified model of the process is used to predict the effect of disturbances before they reach and can upset the controlled variable. In contrast, feedback control must first detect an error in the controlled process variable before it can initiate a corrective action.

A simple example of feedforward control is given in Figure 8.27z. Here the feedback loop consists of the temper-

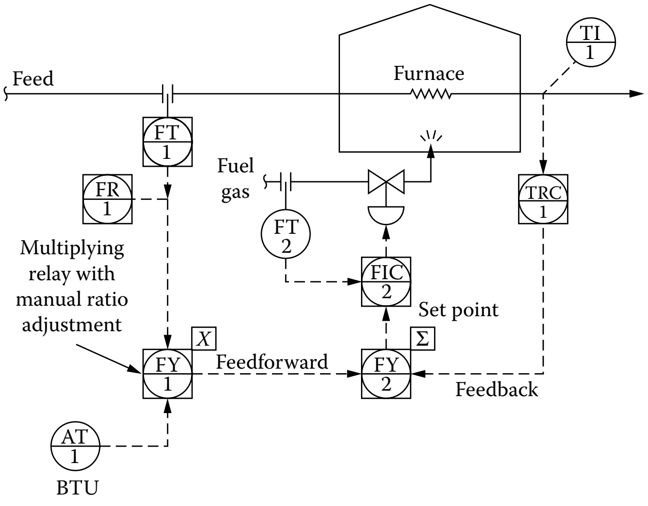


FIG. 8.27z
Feedforward adjustment of furnace firing rate can minimize upsets caused by feed flow changes.

ature controller (TIC-1) providing the set point for the fuel flow controller (FIC-1). When the feed flow to the furnace changes (this flow is an independent variable in this system), there is a need to change the rate of fuel firing.

The feedforward loop includes the feed flow transmitter (FT-1), which detects the change in feed flow rate, and this rate is multiplied by a constant in FY-1. This constant relates to the heating value of the fuel gas, and if a BTU analyzer (AT-1) is used, the correction can be automatically obtained. This multiplying block sets the relationship between a change in feed flow rate and the required corresponding change in fuel header pressure. This is an empirically determined value and can be field-adjusted.

Dynamics may be included (dead time, lead/lags) if warranted (this topic is discussed in full detail in Section 2.9 in Chapter 2). The response of the furnace to an increase in firing rate is often too slow, and therefore such controls are useful only when responding to step changes in feed flow rate.

The modified signal from FY-1 is then summed (FY-2) with the signal from the temperature controller output, thereby setting a new fuel gas flow rate via FIC-2. In effect, advance information is fed forward through FY-1 to the firing controller (FIC-2), indicating that a change in process load will require a change in firing rate shortly and that the firing rate, therefore, should begin to change.

Without the feedforward leg of the loop (FY-1 and FY-2), the required change in firing rate would take place much later, after the temperature controller (TRC-1) detects an error in the controlled variable (the effluent temperature of the furnace). If a constant ratio existed between feed flow and fuel gas flow, the temperature controller would not even be necessary, but with changing ambient and process conditions, this ratio changes with time. Thus, TRC-1 acts in the feedback path as a slow trim-controller, keeping the controlled variable at a desired value.

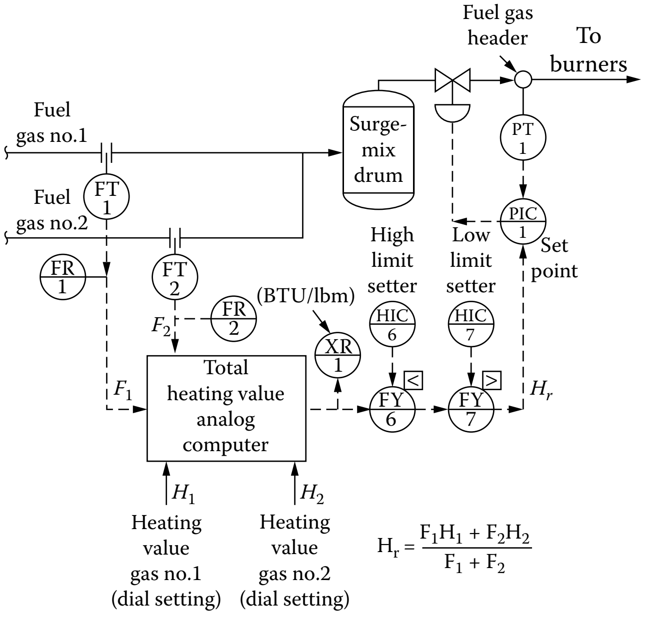


FIG. 8.27aa
Header pressure set by heating value of fuel mixture.

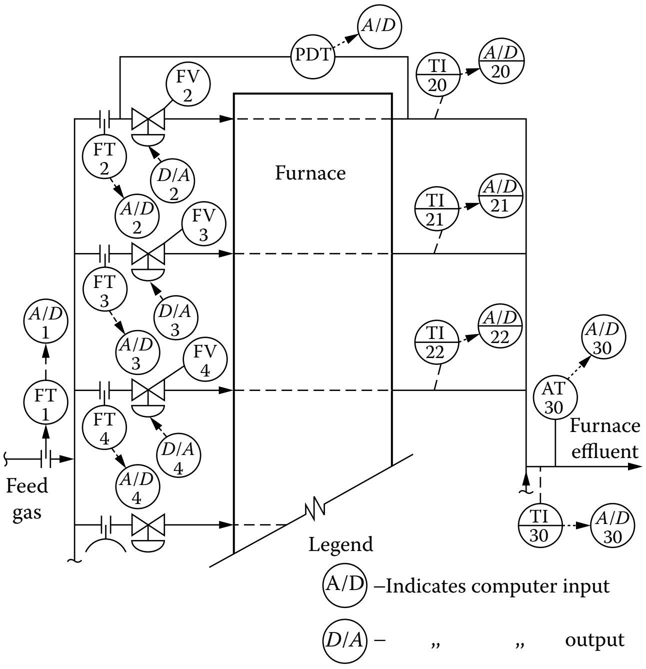


FIG. 8.27bb
The I/O requirements of a digitally controlled pyrolysis furnace.

Coil Balancing Control

One of the main considerations in a multipass furnace control is the proper splitting of feed among the parallel passes through the furnace. In a large ethylene plant, six to eight identical cracking furnaces may be present, each with from 10 to 20 parallel coils. Maintaining proper flow splits in these coils is a formidable task if done manually.

The objective is to maximize furnace energy utilization per unit of feed flow. In general, heat transfer efficiency is highest when the percentage of vaporization measurements in all the coils are equal. However, without a measurement of percentage of vaporization, the best that can be done is equalization of all the individual coil outlet temperatures.

Digital control of the total furnace circulation is implemented by feedback manipulation of all furnace coil flow rates. Figure 8.27bb shows an older computer control system. Here, the total feed flow (A/D-1) and the individual coil flows (A/D-2, -3, -4, etc.) are monitored by the DCS system, which calculates the proper valve settings based on these flows. The proper valve openings are sent by the DCS system (through the digital-to-analog converters (D/A-2, -3, -4, etc.), and the valves divide the total feed flow properly.

The total feed flow rate sets the total charge to the furnace. Each coil flow rate set point is calculated as the total furnace flow (minus the flow of all nonautomated coils), divided by the number of coils that are automatically controlled. This allows for coil balancing of at least a portion of the furnace, if some of the coils cannot be placed under automatic control due to defective instrumentation.

After start-up, the total flow will be divided equally among the passes, but as time progresses, coke builds up at different rates in the individual coils, causing outlet temperatures to vary from coil to coil. This control system keeps the coil effluent temperatures the same (within some tolerance) by adjusting the flows of the feeds through each coil. The DCS system does this by taking the effluent coil temperature information (A/D-20, -21, -22, etc.), comparing them with the desired temperatures, and modifying the "feed-splitting" computation to correct for deviations. The results of the computation are sent out to reposition the feed valves (FV-2, FV-3, FV-4, etc.).

Changes to the furnace charge rate are normally ramped, and their speed of change is limited to avoid rapid upsets in the furnace. Limit constraints, including a differential flow limit between each coil and the average coil flow, are normally included in the overall control algorithm to maintain safe flow rates in all coils, in order to prevent excessive coking or development of hot spots in the furnace.

Many additional optimization functions can be performed using much of the same input data. Some of these are "off-normal" alarms on feed flow, "off-normal" alarms on effluent temperatures, alarms to signal excessive pressure drops across coils, high coil-metal temperatures, and other scanning functions.

Cross-Limiting Firing

A cross-limiting firing control technique ensures that air is always in excess of the amount required to fully combust the fuel. This avoids hazardous (fuel-rich) combustible mixtures in the firebox. When an increase in firing rate is needed, cross-limiting firing controls first increase the airflow, and the fuel flow only after that. When a decrease in firing rate is desired, the strategy reduces fuel flow before reducing the airflow. This is performed through a combination of high- and low-select modules and dynamic exponential lag modules. A control block diagram of the cross-limiting firing circuit is shown in Figure 8.27cc.

The success of the cross-limiting firing is predicated on being able to measure or infer the air and fuel flow rates (see Table 8.6d in Section 8.6 for sensors). If the airflow is controlled only by damper position, this flow rate must be interpolated from the damper position. Also, fuel is often not metered, but only controlled by a pressure controller, so the

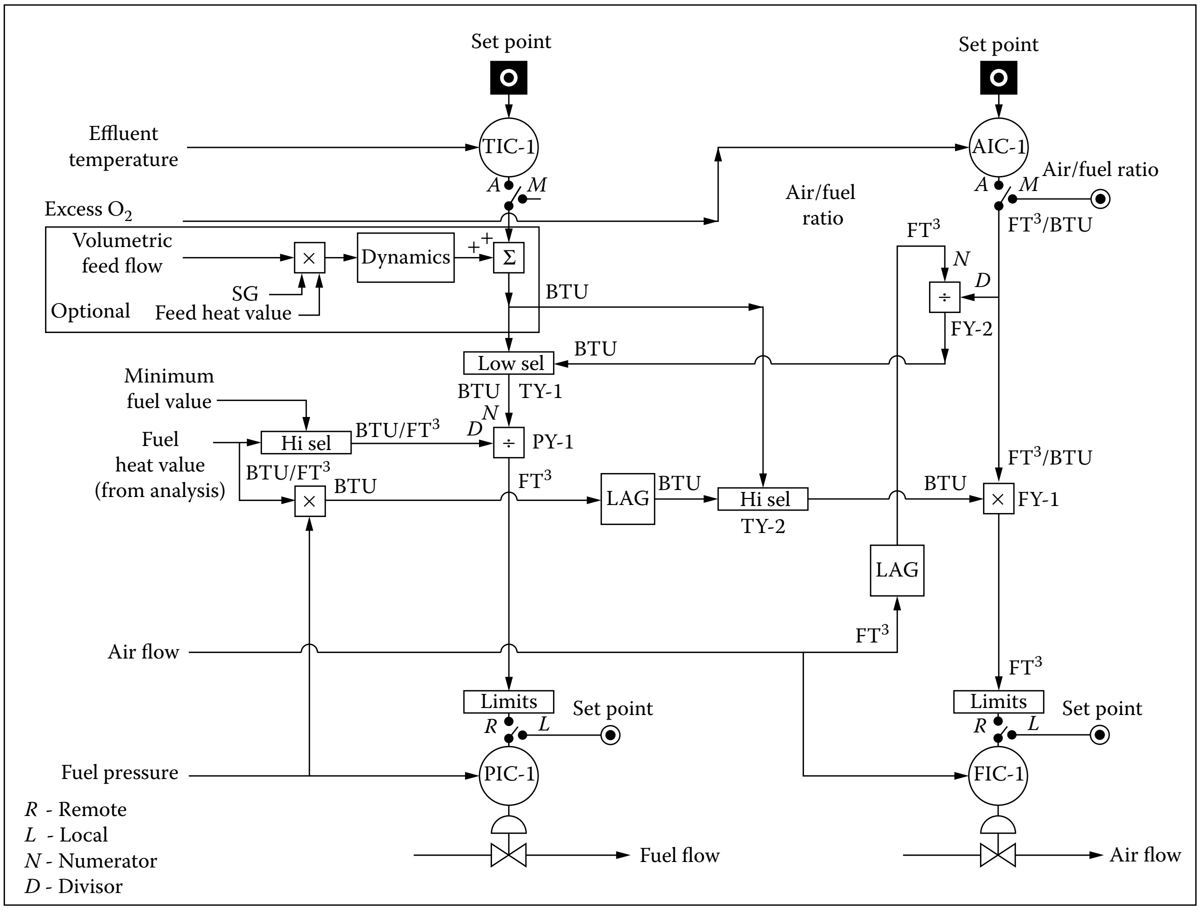


FIG. 8.27cc
Cross-limiting firing control block diagram.

header pressure must be converted into an approximation of flow rate. Such indirect methods of flow approximations are highly inaccurate and should no longer be used.

The cross-limiting firing system described in Figure 8.27cc works in the following manner: When the effluent temperature controller (TIC-1) calls for additional heat to be supplied by the furnace, the output of the controller increases. After feedforward adjustment, this signal goes to both a high- and a low-signal selector (TY-1 and TY-2).

At the high-signal selector, this rising signal will be greater than the signal representing current fuel flow (pressure). The high select (TY-2) will thus call for increased heat by increasing its output, and this increased signal will be multiplied by the current air/fuel ratio in FY-1. The output from this multiplier (FY-1) goes to the set point of the airflow controller (FIC-1). Therefore, the result of an increase in the output of TIC-1 is an immediate increase in the airflow controller set point.

At the low select (TY-1), the increasing TIC-1 output signal will not be chosen because it is greater than the signal supplied by the air/fuel ratio FY-2. The airflow measurement signal is sent through a lag block and is divided by the air/fuel ratio set point in FY-2, which produces the second input to the low selector. The immediate increase in air, which followed the rise in TIC-1 output, is thus lagged with a first-order lag block prior to the low select (TY-1), resulting in a gradual increase in fuel. Therefore, an increase in the firing rate (TIC-1 output) results in a fast increase in air and a slow increase in fuel flow.

When the firing demand drops due to a drop in the feed flow or in the temperature controller output, the reduced signal is sent to both the high and low selector modules (TY-1 and TY-2). The dropping signal is immediately passed through by the low selector (TY-1), resulting in an immediate decrease in the fuel flow set point.

However, at the high selector (TY-2), the decreased signal is blocked, because the lag element keeps the apparent flow of fuel higher. The result is a slow (first-order lagged) reduction in the flow of air. A change in the output of the excess air (or O_2) controller (AIC-1) goes through the low selector on the fuel flow (TY-1) in a similar manner. The output of the controller (AIC-1) is the air/fuel ratio setting for the furnace, which is multiplied by the signal from the high selector in FY-1. Therefore, the air/fuel ratio directly affects the airflow set point. This same air/fuel ratio is the divisor at the airflow in FY-2. The output of FY-2 is the second input to the low selector TY-1, which selects the fuel oil flow set point.

Severity Control

Using the ethylene cracking furnace in Figure 8.27bb as an example, an analyzer measuring the effluent stream composition (AT-30) also sends its measurement signal to the computer, where it is compared with the desired composition. Based on the difference between the two, the desired degree of fuel firing in various furnace zones can be computed.

These firing rates may be printed out as instructions to the operator or used to automatically manipulate the fuel valves, or fuel flow controller set points, thereby changing the heat flow pattern throughout the furnace and thus controlling the cracking operation.

Model-Based Control

Process models can relate the control variables of the radiant coil to operational parameters. Information on furnace operation includes hydrocarbon flow rate, steam-to-hydrocarbon ratio, coil outlet temperatures and pressure, severity of cracking, and fouling parameters. These parameters are used to develop interrelationships between independent and dependent variables.

Matrix calculations can be performed to change the manipulated variables so that the controlled variables are held at their set points. General-purpose computing modules are used to run these models and optimizers. For in-depth coverage of a variety of model-based control approaches, refer to Sections 2.13 to 2.18 in Chapter 2.

Two common applications of model-based control will be briefly discussed below.

Coil Outlet Temperature (COT) Matrix The individual COTs can be simultaneously controlled by solving an interaction matrix that relates the change in each coil temperature to the change in valve openings of the wall burners. The matrix coefficients determine the gains associated with each COT-firing zone pair.

Knowledge of furnace geometry and experimental data is employed to define the matrix. If the matrix accounts for only the steady-state character of the process, the dynamic nature must be handled by feedback trimming of the COT controllers.

It may be necessary to remove the excess degrees of freedom that result from having more wall burners than COT measurements. This can often be accomplished by interpolating the desired wall burner valve positions on the basis of the outputs of adjacent COT controllers.

Provisions should be made to disable controllers whose thermocouples have failed (burned out). The proper installation of tube surface-detecting thermocouples is shown in Figure 8.27dd.

Because of the harsh service, COT thermocouple (T/C) failure is a routine incident that the control strategy must handle. To detect the thermocouple burnout, it is necessary to detect the loop resistance of the installed thermocouple (Figure 8.27ee).

Dual or companion thermocouples can be connected to multiplexers and can be used to validate the COT measurements. When a COT controller is placed in manual, its output is forced to be equal to the average output of the COT controllers that are still controlling the process (the average of those controllers that are not in manual).

In this way, temperature control for the failed coil is still maintained to some extent, and responses are made to

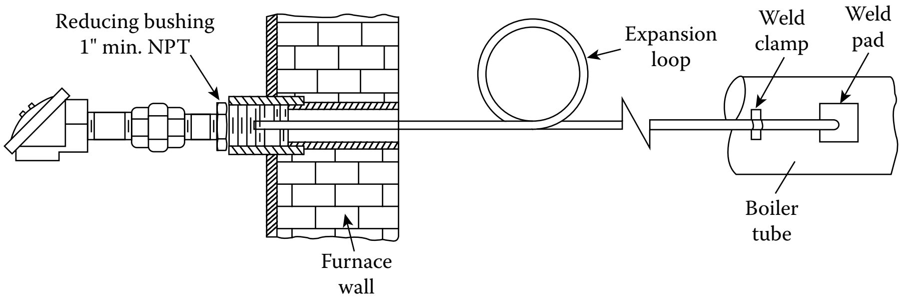


FIG. 8.27dd
The expansion loop allows for thermal expansion in furnace applications.

disturbances that are affecting all the coils. This approach has proven to be an effective solution whenever thermocouples are to be replaced on-line.

The use of a matrix algorithm alone does not necessarily guarantee an even firing pattern, because there is no feedback component in this strategy to correct for errors in the matrix model. Over time, as the effects of coking, air register adjustments, and burner fouling are better understood in a particular installation and when they are better quantified, solutions can be implemented by shifting the duty requirements between the firing zones in such a way as to equalize the bias station positions.

Optimizers Optimizers are used to calculate the required charge rates and the severity of cracking of ethylene furnaces, based upon financial or contractual criteria. The model-based program then simultaneously manipulates several independent variables to achieve these dependent control points.

The model solves simultaneous polynomial equations relating severity to charge rate, steam-to-hydrocarbon ratio, furnace effluent temperature, excess oxygen, and coil fouling rate in a matrix format. Outputs to local controllers or valves are then adjusted all at once.

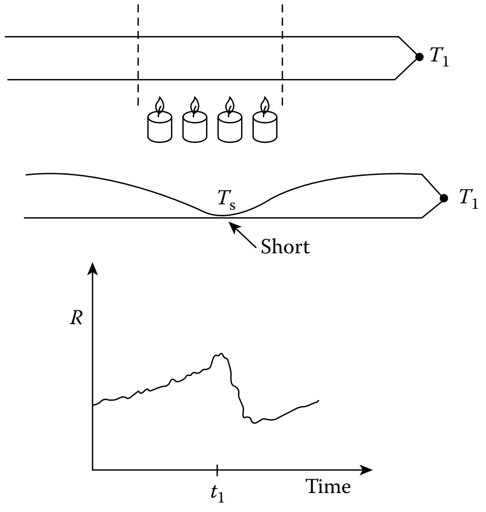


FIG. 8.27ee
Thermocouple burnout can be detected by measuring the resulting drop in thermocouple loop resistance.

CONCLUSIONS

The furnaces and control systems discussed in this section are representative of a wide spectrum of furnaces throughout the process industry. The strategies are valid whether implemented by older analog instrumentation or by more sophisticated DCS or PLC systems.

The control systems that were described in this section did not cover the specifics of advanced process control. For the subjects of optimization and modeling, the reader is referred to Sections 2.7 to 2.20, and for relative gain calculations or for decoupling of interactions to Sections 2.25 and 2.37 in Chapter 2. It should also be noted that the sections describing boiler controls (Section 8.6) and distillation controls (Sections 8.19 to 8.21) should also be studied, because there is a fair amount of overlap between these unit operations and furnace controls.

With the expansion of industrial facilities and with the use of larger furnace units on the one hand and with the increased availability of advanced process control (APC) on the other hand, the use of more advanced and model-based controls should be considered, while taking full advantage of the basic controls described in this section, to obtain more efficient and more profitable furnace operation.

STANDARDS AND GUIDELINES

ANSI/API 536, “NO_x Control for Fired Equipment in Refineries”
ANSI/ISA -84.01–1996, “Application of Safety Instrumented Systems (SIS) for the Process Industries”
API RP 551/1993, “Process Measurement Instrumentation”
API 554/1993, “Process Instrumentation and Control”
API 553/1998, “Control Valves”
API 556/1997, “Instrumentation and Control Systems for Fired Heaters and Steam Generators”
API 557.01/2000, “Guide to Advanced Control Systems”

40CFR Part 60, "New Source Performance Standard" (section for oil refineries)

40CFR Part 61, "National Emissions Standards for Hazardous Pollutants"

IEC 61508, "Functional Safety – Safety Related Systems"

ISA TR84.02-2002, "Fault Tree Analysis"

NFPA 85-2004, "Boiler and Combustion Systems Hazard Code"

OSHA 40CFR1910.119, "Process Safety Management"

Bibliography

AIChE, "Guidelines for Safe Automation of Chemical Processes," 1993.

Batur, C., Sharpless, R. B., Duval, W. M. B., Rosenthal, B. N., and Singh, N. B., "Adaptive Temperature Profile Control of a Multizone Crystal Growth Furnace," *Proceedings of the American Society of Mechanical Engineers,* Dallas, TX, November 1990.

Baxley, R. A., "Computer Control of Severity on Ethylene Cracking Furnaces," *Instrumentation Technology,* November 1971.

Blevins, T. L., McMillan, G. K., Wojsznis, W. K., and Brown, M. W., *Advanced Control Unleashed,* Research Triangle Park, NC: ISA, 2003.

Boyle, W. E. and Hoyt, R. R., "Experimental Application of Combustion Controls to a Process Heater," Annual ISA Conference, September 13–17, 1948.

Calabrese, A. M. and Kreji, L. D., "Safety Instrumentation for Ammonia Plants," *Chemical Engineering Progress,* February 1974.

Clelland, P. J., "Furnace Implosions: Causes and Cures," *Instrumentation Technology,* December 1977.

Dukelow, S. G., "Energy Conservation through Instrumentation," University of Houston Energy Conservation Seminar, February 7, 1993.

Dukelow, S. G., "Fired Process Heaters," presented at ISA/CHEMPID, April 1980.

Dukelow, S. G., "How Control Improvements Save Process Heater Fuel," 1979 Conference on Industrial Energy Conservation Technology.

Edwards, D., "From LR to PC: Recent Developments in Furnace Control Systems," *Metals and Materials,* October 1990.

Garcia-Borras, T., *Manual for Improving Boiler and Furnace Performance,* Houston, TX: Gulf Publishing Co., 1983.

Gibbons, C. and Kania, C., "On-Line Optimization for an Ethylene Plant," *Proceedings of the ISA/90 International Conference & Exhibition,* Vol. 45, Part 2, New Orleans, LA, October 1990.

Gilchrist, J. D., *Fuels, Furnaces, and Refractories,* New York: Pergamon Press, 1977.

Gruhn, P. and Cheddie, H., "Safety Shutdown Systems–Design, Analysis, and Justification," 2001.

Harrison, T. J., "Microcomputer in Industrial Control: An Introduction," Research Triangle Park, NC: Instrument Society of America, 1978.

Hugo, A., "Limitations of Model Predictive Controllers," *Hydrocarbon Proc.,* January 2000, p. 86.

Hurowitz, S., Anderson, J., Duvall, M., and Riggs, J. B., "Distillation Control Configuration Selection," submitted to *J. Process Control,* March 2004.

Johnson, R. K., "Modern Approach to Process Heater Control Systems," Parkersburg-Marietta ISA Section, Parkersburg, WV, February 8, 1977.

Khalil, E. E., *Modelling of Furnaces and Combustors,* Tunbridge Wells, UK: Abacus Press, 1982.

Kostiw, W. B., "How to Instrument Fuel Oil Systems," *Instrumentation Technology,* October 1974.

Kralj, F., "Application of the S88 Model in the Control of Continuous Distillation Facilities," *Proceedings of the World Batch Forum,* Woodbridge Lake, NJ, April 2003.

Lipták, B. G., "Process Control Trends," *Control,* January, March, and May, 2004.

Maeda, M. and Nagao, S., "A New Combustion Sensor for Exhaust Gases for Boilers, Furnaces, and Automobile Engines," *Proceedings of the ISA/90 International Conference & Exhibition,* Vol. 45, Part 1, New Orleans, LA, October 1990.

Marchant, G. R., "Digital Controls for Continuous Copper Smelting," *Instrument Technology,* June 1978.

Normandeau, L. and Bambeck, R. J., "Combustion Control using CO Trim for Coal- and Gas-Fired Boilers," *Proceedings of the ISA/89 International Conference & Exhibition,* Vol. 44, Part 4, Philadelphia, PA, October 1989.

Omata, S., Yamamoto, T., Maruyama, H., and Hotta, M., "Adaptive Control of a Furnace Temperature by using a Microcomputer," Proceedings of the 28th Society of Instrument and Control Engineers, Tokyo, July 1989.

Reed, R. J., *Combustion Handbook,* 3rd edition, Cleveland, OH: North American Mfg. Co., 1986.

Roffel, B., "Distillation: Instrumentation and Control Systems," *Encyclopedia of Separation Science*, London: Academic Press, 2000.

Shinskey, F. G., "Controlling Multivariable Processes," Research Triangle Park, NC: Instrument Society of America, 1981.

U.K. Health and Safety Executive, "Out of Control: Why Control Systems Go Wrong and How to Prevent Failure," 1995.

Willis, M. J., "Selecting a Distillation Column Control Strategy (A Basic Guide)," 2000.

8.28 Header-Supplied Distribution Control

S. BAIN AND B. G. LIPTÁK (2005)

INTRODUCTION

Distribution controls for gases and liquids have already been discussed in this volume in Section 2.23 (Ratio Control), Section 8.1 (Aeration Controls), Section 8.2 (HVAC Controls), Section 8.15 (Compressor Controls), and Section 8.17 (Cooling Tower Optimization). Similarly, the topics of gain compensation and control valve performance have been covered in detail in Chapters 2 and 6.

Yet, it is believed that it is worthwhile to combine in a single section the various control options and considerations that should be taken into account when designing and operating a distribution control system. This is the purpose of this section, which concentrates on controlling header-supplied distribution systems.

Terminology and Coverage

A *header* or a manifold is the pipe that interconnects the supply pumps or blowers with the user valves or other flow distribution elements of the process. A header is typically closed, but the control strategies outlined in this section also apply to open-conduit-type distribution applications. Control strategies will be described to coordinate both the supplying of gases and liquids from blowers and pumps into the distribution header, and the distribution of these materials among their users in the plant.

The main topics of this section will include the most-open valve (MOV)-based valve position control options and their improvements through feedforward and similar control strategies. The section will also discuss gain and hysteresis compensation, methods of failure and fault detection, energy conservation, and some specific applications.

VALVE POSITION CONTROL

Many of the control strategies described in this section are based on the MOV-based valve based valve position control (VPC) systems. The basic motivation for MOV control is to meet the requirements for gas or liquid supply to the process at a minimum of compressor or pump energy. By opening the user valves, the pressure drops and, therefore, the transportation energy requirement is reduced. Therefore, an MOV control system will open all the user valves until one of them approaches full opening. This way, all loads can be satisfied at minimum investment of transport energy.

The advantages of MOV control are not limited to energy savings; they also include the potential reduction for the need for valve maintenance caused by high pressure drop-based abrasion and cavitation. Operating safety can also be improved, because it is safer to work on and operate equipment at lower pressures. If the distributed process liquid is supplied by gravity from an elevated tank, the available maximum pressure drop can limit the maximum load to the most demanding user, and therefore by lowering the required pressure drop, this limit can be raised.

Another advantage of MOV control has to do with the positioning accuracy of control valves, which is about ± 1% of the valve stroke. This positioning error is a fixed quantity, which becomes a larger percentage of the operating stroke (and, therefore, of the flow through the valve) as the valve is throttled down. Therefore, if all valves are opened up, this positioning error will drop, because this error when the valve is 90% open is 1/90 = 1.1%, while with a 10% open valve it is 1/10 = 10%.

In this section the types of valve position control systems will be distinguished by numbers such as MOV1 and MOV2.

When the user valves are throttled on the basis of the flows they deliver, this configuration will be referred to as MOV1. In this case, the MOV1 controller not only controls the individual user flows, but as a consequence, it also controls the flow distribution among users and the total flow supplied to all users. This control strategy is illustrated in Figure 8.28a, where the individual user flows and, therefore, their share of the total flow are both controlled by the needs of the process. In this control configuration, the valve position controller (VPC) measures the total flow required and adjusts the compressor speed to match the demand with the supply, while keeping the discharge pressure at an optimum (minimum) value.

When only the flow distribution among the users is controlled, that configuration will be referred to as MOV2. As can be seen in Figure 8.28b, the flow control loops at the users control only the flow distribution, while the total amount of the flow is determined by pumping capacity. Therefore, in this configuration, the MOV2 controller has to allow the passage of all the flow the distribution header receives, and it only controls its distribution.

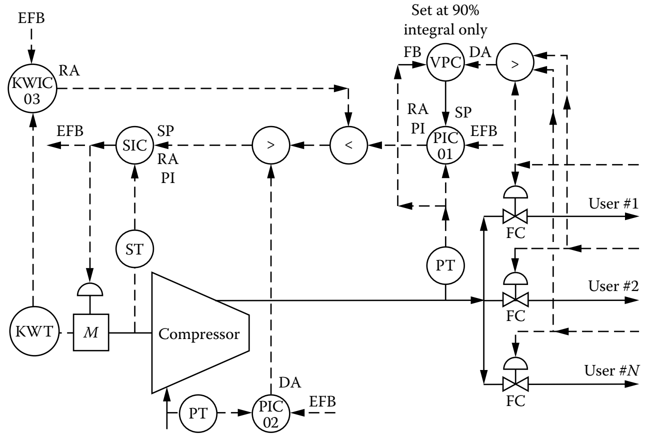


FIG. 8.28a

Illustration of the MOV1 control concept, where both the flows to the individual users and the total flow supplied to the header are controlled by the process demand.

The MOV1 and MOV2 control configurations are fundamentally different and are not interchangeable. They will be discussed in more detail below. Variations of the MOV controllers are many and are continuously evolving, because it makes good sense to automatically switch from controlling the system based on the opening of the MOV to temporarily controlling it by some other variable (pressure, level, or process analyzers), when that is desirable.

The Most-Open Valve

The extent to which the most-open valve should be allowed to open depends on the nature of the process, because the higher that percentage, the less safety margin is available to overcome a process upset. The set point of most valve position controllers is adjustable and is selected to be between 65 and 85%. The higher this set point, the higher the energy efficiency, but the lower the safety margin of operation.

One way to determine the desirable set point (MOVsp) is to make it flow-dependent. In other words, adjust the set point that is desired (say 75%), when the system flow is at the maximum value for which the process has been designed for (Q_d) as a function of the actual flow that is occurring (Q_a) at the time. A suitable relationship between flow and set point can be

$$\text{MOVsp} = 75\% \; (Q_a/Q_d)^k \qquad \textbf{8.28(1)}$$

where k is adjustable from 0 to 1.

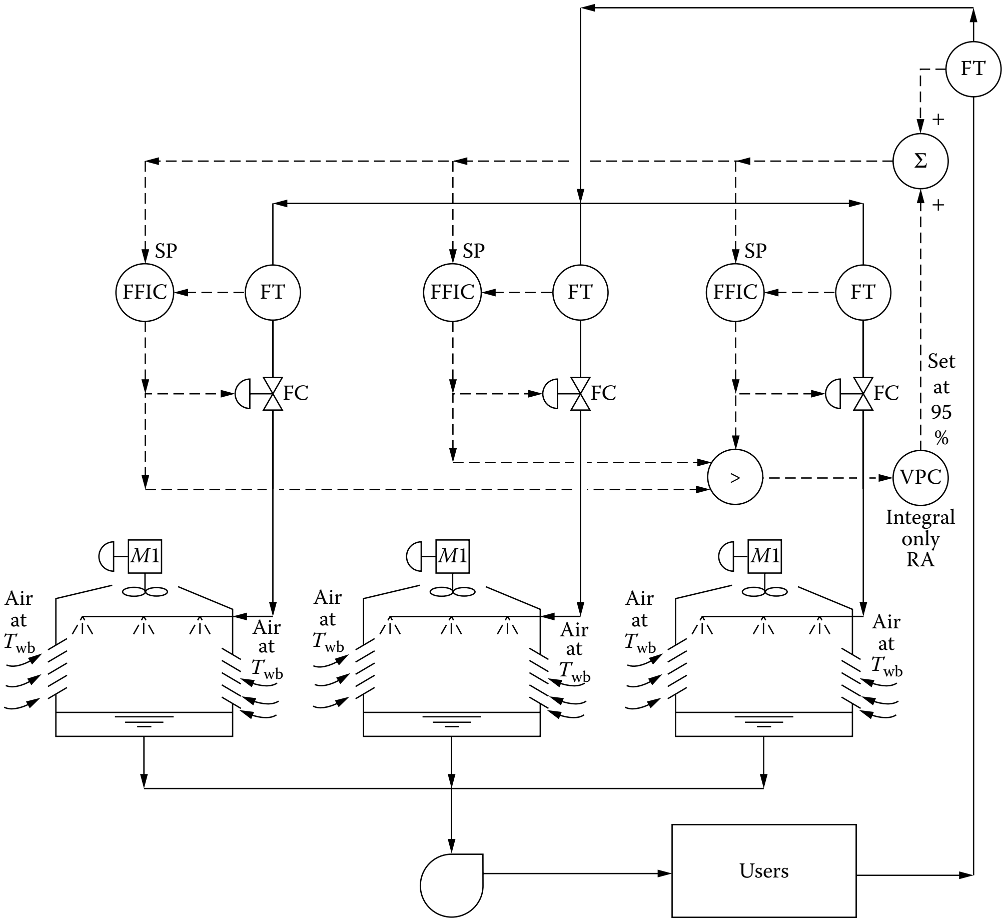


FIG. 8.28b

Water distribution can be controlled automatically while minimizing pumping costs.

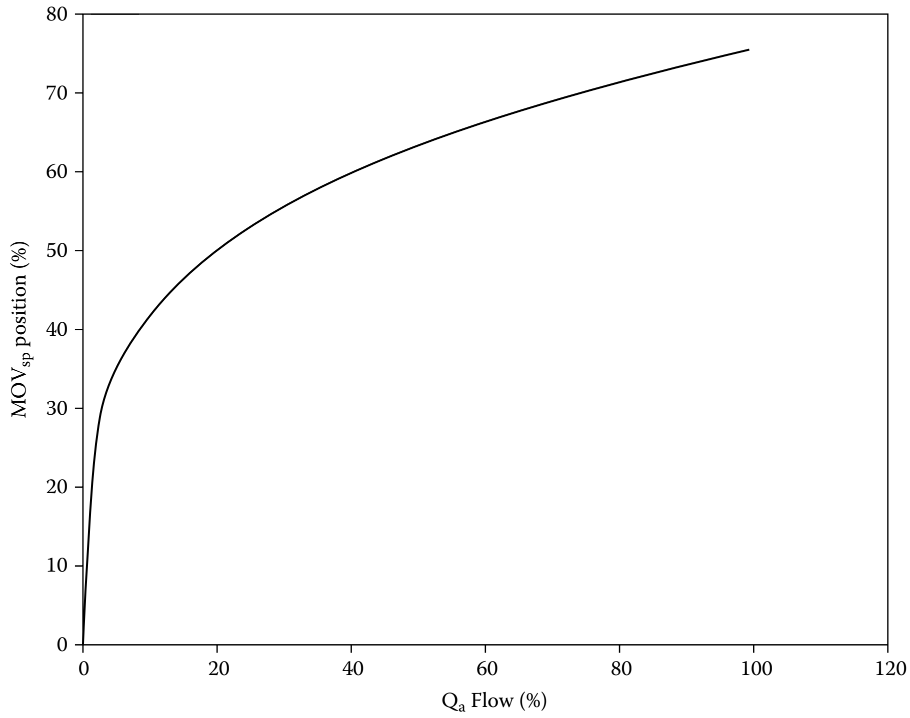


FIG. 8.28c

The relationship between MOVsp and actual system flow, if the value of k in Equation 8.28(1) is set at 0.25.

Figure 8.28c illustrates this relationship when k is selected to be 0.25. If Q_a is greater than Q_d the process is operating beyond its design maximum. The values of Q_a and Q_d can be based on flow through all the user valves or can be based on the flow through only the most-open valve (MOV). It is usually better to use the total system flow, because in that case Q_d is a constant, while if the flow through the selected MOV is used, Q_d has to be changed every time a new valve is selected as the controlling MOV.

In well-maintained systems that are operating with only a few valves, the MOV usually means the selection of a single valve, the one that is the most open. In processes that are not so well maintained (in HVAC or water treatment, where valves can be stuck in full open positions) or in processes involving the selection from among several hundred valves (fuel cells, aeration), selecting a single valve might not be the best. In these processes one might select the three, five, or even ten of the most-open valves and average their openings to arrive at the system MOV.

Energy Savings

A fluid distribution system that is controlled to maintain a traditional constant header pressure is likely to use considerably more energy than is necessary. One way to reduce the energy cost of fluid transportation is to minimize the pressure drops introduced by throttled control valves. This goal is served by valve position control that keeps the most-open valve nearly fully open.

The energy savings associated with using variable-speed or multistage blowers or pumps have been discussed in the sections dealing with compressor, fan, and pump optimization in this chapter. Figure 8.28d illustrates the energy savings potential of controlling the flow to a single user by variable-speed pumping instead of by throttling a control valve on the discharge of a constant-speed pump.

In most fluid distribution applications, the combination of MOV controllers throttling variable-speed supply devices can be expected to bring an energy savings of about 25% or more. Improvements, such as power factor correction or the use of high-efficiency motors, can also contribute significantly to energy reduction.

Similarly to pumping stations in liquid distribution systems, the operating energy costs of gas distribution systems can also be reduced by using variable-speed blower controls. Figure 8.28e compares the power consumption of constant-speed blowers that are throttled by suction or discharge dampers with the power consumption of variable-speed blowers as the load is reduced.

Companies that specialize in energy-related activities include Energy Strategies Corporation (ESCOR), which specializes in blower capacity control, such as adjustable-speed, inlet guide vane throttling and the control of discharge diffusers. H.V.-Turbo A/S (also known as Turblex), Helsingor,

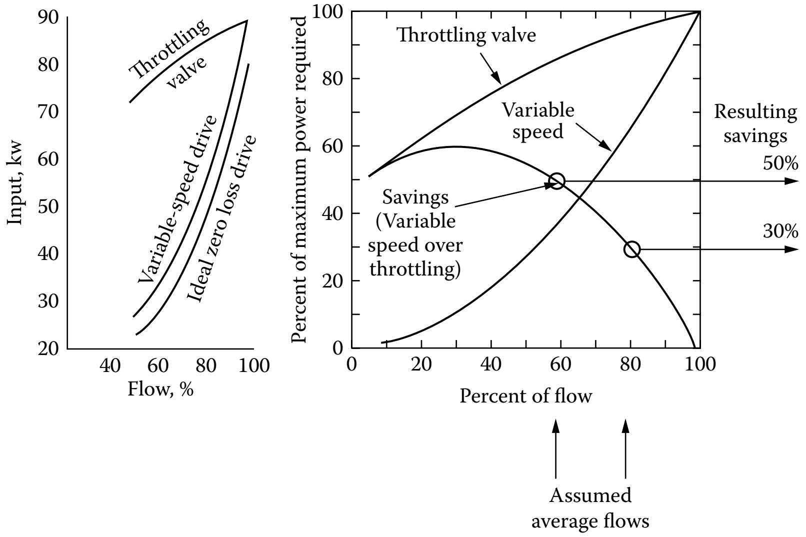


FIG. 8.28d
Energy saving as a function of operating flow when that flow is controlled by modulating the pump speed instead of opening a control valve.

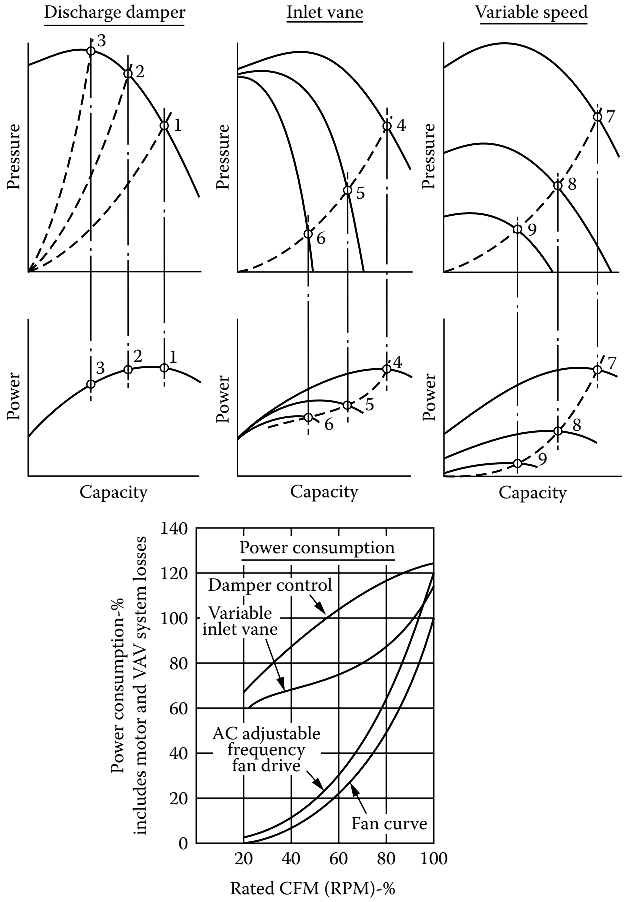


FIG. 8.28e
Operating power requirements of variable-speed blowers and compressors are less than those of constant-speed ones with suction or discharge dampers.

Denmark, is a supplier of high-performance blowers and supplies integrated control packages that can interface with other manufacturers' blowers as well.

Valve and Damper Gain

The gain of any device is its output divided by its input. For a linear (constant gain) control valve, the valve gain (G_v) is the maximum flow divided by the valve stroke in percentage (Fmax/100%).

When a control loop is tuned to provide quarter-amplitude damping, which is the typical goal (see Section 6.7), the controller gain (Gc = 100/% PB) is adjusted until the overall loop gain (the product of the gains of all the loop components) reaches 0.5. If a linear controller and a linear transmitter are used, their gains are constant. Therefore, if the process gain (G_p) is also constant, a linear (G_v = constant) valve is needed to keep the loop gain product constant at 0.5.

Valve or Damper Characteristics The inherent characteristics of a control valve or damper describes the relationship between the controller output signal and the flow through that valve. Some of the widely used inherent lift to flow rate relationships are the linear, equal-percentage, and quick-opening characteristics.

In equal-percentage valves, a unit change in lift will result in a change in flow rate, which is a fixed percentage of the flow rate at that lift. For example, each percentage increase in lift might increase the previously existing flow rate by 3%. Therefore, the theoretical gain of equal-percentage valves is directly proportional to the actual flow rate and increases as the flow rate increases. On a logarithmic chart

TABLE 8.28f
Valve Characteristics Selection Guide

Service	Valve ($\Delta p_{max}/\Delta p_{min}$) under 2:1	Valve ($\Delta p_{max}/\Delta p_{min}$) over 2:1but under 5:1
Orifice-type flow	Quick-opening	Linear
Flow	Linear	Equal %
Level	Linear	Equal %
Gas pressure	Linear	Equal %
Liquid pressure	Equal %	Equal %

the equal-percentage characteristic corresponds to a straight line having a slope that corresponds to a fixed percentage.

In quick-opening valves, the gain decreases with increasing flow rates. The inherent gain of linear valves is constant if the valve pressure drop is constant.

Table 8.28f lists the valve characteristics recommended for some of the most common control valve applications:

Gain Compensation The gain of any damper or valve can be measured at any opening by introducing a 1% step change in the valve stroke and measuring the resulting change of the flow through the valve [G_v = (flow change in units of % of Fmax)/(stroke change of 1% of full stroke)]. In computer-controlled systems, this gain characteristics determination can be performed automatically, and when such data are collected, compensation can be provided.

If it is desired that the apparent valve gain be constant (say $G_{va} = 1.0$), the compensator constant (C) is so selected that the product of the compensator constant and the inherent valve gain (G_{vi}) will be constant ($G_{va} = C \times G_{vi} = 1.0$). The results of a test performed on a 36-in. (900-mm) equal-percentage butterfly valve with a soft-start electric actuator, which was controlled over a digital network, are listed in Table 8.28g.

Modern control systems are provided with the automatic capability for loop component gain determination. Naturally, such tests should only be initiated after the valve stroke and flow sensors have been calibrated. The test to determine the actual, installed valve or damper characteristics should be performed during normal operation of the process, and therefore the data collection might require some time, such as running the program overnight.

It is also recommended that the characteristic curve be recorded when the device is opening and also when it is closing to check for hysteresis and repeat the test several times to establish its repeatability. Once the requirements for the compensator have been identified, one should fit a curve to the collected data. For example, the data in Table 8.28g corresponds to the curve shown in Figure 8.28h and is described by the equation below:

$$Q = -0.011x^2 + 2.036x + 3.094 \qquad \textbf{8.28(2)}$$

TABLE 8.28g
The Inherent Control Valve Characteristics Data Collected by Testing and Finding the Compensator Constant Values Required to Bring the Apparent Valve Gain to 1.0 throughout the Valve Stroke

Control Valve Stroke (x in percent)	% of Maximum Flow Measured (Q)	Inherent Control Valve Gain Measured (G_{vi})	Compensator Values at the Different Valve Openings
0	0		
5	7	1.93	0.519
10	25	1.82	0.551
15	35	1.71	0.586
20	44	1.60	0.627
30	57	1.38	0.727
40	68	1.16	0.865
50	76	0.936	1.068
60	83	0.716	1.397
70	88	0.496	2.016
80	93	0.276	3.623
90	97	0.056	17.86
100	100		

where:
Q is % of maximum flow
x is % of full stroke

To find the valve gain at any point on this curve, one needs to identify the slope of a tangent to the curve, which is the derivative of the valve characteristics. In other words, when using a second-order polynomial approximation for the valve characteristics, the gain equation will be first order, as expressed by the equation below.

$$G_{va} = -0.022x + 2.036 \qquad \textbf{8.28(3)}$$

In order to compensate for the inherent valve gain variation, at each valve opening, the inherent valve gain must be multiplied by the compensation constant C, in order to keep the apparent valve gain constant ($G_{va} = CG_{vi}$ = Constant). If the goal is to keep the apparent valve gain at 1.0, the equation for the compensator curve becomes

$$C = 1/(-0.022x + 2.036) \qquad \textbf{8.28(4)}$$

Hysteresis and Dead Band As was discussed in detail in Section 6.1, the response of the inner valve to a change in control signal is neither instantaneous nor direction-independent. When initiating a stem movement, it takes a finite amount of force to overcome the stem friction, and hysteresis represents the difference between the stem positions resulting from the same control signal when the valve is opening and when the valve is closing.

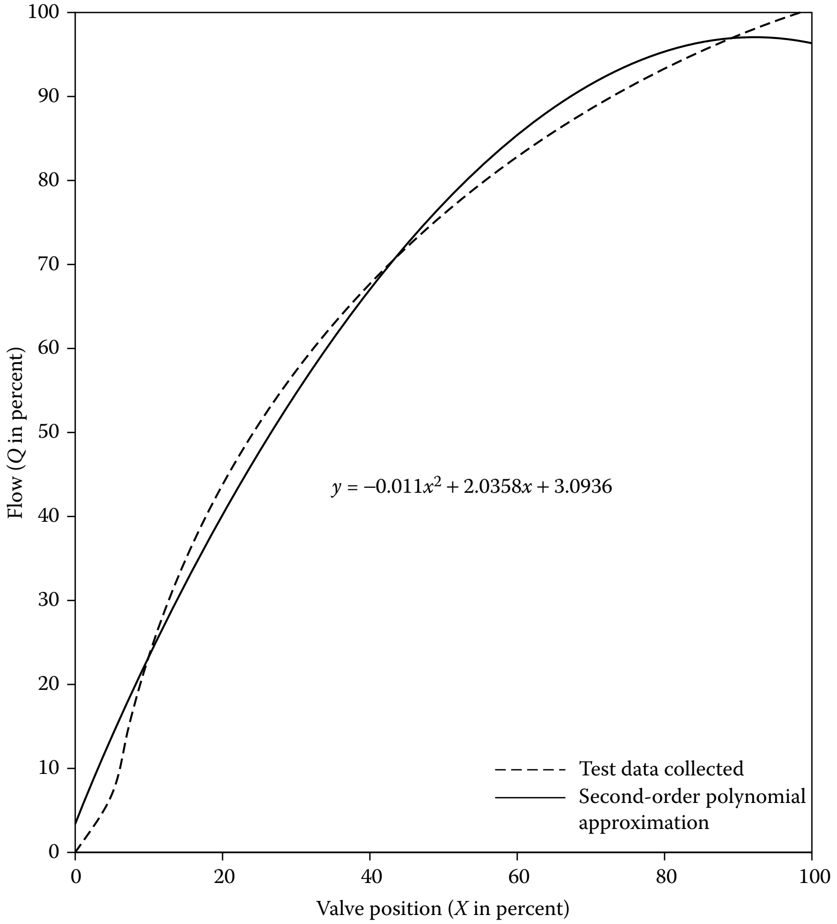


FIG. 8.28h
Fitting a polynomial approximation (———) to the test data collected (- - - - -).

Hysteresis also varies with use. In the case of new valves and dampers, it tends to be high and it tends to stabilize at some lower value after a fair amount of use, but later on, as the packing and other components wear out, it rises again. Hysteresis also varies with maintenance, such as lubrication.

For the above reasons, it is advisable that in critical applications the valve or damper hysteresis also be measured (usually by overnight or more extended automatic data collection). Once the hysteresis of the final control element is known, the compensation algorithm can be corrected by this expected value whenever the direction of stem movement reverses.

SUPPLY-DEMAND MATCHING VPC (MOV1)

As was illustrated in Figure 8.28a, the user valves serve to meet the needs of the process, while the MOV1-type valve position controller (VPC) modulates the total supply to the distribution header (pump or blower) to keep the most-open valve on set point. The set point of the controller sets the safety margin of the operation, because the closer it is to 100%, the less capability the most-open valve will have to respond to a sudden increase in process demand. Therefore, in critical processes, the set point can be as low as 65%, while in inconsequential processes, such as HVAC, it can be as high as 95%.

The header pressure is an indication of the supply-demand balance. If the demand is rising and the supply, therefore, is temporarily insufficient, the header pressure will drop until the supplying device (pump or blower) raises its capacity to match the demand. Inversely, if the demand drops, a temporary overpressure will develop in the distribution header. The time it takes for the system to respond to a change in demand is its speed of response, which can be improved by feedforward control (discussed later).

There are several analog and digital methods of implementing the control system described in Figure 8.28a. A digital and dedicated unit controller will be described below and will be referred to as a MOV1 controller.

MOV1 Controller Algorithm

The MOV1 controller described here is a digital unit controller that periodically identifies the most-open valve. The identity of this valve changes as the process loads vary. The measurement of the valve position controller is the opening of the most-open valve, while the set point of the MOV1 is adjustable. For illustrations of integrating MOV controllers into aeration process control systems, refer to Figures 8.1w and 8.1x in Section 8.1.

For example, if the set point is 80% and the identity of the MOV changes, such as when in the previous period *V*1 was the most-open valve at 80%, while in this period *V*2 is detected to be at 80.1% open, the MOV1 controller selects *V*2 for control. Because the valve opening of 80.1% is above its set point, it increases its capacity control output signal to the pump or blower that supplies the distribution header, which in turn increases the header pressure slightly to bring the new MOV back to set point.

Figure 8.28i shows the control schematic for a MOV1 controller that controls a gas distribution header (serving a number of process users) by throttling the suction control valves on "*n*" number of blowers. The user valves that the header serves are not shown, but their openings are listed on the top right of the figure.

This MOV1 controller algorithm includes a number of features: 1) It provides a feedforward approximation of the total demand of the users, by considering the openings of all user valves. 2) It determines the opening of the most-open valve (MOV) by measuring the average of the valve openings of the five most-open valves. This average then becomes the measurement of the valve position controller (VPC in Figure 8.28i). 3) The position controller VPC operates as the cascade master of the individual pressure controllers (PC), which all measure the header pressure, but control their own inlet control valves. 4) The current drawn by each blower is detected (IT) and controlled (IC) in such a manner so as to equalize them. This approximately equalizes the power consumption of the operating blowers.

Speed of Response and Failure Mode The speed at which the header pressure responds to an upset in the balance between supply and demand is a function of the process fluid and the volume of the header. When an incompressible fluid is being distributed, the response of the header pressure (PT) is fast, and therefore the amplitude and period of the pressure cycling can be limited.

In case of compressible fluids this is not the case, because an upset between supply and demand influences the header pressure (PT) very slowly. Therefore, if the demand changes, the supply does not only need to match the new demand, but also has to change the header pressure to a new value that is required to deliver the new demand. Because the pressure in the header has to rise (or fall) before flow through the control valves will increase (or decrease), a time delay is introduced. This is not the case with incompressible fluids, because if the header is full, whatever is pumped in at the supply end must instantaneously flow out at the delivery end, regardless of the header's volume.

One of the reasons for failure in a MOV1-type control system is the receiving of false user valve opening information. For example, if some failure causes a valve to fully open, it will become the controlling MOV and will cause the system to respond to a false measurement. An example of such a failure is when construction debris is lodged in a pipe bend or valve, restricting the flow.

When this occurs, the MOV1 algorithm will no longer operate at its optimum, but it will still continue to operate, because of two failure response features: One is the approximation of the sum of all valve openings and the averaging of the openings of the five most-open valves. The other is the limiting of the pressure controller (PC) set point values on each blower.

MOV Determination As it was already discussed, the most-open valve position is not necessarily determined as the opening of a single valve. When very large numbers of valves are involved or when some of the valves in the system can be stuck, failed, or manually controlled, it is better to use alternative methods that are selected to match the needs of the process.

For example, in Figure 8.28i, MOV is arrived at as the measurement of the average opening of the five most-open valves in the system, corrected for the total flow (average opening of all valves) in the system. An alternative approach could have been to determine MOV as the average opening of the most-open valves on each of the sub-branches of the distribution system.

It is also recommended to use the more conservative configurations (lower MOV, higher number of averaged valves) during commissioning, because it is more important to gain the trust of the operators than to save energy. In other words, first, the operator's suspicion of such "advanced" controls should be overcome.

After that, the operators should be trained to concentrate on equipment maintenance and on learning the "personality" of the process. They should also develop the habit to first remove the valve from the control algorithm before switching it to manual and, once maintenance is completed, to clear the latched fault switch to return the valve back into the system.

Optimization of the system should occur only after the full confidence of the operators has been gained. This control system also has built-in self-diagnostics capability, because it can be programmed to report any valve that does not move at all for longer periods of time, and can report valves that are cycling or are excessively open (90–100%).

DISTRIBUTION CONTROLLING VPC (MOV2)

As was illustrated in Figure 8.28b, the task of a MOV2 controller is only to distribute a flow that is determined by factors outside of its control. Examples include the distribution of cooling water returning from the plant among a number of cooling towers or chillers or the distribution of wastewater among treatment units.

The incoming wastewater is often received as gravity flow, which is dependable, but its head pressure is low and cannot be changed. In such applications, it is essential to minimize the pressure drop introduced by the distribution valves. In such

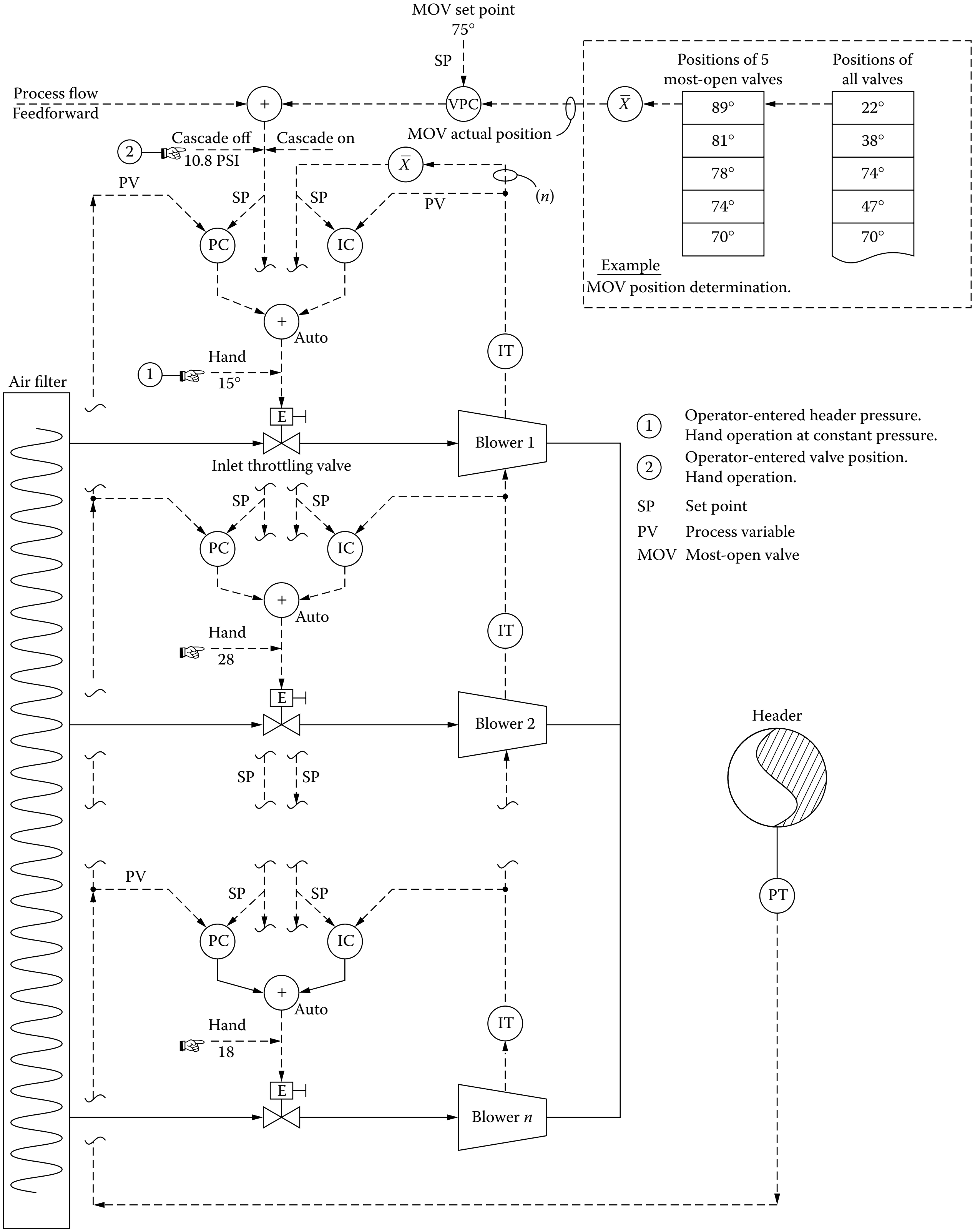


FIG. 8.28i

The structure of a particular unit operations controller (MOV1), serving to match the gas supplied into a distribution header by a number of blowers, to the demand represented by the process users on the header.

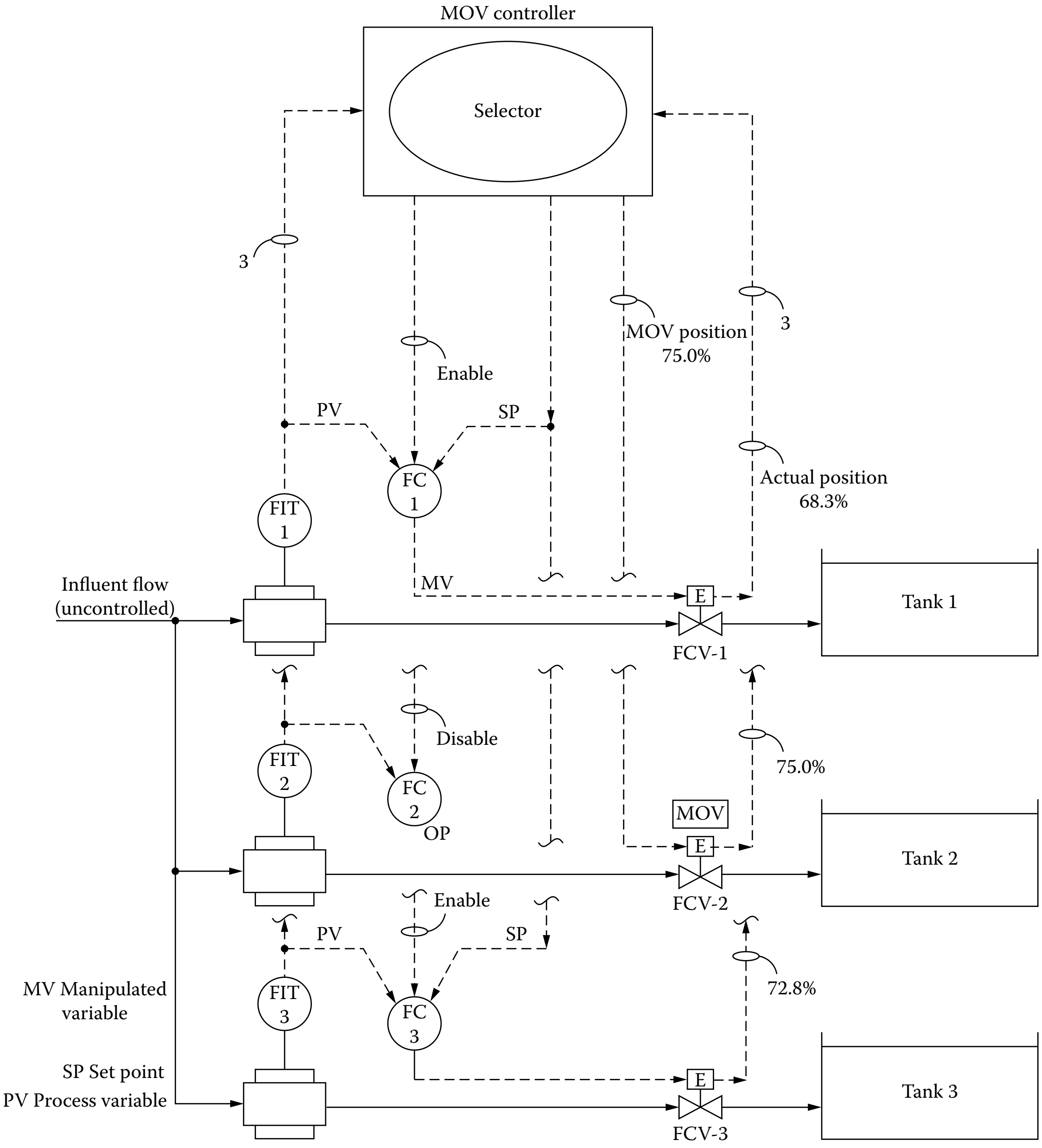


FIG. 8.28j
The structure of a particular unit operations controller (MOV2), serving to distribute an uncontrolled water stream among a number of receivers.

applications, the MOV2 controller serves to identify the most-open valve and compares that opening to its set point, which is the desired maximum opening of, say, 80%. The MOV2 controller also measures the flow through the selected MOV and, if the other users are of equal capacity, uses that flow as the set point for the other flow control loops that are throttling the other user valves. If the other users are of different capacities, the set point is sent through a ratio relay.

Example

If a MOV2 controller is set at 80%, its operation during a time when the identity of the most-open valve is changing is as follows: If FCV1 in Figure 8.28j is being controlled at 75% when FCV2 opens to 75.1% (possibly, because of an increase of pressure drop due to plugging in its piping), the measurement selection of MOV2 is switched to FCV2. As a consequence, the control of FCV2 is switched from its flow controller (FC-2 in Figure 8.28j) to the MOV controller, which keeps it at an opening of 75%. At the same time of this switching, the FVC-1 flow control valve is returned to be controlled by its own flow controller (FC-1).

If the tanks in Figure 8.28j are of different sizes or if they supply treatment facilities of different capacities, the set points of their flow controllers are sent through ratio relays that are set to reflect their capacity ratios.

Failure Modes and Applications

Once the MOV2 controller has positioned the valves to give the required flow distribution ratio, its valves should seldom move. In an MOV2 controller, a failure such as a blockage causes the blocked valve to open and become the MOV. The flow through the plugged valve will be low because of the blockage, yet that flow will then become the set point for the other valves so they will close. This will cause a backup and potentially an overflow on the influent side of the process.

For this reason and other maintenance-related reasons, it is desirable to provide the MOV2 controller with the means to turn off its controls, while equipment failure is being corrected.

MOV2 controllers are suitable for open-channel flow distribution applications. They are also suitable for distributing gases or liquids that are transported by pumps and blowers, including positive displacement types, because it is the pumps or blowers that control the flow; the valves only distribute it.

The MOV2 controllers can maintain other variables besides the opening of the most-open valve. If their measurement signal is switched from an MOV position to the back-pressure detector, it will maintain the back-pressure on the distribution system. Similarly, if the MOV2 measurement is switched to a level measurement signal from the supply tank of the distribution system, it will maintain that level. Needless to say, when the measurement is switched away from the MOV, the controller will abandon the control of the valve position.

Yet another option is to send the measurements of MOV, level, or pressure to a high-low limit envelope and allow the controlling active measurement to be selected on the basis of some predetermined process limit consideration.

The features of MOV controllers can be combined or control can be automatically switched from supply-demand matching (MOV1) to distribution (MOV2) control and back. It is also possible to automatically alter or select other process variables to temporarily replace the MOV measurement as the controlled variable or to combine the effect of some measurements in the set point as trimming variables.

DESIGN, START-UP, AND MAINTENANCE

For an in-depth discussion of the main components of distribution systems, such as the control of blowers, compressors, and pumps and for the dynamic characteristics of actuators, dampers, and valves, the reader is referred to the corresponding sections of this handbook. In addition to that, a few design recommendations that are specific to distribution system controls are listed here.

Actuators and Valves

Adjustable frequency drives (AFDs) are most often used, although DC and pneumatic actuators are also still in use. DC units tend to be high-maintenance items, while pneumatic units generally have high hysteresis. The actuators for the capacity controls of blowers should be fast and be provided with positioners. Electrohydraulic actuators, available from Acutrol or REXA, can meet this requirement.

Control

When designing load-sharing controls, it is preferred that they measure the power consumption rather than the speed in equalizing the power consumption of multiple variable-speed units. The set point of these power equalization controllers should be the average power consumption of the operating units. As shown in Figure 8.28i, the output of the power equalization controllers (IC) is then combined with the output signal of the capacity controller (PC).

When these two signals are combined, the capacity controller should maintain its primary role and the interaction between the two loops should be decoupled. For these reasons, the power controller output is often used only as a slow bias through a low-gain integrator.

When suction dampers or guide vane controls instead of variable-speed controls are used, reciprocal gain compensation is recommended, so that as the gain of the final control element increases (or decreases), the loop gain will stay constant.

One should also consider to modify the control algorithm such that when the blower is operating near its surge or overload lines, the gain of its capacity controller is temporarily reduced.

Operator Interface In many installations, it is desirable to suppress the measurement range of controllers by elevating the zero of the measurement signal. In such cases, the controller might consider, say, a header pressure reading of 70 kPa as 0 and 100 kPa as 100%. In such applications, it is desirable that on the operator's interface the actual header pressure be displayed with a range of 0–100 kPa, so that the operator can distinguish a true depressurized condition from a zero reading on the controller range with elevated zero.

Master Hand Controls If the distribution system consists of many valves and the restart sequence after a power failure is initiated manually, it is desirable to provide master controls for the restarting of all valves and other instruments by a single switch or other control device. In other applications, it might be convenient to approach this on a unit operation basis, so that a tank or other unit operation can be taken out or returned to automatic control by a single switch.

Blowers

When meeting the process load with two blowers, for smooth switching it is desirable to size them, so that at its maximum flow, one would deliver 110% of the sum of the minimum flows of two blowers. This requirement can bring the operation close to either the surge curve (see Figure 8.15i in Section 8.15) or the overload (choke) curve of these units.

During commissioning, the diffusers should be protected against overpressuring and from excessive temperatures. With no water pressure on the diffusers, the blower pressure can be sufficient to pop off or burst the diffusers. Membrane diffusers with PVC piping have a temperature limit of 70°C (160°F).

Fault Detection and Response

Distribution control system design should consider the response to power failure, sensor or actuator failure, valve plugging, and operator errors. The control system should also include the capability of both predicting and signaling the need for maintenance. Such self-diagnostics can be based on the comparison of valve opening and valve flow and also on the comparison between control signal and valve opening.

If such evaluation flags the need for valve maintenance, automatic checking could include the initiation of a 5 or 10% change in control signal and the checking if the valve opening changes in response to it. Similarly, such an automatic test can include checking if the flow has changed in response to the initiation of this temporary step change in the control signal.

Start-Up

The commissioning of distribution system controls is not very different from the start-up of any other unit operation. Commissioning should be started by leak and functional testing of all piping and equipment components. This should be followed by calibration, and when that is concluded, it should be continued by determining the values of multiplier, summer, bias, feedforward, and controller tuning constants.

In case of air distribution, it is desirable to provide the means for air venting, so that header controls can be tested without affecting the process.

When starting up in the winter, the plans for commissioning should take the low temperatures and the probability of ice formation into account.

CONCLUSIONS

The MOV1 controller is generally suited to applications where the flow can be controlled by the MOV1's valves. This is often the case, so that the MOV1 controller is widely used.

However, the MOV1 controller is usually not suitable for applications such as positive displacement (PD) pumps pumping incompressible liquid. In this situation there is nothing to compensate for inevitable differences between required and pumped flows. However, if the flow is compressible gas with PD blowers, the difference can be absorbed by a changing header pressure that the controls detect and correct. Or if the pumps or blowers are centrifugal types, then the changing header pressure shifts their operating points, which corrects the difference.

MOV1 controllers are not typically used with open-channel flow distribution, because usually the pressure in an open channel cannot be varied significantly. In an open channel, pressure equates directly to level, and an open channel suitable for variable pressures is likely to be uneconomical; a pipe is likely to be more suitable.

Bibliography

Buzzard, W. S., *Flow Control,* Research Triangle Park, NC: ISA, 1994.

Buzzard, W. S., "Automatic Flow Splitting Systems," Fischer & Porter Bulletin 91-53G-07, undated.

Carlsson, B. and Lindberg, C.-F., "Some Control Strategies for the Activated Sludge Process," Systems and Control Group, Uppsala University, 1998.

EPA, *Design Manual for Fine Pore Aeration Systems,* Washington, D.C.: EPA/625/1-89/023, 1989.

Harris, G. et al., *A Significant Advancement in the Control of Aeration Systems*. WEFTEC, 1999.

Vinton, R. H., *Most-Open Valve Control and Cascade Control of Multiple Compressors Improve Aeration Efficiency & Cut Costs,* WEFTEC, 1996.

8.29 Heat Exchanger Control and Optimization

B. G. LIPTÁK (1970, 1985, 1995, 2005)

INTRODUCTION

The transfer of heat is one of the most basic and best understood unit operations of the processing industries. Heat can be transferred between the same phases (liquid to liquid, gas to gas), or phase change can occur on either the process side (in case of condensers, evaporators, and reboilers) or the utility side (in case of steam heater) of the heat exchanger. With the exception of evaporators, which are discussed in Section 8.23, the control and optimization of each of these systems will be discussed in this chapter.

The section is started with some general aspects, such as plantwide heat audits, heat exchanger safety, temperature measurement advances, tuning, and heat-transfer dynamics. Next, it describes the basic components and controls of liquid-to-liquid heat exchangers, steam heaters, condensers, reboilers, and vaporizers. The section is concluded by the description of more advanced control techniques, including override, cascade, feedforward, adaptive gain, model-based, and multipurpose systems.

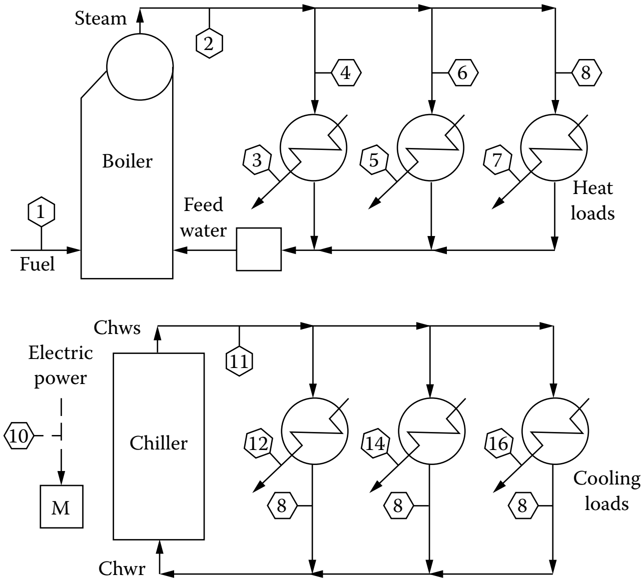


FIG. 8.29a
A plantwide energy audit can determine the useful portion of the energy that is represented by fuel consumed at the boiler(s) and by electric energy used by the chiller(s) that actually reaches the process.

GENERAL CONSIDERATIONS

In a processing plant, the heat exchanger as a unit operation is part of larger plantwide unit operations, serving the heating or cooling needs of all the processes in the plant. By the use of BTU flowmeters (Section 2.3 in Chapter 2 in the first volume of this handbook), one can obtain a plantwide energy audit (Figure 8.29a) and provide historical plots of both the total amounts of energy used and of the efficiency at which it was used. If the individual heat exchanger controls of the plant are optimized, such plantwide information can be useful in such decisions as when to clean heat-transfer surfaces or what the optimum cooling water temperature or steam supply pressure and superheat values are.

Temperature Detection and Transmission

Chapter 4 of the first volume of this handbook is devoted to the description of the capabilities of the various temperature detectors available. The vast majority of temperature measurements are made with either a thermocouple (TC) or a resistance temperature detector (RTD). Higher purity materials and improved manufacturing processes have provided sensors that more closely match theoretical curves and exhibit lower drift than sensors of the 1980s and the 1990s.

More advanced transmitters incorporate the ability to match the sensor to the transmitter to minimize this error. Typical accuracies are about ±0.4°F. For higher precision, at least one vendor offers a bath calibration technique that allows the transmitter to capture actual values output by the RTD or thermocouple at specific temperatures. This method provides system accuracies of about ±0.02°F (0.1°C) for RTDs and about ±2°F (1°C) for thermocouple measurements.

The array of intelligent temperature transmitters on the market seems almost endless. Some common features of the leading models are universal inputs from any TC, RTD, mV, resistance, or potentiometer; loop-powered with 0–20/4–20 mA output; digital outputs; and configuration with pushbuttons, PC software, or a handheld configurator. Choices must be made concerning the protocol required: HART, Foundation Fieldbus, Profibus, vendor proprietary, Ethernet, or just

4–20 mA. Some field locations will benefit from local indication, and this feature is optional with most manufacturers.

Direct connection of temperature sensors to input subsystems of DCSs or PLCs is an alternative to using a temperature transmitter for each measurement. This may sound like a way to cut costs. In actuality, on an installed basis, it is more expensive, far less accurate, and not as robust. The benefits of using transmitters include higher precision, sensor-transmitter systems calibrated to the range of interest, better RFI immunity and noise rejection, transmitter diagnostics, lower wiring costs, less expensive I/O cards, faster loop checks, and shorter start-ups.

Process Characteristics

Heat transfer is one of the best understood processes, and as such, it has good potentials for modeling and optimization. Before discussing the control of heat exchangers, the degrees of freedom, scaling, dynamics, and tuning of this unit operation will be briefly discussed.

Degrees of Freedom Chapter 2's Section 2.1, in connection with Figures 2.1r and 2.1s, provides a detailed discussion of the determination of the degrees of freedom of any process. The degrees of freedom of a process define the maximum number of independently acting automatic controllers that can be placed on a process.

Figure 8.29b shows a steam heater with its variables and defining parameters. The temperatures and flows are variables, and the specific and latent heats are parameters. The available degrees of freedom are determined by subtracting the number of system-defining equations from the number of process variables:

$$\text{degrees of freedom} = (\text{number of variables}) - (\text{number of equations}) \quad \mathbf{8.29(1)}$$

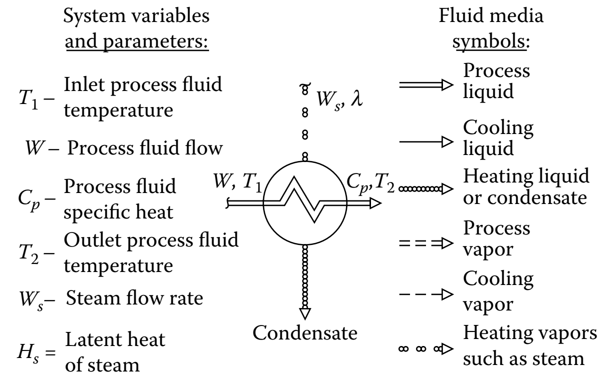


FIG. 8.29b
In a steam heater there are four process variables (two flows, two temperatures) and one defining equation (conservation of energy), hence the number of degrees of freedom of that process is three.[1]

In this case, there are four process variables and one defining equation, which is that of the conservation of energy, based on the first law of thermodynamics:

$$HsWs = C_pW(T_2 - T_1) \quad \mathbf{8.29(2)}$$

Therefore, this system has three degrees of freedom; thus, a maximum of three automatic controllers can be placed on it.

In a liquid-to-liquid heat exchanger, there are four temperature and two flow variables with still only one (conservation of energy) defining equation, resulting in five degrees of freedom:

$$C_pF(T_{h1} - T_{h2}) = C_{pc}F_c(T_{2c} - T_{1c}) \quad \mathbf{8.29(3)}$$

In a steam-heated reboiler or in a condenser cooled by a vaporizing refrigerant (assuming no superheating or supercooling), there are only two flow variables and one defining equation:

$$HsWs = H_1W_1 \quad \mathbf{8.29(4)}$$

In these cases there is only a single degree of freedom, and therefore on these processes only one automatic controller can be used.

In the majority of installations, fewer controllers are used than the available degrees of freedom, but every once in a while mistakes are made by using too many controllers and thereby overdefining the process.

Scaling, Gain, and Time Constant Scaling is the conversion from engineering units to fractions or percentages. For the scaling of a heat-transfer equation, a detailed example has been provided in connection with Figure 8.5r.

The steady-state gain of a heat-transfer process is the ratio of the percentage change in controlled temperature that result from a 1% change in the flow through the control valve that controls that temperature.

Because the heat-transfer surface of a heat exchanger is fixed, the more heat is to be transferred, the less efficient the process becomes and therefore the lower will be the process gain (*Gp*). Assuming that in the control loop the sensor gain (*Gs*) and the controller gain (*Gc*) are constant, a stable loop operation can be obtained if the control valve gain (*Gv*) rises with load, thereby compensating for the drop in the process gain and maintaining the loop gain constant. The gain product of a well-tuned control loop is about 0.5, as illustrated in Figure 8.29c.

The time constant of any process is the result of its capacitance and resistance. In case of a heat exchanger, the process outlet temperature is usually the controlled variable, and the flow rate of the heat-transfer fluid is the manipulated variable. The time constant of an exchanger is determined by the mass and the specific heat of the tube material, the mass flow and the specific heat of the process and utility streams, and their heat-transfer coefficients. A good approximation of

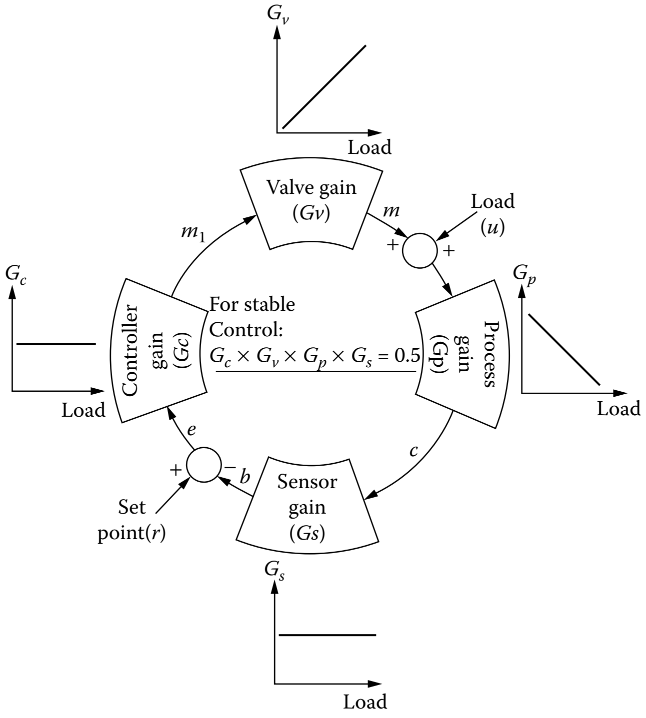


FIG. 8.29c
If the process gain varies with load, such as is the case of heat-transfer processes, the gain product of the loop can be held constant by using a valve whose gain variation with load will compensate for the process gain variation.

a heat exchanger's time constant can be obtained with the empirical formula:

$$\tau = K \frac{m_T \cdot C_{pT}}{F_w \cdot C_{pw} \cdot F_c \cdot C_{pc}} \qquad \textbf{8.29(5)}$$

where

τ = the time constant in seconds
K = 100 for English units, and 200 for SI units
m_T = the total mass of the tubes in lb (kg)
C_{pT} = the specific heat of the tube material in BTU/lb_m (kJ/kgK)
F_w = the flow rate of the warm stream in lb/sec (kg/sec)
C_{pw} = the specific heat of the warm fluid in BTU/lb_m (kJ/kgK)
F_c = the flow rate of the cold stream in lb/sec (kg/sec)
C_{pc} = the specific heat of the cold fluid in BTU/lb_m (kJ/kgK)

In process industry heat exchangers, the time constants are usually between 5 sec and 2 min.

Dead Time and Tuning Before the new flow rate or temperature of the heat-transfer fluid can start to have an impact on the heat balance, first it has to get to the heat-transfer surface of the exchanger. Therefore, dead time is also called transportation lag, because it is the time it takes for the heat-transfer fluid to transport (displace) the contents of the exchanger and its associated piping.

The dead time is the worst enemy of control, because until it has expired, a change in the heat-transfer fluid flow (or temperature) will not even start to have an observable effect on the process. For a heat exchanger, the dead time is usually between 1 and 30 sec. When the equipment is correctly designed, the dead time is much less than the time constant.

Tuning is discussed in detail in Sections 2.35 to 2.38 in Chapter 2 in this volume. The controller gain is adjusted so that the loop gain (the gain product of the loop components) will be around 0.5. In case of variable gain processes, such as heat transfer, one usually tries to compensate for the gain variations of the process by using a control valve characteristic that compensates for it. In case of the more critical processes, adaptive (Section 2.19) and model-based (Sections 2.13 to 2.18 in Chapter 2) controls can be used.

Control loops can be tuned by considering the dynamics of only the process (open-loop method) or evaluating the response of the complete loop (closed-loop method). Because the period of oscillation of all control loops (P) is some multiple (around 3.5 to 4.0) of their dead time, the integral and derivative settings of PID controllers are generally adjusted as a function of the dead time of the process.

For noninteracting control loops with zero dead time, the integral setting (minutes/repeat) is about 50% and the derivative is about 18% of the period of oscillation. As dead time rises, these percentages drop. If the dead time reaches 50% of the time constant, I = 40%, D = 16%, and if dead time equals the time constant, I = 33% and D = 13% of the period of oscillation. For a comparison of the various tuning criteria, refer to Table 2.35l, and for the load and set point responses of the various tuning methods, see Figures 2.35o, p, and q in Chapter 2.

As will be shown later, when tuning the feedforward portion of a control loop, one has to separately consider the steady-state portion of the heat-transfer process (flow times temperature difference) and its dynamic compensation. The dynamic compensation of the steady-state model by a lead/lag element is necessary, because the response is not instantaneous, but affected by both the dead time and the time constant of the process.

Safety

Heat exchangers are a class of process equipment requiring special relief considerations, because of the potential need for protection against thermal expansion, external fire, blocked outlet, and tube rupture cases.

Blocked-In Exchangers Heat exchangers frequently have valves located on both their inlet and outlet piping. When these valves are all closed, the exchanger is "blocked in." If the cold side of the heat exchanger can be blocked in, relief devices are installed to provide protection against thermal expansion of liquids in the exchanger. This is always done for the cold side of an exchanger where the liquid can be heated by the hot fluid on the other side, or can be heated by ambient temperature, while sitting with the inlet and outlet valves closed.

No relief device is necessary for the protection of either side of an exchanger that cannot be blocked in. In such installations it is assumed that the relief of the unit is taken care of by the relief device on the related tank or equipment.

Liquid Refrigerants When liquid refrigerants are used, a relief device should always be provided for the protection of the refrigerant side, if the refrigerant side can be blocked in and if the vapor pressure of the refrigerant can exceed the design pressure of the exchanger, when its temperature rises to that of the hot side.

A relief device should also be provided whenever the vapor pressure of the material flowing at 100°F (37.8°C) is greater than the design pressure of the exchanger. This recommendation is somewhat site-specific and is based on an assumed (maximum) ambient temperature of 100°F (37.8°C). This temperature should be modified according to the geographical areas involved.

Gas-Fired Tubular Heaters Direct gas-fired tubular heaters are always protected by relief valves on their tube side. The valve is normally sized for the design heat-transfer rating of the heater and must initially handle a fluid rate corresponding to the rate of thermal expansion in the tubes when they are blocked in.

When designing fired heaters, there should be no block valve on its outlet. This is because PRVs for high-temperature services exceeding 550°F are not available with dependable seat and seal materials.

Tube Rupture Consideration should be given to relief protection of low-pressure equipment in the event an exchanger tube should rupture because of corrosion or vibration or due to thermal shock. ASME Code, Section VIII, Division 1, Paragraph UG-133(d) requires such protection.

This consideration is particularly critical when the low-pressure side design pressure is less than the operating pressure on the high-pressure side. In terms of high- and low-pressure side design pressures, PRV protection against tube rupture is recommended, if the design pressure of the low-pressure side is less than 77% of the high-pressure side.

For advice on the sizing of PRVs to protect against overpressure caused by tube rupture, refer to Section 7.15 in Chapter 7 in the first volume of this handbook. The PRV that is to protect the exchanger should be located either directly on the exchanger, or very close to it.

BASIC CONTROLS

In the following paragraphs, the basic controls of liquid-liquid heat exchangers, steam heaters, condensers, vaporizers, and reboilers will be discussed. This will be followed by a description of more advanced controls for these unit operations.

Liquid-Liquid Heat Exchanger Controls

Figure 8.29d and Equations 8.29(7) and (8) illustrates that the total heat transferred (Q) from the hot process to the cold coolant fluid is dependent on the overall heat-transfer coefficient (U), the heat-transfer area (A), and the log mean temperature difference (ΔTm). Manipulation of any of these three variables can affect the total heat transfer.[2]

The dead time of a heat exchanger equals its volume divided by the flow rate through it. As process flow increases, the process dead time is reduced, while the loop gain is also decreased. If the controlled variable (Th_2 in Figure 8.29d) is differentiated with respect to the coolant flow (manipulated variable Fc), the steady-state gain of the process is

$$\frac{dTh_2}{dF_c} = \frac{C_{pc}}{F_h C_{ph}} \qquad \mathbf{8.29(6)}$$

where
C_{pc} is the specific heat of the cold fluid
C_{ph} is the specific heat of the hot fluid

If $C_{pc} = C_{ph}$, the steady-state gain equals $1/F_n$. Therefore, as the process flow (load) drops, the same amount of valve adjustment by TIC in Figure 8.29d will have more effect, because the process fluid will spend more time in the exchanger. As load is reduced, the exchanger becomes relatively oversized, and therefore, it becomes a faster and more effective heat-transfer device.

As a drop in load tends to increase the process gain — that is, tends to make the loop more sensitive and more prone to cycling — an increase in load does the opposite. As load rises,

$$Q = UA\Delta Tm = UA\,\frac{(Th_1 - Tc_2) - (Th_2 - Tc_1)}{\text{In}\,((Th_1 - Tc_2)/(Th_2 - Tc_1))} \qquad \mathbf{8.29(7)}$$

$$Q = Fh\;\; Cph\,(Th_1 - Th_2) = Fc\,Cpc\,(Tc_2 - Tc_1) \qquad \mathbf{8.29(8)}$$

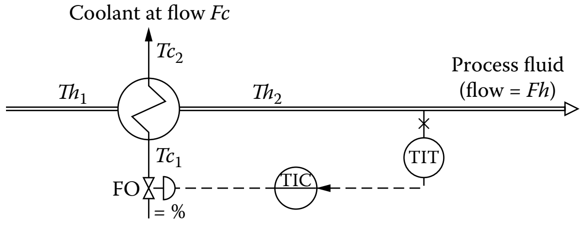


FIG. 8.29d
The feedback-type temperature control of a process cooler involves the control of a nonlinear and variable gain process.[1]

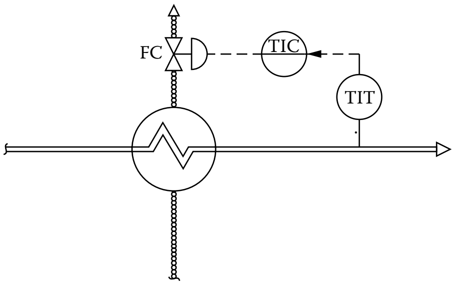


FIG. 8.29e

The feedback temperature controller of a process heater can throttle either the inlet or the outlet of the heating media on this nonlinear and variable gain process.[1]

the exchanger becomes less and less effective, more and more undersized; therefore, it takes more time for the TIC to make a correction. An increase in load thus will make the loop sluggish, as the residence and dead times are reduced and the process gain is lowered.

As the process gain of all heat exchangers varies with load, it would not be possible to tune the TIC in Figure 8.29d for more than one load, if all other loop gains remained constant, while the process gain varied. In order to minimize cycling, the TIC is usually tuned for the minimum load (maximum process gain). Therefore, if not automatically compensated, it can act sluggish at higher loads.

As the process gain drops with rising load, the total loop gain can be held relatively constant by introducing a loop component whose gain rises with load. Such an element is an equal-percentage valve, which can compensate for the variable gain process. This is a good solution if the temperature difference through the exchanger ($Th_1 - Th_2$) is constant.[3] If it is not constant, feedforward compensation of the gain is required, as will be discussed later.

Component Selection Figures 8.29d and 8.29e illustrate cooler and heater control systems with the control valve mounted on the exchanger inlet and outlet, respectively. From a control quality point of view, it makes little difference whether the control valve is upstream or downstream of the heater. The inlet side is usually preferred, because this allows the exchanger to operate at a lower pressure than that of the return header.

It is generally recommended that positioners be provided for these valves to minimize the effects of valve friction (hysteresis and dead band) effects. The use of equal-percentage valve characteristics is recommended, because it contributes to maintaining the control loop gain relatively constant under changing throughput conditions. Equal-percentage trims maintain a constant relationship between valve opening and temperature change (reflecting load variations).

In the majority of installations, a three-mode controller would be used for heat exchanger service. The use of the derivative or rate action is essential in slower (long time-lag) systems or when sudden changes in heat exchanger throughput are expected.

Because of the relatively slow nature of these control loops, the proportional band setting is usually wide to maintain stability (usually approaching 100%). This means that the valve will be fully stroked only as a result of a substantial deviation from the set point. It is for this reason that the integral control mode is required to correct for temperature offsets caused by process load changes. Other variables can give the appearance of load changes, such as inlet temperature and header pressure changes of the heat-transfer medium.

The Thermal Element The selection and location of the thermal element are also important. This element must be placed in a representative location, without increasing measurement time lag. In reference to Figure 8.29d, this would mean that the bulb should be located far enough from the exchanger for adequate mixing of the process fluid but close enough so that the introduced delay will not be substantial. If the process fluid velocity is 3 ft/s (0.9 m/s), then a 1 sec distance-velocity lag is introduced for each 3 ft (0.9 m) of pipe between the exchanger and the bulb.

This lag can be one of the factors that limits the dynamic performance of the system, but it is not the only thermal lag in the system. In order to change a temperature measurement, first heat must be transferred into the thermal bulb through its fixed area.

For example, a typical filled bulb might have an area of 0.02 ft^2 (0.0018 m^2) and the heat capacity of 0.005 BTU/°F (9.49 J/°C). If this bulb, which has a bare-bulb diameter of $^3/_8$ in. (9.375 mm), is immersed in a fluid having a heat-transfer coefficient of 60 BTU/h/°F/ft^2 (1.23 MJ/h/°C/m^2) and the process temperature is changed at a rate of 25°F/min (13.9°C/min), the dynamic lag can be calculated. To do so, first the amount of heat flowing into the sensor is determined:

$$\begin{aligned} q &= \text{(rate of temperature change) (bulb heat capacity)} \\ &= (25)(60)(0.005) = 7.5 \text{ BTU/h } (7.9 \text{ kJ/h}) \end{aligned} \qquad \mathbf{8.29(9)}$$

The dynamic measurement error is calculated by determining the temperature differential across the fluid film surrounding the bulb that is required to produce a heat flow of 7.2 BTU/h (7.9 kJ/h):

$$q = Ah\,\Delta T \qquad \mathbf{8.29(10)}$$

Therefore,

$$\Delta T = q/Ah = 7.5/(0.02)(60) = 6.25°\text{F } (3.47°\text{C}) \qquad \mathbf{8.29(11)}$$

If the rate of process temperature change is 25°F/min (14°C/min) and the dynamic error based on that rate is 6.25°F (3.47°C), the dynamic time lag is as follows:

$$t_o = 6.25/25 = 0.25 \text{ minutes} = 15 \text{ seconds} \qquad \mathbf{8.29(12)}$$

This lag can also be calculated as

$$t_o = \frac{\text{(bulb heat capacity)}}{\text{(bulb area)(heat-transfer coefficient)}} = \frac{60 \times 0.005}{0.02 \times 60}$$

$$= 0.25 \text{ minutes} \qquad \mathbf{8.29(13)}$$

Bulb time lags vary from a few seconds to minutes, depending on their mass, heat-transfer area, and the process fluid being detected. Measurement of gas temperatures at low velocity involves the longest time lags, and measuring water (or dilute solutions) at high velocity results in the shortest lags.

The addition to the sensor lag of a thermowell will further increase the total lag time, but in most industrial installations, thermowells are necessary for reasons of safety and maintenance. When they are used, it is important to eliminate any air gaps between the bulb and the socket.

One method of reducing time lag is by miniaturizing the sensing element. Conventional thermometers are not suitable for all temperature measurement applications. For example, the accurate detection of high-temperature gases (500 to 2000°C) at low velocities (3 to 5 ft/s) is a problem, even for thermocouples. This is because conductance through the lead wires and radiation both tend to alter the sensor temperature faster than the low velocity gas flow can resupply the lost heat. In such applications (fuel cell controls are prime examples), optical fiber thermometry is a good choice.

Thermocouples are usually not accurate enough for the precise measurement of temperature differences and are not fast enough to detect high-speed variations in temperature. RTDs can detect temperature differences of 10°F at a measurement error of ±0.04°F. If high-speed response is desired, thermistors or infrared detectors should be considered.

In the conventional control loop, the measurement lag is only part of the total time lag of the control loop. For example, an air heater might have a total lag of 15 min. Of this lag, 14 min represent the *process lag,* 50 sec the bulb lag, and 10 sec is the control valve's contribution to the total time lag.

Three-Way Valves The limits within which process temperature can be controlled are a function of the nature of load changes. In many installations, the process time lag in the heat exchanger is too great to allow for effective control during load changes. In such cases, it is possible to circumvent the dynamic characteristics of the exchanger by partially bypassing it and blending the warm process liquid with the cooled process fluid, as shown in Figure 8.29f. Increased speed of response and some cost savings are the main motivations for considering three-way valves in such services.

The bulb time lag discussed in the previous paragraph has an increased importance in system configurations, because this lag represents a much greater percentage of the total loop lag time than in the previously discussed two-way valve installations.

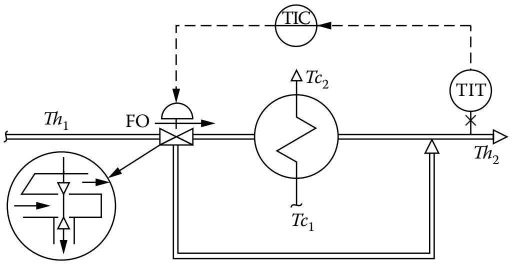


FIG. 8.29f
When a diverter valve is used to control a cooler, it does not eliminate nonlinearity, but it does speed the response and minimize fouling.[1]

The use of a three-way valve such as that shown in Figure 8.29f does not change the nonlinear nature of the heat exchanger, because the process gain will still vary with load, but the dynamic response of the loop will still be improved.

This is because the bypass will shorten the time delay between a change in valve position and the response at the temperature sensor. Another benefit from the use of a three-way valve on the process side is that the coolant is not throttled, which keeps the heat-transfer coefficient up while minimizing fouling. The disadvantages include increased pumping costs and lower temperature differences between supply and return water.

As illustrated in Figures 8.29f and 8.29g, either a diverter or a mixing valve can be used as a three-way valve. Stable operation of these valves is achieved by using flow to open valves in both cases. If a mixing valve is used for diverting service or if a diverting valve is used for mixing, the operation becomes unstable because of the "bathtub effect" (if the flow direction is reversed, the fluid itself will try to push the valve plugs closed). Therefore, it is not good enough just to install a three-way valve, but the selection (mixing or diverting design) has to match the particular service.

Three-way valves are unbalanced valve designs and are normally provided with linear ports. Their unbalanced nature places a limitation on the allowable shut-off pressure difference across the valve, and the linear ports eliminate the potential of them serving to compensate for the variable gain of the process.

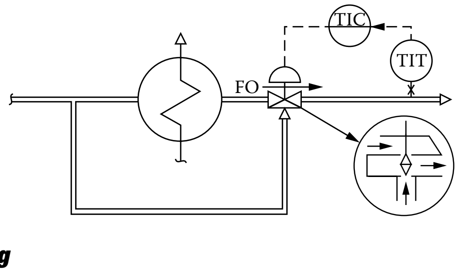


FIG. 8.29g
A mixing valve can be used to control the cooler. It has a flow to open inner valve for stability.[1]

Misalignment or distortion in a three-way control valve installation can cause binding, leakage at the seats, dead band, and packing friction. Such conditions commonly arise when three-way valves are used in high-temperature applications. This is because, having been installed at ambient conditions and rigidly connected at three flanges, the valve cannot accommodate pipeline expansion caused by high process temperatures, and therefore distortion results.

Similarly, in mixing applications, when the temperature difference between the two ports is substantial, the resulting differential expansion can also cause distortion. For these reasons, the use of three-way valves at temperatures above 500°F (260°C) or at differential temperatures exceeding 300°F (167°C) is not recommended.

The choice of three-way valve location is normally based on pressure and temperature considerations, with the upstream location (Figure 8.29f) usually being favored for reasons of uniformity of valve temperature. When the overriding consideration is the desire to operate the exchanger at a high pressure, the downstream location might be selected.

Cooling Water Conservation Figure 8.29h modifies the control systems shown earlier by adding a controller to serve cooling water conservation. The TIC-01 in Figure 8.29h is set at a relatively high value to maximize the outlet cooling water temperature, thereby minimizing the rate of cooling water usage.

When using this approach, one should be careful to make sure that the cooling water contains chemicals to prevent tube fouling under higher return water temperature conditions. If the cooling water supply is sufficient and the TIC-01 set point is properly selected, this system on the one hand will conserve water and, on the other, will also protect against excessively high outlet water temperature.

Unfortunately, this configuration will not yield stable control, because controlling temperature through manipulation of its own flow results in a limit cycle. The cause of this instability is that the output of TIC-01 affects the dead time of the process that it controls. This comes about because if TIC-01 detects a temperature rise, it opens its valve, causing a sudden drop in temperature as the process dead time is also reduced.

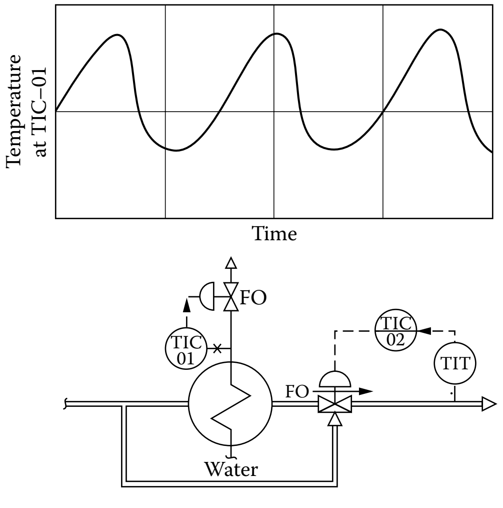


FIG. 8.29h
Conservation of cooling fluid is provided at the price of stability.[1]

As TIC-01 senses this sudden drop in temperature, it will close its valve, but because this also increases the process dead time, the resulting temperature will rise slowly.[4] This limit cycle, consisting of segments of slow rise and fast fall, will also affect the performance of TIC-02 through interaction between the loops. The amplitude of the cycle can be reduced by increasing the proportional band (reducing the gain) of TIC-01, which will also increase the period of oscillation.[4]

The only way to eliminate this oscillation altogether is to fix the process dead time. This can be done by the addition of a recirculating pump, which will return part of the heated water back to the inlet. As can be seen, there is no easy way to make this configuration stable, and for that reason, the best solution is to avoid its use.

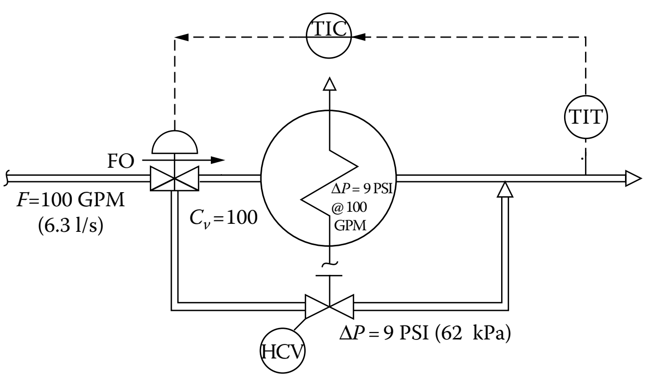


FIG. 8.29i
When three-way valves are used, it is recommended to install a manual balancing valve in the bypass.[1]

Balancing the Three-Way Valve When using three-way control valves, it is recommended that a manual balancing valve be installed in the exchanger bypass, as shown in Figure 8.29i. The opening of this valve should be so adjusted that its pressure drop equals that of the exchanger at normal load.

The resistance to flow in such installations will be maximum when one of the paths is closed and the other is fully open, whereas minimum resistance will be experienced when the valve divides the flow equally between the two paths. For example, if a diverting valve with $C_v = 100$ and process fluid flow rate of 100 gpm (6.3 l/s) has a 9 psi (6.21 kPa) pressure drop at full flow through either the exchanger or the balancing valve, the equivalent coefficient for the exchanger (or the balancing valve) is calculated as

$$C_e = \frac{\text{flow}}{\sqrt{\text{pressure drop}}} = \frac{100}{\sqrt{9}} = 33.3 \qquad \textbf{8.29(14)}$$

Therefore, in either extreme position (closed or full bypass), the total system resistance expressed in valve coefficient units is

$$\frac{1}{(C_t)^2} = \frac{1}{(C_v)^2} + \frac{1}{(C_e)^2} = \frac{1}{(100)^2} + \frac{1}{(33.3)^2} \qquad \mathbf{8.29(15)}$$

$$\therefore C_t = 31.7$$

When the valve divides the flow equally between the two paths, because of the linear characteristics of three-way valves, its coefficient at each port will be $C_v = 50$. The equivalent coefficient ($C_e = 33.3$) of the exchanger and balancing valve being unaffected, the total system resistance in valve coefficient units is $2C_t$:

$$\frac{1}{(C_t)^2} = \frac{1}{(C_v)^2} + \frac{1}{(C_e)^2} = \frac{1}{(50)^2} + \frac{1}{(33.3)^2} \qquad \mathbf{8.29(16)}$$

$$\therefore 2C_t = 55.6$$

If the total pressure drop through the system is calculated when the valve is in its extreme and when it is in its middle position, handling the same 100 gpm (6.3 l/s) flow, the following pressure drops are found:

$$\Delta P_{\text{extreme}} = \left(\frac{\text{flow}}{C_t}\right)^2 = \left(\frac{100}{31.7}\right)^2 = 10 \text{ PSI (69 kPa)} \qquad \mathbf{8.29(17)}$$

$$\Delta P_{\text{middle}} = \left(\frac{100}{55.6}\right)^2 = 3.25 \text{ PSI (22.4 kPa)} \qquad \mathbf{8.29(18)}$$

These results indicate that the system drop in one of the extreme positions is more than three times that of the pressure drop in the middle position.

Two Two-Way Valves Sometimes it is desirable to improve the system response speed, but three-way valves cannot be used. In such situations, the installation of two linear, two-way valves is a logical option.

As illustrated in Figure 8.29j, the two valves should have opposite failure positions, so that when one is open the other is closed, and at a 50% signal both are halfway open. The price of a three-way valve and its installation is about 65% that of two two-way valves.

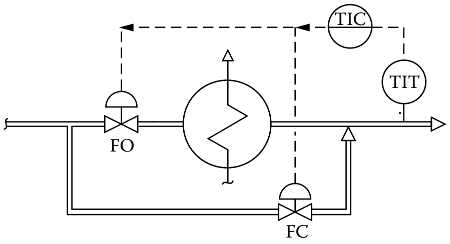


FIG. 8.29j
Exchanger bypass control by using two two-way valves.[1]

TABLE 8.29k
Merits of Two-Way vs. Three-Way Valves in Exchanger Bypass Installations

	One Three-Way Valve	*Two Single-Seated Two-Way Valves*	*Two Double-Ported Two-Way Valves*
Most economical	Yes		
Provides tight shut-off	Yes	Yes	
Applicable to service above 500°F (260°C)		Yes	Yes
Applicable to differential temperature service above 300°F (167°C)		Yes	Yes
Applicable to operation at high pressure and pressure differentials		Yes	Yes
Highest capacity for same valve size			Yes

On the other hand, the capacity of one three-way valve is only about 70% of the capacity of a double-ported two-way valve. This could mean that instead of one 10 in. (250 mm) three-way valve, two 8 in. (200 mm) double-ported two-way valves can be used, if the leakage of such valves (1–5% of full capacity) is acceptable. Therefore, if tight shut-off is required, only the three-way or the single-ported two-way valves can be considered, and their capacity is about the same.

The use of three-way valves is not recommended for high-temperature or high-pressure differential services. In addition, the hollow plug design of three-way valves also contributes to their limitations, because it makes these valves more sensitive to thermal expansion and more difficult to harden than are the solid plugs.

To summarize, bypass control is applied to circumvent the dynamic characteristics of heat exchangers, thus improving their controllability. Bypass control can be achieved by the use of either one three-way valve or two two-way valves. Table 8.29k summarizes the merits and drawbacks of using one or the other option.

Steam Heater Controls

The steam-heated exchanger shown in Figure 8.29l is also nonlinear. Its steady-state gain is the derivative of its outlet temperature with respect to steam flow,[2] having the dimension of °F(lb/h):

$$Kp = \frac{dT_2}{dFs} = \frac{\Delta Hs}{FCp} \qquad \mathbf{8.29(19)}$$

Therefore, its process gain varies inversely with flow. In a step response test, the outlet temperature (T_2) reaches 63.2%

$$Q = UA\Delta \mathrm{Tm} = UA\,\frac{(T_s - T_2) - (T_s - T_1)}{\ln((T_s - T_2)/(T_s - T_1))} \qquad \textbf{8.29(20)}$$

$$Q = F_s\,\Delta H_s = F\,C_P\,(T_2 - T_1)$$

Where Q = Heat-transfer rate
F_s = Steam mass flow
ΔH_s = Latent heat of vaporization
F = Feed rate
Cp = Heat capacity of feed
T_0 = Steam supply temperature
P_1 = Steam supply pressure
P_2 = Steam valve outlet pressure
P_s = Condensing pressure
T_1 = Inlet temperature
T_2 = Outlet temperature
ΔTm = Log mean temperature difference
T_s = Condensing steam temperature

F_s = 500 to 2500 lb/hr (227 kg/hr to 1134 kg/hr)
P_1 = 200 PSIA (1.38 MPa)

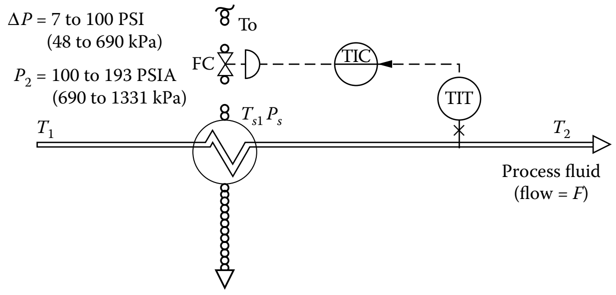


FIG. 8.29l
The feedback control of a steam-heated exchanger and its characteristic equations.[1]

of its final value after a time equivalent of its residence time, which is tube volume divided by flow. Therefore, its time constant, dead time, residence time, period of oscillation, and process gain all vary with flow. As flow (load) drops to 50%, the process gain is doubled.

Therefore, the TIC loops have a tendency to become unstable at low and sluggish at high loads. In order to eliminate cycling, these loops are often tuned at minimum load, resulting in sluggish response at higher loads. One way to compensate for the drop in process gain as the flow increases is to use an equal-percentage valve whose gain increases with load. This is sufficient if the temperature rise ($T_2 - T_1$) is constant. Otherwise, feedforward compensation is needed.

The use of equal-percentage valves is even more important in the case of steam heaters than with liquid-to-liquid exchangers, because the rangeability requirements of steam heaters is usually higher, because of the variations in condensing pressure (steam valve back-pressure) with changes in process load. This can be best visualized by an example.

In Figure 8.29l, both the high and the low load conditions are shown. When the steam flow demand is the greatest, the back-pressure is also the highest (P_2 = 193 PSIA, or 1331 kPa), leaving the lowest driving force (pressure drop) for the control valve (7 PSI, or 48 kPa). Sizing for high flow (2500 lb/h) and low pressure drop results in a large valve.

The back-pressure at low loads is only 100 PSIA (690 kPa), resulting in a pressure drop through the valve that is some 16 times greater (100 PSI, or 690 kPa) than at high loads. The ratio between the required valve coefficients for the high and low load conditions represents the rangeability that the valve has to furnish.

$$\text{rangeability} = S\,\frac{F_{s,\max}}{F_{s,\min}}\sqrt{\frac{[(P_1 - P_2)(P_1 + P_2)]_{\min}}{[(P_1 - P_2)(P_1 + P_2)]_{\max}}}$$

$$= 1.5 \times 5\sqrt{\frac{100 \times 300}{7 \times 393}} = 25.5 \qquad \textbf{8.29(21)}$$

The letter S with the numerical value of 1.5 in Equation 8.29(21) represents the safety factor that is applied in selecting the control valve. A rangeability requirement of this magnitude can create some control problems. One solution is to use a large and a small valve in parallel.

If this is not done, the control quality will suffer for two reasons: First, at low loads the valve will operate near to its clearance flow point (see Section 6.7), where the flow-vs.-lift curve changes abruptly, contributing to unstable or possibly on/off cycling of the valve. Second, for good control the system gain should not vary with changes in load, which an equal-percentage control valve can guarantee only if the control valve ΔP is not a function of load. This being the case, one way to guarantee constant system gain is to install a large and a small valve, both sized to maintain the same gain. The other option is to use an adaptive gain correction in the control algorithm, which is discussed later.

Minimum Condensing Pressure The condensing pressure is a function of load when the temperature is controlled by throttling the steam inlet. As long as the heat transfer area is constant, a reduction in load must result in the lowering of the log mean temperature difference across the exchanger. If T_2 is held constant by the TIC, this can occur only if the steam-side temperature (*Ts*) is lowered. The condensing temperature required can be calculated as follows:[5]

$$T_s = \left(\frac{T_1 + T_2}{2}\right)\left(1 - \frac{L}{100}\right) + \left(\frac{L}{100}\,\mathrm{To}\right) \qquad \textbf{8.29(22)}$$

where L is the load in percentage of maximum. The condensing pressure corresponding to the values of T_s can be plotted as a function of load. Once this pressure drops below the trap back-pressure plus trap differential, it is no longer sufficient to discharge the condensate, and therefore the condensate will start accumulating in the exchanger. As condensate accumulation progresses, more and more of the heat-transfer area will be covered up, resulting in a corresponding increase in condensing pressure.

When this pressure rises sufficiently to discharge the trap, the condensate is suddenly blown out and the effective heat-transfer surface of the exchanger increases instantaneously. This can result in cycling as the exchanger surface is covered and uncovered. In addition, noise and hammering can be

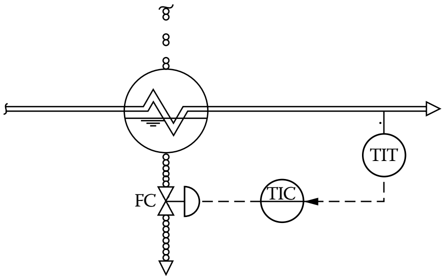


FIG. 8.29m
The problems associated with minimum condensing pressure and condensate removal can be eliminated by locating the control valve in the condensate line.[1]

caused as the steam bubbles collapse on contact with the accumulated cooler condensate. The methods to remedy this are several and are discussed in the following paragraphs.

Condensate Throttling Mounting the control valve in the condensate line, as shown in Figure 8.29m, is sometimes proposed as a solution to minimum condensing pressure problems. There also is a cost advantage to this configuration, because a small condensate valve costs less than a larger one for steam service.

On the surface this appears to be a convenient solution, because the throttling of the valve causes variations only in the condensate level inside the partially flooded heater and has no effect on the steam pressure, which stays constant. Therefore, condensate removal does not present a problem. Unfortunately, the control characteristics of this loop are not so favorable.

The response time of such a control configuration depends on both the load and exchanger geometry. Tests have shown that in such a system, correcting a load decrease takes about three times as long as correcting a load increase.[5] The response time to a 5% change in load is from a few seconds to a few minutes, whereas to a 50% change it can be from a few minutes to nearly an hour.[5]

When the load is decreasing, the valve is likely to close completely before the condensate builds up to a high enough level to match the new lower load with a reduced heat-transfer area. In this direction, the process is slow, because steam has to condense before the level can be affected. When the load increases, the process is fast, because just a small change in control valve opening is sufficient to discharge enough condensate to expose the heat-transfer surface area that is required to match the increased load.

With such "nonsymmetrical" process dynamics, control is bound to be poor. If the controller is tuned for the fast response speeds (corresponding to the increasing load direction), sluggish performance and overshoot can result when load is decreasing. If it is tuned for the slow part of the cycle, cycling can occur when the load rises. Therefore, the replacement of the steam trap with an equal-percentage control valve is usually not a good solution.

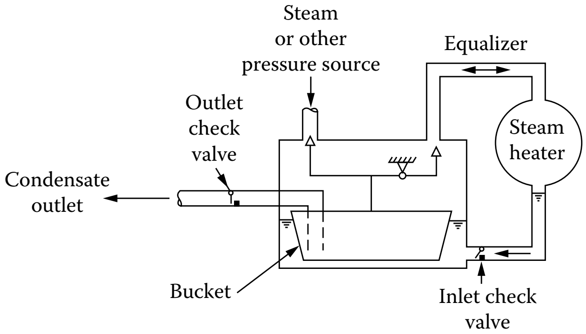


FIG. 8.29n
A lifting or pumping trap, shown here in its filling position, can help prevent condensate accumulation in heaters operating at low condensing pressures.[1]

Pumping Traps It is possible to use lifting traps to prevent condensate accumulation in heaters operating at low condensing pressures. This device is illustrated in Figure 8.29n and depends on an external pressure source for its energy.

The unit is shown in its filling position, in which the liquid head in the heater has opened the inlet check valve of the trap. Filling progresses until the condensate overflows into the bucket, which then sinks, closing the equalizer and opening the pressure source valve. As pressure builds in the trap, the inlet check valve is closed and the outlet valve opens when the pressure exceeds that of the condensate header.

The discharge cycle follows. In this cycle, the bucket is emptied. When the bucket is near empty, the buoyant force raises the bucket, which then closes off the steam valve and opens the equalizer. Once the pressure in the trap is lowered, the condensate outlet check valve is closed by the header back-pressure, and the inlet check is opened by the liquid head in the heater, which then is the beginning of another fill cycle.

The aforementioned pumping trap guarantees condensate removal regardless of the minimum condensing pressure in the heater. If such a trap is placed on the exchanger illustrated in Figure 8.29l, it will make temperature control possible even when the heater is under vacuum. Of course, this does not relieve the rangeability problems discussed earlier, and the use of two valves in parallel might still be necessary.

Level Controllers The control system in Figure 8.29l provided quick response but was unable to discharge its condensate at low loads. The system in Figure 8.29m eliminated the condensate problem, but at the price of worsening control response.

Because the low condensing pressure situation is a result of the combination of low load and high heat-transfer surface area, it is possible to prevent it from developing by reducing the heat-transfer area. One method of achieving this is shown in Figure 8.29o. Here, the steam trap has been replaced by a level control loop. With this instrumentation, it is possible to adjust the size of the heater by changing the level set point

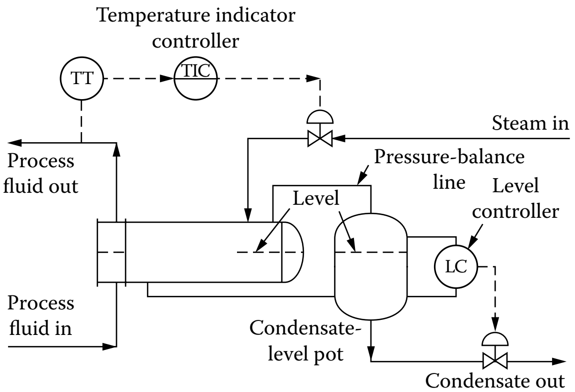


FIG. 8.29o
The use of two control loops can maintain sensitive temperature control without condensate removal problems.[5]

to match the process load. This technique gives good temperature control, if the level setting is correct and if there are no sudden load variations in the system.

As the required level is a function of the load, it can be adjusted automatically. This relieves the operator of the responsibility of manually changing the level controller set point whenever the load changes. More important, this also eliminates the potential for human error, which at low-level settings can result in condensate removal problems and at high levels can prevent temperature control.

Automatic adjustment of the level is illustrated in Figure 8.29p. Here the load is detected by the valve position controller (VPC) having a set point of, say, 50%. When the load rises, the steam valve opens beyond 50% and the VPC increases the active heat-transfer area in the heat exchanger by lowering the set point of the condensate level controller.

The VPC is an integral-only controller and, therefore, will not respond to measurement noise or valve cycling. The integral time is set to give slow floating action that is fast enough to respond to anticipated load changes. The external feedback protects the VPC from reset windup when the LIC is switched to local set point.

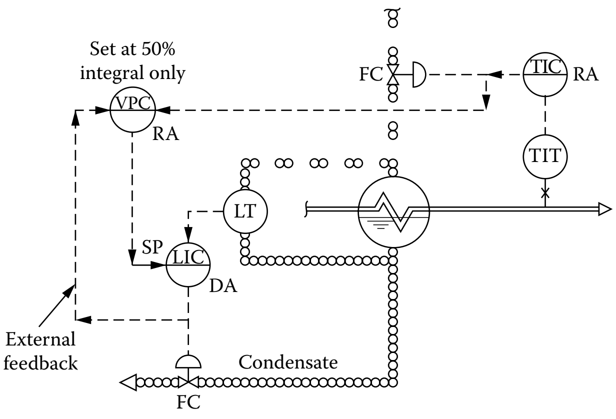


FIG. 8.29p
The level setpoint can be floated to match the load.[1]

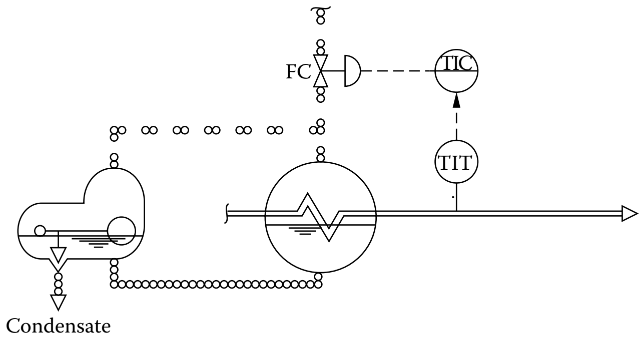


FIG. 8.29q
Continuous drainer traps can serve the same function as level controllers, but can handle only small variations in condensate level.[1]

In less critical applications, in which the cost of an extra level control loop cannot be justified, a continuous drainer trap such as the one shown in Figure 8.29q can serve the same purpose as the level control scheme just described. Its cost is substantially lower, but it is limited to the range within which its level setting can be varied. Another limitation is that its control point is offset by load variations. For this reason, continuous drainer traps are unlikely to be considered for installations on vertical heaters or on reboilers, where the range of level adjustment can be substantial.

Bypass Control Table 8.29k summarizes some of the features of three-way valve installations serving to circumvent the transient characteristics of coolers. Figure 8.29r shows the same concept applied to a steam heater. The advantages and limitations of this system are the same as discussed in connection with liquid-to-liquid exchangers, with the additional advantage that the bypass used has also created an additional degree of freedom.

Therefore, steam can now be throttled as a function of some other property. The logical decision is to adjust the

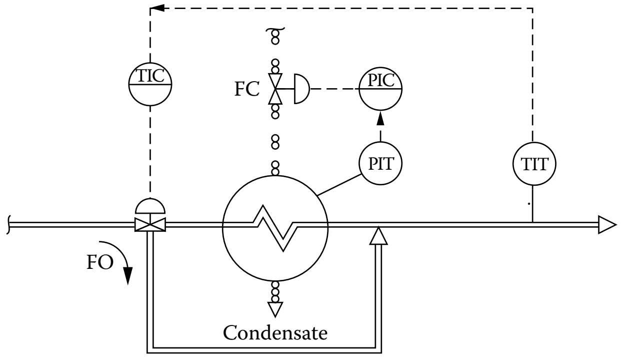


FIG. 8.29r
Bypass control of steam heater provides the added degree of freedom needed for independent control of condensing pressure.[1]

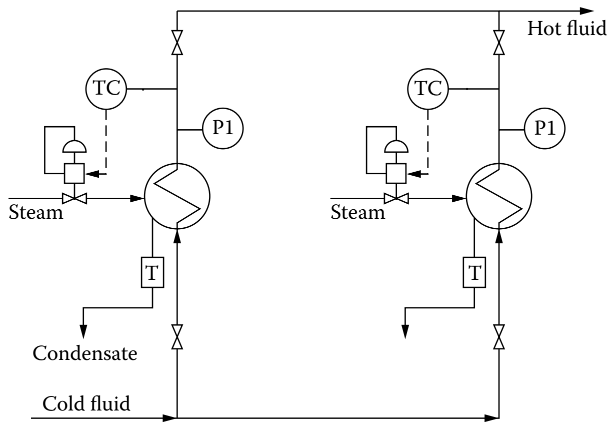


FIG. 8.29s
Parallel heaters operating close to the boiling point of the process fluid can experience serious interaction problems.[4]

steam feed so that it maintains the condensing pressure constant. This then eliminates problems associated with condensate removal. It is also important to realize that when all the process fluid flow is sent through the bypass, the stagnant exchanger contents will be exposed to steam heat, and therefore, unless protection is provided, it is possible to boil this liquid.

Interaction between Parallel Heaters If the process fluid is heated to a temperature approaching its boiling point, serious interaction can occur between parallel heaters (Figure 8.29s). The mechanism of developing this oscillation is as follows: A sudden drop in flow causes overheating and vaporization in one of the heaters. The vapor formation increases the back-pressure and further reduces the flow, eventually forcing all flow through the other "cold" exchanger, while the "hot" exchanger discharges slugs of liquid and vapor.

After a period of noise and vibration, when the "hot" exchanger has discharged all of its liquid, the back-pressure drops and flow is resumed, drawing feed from the "cold" exchanger. This causes the "cold" exchanger to overheat, and the cycle is repeated with the roles of the exchangers reversed.[4]

This type of interaction can be eliminated by distribution controls (Section 8.28). The control system used in making sure that the load is equally distributed between exchangers is the same as the system described for cooling tower balancing (Figure 8.17e).

Condenser Controls

Depending on whether the control of condensate temperature or of condensing pressure is of interest, the system shown in Figures 8.29t or 8.29u can be considered. Both of these systems operate by throttling the cooling water flow through the condenser, causing a potential for high-temperature rise that is acceptable only when the water is chemically treated against fouling.

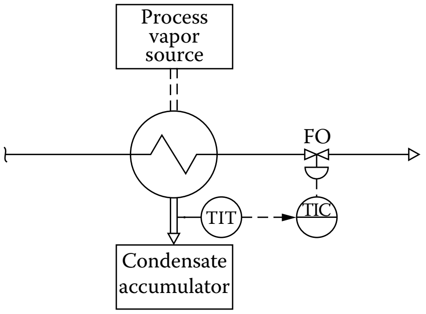


FIG. 8.29t
Condensate temperature control throttles the cooling water flow through the condenser.[1]

For good, sensitive control, the water velocity through the condenser should be such that its residence time does not exceed 1 min. Another rule of thumb is to keep the water velocity above 4.5 ft/s (1.35 m/s). This can be accomplished by the use of a small recirculating pump, such as P1 in Figure 8.13i. In this system, FSHL-3 detects the net forward velocity of the water. As long as it is under 4.5 ft/s, FSHL-3 will keep the recirculating pump running.

When it is not desirable to throttle the cooling water, the system illustrated in Figure 8.29v can be considered. Here, the exposed condenser surface is varied to control the rate of condensation. If noncondensables are present, a constant purge may be used to remove the inerts. The disadvantages of the nonsymmetrical response of such condensate level controls have already been described in connection with Figure 8.29m, and they are the main disadvantage of this design.

Because the heat-transfer efficiency is the highest when both the coolant flow and the heat-transfer area are at their maximums, the goal of optimization in Figure 8.29u is to open the coolant valve fully. In Figure 8.29v, the goal of optimization is to eliminate flooding. Both of these goals can be achieved by slowly lowering the PIC set point until the cooling capacity is fully utilized and none of it is wasted by throttling.

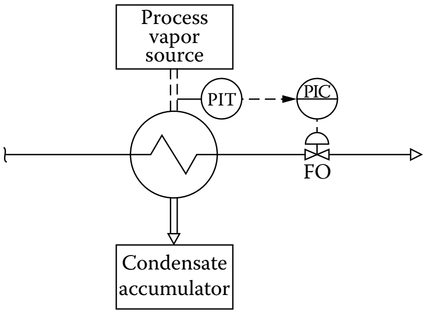


FIG. 8.29u
The cooling water flow through the condenser can be throttled to control pressure.[1]

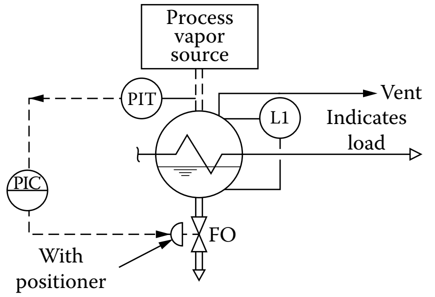


FIG. 8.29v
Condenser control can be achieved by changing the wetted surface area.[1]

The preferred method of condenser control is to vary the heat-transfer area through partial flooding.[2] In that case, the load is indicated by the condensate level (Figure 8.29v). The condensate is subcooled to differing degrees as a function of residence time. Therefore, condensate temperature cannot be used for control.

The controller on a partially flooded condenser can be tuned as a level controller, but because of the noise filtering effect of the system's heat capacity, derivative action can usually be added.[2] The addition of valve positioners is always helpful in improving the response and stability of the loop.

Distillation Condensers A variety of condenser designs is available for use on distillation towers. The disadvantage of air-cooled designs is that they are sensitive to ambient variations, particularly to rainfall. These units are throttled by manipulating their inlet louvers or by adjusting the blade pitch of their fans. If multispeed or variable-speed fans are available, the condenser rangeability can be increased while energy is saved.

When the condenser is water-cooled and its inlet pressure is controlled by partial flooding (Figure 8.29v), the main limitation is the response speed of the system, because a change in valve position does not immediately affect the heat-transfer area. The response speed is slower when the level needs to be increased and faster when it is to be decreased. The consequences of this "nonsymmetricity" have already been explained in connection with Figure 8.29m. In controlling distillation towers, the elimination of this nonsymmetricity is one of the important control goals.

The simplest method of handling this problem is to bleed off a fixed amount of gases and vapors to a lower-pressure unit, such as to an absorption tower, if such is available in the system. If an absorber is not present, it is possible to install a vent condenser to recover the condensable vapors from this purge stream.

It is recommended that the fixed continuous purge be used wherever economically possible; however, when this is not permitted, it is possible to modulate the purge stream. This might be desirable when the amount of inerts is subject to wide variations over time (Figure 8.19ff in Section 8.19).

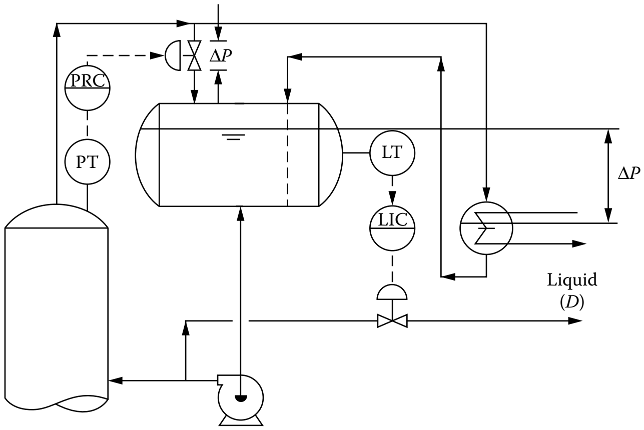


FIG. 8.29w
Condenser control by the throttling of hot gas bypass.

As the uncondensables build up, the pressure controller will tend to open the control valve (PCV-1) to maintain the proper rate of condensation. As shown in Figure 8.19ff, the signal that is opening this control valve can also be used to signal when the pressure control valve (PCV-1) is nearly fully open, and therefore purging is needed. In Figure 8.19ff, this is done by a valve position controller (VPC-2).

Condenser below Receiver Lowering the condenser below the accumulator not only reduces installation cost and makes maintenance more convenient but also eliminates the nonsymmetricity of the process. It is the usual practice to elevate the bottom of the accumulator 10 to 15 ft (3 to 4.5 m) above the suction of the pump in order to provide the required suction head.

In this type of installation, the control valve is placed in a bypass from the vapor line to the accumulator (see Figure 8.29w). When this valve is open, it equalizes the pressure between the vapor line and the receiver. This causes the condensing surface to become flooded with condensate because of the 10 to 15 ft of head that exists in the condensate line from the condenser to the receiver. The flooding of the condensing surface causes the pressure to build up because of the decrease in the active heat-transfer surface available.

Under normal operating conditions, the subcooling that the condensate receives in the condenser is sufficient to reduce the vapor pressure in the receiver. The difference in pressure permits the condensate to flow up the 10 to 15 ft of pipe between the condenser and the accumulator.

When the condensing pressure is to be reduced, the valve closes, resulting in an increase in the exposed condenser surface area. In order to expose more area, the condensate is transferred into the accumulator; this transfer can occur only if the accumulator vapor pressure has been sufficiently lowered by condensation. Therefore, the system speed in this direction is a function of the amount of supercooling of the condensate. Increased supercooling increases system response speed.

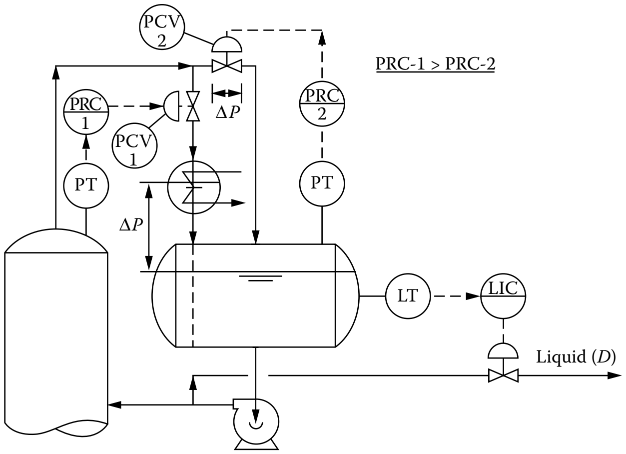


FIG. 8.29x
High-speed condenser control.

Based on these considerations, the unit might or might not be symmetrical in its dynamics. If high speed is desired and there are no inerts in the process vapors, the controls shown in Figure 8.29x can be considered. Here, the pressure is controlled in the accumulator by throttling the condenser bypass flow. The column pressure is maintained by throttling the flow of vapor through the condenser. Controlling the rate of flow through the condenser gives faster pressure regulation for the column.

When the condensing temperature of the process fluid is low, water is no longer an acceptable cooling medium and cooling by refrigerants is required. One standard technique of controlling a refrigerated condenser is illustrated in Figure 8.29y. Here, the heat-transfer area is set by the level control loop, and the operating temperature is maintained by the pressure controller.

When process load changes, it affects the rate of refrigerant vaporization, which is compensated for by level-controlled makeup. Usually the pressure and level settings are made manually, although there is no reason why these set points could not be automatically adjusted as a function of load. This can be done by placing a valve position controller on the PIC output signal. This VPC will detect an increase in load and will respond to it by lowering the LIC set point to increase the heat-transfer area, as was discussed in connection with Figure 8.29p.

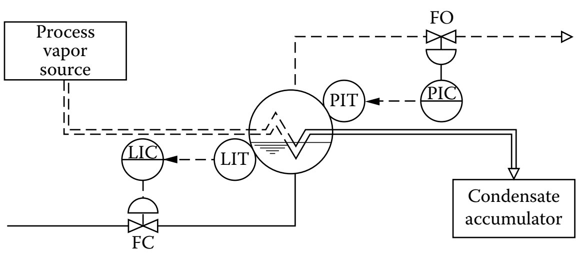


FIG. 8.29y
Condensers can be controlled using refrigerant coolants.[1]

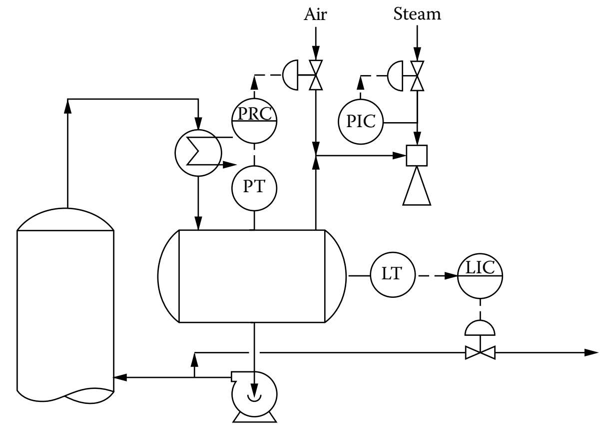


FIG. 8.29z
Typical control system for the operation of vacuum condensers.[1]

Vacuum Systems For some liquid mixtures, the temperature required to vaporize the feed would need to be so high that decomposition would result. To avoid this, it is necessary to operate at below atmospheric pressure. The common means for creating a vacuum is to use steam jet ejectors. These can be used singly or in stages to create a wide range of vacuum conditions. Their wide acceptance is due to their having no moving parts and requiring very little maintenance.

Most ejectors are fixed-capacity devices and work best at a single steam condition. Increasing the steam pressure above the design point will not usually increase the capacity of the ejector; as a matter of fact, it will sometimes decrease the capacity because of the choking effect of the excess steam in the diffuser throat.

Steam pressure below a critical value for a jet will cause the ejector operation to become unstable. Therefore, it is recommended that a pressure controller be installed on the steam inlet to maintain the optimum pressure required by the ejector.

The recommended control system for the operation of vacuum condensers is shown in Figure 8.29z. In this configuration, air or gas is bled into the vacuum line just ahead of the ejector. This makes the maximum capacity of the ejector available to handle any surges or upsets, but at low loads, this represents a substantial waste of steam. Therefore, for processes in which load variations are expected, the operating costs can be lowered by installing a larger and a smaller ejector. This makes it possible to switch to the small unit automatically when the load drops off, and thereby reduce steam demand.

A pressure controller (PRC) regulates the amount of bleed air used to maintain the pressure on the accumulator. Controlling the air bleed by the accumulator instead of the condenser inlet pressure reduces the time lag and, thereby, increases the sensitivity of the loop.

Reboilers and Vaporizers

As noted at the beginning of this chapter, when a steam-heated reboiler is used, only one degree of freedom is available

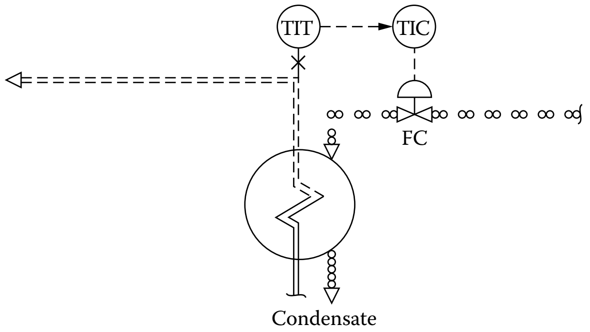


FIG. 8.29aa
Reboilers can be controlled by generating vapors at controlled temperatures.[1]

(Equation 8.29[4]); therefore, only one controller can be installed without overdefining the system. This one controller usually throttles the rate of steam addition. Minimum condensing pressure considerations are the same as have been discussed earlier in connection with liquid heaters.

Figures 8.29aa and 8.29bb show the two basic alternatives for controlling the reboiler. One is to generate the process vapors at a controlled superheat temperature, the other is to generate saturated process vapors at a constant rate by keeping the rate of heat input constant.

Naturally, there are other, more sophisticated reboiler control strategies, using composition, temperature difference, or derived variables as the means of control (see Section 8.21), but in their effects they are all similar to the systems shown in Figures 8.29aa and 8.29bb.

When several reboilers operate in parallel, it is important to balance the load distribution between them according to the type of control system, as was described in Figure 8.28b.

Fired Reboilers When heat duties are great and bottom temperatures are high, fired reboilers are used. The fired reboiler can vaporize the bottoms of a distillation tower as this liquid circulates by natural convection through the heater tubes. The coils are generously sized to ensure adequate circulation of the bottoms liquid. Overheating the process fluid is a contingency that must be guarded against, because most tower bottoms will coke or polymerize if under excessive temperatures for some length of time.

A common control scheme is shown in Figure 8.27q in Section 8.27. This figure depicts the tower bottom along with the fired reboiler. It is not usually practical to measure the flow of tower bottoms to the reboiler — first, because the liquid is near equilibrium (near the flash point), and second, because it is usually of a fouling nature, tending to plug most flow elements. Proper circulation of the fluid is provided for in the careful hydraulic design of the interconnecting piping.

The other important variable is the reboiler return temperature, which in Figure 8.27q is controlled by TRC-1's throttling of the fuel gas control valve (TV-1). The high-temperature alarm (TAH-1) is provided to warn the operator that the process fluid has suddenly reached an excessive temperature, indicating that manual adjustments are needed to cut back on the firing.

The furnace draft can be manually set by the operator, adjusting the stack damper while observing the draft gauge (PI-1), or it can be maintained automatically. The stack temperature (TI-2) and the firebox temperature (TI-3) are detected and can be alarmed to signal excessive temperatures that may develop during periods of heavy firing.

The major dangers in this type of furnace operation include the interruption of process fluid flow or the stoppage of fuel flow. The loss of process fluid can occur if the liquid level in the tower bottom is lost. If this happens, flow to the reboiler tubes will stop and a dangerous overheating of these tubes may result. To protect against this, the low-level switch (LSL-1) is wired to trip the solenoid valve (LY-1) that vents the diaphragm of the control valve (TV-1) to close the fuel valve.

A momentary loss of fuel can also be dangerous, because that can extinguish the flames, and on the resumption of fuel flow a dangerous air/fuel mixture can develop in the firebox. To prevent this occurrence, the low-pressure switch (PSL-1) also trips the solenoid (LY-1) if the fuel pressure is low, and that closes the fuel valve (TV-1). The solenoid valve (LY-1) should be of the manual reset type and should keep the valve (TV-1) closed even after the fuel pressure has been reestablished.

On loss of air supply, the fuel control valve (TV-1) should fail closed (FC), so that firing will be discontinued during such an emergency.

Fired Heaters and Vaporizers The unit feed heater of a crude oil refinery is representative of the class of furnaces that includes fired heaters and vaporizers. As was discussed in Section 8.27, crude oil, prior to distillation into the various petroleum fractions (gasoline, naphtha, gas oil, heavy fuel oil, and so forth) in the "crude tower" must be heated and partially vaporized. The heating and vaporization are done in the crude heater furnace, which consists of a firebox with preheating coils and vaporizing coils.

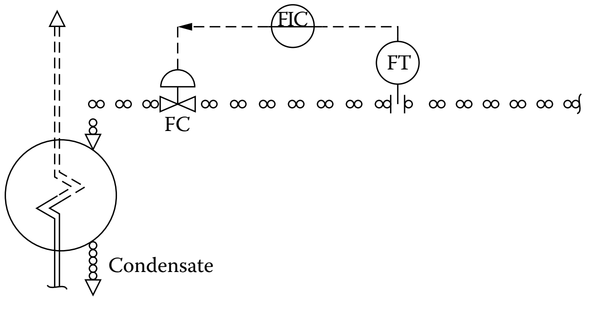


FIG. 8.29bb
BTU control sets the steam flow to reboilers.[1]

The heating is usually done in coils in the convection section of the furnace. The convection section is the portion of the coils that does not see the flame but that is exposed to the hot flue gases on their way to the stack. The vaporizing takes place at the end of each pass in the radiant section of the furnace (where the coils are exposed to the flame and the luminous walls of the firebox).

The partially vaporized effluent then enters the crude tower, where it flashes and is distilled into the desired "cuts." Other process heaters that fall into this category are refinery vacuum tower preheaters, reformer heaters, hydrocracker heaters, and dewaxing unit furnaces.

The prime control tasks in this process are

1. Flow control of feed to the unit
2. Proper splitting of flow into the parallel paths through the furnace to prevent overheating of any one stream, which could result in coking
3. Supplying the correct amount of heat to the crude tower

Figure 8.29cc shows the typical process controls for this type of furnace. The crude feed rate to the furnace is set by the flow controller (FRC-2). This flow can be split through the parallel paths of the furnace by remote manual adjustment of the control valves (HV-1 through HV-4) via the manual stations (HIC-1 through HIC-4) or can be distributed automatically.

In this manual control configuration, the operator manually adjusts (HIC-1 to HIC-4) the distribution valves to equalize the flows (FI-1 through FI-4). Naturally, this can also be done automatically by replacing the four flow indicators with four flow ratio controllers (FFIC-1 to FFIC-4) and sending the total flow transmitter signal (FT-2) as the set point to each.

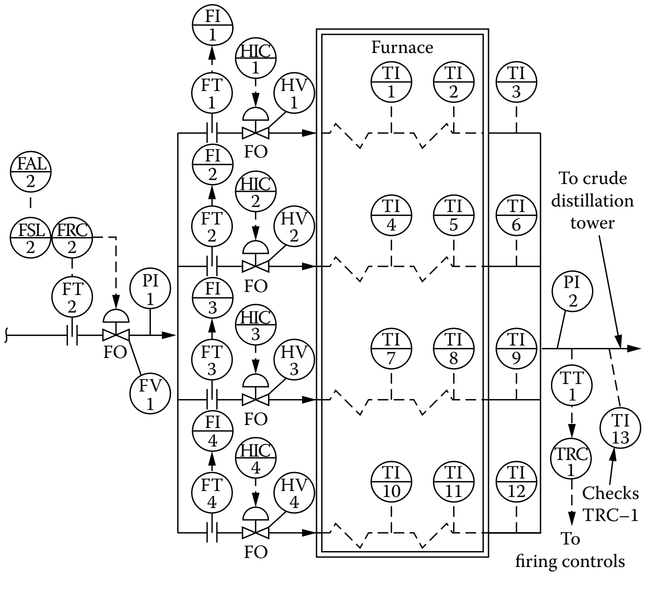


FIG. 8.29cc
Semiannual flow distribution controls of a fired vaporizer.

The temperature indicators TI-1 through TI-12 can be periodically observed to determine if there are any rising or falling trends in any one of the passes. If such a trend develops, the flow through that pass is altered slightly to drive the temperature back toward the norm. If a higher level of automation is desired, pass temperatures can act as cascade masters that will automatically adjust the set points of the flow ratio controllers (FFIC-1 to FFIC-4).

The desired heat output into the feed stream is more difficult to control, because the effluent of the furnace is partially vaporized and the feedstock varies in composition. If the feed were only heated, with no vaporization taking place, the control would require only that an effluent temperature be maintained. If complete vaporization and superheating occurred, this too could be handled well by straight temperature control.

But in the case of partial vaporization with a variable-feed composition, effluent temperature control alone (TRC-1 in Figure 8.29cc) is not a reliable approach. The composition and, hence, the boiling-point curve of the feed is not constant, and therefore required control temperature itself varies. Additional information is thus required and can be obtained from the distillation downstream of the furnace. By observing the product distribution from the fractionation, the need for a change in heat input can be determined.

Because this approach is slow and not precise, more accurate controls have been developed, using sophisticated instrumentation and special analyzers to measure feed composition and computers to optimize the mathematical model and thereby determine the required head input to the feed. Some of these control strategies are described in Section 8.27.

In older plants, the practice is to achieve approximate control with a temperature controller (TRC-1 in Figure 8.29cc). In that case the unit operator periodically changes the TIC set point to account for feed composition variations. The operator depends on experience and on the performance of the fractionator to determine the proper temperature setting. Naturally, in more advanced refineries (Section 8.21), all this is under computer control.

ADVANCED CONTROLS

In the paragraphs that follow, the more advanced heat-transfer controls will be described. These include the more traditional techniques of override, cascade, and feedforward controls and the more recent strategies of adaptive gain, model-based, and multipurpose strategies.

Override Controls

Override controls in general are discussed in Section 2.28 in Chapter 2. The purpose of override controls usually is to protect the operating equipment against unsafe or maintenance-intensive operation. An example of such a case is shown in Figure 8.29dd.

This control system should be considered when river water is being used for cooling and therefore the fouling of

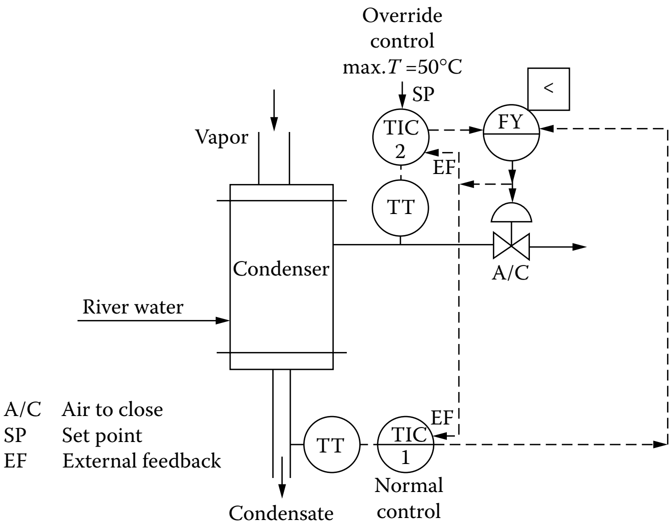


FIG. 8.29dd
When the river water outlet temperature rises to 50°C, the override controller (TIC-2) overrides the normal controller (TIC-1) and prevents the outlet temperature from rising beyond 50°C.

the heat-transfer surfaces tends to speed up when the water outlet temperature exceeds 50°C (122°F). In this configuration, normally the river water flow is throttled by TIC-1 to control the condensate temperature. However, if the cooling water temperature reaches 50°C (122°F), the high temperature override (TIC-2) takes over by increasing the cooling water flow and prevents this temperature from further rising. Naturally, during such override periods, the control of the condensate temperature will be abandoned, and having been left uncontrolled, it is likely to drop.

Similar override configurations can be utilized to prevent steam or process pressures from reaching undesirable limits. It is also possible to apply override controls in an envelope configuration, so that any of a number of limits (temperature, level, valve opening, and so on) can override the normal operation when their limits are reached.

Cascade Control

For a detailed discussion of all aspects of cascade control, refer to Section 2.6.

By definition, cascade loops consist of two controllers in series. In heat exchanger applications, the master detects the process temperature and the slave detects a variable that may upset the process temperature. A cascade loop controls a single temperature, and the cascade master adjusts the set point of the slave controller to assist in achieving this.

In order for a cascade loop to be successful, the slave must be much faster than the master. A rule of thumb is that the time constant of the primary controller should be ten times longer than that of the secondary, or that the period of oscillation of the primary should be three times that of the secondary.

Cascade loops are invariably installed to prevent outside disturbances from entering and upsetting the primary controlled variable. An example of such a disturbance is the steam pressure variation of a steam heater. The conventional single-controller system (see Figure 8.29l) cannot respond to a change in steam pressure until it has first upset the process temperature. In other words, an error in the detected temperature has to develop before corrective action can be taken.

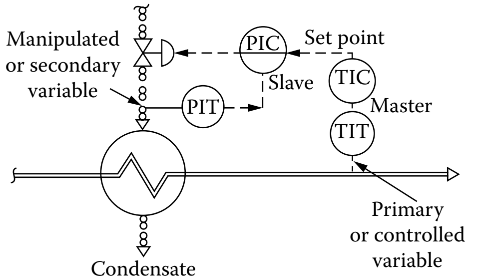


FIG. 8.29ee
A temperature-pressure cascade loop can eliminate the upsets caused by supply pressure variations in a steam heater.

The cascade loop, in contrast, responds immediately by anticipating an upset that would otherwise develop and corrects for the effect of pressure change before it can influence the process temperature (Figure 8.29ee).

The improvement in control quality is a function of relative speeds and time lags of the two loops. A slow primary (master) variable and a quickly responding secondary (slave) variable is a desirable combination for this type of control. If the slave can quickly correct the fast disturbances, they will not be allowed to enter the process and, therefore, will not upset the primary (master) variable.

One of the quickest and, therefore, best cascade slaves is the simple and inexpensive pressure regulator. Its air-loaded variety (Figure 8.29ff) is extremely fast and can correct for steam supply pressure or load variations almost instantaneously.

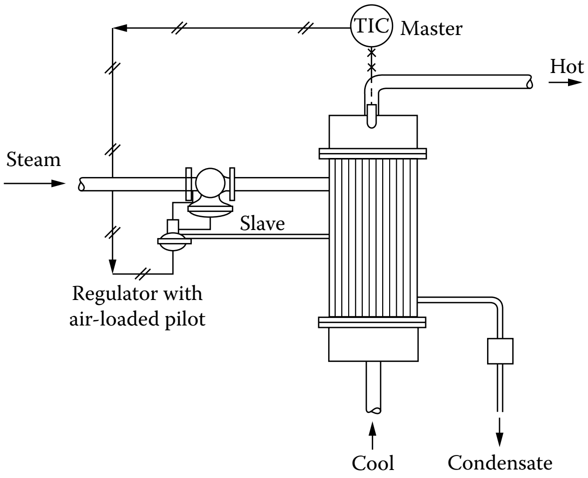


FIG. 8.29ff
One of the best cascade slaves is the pressure regulator.[6]

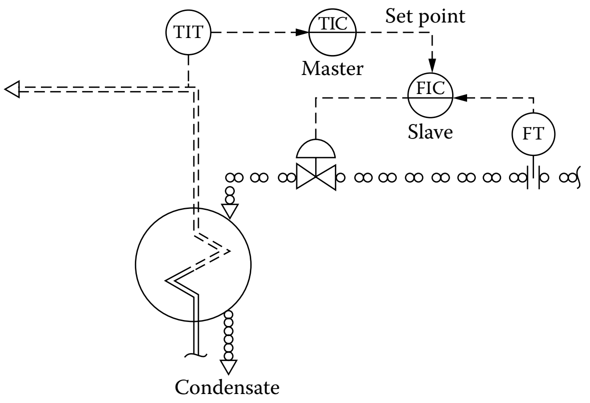


FIG. 8.29gg
Temperature-flow cascade loops are commonly used on steam reboilers.[1]

In Figure 8.29gg, the master controlled variable is temperature and the manipulated slave variable is the pressure or flow of steam. The primary variable (temperature) is slow, and the secondary (manipulated) variable is fast and capable of responding quickly to disturbances. Therefore, if disturbances occur (a sudden change in plant steam demand, for example) that are upsetting the manipulated variable (steam flow or pressure), these disturbances will be sensed immediately and corrective action will be taken by the secondary controller so that the primary variable (process temperature) will not be affected.

Feedforward Control

Feedback control involves the detection of the controlled variable (temperature) and the counteracting of its deviation from the set point by the adjustment of a manipulated variable (the flow of a heat transfer fluid). This mode of control necessitates that an upset occur in the controlled variable itself before correction can take place. Hence, the term "feedback" can imply a correction "back" in terms of time — a correction that occurs after an upset and should have taken place earlier, when the disturbance occurred.

Feedforward is a mode of control that responds to a disturbance instantaneously and compensates for the error that the disturbance would have otherwise caused later in the controlled variable.

Figure 8.29hh illustrates a steam heater under feedforward control. This control system consists of two main segments. The feedforward portion of the loop detects the major load variables (the flow and temperature of the entering process fluid) and calculates the required steam flow (W_s) as a function of these variables.

When the process flow increases, it should be matched with an equal increase in steam flow instantaneously. Because instantaneous response is not possible, the next best thing is to add more steam than needed as soon as possible. This goal of dynamic correction is served by the lead-lag element in the feedforward loop. The resulting improvement in control performance is illustrated in Figure 8.29ii.

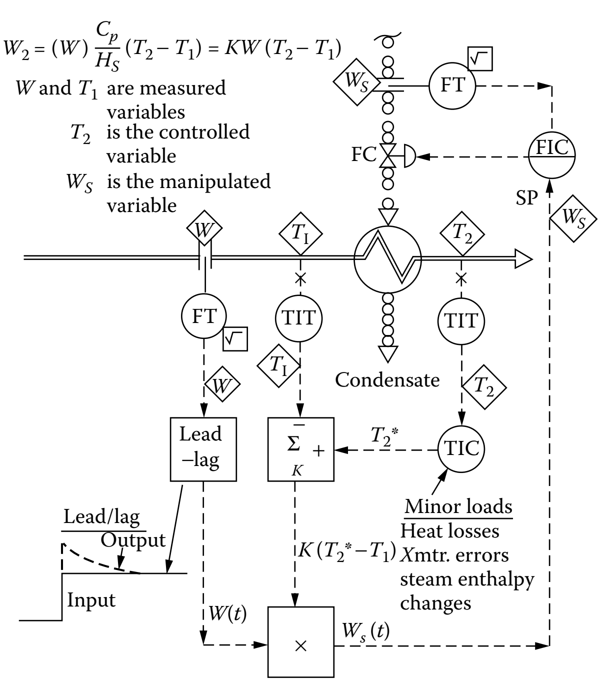


FIG. 8.29hh
In feedforward optimization of steam heaters, major load variations (T_1 and W) are corrected by the feedforward portion of the loop, leaving only the minor load variables for feedback correction.[1]

The feedback portion of the loop (TIC in Figure 8.29hh) has to do much less work in this configuration, as it only has to correct for minor load variables, such as heat losses to the atmosphere, steam enthalpy variations, and sensor errors. The feedback and feedforward portions of the loop complement each other. The feedforward portion is responsive, fast, and sophisticated, but inaccurate. The feedback portion is slower, but is capable of correcting the upsets caused by unknown or poorly understood load variations, and it is accurate.

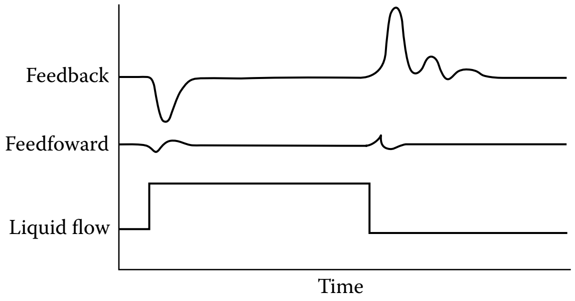


FIG. 8.29ii
Feedforward control is capable of reducing both the area and the duration of the load response transients.[2]

The design variations of feedforward loops, their dynamic compensation, and tuning are covered in more detail in Section 2.9 in Chapter 2.

Adaptive Gain As was already noted in connection with Figure 8.29l, the heat-transfer process is a variable gain process. The steady-state gain of a steam heater process is

$$dT_2/dW_s = \Delta Hs/WCp \qquad \textbf{8.29(23)}$$

Therefore, the process gain varies inversely with the process flow (W). If the temperature rise ($T_2 - T_1$) is also variable, even the use of an equal-percentage valve cannot correct for this nonlinearity in the process. In that case, the only way to keep the process gain constant is to use the feedforward system shown in Figure 8.29hh.

In the feedforward control system shown, a reduction in process flow causes a reduction in the gain of the multiplier, which cancels the increase in process gain because as the process gain varies inversely with flow, it causes the controller gain to vary directly with flow. Thus, the feedforward loop provides gain adaptation as a side benefit. The result is constant total loop gain and, therefore, stable loop behavior.

The feedforward concept described in Figure 8.29hh is not limited to steam heaters but can be used on all types of heat exchangers. The only modifications needed are the ones that are required to correctly reflect the heat balance equation corresponding to the particular heat-transfer unit.

The advantages of feedforward control are similar to those of cascade control, because in feedforward control also the load upsets or supply disturbances are corrected for before their effects are felt by the controlled variable. Feedforward control therefore contributes to stable dampened response to load changes and to fast recovery from upsets.

It is also possible to adapt the controller gain to vary directly with steam flow without using feedforward control. Equations 8.29(24) and 8.29(25) describe such a flow-adapted three-mode controller.

The equation for the flow-adapted three-mode controller is

$$m = K_c w\left(e + \frac{w}{T_i}\int e\,dt + \frac{T_d}{w}\frac{d}{dt}e\right) + b \qquad \textbf{8.29(24)}$$

where w is the fraction of full-scale flow, and K_c, T_i, and T_d are the proportional, integral, and derivative settings at full-scale flow. Equation 8.29(24) can be rewritten to reduce the adaptive terms to two:

$$m = K_c\left(we + \frac{w^2}{T_i}\int e\,dt + T_d\frac{d}{dt}e\right) + b \qquad \textbf{8.29(25)}$$

Model-Based Controls

All expert systems take advantage of the large memory and fast data manipulation capability of computers. There are hundreds of expert systems on the market today, and we should understand what they can and what they cannot do. Their common feature is that they all serve some form of optimization, while their distinguishing features are both in the type of optimization they perform and in their method used in performing it. From the perspective of their methods, one can distinguish model-based and model-free methods.

Model-based control (MBC), model predictive control (MPC), and internal model control (IMC) are all suited for the optimization of such unit processes that are well understood, such as heat transfer. Their performance is superior to that of the model-free systems, because they are capable of anticipation and, thereby, can respond to new situations. In this sense, their performance is similar to that of feedforward control systems, while the model-free systems behave in a feedback manner only.

For details on the nature and operation of the various model-based control systems refer to Sections 2.13 to 2.18 in Chapter 2.

Multipurpose Systems

To this point, the control of the isolated heat-transfer unit has been discussed. In the majority of critical installations, the purpose of such systems is not limited to the addition or removal of heat, but involves making use of both heating and cooling in order to maintain the process temperature constant. Such a task necessitates the application of multipurpose systems, incorporating many of the features that have been discussed individually earlier.

Figure 8.29jj, for example, depicts a design that uses hot oil as its heat source and water as the means of cooling, arranged in a recirculating system. The points made earlier in connection with three-way valves, cascade systems, and so forth also apply here, but a few additional considerations are worth noting.

Probably the most important single feature of this design is that it operates on a *split-range signal.* This means that when the process temperature is above the desired set point,

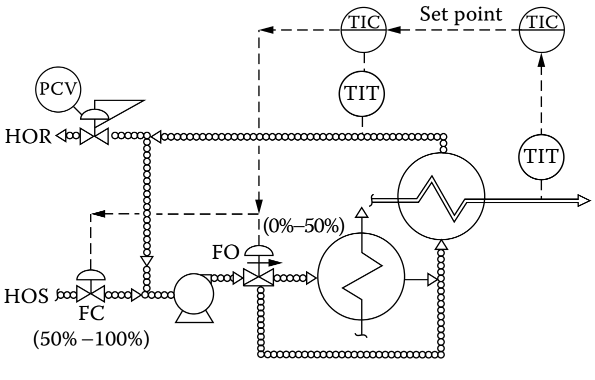


FIG. 8.29jj
This multipurpose temperature control system is capable of both adding and removing heat from the process fluid.[1]

the output signal to the valves will be reduced. When the value of this signal is between 50 and 100%, the three-way valve is fully open to the exchanger bypass, and the two-way hot oil supply valve is partially open.

If the reduction in the two-way valve opening is not sufficient to bring the process temperature down to set point, the signal will further decrease, thereby fully closing the two-way valve at a signal value of 50% and will begin to open the three-way valve to the cooler. At a 0% signal, the total cooling capacity of the system is applied to the recirculating oil stream, which at that time flows through the cooler without bypass.

The limitations of such a split-range operation include the following:

First, at a signal level near 50%, the system can be unstable and cycling. This is because at this point the three-way valve is just beginning to open to the cooler, and the system might receive alternating slugs of cooling and heating because of the limited rangeability of the valves. Such unstable operation occurs at loads that correspond to a flow rate between zero flow (which is not enough) and minimum flow (which is too much) at the particular load condition.

Second, when the signal is between 50 and 100%, the cooler shell side becomes a reservoir of cold oil. This upsets the control system twice: once when the three-way valve just opens to the cooler and once when the cold oil has been completely displaced and the oil outlet temperature from the cooler suddenly rises from that of the cooling water to some higher value.

Finally, most of these systems are nonsymmetrical in that the process dynamics (lags and responses) are different for the cooling and heating portions.

To remedy these problems, several steps can be considered. As was shown in Figure 8.29jj, a cascade loop can be used so that upsets and disturbances in the circulating oil loop will be prevented from upsetting the process temperature. In addition, a slight overlapping of the two valve positioners is desirable. This will offset the beginning of cooling and the termination of heating phases so that they will not both occur at 50%. The resulting sacrifice of heat energy can be justified by the improved control obtained.

To protect against the development of a cold oil reservoir in the cooler, a continuous low flow through this unit can be maintained.

The recirculating design shown in Figure 8.29jj is a flooded system. Therefore, when hot oil enters it, a corresponding volume of oil must be allowed to leave it. The pressure control valve (PCV) serves this function. The same purpose can be fulfilled by elevating the return header. The important consideration is that whatever means are used, the path of least resistance for the oil must be back to the pump suction, in order to keep it always flooded and, thereby, to prevent cavitation.

Most multipurpose systems represent a compromise of various degrees. Figure 8.29kk, for example, illustrates a design in which low cost and rapid response to load changes are the main considerations. These characteristics are provided by using the minimum amount of hardware and by circumventing the transient characteristics of the exchangers. The price paid for this compromise is the full use of utilities at all times and the development of hot and cold reservoirs. In addition, because the cascade slave (which detected the temperature of the heat-transfer fluid in Figure 8.29jj) is eliminated, all supply disturbances have to affect the process temperature before corrective action can be initiated.

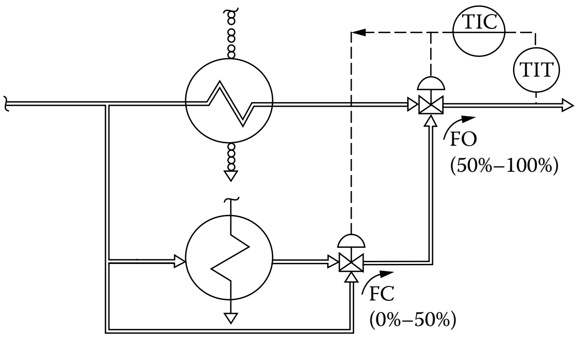


FIG. 8.29kk
This multipurpose temperature control system blends process streams at different temperatures.[1]

Other multipurpose heat-transfer systems are used in the control of jacketed chemical reactors. These can be single-purpose or double-purpose, or configured to operate with three or more different heat-transfer fluids, as was shown in Sections 8.8 to 8.11.

CONCLUSIONS

The temperature upset resulting from the sudden doubling of the load can be reduced by an order of magnitude through the use of feedforward control (Figure 8.29ii). Such a tenfold reduction in temperature error can make a great contribution to both safety and quality of production. Therefore, optimized or dynamically compensated feedforward heat exchanger controls are usually justified not on the basis of increased production or reduced energy costs, but on the basis of more stable, accurate, and responsive control.

Some of the considerations that the designer should consider are listed below:

The effect of supply disturbances on systems performance
The response speed of the system
Valve and sensor rangeability considerations
The quality of cooling water available
Potential problems due to nonsymmetrical dynamics
Potential problems caused by low minimum condensing pressures

In addition, it is always wise to consider the use of equal-percentage control valves furnished with positioners and to evaluate the advisability of using cascade, feedforward, or model-based advanced controls.

References

1. Lipták, B. G., "Control of Heat Exchangers," *British Chemical Engineering and Process Technology,* Vol. 17, No. 7/8, p. 637, July/August 1972.
2. Shinskey, F. G., *Process Control Systems,* 2nd edition, New York: McGraw-Hill, 1979, pp. 186, 227.
3. Shinskey, F. G., *Controlling Multivariable Processes,* Research Triangle Park, NC: ISA, 1981, p. 74.
4. Shinskey, F. G., *Controlling Unstable Processes,* Publication 413–4, The Foxboro Co.
5. Mathur, J., "Performance of Steam Heat-Exchangers," *Chemical Engineering,* September 3, 1973.
6. Driskell, L. R., *Control Valve Selection and Sizing,* Research Triangle Park, NC: ISA, 1983.

Bibliography

Astrom, K. J., *PID Controllers: Theory, Design, and Tuning*, 2nd edition, Research Triangle Park, NC: Instrument Society of America, 1995.

Åström, K. J. and Hägglund, T., *PID Controllers: Theory, Design and Tuning*, Research Triangle Park, NC: Instrument Society of America, 1995.

Bailey, E. G., "How to Save Coal," Bulletin No. 41, Wickliffe, OH: Bailey Meter Company, April 1918.

Batur, C., Sharpless, R. B., Duval, W. M. B., Rosenthal, B. N., and Singh, N. B., "Adaptive Temperature Profile Control of a Multizone Crystal Growth Furnace," *Proceedings of the American Society of Mechanical Engineers,* Dallas, TX, November 1990.

Boyle, W. E. and Hoyt, P. R., "Experimental Application of Combustion Controls to a Process Heater," Annual ISA Conference, September 13–17, 1948.

Dukelow, S. G., "Energy Conservation through Instrumentation," University of Houston Energy Conservation Seminar, February 7, 1975.

Dukelow, S. G., "Fired Process Heaters," presented at ISA/CHEMPID, April 1980.

Dukelow, S. G., "How Control Improvements Save Process Heater Fuel," 1979 Conference on Industrial Energy Conservation Technology, Houston, TX, April 23, 1979.

Edwards, D., "From LR to PC: Recent Developments in Furnace Control Systems," *Metals and Materials,* October 1990.

Forman, E. R., "Control Dynamics in Heat Transfer," *Chemical Engineering,* January 3, 1966.

Gibson, J. E., *Nonlinear Automatic Control*, New York: McGraw-Hill Book Co., 1963.

Goodwin, G. C., Graebe, S. F., and Salgado, M. E., *Control System Design,* Prentice-Hall, 2001.

Hesselgreaves, J. E., *Compact Heat Exchangers,* New York: Pergamon Press, 2001.

Hugo, A., "Limitations of Model Predictive Controllers," *Hydrocarbon Processing,* January 2000.

Johnson, R. K., "Modern Approach to Process Heater Control Systems," Parkersburg-Marietta ISA Section, Parkersburg, WV, February 8, 1977.

Kakac, S. et al., *Heat Exchangers: Selection, Rating, and Thermal Design,* 2nd edition, Boca Raton, FL: CRC Press, 2002.

Khalil, E. E., *Modeling of Furnaces and Combustors,* Abacus Press, 1982.

Kuppan, T., *Heat Exchanger Design Handbook,* New York: Marcel Dekker, 2000.

Levine, W., *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.

Lipták, B. G., "Process Control Trends,", *Control,* March, May, and July, 2004.

Lipták, B. G., "Control of Heat Exchangers," *British Chemical Engineering,* July 1972.

Luyben, W. L. and Tyreus, B. D., *Plantwide Process Control,* New York: McGraw-Hill, 1998.

Mathur, J., "Performance of Stream Heat-Exchangers," *Chemical Engineering,* September 3, 1973.

McKetta, J. J., *Unit Operations Handbook,* Volume 1, Mass Transfer, New York: Marcel Dekker.

McMillan, G. K., *Tuning and Control Loop Performance,* 3rd edition, Research Triangle Park, NC: ISA, 1994.

McMillan, G. K., *Process Industrial Instruments and Control Handbook,* New York: McGraw-Hill, 1999.

Ociai, S., "Creating Process Control Parameters from Steady-State Data," *ISA Transactions,* Volume 36, 1998.

Omata, S., Yamamoto, T., Maruyama, H., and Hotta, M., "Adaptive Control of a Furnace Temperature by Using a Microcomputer," *Proceedings of the 28th Society of Instrument and Control Engineers,* Tokyo, July 1989.

Perry, R. H., *Chemical Engineers Handbook,* 7th edition, New York: McGraw-Hill Book Co., 1997.

Ruel, M. "Loop Optimization: How to Tune a Loop," *Control Magazine,* Vol. 12, No. 5 (May 1999), pp. 83–86.

Seborg, D. E., Edgar, T. F., and Mellichamp, D. A., *Process Dynamics and Control,* 2nd edition, New York: John Wiley & Sons Inc., 2003.

Shinskey, F. G., "Feedforward Control Applied," *ISA Journal,* November 1963.

Shinskey, F. G., *Process Control Systems,* 4th edition, New York: McGraw-Hill, 1996.

Shinskey, F. G., *Process Control Systems: Application, Design and Tuning,* 4th edition, New York: McGraw-Hill, 1998.

Skogestad, S. and Postlethwaite, I., *Multivariable Feedback Control,* New York: John Wiley & Sons Inc., 1996.

Waller, M. H., "Measurement and Control of Paper Stock Consistency," ISA Conference in Houston, TX, October 1992.

8.30 Inert Gas Blanketing Controls

D. S. NYCE, B. G. LIPTÁK (2005)

Applications:	Inerting systems are required for tanks containing flammable gases and vapors. Their goal is to keep the oxygen concentration in the vapor space of the tank below the minimum oxygen for combustion (MOC).
Range:	The MOC of most flammable gases and vapors is below between 5 and 12%; therefore, the setting of oxygen levels above 10% will not serve fire protection (Table 8.30e). Automatic inerting units can control the oxygen concentrations down to 0.1%.
Inaccuracy:	Automatic inerting systems can control the vapor space oxygen concentration within 0.1% of set point. The control system set point should be selected to provide a 30% safety margin below the MOC, because of the uneven mixing of the gases in the vapor space.
Reliability:	Inerting controls that do not include the measurement and control of oxygen may not be reliable. These systems, depending on the control of pressure or flow, provide no positive guarantee of safety and require long time periods of purging. Others are reliable but use a lot of inert gas.
Costs:	A manually operated combination of variable-area flowmeter and valve can be installed for less than \$200. A pressure control system costs \$500 to \$1000, depending on the required flow rate. An automatic system, with oxygen sensing and control, costs \$1000 to \$2000 for a single point of installation. The cost of multipoint systems adds about \$500 per channel.
Partial List of Suppliers:	Air Liquide (www.us.airliquide.com) Air Products (www.airproducts.com) Alpha Omega (www.aoi-corp.net) Delta F (www.delta-f.com/) Linde (www.us.lindegas.com) Neutronics (www.neutronicsinc.com) NTRON (www.ntron.com) Revolution Sensor Company (www.rev.bz) Teledyne Analytical Instruments (www.teledyne-ai.com)

INTRODUCTION

Inerting or inert gas blanketing is a fire- and explosion-prevention method that works by lowering the oxygen concentration of a flammable gas mixture. The technique is used to blanket the vapor space in tanks containing flammable process liquids, combustible dusts, fibers, and particulate solids. The applications for inerting systems include chemical reactor vessels, mixers, centrifuges, web coating lines, and mills. Although inert gas blanketing is also used to protect nonflammable process materials from discoloring or other forms of degradation, this section concentrates on inert gas blanketing (inerting) for the purpose of fire prevention.

In the process of inert blanketing, the vapor space of the tank is filled with inert gas in order to prevent fire and explosion. The term *inert gas* refers to any gas that will not support combustion. The inert gas most commonly used is nitrogen, although it is also possible to use CO_2, argon, or other gases that are oxygen deficient, such as steam, or the products of combustion.

Pollution and Personnel Safety

Many processes use volatile organic compounds (VOCs) as solvents that, when the vapor space of a storage tank is purged, are released to the atmosphere along with the inert gas exhaust. Therefore, the less inert gas is used in the process of inerting, the less VOCs will be lost to the atmosphere.

Minimizing VOC emissions is important for both safety and environmental reasons. As will be seen in the discussion below, the oxygen-based inerting systems are the best candidates for minimizing the use of inert gas and, therefore, also minimizing VOC emissions to the atmosphere.

A vapor space that has been purged by inert gas to obtain a low oxygen concentration, in order to prevent fires, also has too little oxygen for breathing. Therefore, plant operation should make sure that this does not result in a low-oxygen hazard to personnel.

Inert gas that is exhausted to the room normally has little effect on breathable air. A potential problem could arise, however, if an operator opens a hatch of an inerted vessel and leans into it for any reason. Therefore, alarms or interlocks must be provided to protect against such possibility.

THE COMBUSTION PROCESS

Combustion is a chemical oxidation process that occurs rapidly enough to produce heat and light in the form of a flame or glow. Combustion is called deflagration if the rate of propagation of the combustion zone is slower than the speed of sound in the unreacted medium.[1] If the propagation speed is greater than the speed of sound, it is called detonation. An explosion is the bursting of an enclosure due to internal pressure generated by a deflagration.

Inerting

An inerting system can prevent combustion of flammable materials in almost any sealed space by keeping the oxygen concentration below the level that could support the minimum oxygen for combustion (MOC). The most common application of inerting systems is to protect the vapor space (headspace) of a vessel containing a flammable liquid. The flammable liquid and the headspace in an agitated tank are shown in Figure 8.30a.

A flammable gas or vapor is one that will burn in air at normal temperatures and pressures, if it reaches a certain minimum concentration. Liquids can generate flammable vapors if they are above their flash point. The flash point of a liquid is its lowest temperature (at normal atmospheric pressure) at which the rate of evaporation is sufficient to form a combustible mixture near the liquid surface.

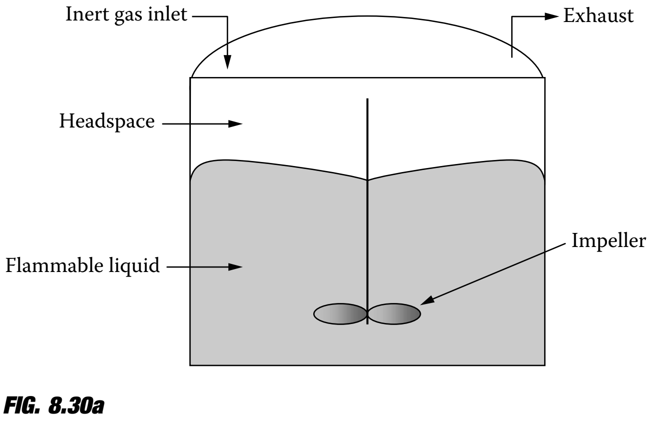


FIG. 8.30a
Fire protection is provided by introducing inert gas into the headspace of a mixing vessel that contains flammable liquids.

Alternatives to Inerting

For each flammable vapor or gas, there is a range of concentrations in air over which combustion is possible. This is the flammable range and is the range between the lower flammable limit (LFL) and the upper flammable limit (UFL)

Below the LFL, the gas concentration provides insufficient fuel to propagate a flame. Above the UFL, the gas concentration is too rich (not enough oxygen) to propagate a flame. If the gas concentration falls between the two limits, then it can be ignited and act as the fuel in a combustion process. In industrial applications, where process vessels contain flammable liquids at temperatures above their flash points, it is assumed that the vapor space can contain a flammable range of vapor concentrations. Therefore, fire prevention is necessary to ensure the safety of personnel and equipment.

As an alternative fire-prevention method to inerting systems, one can maintain the gas or vapor concentration above the UFL, where the flammable vapor concentration is too rich to support combustion (not enough oxygen). This is an acceptable method of prevention only in vessels that are normally sealed, with no entry of air allowed.

Another alternative method is to maintain the gas or vapor concentration below the LFL, where there is insufficient fuel concentration to support combustion. This is accomplished by purging with a nonflammable gas and can be used when the flow of flammable gas is low, or the flammable vapor is not being generated at a high rate during the normal operation of the process.

The Combustion Triangle

In order for combustion to take place, three requirements must be satisfied. These are the presence of a fuel, an oxidizer, and an ignition source (Figure 8.30b). In most cases (except with

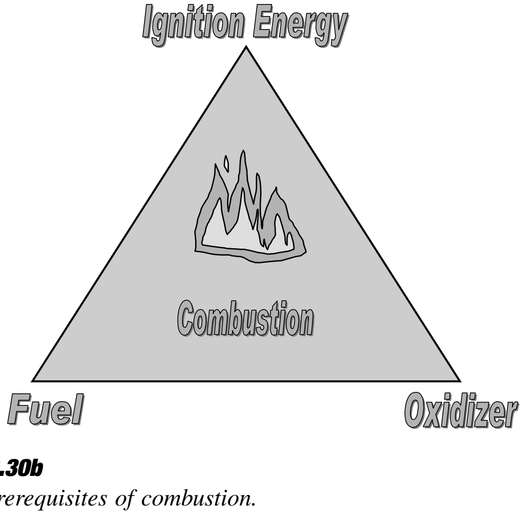


FIG. 8.30b
The prerequisites of combustion.

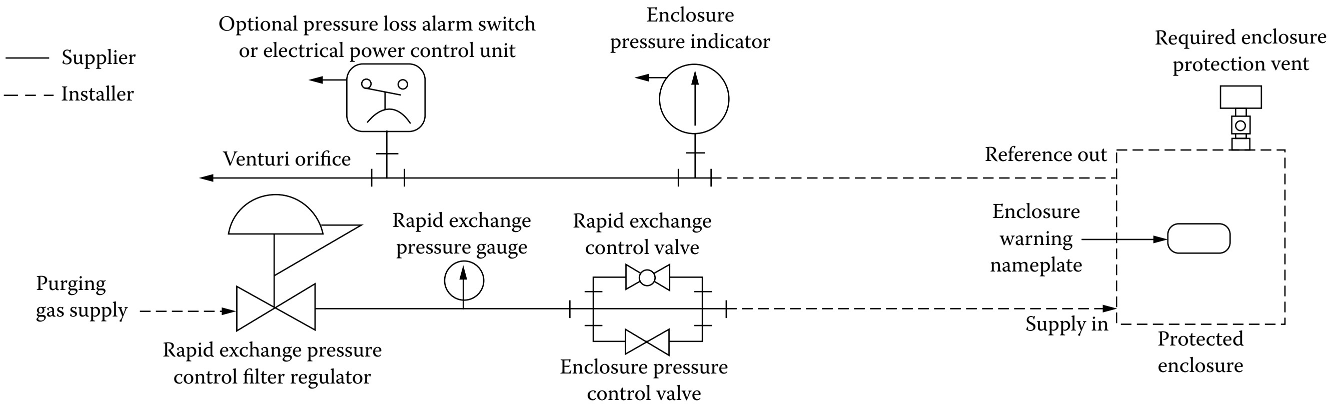


FIG. 8.30c

Packaged purge systems are marketed for Class 1 area and are provided with rapid exchange purging capability. (Courtesy of Bebco Industries.)

certain materials that contain their own oxidant), removing any one of these three prerequisites will prevent combustion.[2]

Fuel can be defined as a flammable gas or vapor within the combustible range of concentration, or a combustible fiber, dust, or other particulate material. This prerequisite of combustion is a given, so the design engineer must concentrate on the removal of at least one of the other two.

Common sources of ignition are electric sparks and heat. An electrical discharge can come from the opening of an electrical contact, a broken wire, or a discharge of static electricity. Heat can be generated by the process, by heaters, or by electrical faults.

The use of intrinsically safe (IS) electrical systems can eliminate some electrical ignition sources, but cannot protect against process-generated heat or static electricity discharge. Many industrial processes generate static electricity by mixing or agitating dielectric materials. For these reasons, the complete elimination of ignition sources, even when an IS electrical system is installed, is not always possible. Therefore, in many instances, the removal of the oxidizer by inert blanketing is the most practical means of fire and combustion protection.

The oxygen concentration of atmospheric air is about 20.9%, the remainder being mostly nitrogen. This percentage concentration remains the same over different pressures (and, therefore, different altitudes), although it is affected slightly by changes in relative humidity. If the air within the tank can be replaced with an inert gas, then the oxidizer prerequisite in the combustion triangle is removed. This is the goal of inert blanketing. A manual purge system is shown in Figure 8.30c.

Fibers, Dusts, and Particulate Solids Some dusts and finely divided powders, such as those of magnesium and zirconium, can deflagrate with no additional oxygen being present in the atmosphere. Therefore, their storage tanks cannot be protected by inerting.

Some magnesium or lithium compounds can react with nitrogen, and so argon is used as the inert gas supply for the gas blanket. Sometimes, fibers can continue to smolder after blanketing with inert gas, only to burst into flame when exposed to air later.

In addition, the mechanical difficulty of ensuring homogenous dispersion of an inerting gas throughout porous solid materials requires special attention. Inerting of the gas space above a process liquid contained in a vessel is a fairly straightforward process, but inerting of combustible fibers, dusts, and particulate solids requires a careful evaluation of the specific properties of the materials involved.

INERTING CONTROL SYSTEMS

Inert blanketing systems can be operated on pressure, flow, or oxygen control. Flow-controlled inerting is also called a "timed-volume" system, while oxygen-based inerting is often referred to as "automatic control."

Flow-Controlled Inerting

A *flow-based* inerting system is also called a *timed-volume* control system. It consists of a purge flowmeter in combination with a differential pressure regulator that acts as a self-contained flow regulator (Figure 8.30d). The inert gas flow through a standard purge flow controller is usually adjustable between 0.2 and 2 SCFH (6 and 60 slph). Over this 10:1 flow range, the flow control error is within 5% of full scale. The standard pressure and temperature ratings are 150–300 PSIG (1–2 MPa) and 212–572°F (100–300°C). The purge flowmeter can be obtained in much larger sizes to control higher flow rates as required.

At start-up, the required total initial charge of the inert gas is set to be about five times the volume of the vapor (dilution ratio[1] of 5) space of the vessel. This amount of inert gas, if introduced at a high enough velocity to create turbulent flow, will usually lower the oxygen concentration in the vapor space of the tank to under 0.5%. The purge flowmeter is

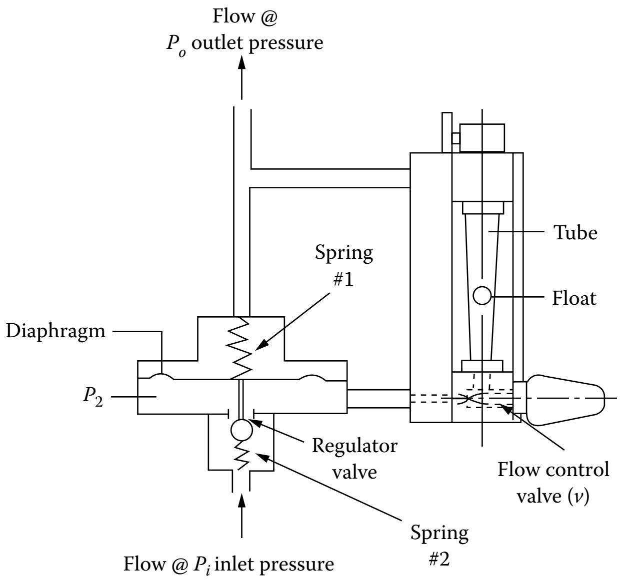


FIG. 8.30d

Purge flow regulator consisting of a glass tube rotameter, an inlet needle valve, and a differential pressure regulator. (Courtesy of Krone Inc.)

usually sized such that it will deliver this initial volume of inert gas into the tank's vapor space in an "initial purge time" of about 5 min.

When the inert blanketing system is manually operated, the operator is usually provided with an instruction sheet that is mounted next to the purge flowmeter. The first line on the instruction sheet gives the initial purge time period and the corresponding flow setting that the operator should use during start-up.

After the initial purge is completed, the purge flow rate is lowered to the continuously maintained rate that is required. This second flow rate is selected to make sure that the oxygen level in the vapor space will not reach the minimum oxygen for combustion level for the material in the tank.[1] If the start-up and the continuous purge rates are drastically different, a single purge meter might not be able to control both, and two purge meters have to be installed in parallel.

The continuous purge flow rate has to be set high enough to keep the oxygen concentration in the vapor space under MOC limits even during periods when a valve or a hatch has to be open to the atmosphere in order to add an ingredient or remove a product. Table 8.30e provides MOC values for a variety of flammable materials. For other material, consult Reference 1.

Flow-based inerting systems are usually used because of their simplicity and low installed cost. Their disadvantage is that the controlled variable (oxygen concentration in the vapor space) is only assumed, but not measured. Another disadvantage is their higher operating cost, because they tend to consume more inert gas than the other inerting systems.

TABLE 8.30e

MOC (Minimum Oxygen for Combustion) Values of Some Flammable Materials

Flammable Material	*MOC (%)*
Carbon monoxide	5.5
Gasoline	12.0
Hydrogen	5.0
Methane	12.0

Naturally, the manual mode of operation is only an option, not a necessity. Inert blanketing systems can also be controlled automatically. In that case, automatic on/off valves are required to start up the system, and timer controls are needed to automatically switch the purge flow rate set point from the start-up to the continuous rate.

Pressure-Controlled Inerting

Figure 8.30f illustrates the configuration of a pressure-controlled inerting control system. The start-up phase of the operation is the same as it was in the flow-controlled inerting system. During the start-up period, the purge valve is opened and the purge flowmeter is set for the initial purge rate. Once the initial purge time has expired and the vapor space of the vessel has been sufficiently diluted, the purge valve is closed.

After the purge valve has been closed, the pressure control valve (PCV) maintains the pressure in the vapor space at a safe value, determined by the design pressure of the tank and its associated equipment, including emergency and conservation vents.

Pressure-based inerting controls can reduce the inert gas consumption if the leakage rate is small because of tight vessel construction. If the tank operation is such that the inert gas flow needs to be drastically increased periodically (say, because of the need to open a hatch), the PCV has to be oversized, which can result in leakage and overpressuring during normal operation.

Similarly to the flow-based purge controls, the operation of the pressure-based inerting systems can also be automated by the use of a timer that operates an automated purge valve.

Oxygen-Controlled Inerting

While the pressure- or flow-based inerting systems indirectly lower the oxygen concentration of the vapor space, the measurement of oxygen is a direct and automatic method of keeping that concentration under the MOC limit. There are a number of methods of detecting oxygen concentration (see Section 8.42 in Chapter 8 in the first volume of this handbook), and once the measurement signal is available, it can be used in on/off or continuous control configurations.

The Oxygen Sensor In an oxygen-based automatic inerting system, the oxygen sensor is usually either the electrochemical

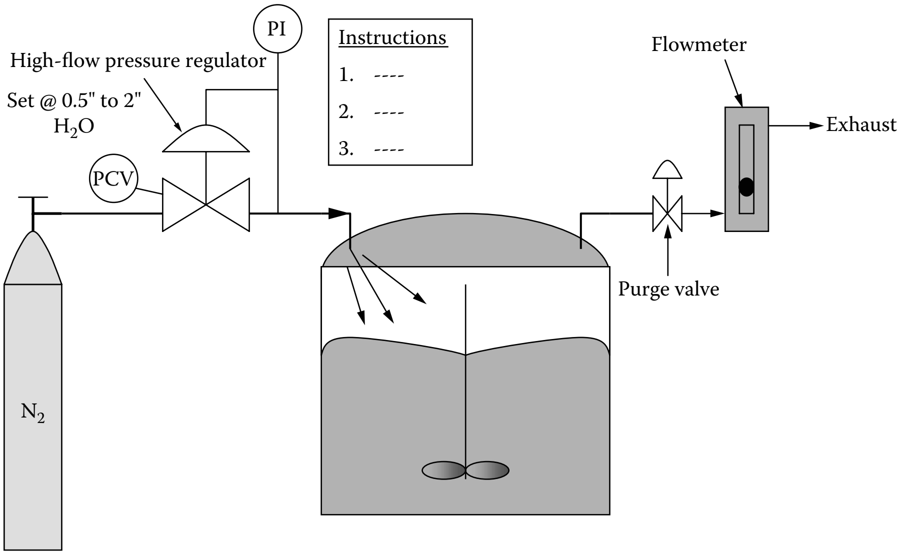


FIG. 8.30f
The main components of a pressure-based inert blanketing control system.

fuel cell-type design or the paramagnetic type. Fuel cells are small and relatively inexpensive, and can measure oxygen as a percentage by volume. They generate an output current by oxidizing an internal material, such as lead. Fuel cells also have an inherent true zero, i.e., no output is generated if no oxygen is present. The true zero means that no zero gas is needed for checking the calibration of the fuel cell.

All gas sensors should be routinely maintained so that their calibration is validated on a schedule. With an oxygen fuel cell, only a span gas is needed for calibration. Almost universally, fresh air is used to calibrate an oxygen fuel cell that reads in percent by volume. It can be adjusted to 20.9% in fresh air (ignoring the small changes that are due to changes in relative humidity). If using a membrane-type fuel cell that measures partial pressure, allowance may be made for changes in the atmospheric pressure.

For its measurement, the paramagnetic oxygen sensor relies on the force of alignment of oxygen molecules along the lines of flux in a strong magnetic field. In one design, this force generates a small motion that is measured through the use of a mirror, a light source, and a pair of optical sensors. Because a paramagnetic oxygen sensor measures the partial pressure due to oxygen, it is affected by atmospheric pressure, similarly to a membrane-type of fuel cell.

One limitation of paramagnetic oxygen analyzers is that a major repair is required if the sensing cell is damaged, and replacement is fairly expensive. In contrast, with a fuel cell, the sensing element is disposable and can be replaced at a nominal cost.

On/Off Oxygen Controls When the vapor space oxygen concentration is detected, there is no need to calculate the time required to complete the initial purge cycle during start-up, because the switching to the continuous mode of purging occurs automatically (Figure 8.30g).

At the beginning of a start-up cycle, the oxygen analyzer is likely to detect 20.9% oxygen (the concentration that is present in a normal atmosphere). Therefore, the solenoid valve on the inert gas supply line will open. As the inert gas displaces the vapors in the headspace of the vessel, the oxygen sensor will detect a steady decrease in the oxygen level.

When the oxygen level has dropped to the set point of the inerting controller, a green light goes on, indicating that it is safe to start up the process. During normal operation, the solenoid valve is cycled on and off as necessary to keep the oxygen level under control. (This system was invented by the author in the late 1970s.)

To reduce the amount of cycling, a dead band around the set point is provided in the on/off control switch. In normal operation, a red light illuminates when the solenoid is open and the oxygen level is being lowered to the set point during start-up. That light goes out and a green light comes on when the oxygen level is within the dead band of the set point's control range. A yellow light cycles on and off as the solenoid cycles on and off during normal operation.

Two control points are set by the user, depending on the MOC of the particular flammable materials present in the vessel. For example, if the published MOC of the stored liquid is 8.0%, the user might set the alarm point at 5.0% and the high and low control points (the control band) between 4.0 and 3.0%. With such settings, the solenoid valve will then open at or above 4.0%, and close at or below 3.0%.

This way, there is a safety margin between the MOC of 8.0% and the alarm point of 5.0%, while the minimum amount of inert gas is used. This control system also provides a continuous indication of the safety and operating status of the vessel.

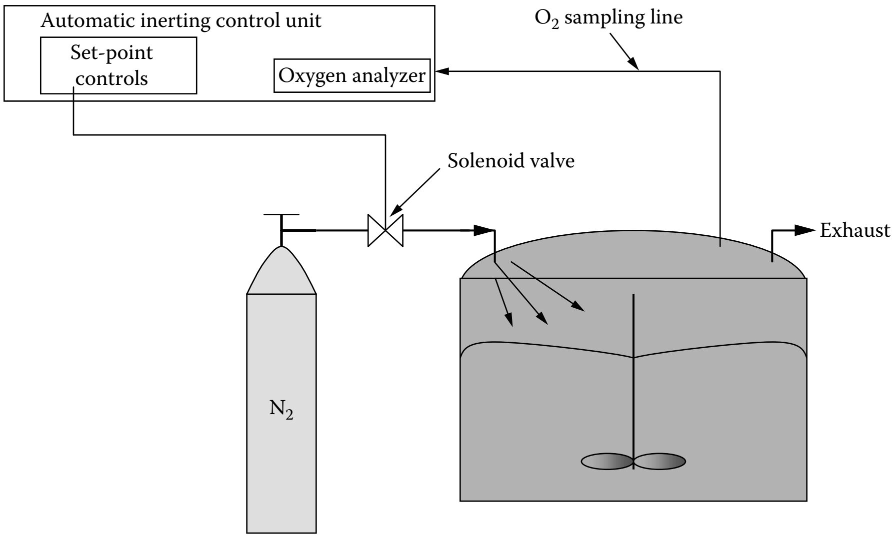


FIG. 8.30g
Automatic, oxygen-based, on/off inerting control system.

This oxygen-based on/off control system uses the least amount of inert gas and provides more safety than the flow- or pressure-based systems, because it directly detects the oxygen concentration. Its disadvantage is higher installed cost. The higher initial investment can usually be recovered through savings due to the reduction in inert gas usage.

COMPONENT DESIGN CONSIDERATIONS

The proper selection and installation of the purge control solenoid valves and of the purge flow rates deserve further discussion.

Do Not Use "Fail-Safe" Solenoids

On the surface it would appear that to select a fail-open solenoid valve (i.e., the valve is open when not energized) would guarantee safety, because it would open the inert gas flow to the protected vessel whenever power failed. In fact, such a design would mean that all of the solenoid valves on all of the inerted vessels would open up during a power failure.

This is undesirable if the inert gas supply system could not keep up with the demand and, therefore, no vessels would be adequately protected. Instead, the inert gas solenoid valves can be the energize-to-open design, and any failure (system or power) can require that the operation of the protected process be terminated until the problem is remedied.

The decision on the failure position of the solenoid should be based on the balance of the costs and consequences of the two solenoid failure options. Fail-closed solenoids can be used if the temporarily shutting down of the process is acceptable, while fail-open solenoids are the proper choice if providing a sufficiently high-capacity emergency inert gas supply system, which could supply all the failed open solenoids during periods of power failure, is feasible.

If fail-open solenoids are used (i.e., the valves open when energized), one area of concern is that the coil circuit wiring of a solenoid valve can open up (i.e., fail as an open circuit), while the controls are applying power to the valve circuit to open the valve when it is not. To protect against this, one can measure the current flow through the coil of the solenoid valve. This way, the solenoid valve will be reported to be energized only if current is flowing in its coil.

Purge Flow Rate Variation

In the flow-based inerting system, after the initial start-up purging cycle is completed, a continuous flow rate of purging is maintained. This rate must be large enough to maintain a low oxygen concentration even if the protected vessel has a substantial leak rate.

In addition, in case of an outdoor storage tank, on a hot and sunny summer day, the vapor temperature in the headspace will be quite high and will drop drastically during a thunderstorm. In that case, the tank will rapidly cool, reducing the gas volume, and tending to create a vacuum. If during such episodes outside air is drawn in, a hazardous condition can evolve.

The same applies to a process vessel from which the product is quickly discharged. During such episodes, it is necessary to quickly and safely break the vacuum by manually or preferably automatically switching to a higher capacity purge system.

The need for a variable flow rate is not a problem with a pressure-based or automatic systems, as long as the pressure-based system has a sufficiently high flow capacity. Similarly, the oxygen-based purge controls can also handle changes in purge flow demand as long as the inert gas supply and the size of the solenoid valve is sufficient.

CONCLUSIONS

The initial cost of flow- and pressure-based inerting systems is lower than that of oxygen-based ones. The flow- and pressure-based systems are indirect in the sense that they do not directly measure the oxygen content of the vapor space, and they also consume more inert gas and release more process vapors (including volatile organic compounds) into the atmosphere. Over the long run, when the operating costs of purge gas consumption are considered, oxygen-based systems can be less costly overall.

Each type of purging system requires routine maintenance to make sure that all valves and lines are clear and operational. In addition, an automatic system with oxygen concentration feedback requires maintenance of the gas sampling system and sensor. If an electrochemical fuel cell-type of oxygen sensor is used, it must be replaced once per year or when it fails to calibrate.

References

1. National Fire Protection Agency (NFPA) 69, "Standard on Explosion-Prevention Systems," 2002 edition, Quincy, MA: National Fire Protection Agency, 2002.
2. Nyce, D. S., *Linear Position Sensors, Theory and Application,* New York: John Wiley & Sons, 2003, p. 140.

Bibliography

Anderson, N. A., *Instrumentation for Process Measurement and Control,* 3rd edition, Boca Raton, FL: CRC Press, 1998.

Harvey, J. F., *Pressure Component Construction,* New York: Van Nostrand Reinhold, 1980.

Lipták, B. G. (Ed.), *Instrument Engineers Handbook, Process Measurement and Analysis,* 4th edition, Boca Raton, FL: CRC Press, 2003.

Magison, E. G., *Intrinsic Safety,* Research Triangle Park, NC: Instrument Society of America, 1984.

Shinskey, F. G., *Simulating Process Control Loops Using Basic,* Research Triangle Park, NC: Instrument Society of America, 1990.

Webster, J. G. (Ed.), *The Measurement, Instrumentation, and Sensors Handbook,* Boca Raton, FL: CRC Press, 1999.

8.31 ORP Controls

D. M. GRAY (1985) **R. R. JAIN** (1995) **R. H. MEEKER, JR.** (2005)

INTRODUCTION

Oxidation-reduction potential (ORP) measurements are useful for the quantitative determination of ions, for monitoring chemical reactions, for determining the extent to which oxidizing or reducing reactions have taken place, and for determining the concentration of chemical species present. The ORP analyzer probes are described in Section 8.41 in Chapter 8 of the first volume of this handbook, *Process Measurement and Analysis*. While ORP measurements are somewhat similar to those of pH, the potential reading must be interpreted even more carefully for meaningful results.

In this section a description is given of the oxidation-reduction (redox) reactions and their control using oxidation-reduction potential measurement. ORP and pH instrumentation is compared and an example of titration curves is provided to serve the understanding of the ORP response. The pH electrode is an ORP sensor that is specific to hydrogen and has been designed to selectively develop a hydrogen ion activity-related potential.

Several major industrial applications for ORP control are described in this section, including cyanide oxidation, chrome reduction, sodium hypochlorite (bleach) production, and removal (scrubbing) of chlorine and chlorine dioxide from gaseous emissions. Both batch and continuous cyanide oxidation and chrome-reduction processes are discussed for waste treatment applications. The applications of bleaching of paper pulp and the treatment of swimming pool and spa water are also covered.

ORP AS A PROCESS VARIABLE

An oxidation-reduction reaction involves the transfer of electrons from one material (*oxidation*) to another (*reduction*). In industrial applications, oxidizing or reducing agents are used to promote the desired reaction. An example of such a reaction occurs between chromium and ferrous ions, as shown in Equation 8.31(1):

$$\mathrm{Cr}^{+6} + 3\mathrm{Fe}^{+2} \rightarrow \mathrm{Cr}^{+3} + 3\mathrm{Fe}^{+3} \qquad \textbf{8.31(1)}$$

In this process, the *reducing agent* (ferrous ion) donates electrons to the chromium, thus reducing chrome while the iron is oxidized.

The Nernst Equation

The measured electric potential developed by oxidation-reduction reactions is described by the Nernst equation:

$$E = E^0 - \frac{2.303RT}{nF}\log\frac{[\alpha_X]^x[\alpha_Y]^y[\alpha_Z]^z\ldots}{[\alpha_A]^a[\alpha_B]^b[\alpha_C]^c\ldots} \qquad \textbf{8.31(2)}$$

where

E = the developed potential (mV)

E^0 = the standard potential (mV)

R = the gas law constant (e.g., 1.987×10^{-3} kcal/mole°K, in SI units)

T = absolute temperature (e.g., 298°K, in SI units, for a reference temperature of 25°C)

n = number of electrons transferred in the reaction

F = Faraday constant (e.g., 23.06×10^{-3} kcal/mV, in SI units)

For a reaction written as a reduction equation:

$$aA + bB + cC + \cdots + ne^- = xX + yY + zZ + \cdots \qquad \textbf{8.31(3)}$$

$[\alpha_A]$, $[\alpha_B]$, $[\alpha_C]\ldots$, and $[\alpha_X]$, $[\alpha_Y]$, $[\alpha_Z]$, … are the *activities* of each respective ion (note, only ions, not neutral molecules, are counted in the equation); a, b, c, … and x, y, z, … are coefficients of the balanced chemical reaction.

Activity α is related to the molar concentration (e.g., moles/liter) as follows:

$$\alpha = \gamma[A] \qquad \textbf{8.31(4)}$$

where

α = the ionic activity

γ = the activity coefficient

$[A]$ = concentration of reactant A

Generally, the more ions there are in solution, the more difficult it is for the exchange of electrons to occur that results in development of an electric potential. The activity coefficient reflects this. Most solutions of interest are dilute enough that the activity coefficient is very close to unity, though this is not always the case. Fortunately, in most process control applications it is ionic activity (the tendency of a component to participate in a reaction) that is of most interest and not the absolute concentration of the ions.

With the constants for R, T, and F inserted and a reference temperature of 25°C assumed, Equation 8.31(2) can be written in a more commonly used form:

$$E = E^0 - \frac{59.16}{n}\log\frac{[\alpha_X]^x[\alpha_Y]^y[\alpha_Z]^z\ldots}{[\alpha_A]^a[\alpha_B]^b[\alpha_C]^c\ldots} \qquad \mathbf{8.31(5)}$$

The standard potential, E^0, is defined for any given half-cell reaction as the potential obtained when all ions or molecules in solution are at a concentration of 1 mole/liter and all gases are at a partial pressure of 1 atmosphere.[2] By convention, the standard potential is normally referenced to a temperature of 25°C and to the hydrogen ion/hydrogen (H^+/H_2) half-cell reaction ($E^0 = 0$ mV).

As indicated by Equations 8.31(2) and 8.31(5), ORP is dependent upon concentrations of all of the reactants involved. If any of the reactants are hydrogen (H^+) or hydroxide (OH^-) ions, then the ORP measurement is pH-dependent. This is evident by further refining the Nernst equation specifically for pH measurement:

$$E = E^0 - 0.198T_k \log\frac{1}{[\alpha_{H^+}]} \qquad \mathbf{8.31(6)}$$

and

$$\alpha_{H^+} = \gamma[H^+] \qquad \mathbf{8.31(7)}$$

The similarities between the ORP and pH measurements are evident in the fundamental principles upon which the measurements are based. It is helpful to realize the pH is just an ion-specific ORP sensor, where the measurement electrode has been designed to selectively develop a potential related to hydrogen ion activity. Similarly to hydrogen, other ion selective sensors can also be built.

Measurement

During the chemical reaction shown in Equation 8.31(1), an inert metal electrode is placed in contact with the solution and detects the solution's ability to accept or donate electrons. The resulting oxidation-reduction potential (redox potential) is directly related to the progress of the reaction. A ferrous reducing ion (Fe^{+2}) provides electrons and tends to make the electrode reading more negative. A chromium oxidizing ion (Cr^{+6}) accepts electrons and tends to make the electrode reading more positive. The resulting net electrode potential is related (logarithmically) to the ratio of concentrations of oxidizing and reducing ions in solution.

ORP is an extremely sensitive measurement of the degree of treatment provided by the reaction. However, it is generally impractical to relate this measurement to a definite concentration of a reaction component. This is due to the dependence of the measurement upon the concentration ratios of all oxidants and reductants. Interpretation of the measurement is further complicated because of its nonlinearity with respect to concentration, plus the detector's inaccuracy and the lack of temperature compensation.

The exact potential to ensure complete treatment can be calculated theoretically, but in practice is subject to variations in reference electrode potential, pH, the presence of other waste stream contaminants, temperature, purity of reagents, and so forth.[2] The target potential is usually determined empirically by testing the treated process flow for trace levels of the material to be eliminated or minimized. The optimum control point (ORP reading) occurs when just enough reagent has been added to complete the reaction. Approximate suggested control points are given later in this section, but should be verified on-line by testing the actual samples.

Instrumentation

The detectors used for ORP measurement are very similar to that used for pH measurement. The reference electrode is identical to the pH reference electrode. The measuring electrode, however, consists of a noble metal instead of a glass bulb. The noble metal electrode is usually made from platinum, gold, or nickel, but carbon may also be used. The noble metal electrode is subject not only to coating, but also to poisoning, which can result in sluggish or inaccurate measurement.

For this reason, it is advisable to use scrubber-type probe cleaners (Figure 8.31a) or retractable probes that in the retracted position can first be cleaned by a cleaning fluid and then can be recalibrated using a standard reference solution.

ORP measurement electrodes must be maintained to provide a very clean metal surface. Routine cleaning of electrodes with a soft cloth, dilute acids, or cleaning agents is needed often to promote fast response. The noble metal comprising the measurement electrode is often platinum or gold. It is important to carefully select the electrode material for compatibility with the process. For example, gold may actually have a greater tendency to corrode than platinum in some solutions; however, gold is generally more resistant to strong oxidizing agents than platinum.

Similarly, the composition and design of the reference electrode must be compatible. Incompatible reference electrode solutions can be quickly *poisoned* by the process fluid, rendering the measurement useless.

At a constant solution temperature of 25°C, the slope of pH measurement will be 59.16 mV per pH unit, so full scale would be approximately +414 to –414 mV for 0–14 pH.

Polarity of ORP measurements can be a source of confusion. Electrodes that measure pH produce a negative potential with upscale pH readings (because pH is defined as the *negative* log of the hydrogen ion concentration, and more hydrogen ion activity means a more acidic solution). Thus, the *reference electrode* connection of a pH instrument is the *positive* input.

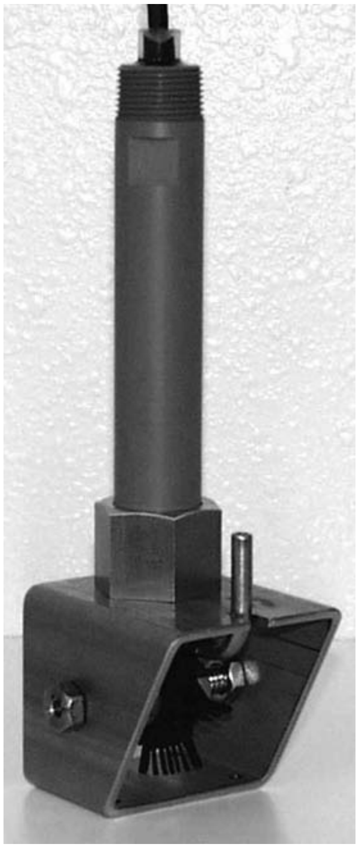


FIG. 8.31a
Scrubber-type probe cleaner. (Courtesy of Universal Analyzers Inc.)

Relative to the reference electrode, the more positive the potential of the ORP noble metal measurement electrode is, the greater the oxidizing environment of the solution in which it is placed, and conversely, the more negative, the more reducing the environment is. When ORP equipment is based on a pH instrument design, it may be necessary to connect the ORP electrode to the reference input and the reference electrode to the measuring terminal to achieve this response.

When sensor connectors are of the BNC style, this may not be practical. Some combination pH/ORP transmitters have a jumper or configuration selection that will swap the polarity internally or permit the mA output range to be reversed. The applications described in this section assume the ORP sensor readings indicate increasing potential with increasing oxidizing environment. This is the most common convention, consistent with most chemistry texts and standard potential tables.

Calibration

ORP instruments are calibrated like voltmeters, measuring absolute millivolts (mV), although a standardized (zero) adjustment is often available on instruments designed also for pH measurement.

TABLE 8.31b
ORP Values (in Millivolts) of Quinhydrone-Saturated pH Buffer Solutions (Using Saturated Silver–Silver Chloride Reference Electrode)

pH Buffer	*68°F (20°C)*	*77°F (25°C)*	*86°F (30°C)*
4.010	267	263	259
6.86	100	94	88
7.00	92	86	80
9.00	–26	–32	–39
9.18	–36	–43	–49

To verify operation of electrodes, it is useful to have a known ORP solution composition using quinhydrone and pH buffer solutions. These must be made up fresh to prevent air oxidation and deterioration. A quinhydrone reference solution prepared in a pH 4 buffer solution should read about 264 mV at 25°C with a platinum measurement electrode and silver–silver chloride (Ag/AgCl) reference electrode. With quinhydrone in a pH 7 buffer solution, the same-type electrodes should read about 87 mV at 25°C.

Quinhydrone-saturated pH buffers are used to establish known potentials as a check when a shift is detected (manually or automatically) in either the span or the potential. See Table 8.31b for ORP values in millivolts in quinhydrone-saturated pH buffer solutions at various temperatures.

A more stable ORP reference solution has been developed (also known as "Light's Solution"), consisting of 0.1 M ferrous ammonium sulfate, 0.1 M ammonium sulfate, and 1.0 M sulfuric acid. Its ORP is +476 mV when measured with a silver–silver chloride, saturated potassium chloride reference electrode.[3]

The ORP measurement is displayed in millivolts. Temperature compensation is not used because the compensation would be different for each reaction, making it impractical to produce a generally useful temperature correction as can be done with pH.

ORP CONTROL

The ORP control process (like pH control) requires that the control system designer understand the chemistry of the process that is to be controlled. In addition to vessel size, vessel geometry, agitation requirements (needed to guarantee uniform composition), and reagent delivery systems, solid removal problems must also be considered.

In cases where one or both of the half-reactions (redox reaction) involve hydrogen ions, ORP measurement becomes pH-dependent. The potential changes measured by the ORP electrode will vary with the redox ratio, but the redox ratio will vary with pH. Therefore, it becomes necessary to experimentally determine the control point, and both pH and ORP measurements are required to control the process.

As with pH, reliable ORP control requires vigorous mixing to ensure uniform composition throughout the reaction tank. For continuous control the tank should provide adequate retention time (filled tank volume divided by process flow rate), typically 10 min or more.[4]

ORP measurement in relation to concentration of reactants (and, therefore, reactant flowrate) produces a nonlinear response similar in shape to the familiar pH titration curve. This nonlinearity in the ORP titration curves can make PID control difficult. The degree of difficulty is a function of how tightly the ORP is to be controlled, where the operating point is on the ORP curve, and over what range of conditions should the ORP be controlled. If necessary, techniques used to improve pH control, such as characterization of the error, can also be applied in the control of ORP (see Section 8.32).

Complete treatment requires a slight excess of reagent and a control point that is slightly beyond the steep portion of the titration curve. Control in this plateau area, where process gain is relatively low, can be obtained by simple on/off control. Reagent feeders are typically metering pumps or solenoid valves. A needle valve in series with a solenoid valve can be used to set the reagent flow more accurately and to improve on/off control.

Chrome Waste Treatment

Chromates are used as corrosion inhibitors in cooling towers and in various metal finishing operations, including bright dip, conversion coating, and chrome plating. The resulting wastewater from rinse tanks, dumps, or cooling tower blowdown contains the toxic and soluble chromium ion (Cr^{+6}), which must be removed before discharge to comply with EPA regulations.

The most frequently used technique for chrome removal is a two-stage chemical treatment process. In the first stage, acid is added to lower the pH, and reducing agent is added to convert the chrome from soluble Cr^{+6} (toxic) to Cr^{+3} (nontoxic). In the second stage the wastewater is neutralized, forming insoluble chromium hydroxide, which can then be removed.

First Stage In the first stage, sulfuric acid is used to lower the pH to approximately 2.5 to speed up the reduction reaction and ensure complete treatment. The most commonly used reducing agents are sulfur dioxide, metabisulfite, and ferrous sulfate, but other reducers may also be used. The reducing agents react and form precipitates as shown in Table 8.31c.[5] Equation 8.31(8) describes the reduction reaction with chrome expressed as chromic acid, CrO_3, which has a +6 charge on the chromium. The reducing agent is expressed as sulfurous acid (H_2SO_3), generated by sulfites at low pH. The result is chromium sulfate, $Cr_2(SO_4)_3$, which has a +3 charge on the chromium.

$$2CrO_3 + 3H_2SO_3 \rightarrow Cr_2(SO_4)_3 + 3H_2O \qquad \mathbf{8.31(8)}$$

TABLE 8.31c
Chrome Reduction and Precipitation Reactions

Reducing Agent	*Reaction*
Ferrous sulfate ($FeSO_4$)	$2H_2CrO_4 + 6FeSO_4 + 6H_2SO_4 \rightarrow$ $Cr_2(SO_4)_3 + 3Fe_2(SO_4)_3 + 8H_2O$ $Cr_2(SO_4)_3 + 3Ca(OH)_2 \rightarrow$ $2Cr(OH)_3 + 3CaSO_4$
Sodium metabisulfite ($Na_2S_2O_5$)	$Na_2S_2O_5 + H_2O \rightarrow 2NaHSO_3$ $2H_2CrO_4 + 3NaHSO_3 + 3H_2SO_4 \rightarrow$ $Cr_2(SO_4)_3 + 3NaHSO_4 + 5H_2O$ $Cr_2(SO_4)_3 + 3Ca(OH)_2 \rightarrow$ $2Cr(OH)_3 + 3CaSO_4$
Sulfur dioxide (SO_2)	$SO_2 + H_2O \rightarrow H_2SO_3$ $2H_2CrO_4 + 3H_2SO_3 \rightarrow$ $Cr_2(SO_4)_3 + 5H_2O$ $Cr_2(SO_4)_3 + 3Ca(OH)_2 \rightarrow$ $2Cr(OH)_3 + 3CaSO_4$

Equation 8.31(8) describes the reaction if sulfur dioxide is the reducing agent.

As shown in Figure 8.31d, the first-stage reaction is monitored and controlled by independent control loops: acid addition by pH control and reducing agent addition by ORP control. Acid is added under pH control whenever the pH rises above 2.5. Reducing agent is added under ORP control whenever the ORP rises above approximately +250 mV. (Refer to Section 8.39 on water treatment for other control configurations.)

The ORP titration curve, Figure 8.31e, shows the entire millivolt range that is covered when Cr^{+6} chrome is treated in batches. With continuous treatment, however, operation is maintained in the completely reduced portion of the curve near the nominal +250 mV control point. The exact set point for a particular installation should be at a potential where all the Cr^{+6} has been reduced but without excess sulfite consumption, which is accompanied by the odor of sulfur dioxide.

To complete the chrome reduction reaction takes 10–15 min. The reaction time increases if pH is controlled at higher levels. Variations in pH also affect the measured ORP readings. Therefore, pH must be held constant to achieve consistent ORP control.

Second Stage The wastewater is neutralized to precipitate Cr^{+3} as insoluble chromium hydroxide, $Cr(OH)_3$, and also to the limits for pH, before the treated wastewater can be discharged. Sodium hydroxide or lime ($Ca(OH)_2$) is used to raise the pH to 7.5–8.5, as shown by the reaction in Equation 8.31(9):

$$Cr_2(SO_4)_3 + 6NaOH \rightarrow 3Na_2SO_4 + 2Cr(OH)_3 \qquad \mathbf{8.31(9)}$$

In the second stage, it is more difficult to provide good pH control than in the first, because the control point is closer to the sensitive region of the titration curve near neutrality.

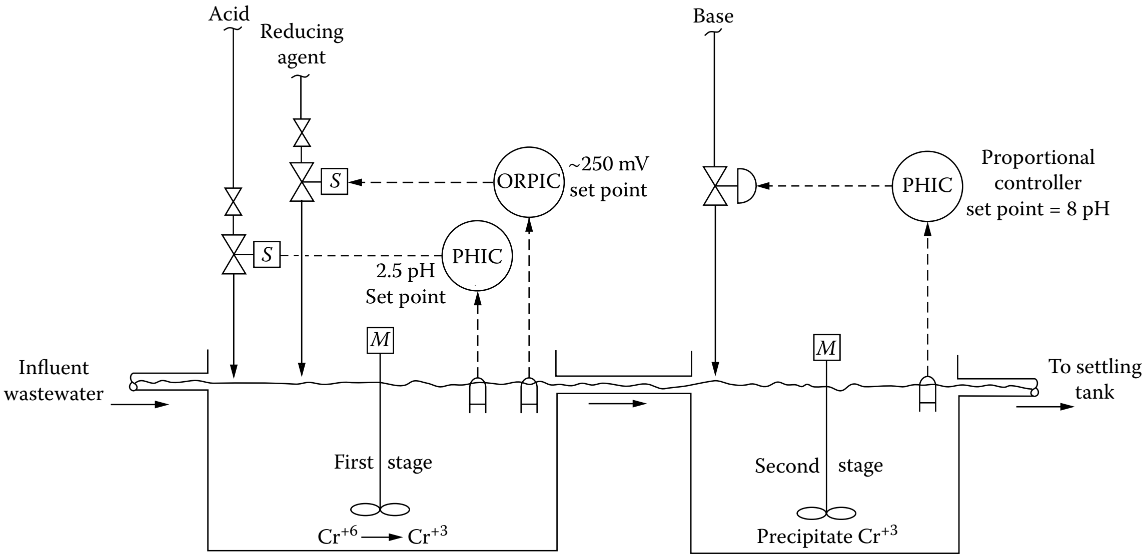


FIG. 8.31d
Continuous chrome treatment.

Although the second-stage reaction is fast, retention time of at least 10 min is usually needed for continuous treatment processes, in order to achieve stable operation. In this stage, the pH controller can be a proportional controller.

A subsequent settling tank or filter removes the suspended chromium hydroxide. Flocculating agents have been found helpful to assist in this separation.

Batch Chrome Treatment Figure 8.31f shows the arrangement for a batch chrome treatment process in which all steps are accomplished in a single tank using a pH and an ORP controller. The steps of the treatment are sequenced, so the pH set point may be changed as needed. In the first stage, acid is added to lower the pH to 2.5, then reducing agent is added to lower ORP to approximately +250 mV.

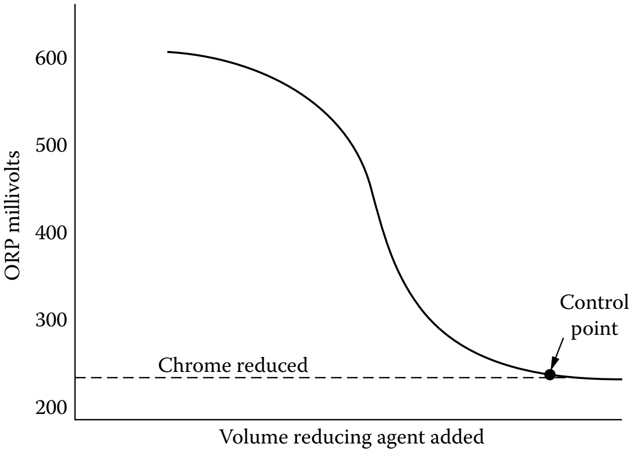


FIG. 8.31e
Chrome reduction titration curve.

After a few minutes have elapsed (ensuring complete reaction) and after a grab sample test for Cr^{+6} has been made, basic reagent is added in the second stage to raise pH to 8. A settling period then follows, or the batch is pumped into a separate tank or pond for settling.

Cyanide Waste Treatment

The metal-plating and metal-treating industries produce the largest amounts of cyanide waste. However, other industries also use cyanide compounds as intermediates. Cyanide solutions are used in plating baths for brass, copper, silver, gold, and zinc. The toxic rinse waters and dumps from these operations require destruction of the cyanide before discharge.

The most frequently used technique for cyanide destruction is a one- or two-state chemical treatment process. The first stage raises the pH and oxidizes cyanide to the less-toxic cyanate form. When required, the second stage neutralizes and further oxidizes cyanate to harmless bicarbonate and nitrogen. The neutralization also allows the metals to be precipitated and separated from the effluent.

First Stage Sodium hydroxide (NaOH) is generally used to raise the pH to approximately 11 to promote the oxidation reaction and ensure complete treatment. The oxidizing agent is generally chlorine (Cl_2) or sodium hypochlorite (NaOCl).

Alternately, oxidation reduction of cyanide wastes to less toxic by-products may be also be achieved by using ozone or hydrogen peroxide as oxidizing agents. The two-step chemical oxidation reaction between ozone and cyanide can be written as

$$CN^- + O_3 \rightarrow CNO^- + O_2 \qquad \mathbf{8.31(10)}$$

$$2CNO^- + H_2O + 3O_3 \rightarrow 2HCO_3 + N_2 + 3O_2 \qquad \mathbf{8.31(11)}$$

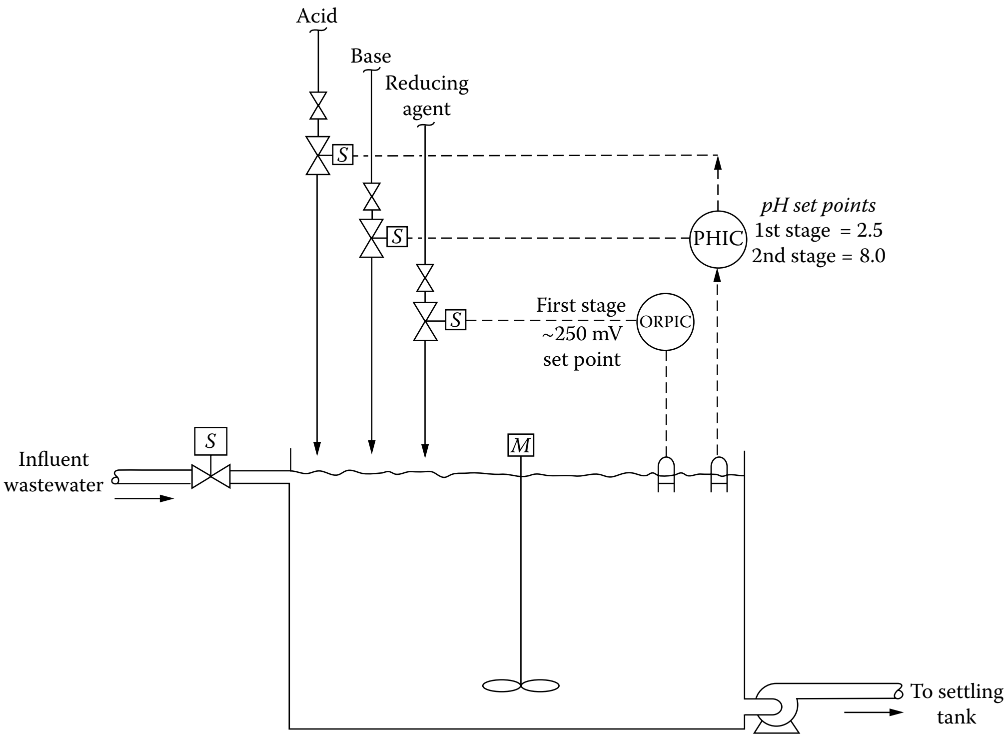


FIG. 8.31f
Batch chrome treatment.

A total ozone dosage of approximately 3–6 O_3/parts per million (ppm) CN is required for near-total cyanide destruction in industrial waste streams.

A one-stage process using hydrogen peroxide and formaldehyde effectively destroys free cyanide and precipitates zinc and cadmium metals in electroplating rinse waters. The chemistry of the destruction of free cyanide cannot be expressed in a simple sequence of reactions, because destruction involves more than one sequence. The monitoring of cyanide rinse water treatment by ORP measurement (using a gold wire electrode) is a useful diagnostic tool for indicating whether the proper quantities of treatment chemicals have been added.

The overall reaction for the first stage using sodium hypochlorite (NaOCl) is given here, with cyanide expressed in ionic form (CN^-), and the result expressed as sodium cyanate (NaCNO) and chloride ion (Cl^-):

$$NaCOCl + CN^- \rightarrow NaCNO + Cl^- \qquad \mathbf{8.31(12)}$$

For the case when the oxidizing agent is chlorine, refer to Section 8.39.

As shown in Figure 8.31g, the first-stage reduction is monitored and controlled by independent control loops: base addition by pH control and oxidizing agent addition by ORP control. The pH controller adds base whenever the pH falls below 11. The ORP controller adds oxidizing agent whenever the ORP falls below approximately +450 mV. A slightly different design is described in Section 8.39.

The ORP titration curve, Figure 8.31h, shows the entire millivolt rage that is covered when cyanide is treated in batches. With continuous treatment, however, operation is maintained in the oxidized, positive region of the curve near the +450 mV set point. The exact set point is determined empirically by measuring the potential when all the cyanide has been oxidized but no excess reagent is present. This point can be verified with a sensitive colorimetric test.

In this reaction, pH has a strong inverse effect on the ORP. Thus, pH must be closely controlled to achieve consistent ORP control, especially if hypochlorite is used as the oxidizing agent. Hypochlorite addition raises pH, which if left unchecked will lower the ORP, calling for additional hypochlorite and causing a runaway situation. To protect against this, the set point of the pH controller should be above the pH level at which hypochlorite has an influence. It is also necessary to move the ORP electrodes away from the hypochlorite addition point to prevent such interactions.

Gold ORP electrodes have been found to give more reliable measurement than platinum for this application.[6] Platinum may catalyze some additional reactions at its surface and is more subject to coating than is gold. The solubility of gold in cyanide solutions does not present a problem, because it is in contact primarily with cyanate. Any slight loss of gold actually serves to keep the electrode clean.

Second Stage In this stage, the wastewater is neutralized to promote additional oxidation as well as to meet discharge

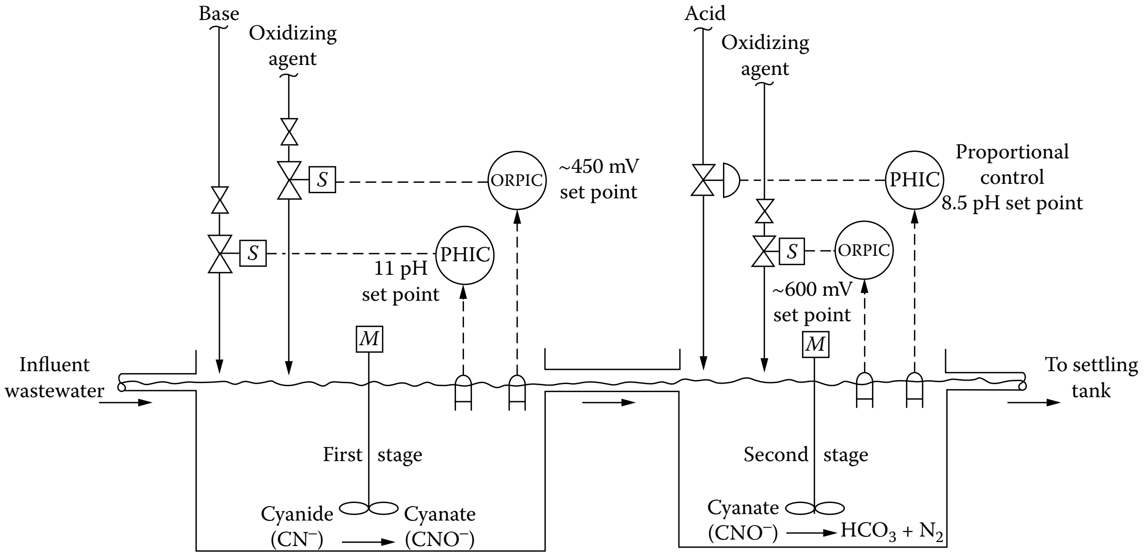


FIG. 8.31g
Continuous cyanide treatment.

pH limits. If the lowering of the pH is not required, that loop can be eliminated. Sulfuric acid is typically used to lower the pH to approximately 8.5, where the second oxidation occurs more rapidly. Acid addition must have a fail-safe design, because below neutrality (pH = 7) highly toxic hydrogen cyanide can be generated if the first-stage oxidation has not been completed.

Hypochlorite is added either in proportion to that added in the first stage or by separate ORP control to complete the oxidation to sodium bicarbonate ($NaHCO_3$), as shown by Equation 8.31(13):

$$2NaCNO + 3NaOCl + H_2O \rightarrow 2NaHCO_3 + N_2 + 3NaCl \qquad \textbf{8.31(13)}$$

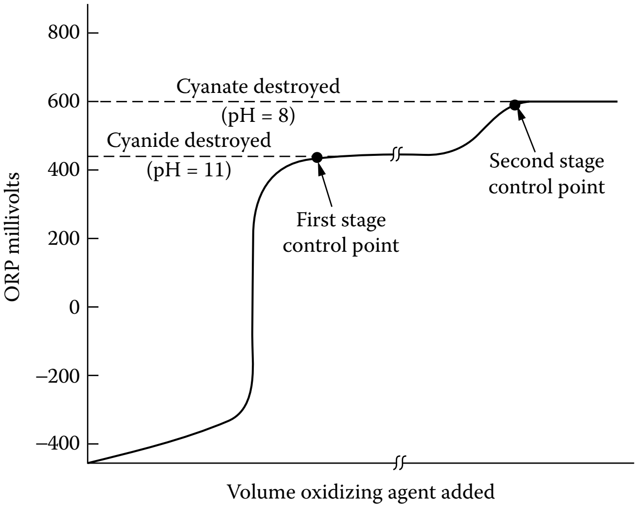


FIG. 8.31h
Cyanide oxidation titration curve.

ORP control in the second stage is very similar to that in the first, except that the control point is near +600 mV. In the second stage, pH control is more difficult than in the first, because the control point is closer to the sensitive region of neutrality. The pH controller can be proportional only.

A subsequent settling tank or filter can remove suspended metal hydroxides, although further treatment may be required.

Batch Cyanide Treatment Figure 8.31i shows the arrangement for batch cyanide treatment with all steps accomplished in a single tank, which is provided with one pH and one ORP controller. (Other control configurations are described in Section 8.39.)

In this control system, the steps are sequenced, changing the pH and ORP set points to obtain the required treatment, with the added assurance that treatment is complete before going on to the next step. Caustic is added to raise the pH to 11. Hypochlorite is added to raise the ORP to approximately +450 mV, simultaneously adding more caustic, as required, to maintain a pH of 11.

An interlock must be provided to prevent acid addition before the completion of oxidation of all cyanide to cyanate. Then acid can be added to neutralize the batch and further hypochlorite oxidation completes the cyanate-to-bicarbonate conversion. A settling period can be used to remove solids, or the batch can be pumped to another tank or pond for settling.

Sodium Hypochlorite Production

Sodium hypochlorite (NaOCl) is produced by reacting chlorine gas (Cl_2) and dilute sodium hydroxide (NaOH). Sodium hypochlorite is used both as an industrial and as a domestic

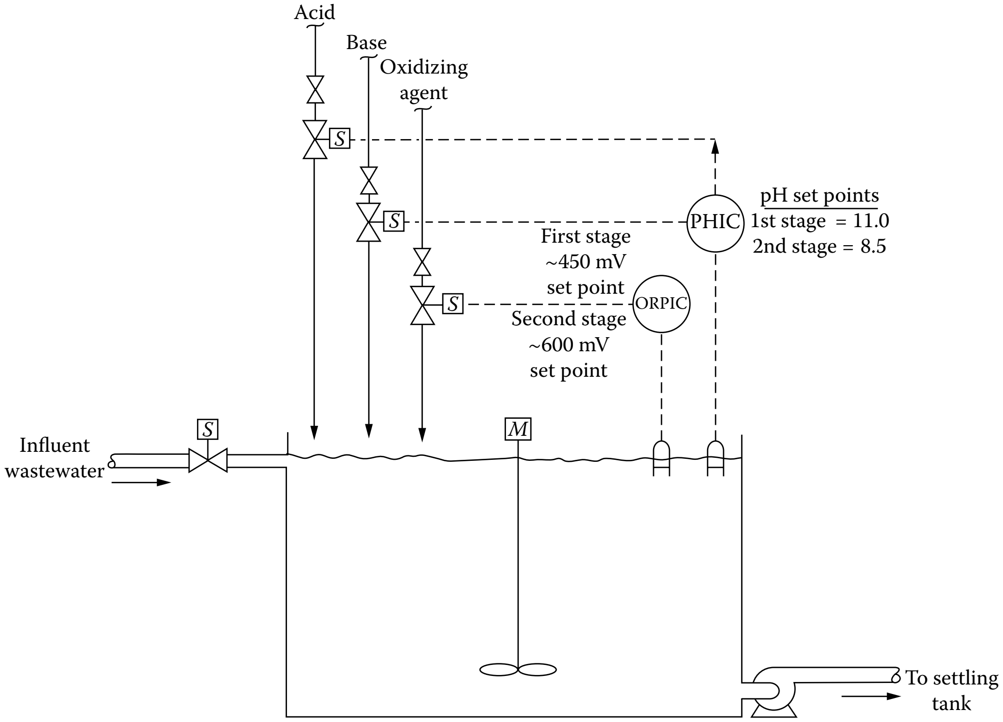


FIG. 8.31i
Batch cyanide treatment.

bleaching agent. For industrial use, it is normally produced on-site and has an available chlorine strength of 12–15%. For domestic use, as household chlorine bleach, the strength is typically 3–6% available chlorine.

ORP is used as a measure of the available chlorine in the final product. The control system of the NaOCl production process is shown in Figure 8.31j. Caustic soda solution (sodium hydroxide) at 5% concentration is admitted on level

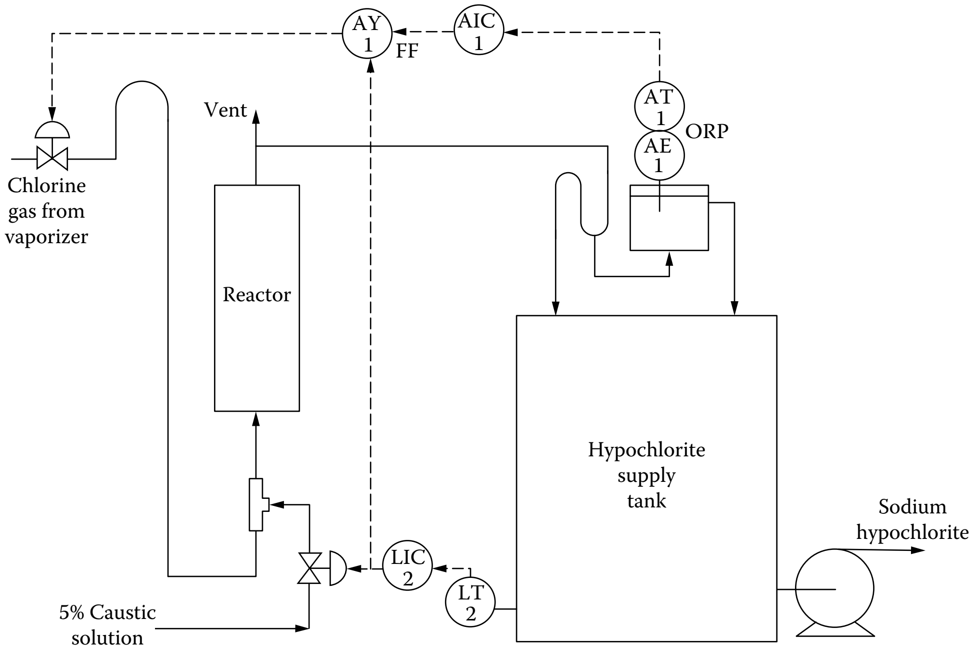


FIG. 8.31j
Control system for the continuous production of sodium hypochlorite.

control. This supply of caustic soda is normally diluted down from 50 to 5% concentration in two steps (not shown). The ORP controller throttles the ratio of the chlorine gas that is added to the flow of 5% caustic solution. The set point for the ORP controller will be somewhere between 500 and 750 mV, depending upon the initial caustic strength and the desired ratio of available chlorine to excess caustic.

On-site sodium hypochlorite production is common in pulp and paper mills, where it is used in the pulp bleaching process. However, the industry trend has been to move away from chlorine bleaching (elemental chlorine or sodium hypochlorite), in favor of chlorine dioxide, hydrogen peroxide, oxygen, and other bleaching agents, which are considered to be preferred from a pollution point of view.

Paper Pulp Bleaching

The pulp and paper industry uses strong oxidizing agents (commonly chlorine, chlorine dioxide, hydrogen peroxide, oxygen, and ozone) for bleaching and delignification of pulp. ORP has been applied both to control the pulp bleaching and the delignification process and as emission controls for chlorine (Cl_2) and chlorine dioxide (ClO_2) gases.

Bleaching and Delignification Control Bleaching and delignification are accomplished in several stages, with the residues of the bleaching operation extracted from the pulp in washers located between the stages. A traditional bleach plant sequence is chlorination ("*C*"), followed by caustic extraction ("*E*"), followed by hypochlorite bleaching ("*H*"), sometimes followed by caustic extraction ("*E*"), and finally, by treatment with chlorine dioxide ("*D*").

To eliminate chlorine, the CEHED or CEHD sequences have in many cases been replaced with DEDED, DED, or other variants that eliminate the *C* and *H* stages. ORP measurement can be used in the pulp discharge vat of the washer following the *D* stage to ensure the chlorine dioxide has been sufficiently rinsed from the pulp. Applying it to control actual chemical addition is more difficult.

Because a variety of organic compounds are oxidized, it is not possible to write exact equations for the pulp bleaching process. The critical control parameters for the bleaching and delignification process are brightness, kappa number (an indication of lignin remaining, or pulp purity), chemical residual, pH, temperature, pulp consistency, pulp flowrate, chemical flowrate, and retention time (determined by tower size and flowrate).

ORP can be used as an indication of the chemical residual. It is more apt to be useful for the measurement of lower concentrations of residuals, such as in the final stages or in the washers. Because ORP is so nonlinear over the potentially wide range of chemical concentrations in pulp bleaching, and because it can saturate at high residual levels, the preferred method of measurement has become the polarographic technique. Sensors based on polarography are now highly evolved for this application (see Sections 8.4, 8.42, and 8.43 in Chapter 8 in the first volume of this handbook).

Polarographic sensors can be located in the line a relatively short distance downstream from the point of chemical addition to rapidly measure how the reaction is proceeding and to be used in closed-loop control. While, in principle, ORP measurement could also be used, it has been found to be inadequate, because the ORP and reference electrodes require excessive maintenance in the harsh environment.

When ORP is used to control the dosage of bleaching (chemical flow), it is best used in combination with on-line brightness, pH, temperature, pulp consistency and flow, and chemical flow. Brightness, residual (ORP or polarographic), and chemical flow can be used to calculate a "compensated brightness" used to actually control chemical dosage (pH or consistency can also be part of the compensation). Controlling dosage on residual alone, as measured by ORP or a polarographic instrument, will tend to overbleach for certain process changes.[7]

It is also often necessary to independently control pH, depending upon the type of bleaching stage. In a ClO_2 bleaching stage, for example, pH must be kept below 3.5 for maximum effectiveness.

Chlorine/Chlorine Dioxide Scrubber Control A natural consequence of treating the pulp with strong oxidizing chemicals is the emission of noxious gases. Wet scrubbers are commonly used to "scrub" Cl_2 and ClO_2 from process gases prior to venting them to the environment. The most effective scrubbing media for this application are the white liquor or the weak wash, both of which contain large amounts of sodium hydroxide (NaOH) and sodium sulfide (Na_2S).[8]

The Cl_2 and ClO_2 gases rapidly oxidize the sulfides in the white liquor or weak wash to form chlorite (ClO_2^-) and chloride (Cl^-). A number of concurrent reactions are possible, the favored ones being:

$$8ClO_2 + 5Na_2S + 8NaOH \rightarrow 5Na_2SO_4 + 8NaCl + 4H_2O \quad \mathbf{8.31(14)}$$

$$4Cl_2 + Na_2S + 8NaOH \rightarrow Na_2SO_4 + 8NaCl + 4H_2O \quad \mathbf{8.31(15)}$$

The scrubber consists of a vertical packed tower. The scrubbing liquid is continuously circulated from the bottom to spray nozzles at the top (Figure 8.31k). Fresh scrubbing make-up solution is added under ORP control to maintain the required excess of reducing agent in the circulated liquid. A continuous overflow provides for changeover of liquid in the system.

The purpose of the double-headed positive displacement-type metering pump is to supply the required amount of scrubbing make-up solution to the scrubber, at the right concentration. The required dilution of the weak wash or white liquor is performed by manually adjusting the stroke of each head, thereby setting a constant ratio between these fluids and dilution water (usually condensate).

If there was no excess reductant remaining in the recirculated scrubber fluid, the concentrations of Cl_2 and ClO_2

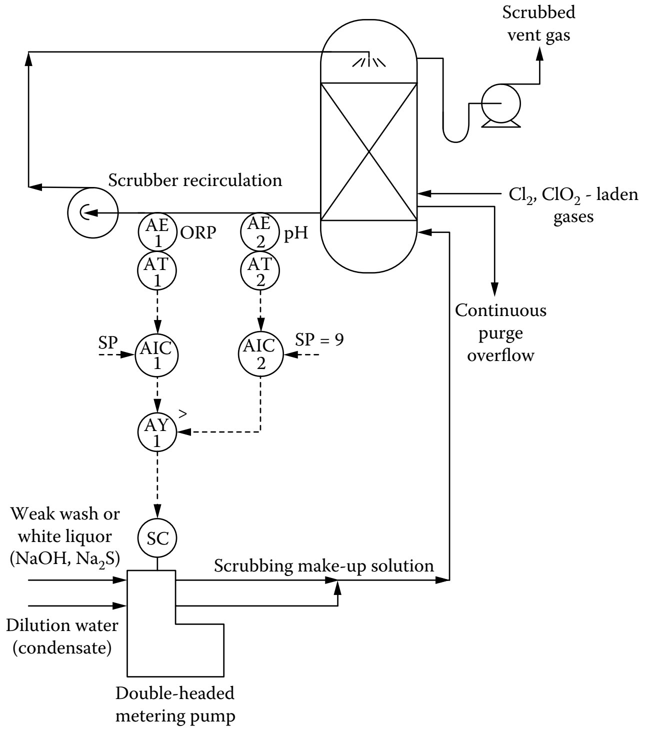


FIG. 8.31k
The control system for a chlorine/chlorine dioxide scrubber.

emissions in the vent gas will rise sharply. The object of the ORP control loop is to keep just enough excess reductant in the scrubbing fluid. The ORP set point would be in the reducing range, somewhere between –100 and –500 mV. For this process it is necessary to run emission tests and vary the ORP by varying fresh make-up to determine the operating point that results in design removal efficiency with only minimal excess chemical.

In scrubbers that use white liquor or weak wash that contains sodium sulfide (Na_2S), pH must be maintained above 7, because otherwise sulfur compounds will tend to form and deposit in the scrubber, increasing chemical consumption, decreasing removal efficiency, and causing plugging. Further, for liquor or weak wash containing sodium hydrosulfide (NaHS), pH must be maintained above 8.5 to avoid the formation of hydrogen sulfide (H_2S) gas.[8]

Consequently, the control system shown in Figure 8.31k includes a pH-based override controller. The output signal of this controller, through a high-signal selector (AY-1), overrides the ORP controller to ensure that enough white liquor or weak wash is added to keep the alkalinity at a pH of 9.0.

Other ORP Control Systems

There are numerous other industrial processes where it is important and useful to know the extent to which an oxidation-reduction reaction has proceeded or the extent to which an oxidizing or reducing environment is being maintained. Treatment of process cooling tower water and the indigo dyeing process (textiles) are two examples.

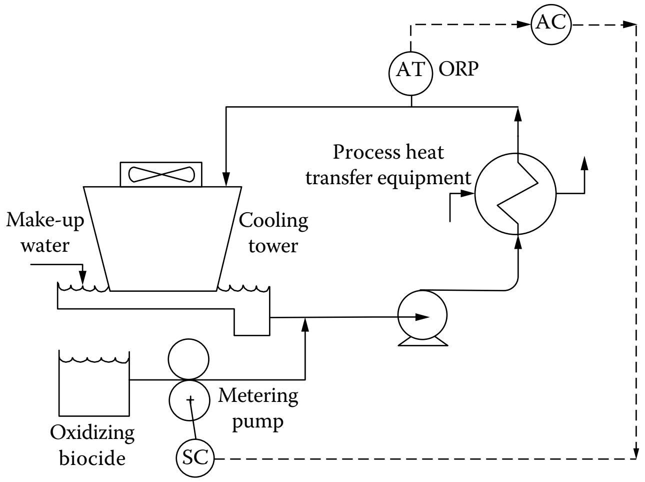


FIG. 8.31l
ORP-based control system for the control of biological growth in cooling towers.

Microbial Control in Cooling Towers Biocides are frequently used in cooling towers to control the growth of algae, bacteria, fungi, barnacles, and even clams and mussels in the water. Typical oxidizing microbiocides for this application are chlorine (Cl_2), bromine (Br_2), sodium hypochlorite (NaOCl), chlorine dioxide (ClO_2), and ozone (O_3). ORP-based continuous control of the addition of biocide helps to maintain effective treatment without wasting chemicals (Figure 8.31l). The desired range of ORP control is typically between 550 and 650 mV.

Indigo Dye Process[9] The process of dyeing cotton and cellulose fibers with indigo dye requires several additives to make the dye soluble in water. Sodium dithionite ($Na_2S_2O_4$) is added to the bath to reduce the dye (*leuco* form), and sodium hydroxide (NaOH) is added to form phenolates, which are water-soluble.

After the dye has been applied, it is reoxidized, fixing it into the fabric and making it once more insoluble (this can be done just by exposing it to air to dry). Control of the pH can be used to vary the behavior of the dye application (e.g., the degree of penetration) and, therefore, the final effect.

ORP is maintained in the range of –760 to –860 mV, to keep the indigo dye in its reduced (leuco) form for solubility. ORP can also be adjusted in and around this range to affect the final shade of the dyed fabric.

Swimming Pool and Spa Treatment ORP can provide a measure of sanitizer activity and water quality in swimming pools and spa water. In the United States many public pools and spas use ORP controllers for the automatic control of chlorine addition, but the controller readouts are usually labeled in parts per million of free chlorine instead of in units of ORP (mV).

When chlorine is introduced into the pool or spa water, it forms the active form of free chlorine, hypochlorous acid (HOCl), which is an excellent bactericide. The process of chlorine addition to water is described by the following equilibrium equation:

$$Cl_2(g) + H_2O \leftrightarrow HOCl(aq) + H^+(aq) + Cl^-(aq) \qquad \mathbf{8.31(16)}$$

The position of the equilibrium in the reaction shown in Equation 8.31(16) strongly depends upon the pH and the other reactions going on in the process of oxidizing and disinfecting. The HOCl is a fast-acting sanitizer (weak acid). Its concentration decreases very rapidly with increasing pH in the range of interest for pools and spas (pH 7–8).

For good bacteriological quality, it is essential to maintain a proper HOCl level in the water at all times. The recommended minimum ORP level should be controlled at between 650 and 750 mV. The ORP standards (650 or 750 mV) should apply to all sanitizers, including all forms of chlorine (with or without stabilizer) and bromine, as well as to systems using ozone or other sterilization methods.

The recommended chlorine control parameters are given in Table 8.31m.

Chemical controllers used in the pool and spa industry normally utilize an ORP sensor to monitor the sanitizer concentration, as well as a pH sensor to monitor the pH. The controller automatically turns the appropriate chemical feeders on and off as required to maintain the proper sanitizer and pH levels. This results in good water quality and elimination of chloramines and other undesirable by-products, as well as in savings in both chemical consumption and labor.

Cyanuric acid, used to stabilize chlorine against degradation by sunlight in outdoor pools, will tend to reduce the ORP (and, therefore, the chlorine effectiveness) at a given pH and free chlorine residual level. Above 50 ppm cyanuric acid, ORP controllers lose effectiveness, because additional chlorine called for by the ORP controller has a diminishing ability to increase free chlorine residual and, therefore, ORP. At about 70 ppm cyanuric acid, additional chlorine is reported to have no effect, and ORP control is rendered useless.[10]

TABLE 8.31m
Basic Control Parameters for the Treatment of Pool and Spa Waters

Parameters	*Recommended Control Levels*
pH	7.4–7.8
Free chlorine residual	Minimum free chlorine residual 0.4 ppm
Total chlorine residual	Maximum total chlorine residual should not exceed free chlorine by more than 0.5 ppm
ORP control standard	650 or 750 mV
Total alkalinity*	80–150 ppm
Calcium hardness	Above 140 ppm

**Note:* Total alkalinity times calcium hardness must equal 25,000–30,000 (this rule works when pH is 7.4–7.6 and temperature is 78–85°F).

ORP control can also be used for changing other oxidizers used in pool and spa water sanitation, such as bromine and ozone. Automated pool and spa water treatment systems represent a major advance in pool operation and maintenance. ORP standards for pool and spa sanitation have been recognized by the World Health Organization (WHO), the Centers for Disease Control and Prevention (CDC), the National Spa and Pool Association (NSPA), and a number of state and local health departments.[11]

CONCLUSIONS

ORP control is useful when the process to be controlled involves reduction and oxidation (redox) reactions. ORP measurement is best used to indicate the extent to which expected reactions have proceeded and to detect the relative strength of the oxidizing and reducing chemicals in the solution being measured.

Successful ORP measurement and control requires an understanding of the ORP measurement principles and of the chemistry that governs the process. ORP control can be highly nonlinear, and therefore the same techniques and strategies that are used in controlling highly nonlinear pH processes (Section 8.32) can also be used for ORP control.

References

1. Masterton, W. L., and Slowinski, E. J., *Chemical Principles,* 4th edition, Philadelphia, PA: W.B. Saunders, 1977.
2. Latimer, W. M., *Oxidation Potentials,* New York: Prentice Hall, 1952, Chapter 1.
3. Light, T. S., “Standard Solution for Redox Potential Measurements,” *Analytical Chemistry,* 44:6, pp. 1038–1039, May 1972.
4. *pH Controllability,* North Wales, PA: Leeds & Northrup, Application Bulletin C2.0001, October 1975.
5. Lipták, B. G., *Water Pollution Environmental Engineers' Handbook,* Vol. 1, Radnor, PA: Chilton, 1974.
6. Shinskey, F. G., *pH and pIon Control in Process and Waste Streams,* New York: John Wiley & Sons, 1973, p. 120.
7. “TAPPI Bleach Plant Operations Short Course,” course notes, Orange Beach, AL: Technical Association of the Pulp and Paper Industry (TAPPI), 1994.
8. Jain, A. K., and Dallons, V. J., “Technical Bulletin No. 616, Bleach Plant Chlorine and Chlorine Dioxide Emissions and Their Control,” New York: National Council for Air and Stream Improvements, (NCASI), 1991.
9. Rosemount Analytical, “pH and ORP Monitoring for the Indigo Dye Process,” Application Data Sheet ADS2820-03/Rev.A, Emerson Process Management, 2002.
10. Williams, K., “Cyanurics: Benefactor or Bomb,” *Pumproom Press,* Professional Pool Operators of America, 1997.
11. Steininger, J. M., “ORP Control in Pools and Spas,” Santa Barbara Control Systems, 1998.

Bibliography

Benefield, L. D., Judkins, J. F., and Weand, B. L., *Prosess Chemistry for Water and Wastewater Treatment,* Englewood Cliffs, NJ: Prentice Hall, 1982, Chapter 13.

Cali, G. V., and Galetti, B. J., "Plating Waste Control Instrumentation," *Pollution Engineering,* March 1976, pp. 48–50.

Chamberlin, N. S., and Snyder, H. B., "Technology of Treating Plating Wastes," Tenth Industrial Waste Conference, Purdue University, May 1955.

Devlin, P. M., Wang, H. J., Winchell, C. J., Day, S. G., Zura, R. D., and Edlich, R. F., "Automated Hydrotherapy Pool Water Treatment System," *Hydrotherapy Pool Management,* Vol. 10, No. 1, January/February 1989.

Eckenfelder, W. W., Jr., *Industrial Water Pollution Control,* 2nd edition, New York: McGraw-Hill, 1989.

Greer, W. N., "Measurement and Automatic Control of Etching Strengths of Ferric Chloride," *Plating,* October 1961.

Henkel Corporation, Energy Group, "Ozone Treatment of Cyanide Wastes," Data Sheet 941C, June 1989.

Kemmer, F. N., *The NALCO Water Handbook,* 2nd edition, New York: McGraw-Hill, 1988.

Lanouette, K. H., "Heavy Metals Removal," *Chemical Engineering,* October 1977, pp. 73–80.

Lavigne, J. R., *Instrumentation Applications for the Pulp and Paper Industry*, California, Miller-Freeman, 1979.

Lawes, B. C., "A Peroxygen System for Destroying Cyanide in Zinc and Cadmium Electroplating Rinse Waters," *Plating,* September 1973.

Leeds & Northrup, Reference Binder 5, Environmental Sciences, "Oxidation-Reduction (REDOX) Potentials," June 1969.

Masterton, W. L., and Slowinski, E. J., *Chemical Principles,* 4th Edition, Philadelphia, PA: W.B. Saunders, 1977, Chapters 22, 23.

Mattock, G., "Automatic Control in Effluent Treatment," *Transactions of the Society of Instrument Technology,* December 1964, pp. 1973–1989.

McMillan, G. K, "Methods of Controlling pH," *Chemical Processing,* July 1997.

Record, R. G. H., "The Use of Redox Potentials in Chemical Process Control," *Instrument Engineer,* October 1965, pp. 65–75, continued April 1966, pp. 95–102.

Rosemount Analytical, "Fundamentals of ORP Measurement," Application Data Sheet ADS43-014, Emerson Process Management, 2001.

Ross, M., "Treating Water at ORP Speed," *Water Technology,* May 1996.

Shinskey, F. G., *pH and pIon Control in Process and Waste Streams,* New York: John Wiley & Sons, 1973, Chapters 1, 5.

Shinskey, F. G., *Problem Solving Software*, Ver. 3.2, Foxboro, MA: The Foxboro Company (Invensys Systems), 1998.

Shinskey, F. G., *Process Control Systems*, 4th edition, New York: McGraw-Hill, 1996, p. 93.

Shinskey, F. G., "Sharpening pH Control," *Control,* July 1998.

Steininger, J. M., "Clean Up with ORP," *The Business Magazine for the Spa & Pool Professional,* June 1990.

Steininger, J. M., "PPM or ORP: Which Should be Used?" *Swimming Pool & SPA Merchandiser,* November 1985.

Weast, R. C., *Handbook of Chemistry and Physics,* 50th edition, Cleveland, OH: The Chemical Rubber Company, 1969.

Yorgey, W. B., "Updating a Wastewater Treatment Plant with Automatic Controls," *Plant Engineering,* February 1973, pp. 1978–1982.

8.32 pH Control

D. L. HOYLE (1985) **G. K. McMILLAN** (1995) **F. G. SHINSKEY** (2005)

Final Element Rangeability:	The rangeability requirement is extraordinary and depends upon the titration curve and on influent flow variability. Metering pumps are capable of 20:1 to 200:1, valves with positioners are capable of 50:1 or more, and a pair of split-ranged valves is capable of 1000:1 or more. For unbuffered titration, the pH variation (on either side of neutral) that valves can accommodate can be determined as the base 10 logarithm of the rangeability. For example, if rangeability is 1000:1 and the controllable pH variation is 3, it can be from pH 3–6 or 12–9.
Final Element Precision and Characteristics:	Requirement is exceptional and depends upon titration curve and desired control band. Electronically set metering pumps and valves with positioners have repeatability from 0.1 to 2.0%. Linear valve characteristics are generally preferred, and the use of smart digital positioners is recommended.
Control Loop Dynamics:	The effect of the extreme nonlinearity and sensitivity of a pH process is diminished by minimizing loop dynamic gain. To dampen oscillations, one needs ratios of loop dead time to time constant of less than 0.05, when controlling on the steep portion of the titration curve. Reagent delivery delay is often the largest source of loop dead time.
Design Considerations:	For proper mixing, process requires premixing of reagent with influent at the entrance of a well-mixed vessel. Close-coupled control valve with loop seal at the point of injection or water flush is needed to minimize reagent delivery delay. If the pH of the influent is more than three pH units away from the control band, or whenever the requirement for the final-element rangeability or sensitivity exceeds 1000:1, an additional neutralization stage is needed.
Reaction Tanks:	Liquid depth should equal diameter, retention time should not be less than 5 min (10–30 min for lime), and dead time should be less than 1/20th of the retention time. For strong acid-strong base neutralizations, one, two, or three tanks are recommended for influent pH limits of 4 to 10, 2 to 12, and 0 to 14, respectively. Influent should enter at top, and effluent should leave at the bottom.
Agitator Choices:	For tanks under 1000 gal (3780 1), use propeller-type agitators, and for larger volumes, axial-flow impellers are preferred. Flat-bladed radial flow impellers should be avoided. Acceptable impeller-to-tank diameter ratio is from 0.25 to 0.4. Peripheral speeds of 12 ft/s (3.6 m/s) for large tanks and 25 ft/s (7.5 m/s) for tanks with volumes less than 1000 gal (3780 1) are acceptable.
Mixing Equipment:	In-line mixer residence time should be less than 10 sec, and a well-mixed vessel residence time should be greater than 5 min. The vessel agitator should provide both a pumping rate greater than 20 times the throughput flow and a vessel turnover time of less than 1 min. Solid and gas reagents require a residence time greater than 20 times the batch dissolution time. Gas reagent injection needs a sparger designed to reduce bubble size and improve bubble distribution.
Set Point Location:	If the set point is located on the flat portion of the titration curve, this reduces the oscillation of the pH process. It also reduces control sensitivity and the requirement for control valve precision.

PH Sensor Location:	Insertion assemblies in pumped recirculation lines are preferred, because they can provide increased speed of response, decreased coating, improved accessibility, and auto on-line washing and calibration. If immersion sensor is used, it must be located in the mixed zone, directly before the discharge.
Continuous Control Techniques:	Nonlinear characterization is recommended for the feedback controller; self-tuning is helpful when the titration curve varies widely. Flow-based feedforward is effective when rapid upsets in flow are expected, while pH-based feedforward is rarely used, because it is effective only for neutralization applications having a single titration curve.

INTRODUCTION

This section begins with an explanation of the difficult nature of the pH process and of the various titration curves. Next, the process equipment used in neutralization facilities is described, including such topics as the selection of reagent delivery systems, mixing equipment, tank sizing, and other considerations. This is followed by a discussion of the special requirements of the pH controller and its tuning. The section is finished with an examination of the role of feedforward control.

TITRATION CURVES

The difficulty in controlling pH stems entirely from its nonlinear relationship to acid and base concentration in a solution. The slope of a single titration curve of pH vs. reagent delivery can vary over several orders of magnitude, and the set point where control is exercised is usually positioned in the steepest region of the curve. Because of this dominance, the nature of the titration curve will be the first topic to be considered in this section.

pH is the negative base-10 logarithm of the hydrogen-ion activity in g-ions/l in a solution. Hydrogen-ion *activity* is not quite the same as its *concentration,* but in dilute solutions, where pH is mostly controlled—especially wastewater—their difference is negligible.[1] For our purposes, then, the relationship will be expressed in terms of concentration, which appears in brackets:

$$\text{pH} = -\log[\text{H}^+] \quad [\text{H}^+] = 10^{-\text{pH}} \qquad \textbf{8.32(1)}$$

Water has a neutral pH of 7 at 25°C, at which point hydrogen and hydroxyl ions are in balance at 10^{-7} g-ions/l each:

$$[\text{H}^+][\text{OH}^-] = K_w = 10^{-14} \qquad \textbf{8.32(2)}$$

where K_w is the equilibrium constant of water. (It does change from 10^{-15} at 0°C to 10^{-12} at 100°C.) A tenfold increase in $[\text{H}^+]$ would lower the pH to 6, and a tenfold decrease would raise it to 8, where $[\text{OH}^-]$ would be at a concentration of 10^{-6} g-ion/l. One g-ion/l is considered a *Normal* solution and is designated 1.0 *N.* Control can be exercised over a pH of 0–7 in the acidic range and 7–14 in the basic range, or over a concentration range of 10^{-7} to 1.0 *N*:

Strong Agents

A strong agent is one that completely ionizes in aqueous solution. Two common strong agents are hydrogen chloride (HCl) and sodium hydroxide or caustic soda (NaOH). When dissolved in water, all of the H in HCl acts as the ion H^+, and all of the OH in NaOH acts as the OH^- ion. It is the ions that are reactive and are sensed by the pH electrode. A solution of 10^{-3} *N* HCl will then read pH 3, and a solution of 10^{-3} *N* NaOH will read pH 11. In any solution, the negative and positive charges must balance—for example,

$$[\text{Na}^+] + [\text{H}^+] = [\text{Cl}^-] + [\text{OH}^-] \qquad \textbf{8.32(3)}$$

Consider a neutralization reaction between these two agents, where HCl is added to water in a concentration of x_A and caustic to a concentration of x_B. All of the Cl^- comes from the HCl, and all of the Na^+ comes from the NaOH, so that the above charge balance becomes

$$x_B + [\text{H}^+] = x_A + [\text{OH}^-] \qquad \textbf{8.32(4)}$$

When the charge balance is combined with equilibrium Equation 8.32(2), the result is the strong-agent titration curve:

$$x_B - x_A = \frac{10^{-14}}{[\text{H}^+]} - [\text{H}^+] = 10^{\text{pH}-14} - 10^{-\text{pH}} \qquad \textbf{8.32(5)}$$

This relationship is plotted in Figure 8.32a, with the outer curve plotted against the bottom scale. Although the curve appears to be linear between pH 4 and 10, it is not— the inner curve, plotted against the expanded top scale, continues the logarithmic shape.

This is the titration curve when a strong acid is neutralized by a strong base; if a strong base were neutralized by a strong acid, the *x*-axis would be reversed to $x_A - x_B$, and the pH would fall. Another strong acid is nitric; sulfuric acid, while very corrosive, is not as strong, but may be considered so in the neutral pH range. Other strong bases are the hydroxides of potassium and lithium; calcium hydroxide is not as strong, and its pH is limited to about 12 by its low solubility, but in the neutral range it is essentially completely ionized.

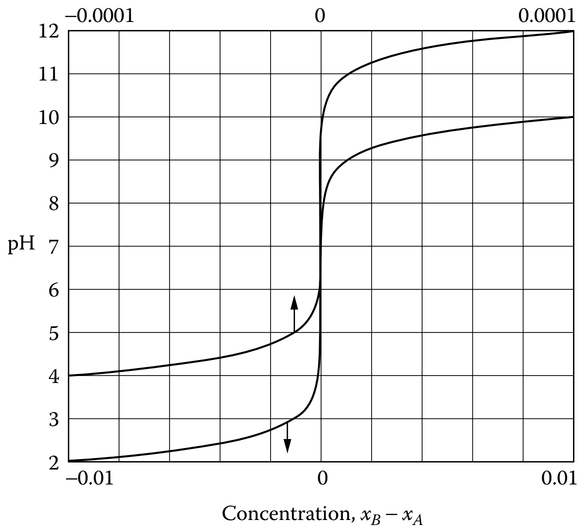


FIG. 8.32a
The strong-agent titration curve has the greatest gain variation of all.

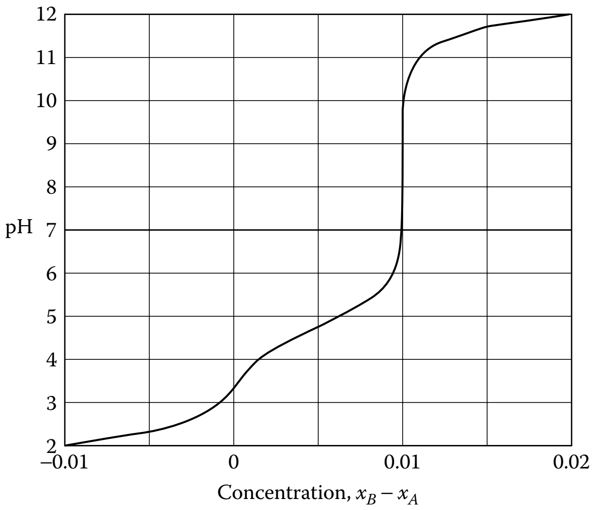


FIG. 8.32b
Acetic acid buffers the titration curve at about pH 4.75.

The difficulty the titration curve in Figure 8.32a presents to the pH controller can be estimated by comparing the size of the control target to the distance that the influent pH is from the target. For example, suppose the target specification that a wastewater must meet is a 6–8 pH and the influent entering the treatment plant is at pH 4.

This requires matching an influent concentration of 10^{-4} *N* with an accuracy of $\pm 10^{-6}$ *N*, or one part in 100 — difficult, but not impossible. However, if the influent pH were 2, then 100 times as much caustic would be required, but it must still be delivered with the same absolute precision, which amounts to control of 1 part in 10,000.

This level of accuracy is beyond the capability of controllers and valves especially, and so in the absence of buffering, multistage treatment would be required. Fortunately, most wastewater streams are buffered to some extent, and some are heavily buffered.

Buffering

Most acids and bases are weak and, therefore, only partially ionized. And most wastewater contains a small amount of carbon dioxide, which is also a weak acid, measured as alkalinity. Its presence greatly moderates the task of controlling pH in the neutral range. In fact, the bicarbonate ion is used as a buffer for stomach acid, as it tends to regulate pH around 7. Acids may have as many as three hydrogen atoms in their molecule, and they tend to be released as ions in different pH ranges. If a weak acid has but a single hydrogen atom, it affects the titration curve as follows:

$$x_B - x_A = 10^{\mathrm{pH}-14} - 10^{-\mathrm{pH}} + \frac{x_C}{1 + 10^{\mathrm{p}K_C - \mathrm{pH}}} \qquad \textbf{8.32(6)}$$

where x_C is the concentration of the weak acid and K_C is its ionization constant; pK_C is the negative base-10 logarithm of K_C and corresponds to the pH value where half of the acid is ionized. For example, the pK_C for acetic acid (vinegar) is 4.75. In this pH range, the titration curve is buffered by the weak acid, as shown in Figure 8.32b.

The curve was plotted for a concentration of $x_C = 0.01$ *N*. At that concentration of $x_B - x_A$, the weak acid is neutralized. The curve has been extended in both directions to show the effects of additional strong acid and base, because most wastewater includes both strong and weak agents to be treated.

Ammonia in water forms the weak base ammonium hydroxide, whose buffer point is pH 9.25, which just happens to be 14–4.75. Therefore, its titration curve is the mirror image of acetic acid's. Adding a strong acid to aqueous ammonia will lower the pH gradually through the range of 10.25–8.25, where it will begin to fall sharply and then continue to follow the strong-base curve.

Carbon dioxide in water forms the weak carbonic acid, which releases two hydrogen ions — its titration curve is described by

$$x_B - x_A = 10^{\mathrm{pH}-14} - 10^{-\mathrm{pH}} + \frac{x_C(1 + 0.5 \times 10^{\mathrm{p}K_2 - \mathrm{pH}})}{1 + 10^{\mathrm{p}K_2 - \mathrm{pH}}(1 + 10^{\mathrm{p}K_1 - \mathrm{pH}})}$$

8.32(7)

where $pK_1 = 6.35$ and $pK_2 = 10.25$. Its titration curve is plotted in Figure 8.32c for a concentration of $x_C = 0.0001$ *N*, which is one tenth of the horizontal scale. In other words, the initial pH of 3 is due to the presence of ten times as much strong acid.

This is typical for industrial wastewater, where strong agents need to be neutralized in the presence of much smaller

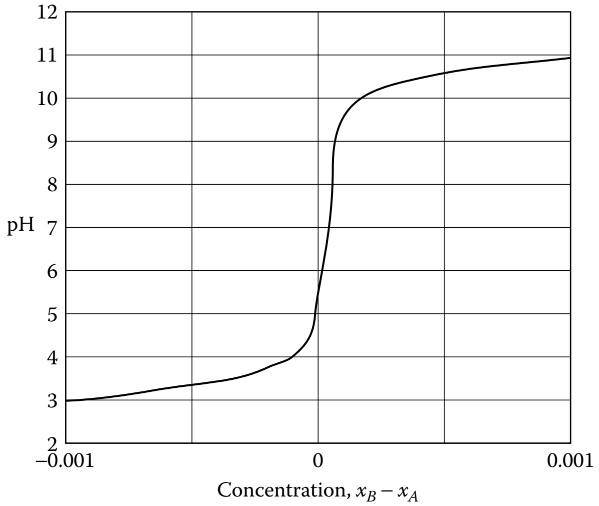


FIG. 8.32c
Carbon dioxide moderates the titration curve between pH 6 and 7.

concentrations of weak agents, principal among them being carbon dioxide. A normality of 0.0001 corresponds to an alkalinity of 5 ppm calcium carbonate, within the common range for surface water.

While the moderation of the curve in the range of pH 6–7 may not seem very noticeable, it is substantial compared with the strong-agent curve in Figure 8.32a. With that curve, neutralization from pH 3 to pH 6–8 required a caustic dosage accurate to ±1 part in 1000. For the curve in Figure 8.32c, the accuracy required is only ±1 part in 28, an improvement by a factor of 36.

Another measure of the effect of buffering is the gain of the titration curve at a set point of pH 7. The reagent valve needs to be sized for twice the anticipated demand. Given a wastewater at pH 3 or 11, the valve must be able to deliver a reagent concentration of 0.002 *N*.

In the absence of any buffering — as in treating distilled or deionized water — the steady-state open-loop gain, assuming a transmitter range of pH 2–12, would be about 420. In other words, a change in valve position of 0.01% could change the pH of the solution by 4.2% of scale, which is 0.42 pH.

The presence of carbon dioxide of as little as 10^{-5} *N*, which is very common in water supplies, lowers the gain to 75; at 0.0001 *N*, which is shown in Figure 8.32c, the gain is 11.4; and at 0.001 *N*, the gain is only 1.6. The lower gain both reduces the sensitivity of the solution pH to disturbances and allows a higher controller gain setting without causing cycling.

The presence of buffers is very common in industrial wastewaters. All cleaning agents are buffers: carbonates, silicates, phosphates, sulfonates, detergents, and surfactants. So are all organic acids and bases such as citric acid and amines. Most metal ions are also buffers, including iron, copper, chromium, cobalt, nickel, tin, zinc, cadmium, and aluminum. As a result, the pH of wastewater from metal-plating operations is very easy to control; regulation at pH 8–9 causes these metals to precipitate from solution as hydroxides.[2]

If the *principal* load on the pH loop is represented by a buffer, then variation in its concentration changes the gain of the curve and the load at the same time. An equal-percentage reagent valve, whose gain increases in direct proportion to flow delivered, can compensate for the gain reduction caused by increasing concentrations of the buffer.

This situation is unusual, however. The principal problem with the presence of buffers in wastewater is that they tend to come and go, while stronger agents remain. At any given time, a wastewater that is only lightly buffered may be mixed with the rinse from a cleaning operation. This changes the load and the gain of the curve, upsetting the controlled pH for some time, given the low gain of the controller — and then the buffer passes on. However, if the controller were tuned for the buffered condition, then after the buffer is gone — especially when the plant is idle on weekends — the loop will cycle until the buffer returns.

This is a problem when treating wastes from regenerating ion-exchange resin beds. Both anion and cation beds are used for treating boiler feedwater and ultrapure water used in manufacturing semiconductors. These beds are regenerated batchwise, acid for the cation bed and base for the anion bed, and the effluents from regeneration must be neutralized prior to disposal. The first effluent leaving a bed is highly buffered, with the buffering gradually declining toward the end of the procedure, making neutralization of this effluent on-line quite difficult. It is better to combine the anion and cation washings in a common tank where they neutralize each other, and then treat the mixture batchwise.

THE PROCESS

The acid-base reaction is considered to be zero-order — that is, it does not take any time or depend on the relative concentrations of the reactants. This would seem to eliminate the need for the kind of careful process design that goes into chemical reactors in general. However, the very difficult aspects of pH control cited above place other burdens on the design of the process.

Vessel sizing and layout are very important, as are mixing and reagent delivery. Experts in pH control have been known to enter a plant and completely redesign a vessel and its auxiliaries in order to achieve a control performance that was not possible with the original configuration. This is one process that *must* be designed for controllability, or it will have to be modified on-site to meet effluent specifications.

Vessel Design

A *static* mixer is a section of pipe fitted with vanes that provide intense *radial* mixing without longitudinal or backmixing — it is a plug-flow device. The step response of a static mixer is between that of 20 and 50 stirred tanks in series, as shown in Figure 8.10i — it is essentially dead time, and as such it has a dynamic gain of almost 1.0.

A step in reagent concentration at the inlet will emerge at the outlet one residence time later virtually unmitigated. As a result, a pH controller will have to face the full steady-state gain of the titration curve without any dynamic help. If the gain of the titration curve at set point is 10, for example, the controller proportional gain will have to be the order of 0.05 to provide a loop gain of 0.5 as required for damped stability.

A proportional gain of 0.05 is a proportional band of 2000% — rarely are controllers successful in such service — and a gain of 10 is considered moderate for a titration curve. As a consequence, pH control is rarely exercised across a static mixer, and then only when the process is heavily and consistently buffered, as in a fermentation process saturated with carbon dioxide, or when the set point is in a flat portion of the titration curve.

An example of the latter is the addition of sulfuric acid to a solution of sodium dichromate preparatory to its reduction to chromium ions by sulfur dioxide.[3] In the range of pH 2–3, where the set point is located, the slope of the titration curve is about 300 times lower than at pH 8.

Another problem with the static mixer is that its dead time varies inversely with flow. The dead time is essentially the residence time, which is its fixed volume V divided by the variable flow (F) through it — and at zero flow, the loop is open. This can be fixed, and the dynamics can be measurably improved as well, by the addition of a circulating pump.

A circulating pump, taking suction from the static mixer's outlet back to its inlet at rate F_a will reduce the dead time from V/F to $V/(F + F_a)$ and convert the balance of the residence time to a first-order lag.[4] A circulation ratio F_a/F of only 2:1 can reduce the dynamic gain of the static mixer from 1 to 0.26, and the ratio of 5:1 reduces it to 0.12. This allows a comparable increase in controller gain or reduction in proportional band.

Neutralizations should preferably be conducted in a back-mixed tank, with a minimum of 5 min of residence time. If the reagent contains a solid material such as lime, the residence time should be up to 20 min to allow the lime to dissolve. If not enough time is allowed in the controlled vessel, any solid particles of lime carried out will continue reacting and raise the final pH of the effluent beyond the control point. A similar dynamic is encountered when the reagent is a gas such as ammonia or carbon dioxide, which also takes time to dissolve.

The shape of the neutralization vessel is important. Ideally, the depth of liquid should be similar to its length and width, to minimize distance traveled for a given volume. A ditch or channel is the worst configuration, because it will be dead time-dominant like a static mixer. Given a ditch, baffle the entrance to create a square chamber where pH is controlled, and let the remainder of the ditch act as an attenuation basin.

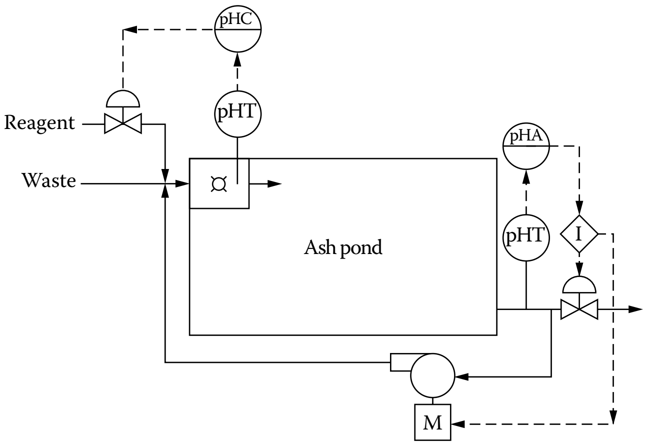


FIG. 8.32d
Ash pond control system, which recycles the pond's effluent until its pH is within specifications.

In some neutralization processes more than one vessel may be needed. If the waste alternates on both sides of pH 7, an upstream tank may allow for self-neutralization of some of the waste, thereby saving reagent. A downstream vessel is also useful to attenuate fluctuations in controlled pH, and may in fact be mandatory for impounding off-specification product. One such arrangement is shown in Figure 8.32d, for the treatment of fly-ash from a coal-fired power plant.

The waste is neutralized in a well-mixed corner of the ash pond where pH is controlled. It then overflows to the balance of the pond, which is unmixed, thereby allowing solids to settle. Hours later it emerges at the outfall, where its pH is again measured. If the effluent falls within specifications, it is discharged to the environment; if not, the discharge valve is closed and a recirculating pump started to return the wastewater for another round of treatment, as long as needed to bring it within specifications.

Mixing

Back-mixing is needed for uniformity of composition within the vessel, and a high velocity is required to minimize dead time between the point of entry and the point of discharge. Figure 8.32e shows two possible flow patterns in stirred tanks.

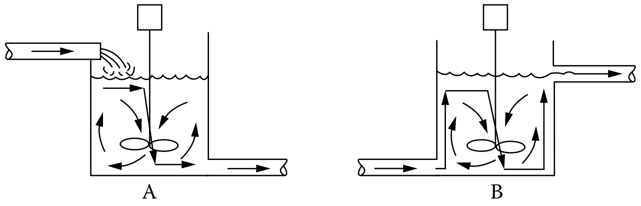


FIG. 8.32e
Flow patterns in stirred tanks. A. Recommended flow path. B. Undesirable flow path.

Reagent should be introduced — actually premixed — with the feed, and the pH should be measured at the vessel exit, but within the mixed zone. Then pattern A gives the shortest flow path, because mixers pump downwards (except those that intentionally aerate the vessel contents). Inflow entering the top of the vessel will be drawn to the bottom, where the pH is measured, whereas inflow entering at the bottom must flow up the side of the vessel first, doubling the dead time of the response of pH to reagent flow.

High-speed *axial* mixing is recommended, using a propeller or axial turbine, to minimize dead time. Low-speed radial mixers are intended to keep solids suspended and are not effective in this service, because they give excessive dead time. The step response for a typical back-mixed vessel is shown in Figure 8.10i. The ratio of dead time to residence time is essentially its dynamic gain.

In general, the degree of back-mixing can be defined in terms of the pumping capacity of an agitator with respect to the flow and volume of the neutralization vessel. In practice, however, this definition has limited usefulness because of variables such as agitator construction and blade pitch, baffling of the neutralization vessel, and placement of inlet piping and outlet measuring electrode. All influence the degree of back-mixing.

Experience shows that the best way to define back-mixing for control purposes is by the ratio of the neutralization vessel's dead time to residence time. A ratio of dead time to residence time equal to 0.05 is adequate for good control.

Figure 8.32f is a plot of tank size against agitator pumping capacity per unit volume on logarithmic coordinates. The family of curves shown for various dead times was developed from empirical data in tanks with capacities of 200, 1,000, 10,000, and 18,000 gal (756, 3,780, 37,800, and 68,040 l). They apply to baffled tanks of cubic shape, with the inlet at the surface and the outlet at the bottom on the opposite side of the tank. The ratio of impeller diameter to tank diameter varied from 0.25 to 0.4. Square pitch propellers at an average peripheral speed of 25 ft/s (7.5 m/s) were used in tanks up to 1,000 gal capacity. Axial-flow turbine impellers at an average peripheral speed of 12 ft/s (3.6 m/s) were used in the larger tanks.

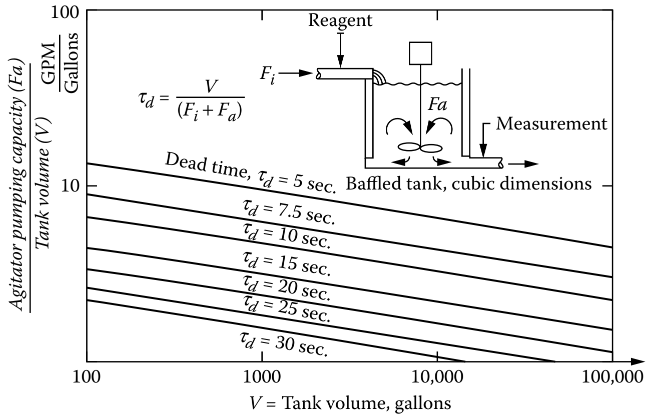


FIG. 8.32f
Dead time (τ_d) as a function of mixing intensity.

To be classified as a well-mixed vertical tank (by the standards of pH control applications), the liquid height should be between 100 and 150% of the vessel width or diameter. The vessel walls should have baffles to prevent liquid rotation, the agitation pattern should be axial, and the agitator pumping rate should be at least 20 times the influent flow rate. The agitation should be great enough to break the surface and pull down the inflow but not enough to cause air entrainment.

Reagent Demand

Titration curves produced from process samples in a laboratory are typically plotted as pH vs. milliliters of reagent, such as 0.01 *N* NaOH, used to treat a 50 ml sample. To convert this information into gal/min of industrial-strength reagent per 1000 gal/min of wastewater, one has to know the reagent concentration in normality. Table 8.32g provides that information for the most-common reagents.

To estimate the reagent requirement from a titration curve, consider the example where 10.2 ml of 0.01 *N* of NaOH was used to neutralize a 50 ml sample of wastewater. The original normality of the sample was then (10.2/50) 0.01 *N*, or 0.00204 *N*. If in the plant, a 10% NaOH reagent will be used, 1000 gal of waste will require 1000(0.00204) / 2.75 = 0.74 gallons of it.

A valve should be selected to deliver twice the maximum anticipated flow to allow for errors in the estimate, and to provide a margin for dynamic correction during recovery from load changes. A similar calculation should be made for the minimum anticipated load, to estimate the rangeability requirements of the system.

Reagent Rangeability

The rangeability needed in a given process is the ratio of the maximum required reagent flow to the minimum required.

TABLE 8.32g
Normalities of Common Reagents

Reagent	*Wt. %*	*Normality (N)*
HCl	32	10.17
	38	12.35
H_2SO_4	62.2 (50° Bé)	19.5
	77.7 (60° Bé)	27.2
	93.2 (66° Bé)	35.2
	98.0	36.0
NaOH	10	2.75
	25	7.93
	50	19.1
$Ca(OH)_2$	5	1.43
	10	2.86
	15	4.30

Metering pumps can be used for reagent delivery, but are of generally insufficient rangeability, typically on the order of 20:1, where *control valves* have rangeabilities of 35–100:1. The penalties for providing insufficient rangeability are two:

1. If the *maximum* reagent flow is insufficient for the largest plant load, the treated effluent will be off-specification—possibly well out of range. If the plant waste can be impounded as in Figure 8.32e, and enough capacity is available to weather the overload, this could be acceptable. But if not, the penalty may be severe.
2. If the *minimum* controllable flow the valve can provide is more than the minimum flow the plant requires, limit-cycling will result. This occurs because when the valve is shut, the small load will drive the pH away from set point, causing the controller to open the valve. However, when the minimum controllable flow is greater than needed to neutralize the current load, the pH will overshoot the set point, causing the controller to close the valve. This cycle will repeat, in a generally saw-tooth pattern, as long as the load remains below the throttling limit of the valve. If the amplitude of the cycle is within the acceptable pH limits, then the plant can live with it; a downstream vessel, even unmixed, will help attenuate the cycle.

Two valves in parallel can be sequenced to extend rangeability, but it must be done properly to be successful. Most importantly, they must not be open at the same time. This is because if the smaller of two sequenced valves is approaching full opening and the larger valve is then opened to its lowest controllable flow, the total flow delivered can almost double.

Therefore, the smaller valve must be closed when the larger is opened. Proper sequencing of valves differing in size by 20:1 or more must be done with equal-percentage characteristics, which are logarithmic, as plotted in Figure 8.32h.

The sequencing of the two valves in the figure produces essentially a single characteristic having a rangeability of almost 1000:1, with a small overlap in the center. The dots on the lines represent the calibration limits for the valve positioners. The smaller valve is throttling over the range of 0–52% of the control signal, and the larger valve from 48–100%. Logic on the control signal must switch one valve off and the other on when the controller output moves outside its range. Switching produces a small transient, but it lasts only for a second or so.

Because linear characteristics are needed for most pH loops, a curve characterizer is required between the controller output and the signal to the valve positioners. Its purpose is for the delivered flow to be linear with the controller output. Figure 8.32i describes the equal-percentage flow characteristic for 1000:1 rangeability, and a compensating hyperbolic characterizer of the form:

$$f(m) = \frac{m}{L + (1 - L)m/100} \qquad 8.32(8)$$

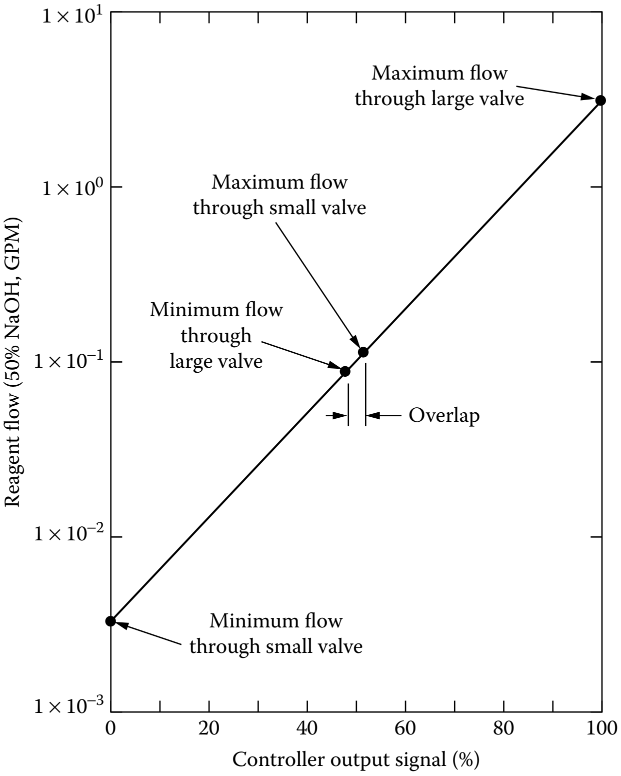


FIG. 8.32h
Reagent flow using sequenced valves.

where m is the controller output and L is the linearity of the characterizer; when $L = 1$, the function is linear. To compensate the 1000:1 equal-percentage valve in the figure, $L = 0.1$ and the compensation is adequate, as shown by the almost-linear combination of the two. Alternatively, an *x-y* function generator can be used.

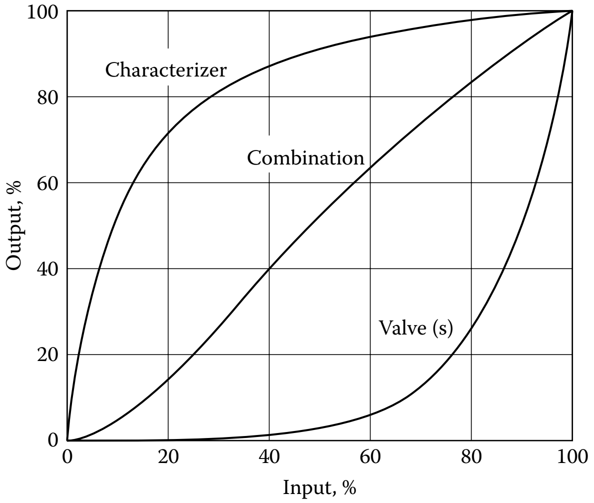


FIG. 8.32i
A hyperbolic characterizer can effectively compensate a 1000:1 equal-percentage valve.

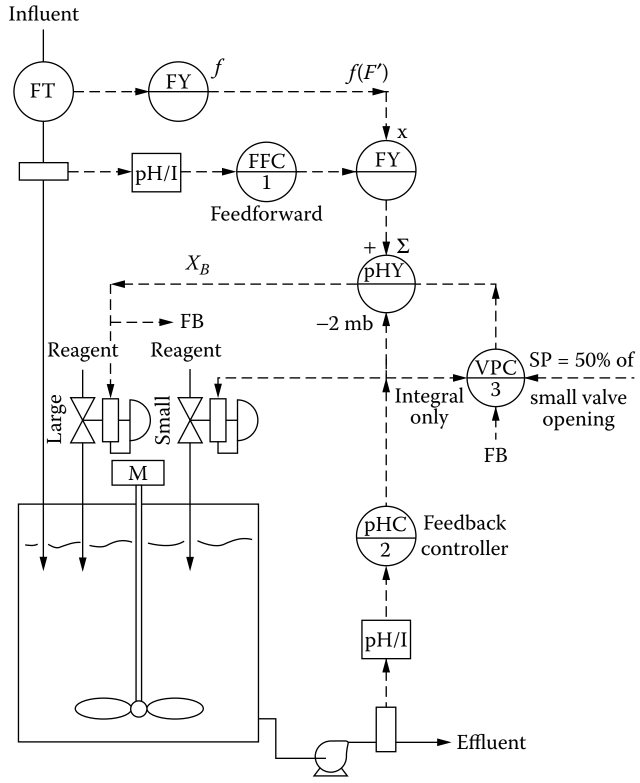


FIG. 8.32j
VPC-3 controlling the small valve opening at 50%. The feedforward portion of the control system is useful only if the relationship between influent pH and reagent demand is consistent, which usually is not the case.

Another, simpler, way to extend valve rangeability is through the use of a valve position controller (VPC) to drive the larger valve while the pH controller drives the smaller. The VPC compares the output of the pH controller to its own set point of 50%, and moves the larger valve to drive it slowly there by integral action. Short-term control is the function of the smaller valve, tending to return to 50% opening as long as the load is within the range of the larger valve. Figure 8.32j illustrates a VPC-based feedback control loop. The figure also shows feedforward trimming of that loop, which is not recommended unless the relationship between influent pH and reagent demand is consistent.

At times, the VPC will drive the larger valve closed, and this may bump the pH, as it will when suddenly opening to its minimum flow. This system may also not be fast enough to cope with load changes outside the range of the smaller valve.

When wastewater pH can vary on both sides of neutral, then acid and base valves must operate in sequence, the acid valve opening from 50 to 100% controller output, and the base valve opening from 50 to 0. Yet both valves must fail closed, and therefore the base valve requires a reverse-acting positioner.

Smart digital positioners are recommended on all valves used in pH control, whether sequenced or not. They combine speed with the elimination of dead band and can be individually characterized if necessary.

Reagent Piping

If reagent is simply dropped on the surface of the water in a tank, it will not be adequately dispersed into its contents, and the pH measurement will reflect this by randomly moving about. Premixing of the small flow of concentrated, viscous reagent with the much larger flow of water is essential for a steady pH reading. The reagent therefore should enter the vessel at the same point as the inflow, taking advantage of entrance hydraulics to provide premixing.

A major source of dead time in a pH loop is the movement of a small flow of dense reagent through the transfer line into the wastewater. When the reagent valve opens, flow into the mixed zone of the vessel should start immediately, and when it closes, flow should stop. The transfer line from the valve to the vessel must remain full at all times to accomplish this. The line should terminate in a loop seal, whether discharging on the surface, or through a dip tube into the bottom of the vessel (in the case where a bottom feed entry cannot be avoided).

Flushing of a transfer line with water is sometimes done, but is not recommended in two common cases. Concentrated sulfuric acid reagent is not corrosive, allowing cast iron and mild steel to be used for valves and piping. But when diluted with flush water, it attacks these materials vigorously. The other problem is with lime slurry. Flush water containing carbon dioxide will react with the lime at their point of contact to form a very hard scale of calcium carbonate, which will eventually plug the line and is very difficult to remove.

Lime is of very limited solubility and is, therefore, always administered as a slurry whose piping must be very carefully designed to avoid plugging — Figure 8.32k illustrates how it should be done. A level switch (LS) on the slurry tank is used

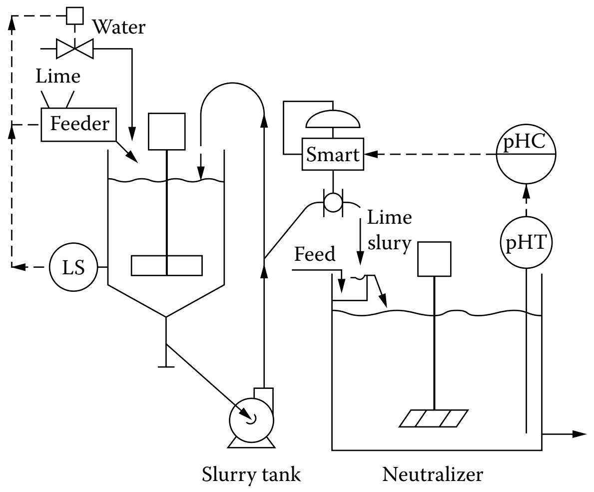


FIG. 8.32k
Control system for a lime-based neutralization process.

to turn on the solid-lime feeder and to simultaneously open the water valve when the level is low and to turn them off on high level, thereby adding the lime and water in consistent proportions.

The circulating pump is always running, keeping the slurry in constant motion. A ball-type control valve draws slurry from the top of its branch to the neutralization vessel, as needed. When closed, any solids in the branch will fall back downhill and away from the valve, avoiding plugging. Ball valves have an inherent equal-percentage characteristic, which must be linearized by a characterizer in their smart positioner.

Multiple Stages

Manipulating a reagent flow to within ±1% or 1 part in 100 is a reasonable expectation in a single well-designed treatment stage. This would take deionized water from pH 4 or 10 to 7±1, slightly buffered water from pH 3 or 11 to 7±1, and well-buffered wastewater from pH 2 or 12 to 6–9, a typical wastewater specification.

When the reagent demand is greater, or the buffering less, a single stage may not be enough. Effluent pH may cycle excessively simply due to valve imprecision. In this case, a second stage is needed, and in extreme cases of solutions at pH 0–1 or 13–14, a third stage. Each stage of treatment has its closed pH loop, which is capable of passing along a cycling load to the next stage.

Unfortunately, the worst type of load disturbance with which a feedback loop must contend is one that is cycling at its own period. To reduce this sensitivity to cyclic load changes, the individual pH loops should have different periods of oscillation. This is best achieved by sizing the stages progressively larger from feed to discharge, as shown in Figure 8.32l.

Each successive stage has a smaller valve, so failure of an upstream stage to control pH in an acceptable range could allow an overload to pass downstream where it cannot be handled. To protect against this possibility, the set points of upstream stages need to be properly positioned. For example, the first of three stages should control pH at 1 or 13, the second stage at 3 or 11, and the last stage at 7–8.

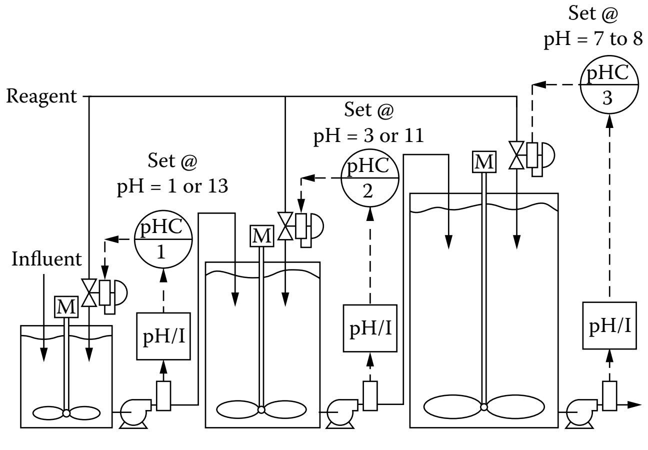


FIG. 8.32l
Classical three-stage pH control system.

THE CONTROL SYSTEM

If the process is well-designed, the control system can perform well with a simple structure. By contrast, even the most complex control system will not perform well or will be prone to failure and operator error if the plant is poorly designed. This applies to feedforward control, as well.

Feedforward adds complexity to the system and is prone to error — more in pH control than in any other process application — owing to the wide variability encountered in titration curves. For these reasons, only the simplest systems are described here.

Measuring pH

For a detailed discussion of pH detectors, refer to Section 8.48 in Chapter 8 in Volume 1 of this handbook.

The measurement of pH used for feedback control must be in the mixed zone, but directly before the exit, so that it truly represents the quality of the effluent. For the preferred flow pattern of top entry, bottom exit, this can mean locating the electrodes under several feet of liquid head.

In Figure 8.32k, this is avoided by locating the probes in the discharge of the transfer pumps. However, if flow were ever to completely stop, they would be in a dead space, causing the pH loop to be open. To avoid this problem, each electrode assembly should be located in a slipstream from the pump back to its tank.

Electrode assemblies have some strange failure modes. The reference electrode and the inside of the glass measuring electrode are both filled with solutions buffered to pH 7. As a result, a measured pH of 7 corresponds to 0 mV across the electrodes — unfortunately, so would a dead short. So a reading of pH 7 is not always good, because it can also be caused by a short circuit, or by a conduit full of water.

Remember that the electrical impedance of a glass electrode is in the order of 100 MΩ, and it only produces 59 mV per pH unit away from 7. Another source of error is a bias caused by contamination of the reference electrode, or by a ground current being rectified by it. Follow manufacturer's procedures for proper grounding, and frequently check electrode calibration against buffer solutions.

Electrodes can also become coated with scale, precipitates, tars, and biological growth. Coatings impede the free flow of ions between the sensitive surface of the glass and the solution at large, creating a dynamic lag that could be up to several minutes long. The effect is like a temperature sensor in a thermowell, but one that grows with time.

If the accumulation is slow, the electrodes may be removed periodically for chemical cleaning: Acids are effective if the

TABLE 8.32m
*Ratings for Various Types of Cleaners**

	Application	*Ultrasonic*	*Water-jet*	*Brushing*	*Chemical*
Slime Microorganism	Food, paper, pulp, aquatic weed	X	O	O	△
	bacteria (activated sludge) whitewash	△	O	O	△
Oil	Tar, heavy oil	X	X	X	△
	Light oil	O	△	△	O
	Fatty acid, amine	X	O	X	O
Suspension	Sediment	O	X	X	O
	Metallic fines	△	X	X	O
	Clay, lime	△	O	X	O
Scale	Flocculating deposit neutralized effluent $CaCo_3$	△	△	△	△

O: Recommend
△: Applicable
X: Not applicable
*Courtesy of Horiba Instruments.

electrodes are in a predominately basic environment, and vice versa. If accumulation is rapid, ultrasonic cleaners are effective against most sources of fouling, and mechanical brushes are also used. A fluid velocity of 5–9 ft/s (2–3 m/s) past the electrodes will tend to keep them clean without excessive abrasion. See Table 8.32m for recommendations on cleaner selection.

The pH Characterizer

The only special feature used in PID controllers for pH is a nonlinear characterizer to help linearize the loop. Titration curves come in a wide variety of shapes, so there is no universal pH characterizer. However, there does not need to be — it is not necessary to match the curve over its entire length, because control is exercised mostly around set point.

Three classes of characterizers are presented here: matched, three-piece, and error-squared. The characterizer matched to the titration curve is the best option, but it is not realistic for most applications, principally because of the variability of the process curve. The three-piece characterizer consists of three straight lines, as shown in Figure 8.32n.

The error-squared curve is a parabolic function that has no adjustments; it is simply $f(e) = e|e|$. It is compared with a typical matched curve in Figure 8.32o.

The error-squared curve has zero gain at zero deviation, which can produce a small amount of offset. This is not a problem with wastewater, whose specification gives some latitude, e.g., pH 6–9. The three-piece function has an adjustable minimum gain, along with an adjustable width to the zone where that gain is applied; its gain outside that zone is 1.0. Other characterizers are feasible, but these have had the most extensive use.

Figure 8.32p plots the results of a simulated neutralization conducted in a stirred tank using the software of Reference 5. A solution containing 0.001 N HCl and 10^{-5} N CO_2 is being neutralized by NaOH, which gives a process gain of 75 at set point. The time scale is normalized to the residence time of the vessel.

With a *linear* PI controller, the 10% step-load change drives the pH from a set point of 7 beyond 9, where it remains for a full residence time before falling, followed by a large overshoot. If the specifications are pH 6–9, then they are violated in both directions.

The length of time required for recovery is critical, because it reduces the effectiveness of any attenuation vessel and can be detrimental to biological life downstream. The proportional band of the controller is 700%, set to provide (light) damping at set point; integral time is $0.32V/F$. The

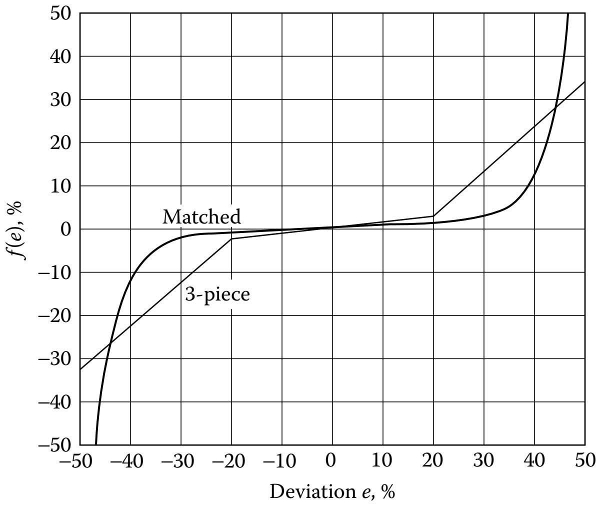


FIG. 8.32n
The three-piece curve has an adjustable low-gain zone.

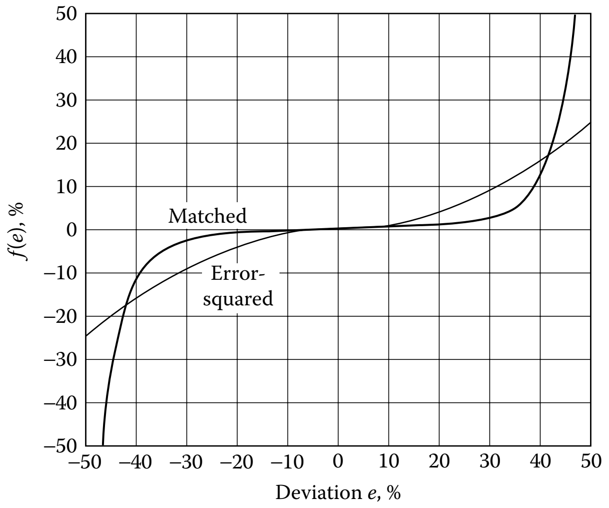


FIG. 8.32o

The matched curve is the most complex, the error-squared curve the simplest.

slow recovery is due to the low loop gain away from set point. Judging from this response, nonlinear compensation is needed to have any chance of satisfying specifications.

A characterizer *matched* to the pH curve across 200 points provides the best response. Because of its very low slope at set point, the controller's proportional band can be reduced to 12% (integral settings for all four controllers are the same). This provides the gain needed for rapid recovery from severe load changes.

Still, the time required to recover is much longer than the period of the cycle around set point. Due to the severely nonlinear nature of the process, even with matched characterization the loop will not behave like a linear loop having the same process dynamics.

The *three-piece* characterizer was given a zone width of ±1.5 pH units, where the gain is 0.10. The proportional band of the controller is 75%, which at set point gives the same gain as a linear controller at 750% band. It performs almost as well as the matched characterizer in this test, although it is a very crude approximation to the titration curve.

The crude fit means that the performance of the controller depends to some extent on the size of the load variations — deviations that lie just within the low-gain zone will not be corrected as quickly as those exceeding it. The width of the low-gain zone is the most effective adjustment and can be used to tune the loop to a particular titration curve. The gain in the zone can be left at 0.1 for most pH curves — it should not be set to 0, or the loop will limit-cycle between the two high-gain zones.

The error-squared characterizer is the least effective of the three, but has no adjustments to tune or points to match. It therefore can be considered as a compromise characterizer for pH control, keeping in mind that P, I, and D are more than enough adjustments for many technicians to tune.

PID Tuning

The tuning of pH controllers is based on the same technology as linear loops, which is described in Section 2.35 in Chapter 2 in great detail. But the extreme nonlinearity of the process makes some modifications necessary. After choosing a characterizer and its settings, the proportional band or gain needs to be adjusted so that the loop is damped at set point, but only lightly.

Several observations need to be made over time of the closed-loop behavior at set point, because as buffers leave the system, limit-cycling may develop — a very common observation in pH control. Limit-cycling may not be objectionable in some pH loops, providing that its amplitude is within the specification limits.

Where it can become costly is in a system where the controller manipulates acid and base valves in sequence. Then, the cycling could cause the controller to alternately add acid and base to the vessel, even with — in fact, especially with — no neutralization load or even any flow of water. To guard against this possibility, the air supplies of the reagent valves serving vessels with pumped flow should be shut off when the pump is stopped — closing both valves — and their controller should be transferred to manual. When the pump restarts, the controller can then resume where it left off.

The length of time required for pH to recover from a major load change, as shown in Figure 8.32p, limits the rate at which the controller can integrate without excessive overshoot. Too short an integral time causes overshoot, and it needs to be increased by a factor of two or more above its optimum value for a linear loop.

The derivative time needs to be optimized for the fastest part of the cycle, which occurs around set point. In this respect, its setting would be about the same as for a linear loop. A noisy pH measurement can preclude the use of derivative action completely, but the source of the noise is worth

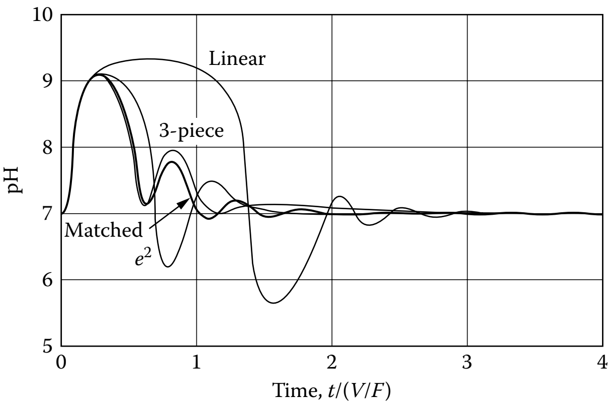


FIG. 8.32p

The response to a 10% step change in load is best with a characterizer matched to the pH curve. Other responses shown are those of "Linear" (linear PI controller), "three-piece" (three-piece characterization), and e^2 (error-squared characterization).

investigating and eliminating, because derivative is of great benefit in composition control, as it is able to substantially reduce the period of oscillation and, hence, the load recovery time.

Self-Tuning Control

Self-tuning controllers are useful whenever process parameters change substantially and frequently. If these variations can be associated with measurable quantities such as flow, then direct compensation should be applied.

An example would be the programming of the integral time of a pH controller on a static mixer to vary inversely with the measured flow, to compensate for dead time variation. When there is no measured variable that correlates to loop behavior, the only remedy is to change controller tuning based on that observed behavior. A device or software that does this is a self-tuning controller.

Given that a variable titration curve is changing the loop behavior, the proper compensation should be applied to the controller's characterizer. But this would require an on-line titration to be run periodically to determine the need for recharacterization. Unfortunately, this is not likely to be helpful, because an automatic titrator is a miniature pH control loop, with similar dynamics — therefore, the information it obtains is too late to be of much use. The first evidence of a change in the curve is likely to be the limit-cycling of the pH loop.

The best way to approach this problem with present technology is to tune the pH controller and to fit its characterizer as well as can be done for the existing or estimated worst-case conditions, and let self-tuning do the rest.

Commercial self-tuning PID controllers are available that observe loop behavior during upset conditions; calculate decay ratio, overshoot, and period of oscillation; and estimate the PID settings that will improve performance. A device that does this without intentionally disturbing the loop is the Foxboro EXACT controller.[6,7,8]

One danger of using a self-tuning controller is its tendency to adapt to unfavorable process changes, like a fouling electrode. It will continue to slow down the controller to accommodate the fouling, where the preferred remedy is to clean the electrodes.

Batch pH Control

The theory behind batch end-point control is covered under Chemical Reactors: Control and Optimization, in Section 8.10. In fact, a pH loop is used there to illustrate the concept. A vessel filled with a solution is held while its pH is adjusted by adding a reagent — there is no load because there is no flow into or out of the vessel. In that section, it was demonstrated that a proportional-only or proportional-derivative controller should be used, with a fixed output bias of zero, corresponding to the zero-load condition. In this way, the valve will be closed when the pH reaches set point.

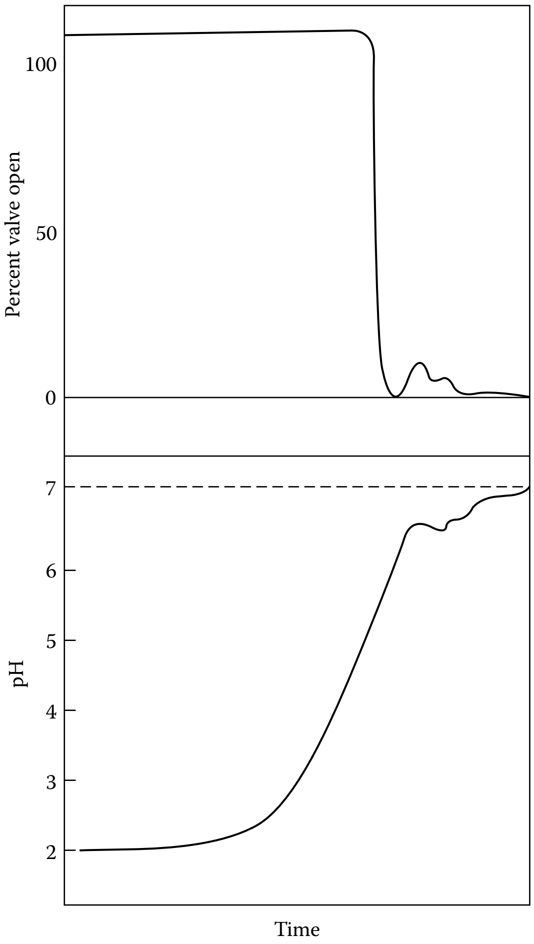


FIG. 8.32q
Measurement and valve opening behavior for a batch process.

The unique feature that a pH measurement brings to batch end-point control is the logarithmic titration curve. Compensation for this nonlinear feature is easily provided through the use of an equal-percentage reagent valve. The valve need not match the titration curve closely at all, for the only time the pH will be at the set point is at the end of the run, when the valve is closed.

So, the valve simply must deliver enough flow when wide open to give a reasonable processing time and throttle finely toward the end of the run, when the pH is on the steep portion of the curve. A record of pH and valve opening for a batch neutralization is shown in Figure 8.32q.

Observe how derivative action closes the valve before the set point is quite reached, resulting in an undershoot. This extends run time slightly, but that is preferred to using too little derivative action, which would result in a permanent overshoot.

Feedforward Control

When load disturbances are sudden and wide-ranging, feedforward control is recommended to reduce variations in the controlled variable, as described in detail in Chapter 2, Section 2.9:

Feedback and Feedforward Control. The method calculates from measurements of the load variable(s) where the manipulated variable needs to be positioned to maintain a dynamic balance in the process — in the case of pH control it is a material balance.

However, in the case of pH control the material balance must be extraordinarily precise — matches of reagent to load of 1 part in 1000 may be required. Equations have been presented at the beginning of this section to show the relationship between pH and the concentration of various agents, but they vary widely with the species and their concentrations. The problem is that all of the acids and bases must be neutralized, but that some are only partially ionized and sensed by the pH electrodes at any given pH. It is much like estimating the size of an iceberg from observing its tip, but with the relationship between the two being quite variable and even unknown.

Consider, for example, HCl in solution at pH 3.0 — its concentration would be 0.001 *N*, and it would require 0.001 *N* of caustic to neutralize it. Now look at Figure 8.32b, the titration curve for 0.01 *N* acetic acid. With no other agents in solution, its pH is about 3.5 — *higher* than that of the above HCl solution and, therefore, seemingly more neutral, yet it would take *10 times more* caustic to neutralize it.

In other words, if a feedforward signal from an influent pH measurement were to be used to estimate the flow of caustic reagent required for neutralization, the change in the influent from 0.001 *N* HCl to 0.01 *N* acetic acid would cause a miscalculation by a factor of –20!

In this light, feedforward is *not* recommended for pH control, unless there is a single species that must be treated, so that the relationship between influent pH and reagent demand is consistent. In wastewater treatment it is not. Feedforward from flow is much less of a problem — it amounts to a pH controller setting a ratio between measured inflow and manipulated reagent flow. But even this has its limitations. For example, in municipal wastewater treatment, the reagent dosage required for domestic waste is very different for stormwater. Yet in most municipalities, the two are mixed in varying proportions depending on the weather, causing the required dosage to vary widely.

References

1. Shinskey, F. G., *pH and pIon Control in Process and Waste Streams,* New York: Wiley-Interscience, 1973, p. 1–2.
2. *Ibid.*, p. 72.
3. *Ibid.*, p. 119.
4. Shinskey, F. G., *Process Control Systems,* 4th ed., New York: McGraw-Hill, 1996, p. 93.
5. Shinskey, F. G., *Problem Solving Software,* Ver. 3.2, Foxboro, MA: The Foxboro Company (Invensys Systems), 1998.
6. Instruction MI 018-885, "672CNA Single-Station Micro Controller," Foxboro, MA: The Foxboro Company, 1996.
7. Kraus, T. W., and Myron, T. J., "Self-Tuning PID Controller Uses Pattern Recognition Approach," *Contr. Eng.,* June 1984.
8. Callaghan, P. J., Lee, P. L., and Newell, R. B., "Evaluation of Foxboro Controller," *Process and Control Eng.* (Australia), May 1986.

Bibliography

Hoyle, D. L., "Designing for pH Control," *Chem. Eng.,* November 8, 1976.

Mahuli, S. K., Rhinehart, R. R., and Riggs, J. B., "Nonlinear Model-Based Control of pH," AIChE Annual Meeting, 1992, Paper 149g.

McMillan, G. K., *pH Control,* Research Triangle Park, NC: ISA, 1984.

McMillan, G. K, "Understand Some Basic Truths of pH Measurement," *Chem. Eng. Progr.,* October 1991.

McMillan, G. K, "Methods of Controlling pH," *Chem. Processing,* July 1997.

Moore, R. L., *Neutralization of Wastewater by pH Control*, Research Triangle Park, NC: ISA, 1978.

Piovoso, M. J., and Williams, J. M., "Self-Tuning pH Control," *InTech,* May 1985.

Shinskey, F. G., "Sharpening pH Control," *Control,* July 1998.

Waller, K. V., and Gustafsson, T. K., "Fundamental Properties of pH Control," *ISA Trans.,* Vol. 22, No.1, 1983.

8.33 Power Plant Controls: Cogeneration and Combined Cycle

J. E. JAMISON, B. G. LIPTÁK, A. ROHR (2005)

INTRODUCTION

The controls of the unit operations used in the power plants (boilers, heat recovery steam generators, compressor, turbines, combustion controls, cooling towers, and condensers) are discussed in detail in other sections of this chapter. Also, these mechanical packages, as purchased, are often already provided with their integral control systems. For these reasons, this section will give more emphasis to the description of the mechanical equipment, its features, possible combinations, operation, and start-up and shutdown procedures.

This section will describe processes that can substantially increase the efficiency of power generation. These processes include cogeneration, combined cycle generation, and the use of both techniques in combination. Cogeneration is a mode of operation when the plant produces both heat (steam or hot water) and electricity. In a combined cycle power plant, electricity is produced by two turbines, a gas and a steam turbine. The gas turbine is operated by the combustion products of the fuel (Brayton cycle), while the steam turbine (Rankine cycle) is operated by the steam generated by the heat content of the exhaust gases leaving the gas turbine.

The efficiency of a traditional power station, consisting of a fired boiler, full condensing steam turbine, and electric generator, can hardly reach 40%. The major losses occur in the steam condenser, which wastes more than 45% of the total thermal energy that is supplied to the plant. Therefore, when the total electric power generated is constant, the overall power plant efficiency can be improved by lowering the inlet steam quantity to the condenser.

COGENERATION

Cogeneration means that there are two marketable products: electricity and heat. The heat is usually in the form of steam, but sometimes as hot water. The heat is sold to process industries (paper mills, tire production plants, petrochemical plants, textile mills, desalination plants, sugar mills, and so on) or for civil applications (district heating), and thereby the latent heat of the steam is not wasted, but utilized.

In order to obtain a significant improvement in the overall plant efficiency, it is necessary that 1) The portion of the steam that can be sold is a high percentage of the steam being generated. 2) The demand for steam is continuous and steady. This means that seasonal users like sugar mills or district heating applications are not as good applications as are paper mills or tire production plants. 3) The steam user should be near the power plant (1000–5000 ft, or 300–1500 m), in order to reduce steam transportation and heat losses.

As a consequence, the large generation units normally are not suitable for cogeneration, because the exported steam would be only a small portion of the total that is generated, making the application unattractive. Also, the distance of the power station from the users would make the installation economically impractical.

In some cases, steam generation can be more critical than electricity production. This might require the installation of an auxiliary boiler to guarantee the continuous availability at least of the minimum steam supply that is required to protect against damage to the user's plant or production, or to avoid incurring expensive penalties.

The steam can be obtained from back-pressure steam turbines or from controlled/noncontrolled extraction in the IP or LP stage of the turbine, in an amount determined by the user. Some percentage (0–100% according to the application) of the steam is returned as condensate to the power plant.

The quantity of the returned condensate can impact the size of the demineralization unit and can require continuous checking of the quality of the returned condensate. If the condensate quality is unacceptable as boiler feedwater, it might be dumped to wastewater treatment or to the demineralization plant, if the pH, and/or conductivity, and/or TOC, and/or silica are beyond the acceptable limits. Before entering the thermal cycle of the boiler in the hot well of the steam turbine condenser, the returned condensate is mixed with the demineralized make-up water.

Cogeneration with Combined Cycles

Smaller combined cycle power plants (up to 60–100 MW) are particularly suitable for cogeneration, because they can be located very close to the thermal user or even within the fence of the industrial plant. Figure 8.33a is a simplified process flow diagram showing a combined cycle with cogeneration, in which some steam is sent to the associated industrial plant.

In many cases, the steam is generated at two pressure levels and the heat recovery steam generator (HRSG) is designed to serve the steam users. In some installations, if the HP or LP steam characteristics are suitable for the steam turbine, flexible operation is obtained by sending the excess steam to the turbine or by drawing steam from the turbine supply to meet the needs of the users in the plant.

Figure 8.33b shows a steam distribution control system where the steam generated by the HRSG is distributed to the steam users in the plant while sending the remaining low-pressure (LP) steam to the steam turbine. In the case when no steam is required by the steam users, PV-1 is fully open and PV-2 and PV-3 are closed.

The system status and operation is summed up in Table 8.33c. When, compared to the HRSG production, the users require only a small amount of steam, PIC-1 output is in the 50–100% range and throttles PV-1, and the PIC-2 output is in the 0–50% range and modulates PV-3 while PV-2 is closed. When the users' demand for steam rises further and, therefore, PIC-2 output rises above 50% (exceeding the steam availability from the HRSG), it throttles PV-2, while the low-signal selector PY-2 selects the under-50% output of PIC-1 to throttle PV-1 and PV-1 is closed. In this mode, the LP steam generated by the HRSG is supplemented by steam from the turbine.

The purpose of the low-signal selector PY-2 is to prevent the depressurization of the LP drum in case of high steam demand from the users. PV-2 and PV-3 operate on a split range, so that PV-2 starts opening only when the PIC-2 output has risen above 50%.

Cogeneration with Internal Combustion Engines

In case of smaller loads (e.g., hospitals or small district heating), when the required amount of heat is small and is at a low temperature, the required electric power can be generated with internal combustion engines. These engines can generate up to 1–2 MW. The heating medium is pressurized hot water that is close to its boiling point.

In this configuration, the heat is obtained by recovering it from the cooling circuit, lubricating oil, and flue gas of the engine, by means of heat exchangers.

COMBINED CYCLES

In a combined cycle power plant, electricity is produced by two turbines, a gas and a steam turbine. The gas turbine is operated by the combustion products of the fuel (Brayton cycle), while the steam turbine (Rankine cycle) is operated by the steam generated by the heat content of the exhaust gases leaving the gas turbine. The name *combined cycles* comes from the fact that the gas turbine operates according to the Brayton cycle and the steam system operates according to the Rankine cycle.

As shown in Figure 8.33d, the dual-shaft combined cycle plant consists of a gas turbine (GT) with its associated electric generator, a heat recovery steam generator (HRSG), a steam turbine (ST) with its associated condenser and electric generator, plus auxiliaries like a demineralization/polishing process, a fuel gas or fuel oil system, and a closed circuit cooling water system. The gas turbine (GT) exhausts into the heat recovery steam generator (HRSG) where the heat content of the flue gas produces steam, which is fed to the steam turbine (ST).

The combined cycles configuration is preferred for base load applications, i.e., to operate continuously at full power or very close to it, even though sometimes they can also be used to meet peak loads. When operating at base load (100% load) with no steam used for cogeneration, 3/4 of the generated electric power will be generated by the gas turbine (GT) and 1/4 by the steam turbine (ST), if aeroderivative turbines are used. When using heavy-duty turbines 2/3 of the electric power will be from the GT and 1/3 from the ST.

At partial loads on a heavy-duty GT, the ratio between the generated power from the two sources changes and can be 60% from the GT and 40% from the ST. Given a specific GT with its associated HRSG, at each load (generated MW) by the GT, there corresponds a certain (proportional) generated power by the ST, such that in steady state, the total generated power can be controlled by controlling the generated power of the GT.

During transients (either an increase or decrease of the total power), the ratio of the generated power between the GT and the ST can be different, because the GT has a much quicker dynamic response (seconds) than the HRSG and ST assembly (minutes). This fact has normally little impact on the operation of the power station, because a combined cycle is always operating in almost steady-state conditions.

If the load increase is large, it is possible that after a first step of a limited value (e.g., 5%), the rate of change in the GT needs to be limited to 2–5%/min. This is because the ST is unable to accept a sudden change in steam characteristics and quantity. The net heat rate for combined cycle units >150 MW is in the range of 5700–6800 BTU/kWh (6015–7175 kJ/kWh), while for 60 MW it is in the range of 6500–7000 BTU/kWh (6860–7385 kJ/kWh).

Single-Shaft Arrangements

In Europe it is quite popular to have a single-shaft arrangement for the rotating machinery, and this design is now also increasingly accepted in North America. This arrangement decreases the overall cost even though it decreases the flexibility of the plant. In this design, there is only one electric generator, driven by both the gas turbine and the steam turbine. This means that the electric power plant is simpler,

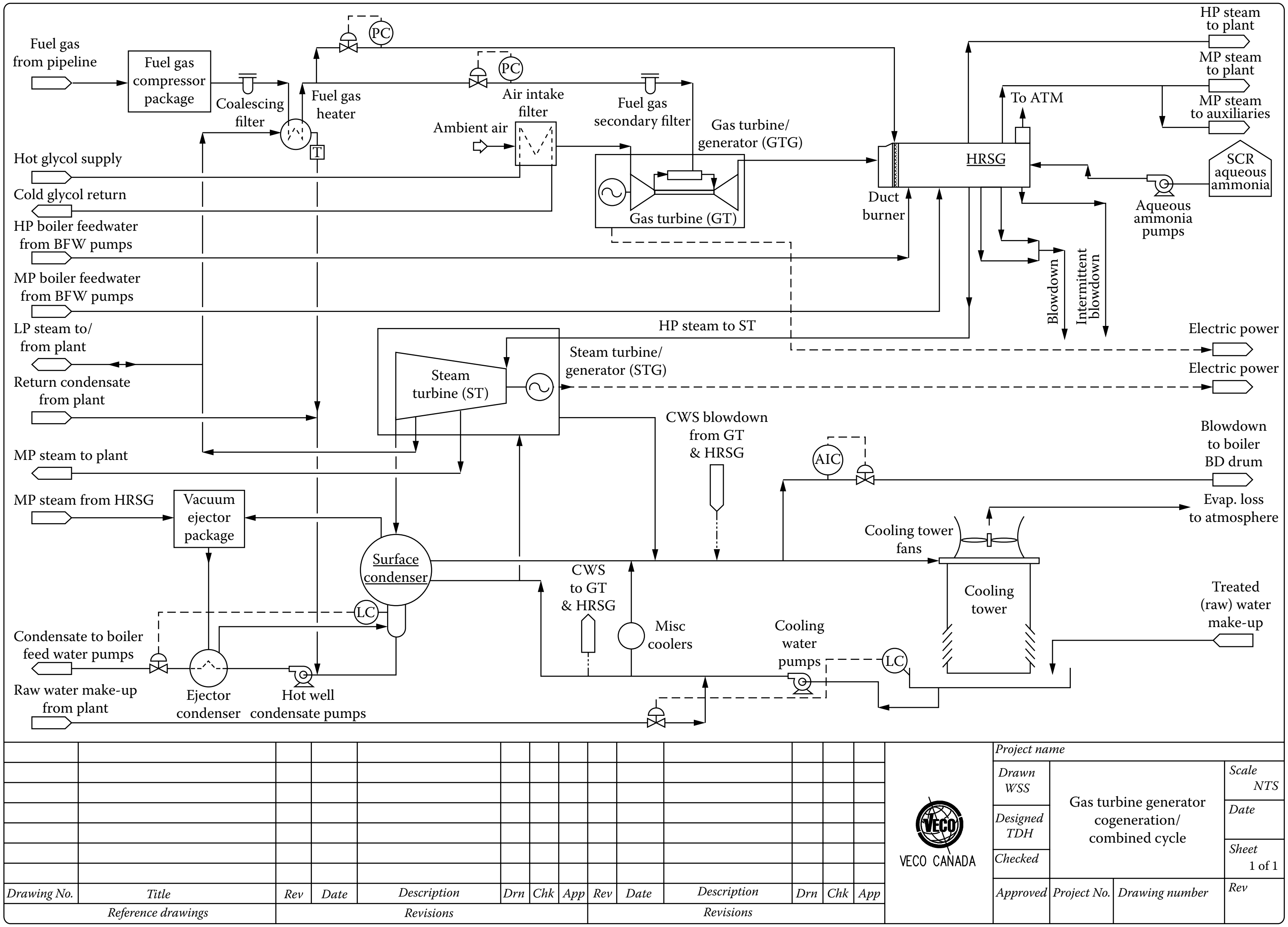


FIG. 8.33a

Piping and instrument flowsheet of a dual shaft cogeneration power plant with combined cycle. (Courtesy of VECO Canada Ltd.)

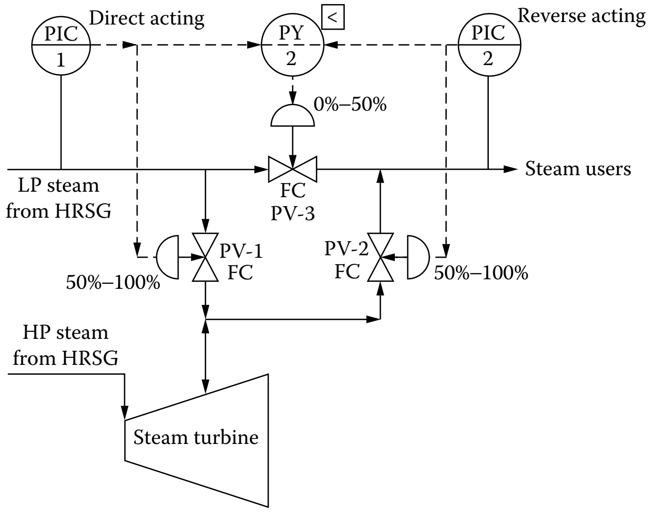


FIG. 8.33b
Control system used to distribute the steam generated by a heat recovery steam generator (HRSG) in such a way that the variable demand for steam can be met by either sending the excess to the steam turbine or by supplementing it from the steam turbine.

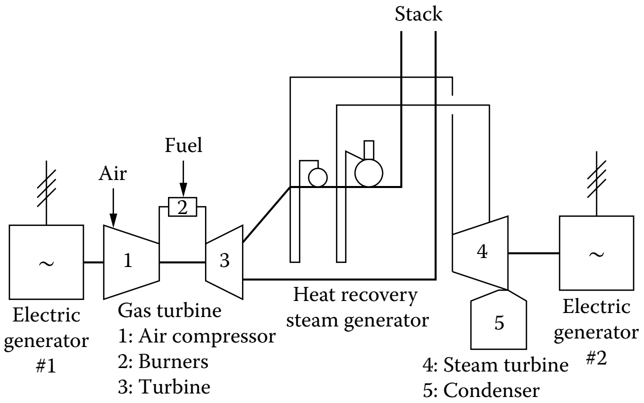


FIG. 8.33d
The main components of a dual-shaft combined cycle power plant.

because there is only one step-up transformer and one bay to connect to the grid.

There are two possible single-shaft configurations, shown in the top and bottom of Figure 8.33e: One is to locate the gas turbine and the electric generator at the two ends of the shaft (top) and the other is to locate the steam and gas turbines at the two ends of the shaft (bottom).

The configuration at the bottom requires a clutch or a special joint (e.g., SSS™) located between the generator and the ST. This is needed in order for the generator to be able to rotate during start, driven by the GT, with the ST stopped until the steam conditions are suitable for starting the ST with its own start-up procedure. If this joint is not used, an auxiliary boiler is needed or steam is derived from other HRSGs to supply steam to the seal glands and to the vacuum ejectors of the condenser, and at least a small amount to the ST.

The configuration on the top of Figure 8.33e always requires auxiliary steam for the vacuum system and for seal glands or from an auxiliary boiler or from other boilers available in the plant.

Alternative Configurations Sometimes, when an even number of GTs is used, it is possible to have two GTs/HRSGs

TABLE 8.33c
Status and Operation of Figure 8.33b Loop Components

Component Characteristics and Status	*User's Steam Demand is Zero*	*User's Steam Demand is < HRSG's LP Availability*	*User's Steam Demand is > HRSG LP Availability*
Direct-acting PIC-1 output	50–100%	50–100%	0–50%
Reverse-acting PIC-2 output	0%	0–50%	50–100%
PV-1 (fail closed, 50–100% range valve) status	Throttling	Throttling	Closed
PV-2 (fail closed, 50–100% range valve) status	Closed	Closed	Throttling
PV-3 (fail closed, 0–50% range valve) status	Closed	Throttled by PIC-2	Throttled by PIC-1
PY-2 (low-signal selector) selects the output of	PIC-2	PIC-2	PIC-1

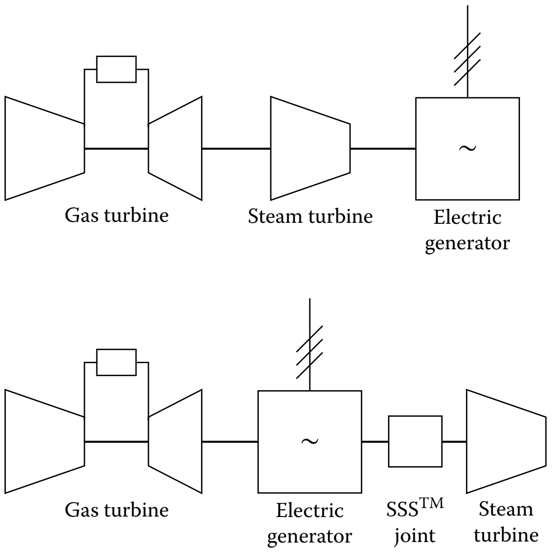


FIG. 8.33e
Single-shaft power plant configurations can locate the gas turbine and the electric generator at the two ends of the shaft (top) or can locate the steam and gas turbines at the two ends (bottom).

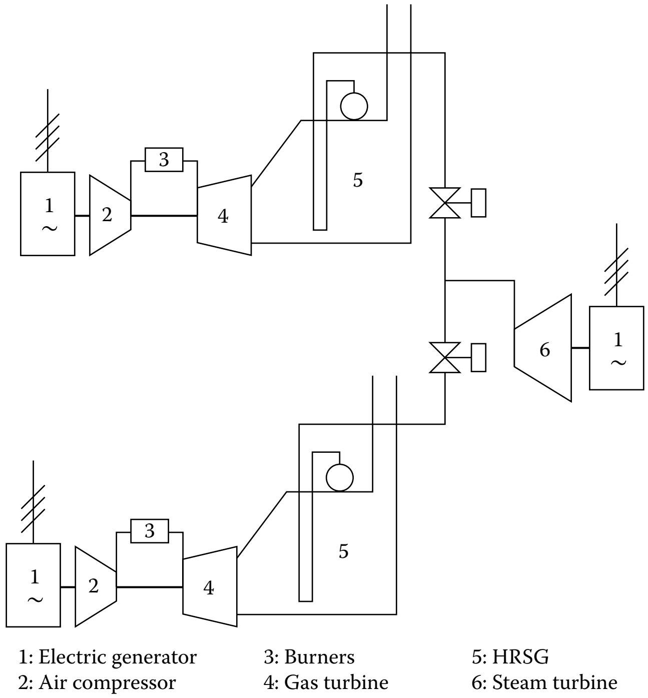


FIG. 8.33f
Power plant equipment configured in a 2 + 1 arrangement where two gas turbines and their two associated HRSGs are sending steam to the same steam turbine.

feeding only one ST. This configuration is known as 2 + 1. As a consequence, the electric generator/step-up transformer driven by the ST is roughly of the same power as the ones driven by the GT (Figure 8.33f). This simplification reduces the investment costs and reduces the electrical installation with the additional benefit of having similar electrical components. However, it is not rare that generators from different manufacturers are used because the GT supplier and the ST supplier could be different.

When aeroderivative GTs are used, sometimes a 3 +1 (i.e., three GT and one ST) configuration is used, with four identical generators. When only a GT is running, for instance during start up, the active HRSG is operating in a sliding pressure mode and the steam is delivered to the ST at about 65% of the normal pressure, with a minimum value of 50%. Sometimes, during the updating and modernization of existing power plants, the existing steam turbines can be reused. This involves a customized reconfiguration of the equipment in the plant as it is converted to operate in a combined cycle.

Main Equipment Blocks

Gas Turbine System The GT system consists of three main parts: the air compressor (axial type), the burners, and the turbine itself. The mechanical power generated by the turbine is partly (about 50%) used to drive the air compressor, so that the net generated power is the difference between the full-generated power and the power required by the air compressor.

The turbine types can either be "heavy duty" or "aeroderivative."

The heavy-duty design is of sturdy construction with the compressor running at the same speed as the turbine. The "aeroderivative" design is based on the design of aeronautical engines that have been modified for ground operation at continuous full load with no axial thrust. The turbine can be split into two parts, one rotating at the rated speed, and the other running at higher speed and driving the air compressor.

In general, these turbines are more delicate and more sensitive to changes in the fuel gas composition and to dirt deposited on the air compressor's blades, with subsequent decreases in the efficiency and in the generated power. The upper power for aeroderivative turbines is in the range of 40 MW at the shaft, but is expected to rise in the near future.

Gas Turbine Characteristics The characteristics of gas turbines are given at ISO conditions, which assumes no losses (i.e., no pressure drop in the inlet and outlet ducts), 59°F (15°C), 14.696 psia (101.325 kPa), 60% relative humidity. It is to be noted that the power generated by the turbine is highly dependent on air temperature. Starting from ISO conditions, roughly an increase of 35°F (20°C) of the inlet air temperature decreases the generated power by 10–13% and the efficiency by 3–5%, while a decrease of the air temperature of 20°F (11°C) increases the generated power by 6%.

In some instances, mainly for aeroderivative turbines, when the ambient temperature is high, the inlet air can be chilled before entering the air compressor to increase the throughput and efficiency. For heavy-duty GTs, the same effect can be obtained with so-called fog systems or wet compression.

The current generation of GTs has an efficiency of 33–38%, referred to as the low heating value of the fuel, when running in open cycle (i.e., exhausting to the atmosphere), thus approaching the overall efficiency of a traditional power station. The heat rate for heavy-duty turbines in sizes > 100 MW is in the range of 9,000–10,000 BTU/kWh (9,500–10,550 kJ/kWh), while for sizes of about 25 MW the heat rate is about 10,000–12,000 BTU/kWh (10,550–12,650 kJ/kWh). For aeroderivative turbines in sizes of about 40 MW, it is in the range of 8,200 BTUtu/kWh (8,650 kJ/kWh).

The GT is supplied as a package in which practically no custom tailoring is possible in the mechanical part because of the high development costs for the optimization of the blades and rotating speed(s) of the air compressor. The custom tailoring is limited to the extent of the supply for the auxiliaries, to the starting method, to the back-up fuel (if requested and feasible), to the voltages for the generator and auxiliary motors, to the cooling of the generator, to the available options for the governor, and so on. The turbine is optimized for operation at a power close to the base load (100% load) and has a turndown of approximately 100:55.

Temperature and Fuel Considerations The temperature in the combustion zone can be quite high [2280°F (1250°C) and even up to 2460°F (1350°C)] and is expected to increase even more with new developments in this technology (mainly due to the cooling of blades). This high temperature results in the nitrogen in the air combining with oxygen, with consequent production of NO_x concentrations that are well beyond the limits stated by almost all local regulations.

Additional nitrogen could be present in the fuel if it is obtained from coal gasification or from tail gas in some process industries. Hence, the need for NO_x abatement, which can be obtained in a wet process (by injecting steam or water into the combustion chamber) or by a dry method, with different flame configurations as a function of load. Nowadays, the latter solution is the preferred one even though it may involve flame instability at certain loads.

A dry method of NO_x reduction anticipates combustion with different flame shapes according to the load. Therefore, the flame shapes from the operating gas nozzles in a certain burner are different depending on the load. Changing from one shape to another involves flame instability, and therefore the loads corresponding to the change of flame shape should be avoided to prevent possible flame outage.

To limit the temperature in the combustion zone, a high excess air ratio is used, so that the oxygen content in the exhaust gas can reach 15%.

The preferred fuel is natural gas, but several types of turbines can accept either gas or liquid. For the dual-fuel-type designs, diesel oil is normally used as the back-up fuel. Some turbines can accept heavy oil as fuel, but the resulting emissions could exceed the limits stated in many states/countries. When liquid and gaseous fuels are used, the gas turbine should be equipped with a dual-fuel system, which is also needed when only gases are burnt, but these gases are characterized by much different "Wobbe Indexes."[1]

Operation The air compressors are very sensitive to pollution, particularly the aeroderivative machines. Thus, large intake filters are necessary. These filters should minimize the pressure loss and should be complete with cleaning systems and antifreeze protection. Depending on the type of turbine and air pollution, the air compressors need to be washed on-line every 8–48 hours, and washed off-line roughly every 1–3 months, when the power plant is out of service (Figure 8.33g). It is a desirable practice to perform the off-line compressor wash every time a long stop is required in order to improve the efficiency. The count down to the next washing is restarted after each wash-down.

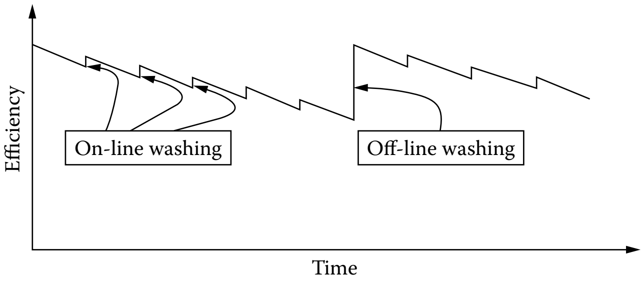


FIG. 8.33g
The gas turbine efficiency is slightly improved after each on-line wash, but periodic on-line wash is needed to return to full efficiency.

The GT is unable to start by itself and requires a launching motor, which can be electric or diesel, or use the electric generator as a motor, fed from the grid via a variable-frequency inverter, also called a static frequency converter (SFC). The starting system is selected according to the availability of the grid. If black starting (i.e., power is not available from the grid) is required, then a diesel engine or a diesel generator feeding an electric motor should be used.

When the GT runs at rated speed and without meeting any load, it is said to be at Full Speed No Load (FSNL). For many GTs this is a stable operating condition, at which the generated power is nil but the exhaust flow and temperature are still relatively high as compared with base-load conditions. For large turbines, the FSNL exhaust mass flow is in the range of 70% of the maximum flow at ISO conditions, and the temperature is roughly 50%, even though the available temperature drop in the HRSG is about 25% due to the increase of the temperature at the stack.

Some new turbines feature at FSNL an exhaust flow of approximately 60% of the nominal and a temperature about 65% of the nominal. The enthalpy per mass unit of flue gas at FSNL is greater than 45% of the one at base load, so that the total enthalpy transferred from the GT to the HRSG at FSNL exceeds 30% of base load and impacts on the start-up procedure of the HRSG. The FSNL-generated HP steam flow is in the range of 20% at a pressure in the range of 20–30%, a temperature (°F) of about 50%, with the superheating temperature in the range of 30%. With the new GTs the SH steam flow is roughly 30% at a temperature of 70% approximately. The fuel gas consumption at FSNL is about 25% of the consumption at base load.

Pollution and Safety At FSNL, the NO_x and CO contents can be beyond the limits accepted by the regulations, so that the FSNL condition can be accepted only as a transient to warm-up of the HRSG and the ST. When a turbine cannot run at FSNL, but needs to feed some load (about 20%), this situation is called minimum technical load. Similarly, at these conditions the emission limits could not be met.

A minimum environmental technical load (about 50–60%, depending on the GT and the local regulations) can be defined as well, where the GT emissions are in line with the regulations

---

[1] Wobbe index W is the ratio between the heating value H_s in standard volumetric units and the square root of the specific gravity G.

$$W = H_S/\sqrt{G} \qquad \textbf{8.33(1)}$$

or can be brought in line by means of selective catalytic reduction (SCR) in the HRSG, if present.

The GT is supplied with its own governor that takes care of all safety and control functions including antisurge control of the compressor, inlet guide vanes (IGVs) control, burner control, start-up and shutdown sequences, and excessive vibrations. For an I&C engineer, the GT is almost a black box interfaced with the overall plant DCS via a serial link (simple or redundant) and relatively few directly hardwired commands.

In several instances the old turbine governors had a standard set of information transmitted over the serial link, and only few could be custom-tailored. Out of the standard set of information, a certain amount was not needed for the specific plant and, therefore, was discarded in the GT governor or in the DCS.

All safety functions of the GT are performed by the turbine governor that normally has a 2oo3 (two out of three) or a 1oo2D (one out of two diagnostics) configuration, which are usually independent from the control functions. The control functions are performed in simple, redundant, or 2oo3 configurations that ensure the degree of safety and availability requirement, which should be consistent with the size and criticality of the plant. The safety functions are also performed in the control processors and are used as back-up for the safety functions performed by the dedicated processors.

The governors are provided with their own operator interface and with processors that store information on the behavior of the turbine, including the sequence of events leading to a shutdown. The governor includes comprehensive self-diagnostics that allow easy maintenance while the GT is running, while still keeping all of the protections active.

The GT runs in temperature control mode when it is at base load. When it is ramping up, it is in speed control, with the temperature of the first row of blades constraint limiting on the generated power.

The GT is normally enclosed in an acoustic cabinet, complete with ventilation, fire- and gas- detecting, and fire-fighting systems.

Steam Injection Gas Turbine (STIG™) When the power generated by the gas turbine (GT) is low (< 15 MW), the combined cycle configuration discussed earlier would be uneconomical. For such applications, small GTs have been developed that can accept the partial or total injection of the steam that is generated in the associated HRSG upstream from the turbine blades of the GT. This results in increasing the mass of the flue gas in the turbine and the generated power, without the need of a small steam turbine.

Part of the steam is injected in the burner chamber for NO_x abatement in a wet mode on a preferential basis. This type of turbine is suitable for small cogeneration plants requiring a variable quantity of steam, while the surplus is sent to the GT. The sizing of the demineralization plant should consider the large amount of demiwater wasted to the atmosphere.

Heat Recovery Steam Generator The HRSG receives the exhaust gases from the GT discharge. The exhaust gas, flowing in counterflow with respect to the steam/water coils, cools down by transferring heat to steam/water. The flue gas temperature at the stack is about 230°F (110°C), even though lower temperatures [200°F (93°C)] can be used if the fuel gas is very clean and sulfur-free. The HRSG is, therefore, similar to a heat exchanger in which the shell side carries the flue gas and the various sections of the tube side carry steam or water.

It has also the characteristics of a boiler because there are one or more steam drums, where the generated steam is separated from boiling water before entering the superheaters. The HRSG can be horizontal or vertical, according to the direction of flue gas path.

The horizontal ones have a horizontal path for the flue gas flow, and the steam/water tubes are vertical, normally with natural circulation in the evaporator(s). The vertical ones have a vertical path for the flue gas flow, and the steam/water tubes are horizontal, thus requiring assisted circulation in the evaporators, at least during the start-up period. The horizontal HRSGs are the most common, with the vertical ones mainly being limited to the revamping of existing boilers or to installations where space is very tight.

Pressure and Temperature Levels The HRSG can have one, two, or three pressure levels according to the size of the plant. For plant sizes of 200–400 MW per line, the pressure levels used are HP, IP, and LP (high pressure, intermediate or medium pressure, and low pressure). Plants down to 30–60 MW usually have two pressure levels (HP and LP), and smaller units only have one pressure level. Sometimes, with three pressure levels, the LP section produces the steam needed for deaeration only.

The following tube banks are used for each pressure level (starting from the GT discharge): 1) steam superheaters, 2) evaporator, and 3) economizer. When more pressure levels are used, the banks of the various levels can be intermingled as to maximize the heat exchange efficiency, taking into consideration the differential temperature between the flue gas at that point and the steam/water in the tubes. In large boilers, the two HP superheaters that are in series can be split in two to four parallel banks. In the large HRSGs the reheating coil for the IP steam can be derived from the steam turbine and mixed with the IP steam, which is generated in the HRSG.

The differential temperature between the exhaust from the GT and the HP superheated steam is called the "approach" temperature, and it is in the range of 50°F (30°C). The differential temperature between the flue gas at the outlet of the evaporator section (inlet for water) and the saturated steam temperature is called the "pinch point" and is approximately 15°F (8°C).

Starting from the turbine discharge, the HRSG casing is mechanically made of a diverging cone to reduce the velocity of the exhaust gases, plus a rectangular section (internally

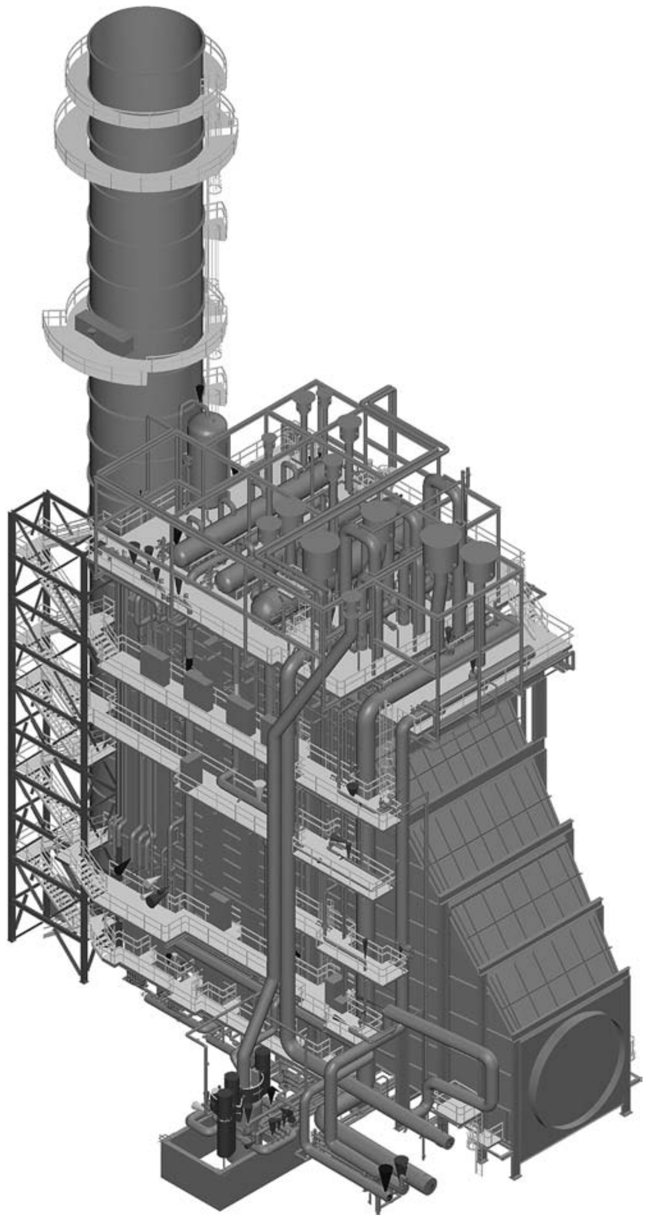


FIG. 8.33h
Heat recovery steam generator (HRSG) generates three pressure levels of steam. (Courtesy of Nooter Eriksen CCT.)

insulated and lined) housing the steam/water coils and the stack. Figure 8.33h shows a horizontal HRSG, with three pressure levels and integral deaerator.

Design Features In more recent times it has become a requirement to insert an empty module in the flue gas ducts of large HRSGs, which serves the possible future installation of a selective catalytic reduction unit for further NO_x abatement.

This empty module is installed in the expectation that in the future more stringent regulations will come into force. This empty module is positioned where the flue gas temperature is about 660–715°F (350–380°C).

Sometimes, a spool piece for future addition of an oxidation catalyst for CO abatement is included for the same purpose as the SCR and located in the same position. Nowadays in North America, almost all HRSGs are fitted with SCRs and occasionally with CO catalyst, while in Europe only provisions for SCRs are required.

In addition, many installations include a silencer in the stack or at least a dummy piece for the future installation of one. If the plant is to be operated in an intermittent mode (e.g., shut down overnight), a damper in the stack is required to protect against excessive cooling of the unit. The damper should be designed to fail open on high flue gas pressure, which failure position is normally obtained by counterweights, or with offset blades so the damper will open without the use of the actuator. Sometimes, actuators are included in order to have the maximum flexibility.

The pressure drop across the HRSG on the flue gas path is in the range of 8–15 in. (200–375 mm) water column. This pressure drop is the back-pressure of the GT and influences its generated power and efficiency by 1 and 2%, respectively.

The water/steam coils are helically finned on the outside to improve their heat exchange efficiency. In some cases, the HP superheater tubes can be nonfinned. Their construction material is carbon or alloy steel, as a function of the pressure/temperature rating required. The coldest coil (LP economizer or feedheater) can be stainless steel to protect it against corrosion on the inside, which might occur if the feedwater is not deaerated. If stainless steel is used, it will also protect against corrosion, which can be caused by acid condensation from the flue gas on the outside.

In some instances, for small plants where the GT exhaust temperature is relatively low, the HRSG can be designed to run dry (i.e., without damage if the water supply fails) in order to comply with some national regulations for unattended boilers.

The HRSGs are provided with a set of motor-operated valves that are installed in the steam and water lines. Each superheated steam line leaving the HRSG is provided with a shut-off valve and, in some cases, also an on/off bypass valve (mainly on HP) to intercept the steam to the user (turbine or industrial load).

The feedwater inlet lines to the economizers are also provided with on/off shut-off valves. Having these shut-off valves allows the “bottling in” of the HRSG by closing all inlet and outlet lines, thereby to keep the boiler pressurized when the shut-down period is expected to be short. Additional motor-operated valves are used to remotely and automatically operate the drains in the superheaters.

The HRSG also includes a pressurized blow-down tank and an atmospheric blow-off tank, and is also equipped with chemical injection pumps to keep the water and steam characteristics within the correct parameters required to maximize the life of the ST and the boiler itself. The HRSG is also equipped with nitrogen connections for purging (dry lay-up) to prevent corrosion in case of long shut-down periods.

Alternative Steam Generator Designs Other steam generator possibilities include an HRSG with the HP section only and a high rate of post-firing. In this case, the water/steam cycle is very similar to that of the standard fired boilers and utilizes water heaters to improve the overall efficiency. In such boilers, at some loads the efficiency can be higher than using a conventional HRSG.

Another possibility is the once-through steam generator OTSG), which is used mainly in combination with GTs in the capacity range of 40–50MW, although larger units are under development. These boilers, built with Incoloy 800 and 825 tube materials, operate without steam drums or blowdown systems and thus require a simplified control strategy.

Their flue gas flow is vertical upwards and the tubes are horizontal, thus allowing for a reduced footprint. These boilers feature the possibility of running dry (if required) and have very short start-up periods due to the thin wall of the tubes and to the missing steam drums. They are delivered in a few prefabricated pieces so that the field erection and installation takes place in a short period of time.

Steam Turbine Steam turbine controls are discussed in Section 8.38. The largest steam turbines today that can be fed by a single gas turbine/HRSG are in the range of 150 MW. Therefore, these steam turbines are still are relatively small for power generation. Only when one ST is fed by two HRSGs (in a 2 + 1 configuration[2]) can the steam turbine's rated power reach 250–300 MW.

The steam turbine is fed with HP superheated steam and is designed to also accept IP and LP steam from the HRSG and also to extract full steam flow at IP level in case of reheating. In addition, if the plant is cogenerative, suitable (controlled or noncontrolled) extractions shall be planned for. For example, if there is a requirement to feed thermal users with steam, it is possible to get this steam through one or more extractions (at the required pressures) from the steam turbine.

In many instances, the LP steam as generated in the LP steam drum has the correct pressure, but not in the required quantity. Therefore, there is a need to send steam from the LP section of the HRSG to the ST, or directly to the users, or to extract steam from the ST to comply with the demands of the users.

The vacuum condenser can be water- or air-cooled. The present trend is in favor of air-cooled condensers, particularly if the water availability is limited. The use of air-cooled vacuum condensers reduces the size of the cooling water towers required for the plant (still needed for cooling the rotating machinery) if closed circuit cooling is selected. This approach also eliminates the need for large flows of cooling water and the thermal pollution of rivers, lakes, or the sea, if open circuit cooling is used. The air-cooled condensers slightly decrease the efficiency of the steam turbine, mainly in hot climates.

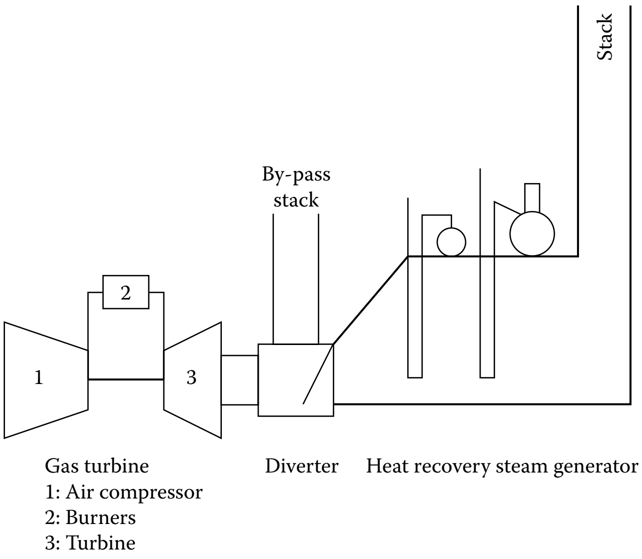


FIG. 8.33i
Diverter provides the flexibility to send the gas turbine exhaust gases to the HRSG or to the atmosphere.

When the plant is located in built-up areas, there might also be an atmospheric condenser that receives the steam that is vented during the start-up to decrease the noise level at the plant. The auxiliary circuits of the steam turbine are the lube oil, control fluid, and vacuum system for the condenser. The vacuum system can be equipped with steam ejectors or with liquid ring pumps (vacuum pumps), or a combination.

Bypass Stack and Diverter In some instances, when the electric power generation is a must, it should be possible to run the gas turbine in open cycle and exhaust the flue gas to the atmosphere instead of sending it to the HRSG, regardless of the overall efficiency (Figure 8.33i). This requires a bypass stack and a diverter that closes the path to the HRSG and opens it to the atmosphere through the bypass stack. The diverter is connected to the GT exhaust duct before the diverting cone of the HRSG, and this implies that the GT has to meet the plant emissions limits, as any SCR in the HRSG is also bypassed.

Throttling by the diverter could also be used to control steam generation in the HRSG. This configuration is rare and limited to STIG applications or to applications in which the steam is generated only for process use and the demand for it is low compared to the amount that could be generated, based on the prevailing demand for electricity (e.g., some desalination plants in the Gulf area).

The most important characteristic of a well-designed diverter is its ability to completely switch the flue gas from the bypass stack to HRSG, under all operating conditions. This requires a good seal, so that the GT exhausts fully to the HRSG without loss of efficiency due to leaking hot flue

---

[2] 2 +1 configuration means two GTs/HRSGs feed only one ST. For more details, see later under the heading "Alternative Configurations."

gas. For this purpose, seal fans are provided that blow air into the gap between the diverter blade and the wall.

Post-Firing (Supplementary Firing) In some cases, it is wise to boost the steam production by supplementary firing in the flue gas duct, at least on a temporary basis. This is possible because of the high oxygen content in the exhaust gas from the GT. The burners are normally in-duct burners located at the end of the diverging cone, upstream of the final HP superheater or between the final HP superheater and the IP reheater. They are fed preferably with natural gas in order to keep the emissions within the regulatory limits and to prevent fouling of the finned tubes.

They require pilot burners and ignitors and a burner management system (BMS). The thermal power delivered by the burners is normally capable of increasing the generated steam in the range of 10–30% so that the overall electric power generated by the GT/ST increases roughly by 5–10%. Post-firing decreases the overall efficiency of the plant but allows it to cover peak demand. It can also be advantageous when electricity can be sold at different prices as a function of the time of the day.

Supplementary firing is inhibited when the GT load is lower than a preset value, ranging between 50 and 80%. The latter value is typical for aeroderivative GTs. These limitations are due to several factors like exhaust flow distribution (not always available at reduced load), flame shape that is longer at reduced flows, and low exhaust temperature inducing increased firing to compensate for reduced energy in the GT exhaust. Allowing the post-firing to be active at lower GT loads could cause excessive temperatures that could damage the HRSG.

Fresh Air Firing (Complementary Firing) When a combined cycle is used for cogeneration, the availability of the steam supply can be very critical for the process and it can be requested that the steam be not used for any other purpose. This requirement can be met by using an auxiliary boiler as a back-up or by fresh air firing in the HRSG.

Fresh air firing requires a diverter to close the outlet duct from the GT, a fresh air fan, and in-duct burners in the diverging cone upstream of the final superheater. All of these system components should be sized to meet at least the minimum contractually guaranteed steam quantity and deliver it for industrial use, while still keeping a good distribution of gases in the flue gas path (Figure 8.33j). The flue gas quantity cannot be decreased too much with respect to the normal flow without impairing the heat transfer. The airflow is normally at least 70% of the flue gas flow at base load.

In this configuration, the HRSG behaves as a fired boiler even though the efficiency is much lower due to high excess air heat losses and a poor thermal cycle. The superheat temperature of the generated steam is usually not suitable for steam turbine supply and, therefore, has to be wasted to avoid damages.

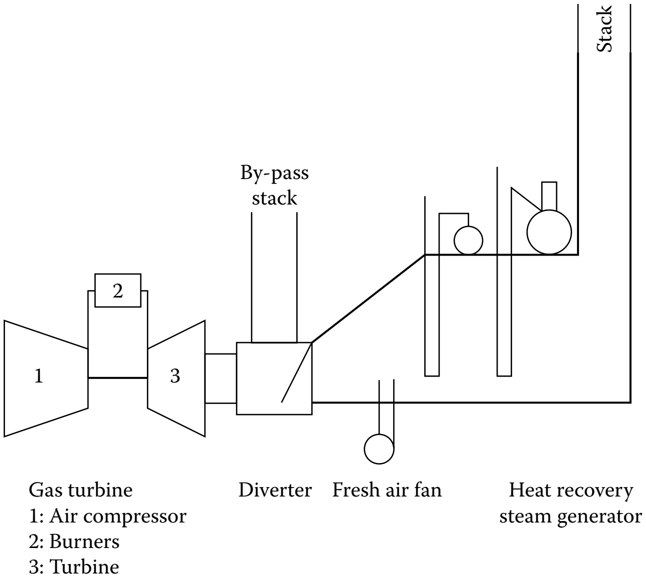


FIG. 8.33j
Equipment configuration when a fresh air fan is added to the combined cycle for complementary firing.

To reduce the pollutant emissions to the allowed limits, it is required to use flue gas recirculation (FGR).

Electric Generator The electric generator is a three-phase unit and can be air- or hydrogen-cooled. The excitation system is static and the voltage/cosφ control is obtained via the automatic voltage regulator (AVR) that is hardwired to the DCS or can be serially linked to the DCS, which is a less frequently used arrangement.

In large groups, at start-up the generator is fed with variable frequency and acts as the launching motor.

The lubricating system is normally common with the driver's system.

Auxiliary Systems

Steam and Water Cycle The steam and water cycle is dependent on the number of pressure levels of the HRSG. If the deaerator is integral with the LP steam drum, the LP drum has also the function of being the suction drum for the boiler feedwater (BFW) pumps that are working on a duty/standby basis. If there are HP and IP sections, the BFW pumps can be multiple-stage centrifugal pumps with an intermediate discharge for the IP section. Otherwise, the pumps have only a discharge nozzle for the HP section.

If required by the pump characteristics (and normally it is), an automatic minimum flow bypass should be installed on the HP discharge nozzle of the pump. It is preferable to use a four-way valve with the manual port used during pre-commissioning and commissioning to extend the life of the automatic port. Alternately, a flow control loop could be

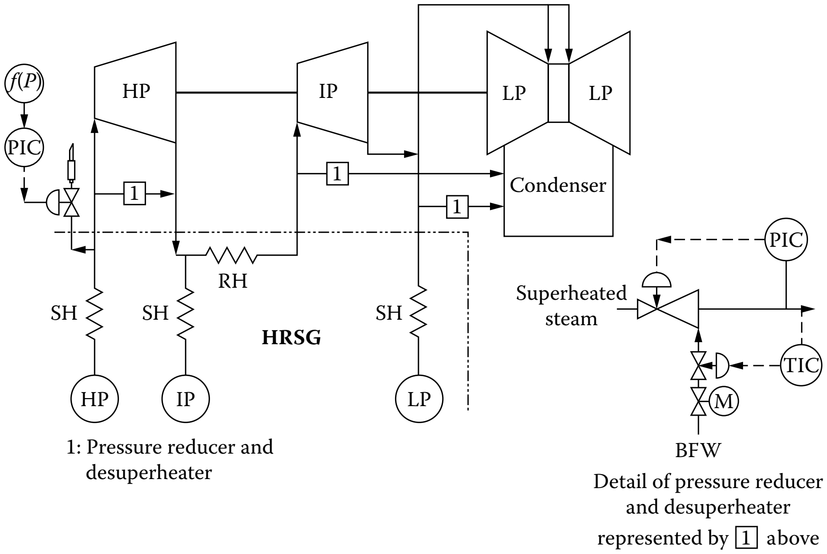


FIG. 8.33k
When high, intermediate, and low pressure (HP, IP, LP) steam turbines and plant users are combined with reheat and superheat (RH, SH) coils, bypasses containing pressure reducers and desuperheaters (1) need to be provided for start-up.

provided to maintain the water flow rate through the pump at a value that is above a preset minimum value and returning the unwanted flow to the suction drum (could also be the LP drum, in the case of an integral deaerator).

The superheated steam is conveyed to the steam turbine that should be bypassed during the start-up phase or in case of turbine shutdown or load rejection. A typical bypass arrangement is shown in Figure 8.33k.

The bypass arrangement includes 1) An HP bypass from HP header to IP header (cold reheat side if reheating is implemented) or LP header or condenser if only two pressure levels are required. 2) IP bypass from IP header (hot reheat side if reheating is implemented) to the condenser or to LP header (less frequently used arrangement). 3) LP by-pass from LP header to the condenser.

Each bypass requires a pressure reduction and desuperheating stage with boiler feedwater or condensate supplies at appropriate pressures so that the physical conditions of the reduced steam are fully compatible with the conditions of the downstream header.

In case of a 2 + 1 configuration, each HRSG has its own turbine bypass system installed upstream of the steam stop valves, so that the starting of the boilers can take place independently. Once the boilers are both on-line and paralleled on the steam header, these bypass valves are operated simultaneously, controlled by the steam header pressure and triggered by the steam turbine trip or load rejection.

An important decision to be made at an early stage of the design that affects the start-up procedure of the plant is how the steam lines are to be heated. It makes a difference if the HRSG starts with the steam stop valves closed or with the steam stop valves fully open and the steam intercepted by the ST admission valves. The latter configuration is preferred when long steam lines are used (e.g., in revamping old conventional power stations).

To keep the required steam purity, a small percentage (1–3%) of the water in the steam drums should be discharged to continuous blow-down and to discontinuous blow-off (as covered in more detail in Section 8.6 on boiler control). For large boilers there is a pressure blow-down tank into which the HP and IP steam drums drain. In addition an atmospheric blow-off tank is also provided to receive the water from the blow-down tank plus the drains from the LP drum and the blow-off from the HP and IP drums.

Small boilers only use one atmospheric tank. The water is cooled to an admissible temperature before dumping it to water treatment or collecting and sending it to the demineralization unit to minimize the quantity of the effluents.

AUXILIARY EQUIPMENT

Demineralization Plant

The water needed for filling the HRSG and as make-up water during normal operation is generated in a demineralization plant. The demineralization plant is usually controlled by its own PLC, which is serially interfaced with a redundant link to the DCS, but sometimes is controlled directly by the plant DCS system.

The demineralized water is stored in a tank that should be sized sufficiently large to provide water in case of disruption in the production. It should also store enough water to supply the quantity needed for pipe blowing in the precommissioning stage, without the need for waiting for the production of new water. This consideration can be the basis for sizing the demineralized water storage tank.

Fuel Gas System

The GT is best utilized when fuel gas, normally natural gas, is burned, because this is the cleanest available commercial fossil fuel and the one releasing the least amount of CO_2. Sometimes, natural gas is distributed at a pressure below that required by the GT, which normally requires 300–650 psig [20–45 bar(g)], as a function of its model.

If the available supply pressure is low, the addition of a compression station is required, which is normally composed of two reciprocating compressors or one centrifugal compressor. If two reciprocating compressors are used, one spares the other. They are followed by a multipurpose knock-out drum, serving the purposes of 1) pulsation damping, 2) residual liquid separation, and 3) surge tank duty to allow for switching the compressors without disrupting the gas supply to the GT.

In the case where the pressure of the natural gas is higher than required, it is necessary to provide a reducing station, which can require preheating with hot water or steam to prevent freezing caused by adiabatic expansion and to keep the gas above the dew point temperature. In this case electric heating or auxiliary hot water generators (boilers) could be required at least for the start-up phase, while during normal operation the heat from the HRSG is used.

The temperature decrease caused by the pressure reduction can be derived from pressure-enthalpy diagrams (see Reference 1 for pure methane). For this reason, the characteristic of the natural gas that is distributed in the grid is very similar to that of methane. As an example, a pressure reduction from 1000 psia (69 bara) at 60°F (16°C) to 435 psia (30 bara) would decrease the temperature to 30°F (–1°C). To obtain a gas at 435 psia (30 bara) and 95°F (35°C), the gas should be preheated to 122°F (50°C), with a heat delivery of 41 BTU/lb (95 kJ/kg) of gas.

The GT requires a very clean fuel gas, thus the gas at proper pressure needs to be further filtered in a location close to the GT. The line could be stainless steel downstream of the filters, to prevent any entrained rust from entering the GT. For safety reasons, servo-actuated double block and bleed valves should be installed upstream of the filters, and they should be independent from the valving supplied with the GT. In addition, the block valves should be fire-resistant.

A point to be checked with the GT manufacturer is the temperature at which the gas is delivered to the turbine. Some gas turbine manufacturers constrain the admissible temperature during the various stages of operations, with special regard to the start-up and shutdown conditions. Fuel gas heaters are usually fitted to the GTs in order to improve the efficiency of the GT cycle. Typically, the gas can be heated up to 365°F (185°C). The heating medium is hot water that is usually received from the IP economizer section of the HRSG, but sometimes from the condensate preheater.

In any case, it is essential that the water pressure be higher than the fuel gas pressure so any leakage that occurs will be water into gas and not vice versa. Some turbines are also very demanding in terms of the steadiness of the controlled gas pressure and its admissible rate of change (pulsation). These constraints impact on the gas system design and relevant instrumentation.

Liquid Fuel System

Some GTs can burn liquid fuel, even though this practice is decreasing because of emission constraints. However, in some countries this is still done. The use of liquid fuel requires careful discussion and agreement with the GT manufacturer on the features and characteristics of the fuel used. Some common requirements include 1) water shall be drained and shall not be fed to the turbine, 2) sediments shall not be pumped to the turbine, 3) the fuel shall be filtered to remove all particles that could clog the fuel injectors, and 4) the storage tank and lines shall be heated if the fuel is too viscous.

Also in the case of liquid fuels, the target is to obtain the cleanest possible fuel to extend the life of the turbine and to obtain a smooth operation. In particular, sodium, lithium, and vanadium should be eliminated, because they are highly corrosive to the GT blades at the combustion temperatures.

Closed Circuit Cooling Water

Water-cooled condensers, users like lube oil, control fluid, and bearings, all require cooling water. The cooling water can be supplied in two closed circuits, both being fed by the cooling tower. One circuit is for the water-cooled condensers and the other for all process coolers and other users (rotating machinery).

The cooling towers can be of modular design, equipped with double-speed air fans (see Sections 8.16 and 8.17). The water-cooled condenser can be cooled by river (seldom sea) water, which requires proper intake works and adequate design to prevent local overheating of the river (or sea). With a cooling water temperature rise of 18°F (10°C) in the condenser, the flow of cooling water is about 55 times the flow of the exhaust steam to be condensed.

If an air condenser is used, the closed-circuit cooling water system becomes much smaller, because the amount of water needed in the rest of the plant is a relatively small percentage of that needed for the water condenser. The air condenser is usually equipped with 2–20 double-speed fans, as a function of the steam flow that is to be condensed, to the required vacuum and to the ambient temperature. In other installations, variable-speed fans are used, which are provided with frequency converters capable of modulating their loading between 50 and 100%.

FIG. 8.33l
The control room of a power plant consisting of four 380 MW groups of generators, located at La Casella, Italy. (Courtesy of ENEL.)

CONTROL EQUIPMENT

Central Control Room

The control system in a combined cycle power plant having one generating group normally consists of a DCS system, provided with two operator stations, each with two CRTs, plus a keyboard, trackball or mouse, and printers.

When more groups are installed, each group can be equipped with three CRTs and possibly with large screens on the wall (Figure 8.33l). The operator stations can be PC-based in a client/server configuration. According to the size and criticality of the plant, the server can be simple or redundant. In general, the cost of controls is small in comparison to the mechanical equipment costs or the costs associated with plant shutdowns, loss of production, and expenses associated with restarting. For these reasons and for increased safety, usually redundant or voting systems are used.

The multiloop controllers are used in redundant or fault-tolerant configurations, and serial gateways should be provided to connect the GT and ST governors, the AVRs, the demineralization water plant, and possibly the electrical protections. In some instances, the ST governor can be part of the DCS itself with a dedicated redundant multiloop controller similar to the ones used to control the HRSG and the cycle. In other cases, the GT governor and the ST governor can be in separate hardware packages.

DCS and I/O Configuration Integration and properly designed interfacing between the DCS, PLC, and other digital control packages is essential. The serial links should be redundant to ensure the maximum operating continuity. If not all external devices are available with redundant links, it is recommended that they be provided. The system bus and the I/O buses should also be redundant, with the goal of guaranteeing the maximum uptime.

The number of continuous and discrete (on/off) I/O points used is in the range of 5000, and the discrete I/O outnumbers the analog. The input signals are often duplicated for control purposes or tripled in a voting system for safety functions, and this requires careful attention in point assignments to the input cards and of the input cards to the multiloop controllers (see Reference 2). If the sequence of events control function is carried out in the DCS, the appropriate input cards have to be correctly anticipated or a dedicated system must be installed and connected to the DCS via a serial link (see Reference 3).

In recent years, the connection of the plant DCS to the company intranet has become common practice, in order to allow for a remote overview of the plant's behavior and to remotely view the main operating parameters and to assist in correctly managing plant maintenance. However, no commands can be issued from the overview location.

When a company owns or operates several power production plants, this configuration allows for the gathering of information for further processing by management or by the engineering support team in real time. Adequate firewalls must be included to prevent unwanted access from hackers or the injection of viruses.

In a few cases, power plants are mechanically designed for operation from a remote location and are unattended. This means that the control room in the plant is equipped as if the plant were attended in order to allow for temporary operation mainly during precommissioning, commissioning, start-up, and tuning, and in case there is a plant upset.

Furthermore, there can be additional redundant links to other control room(s), located a few miles away. Of course, extensive additional information systems serve the goal of plant safety. These systems utilize the fire- and gas-detection and -protection systems and gate control, flooding, and so on. The fire-detection and fire-fighting system cabinets are located in the control room behind or to the sides of the operator's control desks, in order to allow for immediate response to any dangerous situation.

The software packages supplied with the DCS, in addition to the standard operating controls, also include such data sources as the steam tables, the sequence of event (SOE) module, post-trip review modules, and possibly the performance evaluators and stress evaluators.

Control Desks, Operator's Displays Operators find it very useful to have one CRT on the control desk directly connected to the GT governor and another one directly connected to the ST governor, in order to eliminate the transmission delays mainly during start-up and shutdown, or during plant upsets.

The control desk also houses the PC dedicated to continuous emission monitoring system (CEMS) plus the remote direct reading level gauges for the steam drums, where they are required by local regulations.

The engineering station can be located on the same control desk or in a separate room, according to the company policies and also as a function of the available space. It is, however, to be noted that the operations team is very small, has only a couple of operators per shift, a few other maintenance specialists, plus the plant manager. For example, in a power plant with four combined cycle units, the operating team is composed of two operators per shift, each running two units, and a common supervisor.

The control system is also used for the management of spare parts inventory and maintenance of both main machinery and the smart transmitters. The aim is to perform the maintenance on a predictive basis, i.e., when the behavior of a component shows a degradation that could result in a fault. In this case, the use of fieldbus technology connecting field devices is very useful, because of the large volume of diagnostic information to be processed and conveyed to maintenance engineers.

The control system sometimes also has the task of storing technical bulletins and documentation pertinent to the various components of the plant in order to facilitate the finding of information as required.

Monitoring and Control Loops

In a combined cycle power plant, these functions are quite simple and are similar to the boiler and the steam turbine controls that are covered in Sections 8.6 and 8.38 in this chapter. In contrast to boiler controls, HRSG controls do not have air/fuel ratio, steam pressure, furnace draft, and so on, so they are relatively simpler. In the paragraphs below, those measurements and controls will be discussed that are typical to combined cycle controls, or are different from the ones used on boilers and steam turbines.

Fuel Gas Heating Value The heating value of the gas delivered to the users is variable with time, and in several countries the heating bill reflects not only the quantity of fuel delivered, but also its higher heating value. The gas distributors provide average daily values of the upper heating value, which the users should check.

In addition to the commercial issues, this measurement, related to the lower heating value, is very important from an operational point of view, especially when the GT is the aeroderivative type. An increase in the fuel gas heating value increases the flame temperature and can overheat the first row of turbine blades, causing a sudden shutdown of the GT, if the firing rate is not decreased. The consequence of a shutdown is the loss of at least 1 hr of electric production, and if the plant is cogenerative, the steam production is also disrupted.

To prevent this kind of shutdown, a chromatograph or a calorimeter (or both) are installed on the fuel gas. For a detailed discussion of calorimeters and chromatographs, refer to Sections 8.8 and 8.12 of the first volume of this handbook, respectively. If the fuel gas heating value increases, the set point of the GT governor is automatically lowered to reduce the combustion temperature and, thereby, protect the system from a shutdown.

HRSG Start-Up Vent Valve The main purpose of this valve, which is installed downstream of the HP superheater (see item 1 in Figure 8.33k), is to control the pressure in the steam drum during start-up and to let it rise slowly so that the corresponding rate of rise in steam drum temperature will be acceptable. To perform this task, the HP vent valve works in combination with the ST bypass valves.

For large steam drums with a wall thickness > 4 in. (100 mm) operating at 1500 psi (100 bar) and over, the acceptable temperature gradient is 6–11°F/min (3–6°C/min), while for smaller drums operating at 900 psi (60 bar), the temperature gradient can be 15°F/min (8°C/min). The lower figures are used when the HRSG is subject to frequent starts and stops.

The problem that is typical of the HRSG is that during the start-up phase, the heat delivered to the HRSG is quite high even though the GT runs at its minimum capacity, i.e., at FSNL or minimum technical load. The start-up vent valve should be capable of operating with an upstream pressure from nil to the design pressure, with temperature varying from 212°F (100°C) to the superheated steam temperature at the nominal working pressure. The valve has critical pressure drop and should be designed to abate the noise pressure level down to 80–85 dB(a), according to most contractual specifications.

For the purpose of noise abatement, this valve cannot be considered as operating in emergency only, like safety valves, because it operates for long periods of time during all start-ups, and the start-ups can be relatively frequent, at least for the need of off-line washing the air compressor blades. For detailed information on control valve sizing and valve noise calculation, refer to Sections 6.14 and 6.15 of this volume.

Because this valve operates with both variable upstream pressure and variable flow, as a function of HRSG operation and of the requirement to keep the temperature gradient within the allowable limits, it is not easy to prepare sizing data of process conditions describing the pressurization operation in order to size the valve.

The sizing of this valve is dependent upon a number of factors, only one of which is the start-up condition.

It is possible that the plant should be designed to maintain the GT output in case of ST failure or in case of ST condenser failure. In this case, the valve should be sized for 100% unfired duty because the possible supplementary firing is shut down if the vent valve opens. If the plant can be started up by using the ST bypass valves only, because auxiliary steam is available to bring the condenser on-line, then this vent valve can be omitted. Experience has shown that this valve works properly to cope with the start-up needs, if it is sized for at least 30–35% of the boiler steam capacity at nominal conditions.

In some cases this valve can be a motor-operated (inching type) globe valve. An additional requirement for this valve often is to prevent the blowing of the safety valves that are protecting the drum and superheater coils from overpressuring in case of ST trip or load rejection. To be able to do this, the start-up vent valve should be fast opening (opening time of 2–3 sec) and stay open for about 5 sec until the ST bypass valves (if sized for 100% flow) can handle the upset caused by the ST trip.

If there are no ST bypass valves, as in smaller plants or process steam generators, the valve should be able to handle the full steam generating capacity. For high-pressure boilers, the flow can be split into two valves, one suitable for the controlling the start-up conditions up to a pressure of 30% of the nominal one, and the other, used for emergency only, to prevent the blowing of the safety valves. When operating at full pressure, the start-up valve is inhibited in order to prevent uncontrollable level upsets. Whatever valve size is selected, it is essential to consider the amount of make-up water required.

Steam Drum Level Control

The level control of the steam drum of an HRSG is very similar to that of fired boilers. Up to 30% of nominal steam flow, usually a single-element controller (level only as shown in Figure 8.6ww), is used. There have been cases where the level control of large HRSGs was stable, while its load was over 30% and the level was controlled with a single element controller.

Above 30%, the loop usually is bumplessly transferred to three-element control (water level, steam, and water flow as shown in Figure 8.6yy). The phrase *three-element* is used to define a mass balance between input and output flows to/from the steam drum, which are feedback-corrected by the level measurement. In more elaborate algorithms, more than three measurements can also be used.

The feedwater control valve can be located upstream or downstream of the economizer, as a function of the designs of the various HRSG manufacturers. Several HRSG manufacturers do not allow the economizer to steam, and therefore the control valve is installed downstream of the economizer. This solution is very effective in preventing steaming but imposes a higher design pressure on the economizer itself with the associated costs.

The valve in these installations is quite critical because it can be subject to flashing and cavitation. The installation of the control valve at ground level can help reduce this problem, provided the piping layout allows for it. Furthermore, during start-up, due to the water trapped in the economizer, an overpressure builds up, because the boiler feedwater valve is closed due to the minimal amount of steam generation in the steam drum, and the check valve on the BFW pump discharge is closed.

This overpressure needs to be relieved without opening the safety valve through an antiflash valve that should open on the pressure threshold. Some units are designed to maintain a minimum flow through the economizers to prevent this, and the excess water is dumped from the steam drums to a condensate-recovery system.

Some HRSG suppliers accept a little steaming in the economizer and install the control valve upstream of the economizer with an accurate heat-transfer design to reduce steaming during some transient conditions. The valve, in this case, is much less critical as it handles “cold” water without flashing or cavitation problems.

Mainly on the HP circuit, the BFW control valve sometimes requires a smaller valve in parallel for start-up. This start-up valve should be sized for a flow equal to approximately 20–35% of the nominal flow (in order to handle the water flow required during start-up). It should be sized to operate at a pressure drop that is much higher than the one during normal operation through the main control valve. This is because the flow is very low, the BFW pump is practically working at its shut-off pressure, while the friction losses in the line and coils are minimal, and the pressure in the steam drum is much lower than during normal operation.

As the power station can operate under several conditions, it is suggested that the suppliers who bid on the BFW control valve be provided with data on all the different operating conditions. This way they should be able to check the

TABLE 8.33m

Illustration of the Process Data Required for the Proper Sizing of an HP Boiler Feedwater Control Valve

Case	Q Steam (lb/s)	Q Blowdown 2%	Q Sizing (lbs/s)	T (°F)	P Pump (psia) Delivery el. 3 ft.	Dp Line and Head (psi)	P Upstream Valve (psia) el 100 ft.	P Downstream Valve (psia) Pin-eco (psia)	DP (psi) Operation
1	171.3	3.4	174.7	348	2143.6	60	2083.6	1856	227.6
2	169.8	3.4	173.2	348	2146.5	60	2086.5	1813	273.5
3	166.6	3.3	169.9	348	2156.7	60	2096.7	1755	341.7
4	146.0	2.9	148.9	348	2255.3	60	2195.3	1552	643.3
5	146.1	2.9	149.0	348	2265.5	60	2205.5	1581	624.5
6	127.2	2.5	129.7	348	2346.7	60	2286.7	1378	908.7
7	97.0	1.9	98.9	348	2459.8	60	2399.8	1233	1166.8
8	167.4	3.3	170.7	348	2156.7	60	2096.7	1784	312.7
9	146.2	2.9	149.1	348	2264.0	60	2204.0	1566	638.0
10	126.9	2.5	129.4	348	2349.6	60	2289.6	1363	926.6
11	94.2	1.9	96.1	348	2467.0	60	2407.0	1233	1174.0
12	160.2	3.2	163.4	348	2230.7	60	2170.7	1769	401.7
13	228.7	4.6	233.3	348	2290.1	60	2230.1	1552	678.1
14	122.1	2.4	124.5	348	2371.4	60	2311.4	1349	962.4
15	96.1	1.9	98.0	348	2464.2	60	2404.2	1233	1171.2
16	167.8	3.4	171.2	348	2126.3	60	2066.3	1842	224.3
17	166.7	3.3	170.0	348	2065.3	60	2005.3	1784	221.3
18	167.7	3.4	171.1	348	2058.1	60	1998.1	1842	156.1
19	31.4	0.6	32.0	348	2554.1	60	2494.1	377	2117.1
20	169.9	3.4	173.3	348	2135.0	60	2075.0	1827	248.0
21	166.5	3.3	169.8	348	2135.0	60	2075.0	1755	320.0
22	145.9	2.9	148.8	348	2249.6	60	2189.6	1552	637.6

performance of the proposed valve over the complete range of operating conditions (Table 8.33m). The valve body size is frequently determined not only on the basis of the required valve capacity (C_v), but also by considering the maximum acceptable flow velocity [approximately 30 ft/s (9 m/s)]. Additional information can be found in ISA standards (Reference 4).

Steam Temperature Control

Steam temperature is usually controlled with desuperheating valves installed between the first and second superheaters, similarly to the arrangement in regular boilers (Figure 8.33n). If the superheaters are split into more banks in parallel, good practice requires a desuperheating valve per each bank.

The control function is normally implemented by a cascade loop in which the master controller (TIC-111 in Figure 8.33n) senses the final steam temperature and the slave senses the temperature after the desuperheater, before the steam enters the second superheater. The set point of the slave controller has a low limit of roughly 28–30°F (15–16°C) above the saturation temperature obtained by characterizing the steam drum pressure to prevent moist steam in the second superheater, as well as poor temperature control.

The temperature used as the measurement of the slave controller (TT-112) should be located about 60 ft (20 m) downstream of the desuperheater to allow the water droplets to vaporize, which is necessary to obtain a correct measurement and to prevent mechanical damage to the protecting thermowell. Sometimes a feedforward action by steam flow is added (FY-102) to the control system.

Very often, the piping is also protected with internal sheathing to prevent erosion. If the thermal design is correct, in some operating conditions (mainly under base loading) the desuperheaters are not in operation, while in other operating conditions the desuperheating BFW flow can be highly variable. It is suggested that the suppliers who bid on the desuperheater be provided with data on all of the different operating conditions (Table 8.33o).

When a desuperheating valve is not in use, it is isolated by a pneumatically or electrically operated on/off valve that is installed upstream of the desuperheater in the BFW line. This valve is closed a few seconds after the desuperheating valve is closed (or when it is less than 2% open) and is opened

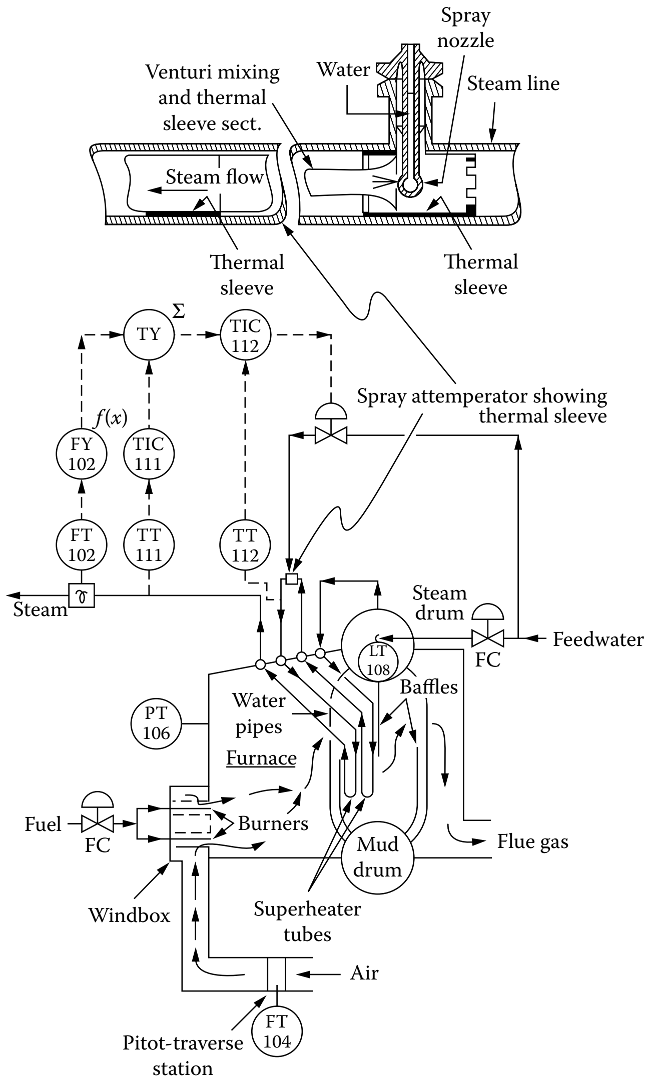


FIG. 8.33n

Cascade configuration for controlling the desuperheater control valve.

when control signal setting the desuperheater valve opening exceeds approximately 4%.

If the desuperheater is of the insertion type, its rating should be determined by considering the design pressure of the BFW and the design temperature of the steam. With today's high pressures [up to 3500 PSIG (241 barg)] and temperatures [up to 1075°F (580°C)], this is not always achievable. This creates serious problems when desuperheaters are selected. Additional information can be found in ISA standards (Reference 5).

Pegging Steam Control

Pegging steam is required to deaerate the boiler feedwater during start-up, when no low pressure (LP) steam is yet available and if no steam is available from an auxiliary boiler or from other HRSGs. The purpose of this control system is to start deaerating the boiler feedwater as soon as possible during start-up.

If the deaerating tower is located on top of the LP drum, then during normal operation, the steam needed for deaeration is generated in the LP drum. Under start-up conditions, the steam first becomes available in the HP section, then in the IP section, and finally in the LP section of the HRSG. If an auxiliary boiler is not available, in order to start degassing as early as possible, a connection can be made between the IP or HP drum and the deaerating head. In this connecting pipe, a motor-operated on/off valve and a throttling control valve are installed.

The control valve serves to reduce the HP or IP pressure to the LP level for the deaerator during start-up. Once the LP section is generating enough steam to be self-sufficient to ensure good deaeration, this start-up phase can be terminated. This happens when the pressure in the deaerator reaches a pressure of about 25 psia (1.7 bara). At this point, the pegging steam line is closed. This is done by closing both the control valve and the on/off isolating valve.

On fired units it is possible that pegging steam is also required during normal operation at high loads, because the LP section does not generate sufficient steam.

Condensate Preheater (Economizer) Temperature Control

This control system is unique to HRSGs and is not used on conventional fired boilers. Its dual purpose is to 1) prevent external condensation on the tubes of the preheater, which could cause external corrosion, and 2) obtain the correct inlet temperature to the deaerator.

The first objective is met by taking some of the hot condensate leaving the preheater and mixing it with the cold condensate entering the preheater (TIC-1 in Figure 8.33p). The hot condensate is recirculated by recirculation pumps, and the flow rate of recirculated hot condensate is throttled by TIC-1, which detects the temperature of the (mixed) condensate stream entering the preheater and keeps that temperature high enough to prevent external condensation.

The correct inlet temperature for the deaerator is obtained by mixing the hot condensate downstream of the preheater with cold condensate. This inlet temperature is maintained by TIC-2, which detects the mixed condensate's inlet temperature to the deaerator and is set a little lower [15°F (8°C)] than the boiling temperature. If large amounts of the hot condensate are recirculated by TIC-1, this could cause TIC-2 to lose control and cause a low temperature at the condensate inlet to the deaerator, thus the need for pegging steam.

A less sophisticated control configuration that has been used to remove sulfur from flue gas is shown in Figure 8.33q. Here, the cascade master TIC-2 maintains the required deaerator inlet temperature by modulating the set point of TIC-1, the slave controller. In this configuration, however, there is no positive protection of the preheater coil against condensation on its external surface.

TABLE 8.33o
Illustration of the Process Data Required for the Proper Sizing of an HP Steam Desuperheater

4 valves per boiler—one for each superheater bank in parallel

Case	*Q Steam (lb/s) (total)*	*Q Steam (lb/s) Per Line*	*T Steam Upstream °F*	*T Steam Downstream °F*	*P Steam (psia)*	*Q Water (total) lb/s*	*Q Water Valve lb/s*	*T Water °F*	*P Water (psia) el. 100 f.*	*Dp Water Steam (psi)*
1	0.0	0.0	921	921	1182.1		0.000	288	1472.1	290.0
2	19.8	5.0	979	842	1241.5	19.832	4.958	293	1472.1	230.6
3	9.3	2.3	961	878	1187.5	9.264	2.316	284	1472.1	284.6
4	32.0	8.0	1022	797	1245.9	32.023	8.006	271	1472.1	226.2
5	32.6	8.2	1044	775	1208.2	32.625	8.156	273	1472.1	263.9
6	0.0	0.0	905	905	1167.6		0.000	297	1472.1	304.5
7	0.0	0.0	912	912	1166.1		0.000	295	1472.1	306.0
8	9.2	2.3	954	882	1211.1	9.231	2.308	295	1472.1	261.0
9	18.1	4.5	973	849	1241.5	18.105	4.526	293	1472.1	230.6
10	21.1	5.3	993	829	1215.4	21.161	5.290	291	1472.1	256.7
11	7.2	1.8	952	889	1179.2	7.208	1.802	293	1472.1	292.9
12	25.0	97.3	1008	813	1212.5	25.000	6.250	286	1472.1	259.6

In some HRSGs, block valves are provided so that the LP preheater can be temporarily valved off (isolated), if water leakage occurs.

If the fuel has a high sulfur content, the primary concern is to keep the inlet temperature of the condensate to the preheater above the dew point of flue gas by proper recirculation of some of the hot condensate.

Supplementary Firing Control

In some plants, the fuel flow to the burners is set manually by the load dispatcher, but errors can cause transient overfiring and overheating the superheater tubes. Therefore, a burner management system, complete with accessories such as flame scanners, purging, and cooling, shall be provided to ensure the boiler safety.

Supplementary firing should limited to a fraction of the GT load. The applicable safety codes in the area of the HRSG depend on the presence of fuel gas in the proximity. It is possible to reduce the area classification by a careful study of the piping layout and by the adoption of some precautions. These would include he use of double stuffing glands with bleeds in between the glands on the fuel gas valves. The intermediate bleeds from such valves are collected and vented to the atmosphere at safe location.

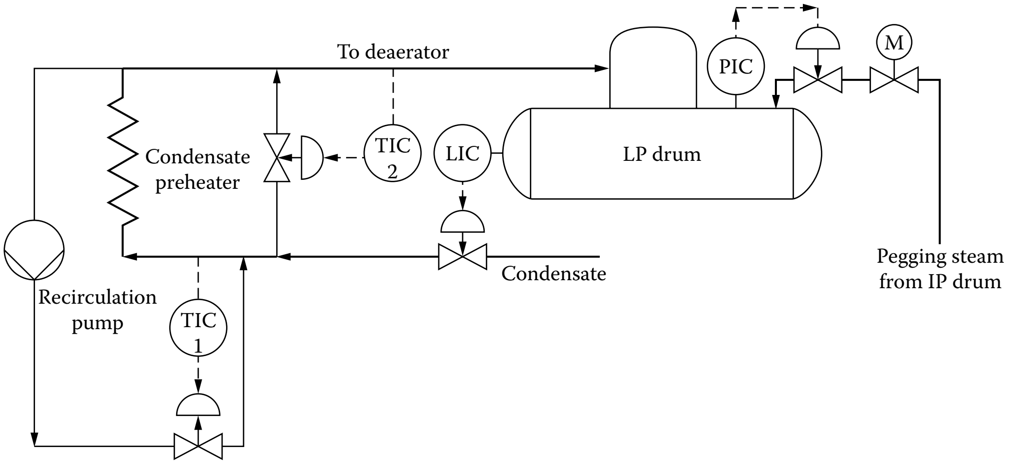


FIG. 8.33p
TIC-1 keeps the condensate preheater surface temperature high enough to prevent external condensation, while TIC-2 keeps the deaerator inlet temperature below the boiling point. Under certain conditions, the two loops can fight each other.

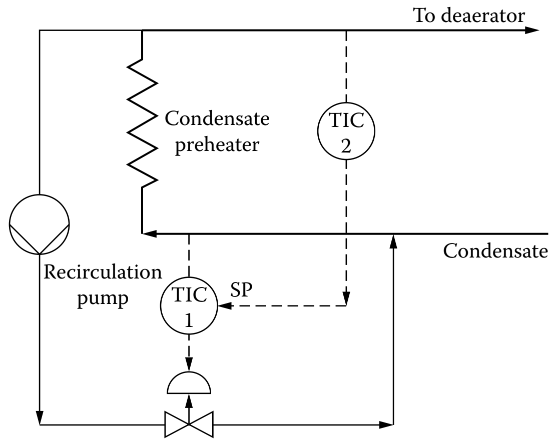


FIG. 8.33q
Cascade control of preheater does not give positive protection against flue gas condensation on the external surface of the preheater.

The firing could also pose some problems with emission control, especially at reduced GT loads, and may require the use of flue gas recirculation (FGR).

Steam Turbine Bypass Control

The bypass valves (Figure 8.33k) around the steam turbine (ST) are closed during normal operation and are throttling during the start-up, after an ST trip, or in the case of load rejection. A pressure controller PIC (with antireset windup) is throttling the bypass valve, so as to keep the upstream (inlet) pressures at the desired HP, IP, and LP settings and to make sure that they do not reach the settings of the corresponding safety valves.

In addition, the pressure control valves in the turbine bypasses should be provided with a quick-opening capability (a solenoid can be used on pneumatic valves) that will be triggered by an ST trip and remain open for a duration of approximately 5 sec.

When the boiler operates in sliding (floating) pressure, the set point of the bypass PICs should track the operating pressure to prevent sudden transients. In these conditions, the purpose of the bypass is to prevent the safety valves from blowing when the steam pressure suddenly increases due to a trip of the ST. Alternatively, if the valves have a hydraulic actuator capable of following the controller output with very little delay (a few seconds), the opening can be proportional to the steam flow prior to the ST trip, thus avoiding any pressure upset.

During start-up, after the vacuum has been established in the condenser, the set points of the bypass PICs are set to follow the acceptable heating curve of the steam drum until the steam turbine is warmed up and can handle the superheated (SH) steam. A desuperheater should be provided to reduce the steam temperature to the value that is compatible with the downstream steam conditions. This desuperheater can be integral with the pressure reducing valve or separate from it.

The bypass valves can be sized to pass partial or full flow at the operating pressure conditions. In combined cycle power plants, the bypass valves are normally sized to pass the full steam flow. If the HRSG can operate on sliding pressure, the sizing is still based on the maximum operating pressure and flow. During the operation in the sliding pressure mode, because the GT is not at base load, the pressure drop across the valve is reduced, and an ST trip can cause a rise in the upstream steam pressure, but always within the normal operating values.

If the plant configuration is 2 + 1 (Figure 8.33f), during plant start-up or if GT is out of service, only one GT and HRSG are feeding the ST. In such cases, the HP pressure is lower than the normal one even though the operating GT is running at base load. A trip of the ST under these conditions will cause a sudden opening of the bypass, and the pressure in the operating HRSG will rise up to a value close to the nominal operating pressure. Additional information can be found in ISA standards (Reference 6).

Load/Frequency Control

The purpose of this control system is to keep the frequency variations within preset limits when the load and the generated power are unbalanced. The system guarantees that each generator unit can automatically and autonomously participate in the frequency control by gradually modifying the generated power in order to keep the frequency constant.

This is completely true if the unit is running at partial loading and is somewhat less so as the loading approaches 95%. If the unit is running at base load, it can only decrease its load if the frequency is in excess of the nominal one but it cannot increase the load in the case of an underfrequency episode.

Due to the offset that is caused by the proportional-only mode of frequency regulation, even when the generated power is balanced with the load, there could still be an offset in the frequency that needs to be corrected. This correction is provided by an external controller (at grid level) that acts as a primary or master of a cascade control loop whose slave is the turbine governor. In such a case, a command is likely to be received from the grid dispatcher to vary the generated power of the units that have spare capacity, in order to restore the nominal frequency of the grid or the correct exchange of power with other grids.

The variation in total generating capacity is split between the GT and ST into 2/3 in the GT and 1/3 in the ST. This means that if the electric power generation should change by x MW, the total change almost immediately (within a few seconds) affects the GT, and then, a few minutes later, there will be an increasing contribution by the ST with simultaneous decreasing contribution by the GT. When the new steady state is reached, the contribution of the GT will be about

2/3x and the incremental power contribution of the ST will be about 1/3x.

Running Permissives

Each machine component is provided with its own internal safety instrumented functions (SIFs) that are defined and often implemented by the manufacturers, because they are inside their scope of responsibility for supply and are also proprietary systems.

As the components of the total plant are very much interconnected from both from the process and from the electrical perspective, it is necessary to provide reliable interconnections with the objective of minimizing the partial or total plant shutdowns that they can cause. These functional interconnections are very much dependent on the arrangement of the plant (single shaft or multiple shaft, 2 +1 configuration, and so on) and the response to the failure of all major pieces of equipment must be planned in advance. Therefore, it is good practice to prepare both a cause–effect table/matrix (sometimes called a shutdown key) and an overall diagram of running permissives, which are also useful to the plant operators during normal operation.

A steam turbine trip causes the sudden temporary opening of the bypass valves around the STs and possibly the opening of the HRSG vent valves (Figure 8.33k). This allows the plant to run at reduced electrical capacity with the GT only, yet retaining the possibility of restarting the ST if the causes for the trip can be quickly removed.

If the condenser is not available, the GT is forced to operate under FSNL condition and either vent valves on the HRSG are opened or the GT must be shut down. This choice is dependent on the design of start-up vents and on the availability of water, as was discussed earlier. If the HRSG is provided with a diverter, the possibility of exhausting the GT to the atmosphere should also be considered, particularly if it is expected that the restarting of the ST will be delayed beyond a certain time period.

An HRSG trip due to unsafe conditions (too low a level, or a "low-low" condition in the drums, or sometimes too high a temperature, or a "high-high" condition of the SH or RH steam) should either open the diverter (if such exists), or otherwise the GT should be shut down. If a high level condition is detected ("high-high"), the ST bypass system can be opened and the ST tripped. In the case of abnormal conditions in the post-firing (flame failure) section, only the burner trip should be considered.

A gas turbine trip causes the trip of the complete train, unless complementary firing is available in the HRSG. In this case, the ST should be tripped, the diverter should be switched, the air fan started, and the burners lit after purging, to avoid disruption of the steam being generated for industrial use.

An electric generator trip should cause the trip of the pertinent driving turbine or of the complete train, in case of single-shaft configuration. In some cases, if the fault is electric, it is possible to try to prevent the trip by opening the circuit breaker of the machine. This way, the GT can be kept in operation at FSNL, with the possible benefit of reducing the time to restart and of reducing the equivalent hours of GT operation.

A step-up transformer fault causes the trip of the pertinent generator, if it is directly connected, and also the opening of the line breaker. If there is a circuit breaker between the generator and the transformer, then it too is opened.

A fault in the bus bar system causes the isolation of the faulty bus bar, possibly without interfering with the generation set.

A fault in the grid causes the opening of the line breakers, the trip of the ST, and the running of the GT in "island mode" (or at minimum load sufficient to feed the pertinent loads in the power house). This kind of electric trip with sudden load loss is claimed to be tolerated by the GT, but sometimes keeping the GT in operation requires a time-consuming fine-tuning.

Safety Functions and Integrity Levels

The manufacturer-supplied GT and ST governors usually include the control and safety systems for the turbines. In fault-tolerant configurations containing a high level of diagnostics, they normally include separate processors at least for safety functions. However, detailed diagnostics are usually not provided by the GT manufacturers, and the safety integrity level (SIL) "capability" of the governors is not guaranteed, either, by their architecture alone. The attainment of the SIL capability also requires the proper selection of the sensors, wiring, final elements, maintenance, and testing.

The turbine suppliers are very reluctant to accept any modification or inclusion of additional functionality to their controls, even if it is strictly correlated with their supplied controls and is external to the GT. This is because the turbine package is provided with control functions inside the governor (designed around the turbine it protects). The supplied packages usually do include several architectures of the furnished control functions, from which the user can select.

Most safety-related controls that are required for turbines in a cycle with an unfired HRSG are furnished within the manufacturer's package. The unsafe conditions that are usually not covered are those of the levels in the steam drums of the HRSG and the fuel gas intercept valves to the GT.

Often the specifications call for a safety integrity level of 3 (SIL 3) for all SIFs, and this could be due to inadequate analysis. Such specifications could result in overengineering of the complete safety instrumented system (SIS) with its associated costs in maintenance, for the whole life cycle of the plant. For the proper selection of SIL levels, see References 7 and 8.

START-UPS AND SHUTDOWNS

Start-Up Procedures

The combined cycle power plants, even if they are operated at base load, are subject to starts and stops more frequently than other industrial plants or power plants. The frequency of start/stop requirements dramatically increases if the combined cycle operates in an intermittent way.

In order to operate the plants with reduced crews of operators, it is necessary to be able to implement a start-up procedure from the control room without people having to walk around the plant. Hence, all valves that need to be operated during the start-up should be motor-operated.

The main concern in starting up a combined cycle power plant is to avoid thermal stresses to the machinery that would shorten its life and produce unsafe conditions. This consideration results in extending the time for start-up, while economics would require that start-up take place in the minimum possible time and with minimum fuel consumption.

For the ST and the HRSG, different starting conditions can be defined depending on their status at the beginning of the start-up (i.e., if they are cold or hot). Of course, a cold start up requires a more cautious and longer procedure to avoid stressing the machinery. However, the cold and hot conditions are different for the HRSG and the ST, so that a start-up can be considered cold for one machine and hot for the other. For large HRSGs, a warm start-up condition is often considered.

Each manufacturer of the main plant equipment sets the requirements for its machine, and then it is up to the process design engineers to combine these requirements with their own to arrive at start-up procedures that will minimize the overall start-up time.

Sometimes, the normal start-up procedure needs to be modified to cope with the environmental constraints as dictated by local regulations. This could involve a long warming period at low load followed by a quick ramp-up, or starting with the GT at high load (even 70–80%) plus a lot of steam being bypassed to the condenser, in order to maintain the proper temperature gradient in the HRSG. This unusual situation can have an impact on the last stages of the ST and needs to be thoroughly investigated with the manufacturers in order to avoid dangerous operating conditions.

The start-up sequences are resident both in the DCS that acts as the overall coordinator, and in the turbine governors that are interfaced with the DCS. As previously mentioned, in several instances the governors are parts of the DCS.

Permissive Conditions The start-up can take place only if some permissive conditions are satisfied, such as the availability of instrument air, cooling water, demineralized water, fuel gas, and electricity to feed all motors (HV and LV) and the DCS. All these permissives should be listed in detail as prerequisites for the start-up procedure.

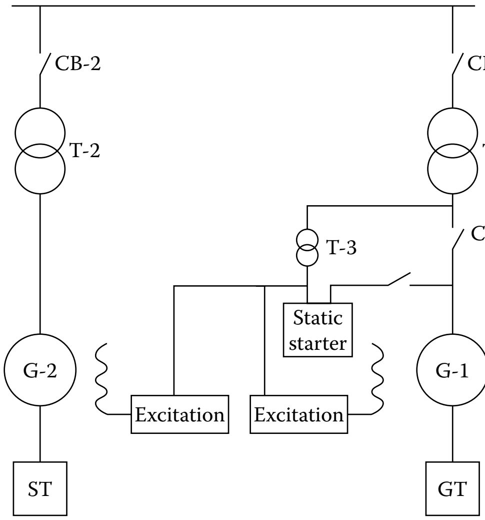


FIG. 8.33r
Simplified one-line electrical diagram of a dual-shaft power station of 250 MW capacity.

The document describing the overall sequence should start with the required position of all valves, status of all controllers (auto, manual, forced to XX value, and so on) and conditions of all pumps (running, standby, available/unavailable) at the beginning of the start-up procedure. The hot, warm, or cold condition of each machine should also be determined.

According to the mechanical and electrical configuration of the plant, the start-up procedure can vary to follow specific requirements. Therefore, each plant has to be designed individually. In other words, each plant is unique and its start-up procedure should also be developed to suit the unique requirements of the GT, HRSG, ST, and condenser manufacturers and, possibly, also of the electrical grid and plant operators. In the case of cogeneration, the requirements of the steam user operators should also be taken into account.

As an example, a typical cold start-up procedure is provided below for a unit designed for 250 MW, dual-shaft, unfired HRSG with three steam pressure levels, a water-cooled condenser, and a GT generator also acting as launching motor, in which the main steps correlating the different equipment are highlighted (Figure 8.33r). In this case, no black starting was required and no steam was available from an auxiliary boiler or other units.

Cold conditions mean that the boiler has an HP temperature lower than 212°F (100°C) and the first-stage metal temperature of the ST is less than 245°F (120°C). Some suppliers also recommend that the shafts of the turbines be cranked for few hours before starting up, with the lubrication system in operation.

For a power plant unit of this size, skin thermocouples are installed on the HP drum, on the final HP superheater, and on the IP reheater, mainly to monitor the metal temperature gradients during the start-up phases.

If all permissives are satisfied and the equipment is ready to be energized, the start-up sequence proceeds as follows:

The Start-Up Sequence

1. Close the line circuit breaker CB-1 in order to energize backwards to the gas turbine step-up transformer T-1 and the unit transformer T-3. The generator circuit breaker CB-3 is open.
2. On the operator's start-up command, the GT governor will put in operation the lube oil system, if it was not in operation because the turbine was cranking. For heavy-duty turbines a slow cranking (few rpm) is necessary for several hours before starting. The temperature of the lube oil shall be higher than the minimum required by the manufacturer.
3. The lubricating oil circuit of the steam turbine is put in operation as well as the jacking oil circuit of the generator.
4. The shaft turning gear of the steam turbine is started, unless it was already operating.
5. All cooling circuits are put in operation.
6. The level of the condensate storage tank is to be at a normal value.
7. The condenser hot well level is to be at a normal value.
8. The condensate extraction pumps are put into operation.
9. The LP drum level is filled to the required start-up value.
10. The IP and HP drums are filled to the required start-up level (minimum level).
11. The boiler blow-down system is started and put in auto; the sampling system is started as well.
12. One main boiler feedwater pump is started.
13. The steam turbine bypass system is prepared for automatic operation.
14. The HRSG configuration is set and checked.
15. The superheater drain valves are open and shall be kept open until a pressure of approximately 30 psig [2 bar(g)] is reached.
16. The steam line drains are open to warm up the piping and shall be kept open until the correct temperature has been obtained.
17. The boiler setup sequence is performed either automatically or manually, step by step. The automatic ramp-up sequence of the HRSG is selected for start.
18. The excitation system of the GT generator and the static starting system of the GT are energized.
19. The block and bleed valves on the fuel gas supply are put in operation in order to deliver fuel gas to the gas turbine inlet valve.
20. Under the control of the GT governor, the gas turbine is now ready to start, waiting for the start command from the operator that can be given through the DCS or directly via governor HMI.
21. When the correct lube oil pressure is established, the static starter will rotate the gas turbine and run it at a preset value without firing during the time needed for the proper purge of the turbine and of the HRSG. It should be noted that the purging time is longer on the combined cycle than on the open cycle, as the volume of the HRSG must be purged as well. The time for the purge (typically, 15–20 min) is calculated on the basis of five air changes with the airflow produced by the compressor running at launching motor speed.
22. At the end of the purge, the GT governor will light the burners, the static starting system will be switched off and the gas turbine will ramp up to the synchronization speed in 15–20 min (for heavy-duty GTs) or a couple of minutes (for aeroderivative GTs). The excitation system regulates the generated voltage to the correct value. The speed gradient of the gas turbine is controlled by the GT governor. The gas turbine is kept at full speed no load, exhausting to the HRSG with the generator disconnected from the grid. For some turbines, it is required that a minimum load be connected; in case the power station auxiliaries are not large enough, the paralleling of the GT generator to the grid shall be performed. Some turbines even at FSNL have an exhaust temperature higher than an acceptable value [usually 770°F (410°C)] for the HP superheater coils (abnormal thermal expansion due to very low steam flow), and therefore it is necessary to activate the exhaust temperature limitation function.
23. The HRSG will start steaming, and the pressure is controlled by actuating the vent valves so that the temperature gradient in the HP drum remains within the limit indicated on the relevant diagrams by the HRSG manufacturer for cold conditions in order to prevent stress to the pressure parts. The vent valves are initially kept at a minimum opening, then switched to automatic control with an automatic ramp of the controller set point. If the vacuum in the condenser can be obtained because some steam is available, the steam pressure is controlled by means of the turbine bypass valves.
24. The HRSG is kept at the warming up condition [HP 800 psig (55 barg), IP 300 psig (20 barg), LP 50 psig (3 barg), approximately] for at least 15 min, while the steam lines are preheated via the main stop bypass valves, line vents, and drains, or alternatively, with the main stop valves fully open.
25. As soon as the differential pressure across the main stop valves and the temperature of the steam lines are acceptable, the main stop valves of the HRSG are opened. If the start-up takes place with the stop valves fully open, this step is skipped.
26. The generated steam from HP or alternatively from IP is conveyed to pressurize the steam glands of the

steam turbine. If auxiliary steam is available, this step takes place much earlier.

27. The start-up ejector is started as soon as steam is available at suitable conditions and is operated until the condenser pressure reaches 4.5 psia [300 mbar(a)]. After that the main ejectors are started in order to reach the proper vacuum 0.4 psia [30 mbar(a)].
28. When the pressure in the condenser reaches 4 psia [280 mbar(a)], the turbine drains to the condenser are opened and the generated steam is desuperheated and dumped to the condenser. The pressure reducers (bypass valves and desuperheaters) in the main steam system are put in operation with an initial set point of approximately 870 psig [60 bar(g)] for the bypass from HP to the cold reheat (CRH), while 200 psig [14 bar(g)] or less is required for the bypass from the hot reheat (HRH) to the condenser and 45 psig [3 bar(g)] from LP to the condenser. The operation is performed by acting through the DCS, starting from the HP to the CRH bypass (sometimes this bypass is opened before the vacuum at the condenser is established in order to flush the RH coils) and ending with the IP and LP to the condenser.
29. At this point, the gas turbine is still running at FSNL (or at minimum load) and the HRSG (HP) stays at 870 psig [60 bar(g)] and 735°F (390°C).
30. When the correct control fluid pressures are established, the ST is started under control of the turbine governor. The turbine increases its velocity with a gradient defined by the governor and is gradually warmed up according to the sequence defined by the manufacturer. It is possible that the acceleration is increased in a range around the first critical velocity of the steam turbine and of the generator. When the turbine velocity is about 90% of the nominal speed, the excitation system will regulate the generated voltage to the correct value.
31. The steam lines' vents and drains are closed.
32. As soon as the steam turbine has warmed up to an acceptable point under control of the synchronizing equipment and on request of the operator, the gas turbine generator is put in parallel with the grid by closing the circuit breaker CB-3.
33. The GT governor ramps up the generated load at least to the minimum allowable working value (about 55%).
34. The gas turbine output is raised up at a rate of about 2%/min. This operation can be done either through the DCS or directly by the GT governor HMI.
35. Meanwhile, the set points of the pressure regulators of the bypass valves are gradually raised, so as to pressurize the boiler, always keeping in mind that the temperature gradient in the steam drums shall not exceed the value stated by the boiler manufacturer.
36. When the boiler reaches a load of 30% the level control is switched to three-element control with the normal operational set point.
37. Drains and chemical dosing system of the boiler are put into normal operation.
38. When the warm-up procedure of the ST is finished, the steam turbine generator is synchronized with the grid, and the circuit breaker CB-2 is closed.
39. The power plant is now ready for ramping up to nominal power; when pressure in the HP drum reaches 1150 psig [80 bar(g)], the ramp limitation for the gas turbine is increased to 5%/min.

The complete cold start-up procedure of such a group of equipment requires about 4–6 hr or even more because of the low acceptable gradients and of the suggested stabilizing time in the HRSG and ST heating. In practice, the GT remains at FSNL or at minimum load for almost the whole start-up period, except the last 1/2–3/4 hour.

The hot and warm start-ups differ from the cold one mainly because several steps in the early stage of the sequence can be skipped and because higher gradients are acceptable in the limits on temperature/pressure rise. On the other hand, some additional precautions must also be taken. For example, it is necessary to protect from the steam becoming cooler than the ST metal by synchronizing the GT and raising the load until a proper steam temperature is reached. It may be necessary to lower the pressure of the reheater (RH) by opening the bypass and possibly also the vent valve, because the bypass is sized to pass the full flow at nominal conditions.

As a result, for the group of equipment considered in the example, the hot or warm start-up time is shortened to 2–3 hr.

Short-Term Planned Stop

A planned shut-down of the complete unit has the purpose of stopping the operation of the group of equipment with minimum stress to the machinery. It is also a goal to keep the electrical system alive to the maximum extent, so as to be able to restart the plant as soon as possible, in the hot start-up conditions. It is advisable to verify with equipment suppliers if there are any process variable limits or constraints that must be obeyed to avoid damaging the equipment itself. In particular, there could be the temperature of the RH steam that should be lowered in a predetermined way to prevent abnormal axial elongation of the ST rotor.

Short Stop Sequence

1. Reduce the generated power with a predetermined gradient through the GT governor by reducing the fuel gas admitted to the gas turbine.
2. The GT governor starts the auxiliary lube oil pump of the GT unless lubrication is obtained by means of separate pumps.

3. Consequential to the reduced power generated in the gas turbine, the steam turbine reduces its generated power. It is necessary to decrease the pressure of the RH beyond the natural pressure decrease by assigning variable set points to the bypass and vents, based on pressure/power curves defined by the ST manufacturer.
4. When the power generated by the steam turbine is close to 30%, the steam turbine bypass valves are opened and shortly after the turbine is tripped. The circuit breaker CB-2 opens, actuated by the minimum load relay. The vacuum in the condenser and the steam to the glands should be maintained as long as possible, at least until the steam turbine is stopped and restarted under turning gear (a sudden inlet of fresh air inside the hot portion of the turbine might generate distortions and thermal stress both on rotor and casing).
5. Once the minimum allowable load is reached, the GT governor shuts the gas admission valves, opens the circuit breaker CB-3, and excludes the excitation circuit.
6. Close the block and open the bleed valves on fuel gas inlet to the gas turbine package.
7. Immediately after the turbines are stopped, the steam turbine jacking oil pump is automatically started and the turbine turning gear is put into service, keeping the AC lubrication oil pumps in service until the journal temperature is reduced to a value compatible with the babbit metal of the bearing.
8. The operator activates the automatic bottling procedure of the HRSG, by which all valves around the HRSG and the damper in the stack (if present) are closed. If the stop is very short, then a couple of minutes after the TG stops, a very slow depressurization procedure is activated by slightly opening the start-up vents, thus keeping the superheaters flushed as to prevent condensation.

At the end of the short planned stop, turbines of the power station are still hot and turning under the turning gear and the HRSG is still pressurized and hot, ready to start in the shortest possible time (a couple of hours).

Long-Term Planned Stop

The long-term planned stop of the complete power generating unit has the purpose of stopping the group of equipment with minimum stress to the machinery and to reach the cold conditions in the shortest time. This has to be done in a manner compatible with the allowable temperature gradients that are needed to perform maintenance by means of internal inspection of the HRSG, or maintenance of the steam turbine.

In case of the long-term stop, the short planned stop is modified by running the ST at minimum load to stabilize its temperature at low values. In addition to the steps listed for the short-term planned stop, the sequence required in this case is as listed below:

Long Stop Sequence

1. Open the vent valves.
2. Break the vacuum in the condenser and take out the sealing steam from the turbine gland, following ST manufacturer instructions.
3. When the pressure in the steam drums reaches about 30 psig [2 bar(g)], open the drain valves in the superheaters.
4. Bring the level in the steam drums to minimum value and close the BFW control valves.
5. Stop the BFW pumps, depending on the pressure in the steam drums.
6. The jacking pump and the lube oil pump of the ST must be kept in operation until the rotor is cold enough to be sure that the bearing shall not be damaged by temperature. This operation takes several days. Following manufacturer instructions, the ST turbine rotor shall be rotated under turning gear periodically to avoid unelastic permanent deformation. During this rotation, the oil system shall be in operation, jacking oil pumps included.

After such a stop the boiler needs cooling down before one can enter the flue gas side. This is accomplished by opening the inspection doors and allowing the air to circulate inside the casing. The boiler is, however, still full of water. If the inside of the steam drums is to be inspected, the blowdown valves should be opened to empty the steam drums. If the boiler must be completely emptied, it is done by also opening all the drains.

Emergency Shutdown

Depending on its cause, the emergency shutdown can be partial or total.

An example of a partial shutdown is a cause requiring the shutdown of the steam turbine. If the exhaust gas from the GT can be diverted to a bypass stack (Figure 8.33i) or if the generated steam is dumped to the condenser, the gas turbine can continue generating electric power, even though in an uneconomical manner.

A fault in the gas turbine or in the HRSG causes a complete stop of the power generating unit, unless a flue gas diverter exists and the fault is in the HRSG. A fault in the utilities can generate partial or total shutdown or lead to a runback of the gas turbine. The emergency shutdown can also be caused by the electric grid, resulting in partial or complete shutdown of the unit.

The operator can also initiate total or partial emergency shutdowns, but this operation should be avoided as much as possible to prevent stresses to the machinery. In terms of equipment life, an emergency shutdown of the GT is considered to be equivalent to at least 50 hours of operation and can exceed 150 hours, according to the formulas of some GT manufacturers. As a rule of thumb, for an emergency shutdown, the equivalent number of hours of operation is equal to the load percentage at the moment of the turbine trip (e.g., if the turbine trips at 75% of the load, the equivalent hours are 75).

A load rejection is accounted as 50% of the corresponding equivalent hours due to an emergency trip. If the causes for the shutdown are external to the generating set, its shutdown should be avoided as much as practical. This can be achieved by operating the gas turbine in an island mode, i.e., disconnected from the grid and feeding the electric loads of the power station only.

The operator can then decide whether to stop the gas turbine or to keep it in island operation. Sometimes, in the case of a sudden load rejection, the gas turbine is unable to go into the island mode of operation, because of the intervention of its overspeed protection interlocks.

PERFORMANCE TESTS

The performance tests are normally carried out in accordance with ANSI-ASME Performance Test Codes (PTCs), namely PTC 22 for Gas Turbines (Reference 9), PTC 4.4 for Gas Turbine Heat Recovery Steam Generators (Reference 10), and PTC 6 for Steam Turbines (Reference 11). For gas turbines see also ISO 2314 (Reference 12).

As part of the performance test, temperature measurements should be made in the HRSG inlet duct, as required by the code. These requirements call for a minimum of 24 temperature sensors in small ducts (up to 48 ft^2 in cross-sectional area). These thermocouples should be supported, mineral insulated, and provided with external stainless steel sheathing, and be without protecting wells.

The thermocouples should be located at the center of the individual areas of a grid, with the number of rows and columns selected according to the ANSI/ASME PTC 4.4. These thermocouples serve a temporary purpose and should be removed after the test is completed. The difference between temperature readings in the section of the flue gas duct can reach 15–20°F (8–12°C). A suitable method of temperature averaging has to be developed in order to correctly evaluate the performances of large units.

It is also necessary to test the BFW flow to the various steam drums (HP, IP, LP) of the HRSG, as the water flow measurement is more accurate than that of the steam flow. The steam flow, therefore, is measured as a reference. During the HRSG performance test, the blow-down valves and the sampling lines to the steam and water analyzers are closed.

The water flow measurement tends to be a little unstable even during steady-state operation, which is required for the performance test. It is, therefore, recommended to prepare a DCS screen page on which the various flows are totalized. This allows the operator to simultaneously start and stop all the totalizations, using software pushbuttons on the page (Table 8.33s).

TABLE 8.33s

Tabulation Allowing the Operator to Compare the Total Flows of Water and Steam In and Out of the Steam Drums

Service	*Total Flow*	*Units*	*Service*	*Total Flow*	*Units*
LP BFW	8 digits	Lbs	LP Steam	8 digits	Lbs
IP BFW	8 digits	Lbs	IP Steam	8 digits	Lbs
HP BFW	8 digits	Lbs	HP Steam	8 digits	Lbs
IP Dsh W	8 digits	Lbs			
HP Dsh W	8 digits	Lbs			
CURRENT TIME		PRESET DURATION		ELAPSED TIME	
hh/mm/ss		hh/mm/ss		hh/mm/ss	
RESET		START		STOP	
ν		ν		ν	

The totalized flows should cover 3- to 4-hr periods of flow integration as a minimum, because the corresponding ANSI standard (PTC4.4) requires that the test should last at least 2 hr. If there are reheat coils and the steam flow through them is not direcly measured, it should be calculated by mass balance around the steam drums.

CONCLUSIONS

In this section, the traditional basic controls used in power plants have been discussed. The potentials for optimizing these processes are somewhat limited, but should still be considered.

For example, when the plant is operated in an intermittent manner, the start-up period should be minimized. The corresponding savings in fuel and production can be substantial if during the design stage and during the subsequent tuning of the system in field, attention is focused on the goal of decreasing the start-up time. To reach this goal, the start-up sequences of the individual equipment should be well coordinated, and when necessary, some of the steps of the start-up sequence should be modified, postponed, or anticipated during commissioning.

Another potential for optimization is noise silencing. This is because if the steam venting during start-up is too noisy, it restricts the times at which the plant can be started up in built-up areas. This consideration can limit the acceptable time period and result in a need to put the power plant on-line well in advance of the required time.

In terms of the potentials for improving equipment efficiency through optimization, keeping the air compressor blades of the GT clean is an important goal. The frequency of the on-line washing required should be determined in the field, based on both the turbine characteristics and on environmental pollution. The off-line washing should be done when the GT performance drops substantially.

The goal of optimization can be served by evaluating the trade-off between the recovery in performance after a maintenance-related shutdown and the loss of production due to that plant shutdown. Computerized control systems can be useful in optimizing the maintenance practices of the plant.

ABBREVIATIONS

AVR	Automatic voltage regulator
bara	Bar absolute
BFW	Boiler feedwater
BMS	Burner management system
CEMS	Continuous emissions monitoring system
CRH	Cold reheat
CRT	Cathodic ray tube
DCS	Distributed control system
FGR	Flue gas recirculation
FSNL	Full speed no load
GT	Gas turbine
HP	High pressure
HRH	Hot reheat
HRSG	Heat recovery steam generator
HV	High voltage
IGV	Inlet guide vanes
IP	Intermediate pressure
KJ	Kilojoule
KW	Kilowatt
LP	Low pressure
LV	Low voltage
MP	Medium pressure
MW	Megawatt
PC	Personal computer
OTSG	Once through steam generator
RH	Reheater or reheated
SCR	Selective catalytic reduction
SFC	Static frequency converter
SH	Superheater or superheated
SIF	Safety instrumented function
SIL	Safety integrity level
SOE	Sequence of event recorder
ST	Steam turbine
STIG	Steam injection gas turbine
TOC	Total organic carbon
1oo2D	One out of two with diagnostic
2oo3	Two out of three

References

1. "Hydrocarbon Processing and Petroleum Refiner," September 1962.
2. Lipták, B. G., *Instrument Engineers' Handbook,* Vol. 1, Measurement and Analysis, 4th edition, Sect. 1.10, Boca Raton, FL: CRC Press/ISA, 2003.
3. Lipták, B. G., *Instrument Engineers' Handbook,* Vol. 3, Process Software and Digital Networks, 3rd edition, Sect. 5.3, Boca Raton, FL: CRC Press/ISA 2002.
4. ANSI/ISA-77.42.01-1999, "Fossil Fuel Power Plant Feedwater Control System—Drum Type," Research Triangle Park, NC: ISA, 2002.
5. ANSI/ISA-77.44.01-2000, "Fossil Fuel Power Plant Steam Temperature Control System—Drum Type," Research Triangle Park, NC: ISA, 2002.
6. ANSI/ISA-77.13.01-1999, "Fossil Fuel Power Plant Steam Turbine Bypass System," Research Triangle Park, NC: ISA, 2002.
7. Marszal, E., and Scharpf, E., *Safety Integrity Level Selection,* Research Triangle Park, NC: ISA, 2002.
8. Lipták, B. G., *Instrument Engineers' Handbook,* Vol. 3, Process Software and Digital Networks, 3rd edition, Chapter 2, Boca Raton, FL: CRC Press/ISA 2002.
9. ASME PTC 22-1997, "Performance Test Code on Gas Turbines," New York: American Society of Mechanical Engineers, 1997.
10. ANSI/ASME PTC 4.4-1981, reaffirmed 1992, "Gas Turbine Heat Recovery Steam Generators," New York: American Society of Mechanical Engineers, 1992.
11. ASME PTC 6-1996, "Performance Test Code 6 on Steam Turbines," New York: American Society of Mechanical Engineers, 1996.
12. ISO 2314-1989, "Gas Turbines—Acceptance Tests," Geneva, Switzerland: International Organization for Standardization, 1989.

Bibliography

Agresti, M., Camporeale, S. M., and Fortunato, B., *Realizzazione di un Programma per la Simulazione Dinamica di Turbine a Gas in Ambiente Matlab-Simulink*, Matlab Conference 2000, Interventi della Terza Conferenza Italiana degli Utenti Matlab, February 8, 9, 2000.

Babcock & Wilcox, *Steam—Its Generation and Use,* 40th edition, Barbertoh, Ott, 1991.

Belding, J. A., "Choosing an Industrial Cogeneration System," *The Cogeneration Journal,* Vol. 2, No. 4, 1987.

Brooks, F. J., "GE Gas Turbine Performance Characteristics," GE Power Systems GER-3567h, October 2000.

Butler, C. H., *Cogeneration: Design, Financing, and Regulatory Compliance,* New York: McGraw-Hill, 1984.

Chapman, J., Wood, D., Williams, M., and Cravey, D., "Ethernet Connectivity, Modular 'Split Architecture' Software, and System Diagnostics in 'Part 75' Continuous Emisson Monitoring Systems in Cogeneration Plants," *Proceedings, ISA EXPO 2003,* Houston, TX, October 2003.

Chase, D. L., "Combined-Cycle Development Evolution and Future," GE Power Systems GER-4206, April 2001.

Crosa, G., Pittaluga, F., Trucco, A., Beltrami, F., Torrelli, A., and Traverso, F., "Heavy Duty Gas Turbine Plant Aerothermodynamic Simulation Using Simulink," *Transactions of the ASME, Journal of Engineering for Gas Turbine and Power,* Vol. 120, 1988.

Diamant, R. M. E., *Energy Conservation Equipment,* New York: Nichols Publishing., 1984.

Diegel, D., Scholz, B., and Kutzner, R., "Turbine-Generator Automation Automatically Better," *Siemens Power Journal On-Line,* November 2001.

Dukelow, S. G., *The Control of Boilers,* 2nd edition, Research Triangle Park, NC: ISA, 1991.

Erikson, D. M., Day, S.A., and Doyle, R., "Design Considerations for Heated Gas Fuel," GE Power Systems GER-4189b, March 2003.

Fruehauf, P. S., and Hobgood, J. V., "Dynamic Simulation of a Steam Generation Process with Cogeneration," *Procedings, ISA 2002,* Chicago, IL, 21–24 October 2002.

Griffin, P. R., and Elmasri, M., "Economically Optimized Loading of a Combined Heat and Power Facility," ASME Paper No. 97-GT-179, New York, 1997.

Gruhn, P., and Cheddie, H. L., *Safety Shutdown Systems: Design, Analysis, and Justification,* Research Triangle Park, NC: ISA, 1998.

Henstorf, B., "The Economics of Control Modernization," *Siemens Power Journal On-Line,* May 2002.

Immonen, P. J., *Mathematical Models in Cogeneration Optimization,* 9th Annual Joint ISA POWID/EPRI Controls and Instrumentation Conference (POWER 1999), 1999.

Immonen, P., and Savolainen, M., *Object-Oriented Approach to Cogeneration Simulation and Optimization,* SIMS 92 Simulation Conference on Simulation of Chemical and Power Plant Processes, Lappeenranta, Finland, June 10–12, 1992.

Johnke, T., and Mast, M., "Gas Turbines Power Boosters," *Siemens Power Journal On-Line,* May 2002.

Johnson, D., Miller, R. W., and Ashley, T., "Speedtronic Mark V™ Gas Turbine Control System," GE Power Systems GER-3567h, October 2000.

Jones, C., and Jacobs, J. A. III, "Economic and Technical Considerations for Combined Cycle Performance Enhancement Options," GE Power Systems GER-4200, October 2000.

Kochenburger, A., "Control in Every Detail," *Siemens Power Journal On-Line,* May 2002.

Koloseus, C., and Shepherd, S., "The Cheng Cycle Offers Flexible Cogeneration Options," *Modern Power Systems Magazine,* March 1985.

Kosla, et al., "Inject Steam in a Gas Turbine, but not Just for NO_x Control," *Power Magazine,* February 1983.

Laskar, S. K., *Gas-Based Cogeneration Plant—an Indian Experience,* 12th Annual Joint ISA POWID/EPRI Controls and Instrumentation Conference, (POWER 2002), 3–7 June 2002.

Legerton, D., Del Signore, S., McMahon, J., and Knowles, R., *Combined Cycle and Cogeneration Applications in the United Kingdom,* 12th Annual Joint ISA POWID/EPRI Controls and Instrumentation Conference (POWER 2002), 3–7 June 2002.

Li, K. W., and Priddy, A. P., *Power Plant Design,* New York: John Wiley & Sons, 1985.

Limaye, D. R., "Cogeneration: Trends and Prospects," *The Cogeneration Journal,* 1985.

Limaye, D. R., *Planning Cogeneration Systems,* Atlanta, GA: The Fairmont Press., 1985.

Lozza, G., *Turbine a Gas e Cicli Combinati,* Bologna, Italy: Società Editrice Esculapio, 1996.

Lu, S., "Dynamic Modeling and Simulation of Power Plant Systems," *Proceedings of the Institute of Mechanical Engineers, Part A: Journal of Power and Energy,* Vol. 213, No. 1, 1999.

Makkonen, S., "Managing the Changing Energy Markets," *Modern Power Systems,* Vol. 13, Issue 11, November 1993.

Orlando, J. A., *Cogeneration Technology Handbook,* Rockville, MD: Government Institutes, 1984.

Orsello, G., and Zambon, V., *Control System Retrofitting to Improve Gas Turbines Reliability and Availability,* PowerGen, Abu Dhabi, 2002.

Payne, W., *The Cogeneration Sourcebook,* Atlanta, GA: The Fairmont Press., 1985.

Reinsch, A. E., and Battle, E. F., *Industrial Cogeneration in Canada: Prospects and Perspectives,* Calgary, AB: Canadian Energy Research Institute, Study #24, March 1987.

Risso, C., and Bottoni, M., *Electrical Systems Control in Combined Cycle Cogeneration Power Plants,* INTERKAMA, 1999.

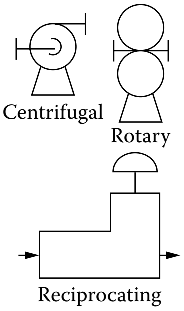
Smith, R. W., Polukort, P., Maslak, C. E., Jones, C. M., and Gardiner, B. D., "Advanced Technology Combined Cycles," GE Power Systems GER-3936a, May 2001.

Snow, C. A., and Realff, M. J., "Mill Steam and Power Optimization in a Real-Time Pricing Environment," *TAPPI Journal,* Vol. 81, No. 12, December 1998.

Wilkes, C., "Gas Fuel Clean-Up System Design Considerations for GE Heavy-Duty Gas Turbines," GE Power Systems GER-3942, 1996.

8.34 Pump Controls

F. B. HOROWITZ (1970) **B. G. LIPTÁK** (1985, 1995, 2005) **S. BAIN** (2005)



Flow sheet symbols

INTRODUCTION

This section describes the basic operation and controls of pumps and pumping stations, while the next section (8.35) concentrates on the optimization of the unit operation of pumping.

This section is divided into three main parts. The first provides a brief discussion of the processes into which the fluid is being transported by the pumps. The second part describes the types of pump designs, including centrifugal, rotary, and positive displacement designs and their basic methods of controls. The third part discusses some aspects of pumping system commissioning and operation.

The general discussion in this section is somewhat abbreviated, because related topics are also covered elsewhere in the handbook. Pumps, pumping stations, and metering pumps are also discussed in Section 7.4, and variable-speed drives in Section 7.10 in Chapter 7 of this volume. In addition, metering pumps are also covered in Section 2.14 in Chapter 2 in the first volume of this handbook.

Some of the pumping system-related terms, abbreviations, and conversion factors are described in Table 8.34a.

THE PROCESS

A pump is a liquid transportation device that must develop enough pressure to overcome the hydrostatic and frictional resistance of the process as it delivers the required fluid. These resistance components are unique characteristics of the process served and can be described by system curves. The system curve of a process relates the pressure (head) required and the amount of fluid flow that is being delivered.

System Curves

The characteristics of the system that is served by a pump or pumping station can be represented by a head-capacity

TABLE 8.34a
*Pump Terms, Abbreviations, and Conversion Factors**

Term	*Abbreviation*	*Multiply*	*By*	*To Obtain*
Length	L	ft	0.3048	m
Area	A	ft^2	0.0929	m^2
Velocity	v	ft/s	0.3048	m/s
Volume	V	ft^3	0.0283	m^3
Flow rate	Q_v	gpm	0.2272	m^3/h
		gpm	0.0631	l/s
Pressure	P	psi	6890	Pa
		psi	6.89	kPa
		psi	0.069	bar
Head (total)	H	ft	0.3048	m
NPSH	H	ft	0.3048	m
Output power (pump)	P_o	water hp (whp)	0.7457	kW
Shaft power	P_s	bhp	0.7457	kW
Input power (driver)	Pi	kW	1.0	kW
Efficiencies (%)				
Pump	E_p	—	—	—
Equipment	E_e	—	—	—
Electric motor	E_m	—	—	—
Utilization	E_u	—	—	—
Variable-speed drive	E_v	—	—	—
System efficiency index (decimal)	SEI	—	—	—
Speed	N or ω	rpm	0.1047	rad/s
Density	ρ	lb/ft^3	16.04	kg/m^3
Temperature		°F	—	°C

* From Reference 1.

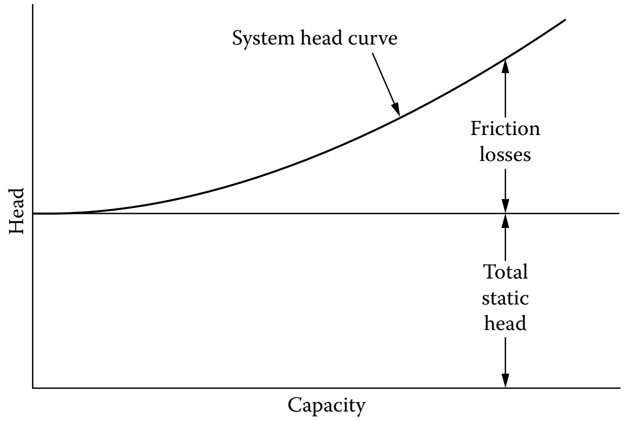


FIG. 8.34b
The system head curve is the sum of the static head and the friction losses that have to be overcome in order to pump liquid into the process.

system curve (Figure 8.34b). The head at any one flow capacity is the sum of the static and the friction heads. The static head does not vary with flow rate, as it is only a function of the elevation or back-pressure against which the pump is operating.

The friction losses are related to the square of flow and represent the resistance to the flow caused by pipe and equipment friction. A system curve tends to be flat when the piping is oversized and steep when the pipe headers are undersized. The friction losses also increase with the age of the plant. Therefore, the system curve for old piping tends to be steeper than for new piping.

A generalized equation describing the system curve of a process is given below:

$$P = H + F_f(Q^x) \qquad \textbf{8.34(1)}$$

where

P is the head pressure required to pump liquid into the process
H is the static or elevation head of the process
F_f is the friction factor of the process
Q is the flow rate of the incompressible fluid
x is an exponent that varies between 1.7 and 2.0, usually 2.0 is used

Static and Friction Pressures Constant static head (H) is the difference in pressure between the pump's intake and its discharge at zero flow, if the discharge piping is full. Usually it is measured between the pump's intake and the discharge side of its check valve, thereby correcting the measurement for any elevation differences between the two.

The constant static head can be the difference in elevation between the piping system's intake and discharge (corrected for liquid density if necessary). If the pump is discharging into a pressurized system such as a potable water distribution network or boiler, it will be the difference between the pump intake pressure and the system's pressure.

In any case, the constant discharge head is the pressure the pump works against when the piping system is full and pressurized, but there is no flow through the pumps' discharge pipes. As its name implies, it is comparatively constant. However, it may change, if for example the pump takes suction from a well and the well level drops, or if pumping to a water storage tower and the water level in the tower rises.

Therefore, the constant discharge head is constant only in relation to variations in one process variable: flow. If other process variables (pressure, temperature, density, level) change, it will be affected; it is not constant.

The friction pressure component in Equation 8.34(1) is the pressure that is lost due to friction between the liquid and the pipe. It includes losses from turbulence in bends and in the conversion of velocity pressure to static pressure in pipe expansions. However, for practical purposes they are combined into a single term.

The value of exponent x in Equation 8.24(1*)* is not critical. However, the controls can be improved by an accurate knowledge of the system curve, so the recommendation is to measure the value of x during commissioning.

Types of System Curves Figure 8.34c illustrates the system curves of three different types of processes. Curve 1 corresponds to the closed-loop circulation of a fluid in a horizontal plane. Here, there is no static head component at all, and the parabola that describes the system starts at zero.

Curve 2 is the system curve for a condenser water circulation network. Here, a limited amount of static head is present, because the pump must return the water to the top of the cooling tower. This curve also illustrates that the friction losses tend to increase when material builds up on the inside of the pipe, because it is no longer new. Curve 3 gives an example of a process dominated by static head. This is the case when feedwater is being pumped into a boiler drum. This curve is flat and is relatively insensitive to changes in system flow.

As will be discussed later in more detail, when the system curve is flat, there is little advantage to variable- or multiple-speed pumping, and the usual response to system flow variations is the stopping and starting of parallel pumps. Inversely, if the system curve is steep, substantial energy savings can be obtained from the use of booster, multiple-, or variable-speed pumps.

Open and Loop-Type Systems Hydraulic systems can be open (noncirculating) or loop-type, as illustrated in Figure 8.34d. Water supply and distribution systems in cities and buildings are typically open systems, whereas hot-and-chilled-water heating and cooling systems of plants are typically loop-type systems.

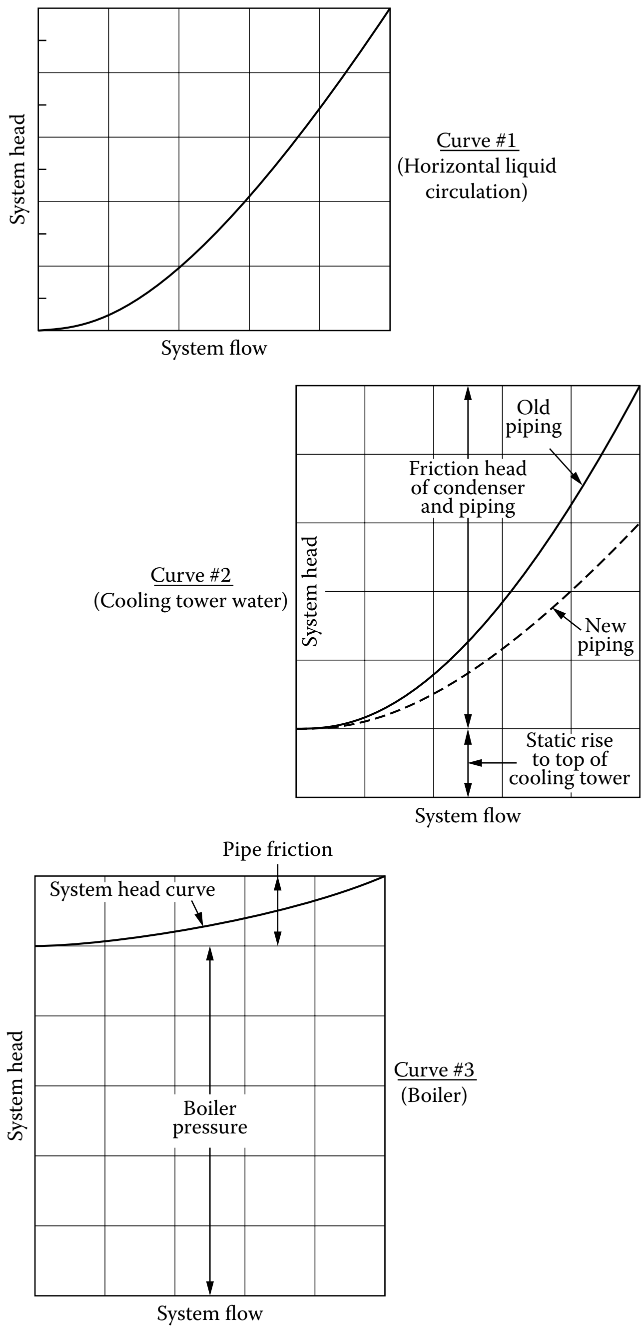


FIG. 8.34c
System curves vary as the static-head components change in various processes. (Adapted from Reference 1.)

Hydraulic systems must also be evaluated as to whether their flow is restricted or unrestricted. Restricted-flow systems are those that include valves that regulate the flow through the system. Hot-and-chilled-water systems, for example, are restricted-flow systems, because manual or automatic valves control their flow. Unrestricted-flow systems include sewage and stormwater lift stations as well as the pumping of municipal water into elevated storage tanks.

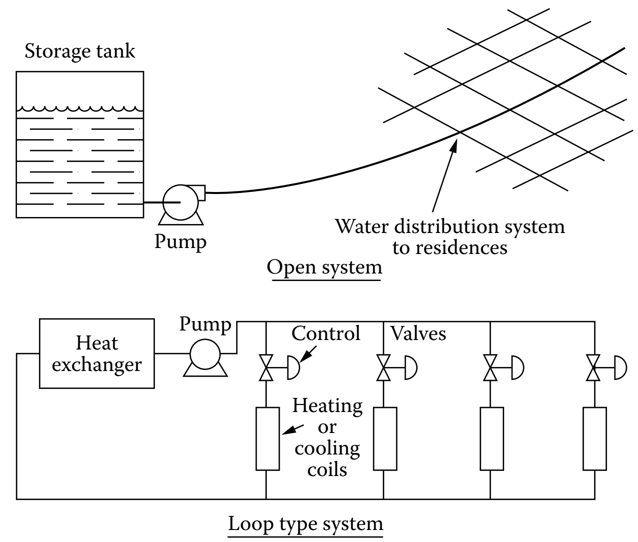


FIG. 8.34d
Pumping systems can be of the open or the loop type.

In actual systems, a single-system head curve may not exist. As illustrated in Figure 8.34e, what often exists is a system head band. This is because the distribution of active loads shifts the system curve within a wide band, as this figure shows.

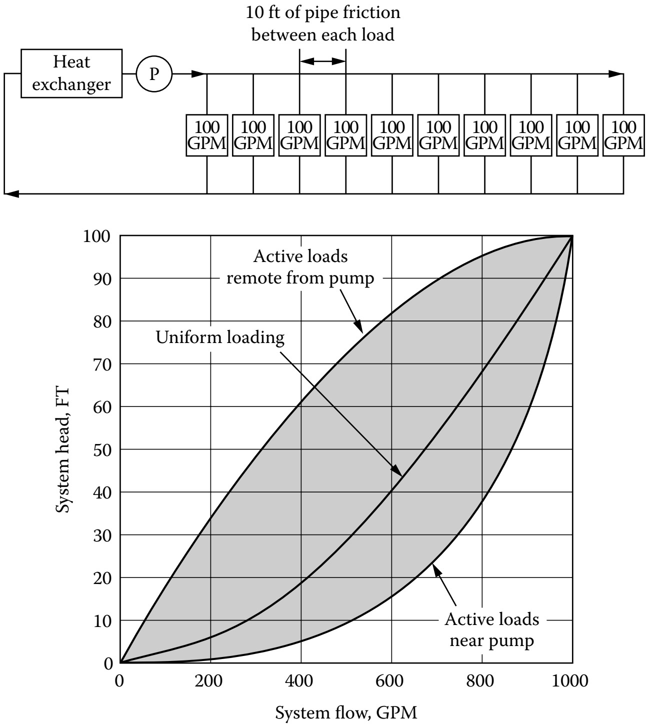


FIG. 8.34e
The system head drops when active loads are near the pump and it rises when the most remote loads are active.[4]

PUMP TYPES

There are many types of pumps; however, almost all flow is pumped by just two types: centrifugal pumps and reciprocating positive displacement (PD) pumps. As a result, this section will focus on these two types, and particularly on centrifugal pumps, which probably pump more flow than all other types combined.

Displacement and Centrifugal Designs

The two main pump designs are the positive displacement and centrifugal pumps. Vertical centrifugal units can pump water from depths up to 2000 ft (600 m), and horizontal units can transport process fluids from clear water to heavy sludge at rates up to 100,000 gpm (6.3 m^3/s). The centrifugal designs are of either the radial-flow or the axial-flow type.

Liquid enters the radial-flow designs in the center of the impeller and is thrown out by the centrifugal force into a spiral bowl. A number of impeller designs are illustrated in Figure 7.4b in Chapter 7. The axial-flow propeller pumps are designed to push rather than throw the fluid upward. Mixed-flow designs are a combination of the two.

In positive displacement pumps, a piston or plunger inside a cylinder is the driving element as it moves in reciprocating motion (Figures 7.4n and 7.4o). The stroke length and, thus, the volume delivered per stroke is adjustable within a 10:1 range. Rangeability can be increased to 100:1 by the addition of a variable-speed drive.

The plunger designs are capable of generating higher discharge pressures than the diaphragm types, because of the strength limitation of the diaphragm. The strain on the diaphragm is reduced if it is not attached directly to the plunger but is driven indirectly through the use of a hydraulic fluid. Because solids will still settle in the pump cavities, these designs are all limited to relatively clean services. For slurry service, the hose-type design is recommended. This design eliminates all the cavities, although the seating of the valves can still be a problem.

Pump Design Variations

If one attempted to give a more-or-less complete list of pump designs, the list would include the following:

Progressing cavity pumps are suitable for pumping liquids with entrained solids that would jam or abrade normal pump components, and liquids with low-shear requirements. The process industries use progressing cavity pumps for chemicals that may crystallize, polymers that need low-shear pumping, and liquids with some entrained solids. Progressing cavity pumps depend on the process liquid for lubrication, and they are easily damaged by low suction pressure. They behave like positive displacement pumps but are considerably less accurate than reciprocating pumps, and their accuracy deteriorates with use, to the 2–5% level.

Peristaltic pumps handle similar liquids as do the progressing cavity designs but are generally more economical in small (ml/s) sizes. The process industries use peristaltic pumps in real-time wet chemistry analyzers that depend on accurately metering small flows.

Gear, lobe, and vane pumps deliver a relatively constant flow at constant speed with large changes in discharge pressure and, therefore, approximate the characteristics of positive displacement pumps. These pumps cover the viscosity range from less than 1 centipoise up to 500,000 centipoises. The usual application of this type of pump is for viscous liquids and slurries that are beyond the capabilities of centrifugal pumps.

Rotary screw pumps are ideally suited for sludge and slurry services, and can safely pump high solids including live fish. A rotary screw pump design is also in development as an artificial heart. Rotary screw pumps behave similarly to centrifugal pumps.

In *lift pumps*, compressed gas, usually air, is blown into the bottom of a submerged updraft tube. The gas bubbles reduce the average density and, therefore, the hydrostatic head inside the tube. As the head is lower on the inside of the tube, the manometer effect induces the surrounding fluid to enter the updraft tube. This is a maintenance-free means of lifting large volumes of slurries over low elevations but is comparatively inefficient if one considers the compressor power required.

The Archimedes Screw is the earliest true pump and is still in use. An advantage of a screw is that it naturally changes its capacity to accommodate changing liquid levels without the complexity of a control system.

Air pumps use compressed air or steam to displace accumulated liquids from tanks (Figure 8.6llll).

Diffuser micropumps are generally specialized for use in small-scale work.

Pumps specialized for high-vacuum work, such as *diffusion and cryogenic pumps.*

CENTRIFUGAL PUMPS

Some centrifugal pumps use centrifugal force to throw liquid radially outward while others, such as propellers, use a screw-type action that results in axial flow. Between these two extremes, there is a whole continuum of impellers that change their pumping action from highly centrifugal radial to axial flow. The line separating centrifugal and axial-flow pumps is vague, and the behavior of these pumps is usually described by the same laws.

Figure 7.4b in Section 7.4 describes a number of impeller designs, including both radial and axial-flow types. Common characteristics of centrifugal pumps are high efficiency (over 90% in case of large pumps); they have only one moving part (the impeller with bearings); they deliver smooth, steady flow; they have a rangeability of about 4:1; and they are relatively insensitive to air-locking, but are susceptible to cavitation.

Vertical centrifugal pumps can lift water from depths of up to 2000 ft (600 m) and horizontal pumps can transport process fluids from clear water to heavy sludge at rates up to 100,000 gpm (6.6 m^2/s). The centrifugal pump is the most common type of process pump, but its application is limited to liquids with viscosities under 3000 centistokes.

Pump Curves

Figure 8.34f illustrates the typical pump curves of a single impeller pump.

Efficiency The typical range of pump efficiencies is from 60 to 85%. Pump efficiency is the ratio of the useful output power of the pump to its input power. Using the symbols defined in Table 8.34a, it is calculated (in both SI and US units) as follows:[2]

$$E_p = \frac{\text{pump output}}{P_i} = \frac{\text{SPQH}_t}{P_i} \text{ (SI units)} \qquad \textbf{8.34(2)}$$

$$E_p = \frac{\text{pump output}}{\text{bhp}} = \frac{\text{SPQH}_t}{\text{bhp} \times 550} \text{ (U.S. customary units)} \qquad \textbf{8.34(3)}$$

where

E_p = pump efficiency, dimensionless
P_i = power input, kW (kN·m/s)
SP = specific weight of water, lb/ft^3(kN/m^3)
Q = capacity, ft^3/s (m^3/s).
H_t = total dynamic head, ft (m)
bhp = brake horsepower
550 = conversion factor for horsepower to ft lb_f/s

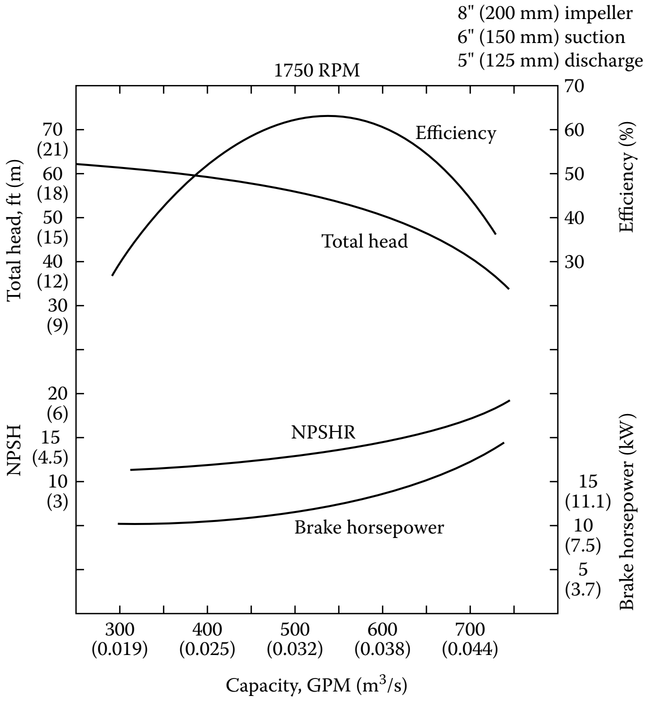


FIG. 8.34f
Typical characteristic curves of a single-impeller centrifugal pump.

Formulas Table 8.34g provides a summary of the more common formulas that can be used in connection with pumping calculations.

TABLE 8.34g
Common Formulas Used in Connection with Pumping Calculation[1]

Formula for	*Conventional Units*	*SI Units*
Head	H = psi × 2.31/SG* (ft)	H = kPa = 9.8023/SG* (m)
Output power	$P_o = Q_v \times H \times$ SG*/3960 (hp)	$P_o = Q_v \times H \times$ SG*/367 (kW)
Shaft power	$P_s = \frac{Q_v \times H \times SG^*}{39.6 \times E_p}$(hp)	$P_s = \frac{Q_v \times H \times SG^*}{3.67 \times E_p}$(kW)
Input power	$P_i - P_s \times 74.6/E_m$ (kW)	$P_i = P_s \times 100/E_m$ (kW)
Equipment efficiency, %	(Constant speed pumps) (Variable speed pumps)	$E_e = E_p \times E_m \times 10^{-2}$ $E_e = E_p \times E_m \times E_v \times 10^{-4}$
Utilization efficiency, %	Q_D = design flow Q_A = actual flow H_D = design head H_A = actual head	$E_u = \frac{Q_D \times H_D}{Q_A \times H_A} \times 100$
System Efficiency Index [see Eq. 8.34 (4)]		SEI = $E_e \times E_u \times 10^{-6}$

*SG = specific gravity

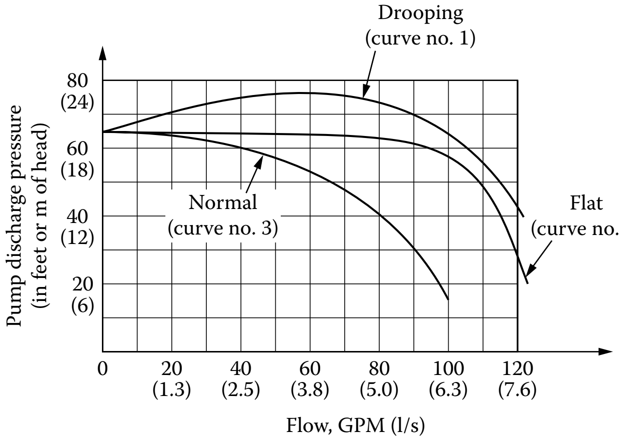


FIG. 8.34h
The centrifugal pump's curve can be drooping (1), flat (2), or normal (3).

Characteristic Pump Curves The head-capacity curve is the operating line for the pump at constant speed and impeller diameter. The characteristic curve of a pump describes the variation of its discharge pressure with volumetric flow. The discharge pressure is the total of the velocity and static pressures.

Three types of head-capacity curves are shown in Figure 8.34h, illustrating various relationships of capacity to discharge pressure. The capacity varies widely with changes in discharge pressure for all curves, but the shape of the curve determines the type of control that may be applied.

The shape of the head-capacity curve is an important consideration in pump selection. Curve 1 is referred as a *drooping* curve, curve 2 is called a *flat* curve, and curve 3 would be considered *normal*. For on/off switching control, curves 1 and 2 are satisfactory as long as the flow is above 100 gpm (6.3 l/s). Below this flow rate, curve 1 allows for two flows to correspond to the same head, and curve 2 may drop to zero flow to obtain a small head increase. Both are, therefore, unstable in this region. Curve 3 is stable for all flows and is best suited for throttling service in cases in which a wide range of flows is desired.

Matching the Pump(s) to the Process Figure 8.34i shows both the head-capacity curve of a centrifugal pump and the system curve of a process. When such a system is uncontrolled, the operating point of the system will be the point at which the pump and system curves cross each other.

If the process flow is controlled, a new system curve has to be artificially generated. This can be done 1) by generating a new system curve through the introduction of extra pressure drop in a control valve, or 2) by changing the pump curve through changing the pump speed. Figure 8.34j illustrates the case where a particular process flow is established by throttling a valve and, thereby, changing the unthrottled (solid) system curve into a throttled system curve (dotted).

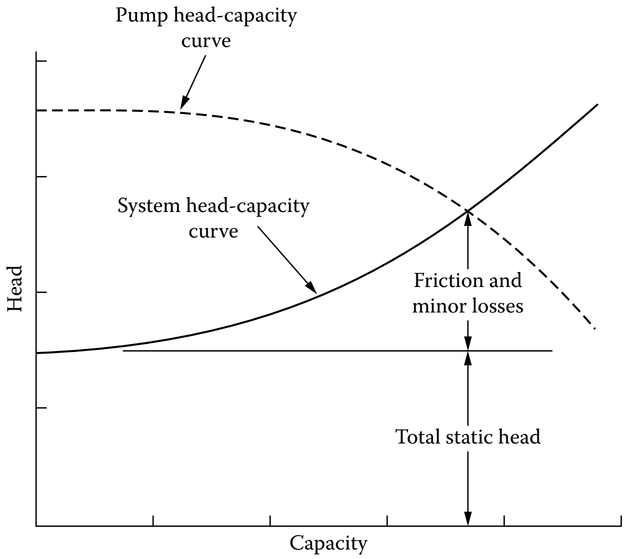


FIG. 8.34i
The system curve crosses the pump curve at the operating point.

Throttling with a control valve makes the apparent pump curve steeper, so that it will cross the system curve at the desired flow (operating point). This modification occurs at the cost of introducing an artificial pressure drop of ($H_p - H_s$), which burns up pumping energy and, therefore, reduces operating efficiency. In addition to that energy waste, the pump will also operate in a less-efficient region, as shown in Figure 8.34k by the throttled operating point #1 (72%) and the unthrottled operating point #2 (80%).

Adjusting the Pump Speed

The affinity laws describe the relationships among changes in speed, impeller diameter, and specific gravity. With a given impeller diameter and specific gravity, pump flow is linearly proportional to pump speed, pump discharge head relates to

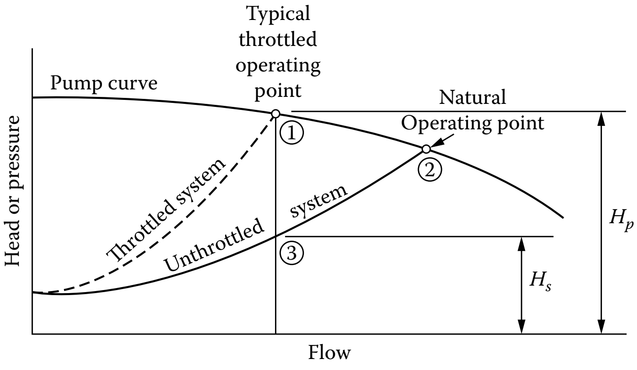


FIG. 8.34j
The addition of a control valve allows the control of flow at the cost of added pressure drop (wasted energy).

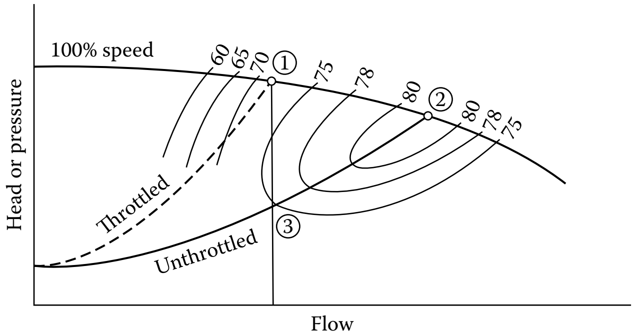


FIG. 8.34k
The throttled system not only wastes pumping energy through valve pressure drop but also operates at a less efficient point on the pump curve.

(approximately) the square of pump speed, and pump power consumption is proportional to the cube of pump speed.

As shown in Figure 8.34l, one can plot the system curves and the variable speed pump curves on a three-dimensional plot (a-pressure, b-flow, c-speed). On such a plot the system curves form one surface (surface A), and the pump curves form another characteristic surface (surface B). This then illustrates how the intersection of surfaces A and B is the operating line on which the variable-speed pump operates. (Section 8.35 will discuss the mathematical definition of the characteristic operating surfaces of pumps.)

Flow control via pump speed adjustment is less common than the use of throttling with valves, because most AC electric motors are constant-speed devices. If a turbine drive is considered, speed control is even more convenient. However, the advent of the pulse-width modulated (PWM) adjustable-speed drive with sensorless flux-vector control has brought adjustable-speed pumping into the mainstream of everyday applications.

In order to vary pump speeds with electric motors, one of the variable-speed drives described in Section 7.10 in Chapter 7 should be used. The efficiency curves (wire-to-shaft efficiencies) of the various variable-speed drives that are used on centrifugal pumps are shown in Figure 7.10z.

Variation of the pump speed generates a family of head-capacity curves, as shown in Figure 8.34m. If the impeller diameter is constant, the volumetric flow through the pump is proportional to its speed, and at reduced speeds, family of speed curves determines the flow rate (points 1, 2, or 3).

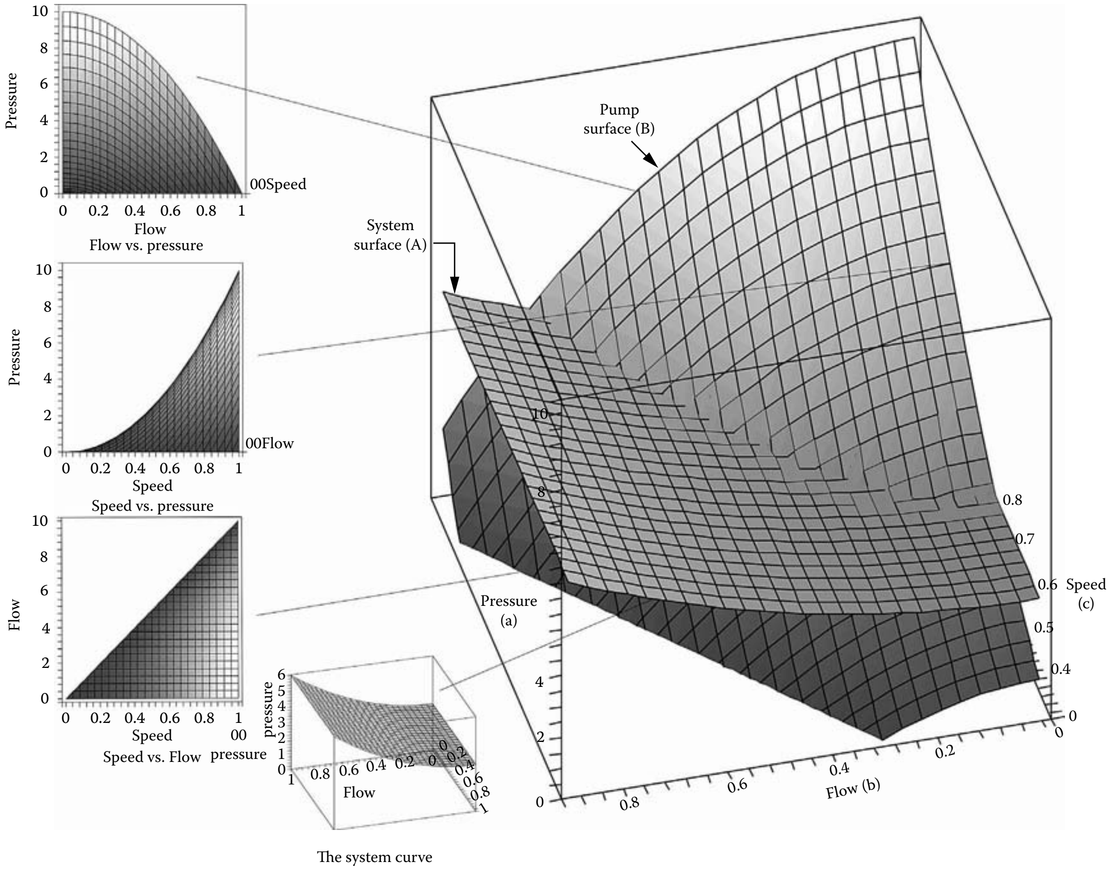


FIG. 8.34l
The variable-speed pump operates on the line where the surface formed by the system curves (A) intersects with the surface formed by the pump curves (B).

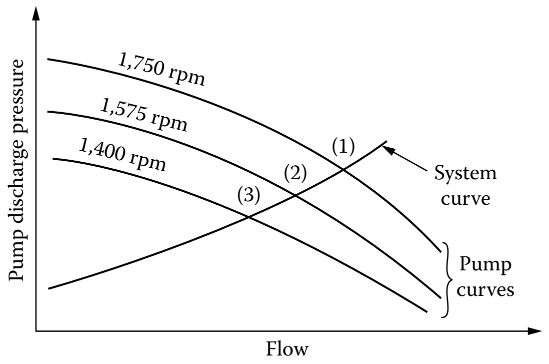
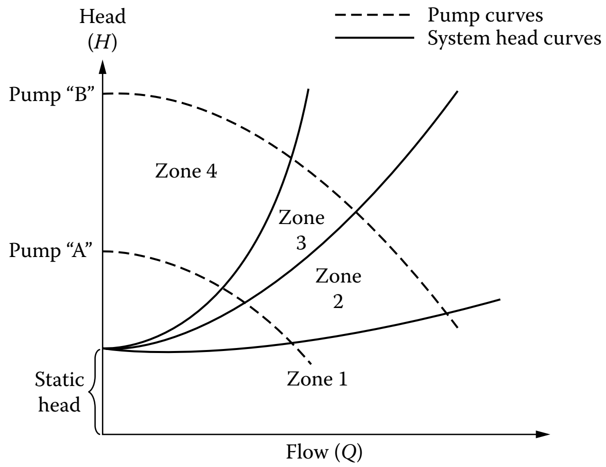


FIG. 8.34m
Variable-speed pump operation can be described by a family of head-capacity curves.

Because the area of peak pump efficiency falls on a parabolic path, speed throttling will usually not reduce the pump efficiency as much as valve throttling (Figure 8.34k). This increases the total energy savings obtained from pump speed control. As shown in Figure 8.34n, when the flow is reduced from F_1 to F_2, instead of wasting the excess pump head of $(P_1 - P_2)$ in pressure drop through a valve, that pump head is not introduced in the first place. Thereby, speed throttling saves the energy that valve throttling would have wasted.

When to Use Variable-Speed Pumps The shape of the system curve determines the saving potentials of using variable-speed pumps. As has been discussed, all system head curves are parabolas ($H \sim Q^2$), but they differ in the steepness of these curves and in the ratio of static head to friction drop. As shown in Figure 8.34o, the value of variable-speed pumping increases as the system head curve becomes steeper.

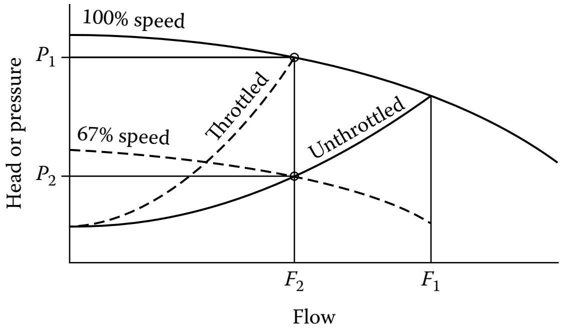
Studies indicate that in *mostly friction* systems (such as zone 4 in Figure 8.34o), the savings represented by variable-speed pumping will increase with reduced pump loading.[7] If, on the yearly average, the pumping system operates at not more than 80% of design capacity, the installation of variable-speed pumps can result in a payback period of approximately 3 years.



Zone	H_t/H_S	Pump sizes	Pump drives
1	< 1.2	Same	Constant
2	< 1.5	Various	Constant
3	< 2.0	Same	One variable
4	> 2.0	Various	All variable

FIG. 8.34o
Pumps and drives should be selected as a function of the steepness of the system curve.

The zones in Figure 8.34o are defined by the H_t/H_s ratio. The higher this number, the higher the zone number and the more justifiable is the use of variable-speed pumps. Figure 8.34p illustrates how the H_t/H_s ratio is calculated. The shaded areas identify the energy-saving potentials of variable-speed pumps. The values of H_t and H_s are identified on the basis of the average yearly flow rate (F_a) and determining its intersections with the pump and system curves. The larger the shaded area in Figure 8.34p, the higher the H_t/H_s ratio will be and, therefore, the shorter the payback for the use of variable-speed pumps is likely to be.



Speed, %	*Flow, %*	*Horsepower required, %*
100	100	100
90	90	73
80	80	51
70	70	34
60	60	22
50	50	13
40	40	6
30	30	3

FIG. 8.34n
Instead of wasting the unnecessarily introduced pump energy, speed is reduced so that such energy is not introduced in the first place.[6]

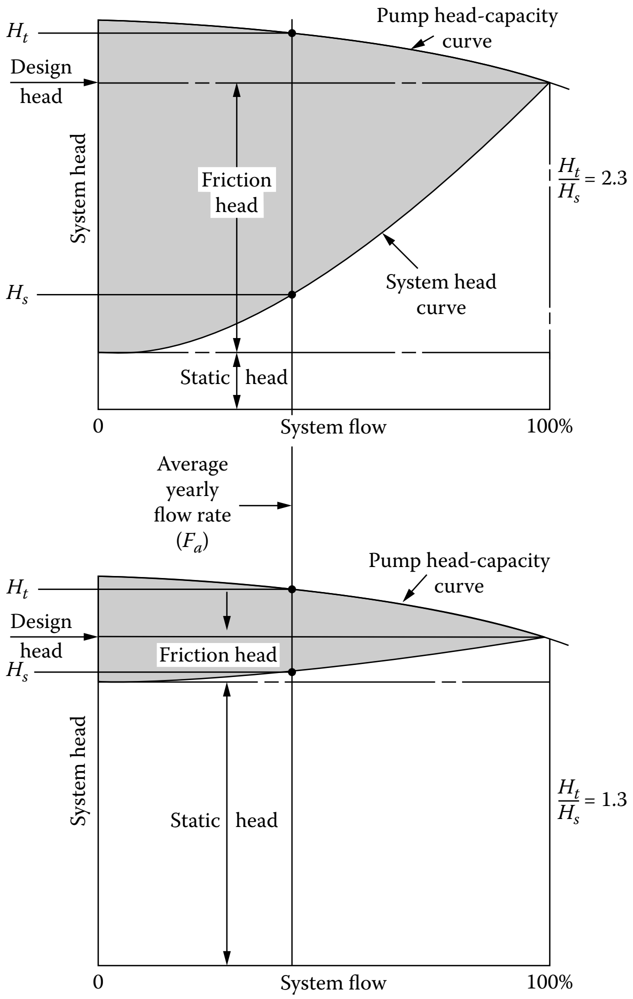


FIG. 8.34p
The higher the Ht/Hs ratio, the more justifiable is the use of variable-speed pumps.

Variable-Speed System Efficiencies The overall system efficiency index (SEI) of a variable-speed pump installation is determined as follows:

$$\text{SEI} = (E_p \times E_m \times E_v \times E_u)10^{-6} \qquad \textbf{8.34(4)}$$

where

E_p = the pump efficiency (%)
E_m = the motor efficiency (%)
E_v = the variable-speed drive efficiency (%)
E_u = the efficiency of utilization (%)

Figure 7.10z gave some typical wire-to-shaft efficiency values for variable-speed drives (E_v). The efficiencies of the variable-speed drives are represented as ranges, because each design has a different efficiency. Figure 8.34q provides some individual efficiency data for a number of variable-speed drive designs.

At 50% of rated speed, the variable-speed drive efficiency can be as low as 40% or as high as 70%, depending on the design selected. Naturally, the less efficient the variable-speed drive, the less expensive it is. Table 8.34r provides some cost information for variable-frequency induction motor drives.

The efficiency of utilization (E_u) is an indicator of the quality of the overall piping and equipment design. For example, in the piping distribution system illustrated in Figure 8.34s, Q_r might represent the required water flow and H_r the required pressure head to transport Q_r. Because three-way valves are used in this illustration, the actual flow (Q_a) is much higher than the required flow, and the head required to transport this actual flow (H_a) is also greater than what is required.

Therefore, the efficiency of utilization, which is defined in Equation 8.34(5), will also be rather low for the design described in Figure 8.34s.

$$E_u = \frac{Q_r H_r}{Q_a H_a} 100 \qquad \textbf{8.34(5)}$$

The numerical value of E_u can be accurately obtained only by testing. Once all four efficiencies are determined, they can be represented by a single system efficiency index curve (Figure 8.34t). By combining different pumps, motors, drives, and system designs, it is possible to arrive at a number of SEI curves. The relative advantages of different devices and designs can be evaluated quantitatively by comparing these curves.

Types of Variable-Speed Drives Section 7.10 in this volume provides a detailed description of the various variable-speed drive designs, their features including efficiencies, and costs. Table 8.34u provides a summary of some of the main features of these variable-speed drives.

Two-Speed Pumps Two-speed motors can be utilized on simple pumping systems in which accurate control of pump pressure is not necessary. Such motors should not be used in systems with high static head and low friction processes. On the other hand, for most friction systems, they can offer a reasonably efficient and inexpensive alternative to variable-speed pumping. As shown in Figure 8.34v, standard two-speed motors are available with speeds of 1750/1150 rpm (29/19 r/s), 1750/850 rpm (29/14 r/s), 1150/850 rpm (19/14 r/s), and 3500/1750 rpm (58/29 r/s).

Cavitation

Cavitation can be a severe problem; it can quickly destroy a pump. The cause of cavitation is that the pumped liquid flashes to vapor at one point in the impeller (where pressure is below the vapor pressure), and as the spinning impeller throws the liquid and vapor outward, the formed bubbles collapse as the pressure rises above the vapor pressure. When the collapsing bubbles come to the wall, they collide with extreme force. This gives rise to the characteristic sound of

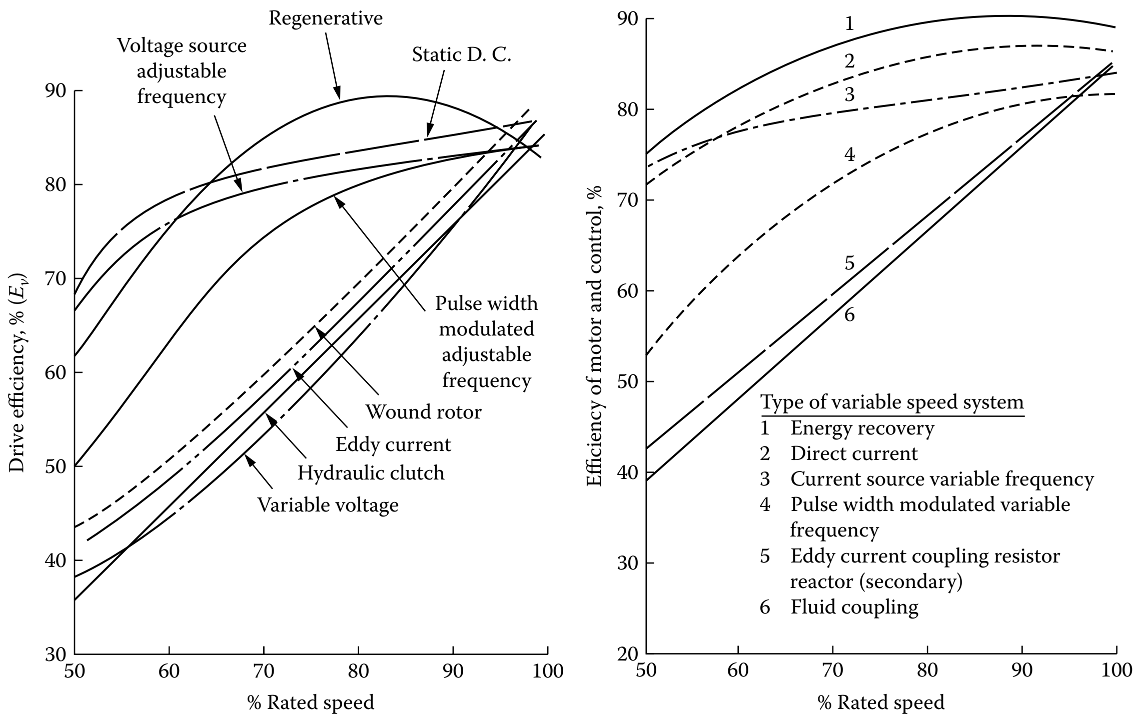


FIG. 8.34q
Variable-speed drives in the 100-HP and larger sizes offer a wide range of efficiencies.[9,10]

cavitation, and also to the consequent erosion, usually of the impeller.

For this reason, it is important to operate pumps only in their operating region where cavitation does not occur. Cavitation usually occurs when the pump is delivering high flow, but some pumps cavitate at low flow too, when liquid recirculating in the impeller passes through points of both low and high pressures (Figure 8.34w).

When one does not have reliable data on the locations of cavitation regions, it might be necessary to (very briefly) force the operating pump into cavitation intentionally. A practical way to force a pump to cavitate is to run it at or above design speed and then to restrict its intake gradually. Cavitation has a very characteristic and memorable sound that has been described as rocks rolling around inside the pump, or hammering on the impeller.

If a pump cavitates, one might try to eliminate or minimize the cavitation by improving inlet conditions. Just as restricting the inlet will force a pump into cavitation, so removing restrictions will often stop cavitation. Reducing the flow range of the pump usually helps also. If the pump cavitates at high flows, try to avoid those flows, possibly by starting a second pump sooner. If cavitation occurs at low flows, one might turn off the pump at low flows, based on some on/off flow control strategy.

Some impellers can accept a part, called an *inducer*, that reduces the pump's susceptibility to cavitation. An extreme option is to inject a compressible gas into the impeller. This reduces pump efficiency and capacity, but it can eliminate cavitation, because the gas acts as a spring inside the liquid, absorbing the drastic localized pressure changes, and so avoiding the pressure extremes that the pure liquid would experience.

However, cavitation is mainly a design concern and should be dealt with during design by ensuring that ample pressure is available around the impeller.

TABLE 8.34r
2004 Cost of Variable-Frequency Induction Motor Drives

Power Rating	*PWM Including Reactor*	*Current-Fed ASCI*
10 hp (7.5 kW)	$1,700	
20 hp (15 kW)	$2,500	
50 hp (37 kW)	$5,300	
100 hp (75 kW)	$7,500	$8,000
200 hp (150 kW)	$11,000	$12,000
500 hp (375 kW)		$18,000
1000 hp (750 kW)		$32,000

Note: The reactor price adds 50–90% to the base price of the variable-frequency drive electronics.

Net Positive Suction Head

When liquids are being pumped, it is important to keep the pressure in the suction line above the vapor pressure of the

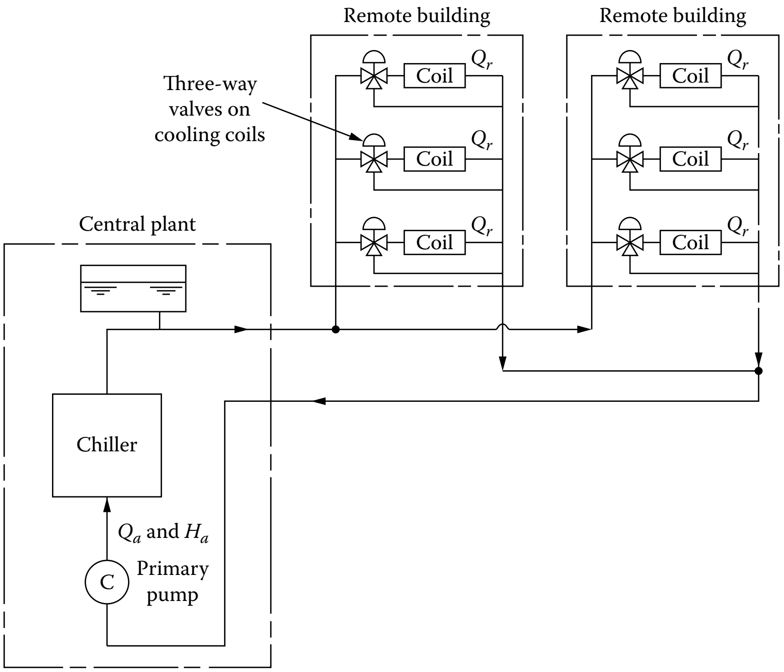


FIG. 8.34s
The efficiency of utilization of this design is low, because the water that is bypassing the coils is being circulated unnecessarily.[11]

fluid. The available head measured at the pump suction is called the *net positive suction head* (NPSH).

A pump at sea level that is pumping 60°F water, which has a vapor pressure of $h_{vp} = 0.6$ ft, and that is operating under a barometric pressure of 33.9 ft, has an available NPSH (NPSHA) of $33.9 - 0.6 = 33.3$ ft. If the impeller centerline is 3.4 ft below the surface elevation of the water being pumped and if the friction losses in the intake piping is 5.3 ft W.C, the NPSHA is $33.9 - 0.6 + 3.4 - 5.3 = 31.4$ ft. As shown in Figure 8.34x, the NPSHA increases with barometric

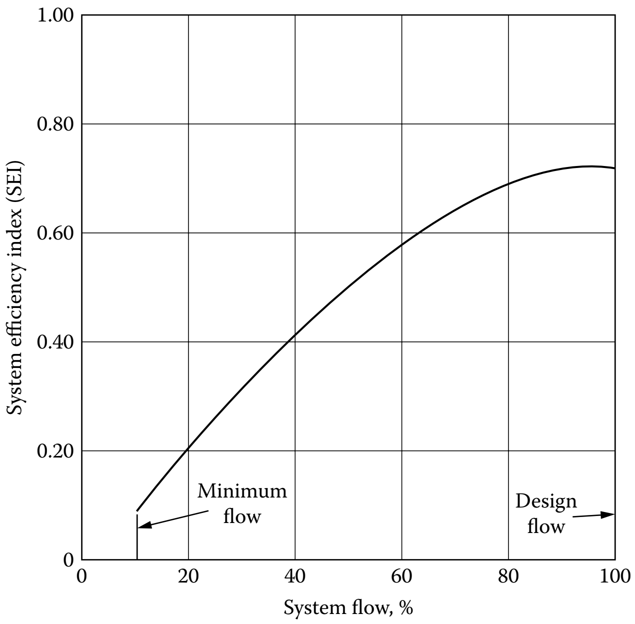


FIG. 8.34t
Overall system efficiency index (SEI) curves describe the total pumping efficiency as a function of load.[1]

TABLE 8.34u
Variable-Speed Drive Comparison

Drive Type	*Efficiency (at 70% Speed)*	*Turndown*	*Sizes (hp)*	*Component Requiring Replacement (Frequency in Yrs.)*
Wound rotor regenerative	High (85%)	2:1*	25–500+	Brush (3–4)
Direct current	High (80%)	Unlimited	1–500+	Brush (1–2)
Variable frequency	High (78%)	3:1	20–500+	—
Wound rotor	Med. (60%)	2:1*	25–500+	Brush (3–4)
Eddy-current clutch	Med. (58%)	5:1	20–500+	—
Fluid coupling	Med. (57%)	3:1	20–500+	—
Variable voltage	Low (52%)	Limited	10–100+	—
Mechanical	Low (50%)	6:1	1–100	Belt or chain (1–3)

*Unstable below 50%.

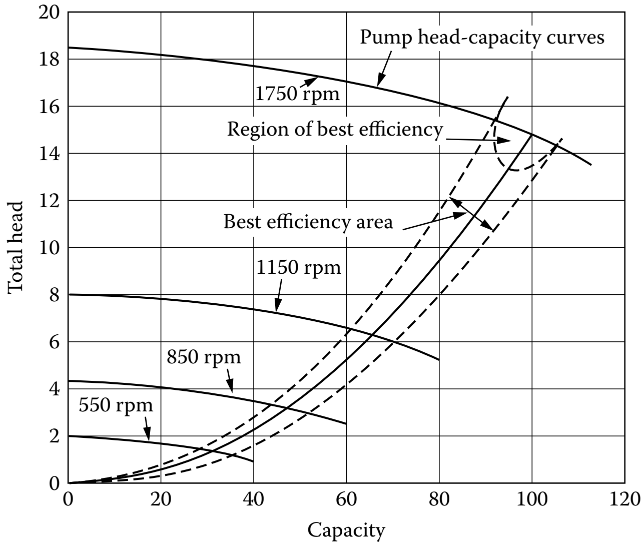


FIG. 8.34v
Several speed combinations are available in two-speed pumps.[1]

pressure and with static head, and it decreases as vapor pressure, friction, or entrance losses rise.

Figure 8.34x also illustrates the difference between *available* and *required* NPSH (NPSHA and NPSHR). Available NPSHA is the characteristic of the process and represents the difference between the existing absolute suction head and the vapor pressure at the process temperature. The required NPSHR, on the other hand, is a function of the pump design (Figure 8.34f). It represents the minimum margin between suction head and vapor pressure at a particular capacity that is required for pump operation. If this minimum NPSHR is not available, the pump will fail to generate the required suction lift and the flow will stop.

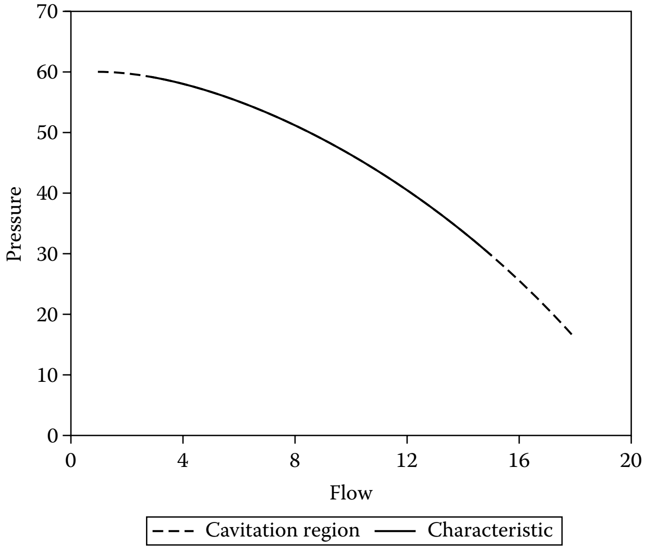


FIG. 8.34w
The dashed segments of the characteristic pump curve show the zones where the regions of cavitation might exist.

The *NPSHR* curve in Figure 8.34f describes the required amount of static head at the pump inlet required to avoid a discharge pressure from dropping more than 3% over the zero-cavitation condition (source of definition: ANSI/HI 9.6.1-1998, Paragraph 9.6.1.1). NPSHR is defined in absolute pressure, and typically for water at 0°C unless otherwise stated.

The available NPSH (the *NPSHA*) increases as 1) the barometric pressure increases, 2) the static pressure of the liquid at the entrance of the impeller, and 3) all other suction-side pressures increase. The NPSHA decreases as 1) the vapor pressure of the liquid increases, 2) friction or entrance losses rise, and 3) all other suction-side pressures decrease.

NPSH and Cavitation Traditionally, the NPSHR curve has been taken to define the onset of cavitation, and designers have concentrated on ensuring that the NPSHR is met under all operating conditions (Figure 8.34x). However, the accuracy of the NPSHR curve in defining the point when cavitation becomes significant is being questioned.

Indications are that cavitation actually begins sooner than previously believed. Allan Burdis (chairman of the Hydraulic Institute's NPSH Margin Committee) writes: "If you want to operate cavitation-free, you need NPSH margin ratios (NPSHA/NPSHR) of 4 to 5." Some evidence suggests that cavitation damage at NPSHA = NPSHR is actually less than the damage when the NPSHA is higher. The Hydraulic Institute's standard report reads: "There are studies that show the maximum cavitation damage can actually occur at NPSHA values that are twice the NPSHR or more for very high suction energy pumps."

Therefore, we might conclude that the science of cavitation prediction on the basis of NPSHA is still evolving, and if one wants to be positive, actual cavitation testing can be necessary.

Water Hammer

If a valve opening is suddenly reduced in a moving water column, this causes a pressure wave to travel in the opposite direction to the flow. When this pressure wave reaches a solid surface (elbow, tee, and so on), it is reflected and travels back to the valve. If, in the meantime, the valve has closed, a series of shocks, sounding like hammer blows, results. An example can illustrate this phenomenon.

Assume that 60°F (20°C) water is flowing at a velocity of 10 ft/s in a 3 in. Schedule 40 pipe, and a valve located 200 ft downstream is suddenly closed. The pressure rise and the minimum acceptable time for valve closure can be calculated. If the valve closes faster than this time limit, water hammer will result. In rigid pipe, the pressure rise (ΔP) is the product of water density (ρ in units of slugs/ft^3), the velocity of sound (c in units of ft/s), and the change in water

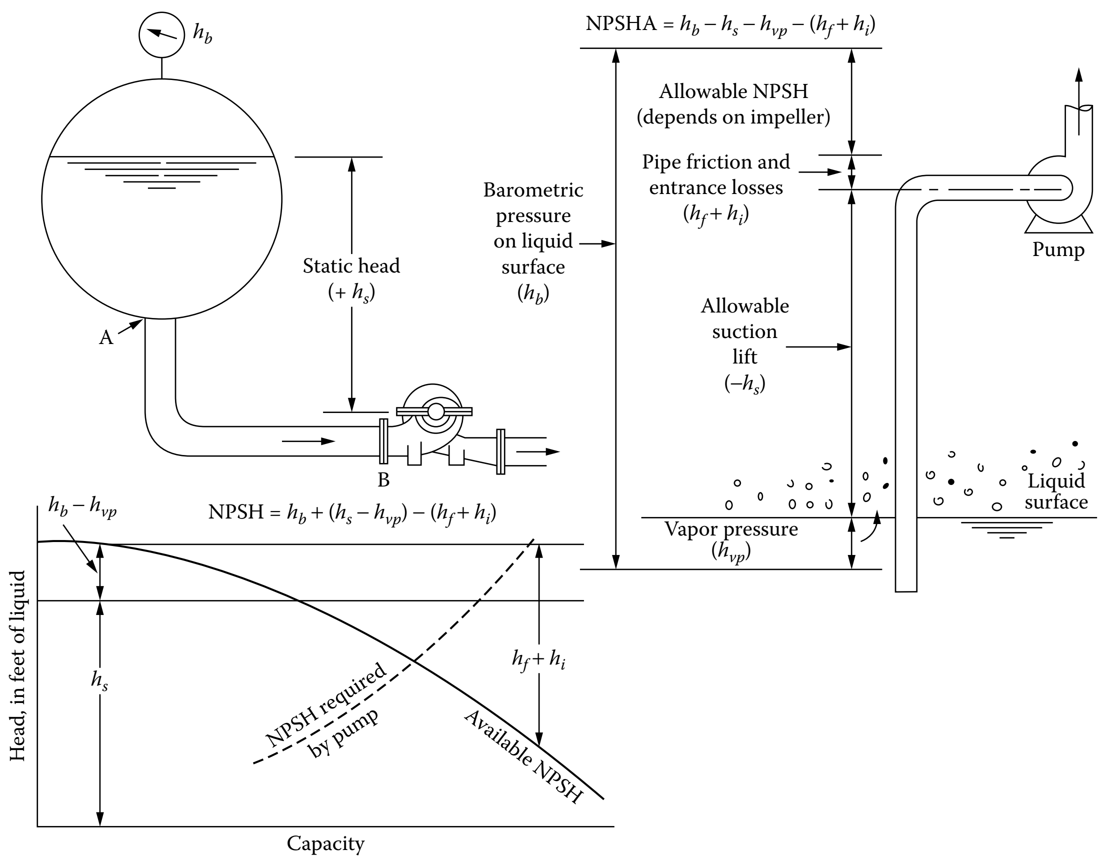


FIG. 8.34x
The available net positive suction head (NPSHA) increases with barometric pressure and static head, and decreases as vapor pressure, friction or entrance losses rise. (Adapted from Reference 3.)

velocity (ΔV in units of ft/s).[12] Therefore, the pressure rise can be calculated as follows:

$$\Delta P = -\rho c\,\Delta V = -(1.937)(4860)(-10)$$
$$= 94{,}138\ \text{lbf/ft}^2 = 653.8\ \text{PSI} \qquad \mathbf{8.34(6)}$$

In order to prevent water hammer, the valve closure time (t) must exceed the ratio of two pipe lengths ($2L$) divided by the speed of sound:

$$t = 2L/c = (2)(200)/4860 = 0.0823\ \text{seconds} \qquad \mathbf{8.34(7)}$$

Therefore, in this example, the valve closure should take more than 0.0823 sec.

The possible methods of preventing water hammer include (1) designing the system with low velocities, (2) using valves with slow closure rates, and (3) providing slow-closing bypasses around fast-closing valves, such as check valves.[15]

When water hammer is already present and the cause of it cannot be corrected, its symptoms can be treated (1) by adding air chambers, accumulators, or surge tanks; (2) by using surge suppressors, such as positively controlled relief valves; and (3) when water flows are split or combined, by using vacuum breakers to admit air and thereby cushion the shock resulting from the sudden opening or closing of the second split stream.

Pump Stations

When either the flow or the pressure requirements of the process are such that a single pump cannot meet them, pump stations consisting of two or more individual pumps have to be used. Multiple pumps operate in parallel and are used if the process flow rangeability exceeds the throttling capability of a single pump. Booster pumps are installed in series and are used to increase the total discharge pressure of the station.

Multiple Pumps in Parallel Individual centrifugal pumps have a rangeability of about 4:1, which can be obtained by either speed control or by discharge throttling. Pump turndown can be increased by 1) bypassing the unwanted flow, 2) turning the pump on and off, and 3) using multiple pumps.

When two or more pumps operate in parallel, the combined head-capacity curve is obtained by adding up their individual capacities at each discharge head, as illustrated in

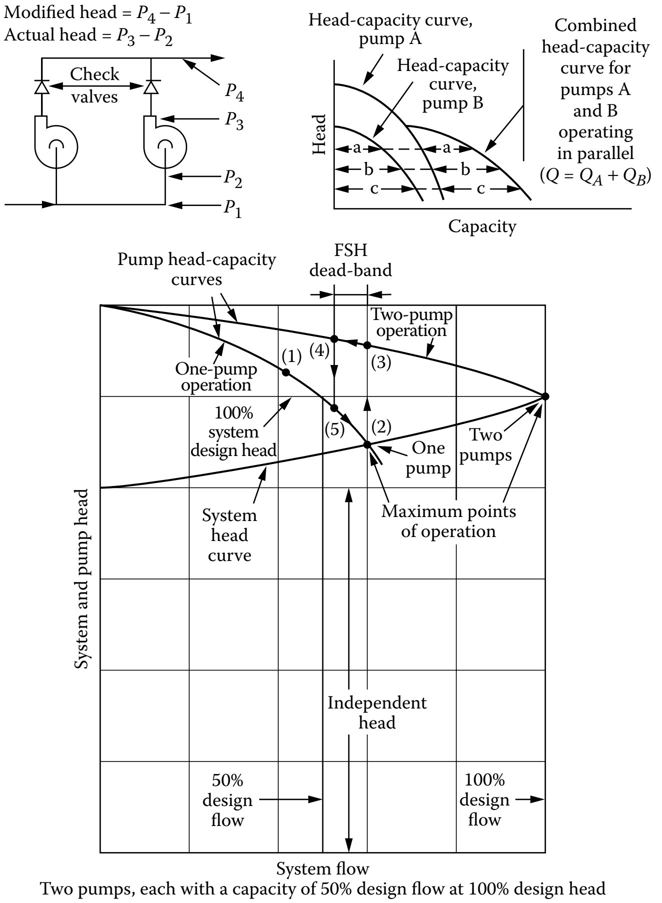


FIG. 8.34y

Pump turndown and rangeability can be increased by operating two or more pumps in parallel.

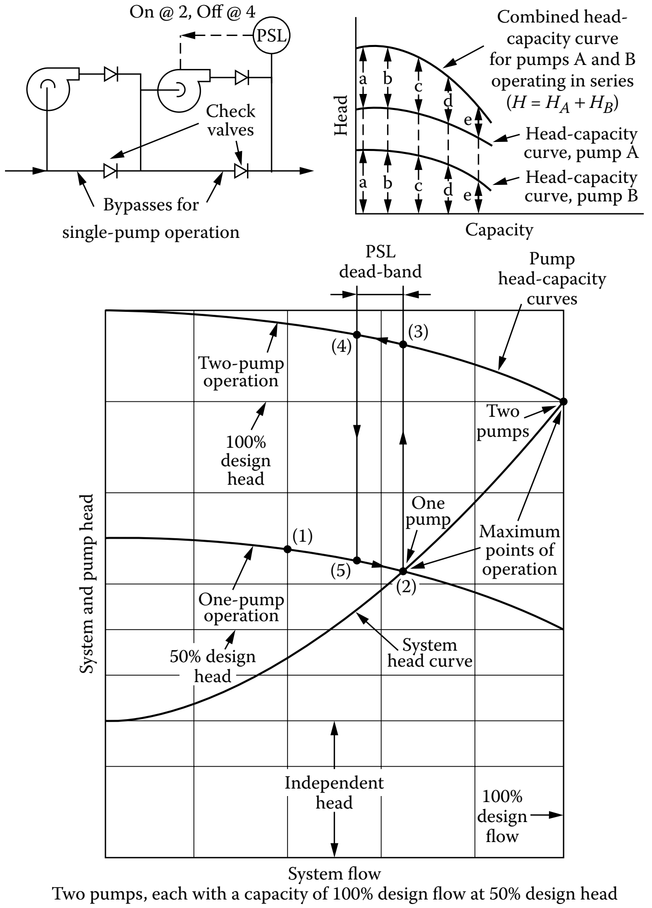


FIG. 8.34z

Multiple pumps in series are effective when the system head curve is steep. The two pumps illustrated by the lower graph are each capable of generating 100% design flow at 50% design head.[1,2]

Figure 8.34y. The total capacity of the pump station is found at the intersection of the combined head-capacity curve with the system head curve. This point also gives the head at which each of the pumps is operating.

If the selection is to be very accurate, the head-capacity curves should be modified by substituting the station losses (the friction losses at the suction and discharge of the individual pumps) so that the resulting "modified head" curve will represent the pump plus its valving and fittings.

When constant-speed pumps are used in parallel, the added increments of pumping can be started and stopped automatically on the basis of flow. As will be discussed later, a dead band is provided in these controls so that if a new pump is started at flow "*x*," the flow will have to drop to, say, "*x*–5%" before that increment is stopped.

Booster Pumps When two or more pumps operate in series the total head-capacity curve is obtained by summing up the pump heads at each capacity. When a booster pump is added to a main fed by several *parallel* pumps, the total head-capacity curve is obtained by adding the booster curve to the modified head of the parallel pumps at each capacity point.

Series pumping is most effective when the system head curve is steep, such as in Figure 8.34z. With such mostly friction loads, series pumping can substantially reduce the overpressure at low loads. Therefore, booster pumps or two-speed pumps can both be considered for the same kind of steep system curves. Multiple pumps in series are preferred from an operating cost point of view, but the capital cost investment of a single two-speed pump is lower.

When constant-speed pumps are used, the booster pump can be started and stopped automatically on the basis of pressure. In this case, an adjustable dead band is provided in the pressure switch. As an example, the normal operating point of the system can be point (1) in Figure 8.34z. As the load increases, the pump discharge head drops and when it reaches point (2) (the set point of the PSL), the booster pump is automatically started. As soon as the booster is on, the system operates at a point to the right of point (3) until the load drops off again.

The booster pump stays on as the load drops below point (3) until the PSL turns it off at point (4). At this point, the system is automatically returned to the single pump operation at point (5). The dead band in the PSL prevents the on/off cycling of the booster pump at any particular load. The width of the dead band is a compromise: As the band is narrowed, the probability of cycling increases, while the widening of the band results in extending the periods during which the booster is operated unnecessarily. If the pumps are identical, their running times can be equalized by alternating them, so that the pump with the higher running time will be the one that is stopped first.

POSITIVE DISPLACEMENT PUMPS

Reciprocating pumps, such as the piston and diaphragm types, deliver a fixed volume of fluid per stroke. The control of these pumps is based on changing the stroke length, changing the stroke speed, or varying the interval between strokes. In all cases, the discharge from these pumps is a pulsed flow, and for this reason they are not suited to control by throttling valves. In practice, the volume delivered per stroke is less than the full stroke displacement of the piston or diaphragm.

This hysteresis is a result of high discharge pressures or high viscosity of the fluid pumped. Under these conditions, the check valves do not seat instantaneously. A calibration chart must, therefore, be drawn for the pump under actual operating conditions. A weight tank or level-calibrated tank is usually the reference standard. Because the discharge is a pulsed flow (Figure 8.34aa), it must be totalized and divided by the time interval to get average flow rate for a particular speed and stroke setting.

Metering inaccuracy is approximately ±1% of the actual flow with manual adjustment and ±1.5% with automatic positioning. Methods of stroke and speed adjustment are covered in detail in Section 7.4 in Chapter 7, and other features of metering pumps are discussed in Section 2.14 in Chapter 2 in the first volume of this handbook.

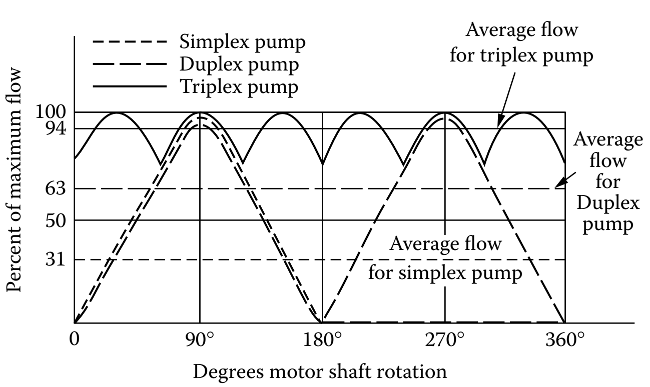


FIG. 8.34aa
Flow characteristics of simplex and multiple plunger pumps.

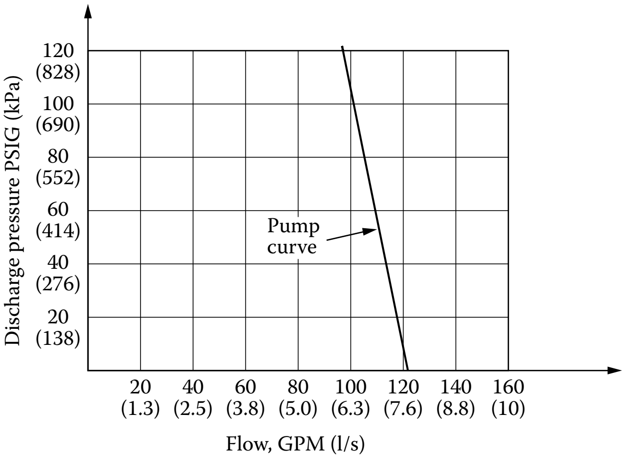


FIG. 8.34bb
The characteristic curve of a positive displacement (PD) pump operating at constant speed and stroke.

Reciprocating Pumps

Reciprocating pumps have a piston or plunger driving element inside a cylinder; the piston moves in and out, in a reciprocating motion (see Figures 7.4n and 7.4o). Liquid is sucked into the pump as the piston moves in, and is forced out as the piston moves out. The liquid flows through one-way valves on the intake to allow flow only into the pump, and on the discharge only to allow it out.

The reciprocating piston action results in each stroke displacing a positive, fixed volume of liquid almost regardless of pressure, hence the name. Figure 8.34bb shows a typical PD pump's flow-to-pressure characteristic. The main feature of this characteristic is that it is essentially constant-flow: The pressure varies greatly while the flow varies very little.

Many reciprocating PD pumps allow the stroke length to be adjusted, to change the pumping rate at constant speed, over a 10:1 range. Manufacturers offer higher stroke ranges, and theoretically they could adjust the stroke down to zero, but accuracy tends to decrease at low strokes, because of fixed losses such as valve leakage.

PD pumps normally develop higher pressures than centrifugal pumps, and their flow is generally unaffected by discharge pressure, which makes them well-suited to metering applications. They can handle higher viscosity liquids than centrifugal pumps, and they can also handle slurries, but they usually do not handle large entrained solids, as many centrifugal pumps can (Figure 8.34cc).

Liquid compressibility can become a consideration at extremely high pressures, so that volume flow (gpm, m^3/s) reduces slightly more than mass flow (lb/min, kg/s). PD pumps are usually considerably less efficient than centrifugal pumps except in high-head applications, so that large PD pumps are far less common than large centrifugal pumps.

Some PD pump assemblies have multiple pumps on a single shaft. This design is normally used to pump reagents

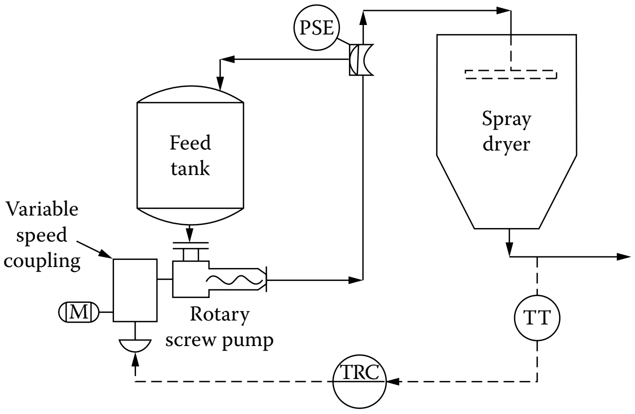


FIG. 8.34cc
When PD pumps are transporting slurries, it is advisable to use rupture discs to protect against the development of excessive discharge pressures.

or ingredients that need to be in a specific ratio. Operators adjust the individual pumps so that their volumes per stroke are precisely in the ratio required, then the pumps are driven by a common shaft so that they all run at the same speed and, thus, deliver flow in a specific ratio.

Because flow is essentially independent of pressure, PD pumps that operate against a closed discharge (closed isolation valve, blocked pipe, and so on) can develop very high discharge pressures that damage the equipment. For this reason, they often are provided with a pressure relief valve, or in applications with highly viscous liquids and slurries, possibly with a rupture disc to relieve the excessive pressures, back to the pump inlet or the intake source.

A PD pump's high pressure applies equally to the suction side of the pump. Vacuum pressure can damage a PD pump, although this is only common in progressing cavity pumps, where an internal vacuum pressure tends to delaminate the flexible stator seal from the rigid stator support. Vacuum pressure can also damage piping and flexible metal couplings.

NPSH and Cavitation It is important to keep the NPSH above 10 psia (69 kPa) or preferably above atmospheric. NPSH can be calculated as follows:

$$\text{NPSH} = P - P_v \pm P_h - \sqrt{\left(\frac{lvGN}{525}\right)^2 + \left(\frac{lvC}{980Gd^2}\right)^2} \qquad \textbf{8.34(8)}$$

where

P = feed tank pressure (psia)
P_v = liquid vapor pressure at pump inlet temperature (psia)
P_h = head of liquid above or below the pump center line (psid)
l = actual length of suction pipe (ft)
v = liquid velocity (ft/s)
G = liquid specific gravity
N = number of pump strokes per minute
C = viscosity (centipoise)
d = inside diameter of pipe (in)

Flow Rangeability Speed adjustment is also an effective method of adjusting a PD pump's output, again generally over a 10:1 range. One can combine both stroke and speed control to achieve rangeability to 100:1, and add on- to off-time (mark-to-space ratio) control to further extend that rangeability.

The PD pump's ability to operate over a wide rangeability may mean that only one pump is needed, which simplifies operation considerably. In process control, a PD pump's rangeability is important for chemical metering pumps, particularly those used for pH control, where the need for adjusting reagents to 1 part in 100, 1000, or more is not uncommon. Rangeability of a hundred to a thousand is seldom reliably achieved even with a PD pump, while it is impossible with centrifugal pumps having a rangeability of only 4:1 or 5:1.

Calibration A PD pump's check valves do not seat instantaneously, so accurate operation needs to be based on a calibration chart for the pump under actual operating conditions. Usually the reference standard for calibration is a calibration column or weigh tank that measures volume or mass. Calibrate the pump's flow by dividing the volume (or mass if using mass flow) by the time interval to get the average flow rate for a particular speed and stroke setting. Figure 8.34dd

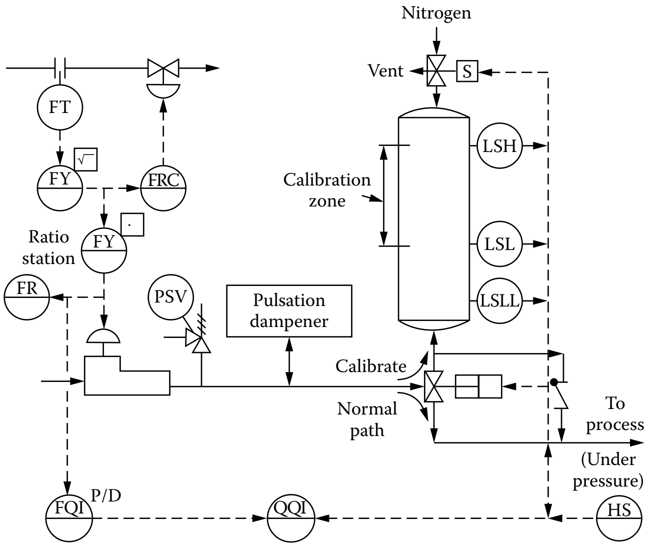


FIG. 8.34dd
Ratio and calibration controls for reciprocating pump. When level has reached LHS, the three-way valve returns to the "normal" path and nitrogen enters the tank to initiate discharge. When level drops to LSLL, discharge is terminated by venting off the nitrogen. Counter QQI is running while rising level is between LSL and LSH. Total count, when compared with known calibration volume, gives total error. Hand switch HS initiates calibration cycle by diverting the three-way valve to the "calibrate" path.

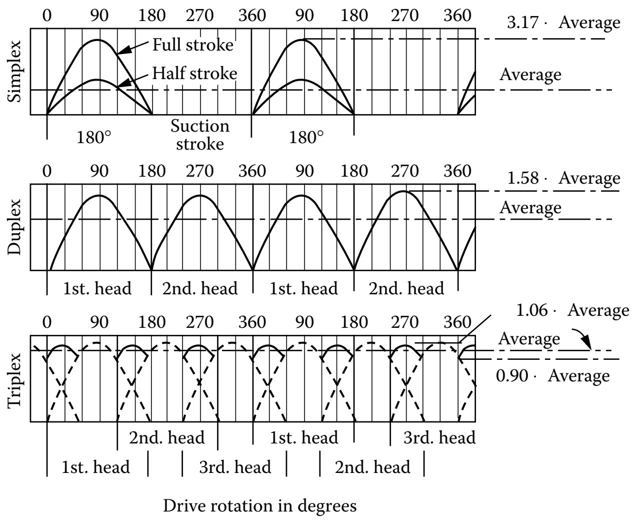


FIG. 8.34ee
Multiple pistons tend to dampen pressure fluctuations.

shows an automatic calibration facility that can be used for more advanced applications, or to calibrate the pump by hand periodically.

If the calibration is done on the suction side of the pump, this usually eliminates errors due to fluid compressibility, because close to the pump intake, the pressure is approximately constant.

Pulsating Flow As illustrated in Figure 8.34ee, the reciprocating pump produces a strongly pulsating flow.

One method of dampening pulsation is to use multiple cylinders, which, similarly to the way a car engine smoothes its pulsating thrust, smoothes out with multiple cylinders. The other option is to use pulsation dampeners, which are illustrated in Figure 7.4u in Chapter 7, Section 7.4.

Dampeners are usually connected into the piping near the pump discharge, and this is effective. However, performance can be improved by minimizing pipe friction losses between the discharge and the dampener, and introducing some friction in the line leading to the process user by a partially throttled valve to provide an acceptably smooth flow.

Valves Reciprocating PD pumps have check valves on their intake and discharge, and this adds several requirements that should be considered in the overall design.

It is necessary to ensure that the discharge pressure is always higher than the suction pressure. This can be guaranteed by placing a back-pressure regulator on the pump discharge, or possibly through the piping arrangement, as shown in Figures 7.4r and 7.4s in Chapter 7, Section 7.4. This is needed because otherwise the valves may open and allow flow through uncontrolled.

It is also recommended to orient the pump vertically, or near-vertically. The check valves are usually seated, at least partially by gravity, so their orientation is important. It is also desirable to strain or filter the pumped liquid to remove particles that may jam the valves, and ensure that the piping is scrupulously cleaned during and after installation.

Air-Locking and Cavitation When the purpose of a positive displacement pump is to meter the flow rate, certain precautions are needed. These include the removal of all entrained or dissolved gases, which otherwise can destroy metering accuracy. Figure 7.4t in Chapter 7, Section 7.4, shows how entrained gases can be returned to the supply tank.

PD pumps airlock comparatively easily: If a large bubble of gas accumulates between the intake and discharge valves, it can simply expand and contract through the pump strokes, and thereby stop the pumping action completely. This is a particular problem if pumping liquids that give off gas, such as sodium hypochlorite, which gives off chlorine; hydrogen peroxide, which gives off oxygen; or biological sludges, which give off methane, hydrogen sulphide, and other gases.

Therefore, one should ensure that the piping inherently intercepts and safely removes gases before they can enter the pump. If the removed gas is poisonous (chlorine and hydrogen sulfide), it must be piped to a safe location.

In the case of hydraulic diaphragm pumps, the gas bubble on either side of the diaphragm has the same effect. If the pump acts as though it is airlocked, but there is definitely no gas between the valves, there may be some on the hydraulics side of the diaphragm.

PD pumps are unaffected by cavitation, because no sudden collapse of the bubbles formed by cavitation is allowed. The piston strokes under complete control, the cavity (the bubble of vapor that is formed during cavitation) only collapses at the rate allowed by the piston. As a result, the destructive velocities that can be achieved by cavitation in a centrifugal pump do not occur in a PD pump.

Chemical Metering Pump Operation A major operational consideration that applies specifically to chemical metering pumps is the distance between the pump and the point where the pumped fluid enters the process. This results in dead time. Assume, for example, that this distance is 100 m and that the liquid's speed in the pipe is 0.33 m/s.

If the pump is transporting a constant dilution water flow, it will take 300 sec from the time when the dilution is changed to the time the new dilution reaches the process. Consequently, dead time limits the best achievable control, and it is always desirable to minimize it.

One method to minimize this dead time naturally is to reduce the distance between the point of control and the point of use. This may require a second pipe to carry the dilution liquid. If using this arrangement, consider running the chemical pipe inside the dilution pipe, to provide inherent secondary containment.

Another solution is to use feedforward based on a measurement that is in advance of the actual process requirement, by at least as much time as is the dead time.

Pacing the dilution or reagent flow to the chemical flow usually requires a separate metering pump, or possibly a flow control valve. If diluting or charging to a number of process users in a specific ratio, one can use fixed valves, such as needle valves, to ensure the correct distribution ratio to individual dilution points.

CONTROL OF PUMPS

Capacity control of pumps must recognize the incompressibility of liquids. For this reason, changes in the volumetric flow rate throughout the system occur simultaneously, and density is constant at constant temperature, regardless of pressure.

Pump capacity may be affected by (1) a control valve in the discharge of a pump, (2) one-off switching, (3) variation in the speed of the pump, or (4) stroke adjustment of PD pumps. Flow control by on/off switching provides only zero or full flow, whereas the other control methods provide adjustable flows in the system. The applicability of these four methods of capacity control is a function of the pump type, such as centrifugal, rotary, or reciprocating. The possible types of capacity controls for the various pumps are summarized in Table 8.34ff.

On/Off Control

On/off switching is the most common capacity control in use. It has many disadvantages such as flow surges that often hinder processing, high friction losses, and high electricity peak demand charges. However, on/off control is simple and can be economical, as its consequences do not require redesign to accommodate the limitations of on/off control.

Pumps that are controlled only by starting and stopping are said to be constant-speed (CS) pumps.

When pumping suspended solids or slurries, when the pump is stopped, a specific disadvantage of on/off control is that the solids may settle out of the liquid and may not go back into suspension when the pump starts again. This can cause plugging.

This can also happen if pumping entrained oil and grease: The grease will float and may stick to the top of the pipe. In these cases, an option is to keep the pump running: deliver flow through a circulating loop and control capacity with a pressure-controlled bypass back to the feed tank. For example, one can provide intermittent flow to feed a centrifuge by opening an on/off valve by a cycle timer. Such a loop is shown in Figure 8.7j. The pressure-controlled bypass shown in this figure allows the normal pump flow to be maintained in the loop, while the centrifuge feed valve is closed.

TABLE 8.34ff
Pump Control Methods

	Possible Types of Controls	
Method of Control	*On/Off*	*Throttling*
On/off switch	Centrifugal, rotary, or reciprocating	
Throttling control valve	Centrifugal or rotary	
Speed control	Centrifugal, rotary, or reciprocating	
Stroke adjustment	Reciprocating	

CS (on/off) pump operation is usually straightforward, except that the pump motor may overheat if it is started and stopped too frequently. Motors are usually rated for a maximum number of starts per hour (sph), because bringing the motor up to speed involves higher currents than keeping it at speed; perhaps ten times higher. Motor heating is proportional to current squared (I^2R), so the heating during start-up may be 100 times higher than normal.

Submersible pumps are usually rated for up to 15 sph, whereas large dry-pit pumps may be limited to 2–4 sph. Control systems must be designed to accommodate starts per hour limitations.

On/Off Level Control Figure 8.34gg illustrates the use of level switches for on/off pump control. The interlocks keep the tank level between the settings of LSH and LSL. In this illustration, the two-probe conductivity level switch operates a relay. When conductive liquid reaches the upper LSH probe, the relay closes contacts *H* and *I*. At this point the pump starts, and although the level then drops below the LSH probe, the pump keeps running because the holding contact (*H*) maintains the circuit.

When the level drops below the LSL probe, both the load (*I*) and the holding (*H*) contacts open, stopping the pump. When the level rises again, no action occurs when the LSL is contacted, because the holding contact is still open. However, when the level reaches the LSH, electrical contact is established, and the relay closes to repeat the pumping cycle.

The bottom portion of Figure 8.34gg shows a simple pump starter circuit that is controlled by the three-position hand-off-automatic switch. When the controls are in automatic (contacts 3 and 4 connected), the status of the interlock contact *I* determines whether the circuit is energized and whether the pump is on. The purpose of the auxiliary motor contact *M* is to energize the running light *R* while the pump motor is on. The parallel hot lead to contact 6 allows the operator to check quickly to see if the light has burned out.

The amount of interlocking provided for pumping systems is usually greater than that shown in Figure 8.34gg. In addition to the overload (OL) contacts shown in Figure 8.34gg, controls often include safety overrides. These usually detect excessive pressure or vibration, low flow, leaks (submersible pumps), or motor winding temperature, and these overrides usually energize remote alarms.

Most pump controls include a reset button that must be pressed after a safety shutdown condition is cleared and before the pump can be restarted. If the same pumps can be

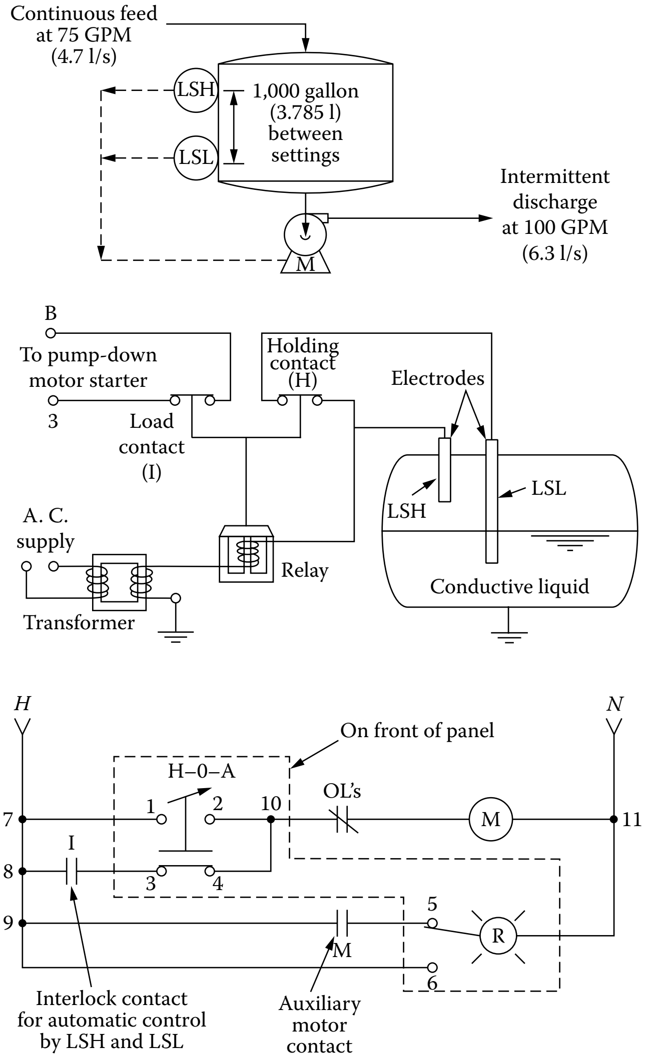


FIG. 8.34gg
On-off pump-down interlocks often utilize two-probe conductivity switches.

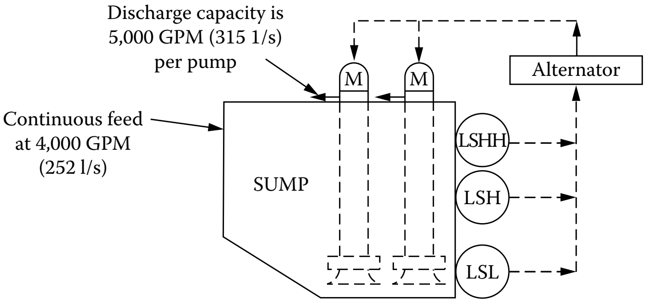


FIG. 8.34hh
On-off level control of dual pump station.

controlled from several locations, interlocks should be provided to resolve conflicting requests. One method is to provide an interlock to determine the location that is in control. This can be a simple local/remote switch or a more complicated system.

When many locations are involved, conflicts between requests are resolved by interlocks that establish priorities between control locations on the basis of selecting the lowest, highest, safest choice of pump operation. In such installations with multiple control centers, feedback should be provided, so the operators not only know the actual status of all pumps, but also are aware of any conflicting requests coming from the other operators or computers.

Multiple Speeds, Multiple Pumps When two-speed pumps are used, added interlocks are frequently provided. One interlock might guarantee that even if the operator starts the pump in high speed, it will operate for 0–30 sec in low before advancing automatically to high. This makes the transition from off to high speed more gradual. Another interlock might guarantee that when the pump is switched from high to low, it will be off the high speed for 0–30 sec before the low speed is engaged, to give time for the pump to slow down.

When several pumps are supplied from a common electrical feeder, it may be necessary to ensure that the feeder is not overloaded by the inrush current, when pumps are started simultaneously. If this feature is desired, a 0–30 sec time delay is usually provided between pump starts.

When two or more pumps are used, an alternator should be interposed between the level switches and the pump motors. The alternator places the pumps in service in an alternating sequence. Thus, if two pumps are used, each will have half as many starts per hour. However, while one pump is out of service, the other must pump all the flow so it will have the same sph as a single pump.

Alternation tends to equalize pump running hours and reduces starts per hour. In Figure 8.34hh, a cooling water return sump is illustrated. In this system, each pump is designed to handle the normal flow by itself, but both pumps operate together if abnormally high flows are required.

On/Off Flow Control Figures 8.34y and 8.34ii show a two-pump arrangement that responds to varying flow demands measured on the discharge side of the pumps. Pump I normally operates at point (1) (at 80 gpm and 36 ft, or 5 l/s and 10.8 m). When flow demand increases to 120 gpm (7.6 l/s), the head drops to 22 ft at point (2), and FSH starts pump II. The combined characteristic gives 120 gpm at 40 ft (7.6 l/s at 12 m) at point (3). In this control scheme, a wide range of flows is possible without serious loss of discharge pressure.

On/Off Pressure Control As was illustrated in Figure 8.34z, a pressure switch may be used to start a spare pump in order to maintain pressure in a critical service when the operating pump fails. In this case, a low pressure switch would actuate the spare pump, which is piped in parallel with the first pump.

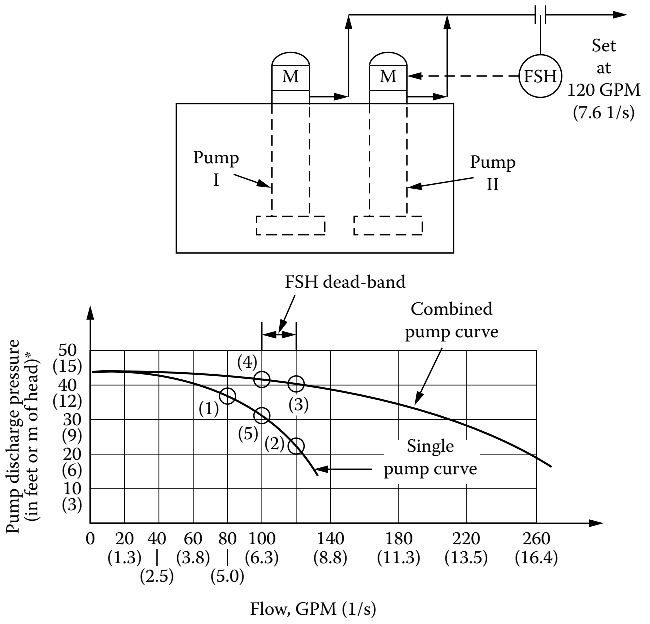


FIG. 8.34ii

On-off flow control can be used with parallel pumps.
**1.0 ft of water = 2.98kPa.*

A second possibility is to boost pressure, as shown in Figure 8.34jj. In this case, if pump I is normally operating at point (1), when the discharge pressure rises to 50 ft (15 m) at point (2), the flow is reduced from 53 to 20 gpm (3.3 to 1.3 l/s). At this point, the pressure switch (PSH) will start pump II and close the bypass valve. The system will now operate at point (3) on the combined characteristic curve, delivering 60 gpm at 50 ft (3.8 l/s at 15 m) pressure.

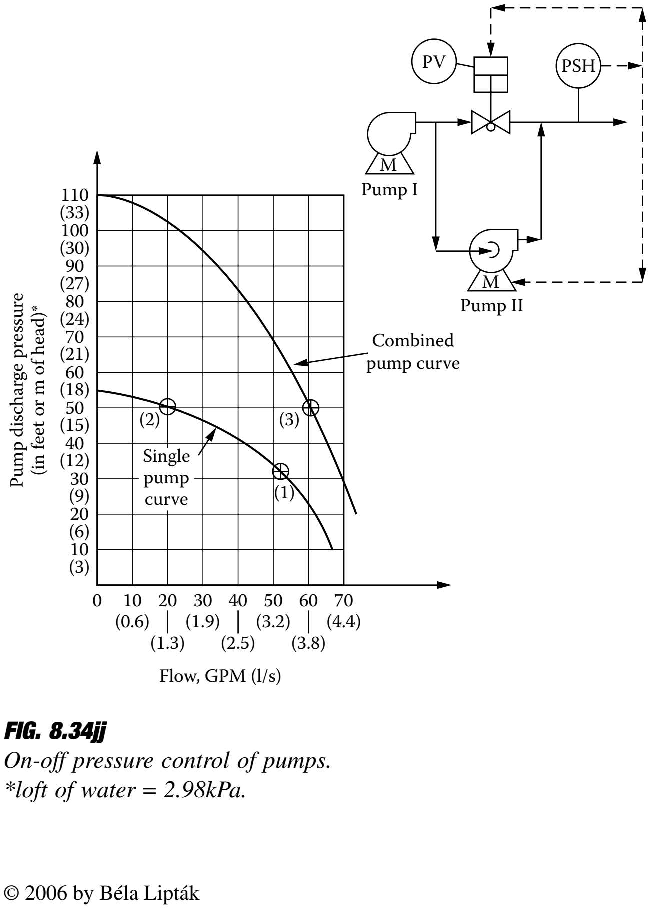


FIG. 8.34jj

On-off pressure control of pumps.
**loft of water = 2.98kPa.*

Using PLCs The previously described simple control strategies can be easily implemented in hardwired controls. With a programmable logic controller (PLC), it is easy to add many other features beyond what is practical in hardwire, such as to equalize the pump run times better, rather than just alternating when it is necessary to stop or start a pump. For example, it is simple to stop the running pump that has logged the most hours or start the pump that has the lower sph.

One can also program the PLC to flush the piping periodically by running both pumps together. If it has been a long time since both pumps ran together, and there is storage capacity available, one can initiate the running of both pumps by delaying the starting of one pump until more demand has accumulated and then run both pumps together. This strategy adds a pump start but is unlikely to result in an sph violation because it only happens if both pumps have not run together for some time.

If a pump has not run for several days because demand is low, and there is some liquid available to pump, it is recommended to run the pump regardless of demand, even if only for a few seconds. This strategy helps to keep the bearings lubricated and avoids developing flats on ball- or roller-bearings. It also adds starts, but again this only happens if demand is low, when starts are unlikely to be a consideration.

Starts per Hour Traditional SPH methods include the use of reduced-voltage or soft motor starters to increase the allowable number of SPH. Other design solutions include the providing of enough storage capacity in the system to accommodate the SPH limitation. One generally tries to avoid the technique of measuring the time between starts, and actively limit these starts, because this is an intrusive approach that is likely to interfere with process performance.

Reduced-voltage starters increase the allowable number of SPH, so they are a successful way of accommodating the sph limitation. However, when this means has been exhausted, the next step is to provide storage capacity. The sph limitation can be met by calculating the pumps' minimum cycle time (minimum start-to-start time), which is given by

$$\mathrm{MCT} = 4V/Q \qquad \mathbf{8.34(9)}$$

where

MCT is minimum cycle time in minutes
V is the available storage volume
Q is the design flow of the pump

Example: A 10,000 gpm pump is rated for up to four SPH. Find the working volume needed to ensure that the rating is not exceeded. Four SPH corresponds to an MCT of 15 min. Therefore,

$$V = \frac{15 \cdot 10{,}000}{4} = 37{,}500 \text{ gallons.}$$

Modulating Control

The flow from a pump or pumping station can be controlled by throttling either the forward flow or the bypass flow around the pump. Capacity control by valve throttling works essentially by throwing away what is not needed. The unnecessarily introduced pumping energy is wasted by either diverting the flow that is not needed through a bypass, or by restricting the pump discharge.

Pumps are usually powered by induction motors, which for a long time have been constant-speed devices. As a result, it was difficult or impractical to adjust pump speed, and alternative ways of controlling the pump's capacity were required. Modulating valves made CS pump capacity control practical, but they wasted energy. The development of reliable adjustable-speed (AS) induction motor drives has changed this situation, although modulating valves are still used.

Bypass Valves With a bypass valve, the pump delivers essentially a constant flow, and the bypass valve returns whatever the process does not need back to the pump inlet. The process determines by the opening of the flow control valve. For example, if the pump is supplying heating water to a reactor, when more heat is called for, the bypass valve is throttled to close, to increase heat to the reactor.

Bypass valves can work with both PD and centrifugal pumps, and they have other advantages, such as 1) When pumping slurries, greases, or mixtures that may separate, solidify, or coagulate, the bypass keeps the flow moving and so tends to keep the liquid homogeneous. 2) When pumping heating water, the water in the piping tends to lose heat. A bypass keeps the flow constant, which tends to ensure a consistent temperature to the process. 3) When pumping liquids with entrained gases, keeping the flow moving tends to avoid accumulation of larger gas bubbles that can airlock the system. 4) The heat gain in the pumped liquid stays constant. This is particularly important if pumping liquids near their vapor pressure.

Throttling Valves Throttling restricts the pump's discharge flow. When pumping incompressible (liquid) flow, throttle the discharge, but when gas (compressible) is transported, throttle the inlet. With a liquid, the problem with inlet throttling is that it causes cavitation, which cannot happen with a gas. With a gas, the advantage of inlet throttling is that the throttled gas expands into the blower at a reduced density (which reduces its power consumption). However, this cannot happen with an incompressible liquid — incompressible also implies inexpandable. So, the arrangement of liquid discharge throttling, and gas inlet throttling, makes the best of both possibilities.

Essentially, throttling changes the friction factor of the system curve. It works well with centrifugal pumps by shifting the pump's operating point, but does not with PD pumps because their characteristic curve is very steep (Figure 8.34bb), and therefore the discharge pressure just increases to force the positive displacement flow through the valve.

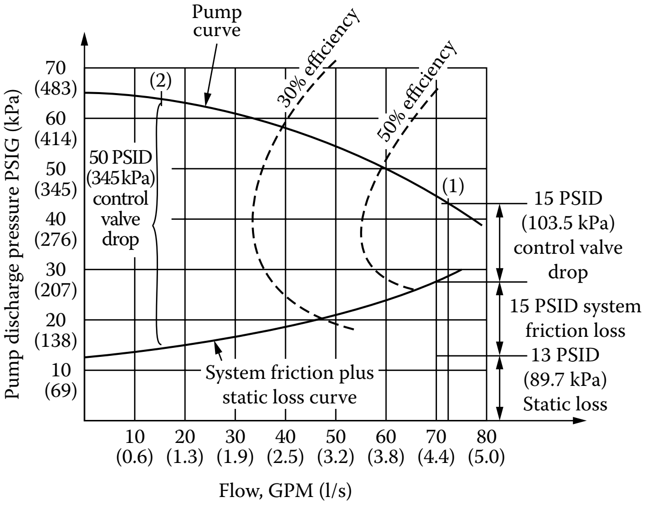


FIG. 8.34kk
Throttling control of centrifugal pump.

In Figure 8.34kk, design point (1) is near the maximum efficiency of the pump. Therefore, when throttling to point (2), the efficiency will drop.

For good controllability, the control valve should be sized to pass the design flow with a pressure drop equal to the system dynamic friction losses excluding the control valve but not less than 10 psi (70 kPa) minimum. (For more details on assigning sizing pressure drops to control valves, refer to Section 6.15 in Chapter 6.)

The control of flow by varying the pressure drop across the valve is illustrated in Figure 8.34kk. Here, when the flow is throttled from point (1) at 73 gpm (4.5 l/s) to point (2) at 15 gpm (0.95 l/s), the differential pressure across the control valve increases from 15 PSID (100 kPa) to 50 PSID (350 kPa).

However, one should be careful not to run the pump at flows low enough to overheat the liquid, or vaporize it, because vaporization can cause the pump to cavitate. One should use a heat balance on the pump to calculate the minimum flow needed through the pump to prevent vaporization. To be conservative, assume that all the motor's power is converted into heat. If this minimum flow was calculated to be 20 gpm (1.3 l/s), then size the PCV in the bypass to pass 20 gpm with a corresponding set pressure of 63 psi (435 kPa).

Figure 8.34ll shows a typical flow control loop with a pressure-controlled kick-back bypass. The rangeability of the control valve is assumed to be 25:1 (see Section 6.7). Thus, if the maximum flow required is 70 gpm (4.4 l/s) through the flow control valve, then the minimum controllable flow would be about 3 gpm (0.2 l/s). If control of lower flows is required, then install a second, smaller flow control valve in parallel with the first, and use it to control these lower flows.

Throttling Valves Waste Energy If there is no throttling valve in the system, a constant-speed pump will always operate at the intersection of its characteristic curve and the system

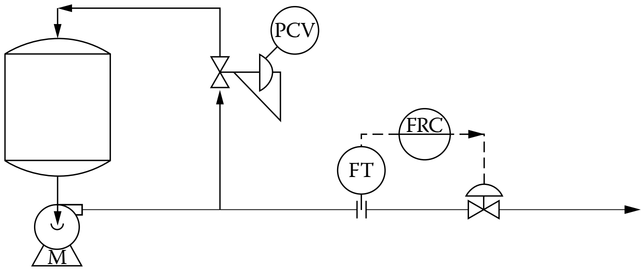


FIG. 8.34ll
Throttling control with pressure kickback.

curve. If the actual system curve has less slope than was designed for, the pump will deliver more flow than was intended. For example, in the case of the pump and system curves shown in Figure 8.34mm, the designers expected one system curve (solid line), but the actual system curve (dashed line) turned out to be flatter. Consequently, the actual operating point will be at point (2) instead of point (1), and therefore the actual flow will be higher, and the pressure lower, than the designers intended.

If a throttling valve on the pump discharge controls the flow, the pump will operate at point (1), and the valve will burn up the differential pressure between points (1) and (3). This drop represents wasted power:

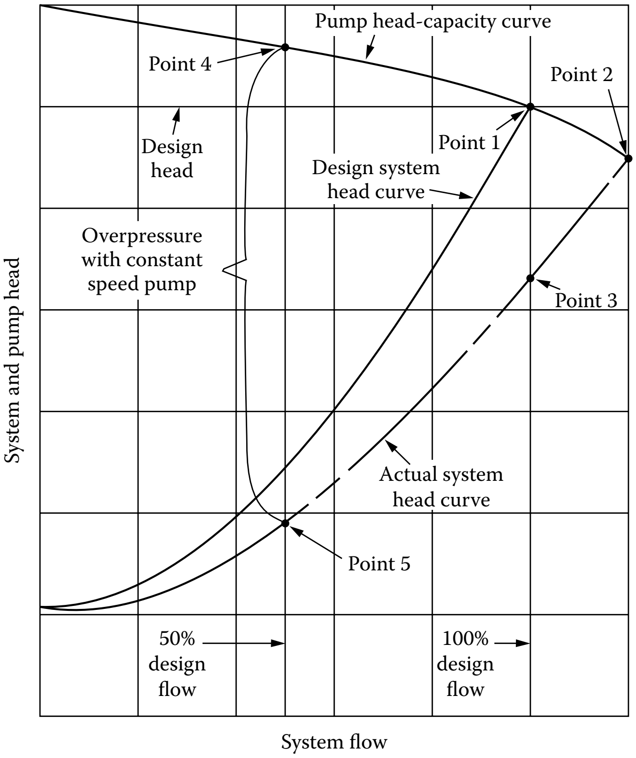


FIG. 8.34mm
Illustration of the consequence when the assumed and actual system curves are not the same.

$$Pw = p \cdot Q \qquad \textbf{8.34(10)}$$

where
Pw is the wasted power (watts, W)
p is the differential pressure across the valve (Pa)
Q is the flow through the valve (m^3/s)

If the valve throttles the flow to 50%, as in points (2) to (4), the energy wasted in the form of valve pressure drop increases, because it becomes the difference between points (4) and (5). This illustrates that throttling always wastes energy. In addition, throttling introduces another source of energy wastage, because it almost always also reduces the pump's efficiency (Figure 8.34k).

In Figure 8.34j, the useful pumping pressure is identified as H_s, and the actual pressure of the throttled system is given as H_p. The $(H_p - H_s)$ difference identifies the energy wasted through throttling. However, this is only part of the total waste, because moving the operating point from (2) to (1) also reduces the pump efficiency from 81 to 71%. As a result, not only is power lost to throttling, but power is also lost to reduced pump efficiency.

One can use Equation 8.34(10) to quantify the loss, by taking the example in Figure 8.34kk. If working with metric units, when the flow is throttled from point (1) at 4.5 l/s to point (2) at 0.95 l/s, the pressure across the valve increases from 100 kPa to 350 kPa. At point (1), the flow is 4.5 l/s (4.5×10^{-3} m^3/s) and required pressure is 190 kPa (4.5×10^3 Pa), giving a required power of

$$Pw = 4.5 \times 10^{-3} \cdot 190 \times 10^3 = 850 \text{ w}$$

However, the consumed power, including the valve's pressure drop, is

$$Pw = 4.5 \times 10^{-3} \cdot 290 \times 10^3 = 1300 \text{ w}$$

Therefore, the control's efficiency at point (1) is $\frac{850}{1300}$ = 65%

We can now compare that result with point (2), which has a flow of 0.95 l/s and required pressure of only 95 kPa, giving a required power of

$$Pw = 0.95 \times 10^{-3} \cdot 95 \times 10^3 = 90 \text{ w}$$

The consumed power, including the valve's pressure drop, is

$$Pw = 0.95 \times 10^{-3} \cdot 440 \times 10^3 = 420 \text{ w}$$

Therefore, the controls efficiency at point (2) is $\frac{90}{440}$ =20%

Figure 8.34kk does not provide the full pump efficiencies, but point (1)'s efficiency is likely to be about 60%, and point (2)'s about 15%. Including these values gives the overall efficiencies of 40% (65 × 60%) for point (1) and 3% (20× 15%) for point (2).

Capacity Control by Stroke Adjustment The flow generated by a PD pump is the product of the volume per stroke and the number of strokes per second. Therefore, one can adjust the PD pump flow by either adjusting the volume per stroke or the strokes per second (the speed). The volume of a piston is its area multiplied by the stroke length. Because the stroke length is readily adjustable, stroke control refers to the pump's flow adjustment method of modulating its stroke length.

The stroke can be adjusted by hand or automatically while the PD pump is operating at constant speed. Hand adjustments are currently more precise, usually rated at about 1% error, compared to 1.5% error when automatic adjustment is used. While manual setting is more accurate, automatic adjustment by closed-loop control will give better performance overall.

The range of flow control by stroke adjustment is 0–100%. However, in order to maintain accuracy, the practical range is 10–100% of maximum flow. The flow is related to stroke length through system calibration.

In chemical metering pump applications, it is often the case that the automatic controls modulate the pumps' speeds, while the operators adjust the strokes by hand. The reason for this technique is to use stroke adjustment to compensate for factors that are hard to measure and remain constant over long periods, such as the reagent concentration, that may change only from delivery to delivery. In such configurations, the feedback controls automatically adjust the speed to compensate for factors that change often and can be measured, such as the process flows or the reagent deterioration over time, possibly caused by temperature and so on.

Capacity Control by Speed Adjustment The cost of reliable and efficient adjustable-frequency drives (AFDs) has dropped rapidly in the past decade. As a result, adjustable-speed pumping is preferable today for both centrifugal and PD pumps. The main advantage of adjustable speed pumping is that it is efficient. It is efficient in two ways: 1) Rather than wasting energy (as do valves), speed control avoids introducing unnecessary energy in the first place, and 2) A modulating valve almost always moves the pump's operating point away from its best efficiency point (BEP). Speed control also moves the pump away from its BEP, but not nearly as much.

The basic concept is that a pump's speed controls its discharge flow: increase speed to deliver more flow, and reduce it to deliver less.

Multiple-Pump Controls

Distribution Controls From a control quality point of view, distribution controls have already been discussed in Section 2.23 in Chapter 2 and Section 8.28. Here, their description will be from the perspective of the performance of the pump station. The overall efficiency of the water distribution system, which was shown in Figure 8.34s, is described in Figure 8.34nn by curve 1.

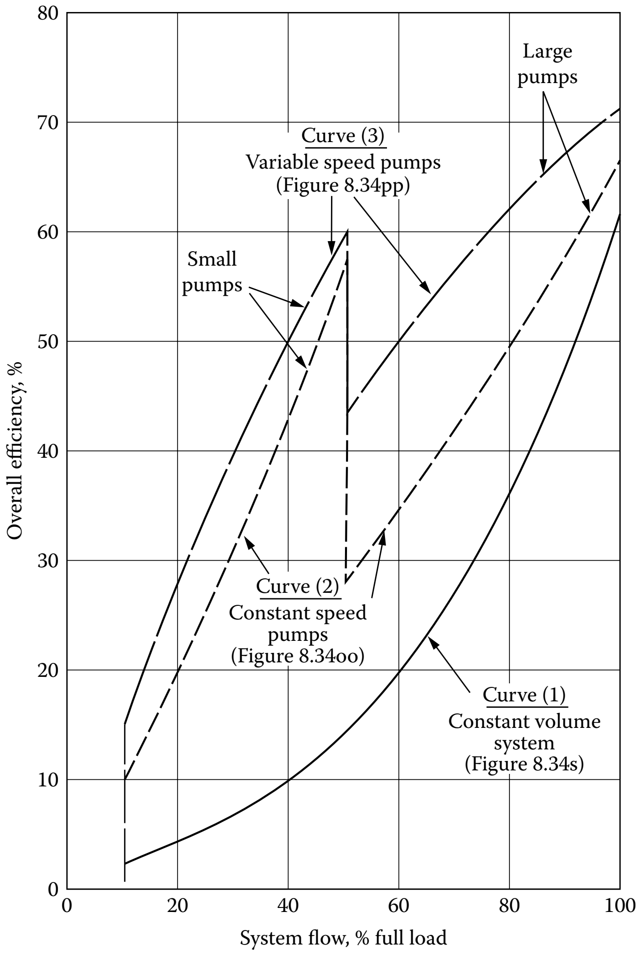


FIG. 8.34nn
Overall efficiency is maximum if two variable-speed pumps are used and is minimum with a constant volume installation.[4]

A substantial increase in efficiency (reduction in operating cost) can be obtained by replacing the three-way valves in Figure 8.34s with two-way ones and by replacing the single large pump with smaller ones. Figure 8.34oo shows such a system. Here, a small and a large primary pump are provided at the main supply point in the central plant, and a small and a large booster pump are furnished in each of the user buildings.

When the load is low, the small pumps are operating; when it is high, the large pumps take their place. The minimum flow requirements of the chiller are guaranteed by a bypass valve, and the chilled water makeup into the recirculating loop of each building is under temperature control (TC). The resulting improvement in overall efficiency is shown by curve 2 in Figure 8.34nn.

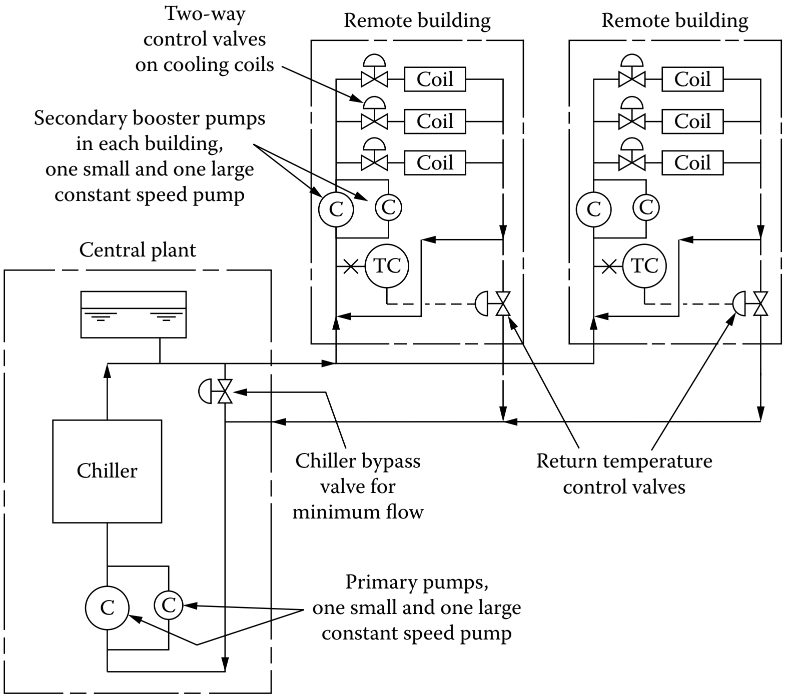


FIG. 8.34oo
Supply-demand matching can be achieved using constant-speed pumps of different sizes.[2]

The highest overall efficiency can be obtained through the use of variable-volume load-following. Figure 8.34pp illustrates such a system, utilizing variable-speed pumps in two sizes. In this system, all waste is eliminated except what is generated by the small minimum flow bypass around the chiller, which is guaranteed by a small constant-speed pump. The resulting increase in overall efficiency is illustrated by curve 3 in Figure 8.34nn.

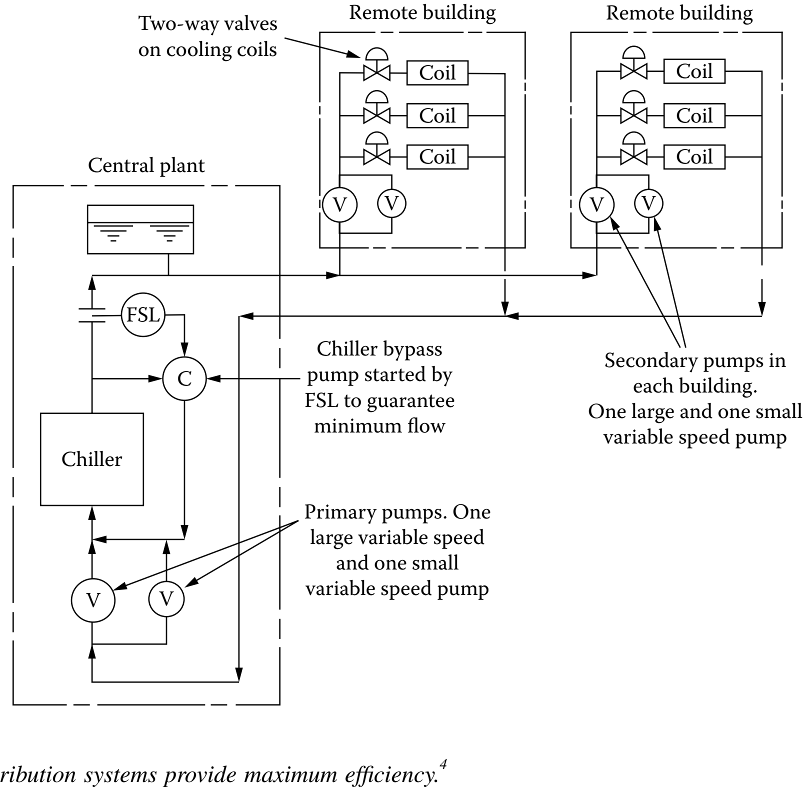


FIG. 8.34pp
Variable-volume water distribution systems provide maximum efficiency.[4]

Starting and Stopping Pumps When the running adjustable speed pumps are near their maximum speeds and when more flow is needed, the capacity of the pumping station must be increased by starting another pump. Similarly, when all pumps are at a low speed and less flow is needed, a pump must be stopped.

It is necessary to know the maximum pump speed a pump can run at and still deliver zero flow (called omega zero, ω_o), the pressure up to which the flow remains zero (shut-off pressure). This is the pressure that a starting pump must overcome before it can start delivering flow (superimposed back-pressure), and the flow delivered by an adjustable-speed centrifugal pump, which is running against the back-pressure, which is superimposed by the other operating pumps.

The subject of determining the above values for particular pumps and processes and to use these values in developing a detailed strategy for starting and stopping individual pumps is beyond the scope of this section, but if the reader needs such information, it can be found in Reference 15.

When to Start or Stop a Pump Accumulated running hours should be recorded on an elapsed time meter (ETM), because pumps are subject to wear while running, so recording their running hours can help distribute wear among them evenly, and it provides a guide to planned maintenance.

The time since the last stopping of the pump, called idle time, also should be recorded. During idle time, pump bearings lose lubrication, and they can be flattened out of round, so long idle periods are undesirable. Recording the idle time helps deal with this. A pump accumulates idle time while stopped, in the same way that it only accumulates running hours while running.

In critical applications consider always keeping an extra pump running at the same speed as the other pumps, so that if a pump fails it is only necessary to speed up the running pumps. One of the pumps should be stopped when the speed of the operating pumps has dropped to its low limit for over 30 sec.

The operator or the control system should initiate a start immediately when a running pump fails, unless a start is already in progress. A pump should also be started when the running AS pumps have been at their maximum speeds (ωm) for over 30 sec. Similarly, if an operating pump has traveled on its characteristic curve to the point where cavitation starts (Figure 8.34w), an additional pump should be started.

Start a pump if its idle time indicates that its bearings would benefit from being rotated. Preferably do this when the pumps are running around the middle of the speed range, so that a normal start or stop is unlikely to occur at the same time. Probably also stop another pump at the same time, to avoid a surge or having too much running capacity.

Selecting the Pump to Start or Stop When it is time to start a pump, start the AS pump that has accumulated the longest idle time. If all the AS pumps are running, start the CS pump that has the longest idle time. This strategy avoids accumulating idle time, consistent with not adding any extra starts.

When it is time to stop a pump, stop the running CS pump that has accumulated the most running hours. If all CS pumps are stopped, stop the AS pump that has accumulated the most running hours. This strategy tends to equalize the pumps' running hours, but it is not extremely successful because equalizing hours has the lowest priority—everything else takes precedence.

A highly successful strategy is to start the pump with fewest hours, stop the pump with the most hours. However, this can allow equipment to stand idle for long periods, particularly when other equipment is brought back into service after a long downtime.

Starting and Stopping When the pumps are of equal size and a pump is being started, accelerate it up to the omega zero speed (ω_0) quickly on its AFD's acceleration ramp. Then, gradually increase the flow through the starting pump, and reduce it through the running adjustable-speed pumps, until they are all running at the same speed and passing a similar flow. Similarly, when stopping, control the pump's deceleration down to ω_0, while accelerating up the running pumps. When the stopping pump reaches ω_0, stop it quickly.

When starting an AS pump, it is necessary to know how many AS pumps are running. Next, determine the starting pump's $\omega_{0.}$ If necessary, perhaps because as-commissioned pump measurements may not be available when the controls are configured, use the characteristic and system curves to estimate ω_0 against *n*. Start the pump and accelerate it up to ω_0 quickly.

When the starting pump reaches ω_0, start a timer, the *transition* timer, to bring the starting pump into action. Let the duration of the timer be *T* sec and the elapsed time since the start be *t* sec.

Depending on the system, the timer's duration may be faster or slower; it is usually not critical, provided it is slow enough that the pumps can follow the flow changes it requires, and fast enough to ensure it will have timed out before there is any need to start another pump.

While the transition timer is running ($0 < t < T$), calculate the starting pump's flow increase, and the running pumps' flow decrease, to pump the instantaneous flow required by the control loop output, *Q*, throughout the start transition. Then convert these flows to pump speeds, as follows:

When the transition timer has timed out ($t = T$), the starting pump has joined the running pumps.

Update the count of the number of running pumps, *n*.

When stopping an AS pump, the principles are the same as was for starting an AS pump, except in reverse: decelerate the stopping pump to ω_0 while accelerating the running pumps, so that total flow is controlled while it also transitions smoothly from the stopping pump to the running pumps.

CONCLUSIONS

This section covered the basics of pump control, while the next section will concentrate on the optimization of this unit operation. Pumping controls are a prime example of applications where it is essential to fully understand the personality of both the process and the pumping equipment used, before a successful control system can be designed.

Another unique characteristic of the pumping process is that over the life of the plant, the operating cost of a pump is much greater (sometimes a hundred times greater) than the first cost of the pumping equipment. It is for this reason that good process controls and optimization can have much higher returns when operating pumping stations than on other unit operations.

The goal of a well-designed pumping control system is good supply–demand matching, which will not only lower operating costs, but also reduce maintenance and cycling. The full automation of pumping stations — including automatic start-up and shutdown — not only will reduce operating costs but will also increase operating safety as human errors are eliminated.

References

1. American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE), "Centrifugal Pumps," in *ASHRAE Handbook, 2003 Equipment Volume,* Atlanta, GA: ASHRAE, 2003, Chapter 39.
2. Tchobanoglous, G., *Wastewater Engineering*, Metcalf and Eddy, Inc., New York: McGraw-Hill, 1981.
3. Karassik, I., *Centrifugal Pumps,* F. W. Dodge Corp., 1960.
4. Rishel, J. B., "Water System Head Analysis," *Plant Engineering,* October 13, 1977.
5. Conzett, J. C., "Adjustable-Speed Drives," Bulletin D-7100, Reliance Electric, July 1981.
6. Lipták, B. G., "Save Energy by Optimizing Your Boilers, Chillers, and Pumps," *InTech,* March 1981.
7. Langfeldt, M. K., "Economic Consideration of Variable Speed Drives," ASME Paper 80–PET–81, 1980.
8. Liu, T., "Controlling Pipeline Pumps for Energy Efficiency," *InTech,* June 1979.
9. Merritt, R., "What's Happening with Pumps," *Instruments and Control Systems,* September 1980.
10. Schroeder, E. G., "Choose Variable-Speed Drives for Pump and Fan Efficiency," *InTech,* September 1980.
11. Rishel, J. B., "Wire to Water Efficiency of Pumping Systems," Central Chilled Water Conference, 1975, Purdue University.
12. Baumeister, T. (Ed.), *Mark's Standard Handbook for Mechanical Engineers,* 8th edition, New York: McGraw-Hill, 1978, p. 3.69.
13. Shinskey, F. G., *Energy Conservation Control,* New York: Academic Press, 1978.
14. Systecon (Division of CEC), Bulletin No. 10-320-1.
15. Karassik, I., "What are the Characteristics of the Unknown Pump?" *Center for Professional Advancement*.
16. Bain, S., "Variable-Speed Pumping Requires Careful Study of Conditions," WEFTEC 65th Conference, Collection Systems Proceedings.

Bibliography

Avallone, E.A., and Baumeister, T., *Mark's Standard Handbook for Mechanical Engineers,* New York: McGraw-Hill, 1996

Baumann, H. D., "A Case for Butterfly Valves in Throttling Applications," *Instruments and Control Systems,* May 1979.

Baumann, H. D., "Control Valves vs. Speed Controlled Pump," Texas A&M Symposium, 1981.

Baumann, H. D., "How to Assign Pressure Drop Across Liquid Control Valves," *Proceedings at 29th Annual Symposium on Instrument Engineering for the Process Industry,* Texas A&M University, January, 1974.

Bose, B. K., *Power Electronics and AC Drives,* Englewood Cliffs, NJ: Prentice Hall, 1986.

Bower, J. R., "The Economics of Variable-Speed Pumping with Speed-Changing Devices," *Pump World*.

Cheremisinoff, N., *Pumps and Pumping Operations*, Engelwood Cliffs, NJ: Prentice Hall, 1992

Fischer, A. K., "Using Pumps for Flow Control," *Instruments and Control Systems,* March 1983.

Gibson, I. H., "Variable-Speed Pumps for Flow Control," ISA/93 Technical Conference, Chicago, IL, September 19–24, 1993.

Gottliebson, M., "Explore the Use of Variable-Speed Water Booster Pumps," *Water and Wastes Engineering,* May 1978.

Gray, C. B., *Electrical Machines and Drive Systems,* Longman Scientific and Technical, 1989.

Greene, *Pumping Machinery,* New York: Wiley.

Hall, J., "Motor Drives Keep Pace with Changing Technology," *Instrument and Control Systems,* September 1982.

Hall, J., "Pump Primer," *Instruments and Control Systems,* April 1977.

Holland, F. A., "Centrifugal Pumps," *Chemical Engineering,* July 4, 1966.

Janki, C., "What's New in Motors and Motor Controls?" *Instruments and Control Systems,* November 1979.

Karassik, I.J. et al., *Pump Handbook*, 3rd Edition, New York: Mc Graw-Hill, 2000.

Leonhard, W., *Control of Electrical Drives,* Springer-Verlag, 1985.

Lipták, B. G., "Costs and Benefits of Boiler, Chiller, and Pump Optimization," *Instrumentation in the Chemical and Petroleum Industries,* Vol. 16, Research Triangle Park, NC: Instrument Society of America, 1980.

Merritt, R., "Energy-Saving Devices for AC Motors," *Instruments and Control Systems,* March 1980.

Murphy, J. M. D., and Turnbull, F. G., *Power Electronic Control of AC Motors,* Pergamon Press, 1988.

Papez, J. S., and Allis, L., "Consideration in the Application of Variable-Frequency Drive for Pipelines," ASME Paper 80-PET-78, 1980.

Poyton, *Metering Pumps,* Marcel Dekker.

Rishel, J. B., "The Case for Variable-Speed Pumping Systems," *Plant Engineering,* November 1974.

Rishel, J. B., "Matching Pumps to System Requirements," *Plant Engineering,* October 1975.

Stewart, R. F., "Applying Adjustable-Speed AC Drives," *Instruments and Control Systems,* July 1981.

Thurlow, C., "Pumps and the Chemical Plant," *Chemical Engineering,* May 24, 1965.

8.35 Pump Optimization

S. BAIN AND B. G. LIPTÁK (2005)

INTRODUCTION

The previous section covered basic pump controls, while this section will concentrate on pumping system optimization. The savings resulting from pumping system optimization are greater than from most other unit operations, because pumping is a very energy-intensive operation. Over the lifetime of a pumping system, the operating cost is about a hundred times greater than the first cost of the pumping equipment. It is for this reason that optimization can have much higher returns here than on other unit operations.

Pump optimization includes the goal of introducing only the energy that is needed to transport the fluid, but no more. The elimination of energy waste and the providing of good supply–demand matching will not only lower operating costs, but will also reduce maintenance. The full optimization of pumping stations — including automatic start-up and shutdown — will not only reduce operating costs but will also eliminate human errors and increase operating safety.

Before starting on the optimization phase of designing a pumping station control system, the design engineer should make sure that power-factor correction, the use of high-efficiency motors, and operating against the lowest possible static and friction heads have been already implemented.

In this section, the subject of pump optimization will be first discussed without using a model. This discussion will focus on keeping throttling valves as open as possible, in order to minimize the waste of pumping energy.

In the later paragraphs of this section, a mathematical model of the pump's performance will be described. This model is stored in the digital memory of the control computer and can be used to predict and optimize future operation. By comparing the theoretical and the actual results, the model can also update itself automatically.

NOMENCLATURE

ω	Rotational speed. The instantaneous rotational speed the pump is running at. Typically in *rpm*.
ωd	The pump's design speed. Bearings, seals, and so on will probably not fail catastrophically immediately above this speed, but it is a guide.
ωo	The highest speed a pump can run at and still deliver zero flow.
Q	Flow; incompressible liquid flow.
P	Pressure; usually the pump's pressure gain.
Ff	Friction factor; used to calculate friction pressure loss in pipes.
N	The number of pumps that are running.
$c_1, c_2, \ldots$	Constants that describe a pump's characteristic curve.
CSH	Constant speed head.
Pw	Power absorbed by pump shaft or wire power. (The symbol for power is *P*, and for pressure is *p*, but the upper- and lowercase letters are similar, so to help distinguish them, this text uses *Pw* for power.)
$pw_1, pw_2 \ldots$	Constants that describe a pump's power curve.
$r_1, r_2 \ldots$	Constants that describe a pump's NPSHR curve.
$a, b, c, d \ldots$	General-purpose constants that are reused as needed without being redefined; their values are local to the expression being developed.
Ns	Specific speed

PUMP EQUATIONS

As the subject of pump and system curves has been covered in Section 8.34, here only a summary will be given, using the above-listed nomenclature.

Affinity 1, pump speed vs. flow:

$$Q_2 = Q_1 \cdot \frac{\omega_2}{\omega_1} \qquad \textbf{8.35(1)}$$

Affinity 2, pump speed vs. pressure:

$$p_2 = p_1 \cdot \left(\frac{\omega_2}{\omega_1}\right)^2 \qquad \textbf{8.35(2)}$$

Table 8.35a gives a summary of the affinity laws for both speed and impeller diameter:

The general form of the system curve is

$$p = CSH + Ff \cdot Q^x, \qquad \text{where} \qquad 1.7 < x \leq 2.0 \qquad \textbf{8.35(3)}$$

TABLE 8.35a
The Affinity Laws Describe the Relationships between Pump Impeller Speed and Diameter to the Resulting Fluid Flow, Pressure, and Power Consumption

Affinity Laws for Speed	*Affinity Laws for Diameter*
Flow ∝ Speed	Flow ∝ Diameter
Pressure ∝ Speed^2	Pressure ∝ Diameter^2
Power ∝ Speed^3	Power ∝ Diameter^3

Figure 8.35b graphically illustrates the mathematical expressions for the affinity laws and the system curve equation.

The pump's characteristics can be expressed as a continuous function (three-dimensional surface) of speed, flow, and pressure. Therefore,

$$p = a + b \cdot Q + c \cdot Q^2 + d \cdot Q^3 \qquad \mathbf{8.35(4)}$$

Characteristic pump surface equation:

$$p = c_1 \cdot \omega^2 + c_2 \cdot \omega \cdot Q + c_3 \cdot Q^2 + c_4 \cdot \frac{Q^3}{\omega} \qquad \mathbf{8.35(5)}$$

Similarly, the power consumption and the required net positive discharge head (NPSHR) of the pump can also be expressed in equation form.

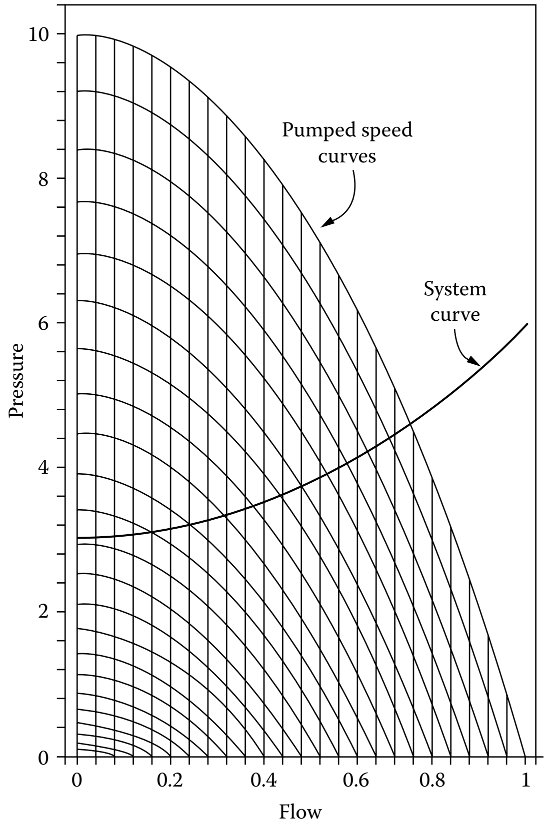


FIG. 8.35b
The system curve of the process is superimposed on a family of pump curves, corresponding to a variety of pump speeds.

The power consumption surface:

$$Pw = pw_1 \cdot \omega^3 + pw_2 \cdot \omega^2 \cdot Q + pw_3 \cdot \omega \cdot Q^2 + pw_4 \cdot Q^3 \cdots \qquad \mathbf{8.35(6)}$$

Equation describing the NPSHR surface:

$$\text{NPSHR} = r_1 \cdot \omega^2 + r_2 \cdot \omega \cdot Q + r_3 \cdot Q^2 + r_4 \cdot \frac{Q^3}{\omega} \cdots \qquad \mathbf{8.35(7)}$$

MODEL-FREE OPTIMIZATION

A pumping system is optimized when it meets the process demand for liquid transportation at minimum pumping cost and does that in a safe and stable manner. Once the equipment is installed, the equipment and piping is fixed and the potential for optimization is limited to selecting the best control system configuration. As an example, Figure 8.35c illustrates the optimization of a pump station consisting of a variable-speed and a constant-speed pump. (It was explained in connection with Figure 8.34o, in Section 8.34, that the use of only one variable-speed pump is the right selection if the static head of the system is between 50 and 67% of the total pump discharge head.)

In Figure 8.35c, PDIC-01 maintains a minimum of 10 psid (69 kPa) pressure difference between supply and return liquid pressures of each group of users. Therefore, if the control valve of any user opens up further, the flow will increase. On the other hand, if demand drops and, therefore, the pressure drop rises across all users, PDIC-01 will slow down the variable-speed pump to keep the differential from rising beyond 10 psid (69 kPa).

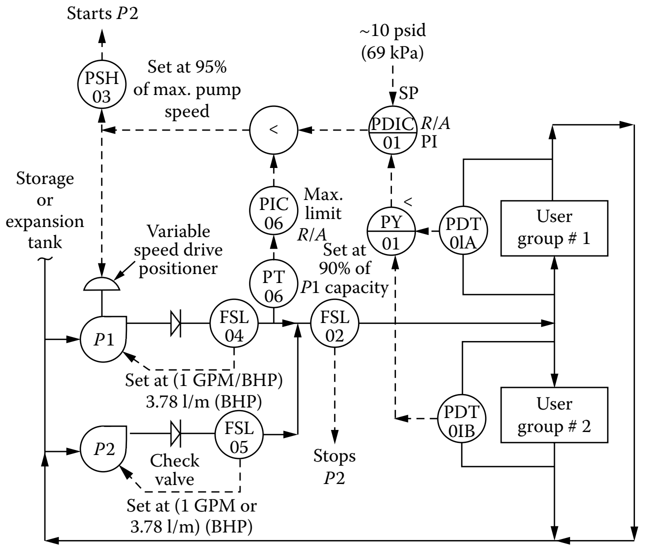


FIG. 8.35c
Optimization controls of a pump station consisting of a constant- and variable-speed pump.

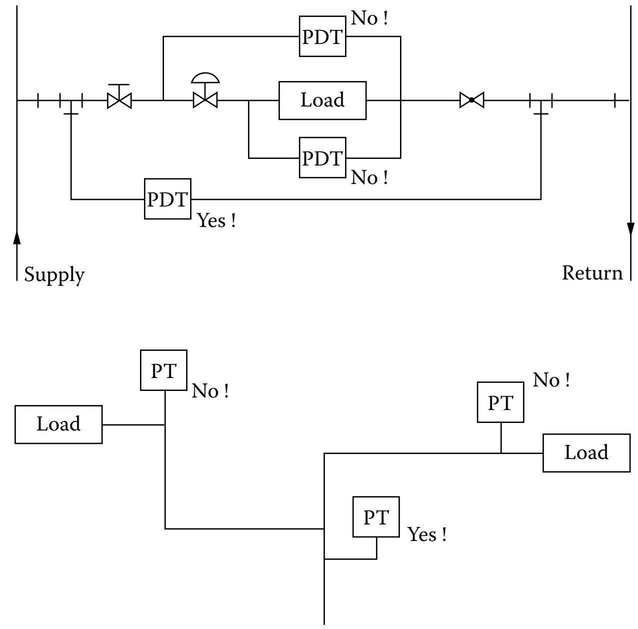


FIG. 8.35d
The pressure and differential pressure transmitters must be correctly located in the pipe distribution network.

When the variable-speed pump approaches its maximum speed, PSH-03 will automatically start the constant-speed pump (*P*2). When the load drops down to the set point of FSL-02, this second pump is stopped. One important recommendation to remember in connection with the operation of multiple-pump controls is that extra increments of pumps are started by pressure but stopped on flow.

FSL-04 and FSL-05 are safety devices that will protect the pumps from overheating or from going into low-flow cavitation. As long as the pump flow rate is greater than 1 gpm (3.78 l/m) for each pump break horsepower, the operation is usually safe from this perspective. If the flow rate drops below this limit, the pumps are stopped.

In Figure 8.35c, the pump speed is set to keep the pressure drop across the lowest user above some minimum limit. For this reason, it is important that the pressure drop across the users (PDT) be accurately detected. Figure 8.35d illustrates both the correct and the incorrect locations for the pressure-detection points in the system. The main consideration is that differential pressure transmitters should measure the head loss not only through the process load but also through the control valve, hand valves, and piping.

Similarly, as the lower part of Figure 8.35d shows, if the header supply pressure (instead of the pressure drop across the load) is to be controlled, the pressure transmitters should be located on main raisers or headers and should not be near major on/off loads.

Valve Position-Based Optimization

One of the best methods of finding the optimum pump discharge pressure is illustrated in Figure 8.35e. The optimum discharge pressure is selected as the one that will keep the most-open user valve at a 90% opening. As the pressure rises, all user valves close; as it drops, they will all open. Therefore, opening the most-open valve to 90% causes all others to be opened also. This keeps the pump discharge pressure and the use of pumping energy at a minimum.

As the valve position controller (VPC-02) lowers the set point of PC-01 and, thereby, opens the user valves, it not only minimizes the valve pressure drops but also reduces valve cycling and maintenance. This is because cycling is more likely to occur when the valve is nearly closed, and maintenance is higher when the pressure drop is high. The other important advantage of this control system is that no user can ever run out of coolant, because no user valve is ever allowed to reach 100% opening. This increases plant safety.

In order to make sure that the pressure controller (PC-01) set point is changed slowly and in a stable manner, the valve position controller (VPC-02) is provided with integral action only, and its integral time is set to be about 10 times that of PC-01. In order to prevent reset windup when the PC-01 is switched to manual or to local control from cascade, the valve position controller is also provided with an external feedback signal from PT-01.

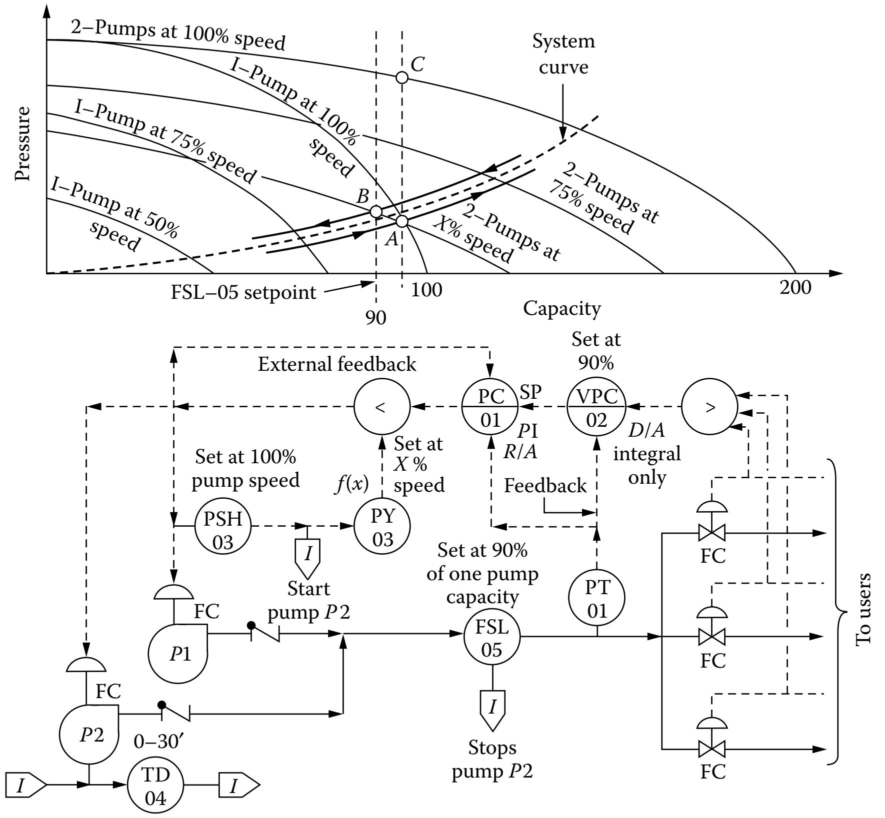
The pump station in Figure 8.35e consists of two variable-speed pumps. Their speed is set by PC-01. When only one pump is in operation and the PC-01 output approaches 100%, PSH-03 is actuated and the second pump is started, as shown by interlock #1 and the table at the bottom of Figure 8.35e. When both pumps are in operation and the flow drops to 90% of the capacity of a single pump, the second pump is stopped if this condition lasts longer than the setting of TD-04. The purpose of the time delay (TD-04) is to make sure that the pump is not started and stopped too often.

The top portion of Figure 8.35e shows the starting and stopping of the second pump on the pump and system curves. As the demand for water rises, the speed of the single operating pump increases until it reaches 100% at point *A*. Here, PSH-3 starts the second pump.

However, if the speed-control signal was unchanged when the second pump was started, an upset would occur, because the pressure and the corresponding pump speed of both pumps would instantaneously jump from point *A* to *C*. In order to eliminate this temporary surge in pressure, PY-03 is introduced. This is a signal generator that, upon actuation by interlock #1, drops its output to *x*. "*x*" corresponds to the required speed for the two-pump operation at point *A*.

Therefore, when interlock #1 is actuated, the low-signal selector immediately selects signal *x* for control, thereby avoiding the upset. After actuation, the output signal of PY-03 slowly rises to full scale. As soon as it rises above the output of PC-01, that signal is blocked by the low-signal selector and control is returned to PC-01.

Once both pumps are operating smoothly, the next control task is to stop the second pump when the load drops back to the point where it can be met by a single pump. This is



Interlock I

Condition			Action
Pump speed detected by PSH–03	Flow detected by FSL–05	Status of TD-04	Pump $P2$
At 100%	——	——	Start
Below 100%	Above 90%	——	Running
	Below 90%	Timing	Running
		Timed out	Stop

FIG. 8.35e

llustration of a control system that optimizes the energy consumption of a pumping station, consisting of two variable-speed pumps, by keeping the most-open user valve at near 90% opening.

controlled by the low-flow switch FSL-05, which is set at 90% of the flow capacity of one pump (point *B* on the systems curve).

The pump-cycling controls described here can be used for any number of pumps. For each additional pump, another PY and FSL must be added, but otherwise the system remains the same.

Optimization Alternatives

In fluid distribution systems in which the number of users served is large, Shinskey suggests a simpler system than the one described in Figure 8.35e. That control system is shown in Figure 8.35f. This control system configuration is based on the assumption that the terminal pressure will be kept constant if the pump discharge pressure is varied in proportion to the pressure drop across an orifice plate.

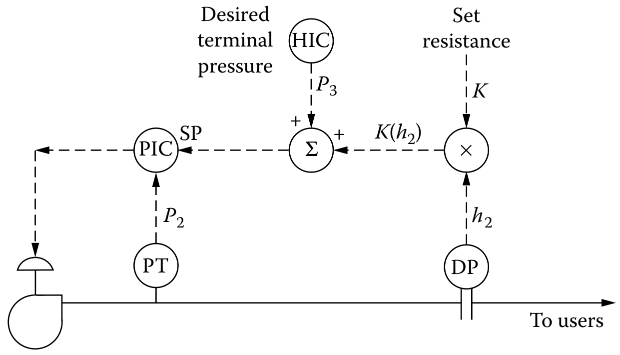


FIG. 8.35f

This control system keeps the terminal pressure of a mostly friction distribution system constant by varying it in proportion to the pressure drop across an orifice in the main header.

The pressure drop through an orifice varies in the same way as does the pressure drop in a mostly friction system. Therefore, if the orifice drop (h_2) is measured, this measurement can be related to the pipeline pressure losses by multiplying the drop by a set resistance ratio (K). Therefore, a set terminal pressure at the user can be maintained under variable load conditions by adjusting the pump speed, as shown in Figure 8.35f.

It is also advisable to continuously monitor the electric power consumption of all the pumps, so that empirical pump efficiency data will be always available and up to date. This will allow any load to be met with the most efficient pump or pump combination. In addition, the continuous monitoring of efficiency can also be used for maintenance scheduling purposes.

Calculating the Savings

The savings resulting from variable-speed pumping can be calculated at any operating point on the pump curve. Based on this information the relationship between the demand for flow and the power input required to meet that load can be plotted. Figure 8.35g shows these curves for both constant speed (control valve throttling)-type and variable-speed-type pumping systems. The difference between the two operating energy costs is the savings potential for optimization.

Once the saving curve shown in Figure 8.35g has been established, the next step is to determine the operating cycle. The operating cycle identifies the percentages of time when the load is 10%, 20%, etc., up to 100% (Figure 8.35h).

When both the savings and the operating curves are available, all that needs to be done is to incrementally calculate the savings as illustrated in Table 8.35i.

Knowing the total horsepower of the pumps and the cost of electricity makes it possible to convert the resulting percentages into yearly savings.

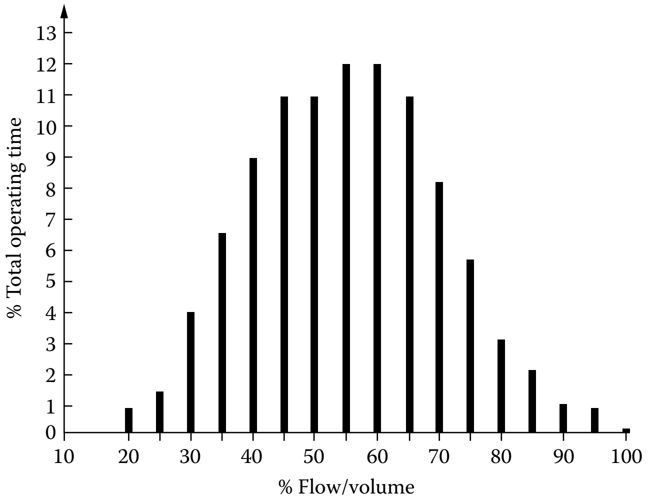


FIG. 8.35h
The pump operating cycle identifies the percentages of time when load is 10 percent, 20 percent, etc.

MODEL-BASED OPTIMIZATION

Equations 8.35(5) to 8.35(7) provide a model of a centrifugal pump that can run in a real-time controller. The basic concept is to add up the flows that are being pumped and then use the model of each of the pumps to calculate the power being consumed, if different pump combinations are used to meet that load. By the use of this model-predictive approach, the pump combination that consumes the least power can be identified and implemented.

Optimization of Pump Selection

A step-by-step optimization process involves the following steps:

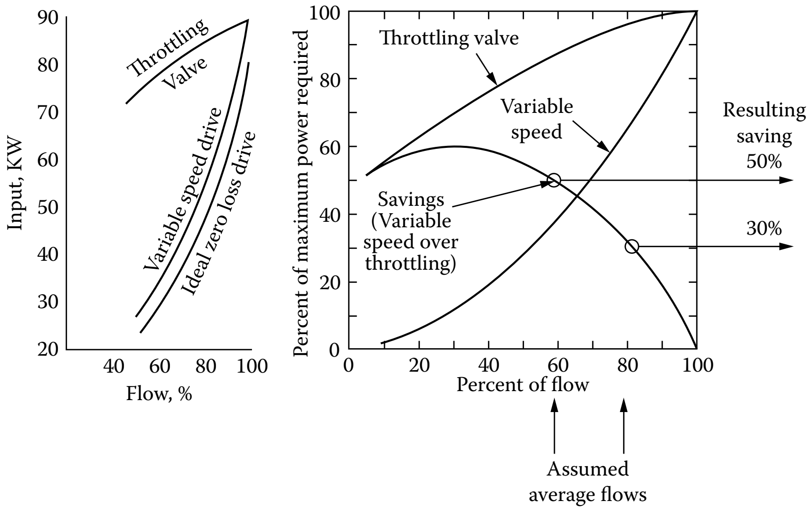


FIG. 8.35g
The savings generated by variable-speed pumping increase as the load drops off.

TABLE 8.35i
Illustration of Savings Calculation Based on Figures 8.35g and 8.35h.

% Flow	% of Time	% Savings	Total Savings (%)
Below 25	2.5	57	1.4
25–35	10	60	6.0
35–45	20	56	11.2
45–55	23	52	11.9
55–65	23	50	11.5
65–75	14	40	5.6
75–85	5.5	30	1.6
85–95	1.5	15	0.2
95–100	0.5	0	0.
	100%		49.4%

The first step is to measure the speed at which each pump is running and measure the pressure differential across each operating pump. Next, substitute the pressure and speed measurements into Equation 8.35(5) for each pump and calculate each pump flow. If individual flowmeters are available, it is advisable to cross-check these calculated flows against them. After this check, add the pump flows together to obtain the estimated total pump station flow.

The second step is to substitute the measured speed and the estimated flows into Equation 8.35(6) for each pump and calculate their power consumption. Sum up the powers to get the estimated pump station power consumption, and cross-check it against the actual power consumption, if installed power metering is available.

The third and last step is to assume that the required flow will be generated by a predetermined large variety of flow distributions among the pumps, and for each combination calculate the pump speed and the power consumption corresponding to the allocated flow. Compare the power requirements for each of the flow distributions and select the pump combination for operation that consumes least power.

The main advantage of having the pump models is speed (20+ combinations can be evaluated per second) and the ability to also automatically check for other criteria, such as NPSH availability. While much of this data can be condensed in look-up tables for manual use, such look-up tables can also be built into the controller, and it is not really necessary, because the digital controller can do the optimizing in real time and adapt to changing conditions.

The controller could also be programmed to find optimum pump selections. Such a heuristic solution would be more complicated than the controller described here, but it would configure itself to suit any set of pumps. See Section 2.18 in Chapter 2, Neural Networks, for more detail on this possibility.

Starting or Stopping Pumps

The strategy when starting or stopping pumps is to speed some up and slow others down smoothly during the *transition.* This smooth speed adjustment should always consider the speed limit represented by omega zero, which is the maximum pump speed that the pump can run at without delivering any flow, as was discussed in the previous section.

Therefore, when starting a pump, accelerate it up to ω_o quickly. Similarly, when stopping a pump, decelerate it from ω_o to zero quickly. This quick acceleration (and deceleration) does not cause flow surges because the pump does not deliver flow at speeds under ω_o.

In pump stations with pumps of different sizes, usually the optimum operation requires that one pump should stop at the same time as another starts. These pumps have different capacities, and after the transition the started pump will need to run at a different speed from the stopped one.

For a detailed description of the strategy to be used to optimize the selection, starting, and stopping of equal-sized and differing-sized pumps in a pumping station, the reader should study Reference 1.

CONCLUSIONS

Model-free pump station optimization, as described in the first part of this section, has been successfully used for decades. Model-based and neural network-based pump station optimization is relatively new and is expected to go through further development.

The adjustable-speed centrifugal pump optimization techniques described in this section will reduce the energy consumption of pumping stations by 12% or more, depending on the nature of the load served. They will also reduce pump wear commensurately.

Combined with surge-free pump starts and stops, and with their inherent predictive maintenance capability, these strategies can also improve the overall plant operation. In addition, they can eliminate the need for storage and can improve safety, if the particular storage volumes are hazardous in confined spaces.

Each model-based application can form a program shell that is reused, so that the development cost is amortized across many applications. Each implementation needs to be configured with specific pump curves and tuned for the specific application. However, this effort is insignificant compared to the initial cost of developing the program shell and the potential savings that can be obtained by these controls.

REFERENCES*

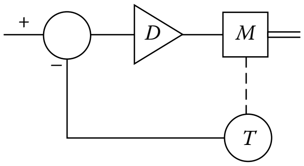
1. Bain, S., "Optimizing the Control of Multiple-Pump Variable-Speed Pumping Systems," *Proceedings of the WEF 66th Annual Conference, Collection Systems,* October 1993.
2. Karassik, I., *Pump Handbook,* Third Edition, New York: McGraw-Hill, 2000

---

* For bibliography, see Section 8.34

8.36 Rolling Mill Controls

K. J. JENTZEN (1985, 1995) **B. G. LIPTÁK** (2005)



Flow sheet symbol

Types of Drives:	Eddy-current clutches, DC static converters, AC variable frequency
Control Techniques:	Analog, digital, programmable
Partial List of Suppliers:	Allen-Bradley; Avtek Systems; Eaton; General Electric; Louis Allis; Parametrics Div. of Zero-max Industries; Reliance Electric; Wer Industrial
Suppliers of Services:	ALSTOM Drives & Controls, engineering design, automation, and control systems Brock Solutions, AC/DC Drive Systems Engineering, Industrial Control System (PLC, DCS, PC) BWG Machinery, mill modernization, turnkey installation, commissioning and personnel training Carlen Controls, feedback devices such as pulse tachometers, resolvers, strip speed sensors, and linear sensors ContRolling Technology International, level-2 process control for hot strip mills, rolling mill simulations, multistand and multipass mills Danieli Automation, USA, automation and process control systems for metals industry development, installation/start-up service Danieli Corus, automation and control, and consulting services for rolling mill and processing line, environmental Ferrex Engineering, rolling mill engineering, equipment supply Mill Equipment & Engineering, engineering, process, and turnkey automation systems for all types of mill control Morgan Construction, construction engineering services, mill audits, and field service programs for rolling mill plants Siemens Energy & Automation, drives and automation systems for hot and cold rolling mills, profile rolling mills, plantwide automation, computer modeling, monitoring and data logging, AGCs, flatness measuring SMS Demag rolling mill technology for the steel and nonferrous metals industries Tippins rolling mill equipment and automation systems, hot strip mill modernization, and aluminum mills VAI Automation, integrated control and information systems for rolling, complete hardware and software integration Zumbach Electronics, laser-based instruments for noncontact dimensional gauging, including profiles or special shapes

INTRODUCTION

Because of the high processing speeds involved and the increasingly stringent product quality specifications, modern automation systems can have a considerable impact on the profitability of a rolling mill operation. The high value of the strip produced means that short payback periods are often possible.

Multiple-drive systems for process lines are available using both analog and digital control techniques and are continually being advanced through the use of programmable digital hardware. These systems can use all available types of electrical drives and can be applied wherever speed is used for control of process variables.

Advanced mill setup models can generate actuator references for tandem and single-stand cold and temper mills rolling any combination of ferrous, aluminum, and brass alloys. The references calculated include interstand thickness and tensions, roll speeds, roll force, roll positions, work roll side shift, roll bending, and roll coolant flows. For continuous rolling mills, the model can calculate how to make a transition from the current coil conditions to those required for the new coil order.

Table 8.36a describes the features and characteristics of a typical rolling mill, which specification defines the equipment being controlled.

MULTIPLE DRIVES IN STRIP MANUFACTURING

The simplest application of two drives in a speed-controlled system is a conveyor line in which more than one set of rollers must handle a common strip of material. This strip can be a belt or a web of product, such as paper, plastic film, or foil. Roll and motor speeds must be maintained either at identical levels or at a constant ratio to allow for gearing differences.

Many process lines require the ability to vary and control speed relationships between machine sections. This ability is a key part of the process, regulating such things as stretch ratios of calendaring systems, proportions of various ingredients in a product, and printing synchronization. Regulation of some process variables is accomplished through coordination of drive speed as well as of torque.

TABLE 8.36a
Typical Rolling Mill Specification

Mill type..................................	1-2-3-4 reversing cold rolling mill
Material	Stainless steel, silicon steel, carton steel, nickel alloys
Strip width, maximum............	18 in. (457 mm)
Strip width minimum..................................	9 in. (230 mm)
Entry gauge maximum............	0.60 in. (1.5 mm)
Exit gauge minimum................	0.001 in. (0.025 mm)
Maximum coil size (winder).....	16 in. (406 mm) I.D./37 in. (940 mm) O.D.
Absolute coil weight.................	4,000 lb (1,820 kg)
Specific weight..........................	250 PIW (4.5 kg/mm)
Mill speed...................................	0/250/500 FPM (0/76/152 MPM)
Mill rolls: work roll nominal diameter	1.125 in. (28.6 mm)
1st Intermediate roll nominal diameter	2.062 in. (52.4 mm)
2nd Intermediate roll nominal diameter	3.600 in. (91.4 mm)
Mill bearings: diameter................	6.299 in. (160 mm)
Width...	3.543 in. (90 mm)
Mill RSF	17,700 lb/in. (316 kg/mm)
Mill power.......................................	150 hp at 0/650/1500 rpm
Winder power.................................	150 hp at 0/650/1500 rpm
Winder mandrel type.....................	Solid block
Winder gearbox ratio....................	12.9:1
Winder tension maximum................	9,000 lbs (4,050 kg) to 500 FPM (152 MPM)
Winder tension minimum................	140 lbs (64 kg)
Payoff mandrel diameter..................	16 in. (406 mm) nominal
Payoff/wind-on speed......................	400 FPM (122 MPM) @ 37 in.(940 mm) O.D. coil
Rewinder mandrel diameter..............	16 in. (406 mm) nominal
Payoff/wind-on tension.....................	3,000 lb (1,364 kg) max.@ 37 in.(940 mm) O.D. coil
Rewind payoff/rewinder speed.....	400 FPM(122 MPM) @ 37 in. (940 mm) O.D. coil
Rewind payoff/rewinder tension.......	1,100 lb (500 kg) max. @ 37 in.(940 mm) O.D. coil
Coolant system...................................	Mineral oil
Mill direction	Left to right

Multiple-drive systems are also used in advanced process control, in which speed is varied in response to changes in product properties, such as temperature or thickness. The refinement of process control computers has increased the potential applications of speed-controlled systems as effective control loops are developed for manufacturing lines.

Electrical Drive Systems

The techniques used for motor speed control in multiple systems include the eddy-current clutch, the DC static drive, and the AC variable-frequency drive.

Eddy-Current Drive

An eddy-current clutch drive applies variable voltage to a field coil, regulating the slip of an output rotor with respect to a constant-speed input rotor. The result is a smoothly operating unit with wide torque capability and simple, compact electronic controls. Its low efficiency, however, has limited its use to a few specialized applications.

DC Static Drives

The DC static converter is probably the most widely used speed control device. Because DC motor speed is proportional to voltage, and motor torque is proportional to current, the output of the system is easily regulated.

The available horsepower in these drives ranges from less than one to several thousand horsepower. Models currently on the market can give full output torque over the entire speed range with proper motor cooling. High efficiencies are obtained from the use of solid-state components, and speed regulation to 0.1% is available using conventional tachometers (for details on tachometers, refer to Section 7.19 in the first volume of this handbook).

The DC drive also lends itself to tension control because of its simple current-torque characteristic. It can also provide full negative, or braking, torque to a load when equipped with regenerative capability. One disadvantage of this type of drive is its poor power factor at low speeds. Installation of correction capacitors is sometimes required to avoid penalties from the utility. DC motors are also typically more expensive than corresponding induction motors (for details on tension and torque detectors, refer to Sections 7.21 and 7.25 in Chapter 7 in the first volume of this handbook).

AC Variable-Frequency Drives

AC variable-frequency drives are the beneficiary of many recent electronic advances. In these drives, the speed of the AC motor is proportional to the output frequency of the controller, with voltage varied in proportion. The development of power transistors and logic devices has enabled these drives to produce variable-frequency power compatible with virtually all AC motors up to 40 hp (30 kW).

In larger power ranges, silicon-controlled rectifier (SCR) choppers are used in place of transistors; efficiency deteriorates, however, because the power waveform is degraded. The AC transistorized drive has a high power factor (typically over 90%) and the ability to produce full motor torque over the entire speed range with proper motor cooling.

Variable-frequency drives are inherently self-regulating when used with synchronous motors. Speed regulation with induction motors depends on slip and is typically 20%. Closer regulation requires tachometer feedback.

Analog Multiple-Drive Systems

A conventional analog drive is speed-regulated through the comparison of a desired speed, represented by a reference voltage, with the actual speed, represented by a tachometer-generated voltage. The difference between the two speeds is the error used to raise or lower the actual drive power output to the proper level.

Speed coordination between two drives is easily implemented using the wide variety of amplifiers, multipliers, and meters developed for use with analog voltages. The techniques described here can be used with any of the basic drive types and are widely available.

Speed Control A typical speed control system is shown in Figure 8.36b. Here, the regulation of a metal foil rolling mill is illustrated, where one set of draw rolls are operating at a differential but constant speed ratio to the speed of another, prior set of rolls.

In the analog controls shown, the voltage used as a reference for the first, or lead, drive is fed into a potentiometer. The reference for the second drive is taken from the wiper of the pot, giving a reference that will vary with both the reference of the first drive and the wiper position. The ratio will be constant between the two drives for a given position. An amplifier is often used between the two drives to give the second drive the ability to go faster than the first as well as to prevent interaction between drives.

Additional amplifiers and control potentiometers enable the coordination of any number of drives. The system described in Figure 8.36b will maintain the desired speed ratios of the three sets of draw rolls as the process speed is changed. A variation often seen is a speed-follower system, in which the actual speed of a lead motor becomes the reference for the next drive or set of drives. This actual speed is generated by the lead motor tachometer.

Motor-operated reference potentiometers, acceleration control amplifiers, and upper and lower limit settings are usually incorporated into multiple-drive systems to give total control of reference voltages. The components used in these systems typically give accuracies of 5–0.1%, depending on their quality and complexity.

Dancer Tension and Speed Controls Winding and unwinding of a web of material, such as paper, is a common application of

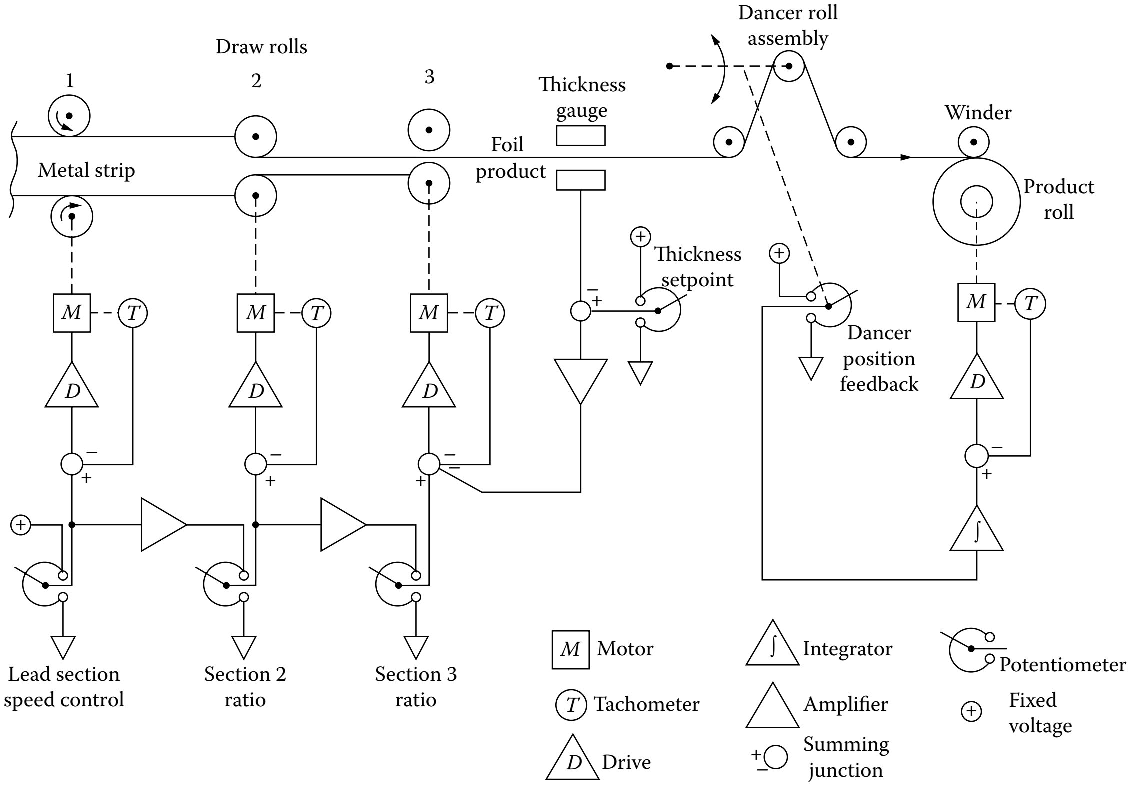


FIG. 8.36b
Analog speed-controlled process line metal foil rolling mill.

multiple drives. These systems often use movable rolls, called dancers, which accumulate a certain length of material and hold a preset tension on it. The dancer assembly is located between a constant-speed output roll and a roll of material being wound, as in Figure 8.36b.

As the diameter of the roll being wound increases, its rotational speed must decrease in order to maintain a constant surface speed. If the input speed to the assembly from the draw rolls has not changed, a gradual surface speed increase in the winder will pull the dancer down. Therefore, the dancer motion mechanically drives a feedback potentiometer, producing a voltage that is integrated and used to correct winder speed to the proper level.

Dancer systems can also be used in the same way between two machine rolls to maintain speed synchronization in a system with otherwise low accuracy. Slight changes in the speed of one section will raise or lower the dancer, which will produce a voltage correcting the speed error.

Thickness Control Analog systems can also be used for more complex process control. The foil thickness control loop shown in Figure 8.36b is an example of the control of two process variables in an interrelated system.

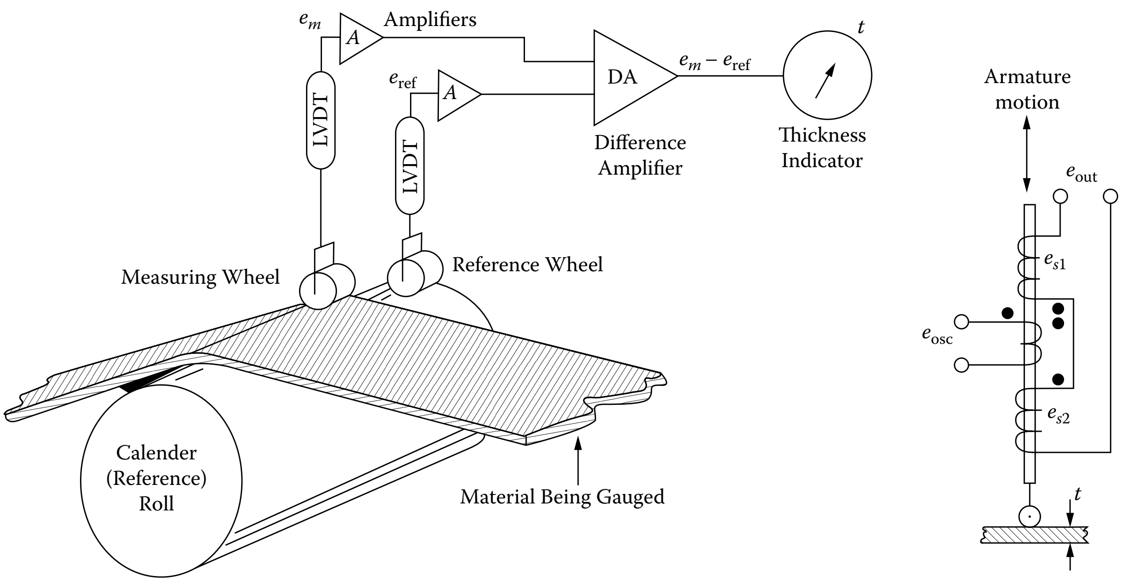
Figure 8.36c describes the operation of a linear variable differential transformer (LVDT)-type thickness gauge, while Table 8.36d provides a summary of the capabilities of a number of thickness detectors.

Digital Control Techniques

The availability of sophisticated and cost-effective digital equipment has greatly expanded the capability of process control systems. In many applications, digital signals are superior to analog signals, because of their precision and high noise immunity. Their processing by high-speed computers can also provide the basis of model-based advanced controls.

A digitally regulated drive and motor will typically be provided with a pulse generator as a feedback device. The pulse generator is driven by the motor and produces a frequency proportional to the motor speed. The speed reference for such a drive will be either a precision frequency or a numerical value. Systems using a frequency reference often include a phase-locked loop for comparison of the reference and actual speeds. These systems can also use up–down counters to detect any build-up in speed error.

Systems using a numerical reference compare the desired speed with the number of pulses generated by the feedback device in a given period; this allows numerical treatment of any speed error and great flexibility in processing the correction. Digital systems typically operate with zero average



$t = k(e_m - e_{ref})$
t = Thickness (mm)
k = Calibration constant (mm/Volt)
e_m = Output of measuring LVDT (Volt)
e_{ref} = Output of reference LVDT (Volt)

$e_{out} = e_{s2} - e_{s1}$
e_{osc} = Output signal of LVDT (Volt)
e_{s1} = Secondary #1 signal (Volt)
e_{s2} = Secondary #2 signal (Volt)

FIG. 8.36c
Left, differential roller gauge with LVDT; right, LVDT schematic.

speed error, meaning that no speed difference is allowed to accumulate.

Two or more motor drives are easily coordinated with digital references. Analog components (which are less accurate) are replaced by thumbwheel switches and microprocessor-based computation, as illustrated in Figure 8.36e. Numerical speed ratios can be set and used precisely, and synchronization is held exactly, because no pulses are lost. The history of the

TABLE 8.36d
Thickness Measurements

Instrument	*Type**	*Sensitivity***	*Application*****
Micrometer caliper	C	0.0001 in.	Sampling, quality control, calibration, foils, web, sheet, film materials (N)
Interferometry gauge	NC	1% of reading	Measures thin-film deposition layer thickness and uniformity of semiconductors and dielectric layers (Y)
Optical micrometer	NC	0.1 μm	Provides thickness monitoring for silicone wafers (Y)
Laser gauge	NC	0.05 in.	Mounted directly on production line and is used to alarm on any out-of-tolerance condition (Y)
Differential roller gauge	C	20 μ in.	Low-speed continuous foils, web, sheet, film materials, calibration, sampling (Y)
Sonic and ultrasonic gauges	C	0.01 in.	Rigid, relatively thick sheets, or pipe walls accessible from one side only (Y)
Capacitance gauge	NC	0.001 in.	Insulating sheets, films (Y)
Radiation gauge	NC	50 mg/cm^2***	Metal foils and plastic films (Y)

* C = contacting; NC = noncontacting.
** 1 in. = 25.4 mm; 1 μ in. = 2.54×10^{-5} mm.
*** Actual sensitivity depends on the specified value for a given instrument divided by its density (g/cm^3).
**** Whether a device is suitable for continuous process instrumentation is indicated by N (no) or Y (yes).

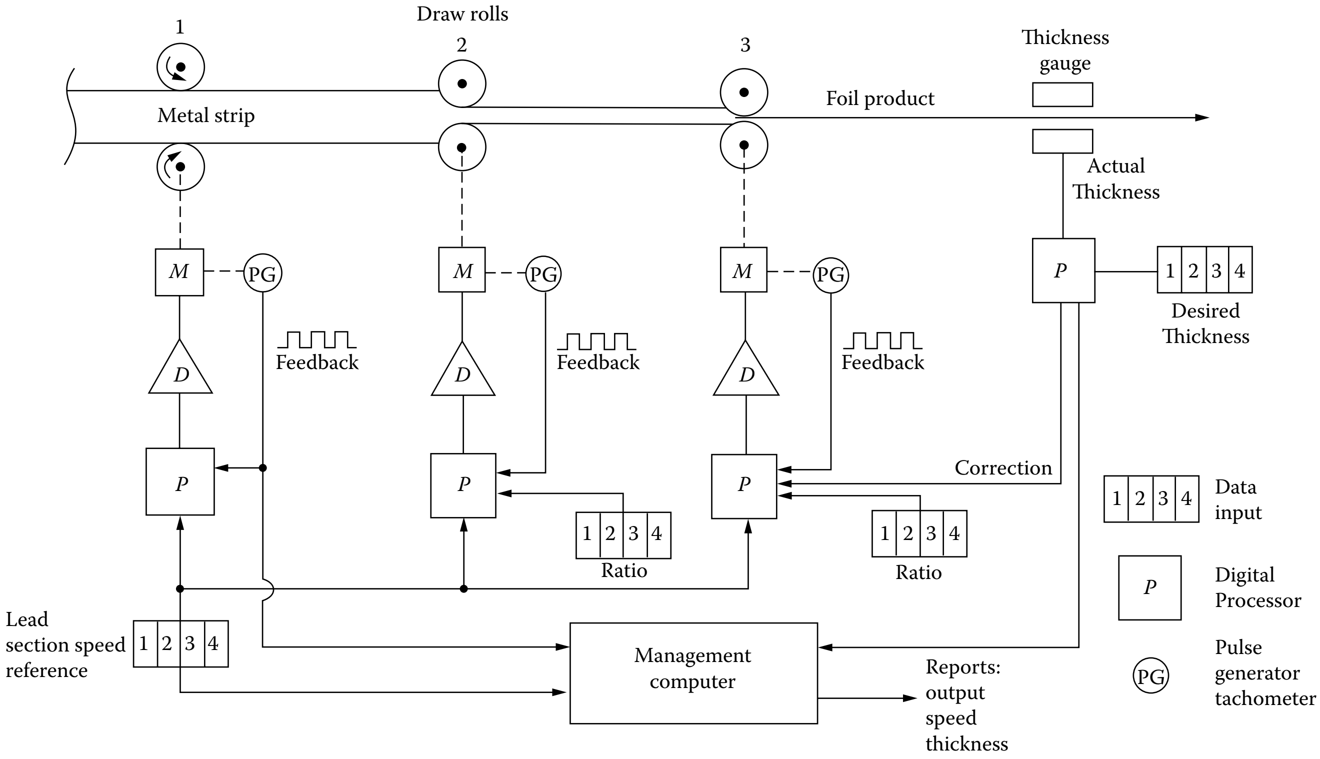


FIG. 8.36e
Digital rolling mill controls.

speed of the machine is always available for subsequent correction.

Thickness Controls The computation capability of a digital system allows complex functions to be easily implemented. An example is the control of the thickness of a strip of metal being drawn to a given gauge between sets of rollers, shown in Figure 8.36e. A thickness-measuring device outputs a number representing the actual gauge of the product, which is then compared with the desired gauge, and in case of an error, a correction signal is fed to the speed control of the draw rolls.

An increase in the speed of the last roll will reduce the thickness of the product foil, because the input rate has not changed; a decrease in the speed of the last roll will increase its thickness. Limits are set upon the maximum and minimum speeds and their rates of change. Tuning of the control system is simplified for the operator by the parameters being entered numerically.

Advanced Thickness Control In more advanced control systems, the noninteractive tension and thickness/extension and tension controls ensure the consistently high performance for any combination of mill type, product range, and operating practice.

A variation of feedback and feedforward control algorithms in combination with a variety of thickness gauges and mass flow thickness controls can optimize the performance of the available mill actuators.

Roll eccentricity compensation (REC) and different tension control strategies can be employed for threading and rolling in conventional tandem mills. The controls can also utilize speed-dependent actuator references generated by the setup calculation. The control systems usually include current overload protection, rounding of acceleration ramps and separate speed ramps for threading and tailout. A dynamic order change function is usually also available for continuous rolling mills.

Advanced Modeling An advanced mill setup model includes comprehensive sequences of calculations that involve the derivation of actuator references and schedule-dependent control system gains appropriate for the particular application.

One application might involve a thick product where the mill setup may aim to achieve the correct thickness on the coil head end as it is threaded with minimal out-of-tolerance lengths. In this case, the setup calculation can be made to achieve the target thickness on the coil head end and at thread speed with minimal yield losses.

In another case, it might not be possible to roll the head end to the target thickness until tension is established on each side of the stands. In such a case, two separate sets of mill actuator references are required for the threading operation. First, a short length of strip will be rolled that is thicker than desired due to the lack of front tension while the head end is produced, even with increased back tension.

Once the head end is coiled, the mill can achieve target thickness at thread speed with acceptable flatness. For this situation, the threading setup calculation must sacrifice a controlled length of out-of-tolerance material until tension is established and then the correct thread speed, roll gap, and roll bending references are applied to achieve the desired thickness reduction on each stand.

Operator Interfaces State-of-the-art digital control systems also improve the operator access to the system, because changes can be initiated directly through a keyboard or a video display using easily learned commands. Operator interfaces include mimic displays, performance monitoring, diagnostics, alarms, and screens to interact with the setup functions and dynamic controls.

Operator interfaces can be implemented on PCs using Windows NT and connected via an Ethernet TCP/IP network. The screens usually update rapidly, and the package is sufficiently flexible that the client's engineering staff can reconfigure screens where that becomes necessary.

Management Support Digital techniques in speed control also allow the monitoring of all other functions of a production process. As shown in Figure 8.36e, the information gathered about speed and thickness can be stored and used to generate reports. This can be used to inform both the equipment operators and management of the quantity and quality of the foil being produced.

The precise control of motor speed has thus led to accurate information that is vital to production management.

Bibliography

Andriev, N., "On-Line Optical Gaging Assures Perfect Fit of Jeep Doors," *Control Engineering,* November 1980.

AISE 2004 Annual Convention proceedings: www.SteelLibrary.com.

Bose, B. K., *Power Electronics and AC Drives,* Englewood Cliffs, NJ: Prentice Hall, 1986.

Butow, R., "Dimensional Inspection," *Measurements & Control,* February 1993.

"Dimensional Control in Rolling Mills.," London: Ashgate, 1991.

Doll, T. R., "Making the Proper Choice of Adjustable-Speed Drives," *Chemical Engineering,* August 9, 1982.

Ginsburg, V. B., *Flat Rolling Fundamentals*, New York: Marcel Dekker, 2000.

Gray, C. B., *Electrical Machines and Drive Systems,* Longman Scientific and Technical, 1989.

Hall, J., "Motor Drives Keep Pace with Changing Technology," *I&CS,* September 1982.

Julita, J. M., "New Dimensions in Optical Gaging," *InTech,* June 1980.

Kloeppel, J., "Liquid Film Thickness Measurement," University of Illinois, 2000.

Kongoli, F., Thomas, B. G., and Sawamiphakdi, K. (Editors), *Modeling, Control, and Optimization in Ferrous and Nonferrous Industry*, 2003.

Plog, H., *Guide to Methods and Instruments for Coating Thickness Measurement,* Draper, 2001.

Poppe, F. W., "Photoelectric Sensing," *Measurements & Control,* October 1991.

Stewart, R. J., "Applying Adjustable-Speed AC Drives," *I&CS,* September 1982.

Swientek, R. J., "On-Line Thickness Control Improves Product Quality," *Food Processing,* August 1986.

"Tachometers, Counters," *Measurements and Control*, April 1993.

Thickness Testing, Portcullis Press, 2000.

Ultrasonic Thickness Measurement, Witherby & Co., 1999.

Wohlstein, S., "Laser Gaging," *Measurements & Control,* June 1991.

Young, R. D., "Eight Techniques for the Optical Measurements of Surface Roughness," NBSIR 73-219, NBS, May 1973.

8.37 Separation Controls, Air

M. H. MACCONNELL (2005)

INTRODUCTION

This section will focus on the different processes and associated control strategies that are used in the air separation processes. The nominal composition of air is provided in Table 8.37a.[1] Air separation processes primarily produce oxygen, nitrogen, or argon in either the gas or the liquid phase. This section includes discussions of process controls that are associated with three air separation technology areas: membrane, adsorption, and cryogenic distillation in the manufacture of oxygen, nitrogen, and argon.

The typical product purities and production rates of the three highlighted technologies are provided in Table 8.37b. Production volume, product mix, and product purity determine the design of the individual air separation facilities. From a production volume point of view, oxygen is the most significant product.

Oxygen is the third-most-produced industrial chemical in the world today; the majority of the production plants use cryogenic air separation facilities.[2] The consumption of oxygen by market is illustrated in Table 8.37c.[3]

ADSORPTION TECHNOLOGY

The Process

The adsorption technologies used are either the vacuum swing adsorption (VSA) or the pressure swing adsorption (PSA) process. In either case, the principle of operation involves creating physical conditions in a vessel that permit an adsorbent material that is packed in the vessel to temporarily remove or adsorb the impurities of interest. Once the adsorbent bed is full, the physical conditions in the vessel

TABLE 8.37a
Nominal Composition of Dry Air

Component	*Percent by Volume*	*Component*	*Parts per Million by Volume (ppmV)*
Nitrogen	78.084	Carbon dioxide	350–400
Oxygen	20.946	Neon	18.2
Argon	0.934	Helium	5.2
		Krypton	1.1
		Xenon	0.09
		Methane	1–15
		Acetylene	0–0.5
		Other hydrocarbons	0–5

TABLE 8.37b
Production Attributes of Air Separation Technologies

Technology	*Purity Range*	*Production Range Tons per Day (TPD)*
Nitrogen membrane	5%–1000 ppm O_2 in N_2	8–10,000 SCFH
Nitrogen PSA	5%–500 ppm O_2 in N_2 200—5 ppm O_2 in N_2	4–116 MSCFH @ 95% N_2 0.4–12 MSCFH @ 5 ppm
2-Bed oxygenVSA	90–93% O_2	15–110 TPD
Single-bed oxygen VSA	90–93% O_2	4–5 TPD
Cryogenic distillation	99.9% O_2 < 0.5 ppm O_2 in N_2	60–4000 TPD

TABLE 8.37c
Consumption of Oxygen by Market

United States Market	*Users %*	*Western Europe Market*	*Users %*
Primary metals production	49%	Primary metals production	40%
Chemicals and gasification	25%	Chemicals and gasification	27%
Petroleum refineries	6%	Fabricated metal products	6%
Welding and cutting	6%	Health services	6%
Clay, glass, and concrete products	6%	Petroleum refineries	5%
Health services	4%	Pulp and paper	4%
Pulp and paper	2%	Water treatment	3%
Water treatment	1%	Other	4%
Other	1%		

are switched to a different state to permit the initiation of desorption of the impurities from the adsorbent bed.

Adsorption processes depend on the affinity of certain natural and synthetic materials to preferentially adsorb the nitrogen or the oxygen molecules. Zeolites, for example, are aluminosilicates that have nonuniform electric fields in their void spaces that cause preferential adsorption of polar molecules relative to nonpolar ones. For this reason, when air is passed through the zeolite, nitrogen is more strongly adsorbed than are oxygen or argon molecules.

As air passes through the bed, nitrogen is retained and an oxygen-rich stream is produced in the vessel. Carbon molecular sieves can also be used as the air separation media. In carbon molecular sieves, the pore sizes are approximately the same size as the air molecules. Because oxygen molecules are smaller than nitrogen molecules, the oxygen molecules diffuse more quickly into the pores than do the nitrogen molecules. This is a kinetic adsorption process. In short, carbon molecular sieves are selective for oxygen, and zeolites are selective for nitrogen.

Control of the Adsorption Process

Typically, programmable logic controllers (PLCs) are used to control the VSA or PSA units. Analog, on/off, and timer functions are all utilized to control product purity, to operate the compressor and regeneration valves, and ultimately to obtain maximum productivity and efficiency.

Locating the Adsorbate Front If the position of the adsorbate front and the associated mass-transfer zone along the axial length of the adsorber is known, it can be very useful, because it determines the composition and ultimate product purity as well as the required timing of the regeneration step.

Because of the temperature and pressure variations in this process, the commercially available oxygen sensors are not useable for direct on-line measurement. Therefore, the operating parameters are established mathematically from pilot studies that infer the composition on the basis of energy balances and axially distributed temperature and pressure measurements.[4] After the product leaves the unit, its quality is analyzed to ensure that product specifications are met. This analysis is also used to determine regeneration valve timing based on the pilot scale-up process and operating experience.

Vacuum Swing Adsorption

Vacuum swing adsorption is principally used in the generation of oxygen gas in the purity range of 90–93%. Manufacturers of VSA units provide adsorbent materials that adsorb nitrogen at atmospheric and slightly elevated pressures. This permits generation of an enriched oxygen stream that can be accumulated in a receiver. Depending on the amount of adsorbent, the bed will be allowed to continue adsorption of nitrogen for a fixed period of time until it is nearly saturated. After this, the beds are switched to vacuum, and the desorption of nitrogen occurs and the nitrogen is vented.

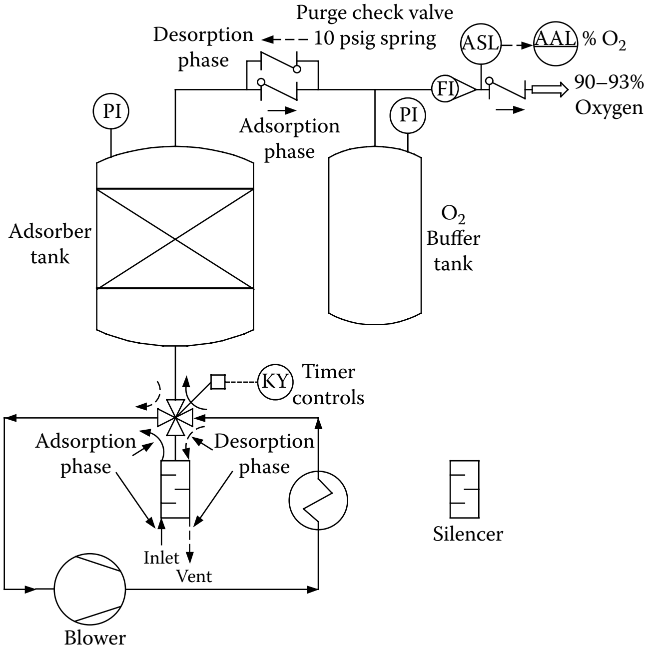


FIG. 8.37d
Components and controls of a small-scale vacuum swing adsorption unit. (Courtesy of Air Products and Chemicals, Inc.)

In a small-scale VSA oxygen generator, a timer (KY in Figure 8.37d) controls the blower that, during the desorption phase, generates the vacuum in the adsorber bed, causing the nitrogen, moisture, and other contaminants to be desorbed and vented. When the vacuum level in the adsorber bed is low enough, the adsorber vessel is purged with oxygen to sweep the remaining nitrogen from the vessel. This purge continues for a timed period, after which the automatic four-way valve returns the system into the adsorption phase, and the blower pushes fresh air into the adsorber bed.

When the pressure in the adsorber bed exceeds the pressure in the oxygen storage tank, oxygen will start flowing into the storage tank and will continue to do so until the feed timer times out and the cycle starts over again. This process is depicted in Figure 8.37d.

Paramagnetic oxygen analyzers (single or redundant) are typically used to continuously monitor the product quality and to alarm (AAL on Figure 8.37d) and initiate automatic interlocks to trip the equipment (which protects the customer by switching the oxygen supply to a backup), if purity falls below the 93–95% range. Larger-scale VSA units operate in a similar manner.

Pressure Swing Adsorption

The PSA technology utilizes co-adsorption, a phenomenon where both nitrogen and oxygen are adsorbed, but they are

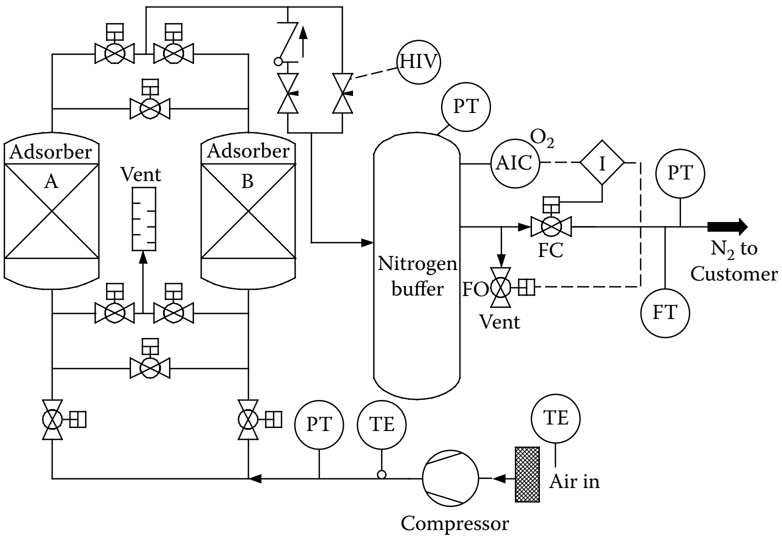


FIG. 8.37e

Process and instrumentation diagram of a pressure swing adsorption-type nitrogen generator. (Courtesy of Air Products and Chemicals, Inc.)

adsorbed at different rates. The process utilizes a two-tank (or tower) arrangement (see Figure 8.37e) in which two primary processing steps, adsorption and transfer, are performed. Unlike a desiccant-type adsorber that removes essentially all of the target molecules, the PSA carbon molecular sieve (CMS) takes advantage of the difference in adsorption rates (kinetic adsorption) to achieve the separation objectives.

The production rate of the PSA process can be limited by the adsorptive capacity of the adsorber, or when the bed is very cold, by the rate of desorption of oxygen from the carbon molecular sieve. The PSA units are often installed out of doors and need to be retuned for winter and summer operations, if ambient temperature conditions are substantially different.

A simplistic explanation of this temperature sensitivity is that hot molecules vibrate faster than cold molecules and, therefore, do not adsorb as readily into the carbon pores when hot. When the beds are cold, the opposite occurs, and therefore, it takes longer for the molecules to desorb from the carbon pores.

Design Considerations The carbon in the CMS is typically an extruded pellet approximately 2 mm in diameter and 1 cm in length. The carbon in the CMS bed is soft and can easily be pulverized into dust by agitation. Therefore, the bed must not be agitated.

If the carbon pellets turn to dust and the dust is vented, this is uneconomical because the CMS is relatively expensive and because the carbon dust shortens the life of the valve seats. Therefore, PSA vessel sizing and gas flow controls must guarantee that the gas velocity through the PSA unit is low, on the order of 1–5 ft/s. The product purity is a function of the gas velocity, the timing cycle of the tower, and the CMS selectivity grade. The sizing of the PSA vessel should also consider the required production rate and the CMS efficiency rating.

In a PSA-type nitrogen generator, it is also important to control the air-to-nitrogen ratio. If the nitrogen product has to be 0.5% pure, an air-to-nitrogen ratio of 3:1 is acceptable, whereas if high-purity (5–10 parts per million [ppm] O_2) nitrogen product is required, the ratio of air to nitrogen should be around 5:1.

The PSA vent dump volume largely impacts the air-to-nitrogen ratio. It is sometimes possible to increase efficiency by using a little more air or cycling the beds faster.

Process Measurements The product purity is generally measured with fuel cell-based electrochemical oxygen analyzers (for a detailed description of the different oxygen analyzers and their features, see Section 8.42 in Chapter 4 in the first volume of this handbook). For critical applications, it is often appropriate to install two oxygen analyzers in parallel to provide redundancy and improved reliability.

Given that the PSA units are subject to wide ambient temperature variation, it is important that the oxygen analyzer be provided with good temperature compensation. The PSA unit should always be provided with the capability for remote monitoring of the process conditions and of the product quality. Because many of the PSA units are located in remote locations, the remote monitoring of PSA installations is also used to dispatch maintenance crews from a central location as needed.

As shown in Figure 8.37e, the analog measurements typically include the measurement of product composition, product flow rate, compressor discharge (feed) pressure and temperature, N_2 supply pressure to the customer, product

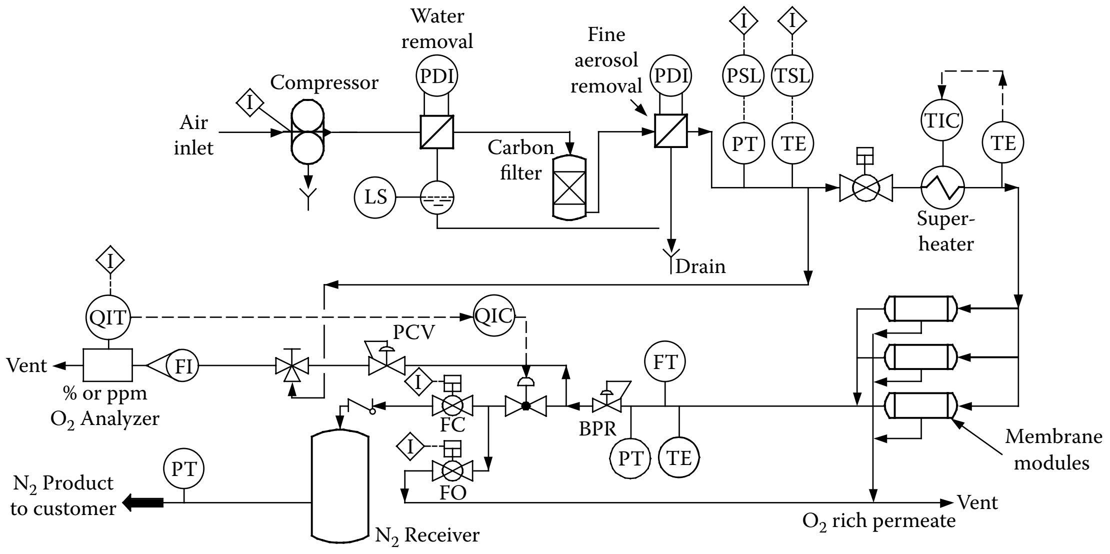


FIG. 8.37f

Process and instrumentation diagram of a pressure membrane-type nitrogen generator. (Courtesy of Air Products and Chemicals, Inc.)

buffer tank pressure, and ambient temperature. A number of on/off status measurements are also made to monitor valve positions and compressor status.

Programmable logic controllers are used to control the adsorption–desorption cycle of the plant, which essentially involves opening and closing the eight or so on/off valves on a time cycle sequence. These valves are either ball or butterfly valves, depending on the line size. The valve actuators are typically double-acting air cylinders, except for the fail closed (FC) nitrogen product valve and the fail open (FO) nitrogen product vent valve, which are spring-loaded to provide the required safe failure positions in the event of a product quality excursion or unit shutdown.

MEMBRANE AIR SEPARATION

Membrane systems are used for the continuous separation of components in mixed gas streams. The production volume and the achievable purity are generally lower than from adsorption or cryogenic processes. The most commonly used membrane systems are in the production of nitrogen, although membrane separation is also used in gas drying applications where water molecules permeate very quickly through the membrane material.

The predominance of nitrogen generators as membrane separators is because the smaller size of the oxygen molecules guarantees the superior permeability of oxygen through the membranes. As a result of this difference in permeability, membrane systems can only be used to produce oxygen in a concentration range of 25–50%, whereas 99.9% nitrogen purity is achievable.

Process Description

The membrane separation process starts with an air compressor (typically, an oil-flooded screw design) that compresses air to 100–200 psig (Figure 8.37f). The moisture and aerosols are removed from the compressed air by a carbon filter, and then it is superheated to prevent any condensation in the membrane modules. The superheated gas passes through a bank of parallel membranes, where oxygen permeates through to the low-pressure side of the membrane and is either vented or recycled.

One or two electrochemical gas oxygen analyzers (QIT in Figure 8.37f) continuously sample the high-pressure nitrogen product as it leaves the membrane modules. The product quality is controlled by the manipulation of the product flow control valve that is feeding the accumulation vessel. If the oxygen concentration in the product exceeds the allowable limits, the FC on/off valve on the receiver inlet trips closed and the FO on/off valve vents the product, until purity is reestablished. A back-pressure regulator (BPR) on the membrane system outlet prevents drawing more product than is available at the specified purity.

Flux and Selectivity

The membrane separation process is based on the difference in permeation or diffusion rates of nitrogen and oxygen molecules through a polymeric membrane that separates the high- and the low-pressure streams. Flux and selectivity are the two properties that bear directly on system design.

Flux determines the membrane surface area, and for a given gas, flux is a function of the partial pressure difference and inversely of the membrane thickness. In addition, each type of membrane material has a characteristic proportionality constant called permeability. Selectivity is the ratio of the permeability of the gases that are to be separated. In other words, permeability is the ratio of the speeds at which the respective gases diffuse through the membrane.

Membrane System Sizing

Sizing of the membrane system is based primarily on two parameters: the length of the membranes in each module and the number of modules installed in parallel. For a given amount of flow, the length of the membrane modules tends to determine the purity of the product. In general terms, the product purity is a function of the flow rate through the modules, while the number of parallel modules determines the volumetric production capacity.

The efficiency of a particular membrane module is a measure of the amount of nitrogen produced from a unit of air consumed. The efficiency and production capacity depend on the overall permeability and selectivity of the membranes, because there is always a certain amount of nitrogen that is lost with the oxygen. Therefore, a given membrane combination will produce a known amount of nitrogen product of a specific purity, if the unit is operating normally.

Physical Description The membranes are hollow polymeric fibers. The diameter of the fibers is on the order of some hundreds of microns. The fibers are composed of any of a number of polymers, including polysulfones, polycarbonates, and polyimides. The fibers are bundled together in a fashion similar to that of a shell and tube heat exchanger The bundle of fibers of specified length is effectively sealed on both ends to tube sheets.

During operation, the higher-pressure air passes through the inside of the fibers, and oxygen is permeated to the outside (the shell side). The entire membrane adsorption system is packaged in a cabinet with operator controls conveniently located on the front panel. Cabinets are normally installed indoors at the customer location.

CRYOGENIC AIR SEPARATION

The term *cryogenic* refers to any material or process that operates at very low temperatures, typically below 120°K (–243°F, or –153°C). The use of distillation to separate air into its component is a mature technology. The use of cryogenic engineering in the liquefaction of air was initially used at the beginning of the 20th century.

Cryogenic processes produce oxygen, nitrogen, argon, and other rare gases from atmospheric air. The selection from among these processes and cycles depends on the product volume, phase, and purity requirements as well as on the capital cost and energy consumption constraints.

An advantage of cryogenic over adsorption processes is their ability to coproduce oxygen, nitrogen, and argon as well as their ability to liquefy a portion of the total output. Storage of liquefied product in vacuum-insulated tanks serves as backup for gaseous product systems during outages or for the liquid filling of truck trailers for merchant sales.

This section will describe the low-pressure (LP) cycle and the associated process controls in a typical operation. Cryogenic processes utilize the same distillation principles as do higher temperature processes, except that the distillation equipment, heat exchangers, and associated piping are located inside an insulated enclosure referred to as a cold box or can, depending on whether the enclosure is rectangular or cylindrical in shape, respectively.

Front-End Air Purification

The cryogenic air separation process consists of two main operations. The first is the front-end purification of the air feed before it is sent to distillation. This step is described in Figure 8.37g. Once the air is purified, the air stream is split in two and both streams are sent to the cryogenic distillation (cold box) operation, which is shown in Figure 8.37h. The main heat exchanger is shown on both figures, completely on Figure 8.37h and partially on Figure 8.37g.

Feed Air Preparation The most commonly used process for the production of oxygen and/or nitrogen is the low pressure or double column cycle. In this process, the main feed air is compressed to a range of 60–200 psia, the heat of compression is removed by a cooler, and the high boiling impurities (H_2O, CO_2, N_2O, and certain hydrocarbons) are removed by an adsorption system (Figure 8.37g).

The purified air is divided into two streams, one of which is compressed with a booster compressor (shown on Figure 8.37h, which is a continuation of Figure 8.37g) to a range of 100–1200 psia. Both feed air streams are cooled in the main heat exchanger (schematically shown in both Figures 8.37g and 8.37h) by indirect heat exchange with cryogenic return streams.

Reversing Heat Exchangers Prior to 1980, most air separation plants used reversing main heat exchangers to remove the high boiling impurities such as moisture and carbon dioxide (CO_2) from the feed air. In these early plants, the moisture and CO_2 in the feed air to the cold box were allowed to freeze in the main heat exchanger that was cooled by the counterflow of cryogenic gas.

This accumulation was allowed to continue for a set time period or until the exchanger began to increase its pressure drop or began to lose some of its heat-transfer capability. At that point, the flow direction through the exchanger was reversed, and dry, CO_2-free waste gas was sent through it to purge the accumulated impurities to an atmospheric vent.

The reversing heat exchangers were very large and tended to require substantial maintenance due to leaks, which were caused by frequent and severe temperature cycling. Moreover, unlike the type of front-end adsorber systems shown in Figure 8.37g, the reversing exchanger also suffered from an

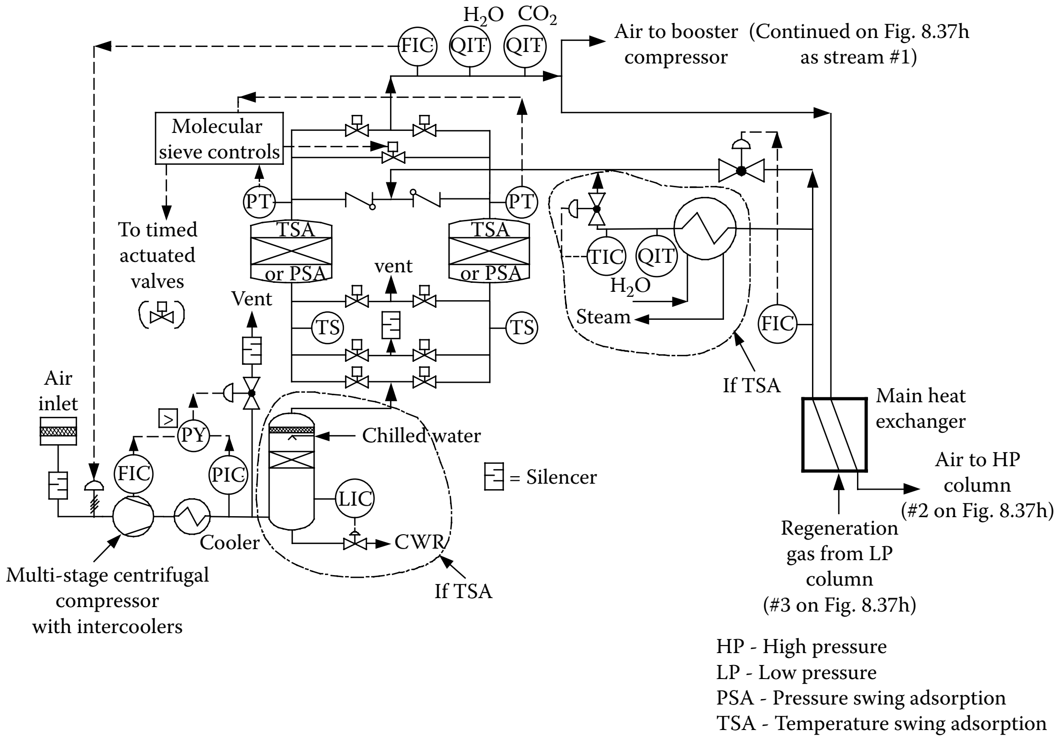


FIG. 8.37g
The equipment and main instrumentation used in the feed section of the cryogenic distillation process. (Courtesy of Air Products and Chemicals, Inc.)

inability to remove acetylene and other particular hydrocarbons, which could then concentrate in the liquid oxygen and, thereby, cause an unsafe condition, having the potential for energy release.

To address this safety concern, plants that use reversing heat exchangers continuously process all their liquid oxygen in the distillation system through adsorption beds to remove the hydrocarbons. The adsorbers are then periodically regenerated with nitrogen.

Feed Air Adsorption Systems Today, most cryogenic air separation plants are built with an adsorption system front end that features either temperature swing adsorption (TSA) or pressure swing adsorption (PSA), similar to the process illustrated in Figure 8.37g.

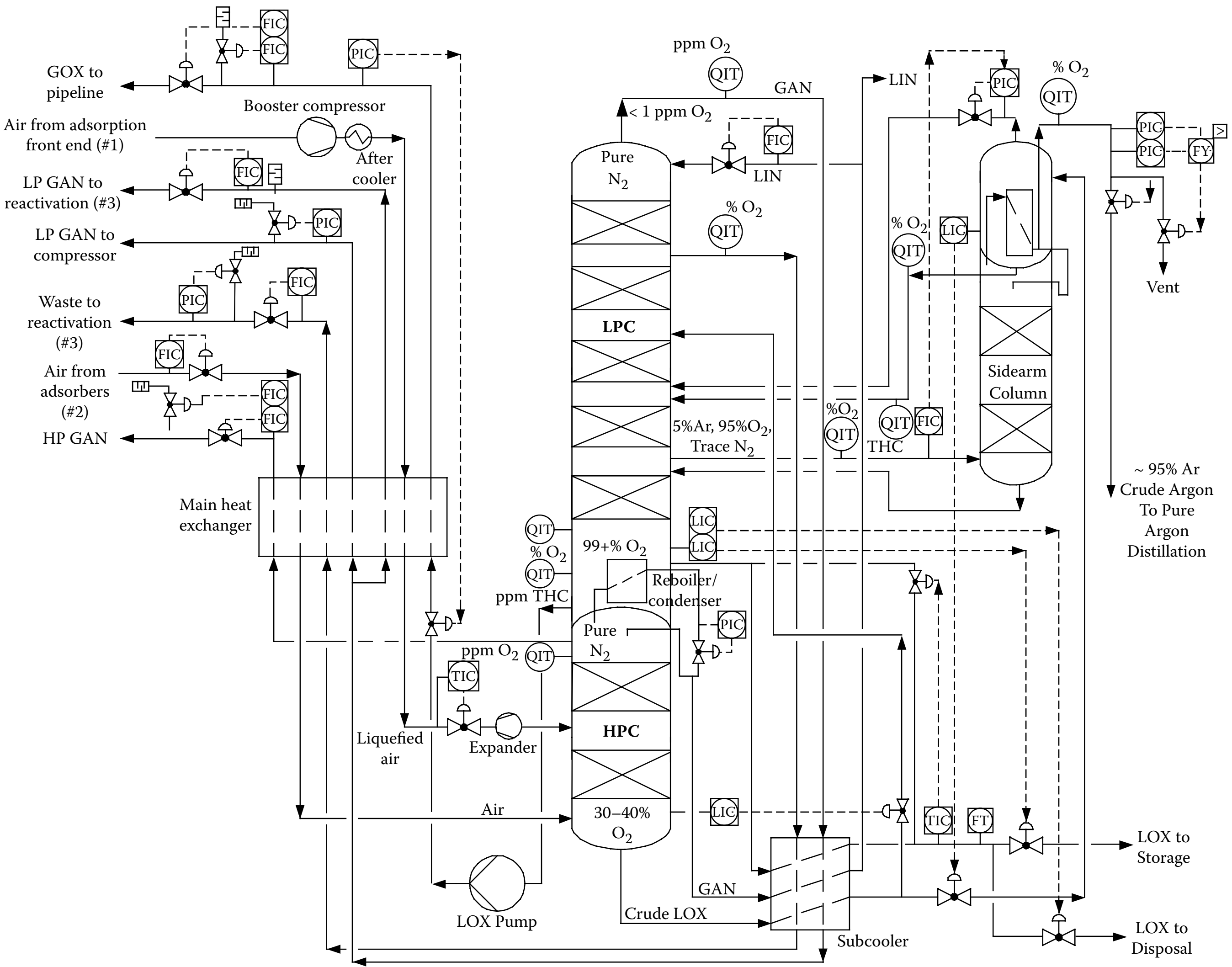
In either case, two or more vessels packed with alumina beads or with a molecular sieve are used, which packing has an affinity for high boiling impurities such as moisture, carbon dioxide, nitrous oxide, acetylene, and a number of other impurities. Removal of these impurities is critical, because if they freeze in the cryogenic section of the plant, they obstruct the flow passages.

Both the TSA and the PSA processes consist of two or more vessels packed with adsorbant beds. During operation, one vessel is on-line and is removing the impurities from the compressed feed air, while the other vessel is being regenerated by being purged with dry and CO_2-free waste gas, which is then vented to atmosphere. There are a number of operational differences between the two technologies that will be described in the paragraphs that follow.

Programmable control systems provide the required valve switching logic for the operation. This logic is based on a variety of process measurements and timing functions. Because the front-end system does not completely remove all impurities, the distillation system must be periodically shut down and defrosted. This defrosting operation is either scheduled on a preventive maintenance basis (typically 3–5 years) or scheduled sooner if dictated by deteriorating operating conditions such as increasing pressure drops across heat exchangers.

Temperature Swing Adsorption The TSA process consists of two or more vessels that are packed with a bed of adsorbant material. These vessels are switched from on-line to regeneration modes every 5 to 12 hr, depending on the design of the plant. While one vessel is on-line and is removing impurities from the compressed air feed, the other is being regenerated by a countercurrent purge flow of high-temperature dry and CO_2-free waste gas (stream #3 on Figure 8.37g).

The TSA waste gas heater uses whatever heat source is most economical. This could be low-pressure steam, electrical, or direct-fired natural gas heat. The TSA beds have greater affinity for adsorption when operated at lower temperatures. For this reason and to enhance removal efficiency, the compressed feed air is cooled with a direct contact after



GAN - Gaseous nitrogen
GO - Gaseous oxygen
HP - High pressure
HPC - High pressure column
LIN - Liquid nitrogen
LOX - Liquid oxygen
LP - Low pressure
LPC - Low pressure column

FIG. 8.37h

The equipment and main instrumentation used in the cryogenic distillation process in which the pumped liquid oxygen (LOX) cycle produces N_2, O_2, and crude argon. (Courtesy of Air Products and Chemicals, Inc.)

cooler (DCAC) that uses chilled water that is generated by the cooling effect of the excess cryogenic nitrogen waste gas that is not used to regenerate the beds. By operating at lower temperature, the impurity removal capacity is increased. This is because the required size of the TSA unit increases with water load, which in turn increases with air temperature.

Pressure Swing Adsorption The PSA process also consists of two or more vessels that are packed with a bed of adsorbant material. These vessels are also switched from on-line to regeneration modes but they are switched several times per hour, much more frequently than with TSA. While one vessel is on-line removing impurities from the feed air, the other is being regenerated by being purged with dry CO_2-free waste gas, which is then vented.

Comparison of TSA vs. PSA TSA adsorber beds are most commonly used in the newer air separation plants while PSA is considered a niche application that is most often used in smaller plants. The PSA process is sometimes favored if there is no economical heat source available for regenerating the TSA, such as low-pressure steam or natural gas.

The PSA does utilize a simpler, less capital-intensive compressor after cooler, as compared to the more expensive

DCAC, which is most often associated with TSA. The PSA process depends on pressure difference to separate impurities from the air and requires more regeneration gas than does TSA, which uses heat to regenerate.

Because the PSA process is often oversized to obtain the required capacity, PSA cycling can result in a variation in the feed airflow to distillation. Finally, TSA provides a secondary benefit if there are airborne contaminants that need to be scrubbed, because its DCAC tower does also provide scrubbing.

Cryogenic Distillation (Cold Box) Process

The equipment, piping, and basic instrumentation used in the cryogenic distillation (cold box) process is shown in Figure 8.37h. Here, the higher-pressure air from the booster compressor on Figure 8.37h is liquefied and fed through a Joule Thompson (JT)-type expander valve. In Figure 8.37h, the liquefied air stream is shown as a feed to the high-pressure column (HPC), but it could also be fed to either or both the HPC and LPC columns. Stream #2, the lower-pressure air stream, is shown as a feed gas to a point lower in the HPC.

Oxygen Production Figure 8.37h describes the design of the pumped liquid oxygen cycle, the "Pumped LOX" cycle. As the name implies, in a pumped LOX plant, the liquid oxygen (LOX) is pumped from the low-pressure column (LPC) sump to serve other process needs; this will be described in more detail later.

Oxygen is purified at the bottom of the LPC and is removed to storage either as gaseous oxygen (GOX) or LOX, depending on the customer requirements. The boiling point and relative volatility of argon is between that of nitrogen and oxygen. In the mid-section of the LPC column, the argon concentration is about 10%, and it is drawn off as feed to a sidearm column that produces crude argon.

All the oxygen, which enters with the gaseous air (stream #2 in Figure 8.37h), leaves in the bottoms from the high-pressure column as impure or crude LOX, in which the oxygen concentration is in the range of 30–40%. The crude LOX is flashed down to a lower pressure and is either fed to an intermediate stage in the low-pressure column, or if argon is to be recovered, it is used to operate the sidearm condenser and then fed to the low-pressure column as a more vapor-rich stream.

Nitrogen Production Nitrogen, having a greater relative volatility than oxygen, concentrates as it rises through the HPC. The vapor from the top of the column is condensed against boiling liquid oxygen in the reboiler/condenser. The condensed overhead is divided into a reflux stream that is returned to the HPC and a reflux stream that is sent to the top of the LPC.

Nitrogen is purified at the top of both the HPC and LPC, and this product can be extracted from either column overhead, depending on the process cycle, but is often withdrawn from the LPC overhead to prevent limiting the HPC reboiler/condenser of reflux.

Argon Product Argon is produced at the top of the sidearm column where impurities include oxygen at a range of 1 ppm to 4.0% and nitrogen at the ppm levels. If pure argon is required and the product from the sidearm column contains unacceptable oxygen levels (e.g., up to 4%), then oxygen is removed through additional purification steps. Pure argon is produced either directly from distillation or through a process that uses catalytic deoxidation of a crude argon stream.

Waste Gas The last remaining major process stream is the waste stream that is extracted from the upper portion of the LPC and contains principally nitrogen but also some small fractions of oxygen and argon, depending on the recovery characteristics of the plant. This purge stream provides refrigeration for the main heat exchanger and dry, CO_2-free regeneration gas for the reactivation of the front-end purifiers (stream #3 in Figures 8.37g and 8.37h).

CRYOGENIC INSTRUMENTATION

Severe-Service Control Valves

The LOX control valves or air JT valves often operate at pressure drop of 1000 psig or more, where flashing or cavitation normally occurs (Figure 6.1y in Chapter 6 describes some anticavitation valve designs). The valve seats in such severe services must be hardened, using such material as stellited trim. It is also desirable to keep the flow velocities as low as possible to reduce erosion of the valve trim and to minimize noise (see Section 6.14 in Chapter 6 for details).

Flashing and cavitation can cause severe erosion of the control valve trim, which can destroy the valve. It is for this reason that special valve designs are required for such services. These valves always require positioners.

The outlet area of severe-service valves should be maximized to better handle the expanding gas and to limit or prevent cavitation as much as possible. Design is complicated in multicomponent streams when flashing and cavitation process conditions exist. Still, one can properly specify and size control valves for cavitating service, while flashing is a consequence of process conditions and cannot be eliminated through valve selection or design.

Impulse and Sample Lines

All impulse or sample lines must include blow-down connections for periodic defrost operations. The impulse lines serving pressure transmitters, differential flow transmitters, and differential pressure level transmitters should be provided with liquid sealed legs inside the cold box. This design prevents large frost accumulations on the impulse lines and on the valves that are located outside the cold box, caused by the boiling of cryogenic liquid and freezing of ambient moisture.

On the other hand, continuously flowing cryogenic liquid samples to analyzers should not have seal loops, because these loops can cause the boiling of the sample inside the cold box and, thereby, cause the distilling of the sample, which makes it nonrepresentative.

The design of impulse and sample tubing/piping must also consider the stress caused by thermal expansion and contraction of cryogenic equipment, as well as the weight loads imposed by the insulation (typically perlite) that is packed in the cold box.

Flow Elements

Venturi and nozzle-type flowmeters (Section 2.29 in Chapter 2 in Volume 1) work well and are frequently used for the measurement of cryogenic liquid and gas flows. In some cases, cryogenic flows are approximated on the basis of the opening of control valves and of the pressure drop across them.

Outside the cold box, elbow meters (Section 2.29 in Chapter 2 in Volume 1) are occasionally used where line sizes are very large and where it is desired to minimize the pressure drop across the flow sensor, such as in the case of measuring TSA or PSA regeneration gas flows. The nominal 4:1 turndown capability of the differential flow devices is usually acceptable, because this rangeability exceeds the turndown requirement of the ASU plant itself.

The turndown capability of cryogenic liquid flow measurement applications is often limited by the flashing that occurs at low flows, as the control valve is throttled to the point where flashing starts. Nonetheless, smart transmitters can provide wider turndown than is possible with conventional differential pressure transmitters in flow measurement applications.

In order to increase plant efficiency, it is important that the instrumentation and control valves used will provide the required accuracy and rangeability. Inversely, if the control loop components are improperly specified or sized, they will not be able to accurately meet the controller set points, and this can cause sustained upsets.

Temperature Measurement

Temperature measurements are made by thermocouples and resistance temperature devices (RTDs). They are described in detail in Sections 4.10 and 4.13 of the first volume. They are provided with bar stock thermowells that are welded into the process lines and vessels as needed.

Thermocouple extension wire or RTD cables are run from the temperature element head to junction boxes on the outer face of the cold box.

Process Analysis in ASU

The three major purposes of using analyzers in an air separation plant are the monitoring and control of product quality, the performance of process control, and the continuous maintenance of process safety.

In a typical sampling system, the pressure of the sample drawn from the process is reduced, and this sample gas is sent to a remote analyzer panel. In order to lower the dead time of the measurement, bypass purge flow is provided around the analyzer. This also helps to minimize the impact of leaks in the sample line.

The sample system response time is computed from the sum of the sample and bypass flows, taking into account the transport system volume and pressure. The primary sample is taken from a tap that is situated in an ideal location, and sample stream-switching capability is provided at the analyzer panel, to use alternative samples taken from auxiliary sample taps, when necessary.

Product Quality Analysis Product quality analysis is required for the monitoring and control of the oxygen, nitrogen, and argon products of the plant.

The primary analytical method used for measuring the purity of the oxygen product is paramagnetic (Section 8.42 in Chapter 8 in Volume 1). The sensor in the paramagnetic analyzer is pressure-compensated and is either the dumbbell or the thermal wind variety. The specification for the oxygen product calls for 99% + purity, and the main impurity in this stream is argon.

The specification for the nitrogen product requires that its oxygen impurity be at the parts per million level. This is measured electrochemically by fuel cell, zirconium oxide, or Coulometric sensors (Section 8.42 in Chapter 8 in Volume 1).

The product specifications for the argon product limit the impurities of oxygen nitrogen and total hydrocarbon (THC) to trace amounts. The trace oxygen analysis is made by the same methods that are used for the measurement of trace oxygen in the nitrogen product.

The trace nitrogen in argon can be measured by either gas chromatography (Section 8.12 in Volume 1) or by spectrographic quartz plasma-type analyzers. THC is measured with a flame ionization detector (FID), and the results are reported as methane-equivalent THC.

Normally, all three products are analyzed for moisture (dew point or ppm), and this is typically done by analyzers having aluminum oxide or quartz crystal microbalance sensors (Section 8.33 in Chapter 8 in Volume 1).

Trace levels of carbon monoxide, hydrogen, helium, and neon are also present in atmospheric air. They are likely to pass through the HPC and concentrate in the nitrogen product and reflux. It is sometimes necessary to install a noncondensable purge on the condensing side of the LPC reboiler to purge these gases.

Similarly, small traces of krypton and xenon, which are present in ambient air, will also process through the system and concentrate in the LOX product in the HPC sump. These trace components can also be recovered and concentrated by specialized processes. The processing of krypton and xenon must also include the removal of hydrocarbons, because if they are allowed to concentrate in the oxygen, unsafe conditions can evolve.

Safety Analyzers For the purpose of safety, carbon dioxide and total hydrocarbons are measured in the ppm range or lower.

Samples are typically taken from the feed air supply to the plant and from the liquid oxygen streams. The trace carbon dioxide is typically measured with non-dispersive infrared (NDIR) methods that utilize the Luft principle or by gas filter correlation (GFC) methods (Section 8.27 in Volume 1).

Analysis of total hydrocarbon is typically performed using flame ionization detector analyzers, and the results are reported as methane equivalent hydrocarbon in ppm.

If thermal deoxidation units (Deoxo) are used to remove oxygen from crude argon, a safety analyzer is required to monitor the oxygen level in the crude argon feed to the Deoxo process to prevent high oxygen levels from causing unsafe thermal runaway of the catalytic bed. A paramagnetic or electrochemical method is used to monitor the percentage of oxygen in the feed to the Deoxo process.

Occasionally, safety analyzers are used to monitor for the unlikely situation where high levels of oxygen (> 21%) contaminate the waste gas stream used to purge the front-end adsorbers. Such monitoring can become necessary because adsorbers that are constructed of incompatible materials may present a risk of fire when exposed to enriched oxygen.

Finally, area monitors are used to analyze the air quality in occupied rooms to ensure that ventilation is adequate and to provide alarms if either an oxygen-depletion or an oxygen-enrichment condition is evolving, with the corresponding risk of asphyxiation or oxygen fire.

Process Control Analyzers

The product purity analyzers discussed earlier are also used in controlling the air separation process. The product purity or safety analyzers also trigger alarms and safety interlocks.

Other process control-related analytical loops include the control of excess hydrogen in the crude argon from the Deoxo process. The concentration of hydrogen is measured by thermal conductivity type analyzers (Section 8.57 in Chapter 8 in Volume 1).

Carbon dioxide in the feed air is measured after the front-end adsorption system and is monitored by NDIR-type analyzers (Section 8.9 in Chapter 8 in Volume 1).

Dew point is detected by aluminum oxide sensors (Section 8.33 in Chapter 8 in Volume 1) on the outlet of the TSA regeneration gas steam heat exchangers and compressor aftercoolers. In case a leak causes the dew point to rise, an alarm is actuated.

The hydrogen supply to the Deoxo systems is monitored for trace methane, because this methane is not removed by the Deoxo process or by argon distillation, and it would become an impurity in the pure argon product.

Nitrogen in the sidearm column feed is either inferred on the basis of LPC operating conditions or is directly measured by gas chromatography or by ion mobility spectrometers. The gas chromatograph first removes the oxygen and then chromatographically separates the nitrogen and argon for analysis. The ion mobility analyzers provide an effective method of nitrogen analysis in this mixed gas stream but have the disadvantage of using a radioactive source and therefore requiring special permits and are rarely used today.

REGULATORY AND FEEDFORWARD CONTROLS

The controls described in the following paragraphs refer to the subsystems of the process shown in Figures 8.37g and 8.37h. The control loops are not shown in detail, and the feedforward and logic/safety controls are not shown at all. Therefore, the reader is asked that for in-depth, detailed discussions of the related algorithms, dead times, time constants, and tuning, refer to Chapters 2 and 8, where they are discussed.

Both regulatory and advanced process controls are used to optimize the production of a specified product mix and to maintain product purity at minimum operating cost. In controlling the air separation processes, these control strategies have to be adapted to optimize production against purity and energy constraints, while also correcting disturbances that would upset the steady-state operation.

Regulatory controls serve to respond to upsets caused by feed airflow disturbances, which can be caused by upsets resulting from TSA or PSA regeneration or from diurnal variations in cooling water temperature, which can affect heat transfer. Advanced control technology is also used to provide load-following optimization strategies so as to ramp plant production in response to variations in customer gas demand.

Process control systems used to implement these objectives include large distributed control systems (DCSs), which are often linked to supervisory control computers for implementation of advanced controls by means of various forms of model predictive control (MPC).

In the following paragraphs, some specific control problems and solutions will be discussed.

Main Air Compressor Flow

As illustrated in Figure 8.37i, the compressed air is vented to provide the added throughput required to keep the main air compressor (MAC) out of surge. (See Section 8.15 for details on compressor control and optimization.)

The surge flow measurement is typically inferred from differential pressure measurement across the compressor stages. The surge line for the compressor is confirmed by surge tests. Some compressor suppliers do not use surge control, but only open the vent when a maximum operating pressure is reached. The level of control sophistication applied is a function of the relative importance to save energy by reducing the surge margin.

Stabilization of Pressure Surge

During regeneration of the front-end adsorbers, it is important to minimize the pressure surges, because these upsets have the potential of causing a disruption of the downstream distillation process, which in turn can disrupt the argon product

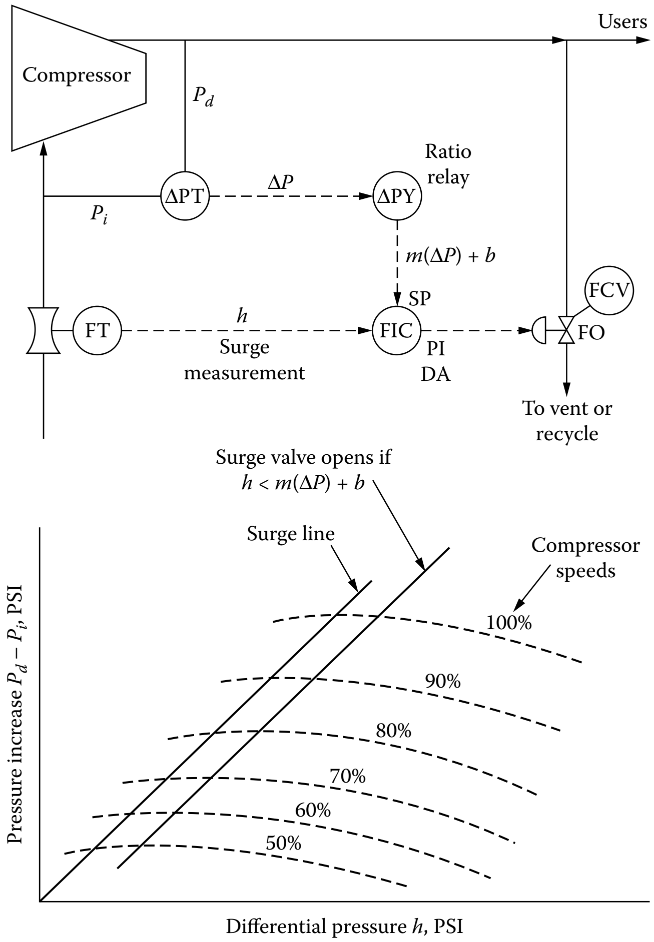


FIG. 8.37i

The flow across the compressor (h) is kept above the surge limit by venting, if the user demand is insufficient to keep it out of surge.

quality profile. Unlike the reversing exchangers that cycled multiple times per hour, the TSA adsorbers are regenerated only once in several hours.

If one has properly determined the process dead times and time constants, feedforward control can be used to attenuate these disturbances. The feedforward loop might include a ramping function to increase the main air compressor flow (the multistage centrifugal compressor in Figure 8.37g), in order to compensate for the increased airflow required when the off-line adsorber vessel is being pressurized after regeneration. After the regenerated unit is on-line and the beds are switched to regenerate the other adsorber vessel, the main air compressor flow is ramped back to its normal operating flow.

Sidearm Nitrogen Control

The feed flow disruption caused by adsorber regeneration is particularly important when argon is being produced, because the argon composition profile in the LPC is sensitive to feed flow, and feed flow upsets introduce a risk of excess nitrogen entering the sidearm column. Nitrogen impurity is normally in the low ppm range. Nitrogen excursions in excess of fractions of a percent in the sidearm column could vapor-lock the condenser and cause loss of vapor flow and reflux, resulting in the draining of the contents of the sidearm column into the LPC.

After a dumping episode, the reestablishing of oxygen purity in the LPC and also of the proper operation of the sidearm column can take many hours to complete. The challenge is in managing the sidearm column feed to maximize argon concentration while preventing too much nitrogen, which will cause the sidearm column to dump.

On the other hand, if the plant is operated conservatively with little or no nitrogen in the sidearm, argon will ultimately be lost in the waste and recovery will be lower than otherwise possible. Control of nitrogen in the sidearm is implemented by determination of composition in the sidearm or at a location several stages above the sidearm feed point by direct analytical measurement (see QIT in Figure 8.37h).

Operating personnel will tend to run the distillation column more conservatively to limit the nitrogen concentration in the sidearm to very low levels, but this also reduces recovery because incrementally more argon will be lost in the waste stream. Therefore, to maintain high argon recovery, the plant requires precise control of sidearm composition, and advanced control systems are capable of providing it.

Product Purity Control

In general, for a plant with a liquid product and a fixed air flow, the product purity is inversely related to the flow rate of product that is being removed from the plant. Therefore, product flow control is often used to control product purity. This control strategy is complicated by the fact that control loops often interact, necessitating feedforward control algorithms and model predictive control strategies for optimal load-following.

The performance of a load-following optimization system is dependent upon the maximum ramping rate of the air separation plant, the severity of demand changes, and the capacitance of the system. The objective is to minimize the venting of overproduction, while avoiding (minimizing) supplemental liquid vaporization, which makes up for production deficiencies.

Oxygen Recovery Controller

In steady-state operation, where the customer has specified a constant GOX production rate, the oxygen purity is a function of the oxygen recovery. The fixed GOX customer flow being the independent variable, the supplied airflow to the plant is the primary manipulated variable that impacts oxygen recovery.

Therefore, the oxygen purity controller (QIT at the bottom of LPC in Figure 8.37h) can act as a cascade master to manipulate the ratio of feed air to product flow. This manipulation affects the power consumption of the plant, and for a fixed number of distillation stages, increasing the feed airflow will improve the product purity. The control loop is configured as a triple cascade where the O_2 recovery controller (QIT, %O_2) is the cascade master of the purity controller (QIT, ppm THC), which is the cascade master of the feed flow controller (FIC on the compressor in Figure 8.37g).

Front-End Purity and Hydrocarbon Accumulation

Methane, ethane, propane, and to a certain extent ethylene are not adsorbed in the TSA or PSA-type front-end air purification processes. Acetylene, which is dangerous due to low solubility, is captured in the TSA or PSA molecular sieve. In addition, there are impurities that can accumulate and plug equipment, such as traces of CO_2 that slip through the front-end adsorbers.

Plugging increases the potential for accumulating local concentrations of hydrocarbons by blocking passages in the LOX sump reboiler/condenser section (bottom of HPC column in Figure 8.37h), which normally are continuously flushed by siphon action.

Trapped oxygen may create a localized concentration of hydrocarbons as a result of dry boiling in the blocked passages and may eventually leave a residue of pure hydrocarbon in a pure oxygen atmosphere that could potentially autoignite, with potentially severe consequences.

Normally, in a pumped LOX plant, it is the front-end CO_2 removal characteristics that determine the minimum amount of LOX that must be removed as a product or in a purge stream.

Hydrocarbon Concentration Factor

A hydrocarbon concentration factor can be determined on a material-balance basis by determining how much LOX is being removed from the LPC reboiler compared to the total oxygen in the air that is entering the system. In a pumped LOX plant, the concentration factor is approximately 5 because the entire oxygen product is leaving the column as a liquid. In a plant that produces gaseous oxygen, where most of the oxygen product is generated as GOX, with only a small amount as LOX being produced, the concentration factor might be in the range of 100–500.

Total hydrocarbon is continuously measured at several points in the plant. These include either the LPC sump (pure oxygen) or the crude liquid oxygen from the sidearm column condenser sump.

REFRIGERATION CONTROLS

In the early days of cryogenic air separation, the feed air was liquefied using a so-called “split cycle” that involved compressing the air to approximately 2000 psig and then cooling the compressed gas. After that, the cold gas was sent through an expander valve to cool it by the Joule Thompson effect. This expander valve is referred to as a Joule Thompson valve or JT valve.

Today, taking a portion of the air after it has been purified (stream #1 in Figure 8.37g), boosting the pressure with a booster compressor, cooling the stream, and then expanding it through a mechanical expander provides the refrigeration required for liquefaction.

Expander Configuration

The refrigeration required in oxygen plants to make up for the heat leak from the environment is provided by machinery that expands a portion of the feed airflow to the low-pressure column. The work done by the expander can be recovered as electrical power. The recovery can be achieved by using a generator-loaded expander, or by taking that portion of the feed that is to be expanded and first compressing it to a higher pressure by a compressor that is mounted on the same shaft as the expander. This machine is commonly referred to as a compander.

In a plant making GOX only, gaseous air would be taken from the TSA, cooled in the main heat exchanger, and expanded directly into the LP column. More refrigeration is needed if either some amount of liquid oxygen product is also needed or if the expander flow is starting to impact the oxygen recovery. In such a case, a compander can be used (instead of recovering the work as electricity), to boost that fraction of the air that is to be expanded. This way, a higher pressure ratio can be obtained across the expander.

Refrigeration Balance Controller

Refrigeration to the plant is typically controlled by manipulation of expander flow. Normally, an expander flow is so selected as to match the cooling needed to produce a little more than the minimum LOX production required. At a constant expander flow, the liquid oxygen sump level controller (LIC at bottom of HPC in Figure 8.37h) will provide steady-state operation.

Main Heat Exchanger Control

On a pumped LOX plant, high-pressure air from the booster compressor is condensed in the main heat exchanger, and the JT valve essentially controls the liquid level in the exchanger, while the booster compressor discharge pressure effectively controls the vapor inventory in the exchanger.

The oxygen in the feed to the plant is an independent variable based on customer oxygen flow requirements, and the booster compressor discharge pressure is held constant. By maintaining a constant discharge pressure, the molecules that condense in the exchanger are effectively offset by the oxygen molecules that are vaporized.

The liquid level measurement in the exchanger is difficult, particularly, because above the critical pressure, the vapor/liquid interface disappears. Knowledge of the liquid level is important, as this determines the available surface area for condensation as well as the magnitude of the temperature approach. The temperature approach translates into a pressure ratio between the high-pressure air and the oxygen that is boiling in the exchanger.

There are multiple complex positioning strategies for the temperature element of the temperature controller (TIC) that throttles the JT control valve. Proper tuning of this controller is also important to provide the required refrigeration.

ADVANCED CONTROL

There is no doubt that advanced controls can improve the plant efficiency, as well as argon recovery, over what is possible through regulatory controls alone. Advanced control systems can interface with the distributed control system using supervisory computers that are used for regulatory process control to assist in control and optimization. Advanced control systems, using model predictive control, can predict evolving conditions and, as such, can optimize the operation through model-based anticipation.

Without advanced control, the plant will still run reliably and will deliver good product, but operators tend to pick safe controller set points, which can be further away from operating or safety constraints than necessary. The safety margins used by the operator are usually a function of both their motivation and their experience/education. Therefore, while the plant will run reliably under manual operator control, its operation will not be optimized. Such unoptimized operation is likely to result in significant losses in argon recovery and is a significant increase in air compressor power consumption.

The advanced controls monitor the operating constraints while meeting the production target at minimum waste of either product or energy. In the air separation process, advanced controls can improve the production and purity of GOX, the sidearm feed composition, the flow and purity of argon, and the total cryogenic liquid production.

Characterizers to Compensate Process Nonlinearity

Nitrogen purity can be controlled by using the nitrogen purity controller as the cascade master of the airflow controller that sets the gaseous air feed flow to the HPC as its slave, if its set point is not already controlled by GOX purity. While linear response is assumed between the composition of the overhead product (GAN) and changes in boil-up and reflux ratio, these relationships are not linear over the whole range of the operation.

As the column approaches a state of operation that corresponds to minimum reflux and maximum recovery, it becomes difficult to recover any more nitrogen, and the bottom liquid (LOX) will approach equilibrium with the air feed. If, under these conditions, more nitrogen is taken off, the process gain drastically changes and product purity drops (oxygen impurity in the nitrogen rises).

In general terms, if the process gain varies with load, stability of the loop can only be maintained if the variation in the process gain is compensated, because stable control requires a stable gain product of the loop components of about 0.5 (Figure 8.37j). One way to achieve this is to remove the nonlinearity from the measurement (of nitrogen purity, in this specific case) by a compensator. Such characterizer can take the logarithm of the product quality signal, for example, to arrive at a constant loop gain.

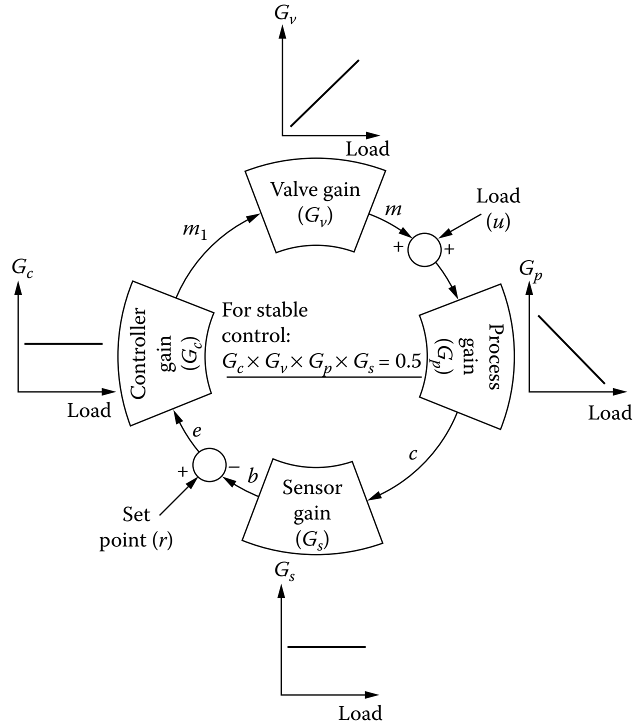


FIG. 8.37j
If the process gain varies with load, it is necessary to compensate for that nonlinearity. This can be done by characterizing the control valve (shown above) or can be done by characterizing the measurement.

Operating Close to Constraints

The most efficient plant operation is obtained if the process is run close to the limitations that are set by the equipment used, the available raw material and utility supplies, market conditions, product quality requirements, power consumption, and so on. If one focuses on the objective of minimizing energy consumption, then airflow should be reduced until the limitations set by the purity requirements of the product are approached. This is because an increase in airflow also increases the boil-up and reflux rates, which in turn results in higher product purity, but at a higher investment of energy.

The ratio of the power consumption and the production rate of a plant expresses its energy efficiency. Air separation plants can have six or more constraints that are alternately approached during operation. When a constraint is approached, the control system should load or unload the plant, as needed to maintain a margin from the constraint. The more constraints are approached at the same time, the more difficult the control challenge becomes, but selective control can handle that task.

Therefore, good plant design should eliminate the possibility of the plant operation being able to approach multiple constraints that require conflicting control responses. For example, no extra stages of distillation should be added in order to obtain better purity than the purity needed at near ideal recovery.

CONCLUSIONS

In this section, the process controls for air separation processes using adsorption, membrane, and cryogenic technologies have been discussed.

Cryogenic plants are the economic choice for most medium to large-scale oxygen and nitrogen production applications, because of their lower unit power consumption and economies of scale. They represent the only choice for argon production. LOX, LIN, and LAR facilities are available for backup and peak shaving to gaseous supply systems and are readily available on-site, using the same equipment that produces the gaseous products. This eliminates the need for add-on liquefiers or the use of supply contracts.

Adsorption systems for the production of oxygen can supply 85–95% oxygen and are best suited for applications requiring less than 100 tons per day production. Nitrogen-producing PSA and nitrogen membrane systems are economical for smaller scale production volumes and where the high purity products of cryogenic systems are not required.

ABBREVIATIONS

ASU	Air separation unit
CMS	Carbon molecular sieve
DCAC	Direct contact after cooler
Deoxo	Deoxidation unit
GFC	Gas filter correlation
GOX	Gaseous oxygen
HPC	High pressure column
JT	Joule Thompson
LAR	Liquid argon
LIN	Liquid nitrogen
LOX	Liquid oxygen
LPC	Low pressure column
MAC	Main air compressor
MPC	Model predictive control
NDIR	Non-dispersive infrared
ppmV	Volumetric parts per million
PSA	Pressure swing adsorption
THC	Total hydrocarbon
TPD	Tons per day
TSA	Temperature swing adsorption
VSA	Vacuum swing adsorption

References

1. Parker, S. P., Ed., *Encyclopedia of Science & Technology,* Volume 3, 8th edition, New York: McGraw-Hill, 1997.
2. www.airproducts.com.
3. SRI International, CEH, November 1999.
4. Beh, C. C. K., and Webley, P. A., "A Method for the Determination of Composition Profiles in Industrial Air Separation, Pressure Swing Adsorption Systems," Department of Chemical Engineering, Monash University, October 2002.

Bibliography

Butricia, A. J., *Out of Thin Air,* Praeger, 1990.

Castle, W. F., "Air Separation and Liquefaction: Recent Developments and Prospects for the Beginning of the New Millennium," *International Journal of Refrigeration,* 25, pp. 158–172, 2002.

Murphy, K., Odorski, A., Smith, A., and Ward, T., "Oxygen Production Technologies for Non-Ferrous Smelting Applications," Air Products and Chemicals, Inc., 1987.

"Safe Operation of Reboilers/Condensers in Air Separation Units," Document 65/99, Brussels, Belgium: European Industrial Gas Association, Industrial Gas Committee.

"Safe Practices Guide for Cryogenic Air Separation Plants," CGA P-8, Chantilly, VA: Compressed Gas Association.

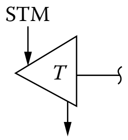
Schmidt, W., Kovak, K., Licht, W., and Feldman, S., "Managing Trace Contaminants in Cryogenic Air Separation," AIChE Spring Meeting, Atlanta, GA, March 5–9, 2000.

Schmidt, W., Winegardner, K., Dennehy, M., and Castle-Smith, H., "Safe Design and Operation of a Cryogenic Air Separation Unit," *Process Safety Progress,* Vol. 20, No.4, pp. 269–279, December 2001.

Standen, A., Ed., *Encyclopedia of Chemical Technology,* 2nd Edition, New York: Interscience, Volume 14, 1967.

8.38 Steam Turbine Controls

E. J. FARMER (1985) **B. G. LIPTÁK** (1995, 2005)



Flow sheet symbol

INTRODUCTION

Steam turbines are used in the generation of electric power, particularly in combined cycle power stations (Section 8.33), and also as variable- or constant-speed drives (Section 7.10) for larger rotary equipment, such as compressors and pumps (Sections 8.15 and 8.34). Their controls are often integrated with the steam generator (Section 8.6), which can use fossil or nuclear fuel.

In the first part of this section, the nature and the operating principles of steam turbines will be discussed first. This will be followed by a description of the various types of turbine designs, efficiencies, and applications.

In the second part of this section, the control of steam turbines will be covered. This discussion will be started with the description of the traditional steam governors and basic regulatory turbine controls. The controls of pressure let-down and extraction turbine applications and the methods of eliminating interaction and decoupling will also be discussed. The section will also describe the steam turbine safety systems and will be concluded with a description of optimization and advanced controls.

The optimization strategies described will consider the nature of the installation, which might have the goal of maximized electric power production while meeting the steam demand of the plant or the goal of maximizing direct steam utilization.

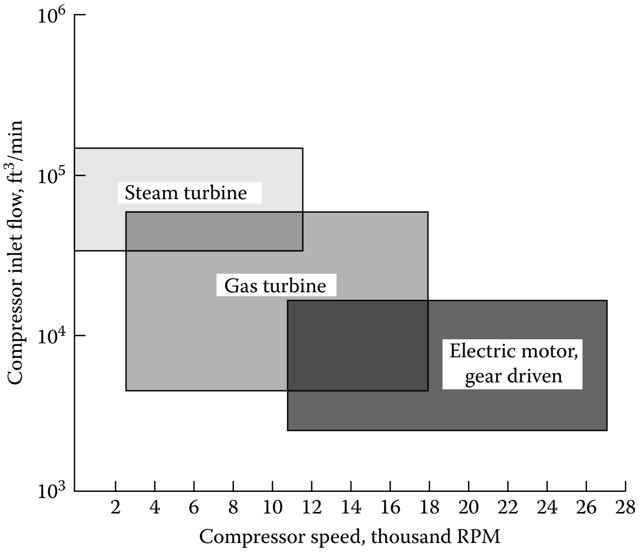
The discussion of advanced controls will cover the use of model-based predictive control and self-diagnostics for maintenance purposes. It will also touch upon thermal stress monitoring (TSM), turbine protection (TP), and monitoring/sequential control (MSC).

Characteristics

Steam turbines are energy conversion machines. They extract energy from the steam and convert it to work, which rotates the shaft of the turbine. Steam turbine sizes range from shaft output energies of a few kilowatts to well over 1000 megawatts, and there is no reason why still larger machines could not be built. No other prime mover can achieve the shaft output capability that is easily attained by large steam turbines.

Rotational speeds vary from approximately 1,800 to 14,000 rpm; this speed can be modulated over a wide range. Such variable-speed operation is an advantage, if the turbine is used to drive pumps or compressors. Figure 8.38a shows the relative operating ranges of different types of compressor drives. When provided with the appropriate speed governor, steam turbines can provide excellent speed stability, which is desirable when the turbine serves as the prime mover in electric generators.

In comparison to all other prime movers, steam turbines are very reliable. Their availability factors are high, and their maintenance costs are low. This is largely a result of the inherently balanced design that is completely free of



Speed governor classification

Governor class	*Regulation*	*Variation*
A	10%	¾%
B	6%	½%
C	4%	¼%
D	½%	¼%

FIG. 8.38a
The throughput of a compressor with a steam turbine drive is higher than those of gas turbine or electric motor-driven units.

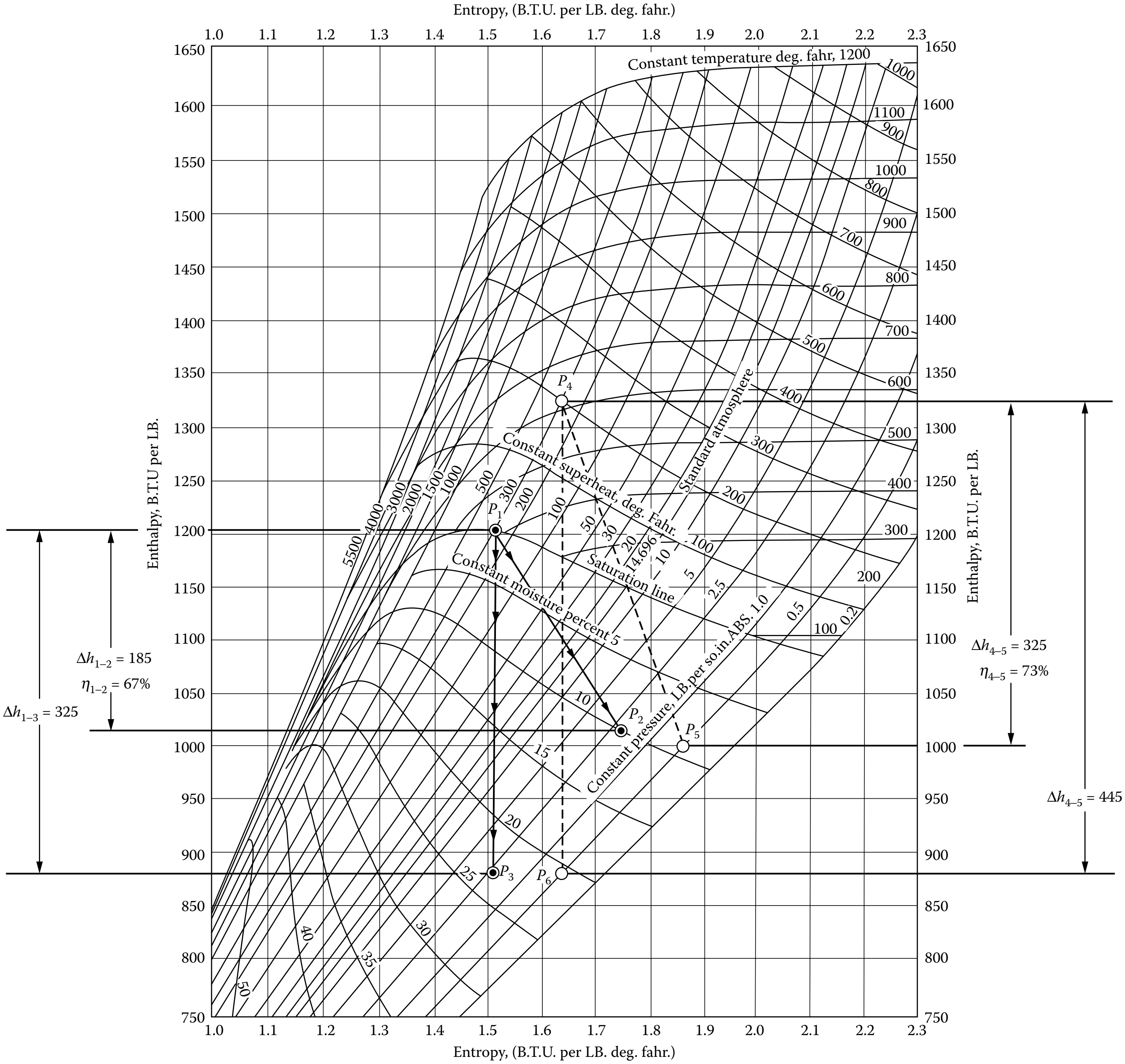


FIG. 8.38b
Mollier diagram showing performance of a steam turbine.

reciprocating or rubbing components (except, of course, for the bearings).

OPERATING PRINCIPLES

The amount of energy that the steam turbine extracts from the steam depends on the enthalpy drop across the machine. The enthalpy of the steam is a function of its temperature and pressure. Because the operating conditions of steam turbines are generally known in terms of inlet and outlet temperature and pressure, one can use a Mollier diagram as a graphic tool to determine the amount of energy available under a particular set of conditions (Figure 8.38b).

As an example, one might consider the case where the turbine inlet conditions correspond to point P_1 and the outlet conditions to point P_2 in Figure 8.38b. A line drawn between these two points is called the "expansion line" and represents the operation of the turbine as it is extracting energy from the steam. In actual operation, this line is not straight, and its shape does depend on the internal operation of the turbine.

In an ideal turbine, the steam would expand at a constant entropy (isentropically). The condition of the exhaust steam from an ideal machine that has no losses would correspond to point P_3 on Figure 8.38b. This P_3 point is found by drawing a vertical line from the inlet point P_1, down to the same exhaust pressure as the actual machine is operating at.

Steam Turbine Efficiency

If one refers to the change in enthalpy from P_1 to P_2 as Δh_{1-2} and to the change in enthalpy between P_1 and P_3 as Δh_{1-3}, one can use these quantities to calculate the turbine's efficiency and steam rate. The efficiency of the turbine, neglecting mechanical losses, is found by

$$\eta = \frac{\Delta h_{1-2}}{\Delta h_{1-3}} \qquad \textbf{8.38(1)}$$

The steam rate that an ideal machine would require to operate as theoretically predicted is called the theoretical steam rate, Q_T, and may be found as

$$Q_T = \frac{2545}{\Delta h_{1-3}} \frac{\text{lbm}}{\text{HP-hr}} \qquad \textbf{8.38(2)}$$

The actual steam rate, Q_A, is calculated by dividing the theoretical steam rate with the efficiency, hence,

$$Q_A = Q_T/\eta \frac{\text{lbm}}{\text{HP-hr}} \qquad \textbf{8.38(3)}$$

This somewhat simplified perspective of steam turbine performance can help the control engineer in both understanding and optimizing steam turbines.

In order to operate the turbine at maximum efficiency, the steam should leave the nozzles and impact on the turbine blades at sonic velocity. In order to maintain the nozzle jets at sonic velocity at partial loads, it is necessary to shut off some blocks of nozzles so that the active ones will receive approximately the same amount of steam all the time. The efficiency of steam turbines increases with size, with the superheat temperature of the steam, and with the level of vacuum at the turbine exhaust.

Example As shown in Figure 8.38b, if saturated steam enters the turbine at 300 psia and is exhausted at 2 psia, then from each pound of steam, 325 BTUs of energy can theoretically be recovered in an "ideal" turbine. If the actual expansion line is $P_1 - P_2$, the actual energy recovery is 185 BTUs per pound of steam, and therefore the efficiency of the installation is 67%.

In actual practice, turbines are not supplied with saturated steam. This is because condensation could cause turbine blade erosion due to the presence of water droplets at the high jet velocities, and also because much more energy can be recovered if the steam supply is superheated and if the exhaust pressure is reduced.

If, for example, the same 300 psia steam enters the turbine with 200°F superheat and is exhausted at 0.5 psia (instead of 2 psia), the theoretical energy recovery will rise to 445 BTUs per pound ($P_4 - P_6$). Under these operating conditions, the actual energy recovery will rise to 325 (from 185) BTUs per pound. This obviously represents a major improvement in the performance of the machine.

This improvement can also be expressed in terms of steaming rate. The steam rate of the turbine is the amount of steam needed to provide one horsepower-hour (hp-hr) of work, which is the equivalent of 2545 BTUs. If the actual operation of the steam turbine corresponds to $P_1 - P_2$ on Figure 8.38b, the steaming rate is 13.8 lb of steam per horsepower-hour, while if it corresponds to $P_4 - P_5$ it is only 7.8 lb of steam per horsepower-hour.

Advantages and Limitations

Even though the initial cost of steam turbines is usually higher than that of alternative prime movers, there are some benefits that can mitigate this cost difference. Especially in larger sizes, steam turbines are physically smaller than most other prime movers, consequently, they require less floor space. This can decrease the cost of the building in which they are to be installed.

Because the inherently balanced design of steam turbines produces considerably less vibration than do reciprocating machines, equipment foundations can also be considerably lighter. Because steam turbines are less likely to cause fires than are other prime movers, they also have an advantage in applications in locations where flammable materials are present.

In applications where the prime mover has to endure substantial overloads, the steam turbine also has an advantage. It can tolerate, without damage, such overloads that would severely shorten the service life of alternative prime movers, if they could tolerate them at all.

The optimization of steam turbine systems is well understood and more widely practiced than the power consumption optimization of alternatives such as electric motors and gas turbines. Less energy will usually be required for a steam turbine prime mover to drive a load than would be necessary if the size of a turbine driving an electrical generator were increased to provide the electric power to drive the same load with an electric motor.

Besides initial cost, the major disadvantage of steam turbines is their low tolerance for wet or contaminated steam. Wet steam can cause rapid erosion, and contaminants can cause fouling. Both will reduce the turbine's efficiency and will shorten its life. Steam quality monitoring is therefore an important requirement to maintain the reliability and to reduce the operating cost of steam turbines.

TABLE 8.38c
Performance Data for Condensing and Noncondensing Steam Turbines

Technical Data	*Units*	*Condensation Turbine*	*Back-Pressure Turbine*
Output	MW	5–120	5–120
rpm	min^{-1}	Up to 14,000	Up to 14,000
Inlet pressure	bar (psi)	Up to 130 (1,885)	Up to 130 (1,885)
Inlet temperature	°C (°F)	Up to 570 (1,058)	Up to 570 (1,058)
Discharge pressure	bar (psi)	Up to min. 0.02 (0.29)	Up to 40 (580)
Exhaust flow	M^3/s (ft^3/s)	Up to 1,300 (45,930)	—

TURBINE DESIGN CONFIGURATIONS

The energy source for steam turbines is the pressure difference between the supply and exhaust steam. The higher this pressure difference and the higher the superheat of the steam, the more work the turbine can do.

The two main categories are the condensing and the back-pressure turbines (Figure 8.38d). Table 8.38c provides a range of operating conditions and electrical power production rates for both types of steam turbines.

The exhaust pressure of a "condensing" turbine is usually subatmospheric, while that of a "noncondensing" or "back-pressure" turbine is greater than atmospheric. Condensing turbines are most often used for electric power generation, while back-pressure turbines are utilized in cogeneration power plants (Section 8.33), which simultaneously supply steam and electricity for the users.

Steam turbine installations can also be configured with more than one exhaust steam streams. A second, and sometimes even a third, outlet may be provided to allow the "extraction" of steam at different pressures (Figure 8.38e).

A rare combination of a mixed input and an extraction turbine is one in which steam can be removed when load conditions permit, or steam can be inducted when additional energy is necessary. This unit is usually referred to as *induction–extraction turbine.*

Application Configurations

A back-pressure turbine with its inlet connected to the plant's high-pressure header and its outlet supplying steam to an intermediate header is called a *topping turbine*, because it is

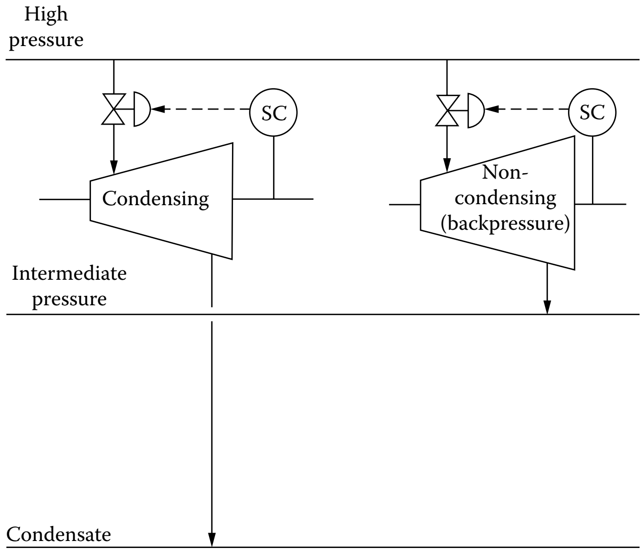


FIG. 8.38d
Typical installation of condensing- and backpressure-type steam turbines.

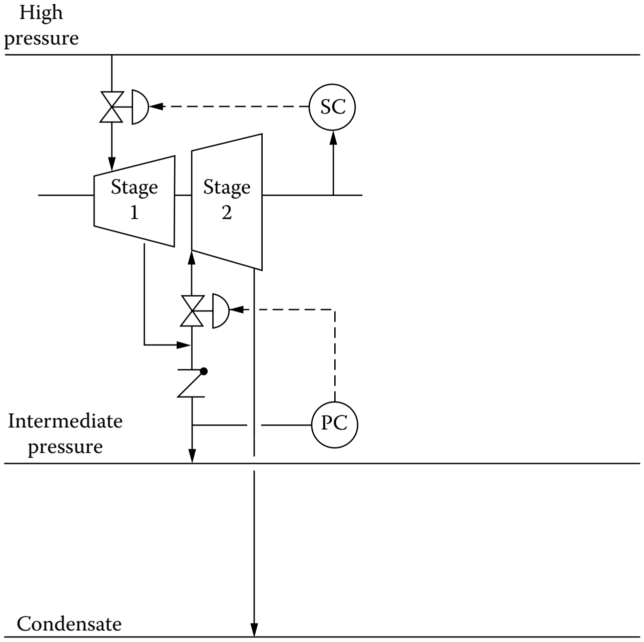


FIG. 8.38e
The extraction turbine generates as much intermediate pressure steam as required by the plant, while taking as much high-pressure steam from its supply header to meet the load by maintaining its shaft speed constant.

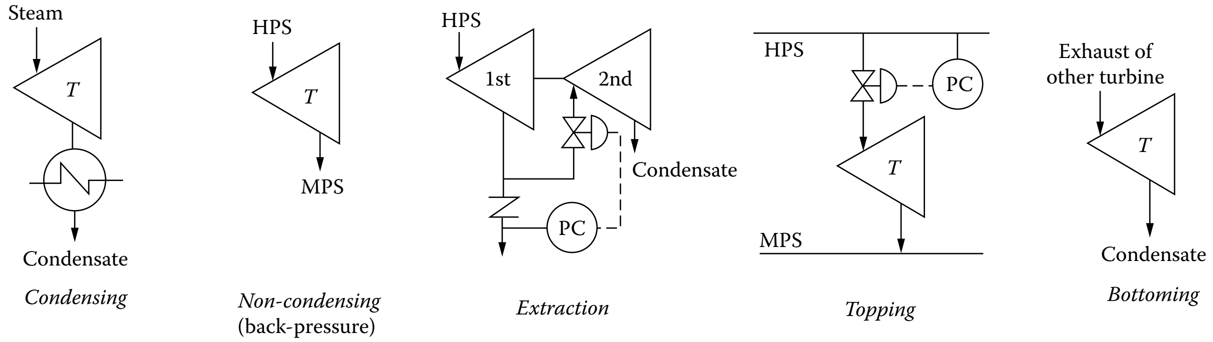


FIG. 8.38f
Terminology used to describe various steam turbine installations.

using steam from the "top" of the plant's thermodynamic cycle.

Similarly, a turbine installed between the noncondensed exhaust of another turbine and the condensate system of the plant is called a *bottoming turbine*. A topping turbine could also be described as a noncondensing or back-pressure unit, and a bottoming turbine as merely a condensing turbine. Figure 8.38f describes the main variations of turbine configurations.

The terms "topping" and "bottoming" as generally used describe turbines that are designed to work over unusual ranges of pressures. For example, a bottoming turbine is unusual in that it is generally designed for operating at a very low inlet pressure, and hence the term "bottoming" is more descriptive than the more generic term of "condensing."

Steam turbines can also be classified according to the purpose of their installation. If a turbine's purpose is to generate electricity, it might be called a "generator drive," while if its purpose is to drive a pump or compressor, it can be called a "mechanical drive."

Internal Design Configurations

From an internal design perspective, the steam turbine is either an "impulse"- or a "reaction"-type design. In the United States, almost all turbine designs are of the axial flow variety, and only a small number are the tangential flow variety. In Europe, a significant number of turbines are the radial flow design.

The steam turbine can also be single-stage or multistage; if multistage and as a function of the number of parallel exhaust stages, one would refer to the turbines as single-flow, double-flow, and so forth.

The casing and shaft arrangement is also an important way of categorizing turbines. In a single casing machine, there is a single casing and one shaft. In a "tandem" design, there are two or more casings connected end to end by a shaft. In "cross-compound" configurations, there are two or more casings connected by multiple shafts.

As can be seen from the above, a turbine can be described on the basis of at least three methods of classification.

STEAM TURBINE GOVERNORS

There are two broad categories of steam turbine controls: "safety systems" and "process systems." Both will be described in the paragraphs that follow. Safety systems are intended to eliminate or, at least, to minimize the possibility of damage to the machine or the hazard to operators. Process control systems serve to control the operation of the machine, so as to follow the load in a stable and efficient manner.

The Governor Valve

One valve, the governor valve, is common to all turbine applications. This is the valve between the main steam supply and the turbine. This valve is the primary means of controlling the unit. When the demand for energy from the turbine is changed, it is the opening of this valve that changes to match the new demand by introducing a new supply of steam energy. The energy supply and demand is matched when the turbine speed is constant.

If the supply valve is too far open, the turbine will run at a speed above that desired. If the valve is too far closed, the turbine will slow down. In essence, the governor valve controls the flow of steam, generally measured in pounds per hour, into the turbine with the assumption that inlet and outlet conditions are constant. When that is the case and the shaft speed is constant, there is a balance between the steam flow and the shaft horsepower.

Some of the steam turbines can only be operated at constant speed because the characteristics of their steam supply valves (the governor valves) are not suited for throttling, because of their quick-opening characteristics. A quick-opening valve plug (Figures 6.19a and 6.19f) is like the plug of a bathtub in which a slight lift results in nearly maximum flow.

By changing that characteristic, a constant-speed turbine can be changed into a variable-speed one. Figure 8.38g illustrates how, by welding into the governor seat a characterizing ring, the initially quick-opening characteristic can be altered and the turbine can be changed to operate at variable speed.

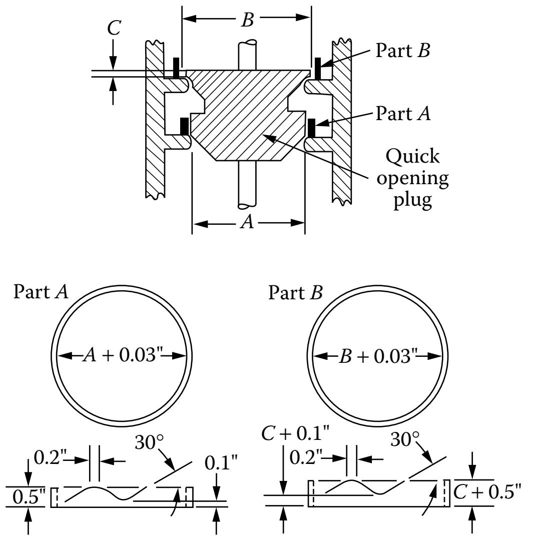


FIG. 8.38g
A double-seated steam governor valve can be rebuilt for optimized variable-speed service. The notched rings provide the necessary rangeability.

The Early Speed Governors

The speed of the turbine is controlled by the "governor." Governors can be mechanical, hydraulic, and electrical. They all include a pilot valve, or a more sophisticated controller, which modulates the turbine's inlet valve in order to keep the shaft speed on set point.

Mechanical governors have been developed from James Watt's original flyball governor shown in Figure 8.38h. The assembly consisted of two weights on the end of short arms, with a hinge in the center to allow vertical motion of each. By gearing to the shaft of the turbine, the assembly was made to rotate. As the shaft speed increased, the weights lifted up toward horizontal.

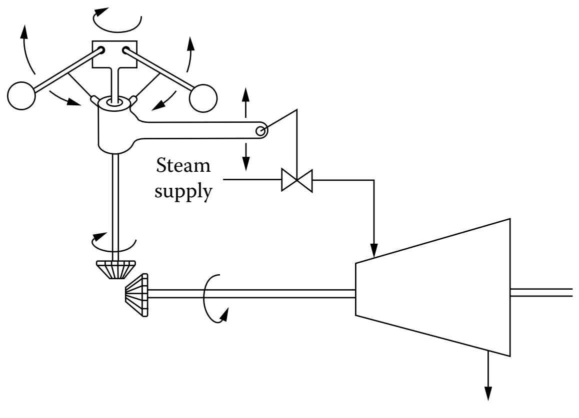


FIG. 8.38h
Schematic representation of flyball governor. As shaft changes speed, rotating fly balls move up and down. Linkage then controls steam supply valve to regulate steam rate. If shaft speed is too fast, balls move up, raising linkage, which then closes down the inlet valve. If shaft speed is too slow, balls drop, lowering collar and opening the inlet valve.

A linkage controlled the throttling valve by admitting less steam as the weights rose and more as they fell. This system was the beginning of automatic machine control and is still used almost unchanged in some modern mechanical governors.

Hydraulic Droop Governors

In hydraulic governors, the shaft speed is generally detected by a flyball, but instead of a direct mechanical linkage between the position of the flyballs and the control valve, a hydraulic system is used to amplify the force generated by the flyball position. The amplification feature improves control sensitivity, because very small changes in flyball position, corresponding to small changes in shaft speed, are sufficient to produce effective control actions. In addition to amplification, the signal can also be characterized as needed for stability.

Two types of governors are distinguished: In the "isochronous" design, the objective is to maintain the speed of shaft rotation constant regardless of load, while in the "droop" design, the speed of the machine is deliberately decreased as load increases (Figure 8.38i). The droop governor is a proportional-only controller, which cannot change its output control signal without first developing an error. This "offset" phenomena was explained in connection with Figure 2.2e in Chapter 2. The terms "offset" and "droop" are interchangeable.

On an increase in speed, the flyballs of a droop governor move out, which raises the stem on the pilot valve. This movement is opposed by the spring. Pressure is applied from the pilot valve to the top of the actuator; the bottom of the actuator drains to the sump through the pilot valve. This decreases the throttle setting. As the linkage moves down, the spring force increases and the force provided by the flyballs is exactly opposed. This moves the pilot valve back to the null position, which maintains the new lower speed.

Droop governors can be an advantage in some applications. For example, if two steam turbines are used to drive two electric generators that are electrically connected in parallel, the droop characteristic will allow the generators to share the total load, whereas isochronous governors would not.

If both generators were to run at exactly the same speed, the division of load between them would depend only on the electrical characteristics of the generators. If they were not precisely identical, which is always the case, the division of load between them would be unequal. If the speeds were not perfectly matched, one might carry all of the load.

A droop governor, however, would cause the one turbine with the heavier load to tend to slow down. When the load carried by the first generator matched that of the other unit and began to surpass it, the first unit would slow down. An

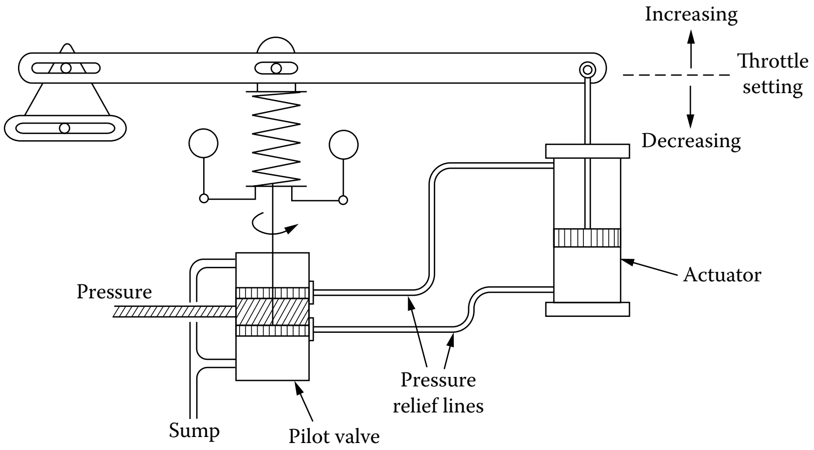


FIG. 8.38i
The design of a droop governor.

equilibrium would be quickly established with each unit carrying a share of the load.

The problems requiring compensation generally include instability that occurs during speed changes. In high-gain control systems, buffering is used to minimize instability. In this case, the governor introduces droop on all speed changes and controls the rate at which the temporary droop characteristic is removed. In this way, speed transitions can smoothly be made. The droop due to buffering can be built into a governor whether the device is a droop type or not.

Electronic Governors

Electronic governors perform the same functions as their mechanical-hydraulic counterparts, but in a somewhat different way. The flyballs are replaced by an electronic tachometer input that is usually generated by a magnetic sensor. The sensor can be triggered when the teeth of a gear connected to the machine's shaft pass by it (Figure 8.38j).

The varying reluctance of the magnetic circuit is used to generate a periodic function with a frequency proportional to the rotational speed of the shaft. The control valve is most often throttled hydraulically, although its actuator can also be electronic or pneumatic. The characteristics of the controller are determined by its transfer function, which usually provides similar performance as that of the earlier designs.

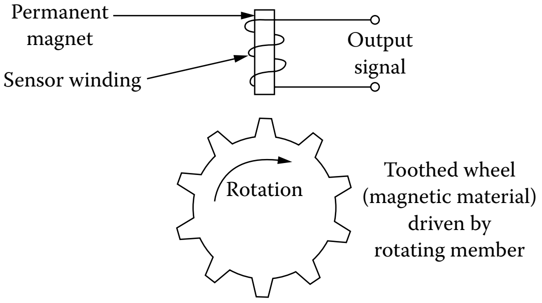


FIG. 8.38j
Schematic of an induction-type speed sensor.

The main difference is that a wider range of features can be built into a single unit, and the same design can be easily adapted to a variety of applications. For example, in addition to speed control, the governor can maintain either the inlet or the exhaust pressure or can control or manipulate other process conditions. Further, it can automatically parallel generator sets, provide overspeed protection, and monitor other machine safety devices so a shutdown can be effected in the event of an unsafe condition. Electronic governors provide versatility rather than improved performance.

Advanced Governors

Advanced control systems perform speed control, load control, steam pressure control, valve testing, remote control, and turbine protection. Normal operation, in addition to speed, load, and steam pressure modulating control, also includes valve testing and remote control (autodispatch, autosynchronizer, and so on). Even the advanced controls usually operate the high-pressure and low-pressure valves through the existing electrohydraulic controls of the turbine.

Further information on steam turbine performance and speed control can be found in API Standard 611, "General-Purpose Steam Turbines for Refinery Services," API Standard 612, "Special-Purpose Steam Turbines for Refinery Services," NEMA Standard SM21, "Multistage Steam Turbines for Mechanical Drive Service" (Table 8.38k), and NEMA Standard SM22, "Single-Stage Turbines for Mechanical Drive Systems." Performance objectives of governors are covered in NEMA Standard SM22-3.13.

Governor systems are classified as A, B, C, or D, depending on performance objectives. Table 8.38k summarizes the basis for these classifications.

TABLE 8.38k
Governor Classification and Performance per NEMA SM21

Class	Speed, Range, %[a]	Maximum Speed Regulation, %[b]	Maximum Speed Variation, ±%[c]	Maximum Speed Rise, %[d]	Trip Speed Setting, %[e]
A	10, 20, 30, 50, 65	10	0.75	13	115
B	10, 20, 30, 50, 65, 80	6	0.50	7	110
C	10, 20, 30, 50, 65, 80	4	0.25	7	110
D	10, 30, 50, 65, 80, 85, 90	0.5	0.25	7	110

[a] Governor may be adjusted to produce any speed within this percentage of rate speed.
[b] Maximum speed regulation from no load to full load.
[c] Maximum speed variation when operation is at constant load.
[d] Maximum overspeed that can occur under any operating conditions.
[e] Proper overspeed trip setting to coordinate with governor maximum speed rise.

CONTROLS AND OPTIMIZATION

A number of control systems will be described here in an order of increasing sophistication. These will include the controls for pressure let-down and extraction turbine controls. Controls systems will also be described for the decoupling of interaction between control loops and for optimization purposes.

The Basic Turbine Controller

The simplest application is one in which a turbine is used to operate a mechanical load at constant speed. Here, steam is supplied from a header and is condensed in the turbine (Figure 8.38l). In this case, the speed controller (governor) senses the shaft speed and manipulates the steam supply valve to keep the speed on set point.

Variations in load, caused by either shaft loading or variations in supply header pressure, affect the balance between the energy supplied to the turbine from the steam system and the work removed from the turbine's shaft. If more energy is available than is being used, the shaft will speed up. The governor will detect this increase in speed and act to eliminate it. Its means of doing so is to reduce the energy supplied to the turbine by closing the supply valve.

If the net change in the energy balance were negative, the shaft would slow down and the governor would respond by opening the supply valve.

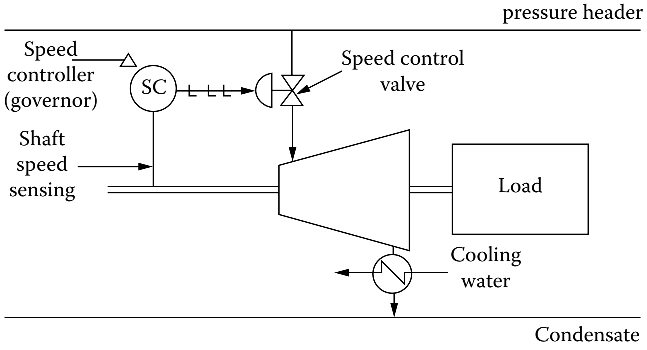


FIG. 8.38l
Simple mechanical drive.

Pressure Let-Down Control

A noncondensing turbine is generally less expensive to buy and to operate than a condensing one, because the energy is extracted from the steam while it has a higher enthalpy, and hence, it has a smaller volume per unit of energy. This has the desirable effect of reducing the size of the turbine and, frequently, also increasing its efficiency.

Because plants usually have a requirement for low-pressure steam for various loads, such as heating, the low-pressure steam generated by noncondensing turbines can often be used to advantage. (Figure 8.33b illustrated the control system to be used if the plant also has a requirement for high-pressure steam, and therefore total HP steam must be shared between the HP steam users and the steam turbine.)

In designing any let-down steam turbine controls, it is important to evaluate the relative amounts of the exhaust steam flow from the turbine and the demand for low pressure steam in the plant and make sure that one has considered all possible combinations.

Optimized Electricity Recovery For example, if the plant's demand for low-pressure steam is variable, it is desirable to send that variable amount of steam through a let-down turbine and to recover its energy content in the form of electricity. On the other hand, one should not send more high-pressure steam to the turbine than the amount of low-pressure steam demanded by the process. These two goals can be converted into two control loops as shown in Figure 8.38m.

In this control configuration, the pressure controller (PC) serves to make sure that all low-pressure steam users in the plant are always satisfied, because if the LP steam pressure

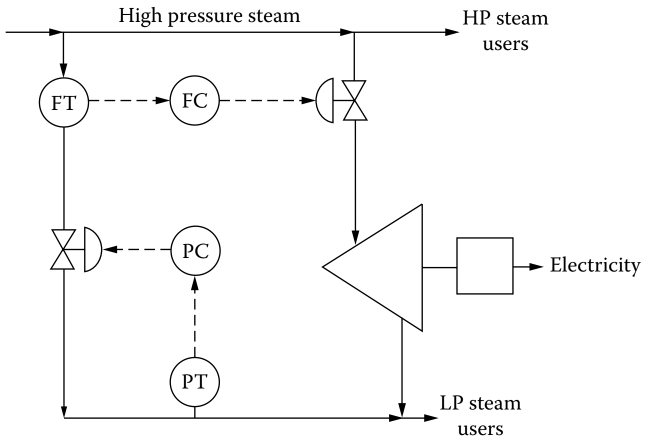


FIG. 8.38m
This control system will follow the variable low-pressure steam demand, while sending most of the HP steam through the turbine to convert its energy content into electricity.

drops, it opens up the turbine bypass to the HP steam header. If the LP steam users cannot tolerate superheating of their steam supply, the bypass has to be provided with a desuperheater.

The task of the flow controller (FC) is to make sure that whatever happens to be the LP steam demand of the plant, it is satisfied mostly by exhaust steam, from which the excess energy has already been recovered in the form of electricity. The flow controller does that by keeping the flow in the bypass at some minimum rate and increasing the HP flow to the turbine as soon as the bypass flow starts to increase.

The control configuration in Figure 8.38m can only be used when the load on the turbine is not determined by the process, but is freely variable, such as in plants where the excess steam energy is utilized for the cogeneration of electricity. While energy conservation dictates that the flow through pressure let-down line be minimized, control dynamics suggest that it should not be completely eliminated. This is because the speed of response of a let-down valve is much faster than that of a turbine. Therefore, the sensitive control of the LP steam pressure is provided by the let-down pressure controller, while the bulk of the steam passes through the turbine and is used to make electricity.

Under the discussion of extraction turbine controls in a later paragraph, Figure 8.38r describes the control configuration required if the LP load of the plant can exceed the full capacity of the turbine or when the demand for LP steam can drop below the steam flow from the let-down section of the turbine.

Valve Position-Based Optimization It was noted in connection with Figure 8.38m that if the low-pressure header is supplied by just throttling the high-pressure steam through a control valve, a considerable amount of energy is lost. It was shown that if the LP steam is supplied by a noncondensing turbine, much less energy is lost. In that configuration, the low-pressure header is supplied preferentially by the steam turbine.

Figure 8.38n illustrates the control system for an installation where the flow controller (FC) sets the turbine's speed controller and the objective is to maintain shaft speed relatively constant. In such a case, the amount of LP steam available from the turbine depends on its load. Therefore, if LP demand exceeds the turbine's ability to supply it, additional steam has to be supplied through the pressure let-down valve in the turbine's bypass. This bypass valve is controlled by the LP header pressure controller (PC).

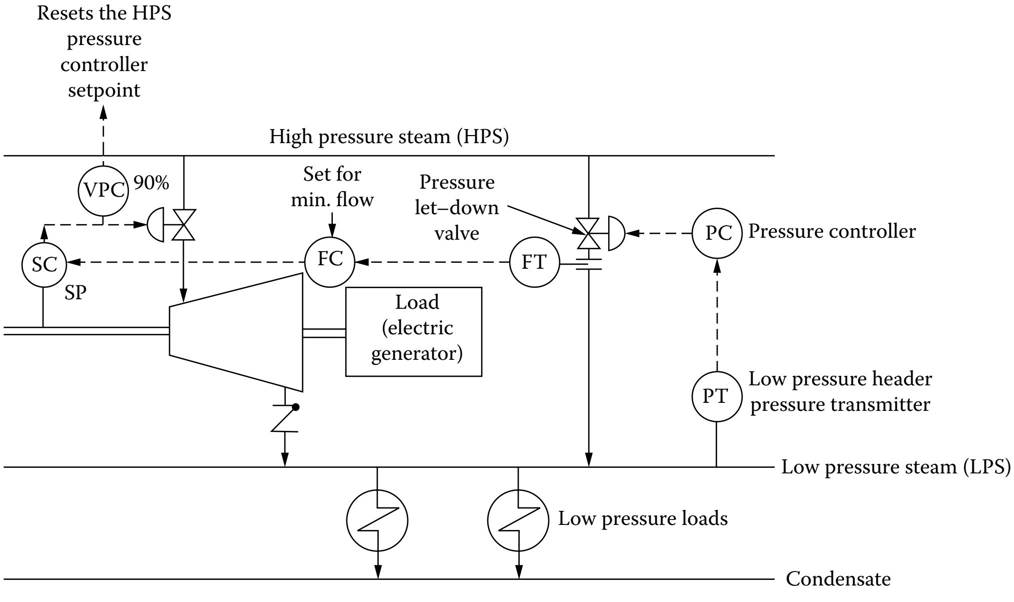


FIG. 8.38n
Back-pressure turbine control system for the generation of LP steam, provided with valve position-based optimizer.

If the only user of high-pressure steam in the plant is the turbine shown in Figure 8.38n, then the HPS pressure controller (not shown) set point can be adjusted to keep the steam governor valve always nearly open (90%). This is done by an integral-only valve position controller (VPC) that reduces the HPS pressure controller set point whenever the governor valve is less than, say, 90% open.

As the HPS supply pressure is reduced, more HP steam will be needed to meet the same electric load on the turbine, and therefore, the governor valve will open. This control strategy, which keeps the governor valve nearly open, is an optimization strategy, because the same load is being met with less pressure drop through the governor valve, and therefore it is being met at a higher efficiency.

The flow controller in the turbine bypass, as was explained in connection with Figure 8.38m, serves to make sure that most of the steam is sent through the turbine. Therefore, whenever excessive amounts of steam pass through the let-down valve, the flow controller (FC) increases the speed set point and thereby increases the amount of steam passing through the turbine.

Extraction Turbine Control

In addition to the governor valve, in extraction turbines, a second "valve" is required. It controls the steam flow rate that is extracted from the first stage of the turbine and is sent to the second stage. The extraction rate can be controlled to keep the pressure of the LP header constant, but it can also be a function of shaft speed, or a combination of the two.

If the turbine incorporates the controls as a built-in feature, the turbine is referred to as an "automatic-extraction" type. Such turbines are generally designed to deliver 100% shaft power and to provide extraction steam only if the load requirements permit. This is the most common type of extraction machine.

Extraction turbines may be visualized as two-stage units from which steam can be removed at a pressure between that of the supply and that of the exhaust. When the demand for work (load) on the turbine is small, the high-pressure stage may be adequate to meet the "work load," and consequently, a large amount of extraction steam may be available to supply the low-pressure header. As work load increases, the second stage becomes necessary to meet the demand for added work and begins to compete for the steam previously being extracted. The control system must allow for this to occur, if meeting the work load is the first priority.

At least a minimal amount of steam must be maintained through the second stage to prevent overheating. This requirement may necessitate limiting extraction, but it can also require the maintaining of a specific second-stage discharge pressure. These requirements are given in the manufacturer's operating specifications.

LPS Demand Exceeding First-Stage Exhaust Figure 8.38o shows an extraction turbine in a pressure let-down application.

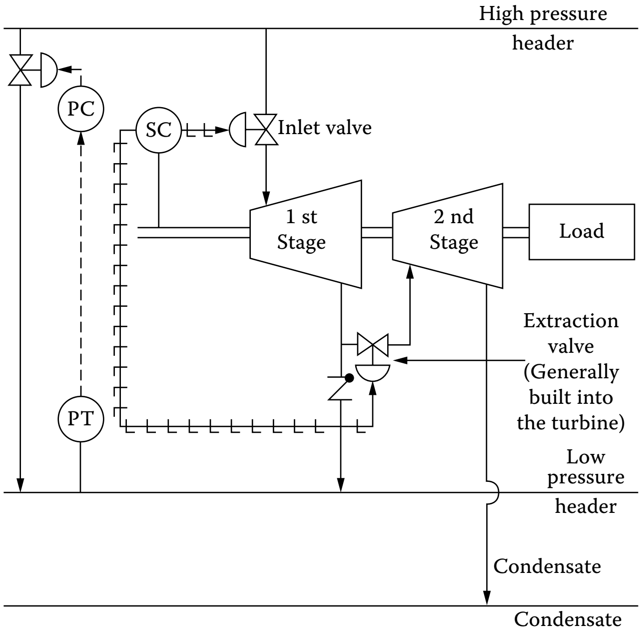


FIG. 8.38o
The controls of an extraction turbine in a pressure let-down application, where the demand for LPS always exceeds the steam available from the interstage of the turbine.

In this example, the exhaust of the first stage is used to supply a low-pressure header, while the second stage is condensing. The speed controller (governor) is arranged by hydraulic or mechanical linkage to close (or nearly close) the extraction valve, if the turbine speed can be maintained by the governor throttling the HP steam supply valve to the first stage (the steam inlet valve).

If speed cannot be maintained by the first stage alone, the extraction valve starts to open, admitting more steam to the second (condensing) stage and, consequently, starving the low-pressure header. When the available extraction steam is insufficient, similarly to the arrangement in the previous figures, a pressure-controlled bypass valve is used to maintain the pressure in the low-pressure header.

LPS Demand Less Than First-Stage Exhaust So far, it has been assumed that the low-pressure header can use all the steam that is available from the turbine. If this is not the case — if the interstage steam supply is in excess of the low-pressure header's requirements — the excess steam must be condensed or vented to protect from overpressuring the LPS header. Figure 8.38p illustrates the controls that will protect against either unnecessary condensing or wasting treated water by venting the steam.

In this control configuration, if the low-pressure header does not need the steam, the HP steam supply flow to the first stage of the turbine is reduced. Because this reduces the energy available from the first stage, the extraction valve is opened to the condensing stage to supply the additional horsepower required to maintain shaft speed. The main

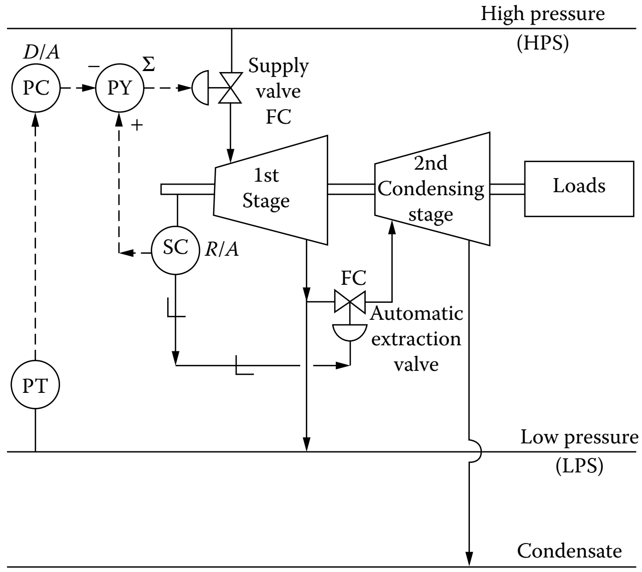


FIG. 8.38p
The controls of an extraction turbine in a pressure let-down application, where the demand for LPS is always less than the steam available from the interstage of the turbine.

difference with the controls in Figure 8.38o is that the pressure controller (PC) that maintains the pressure in the LPS header does not modulate a bypass valve, because there is no bypass. Instead, it reduces the HP steam supply to the turbine if the LPS pressure rises, by sending a negative bias to the summing relay (PY).

The speed controller (SC) in this configuration controls the extraction valve to maintain the shaft speed. The speed controller's output to the HPS supply valve is sent to the positive input of a summing relay (PY). In the absence of a signal from the PC to the negative input of PY, this scheme would operate the same way as the controls in Figure 8.38o.

The pressure controller on the LPS header in Figure 8.38p is direct-acting, so its output will increase when the LPS header pressure rises. This increasing control signal will be subtracted from the speed controller's output and will cause the FC inlet valve to close slightly. This, in turn, will cause the turbine to slow down.

As the shaft speed drops, the speed controller will attempt to open the inlet valve by increasing its output signal, but the change will again be resisted by the pressure controller. As speed falls off, the reverse-acting speed controller will open the extraction valve. As a consequence, the pressure in the LPS header will decrease. This will cause the pressure controller to reduce its output, which in turn will slightly open the HPS supply valve, which will increase the shaft speed, and therefore, the speed controller will close the extraction valve somewhat. Eventually, after much interaction, a new equilibrium will be achieved.

Improving the Control Dynamics The dynamics of the pressure control system in Figure 8.38p is dependent on the dynamics of the turbine's extraction control system. The characteristic of the extraction valve is often not suitable for throttling control and can be improved by use of a characterizing positioner.

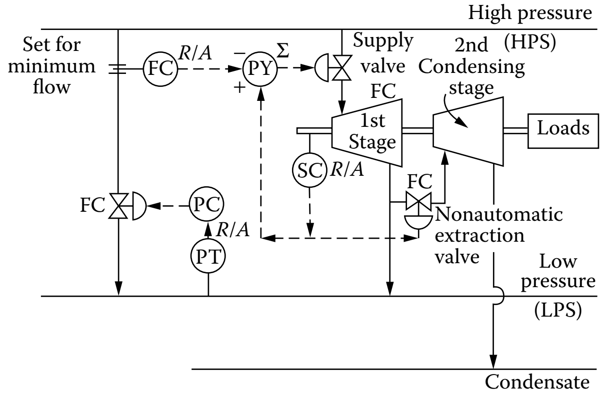


FIG. 8.38q
The addition of a pressure-controlled let-down line increases the speed of response, while the flow controller minimizes the energy waste through that line.

Another method of improving the dynamic response of the control system in Figure 8.38p is to provide a pressure let-down bypass line, as shown in Figure 8.38q. Here, the pressure controller (PC) provides the sensitivity required for quick response, while the flow controller (FC) keeps the flow in the bypass line at a minimum, as it slowly opens the HPS supply valve to the turbine whenever the flow in the bypass line exceeds its set point.

Flexibility by Controller Sequencing If the relative sizes of the work load on the turbine and the user's demand for LPS are unpredictable, the previously described control systems will not work. Figure 8.38r illustrates a control strategy that utilizes controller set-point sequencing to allow optimized and stable operation under any combination of relative load sizes.

If the demand for LPS exceeds the amount of exhaust steam available, the control system shown in Figure 8.38r will operate in a similar manner as did the control system in Figure 8.38o, by the pressure controller (PC) providing the additional steam through the pressure let-down bypass line. The main difference between the two control systems is the addition of the optimizing controller FC-3, which is set at a bypass flow rate slightly exceeding the set point of FC-1.

Therefore, if the pressure controller (PC) opens the bypass and the let-down flow rate exceeds the set point of FC-1, the previously inactive (saturated) FC-3 becomes active and starts cutting back the LPS steam flow to the boiler feed water preheater and, thereby, reduces the plant's demand for LPS. This is an energy-efficient response because the energy recovered from the LPS supplied to the feedwater preheater is less than the energy content of the HP steam that is needed to produce that LPS.

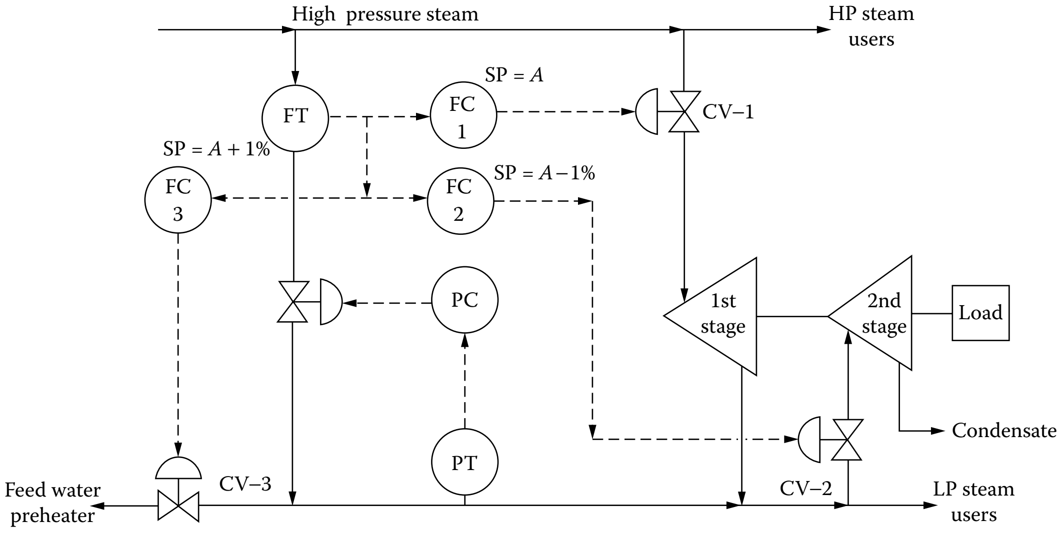


FIG. 8.38r
This control system is both flexible and optimized: FC-1 keeps the bypass flow to a minimum, while FC-3 reduces the LPS demand if it exceeds the work load on the turbine, and FC-3 makes more steam energy available to the turbine if the LPS demand is below the work load.

If the relative loads are reversed and the LPS availability exceeds the demand for LPS, this will cause the pressure in the LPS header to rise and the pressure controller (PC) to reduce down the bypass flow. When this let-down flow drops below the set point of FC-1, the previously inactive (saturated) FC-2 becomes active and admits that part of the LP steam that is not needed in the LPS header into the second, condensing stage of the turbine.

Any number of such bypass flow controllers can be used to sequentially respond to changes in the relative sizes of the work and LP steam loads. These controllers should be provided by integral action only, so that they will be saturated (and their control valves closed) until their set points are reached.

Decoupling the Interaction If the turbine is a nonautomatic extraction type, and therefore one can send a control signal to the extraction valve (as was the case in Figure 8.38r), the interaction between the pressure and turbine speed controllers can be decoupled. In the control system configuration shown in Figure 8.38s, a drop in the speed of the shaft opens both the inlet and the extraction valves, and an increase in shaft speed closes both of them.

In this control system, when the LPS header pressure rises, the pressure controller (PC) output rises, and therefore the PY-1 output drops and the supply valve closes. At the same time, the increase in the PC output increases the output of PY-2, which opens the extraction valve. When the LPS header pressure drops, the opposite is the response: The supply valve opens and the extraction valve closes.

If the weighing of the combining algorithms (PY-1 and PY-2) is correct (if they properly consider features of the valves and the time constants of the loop components), the response to changing pressure conditions, the supply, and extraction valves will complement each other.

Therefore, if the control model is properly tuned and the gains of the summers PY-1 and PY-2 are properly set, there will be no interaction between speed and pressure control,

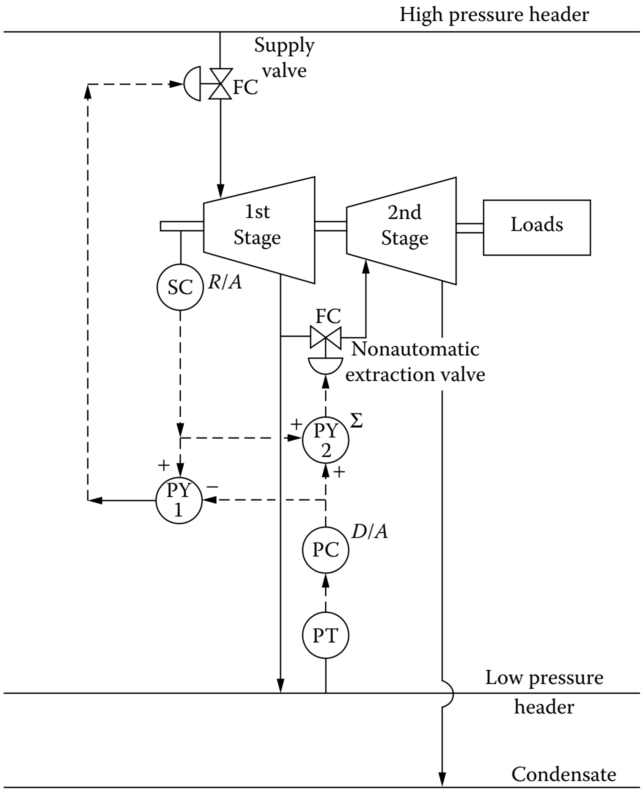


FIG. 8.38s
One way to eliminate interaction between flow and pressure loops is to allow the pressure controller to throttle both the supply and the extraction valves.

and the speed control will not be adversely affected by the responses to pressure disturbances, and vice versa. The speed of the response of this system can also improved by the addition of bypass let-down controls be shown in Figure 8.38q. On the other hand, if the speed of response of the turbine is sufficient to respond to pressure variations, the let-down station (shown in Figure 8.38q) can be eliminated.

An intriguing aspect of this control configuration is the possibility of eliminating the need to throttle steam completely, if the turbine's operating capacity sufficiently matches the low-pressure header's demand for LPS. In that case the supply valve is kept fully open, as was the case in the control configuration shown in Figure 8.38n, and the pressure controls of the LPS header determine the distribution of the extracted steam between the turbine's second stage and the LPS header. Obviously, this configuration is only viable if the size and characteristics of the turbine are properly selected.

SAFETY CONTROLS

The turbine protection system protects the turbine from overspeeding, monitors all critical turbine parameters, and trips the turbine if a condition exists that could cause equipment damage. Thermal stress monitoring performs the calculations needed to determine thermal conditions of the turbine and safe parameters for control operation. It also provides the operator with information on rotor thermal stress (acceleration and load rates, and maximum allowable initial load pick-up).

The main safety control element on a turbine is the steam supply valve. This safety valve can be a separate on/off valve, or the shut-off function can be incorporated into the controls of the steam supply valve that is used for speed control. Taking the turbine off-line is accomplished by closing this valve. Consequently, the safety control system should be so designed as to require that all interlocks be satisfied before this valve is allowed to open.

As shown in Figure 8.38t, the safety interlocks usually include the safety response to lube oil failure, high bearing temperature, overspeed, and vibration. Lube oil is generally monitored by a pressure switch in the case of pressure lubrication systems or by a level switch in nonpressure systems. If the lube oil failure switch is not satisfied, the turbine shutoff valve is closed.

The sudden loss of load will cause the turbine to overspeed. This can happen in mechanical drive applications, but it is a more common occurrence in electrical generator drives. Abnormal electrical conditions in a distribution system can

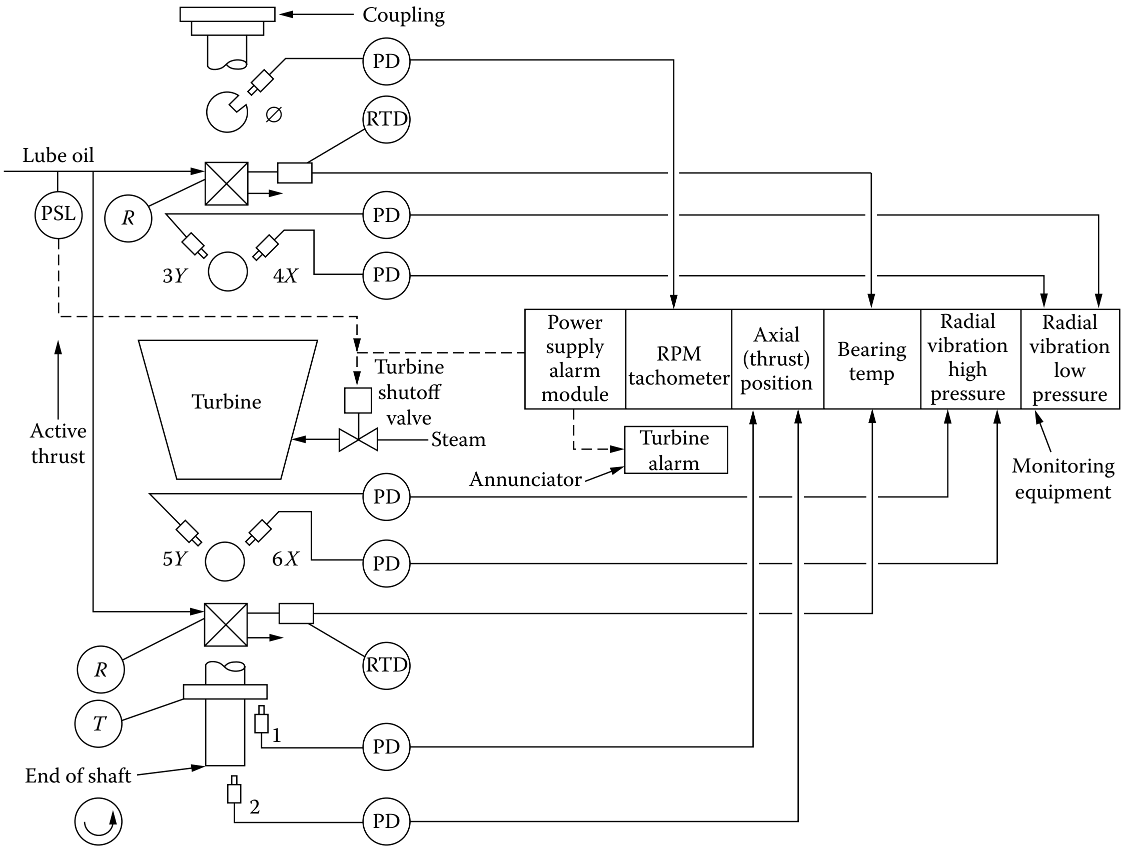


FIG. 8.38t

Turbine "health-monitoring" system. PD: Position detector, proximity sensor. (Adapted from Figure C-1 of American Petroleum Institute Standard 670, Noncontacting Vibration and Axial Position Monitoring System.)

cause protective devices to separate the generator from the system. In that case, a generator that may have been supplying megawatts of power and, consequently, megawatts of load to its prime mover can suddenly disappear, leaving only a small demand for energy that will serve to overcome the friction of its bearing and the windage on its moving parts.

When this occurs, if the control system is not fast enough to reduce the steam supply to the turbine, it will overspeed. This condition can be detected by a variety of devices, such as centrifugal switches, electronic tachometers, or strain-detecting devices installed on or near such components of the machine that are affected by the overspeed condition.

Bearing temperature is an additional indication if lubrication is functioning properly. It is also a way of detecting the deterioration in bearings before complete mechanical failure. Mechanical wear accelerates as a result of improper lubrication or because of mechanical stresses that cause deformation of the bearing's component parts. In either case, the bearing begins to dissipate abnormally large amounts of energy, which in turn results in heating.

Consequently, a sudden rise in bearing temperature is generally an indication of incipient failure. It is important to quickly stop the machine when a potential bearing failure is detected. This is because some turbine designs maintain very small clearances between stationary and rotating parts. If the bearing deforms, it may mean the total destruction of the machine.

Especially on larger machines, a stationary vibration-monitoring system is usually installed. Such a system generally consists of accelerometers or proximity sensors located radially in each bearing and axially on the end of the shaft or the thrust collar. Two sensors positioned at right angles are typically used in bearings.

Electronic monitoring equipment is used to measure the acceleration or displacement that occurs at each monitoring point. The monitoring equipment generally incorporates an "alarm" setting, which is intended to warn an operator of an impending problem, and a "danger" setting, which is intended to shut the machine down. In many instances, a vibration-initiated shutdown can prevent major damage in situations in which, without prompt action, equipment could be lost.

Sequential Controls

In power generating installations, the monitoring and sequential controls serve to automatically bring the turbine from turning gear to generator grid synchronization. These controls evaluate such parameters as bearing temperatures and vibration, water detection, and differential expansion. The control package advises the operator on current turbine status and provides recommended actions.

Vibration monitoring equipment is covered extensively in Section 7.22 in Chapter 7 of the first volume of this handbook.

CONCLUSIONS

Steam turbines are versatile energy conversion devices that, in addition to powering a variety of mechanical loads, can do an excellent job of extracting energy that might otherwise be wasted from a plant's thermodynamic cycle. They provide opportunities for process improvements and energy savings. Therefore, their application should be carefully and insightfully considered. The controls of the high-pressure and low-pressure steam admission valves most often are implemented through the existing (furnished electrohydraulic) turbine controls.

Advanced controls include the features of redundant control, on-line tuning, field-proven hardware, and remote operator displays, including custom graphics, report generation, and on-line, systemwide integration. There usually are four separate and redundant control packages, which perform operator automatic control (OAC), thermal stress monitoring, sequential control, and turbine protection functions.

As has been discussed in this section, OAC control includes speed, load, and steam pressure modulating control, as well as valve testing and remote control operation (autodispatch, autosynchronizer, and process interface).

Bibliography

Adamski, R. S., "Improved Reliability of Rotating Machinery," *InTech,* February 1983.

Anderson, P. M. and Fouad, A. A., *Power System Control and Stability*, New York: Wiley-IEEE Press, 2002.

API Standard 611, "General-Purpose Steam Turbines for Refinery Services," Washington, D.C.: The American Petroleum Institute, 1997.

API Standard 612, "Special-Purpose Steam Turbines for Refinery Services," Washington, D.C.: The American Petroleum Institute.

API Standard 670, "Noncontacting Vibration and Axial Position Monitoring System," 1st edition, Washington, D.C.: The American Petroleum Institute, 2000.

Bachmann, R., et al., *Combined-Cycle Gas and Steam Turbine Power Plants*, 2nd Edition, Tulsa, OK: PennWell, 1999.

Bentley, D. E., *Machinery Protection Systems for Various Types of Rotating Equipment*, Minden, NV: Bentley Nevada Corp., 1980.

Bloch, H. P., *A Practical Guide to Steam Turbine Technology*, New York: McGraw-Hill, 1995.

Boyce, M.P., *Handbook for Cogeneration and Combined Cycle Power Plants*, ASME, 2001.

Champigny, R. A., "Retrofitting Turbine and Boiler Controls," ISA/93 Technical Conference, Chicago, IL, September 19–24, 1993.

Feuell, J., "Single-Stage Steam Turbine-Generator Set Replaces Pressure Reducing Station to Reduce Plant Energy Costs," *Turbomachinery International,* January–February 1980.

Feuell, J., "Steam Turbine Induction Generator Set," *Turbomachinery International,* May–June 1982.

Gires, T. C., and Birnbaum, M., "Digital Control for Large Steam Turbines," *Proceedings of the American Power Conference,* Chicago, IL, April 1968.

L&K International Training, Institute of Electrical & Electronics Engineering, 1999.

NEMA Standard No. SM21, "Multistate Steam Turbines for Mechanical Drive Service."

NEMA Standard No. SM22, "Single-Stage Turbines for Mechanical Drive Service."

Oetkin, A., “Steam Turbine Applications of 2301 Control Systems,” Ft. Collins, CO: The Woodward Governor Co.

Osborne, R. L., “Controlling Central Station Steam Turbine Generators,” *Instruments and Control Systems,* November 1975.

Podolsky, L. B., “A Feedforward System for Digital Electrohydraulic Turbine Control,” 1973 ISA Conference, Reprint No. 73–661, 1973.

Podolsky, L. B., Osborne, R. L., and Heiser, R. S., “Digital Electrohydraulic Control for Large Steam Turbines,” *Proceedings of the 14th International ISA Power Instrumentation Symposium,* New York, May 1971.

Shapiro, L. J., “Condensing Turbine Can Improve Economics of Cogeneration,” *Power Magazine,* Vol. 126, No. 8, pp. 73–74, August 1982.

Still, U., and Zorner, W., *Steam Turbine Governors, Process Control,* John Wiley & Sons, 1996.

The Woodward Governor Co., “Analytic Representation of Mechanical-Hydraulic and Electrohydraulic Governors,” Bulletin 25067, Ft. Collins, CO: The Woodward Governor Co.

The Woodward Governor Co., “43027 Electric Control for Steam Turbine Applications,” Ft. Collins, CO: The Woodward Governor Co.

8.39 Wastewater Treatment Controls

A. J. BAKER (1970, 1985) **B. G. LIPTÁK** (1995, 2005) **S. E. GAERTNER** (2005)

INTRODUCTION

This section concentrates on the description of the controls used in industrial wastewater treatment. It also covers, but in less detail, the controls used in municipal wastewater and cooling water treatment processes. The controls of water supply plants, including water softening and purification, are covered in Section 8.40.

In the United States, the operations of all water treatment plants are directed by the Federal Water Pollution Act. Together with its amendments, it is known as the Clean Water Act (CWA). Water treatment plants can be privately or publicly owned. Private treatment plants can discharge into publicly owned treatment works (POTWs) if they have the proper permits.

The goal of the water treatment plants is to remove organic matter and to reduce the water's nitrogen content to less than 20 mg/l in all forms. The water quality targets for the treated plant discharge usually include BOD concentration under 5 mg/l, suspended solids concentration under 3 mg/l, PO_4 under 0.05 mg/l, and NO_3 under 10 mg/l. If the plant discharge is for golf course use, the allowable BOD concentration is under 10 mg/l and the suspended solids concentration is under 12 mg/l.

The coverage of this section overlaps with a number of other sections in this chapter. These are:

Section 8.1, Aeration and DO Controls, Section 8.31, ORP Controls, and Section 8.32, pH Control. It is recommended to the reader that for an in-depth coverage of the control strategies available for dissolved oxygen (DO), oxidation-reduction potential (ORP), and pH control and optimization, they also refer to the above sections.

Lastly, the reader is reminded that the various sensors and analyzers used in the wastewater treatment processes are all covered in the first volume of this handbook. Their section numbers are as follows:

8.2	Analyzer Sampling, Process Samples
8.6	Biometers
8.7	BOD, COD, and TOD Sensors
8.11	Chlorine
8.15	Colorimeters
8.23	Fiber-Optic Probes
8.28	Ion-Selective Electrodes
8.36	Nitrate, Ammonia, and Total Nitrogen
8.38	Odor Detection
8.39	Oil in or on Water
8.41	Oxidation-Reduction Potential (ORP)
8.43	Oxygen in Liquids (Dissolved Oxygen)
8.45	Ozone in Water
8.48	pH Measurement
8.49	Phosphorus Analyzer
8.54	Streaming Current or Particle Charge Analyzer
8.58	Total Carbon Analyzers
8.60	Turbidity, Sludge, and Suspended Solids
8.65	Water Quality Monitoring
8.66	Wet Chemistry and Auto-Titrator Analyzers

General Considerations

Some of the processes used in the treatment of wastewater are not well suited to automatic process control. This is because many continuous analyzers are not reliable enough for the control of such processes as coagulation and flocculation of water or the biological treatment of wastewater. These processes are, for the most part, an art, not a science, still requiring some human judgment to determine chemical application rates and process control parameters.

The measurements that are reliable and in wide use are pH, oxidation-reduction potential, residual chlorine, and flow rate. Other water properties that are measured to assist the operator in controlling the system are conductivity, alkalinity, temperature, suspended solids, dissolved oxygen, and color. Figure 8.39a illustrates the design of the sampling system used when measurements obtained by inserting the sensors directly into the process are not reliable.

Water and wastewater treatment consists of unit operations that may be classed as mechanical, chemical, biological, and any combinations of these. The mechanical operations most often consist of screening, filtration, and separation by gravity.

Industrial Wastewater Treatment

Environmental regulations restrict the discharges of chlorine, heavy metals, arsenic, cyanides, biological pollutants, and excessively acidic or basic water. These regulations are

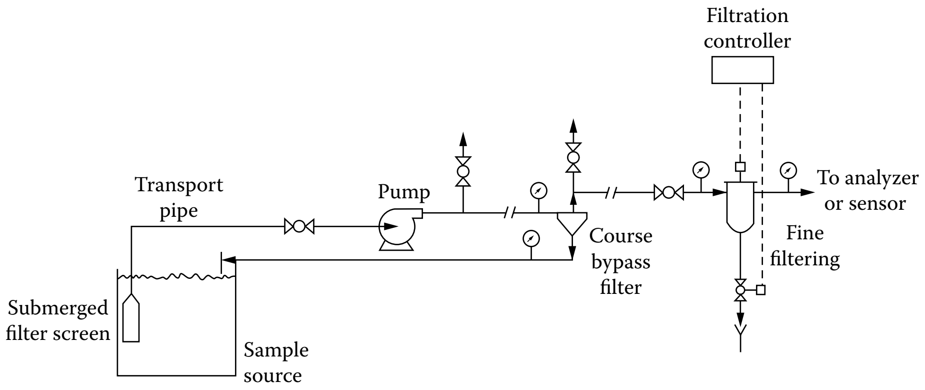


FIG. 8.39a
Schematic of a typical sample transport system.

included in the Clean Water Act and its amendments, including the effluent limitations and pretreatment standards for centralized waste treatment (CWT) facilities.

The majority of industrial wastewater treatment processes are continuous rather than batch-type operations. In the discussion that follows, the control of the following types of water treatment processes will be described: (1) chemical oxidation, (2) chemical reduction, (3) neutralization, (4) precipitation/flocculation/filtration, and (5) biological control by chlorination.

CHEMICAL OXIDATION

Wastewater is treated by chemical oxidation when the contaminant can be destroyed, its chemical properties altered, or its physical form changed. Examples of chemicals that can be destroyed are cyanides and phenol. Sulfides can be oxidized to sulfates, thus changing their characteristics completely. Iron and manganese can be oxidized from the soluble ferrous or manganous state to the insoluble ferric or manganic state, respectively, permitting their removal by sedimentation.

Strong oxidants, such as chlorine, chlorine dioxide, ozone, and potassium permanganate, are used. Chlorine is preferred when it can be used because it is the least expensive and is readily available. Ozone is a strong second choice and is now favored because its excess converts to oxygen, while the excess chlorine could react with industrial waste to produce cancer-causing substances.

The time required for these reactions to proceed to completion is usually pH-dependent. Most often, either residual oxidant or ORP measurement is used to control these processes.

Cyanide Destruction Process

Oxidation-reduction implies a reversible reaction. For a detailed discussion of ORP-based controls, refer to Section 8.31. Because these reactions are carried to completion and are not reversible, the term is misleading. In practice, control is by what may be called "electrode potential readings." An illustration is the oxidation of cyanide into cyanate with chlorine, according to the following reaction:

$$\begin{array}{ll} \mathrm{2Cl_2} & \text{Chlorine} \\ + & \\ \mathrm{4NaOH} & \text{Sodium hydroxide} \\ + & \\ \mathrm{2NaCN} & \text{Sodium cyanide} \\ \downarrow & \\ \mathrm{2NaCNO} & \text{Sodium cyanate} \\ + & \\ \mathrm{4NaCl} & \text{Sodium chloride} \\ + & \\ \mathrm{2H_2O} & \text{Water} \end{array} \qquad \mathbf{8.39(1)}$$

The electrode potential of the cyanide waste solution will be on the order of –200 to – 400 mV. After sufficient chlorine has been applied to complete the reaction described in Equation 8.39(1), the electrode potential will be on the order of +300 to +450 mV. The potential value will not increase until *all* cyanide has been oxidized. Control of pH is essential, with the minimum being about 8.5. The reaction rate increases as the pH rises.

Complete oxidation (the complete destruction of cyanide) is a two-step reaction. The first step is oxidation to the cyanate level described in Equation 8.39(1). The endpoint of

TABLE 8.39b
Set Points and Operating Parameters of the Two-Stage Cyanide Destruction Process

	Process Stages	
Parameters and Set Points	*First Stage: Cyanide to Cyanate*	*Second Stage: Destruction of Cyanate*
pH set point	10–12	8.5–9.5
Reaction time (minutes)	5	45
ORP (mv) set point	+300 to +450	+600 to +750
Maximum concentration of cyanide (cyanate) that can be treated	1000 mg/l	1000 mg/l

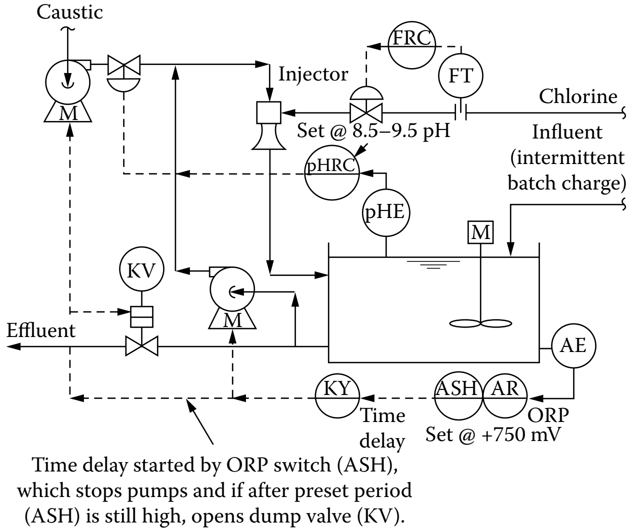
the second-stage reaction is at an electrode potential of about +600 to +750 mV. The overall reaction is

$5Cl_2$ Chlorine
+
$10NaOH$ Sodium hydroxide
+
$2NaCN$ Sodium cyanide
$\downarrow$ **8.39(2)**
$2NaHCO_3$ Sodium bicarbonate
+
N_2 Nitrogen
+
$4H_2O$ Water

Cyanide destruction is the only chemical oxidation reactions that takes place in two steps, as described in Table 8.39b.

Batch Cyanide Control Because of the complexity of a two-stage process, the toxicity of cyanide, and the rigid requirements on waste discharge, a batch-type treatment is often recommended. Such a batch type control system has already been described in Figure 8.31i, and a slightly different one is shown in Figure 8.39c.

In Figure 8.29a, chlorine is charged at a constant flow rate (FRC), while caustic is being added under pH control to maintain the batch pH at about 8.5–9.5. When the ORP set point of +750 mV is reached, the high ORP switch (ASH) actuates a delay timer (KY) and the chemical feed systems are shut down.



Time delay started by ORP switch (ASH), which stops pumps and if after preset period (ASH) is still high, opens dump valve (KV).

FIG. 8.39c
Batch oxidation of cyanide waste with chlorine.

If, after a 30-min delay, the ORP of the tank contents is still at or above the +750 mV set point, the batch is discharged (by opening KV). If further reaction has been taking place during the delay period and the ORP potential value has dropped below the set point, the system is reactivated and the cycle is repeated. When this batch approach is used, additional, usually duplicate, storage tanks are required to receive the incoming waste while the batch tank is used for treatment.

Continuous Cyanide Control As was shown in Figure 8.31g, the continuous flow-through cyanide destruction control systems have the advantage of requiring less space, but this advantage is often offset by the capital cost of the additional equipment that is required.

In the control system configuration shown in Figure 8.39d, the two steps in the cyanide destruction process are separated. In the first step, the ORP controller set point is approximately +300 mV, and this ARC controls the addition of chlorine to oxidize the cyanide into cyanate. The pH in the first-stage tank is controlled by the pHRC at approximately 10 by throttling the caustic flow into the recycled effluent into the injector.

The reaction time of this first-stage process is on the order of 5 minutes. The set point for the chlorine flow controller (FRC) that charges the chlorine into the second-stage tank is ratioed (FY) to the chlorine flow into the first stage. The caustic requirement of the second stage is dependent solely on the chlorine flow rate (pH control is not necessary), and therefore the same chlorine flow signal can be ratioed (FY) to generate the set point for the caustic feed flow controller (FRC).

The ORP detector (AE) is provided to signal process failure and to actuate alarms or initiate emergency actions, if the potential level drops below approximately +750 mV.

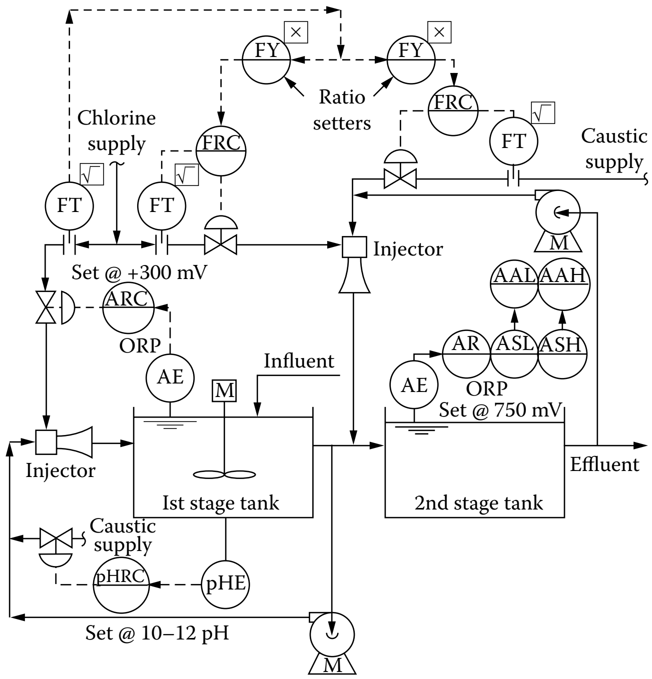


FIG. 8.39d
Continuous oxidation of cyanide waste with chlorine. Influent here has continuous constant flow rate and variable quality.

One might note that in the control system described in Figure 8.31g, an ORP controller (ORPIC) was used to charge the chlorine into the second stage. It is debatable if that is necessary, because the ratioing accuracy available between the first-stage chlorine flow rate and secondary addition (which is approximately 1:1) is, in most cases, sufficiently high. The probability is that the system shown in Figure 8.39d might apply a little more chlorine than is actually required.

In the cyanide destruction process, fixed flow rates are preferred, because they will provide constant reaction times. Residual chlorine analyzers are seldom used in this process, because the metal ions usually present in the waste interfere with their accurate operation. They are used in processes in which the presence of excess residual chlorine indicates a completed reaction. In such processes the set point is usually 1 mg/l or less.

Chlorinator, Sulfonator, and other Controls Figure 8.39e describes the main components of the feeders that can be used in chemical oxidation operations to charge chlorine, sulfur dioxide, or carbon dioxide.

The chlorinator shown in the figure has two operators, two control valves that can be throttled. One is operated in a feedforward mode (PVC) and responds to changes in the influent flow rate. The other control valve (FV) is throttled in a feedback configuration by the effluent quality controller (ARC), which can be detecting ORP, pH, residual chlorine, or other related indicators.

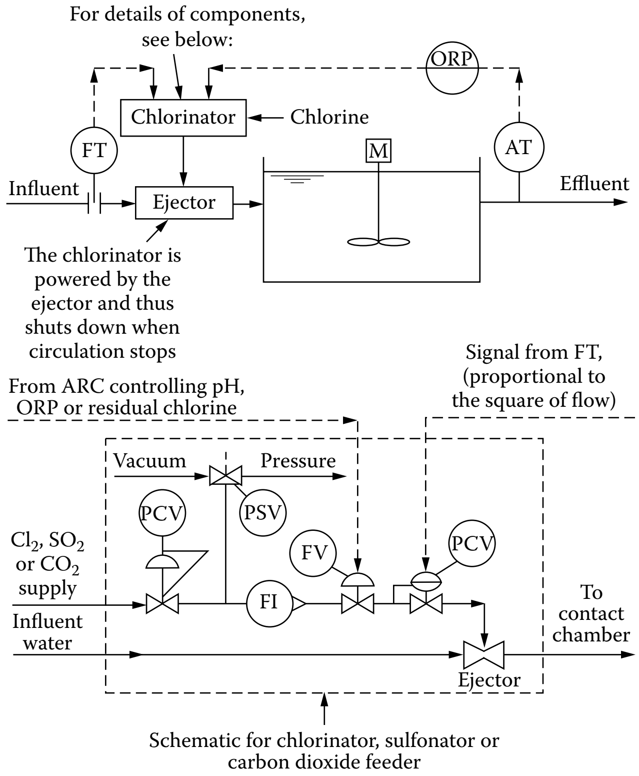


FIG. 8.39e
The lower part of this figure shows the main components of a gas feeder, such as a chlorinator. The top part of the figure shows how a chlorinator is integrated into a chemical oxidation control system.

Most reactions are completed within 5 min, and except for cyanide treatment, most all other chemical oxidation operations are carried out simultaneously with other unit operations, such as coagulation and precipitation, which govern the pH value. Thus, whereas the pH value affects the rate of reaction, it is seldom controlled solely to serve the oxidation process.

Cyanide Destruction by Ozonation The oxidation of cyanide to cyanate by ozonation is extremely fast and is carried out at a pH of 9 to 10.

$$CN^- + O_3 \rightarrow CNO^- + O_2 \qquad \mathbf{8.39(3)}$$

The further reaction to cyanates is much slower and can be accelerated by the addition of copper (2+) salt catalysts. Figure 8.39f illustrates how cyanide oxidation by ozone can be controlled, using either a stirred reactor or a packed or sieve plate tower.

Chlorination is a better developed and more frequently utilized process than ozonation. On the other hand, ozone is faster and more powerful as an oxidizing agent and requires smaller holding and reaction tanks. Chlorine gas is also

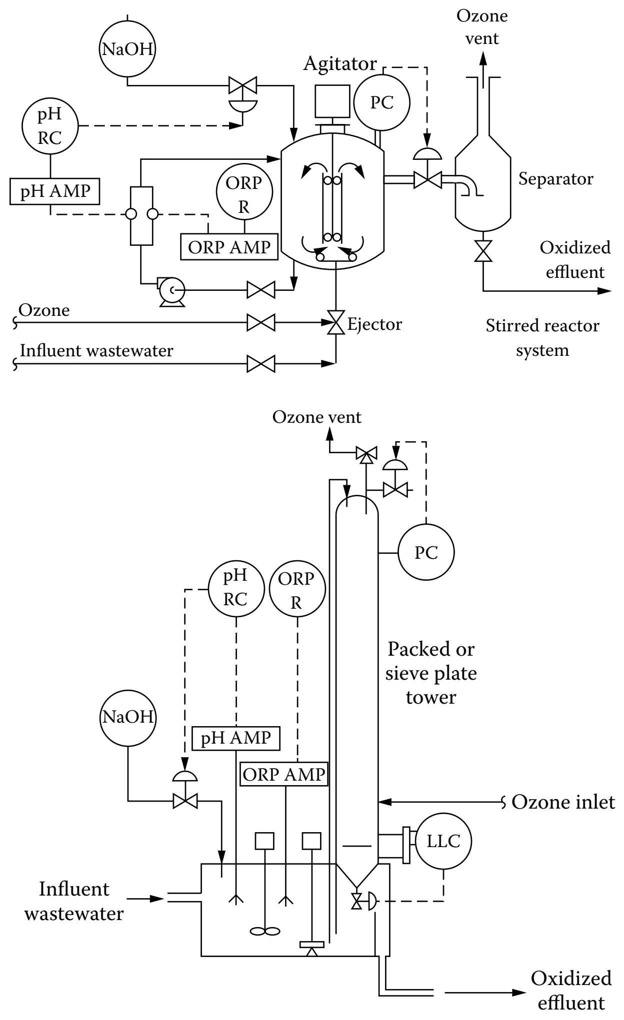


FIG. 8.39f

Control systems for the oxidation of cyanide by ozone, utilizing a stirred reactor (top) or a packed or sieve plate tower (bottom).

hazardous. In order to remove 1 lb of cyanide, one usually requires 2.7 to 6.8 lb of chlorine gas or 1.8 to 4.6 lb of ozone.

CHEMICAL REDUCTION

Wastewater treatment by chemical reduction is quite similar to chemical oxidation. Commonly used reductants are sulfur dioxide and its sodium salts, such as sulfite, bisulfite, and metabisulfite. Ferrous iron salts are infrequently used. Typical examples are reduction of hexavalent chromium, dechlorination, and deoxygenation. Table 8.39g lists some of the reduction and precipitation reactions that take place in the process of chrome treatment.

TABLE 8.39g

Chrome Reduction and Precipitation Reactions

Reducing Agent	*Reaction*
Ferrous sulfate ($FeSO_4$)	$2H_2CrO_4 + 6FeSO_4 + 6H_2SO_4 \rightarrow$ $Cr_2(SO_4)_3 + 3Fe_2(SO_4)_3 + 8H_2O$ $Cr_2(SO_4)_3 + 3Ca(OH)_2 \rightarrow$ $2Cr(OH)_3 + 3CaSO_4$
Sodium metabisulfite ($Na_2S_2O_5$)	$Na_2S_2O_5 + H_2O \rightarrow 2NaHSO_3$ $2H_2CrO_4 + 3NaHSO_3 + 3H_2SO_4 \rightarrow$ $Cr_2(SO_4)_3 + 3NaHSO_4 + 5H_2O$ $Cr_2(SO_4)_3 + 3Ca(OH)_2 \rightarrow$ $2Cr(OH)_3 + 3CaSO_4$
Sulfur dioxide (SO_2)	$SO_2 + H_2O \rightarrow H_2SO_3$ $2H_2CrO_4 + 3H_2SO_3 \rightarrow$ $Cr_2(SO_4)_3 + 5H_2O$ $Cr_2(SO_4)_3 + 3Ca(OH)_2 \rightarrow$ $2Cr(OH)_3 + 3CaSO_4$

Reduction of Hexavalent Chromium

The ORP-based controls of the reduction of highly toxic hexavalent chromium into the innocuous trivalent form has already been discussed in some detail in Section 8.31. If the reducing agent is sulfur dioxide, the following reaction takes place:

$$\begin{array}{ll} 3SO_2 & \text{Sulfur dioxide} \\ + & \\ 2H_2CrO_4 & \text{Chromic acid} \\ \downarrow & \\ Cr_2(SO_4)_3 & \text{Chromic sulfate} \\ + & \\ 2H_2O & \text{Water} \end{array} \qquad \textbf{8.39(4)}$$

Most hexavalent chrome wastes are acidic, but because the rate of reaction is much faster at very low pH values, pH control is essential. Sulfuric acid is preferred because it is cheaper than other mineral acids. The set point of the pH controller is approximately 2.

Similarly to the treatment of cyanide, the chrome reduction reaction is not reversible, either. As shown in Figure 8.39h, the control of sulfur dioxide addition is by electrode potential level, using ORP instrumentation. The potential level of hexavalent chromium is +700 to +1000 mV, whereas that of the reduced trivalent chrome is +200 to +400 mV. The set point on the ORP controller is approximately +300 mV.

The illustrated batch treatment of chrome is controlled at a pH of 2 by feedback control of the acid addition (pHRC), while the sulfur dioxide addition is under ORP control (ARC), which feedback loop is set at 300 mV The reaction time is about 10 min at a pH of 2, and it drops to 5 min if the pH is reduced to 1.5. The control system used in continuous chrome treatment is shown in Figure 8.31d.

The trivalent chromic sulfate is removed from solution by subsequent raising of the pH to 8, at which point it will precipitate as chromic hydroxide (see Figure 8.31d). The control system for this step can be identical with the one used in Figure 8.39v.

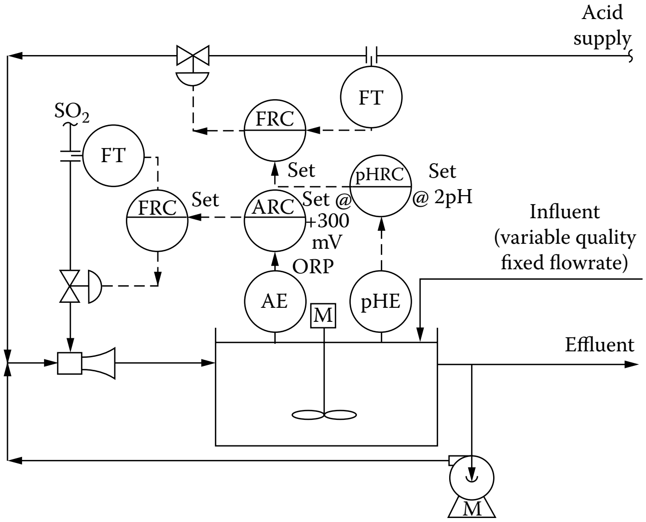


FIG. 8.39h
Batch reduction of chromium waste using sulfur oxide as the reducing agent.

Other Reduction Processes

In dechlorination or deoxygenation controls, the reducing agent is usually added in proportion to the oxidant concentration but by maintaining a slight excess. In most cases, a slight excess of reducing agent is not detrimental. The pH value is not critical and can be determined by corrosion control considerations.

Dechlorination to a fixed residual value is controlled, as illustrated in Figure 8.39e, except that sulfur dioxide is used instead of chlorine.

NEUTRALIZATION CONTROLS

For a detailed discussion of pH control, please refer to Section 8.32. It should be emphasized that pH is a measure of hydrogen ion activity and not acid concentration. A weak sulfuric acid solution will have a low pH value because of the high degree of hydrogen ion activity (disassociation), whereas some strong organic acids may show a pH value as high as 3 or 4.

Strong alkalis react quickly and efficiently to neutralize strong acids. The simplicity of this fact is misleading in its consequences. Acid neutralization is a common requirement in wastewater treatment, but few operations can be as complex. The information needed for the proper design of neutralization systems includes (1) flow rate and range of flow variations; (2) titratable acid content and variations in acid concentration; (3) rate of reaction; and (4) discharge requirements for suspended solids, dissolved solids, and pH range.

Figure 8.39i shows the pH values corresponding to the mixtures of a strong acid and a strong base. The slope of this pH curve near neutrality (pH = 7) is so great that there is no likelihood of controlling such a system. Fortunately, plant effluents usually contain weak acids or bases that are neutralized by strong reagents. (The dotted pH curve shows that these are much easier processes to control.)

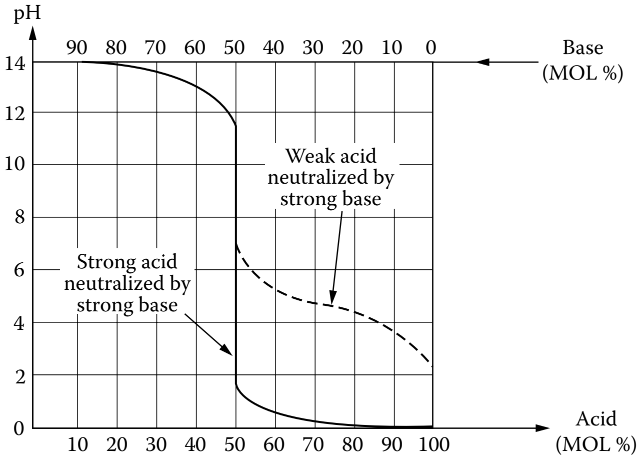


FIG. 8.39i
pH curve of strong acid neutralized by strong base (solid line) and of weak acid neutralized by strong base (dashed line).

The slope of the pH curve is affected by the ionization constants of the acid and base involved and by buffering. Buffering compounds are those that contain no hydrogen or hydroxyl ions but are capable of suppressing the release of these ions from other solutes and, thereby, affect the solution acidity or alkalinity.

Neutralization control of wastewater is difficult because their acid or base contents can vary by several decades and because as the type or amount of buffering of acid (or base) varies, it changes the applicable pH curve. In some plants, the flow rate of the wastewater is also highly variable and the effluent itself can change from acidic to basic, in which case two reagents are required.

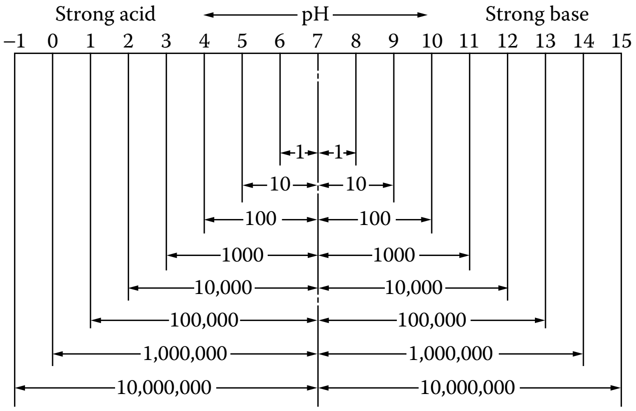
Equalization Tanks

An equalizing basin should be installed ahead of the neutralizing system whenever possible. This will tend to level out fluctuations in the wastewater flow and concentration. This point cannot be overemphasized, because the lack of such a basin has been the cause of many failures. Pumping from an equalizing basin at a constant rate eliminates the need for high rangeability flow rate instrumentation. This, in combination with reduced variations in base or acid content, reduces the reagent feed range requirements.

The obvious disadvantage is in the capital cost for large basins. Most systems are designed with as large an equalization tank as possible according to the available space. Any equalization that can be installed will pay some dividends.

In sizing the reagent equipment, the maximum capacity of the alkali feed system should be determined on the basis of the maximum acidic wastewater concentration and

	pH	Hydrogen ion activity gr mols/liter	Hydroxyl ion activity gr mols/liter
Acidic	0	1	0.00000000000001
	1	0.1	0.0000000000001
	2	0.01	0.000000000001
	3	0.001	0.00000000001
	4	0.0001	0.0000000001
	5	0.00001	0.000000001
	6	0.000001	0.00000001
Neutral	7	0.0000001	0.0000001
Basic	8	0.00000001	0.000001
	9	0.000000001	0.00001
	10	0.0000000001	0.0001
	11	0.00000000001	0.001
	12	0.000000000001	0.01
	13	0.0000000000001	0.1
	14	0.00000000000001	1



Graph of reagent demand. Reagent addition units are 10^{-6} mols/liter.

FIG. 8.39j
The logarithmic nature of pH.

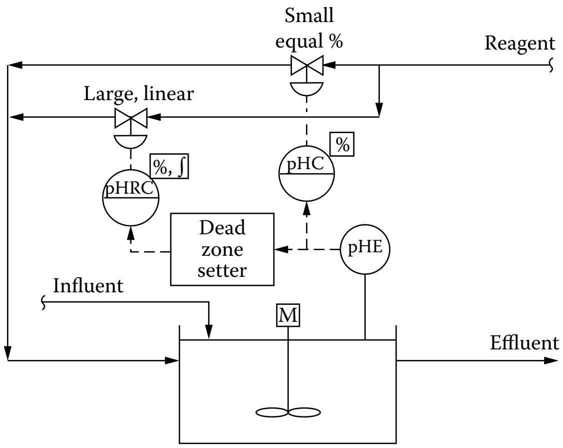


FIG. 8.39k
For accurate neutralization, the required high rangeability is provided by parallel control valves.

maximum flow rate. The minimum flow and minimum acid concentration determines the minimum capacity requirement of the reagent feed system.

Valve Rangeability Required As shown in Figure 8.39j, the reagent demand to neutralize a unit of wastewater increases tenfold every time its pH changes by a single unit. In other words, neutralizing a unit of wastewater at a pH of 3 will take 1000 times more reagent than would an effluent at a pH of 6.

As a consequence, a regent addition rangeability of several hundred to one or more may be required. This is accomplished by the use of two or more control valves in parallel, as has already been discussed in detail in connection with Figure 8.32h.

As shown in Figure 8.39k, the smaller valve has equal percentage characteristics and is throttled by a proportional only controller (pHC). This is desirable to match the pH characteristics near neutrality with that of the valve. If the pH measurement moves outside a preset and narrow "dead zone," this causes the second controller (pHRC) to make an adjustment in the opening of the large linear valve, thereby compensating for load changes. This second controller is provided with two control modes, with the integral action serving to bring the system back to set point after a load change.

A valve position control-based method of using multiple valves in a feedback-trimmed feedforward neutralization system was shown in Figure 8.32j in Section 8.32.

Sequenced Valves A wider reagent delivery capability can be obtained by using the sequenced valve approach where the controller output can be switched to either valve by a pressure switch (PS) or its electronic equivalent. With sequenced controls, only one valve at a time is operating, while the other is closed. In this way, the overall characteristics of a pair of equal-percentage valves are still equal percentage, which is represented by a straight line on a semilog plot.

Figure 8.39l illustrates various combinations of different pairs of sequenced valves. On such sequenced valves, positioners must be used because each valve must be calibrated to stroke over only a portion of the controller output signal range.

Table 8.39m lists the various flow rangeabilities for some of the valve pairs in Figure 8.39l, assuming a constant pressure drop across the valves (equivalent to 9.5 ft, or 2.85 m, of 66° Be sulfuric acid) and assuming an individual valve rangeability of 35:1. For the four valves, the valve capacity coefficients ($c_v s$) are 1.13, 0.14, 0.08, and 0.04, respectively, for CV-1, 2, 3, and 4.

The overlap between each valve pair becomes smaller as the rangeability increases. The pressure switch to transfer the valves can be set anywhere in the overlap region, because in this region the process loads can be satisfied by either valve.

Reaction Rates and Tank Sizing It is essential that reaction rates be determined so that suitable reaction tank sizes (residence time) can be calculated. Once the reaction rates are

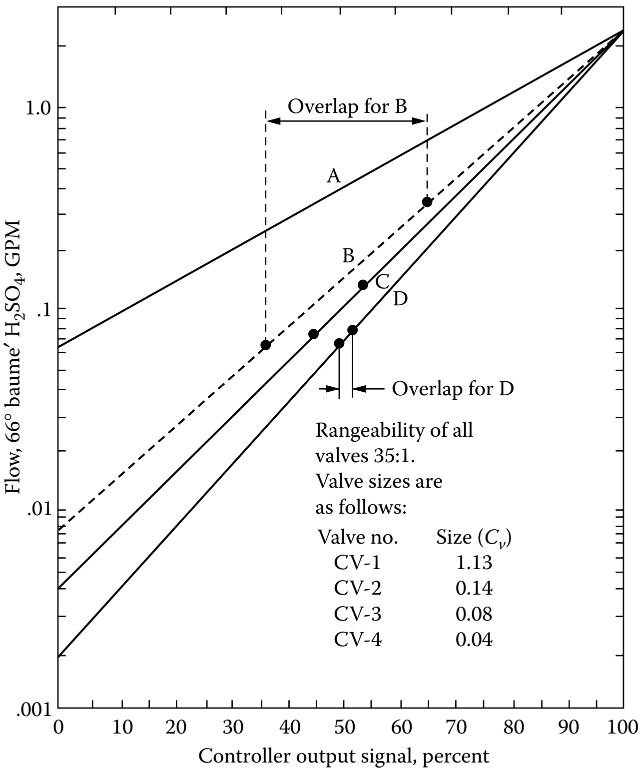


FIG. 8.39l
Delivery capability for various valve pairs.
Key: A = CV-1 alone; B = CV-1 + CV-2; C = CV-1 + CV-3; D = CV-1 + CV-4.

known, one can determine the residence time required to make sure that the reaction has time to go to completion. One obtains these rates by first determining the total amount of alkali (in case of an acid waste) that is required to neutralize a sample of the wastewater and then adding this amount to a second samp1e in a single dose.

Plotting the pH of that sample show the pH rise as a function of time (Figure 8.39n). The reaction vessel volume should provide at least 50% more holding time than the time it took for the sample to reach neutrality.

TABLE 8.39m
Reagent Delivery Turndown (Rangeability) for Sequenced Pairs of Equal-Percentage Valves

Valve Pair	*Line on Figure 8.22hh*	*Turndown*	*Log Turndown**	*Valve Positioner Calibration(s) (%)*
CV-1 (alone)	A	35:1	1.54	0–100
CV-1 + CV-2	B	275:1	2.44	0–63; 37–110
CV-1 + CV-3	C	570:1	2.76	0–58; 44–100
CV-1 + CV-4	D	1150:1	3.06	0–51; 50–100

* Signifies the approximate pH swing that valves will accommodate.

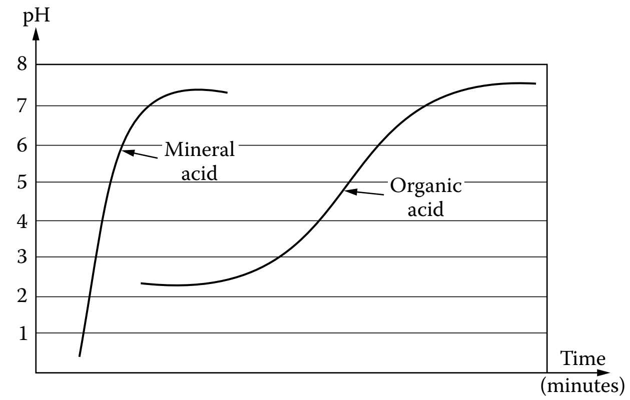


FIG. 8.39n
Acid neutralization reaction rates.

The ideal aim is obtain sufficient equalization to provide a homogeneous wastewater pH at a constant flow rate. To neutralize such a system, a lime feeder operating at a preset constant rate would suffice. Unfortunately, this seldom occurs, and provision must be made for reagent throttling.

Single Reagent Control

Figure 8.39o illustrates a system that can handle changes in both the flow and the acid concentration in the wastewater. In this equipment configuration, the equalization tank is followed by three treatment tanks, and to each of these tanks, the lime slurry feeding controls have identical capacities.

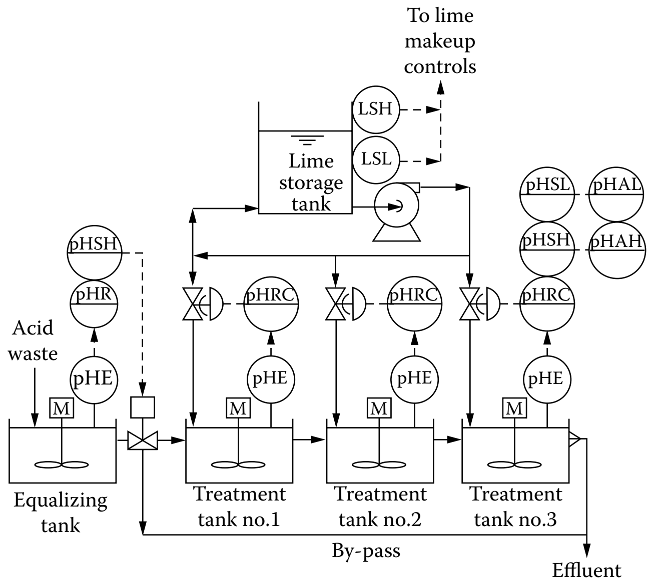


FIG. 8.39o
Simple acidic wastewater neutralization system with a 30:1 rangeability.

Assuming a 10:1 range for each lime slurry control valve, the range of the system is 30:1. Thus, it can handle any combination of flow and acid concentration within that range.

The three pH controllers are set at the same set point of usually between pH 6 and 8. At periods of low flow, or when the acid content is low, treatment tank number 1 can handle the entire neutralization load. Under other conditions, all three tanks may be required, with the first one or two satisfying a major portion of the reagent requirement and the third serving a "polishing," or final trim, function.

Each tank in this system should be sized for a minimum of 50% of the total retention time determined at maximum flow rate. Where mixtures of acids are involved, the maximum time (not average) must be used. Figure 8.39o also has a provision for the occasional case when the incoming wastewater is self-neutralizing and the treatment system can be bypassed.

High maintenance costs of pH electrodes have been reported when lime is used as the reagent, because of the formation of calcium sulfate coatings on the electrodes. For the designs of retractable and self-cleaning pH detectors, refer to Section 8.48 in Chapter 8 of the first volume of this handbook.

Two Reagent Control Systems

In some plants, the wastewater can be either acidic or basic. Figure 8.39p illustrates the two-sided feedback control system required to neutralize such wastewaters. Although only one valve for each side is shown, it would be possible to have a sequenced pair for one side of neutrality and a single valve for the other, or a sequenced pair for both sides.

Because this is a feedback control system, load changes cannot be frequent or severe in order for this system to give acceptable performance. For those applications in which load changes are frequent and severe, a combination feedforward-feedback should be considered, such as the system shown in Figure 8.32j in Section 8.32.

If sequencing is used, the reagent delivery system will have a high gain characteristic, because the stroking of the pair (moving from closed to open) is accomplished with only half the controller output signal, thereby doubling their gain (making them twice as sensitive). The valve gain will vary with the turndown, and a characterizer will be required for each set of sequenced valves to provide constant loop gain.

Ratio Control

Ratio control of pH can be effective when the process flow rate is the major load variable, and the objective is to meet increased flow with a corresponding increase in reagent. Because the errors in flow measurements and because reagent concentration may vary, a means for on-line ratio adjustment must be provided.

Figure 8.39q illustrates a ratio control system in which the reagent set point is changed proportionally to changes in wastewater flow. A feedback signal supplied by the feedback controller (pHC) also adjusts the reagent flow set point proportionally to a nonlinear function of the deviation between desired and actual effluent pH.

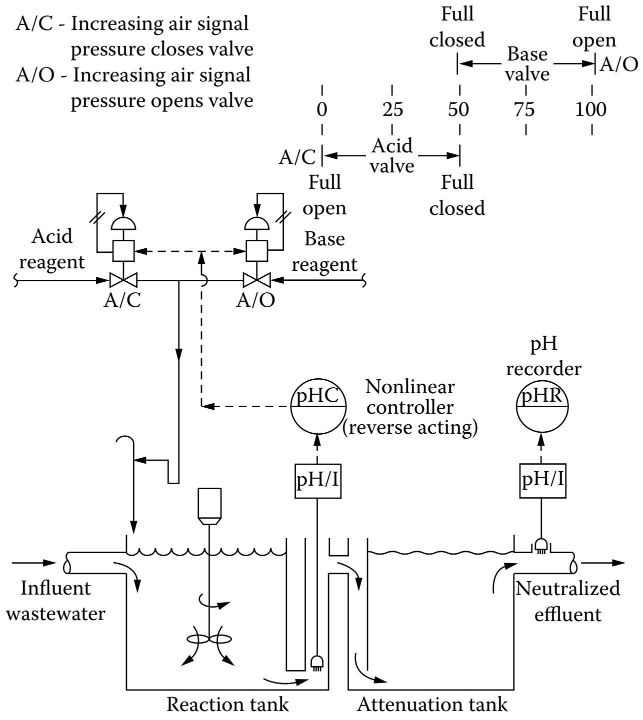


FIG. 8.39p
Two-sided feedback control of pH.

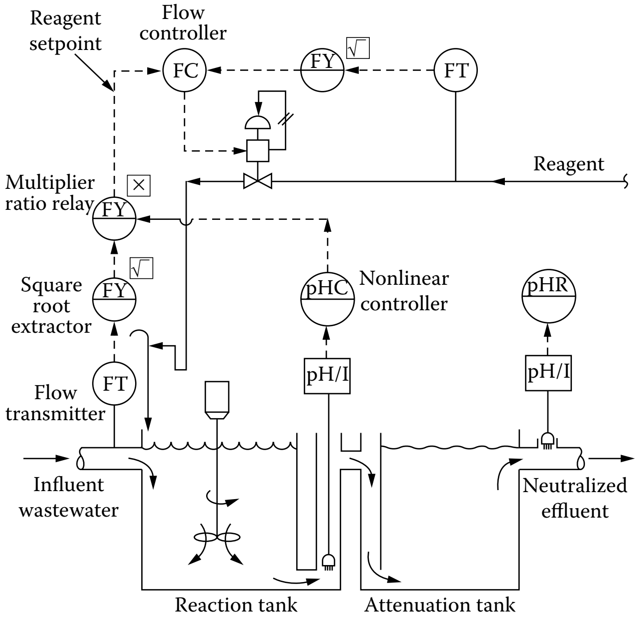


FIG. 8.39q
In this control system, the reagent flow is ratioed in a feedforward manner to the influent wastewater flow, whose ratio is feedback-trimmed by the pH controller.

Note that the rangeability of the ratio system is limited by that of the flowmeters, typically 4:1 for orifice meters to 30:1 for some turbine meters.

Cascade Control

Cascade control as applied to pH control systems can take two forms. In addition to the usual condition in which the output of one controller serves as the set point to another controller, it is also possible to have two vessels arranged in series, each with its own control system. The latter arrangement is referred to as *cascaded residences.*

The conventional cascade control system is shown in Figure 8.39r, wherein the output of controller pHC-1 is the set point of the slave, or secondary controller, pHC-2. This arrangement is particularly useful when lime is the reagent, because of the finite reaction time between the acid and reagent. The set point of pHC-2 may have to be lower than the desired pH of the final effluent, because the materials are still reacting with each other after they have left the first tank. If the set point pHC-2 is too high, the pH of the final stream will be greater than desired. When flocculation is to be carried out downstream of the pH treatment facility, stable pH values can be extremely important.

A delicate balance must be struck in this type of system with respect to the size of the first vessel. A long residence time in the first tank ensures long contact time between reagents, thereby producing an effluent pH that is close to the desired value, but at the same time it may result in a sluggish control loop around this vessel. For efficient cascade control, response of the inner loop (control loop around the first tank) must be fast.

The other control loop (pHC-2), sometimes referred to as the master, or primary, control loop, is usually tuned so as to be less responsive than the inner loop. The tuning of pHC-1 will be a result of the dead time (a delay between a change in reagent flow and the time when its effect is first felt), capacity, and process characteristics.

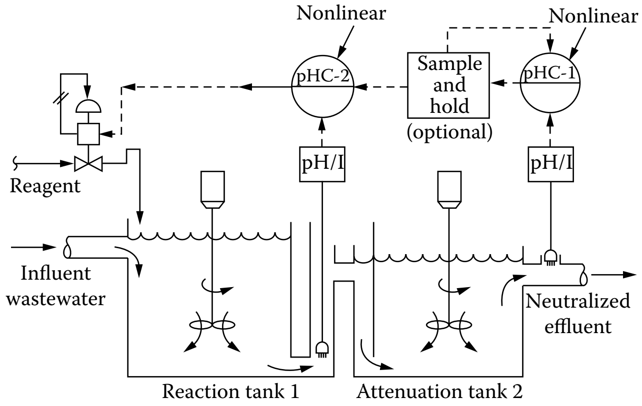


FIG. 8.39r
Cascade control of pH.

When this part of the process is dominated by dead time, the technique of sample data control may be useful in stabilizing the control system by a sample and hold device. This device may be a timer that automatically switches the controller between automatic and manual modes of operation (see Figure 2.2q in Chapter 2). This can allow the controller to be in automatic for a fraction (x) of the cycle time (t) and then can switch it to a fixed-output, manual condition for the rest of the cycle $(1 - x)t$.

Feedforward Control

In those systems in which equalizing basins or other averaging techniques cannot be applied and accurate pH control is required, control by feedback alone is insufficient, and feedforward schemes can be considered. It should be kept in mind, that feedforward based on influent pH is only recommended *if only a single specie in the wastewater must be neutralized*, because if there are several species, the relationship between influent pH and reagent demand cannot be predicted.

A feedforward control system is dedicated to initiating corrective action as soon as changes occur in process load. The corrective action is implemented using a control system that is essentially a mathematical model of the process. For example, the amount of base required to neutralize an acid can be predicted using the following formula:

$$\log(B) = \log(K) + \log(F) + (7 - \text{pHi}) \qquad \textbf{8.39(5)}$$

where

B is the base in normal equivalents
F is flow rate
K is a constant specific to the chemical reaction
pHi is the pH value of the influent

Ordinarily, the inclusion in the model of each and every load to which the process is subjected is neither possible nor economically justifiable. This means that a feedback control loop (usually containing the nonlinear controller for pH applications) is required in conjunction with the feedforward system. The function of the feedback controller is to trim and correct for minor inaccuracies in the feedforward model.

Feedback–Feedforward Combination Control Figure 8.39s describes a feedback-trimmed feedforward neutralization control system, using three control valves for high rangeability. In this system, the control signal to the reagent valves (X_b) is the sum of the characterized flow signal $f(F')$ of the influent wastewater and the output of the feedforward pH controller (pHC-1). Therefore, the total output signal to the sequenced pair of valves (X_B) becomes the sum given in Equation 8.39(6), minus twice the feedback controller's (pHC-2) output:

$$X_B = f(F') + \text{feedforward output} - 2(\text{feedback output}) \qquad \textbf{8.39(6)}$$

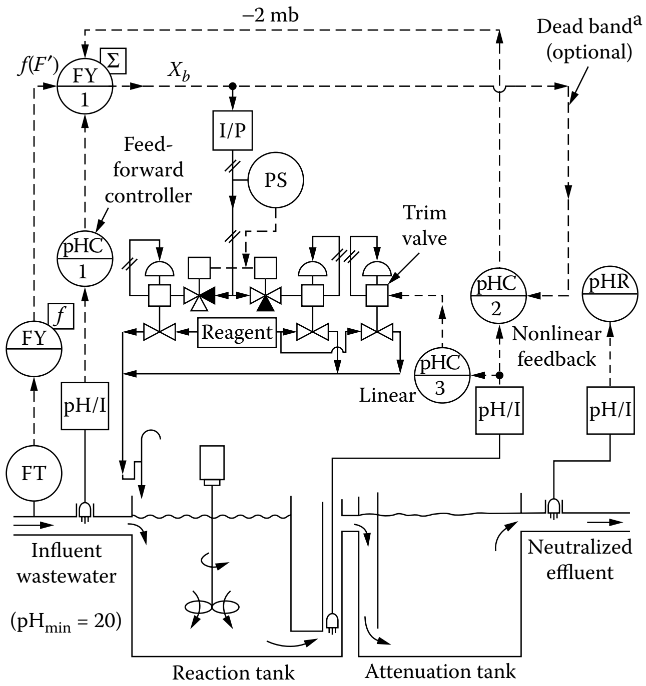


FIG. 8.39s
High rangeability feedback-trimmed feedforward pH control system utilizing three control valves, which are so configured that only one is throttling at any one time.

If the total rangeability of the pair of sequenced valves is 1500:1 and if the minimum expected influent pH is 2.0, then the maximum controllable influent pH value is pH(max) = 2.0 + log(1500) = 5.2 pH. Therefore, in this control configuration, when the influent pH rises to over 5.2, both sequenced valves will be completely closed and the linear feedback controller (pHC-3) will throttle the trimming valve. Figure 8.39t describes the operating ranges of the three valves and their total rangeability in terms of reagent flow.

The nature of a neutralization system is such that the process gain (slope of the pH curve) is likely to be the highest at set point (around pH = 7) and it decreases as the deviation from set point increases. To compensate for this inverse relationship between process gain and deviation, the feedback controller that is trimming the feedforward model (pHC-2) should be a nonlinear one, with its gain varying (increasing) with deviation.

With this controller, the greater the deviation from set point, the larger the controller's correction, but this does not create an unstable condition, because at high deviations the process gain is low. In the control system shown in Figure 8.39s, the feedback control signal (mb), which is trimming the feedforward signals in summer FY-1, is generated by the nonlinear pHC-3 controller. This controller can be provided with a dead band around its set point, so that it does not change its output until its measurement moves outside that gap.

Figure 8.39u illustrates a similar feedforward–feedback neutralization control system, where the expected rangeability

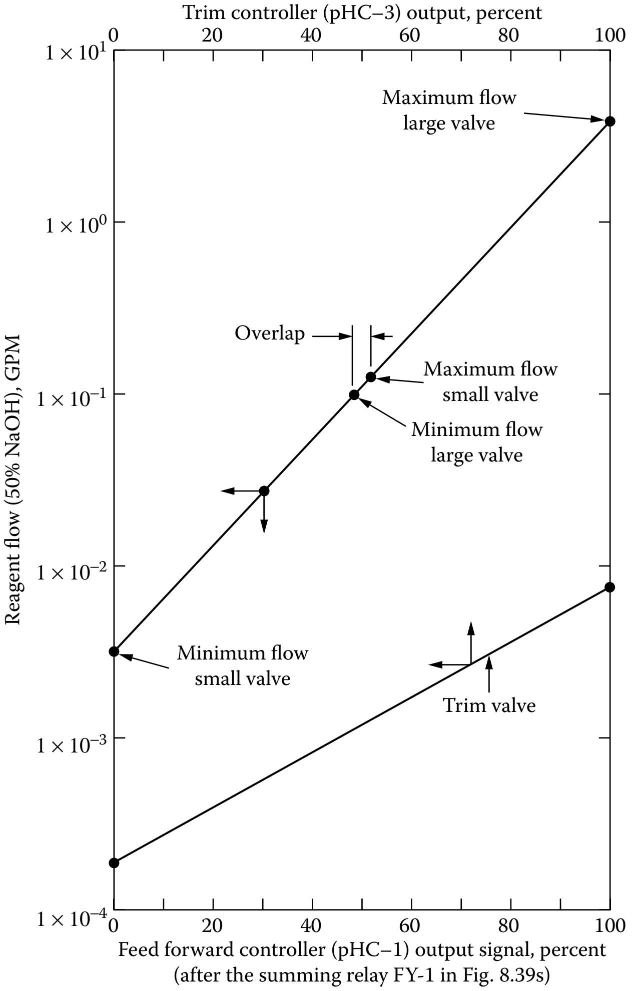


FIG. 8.39t
When the influent pH is below 5.2, the sequenced large and small valves throttle the reagent (top line). When the influent pH rises above 5.2, only the trim valve is modulated (bottom line)

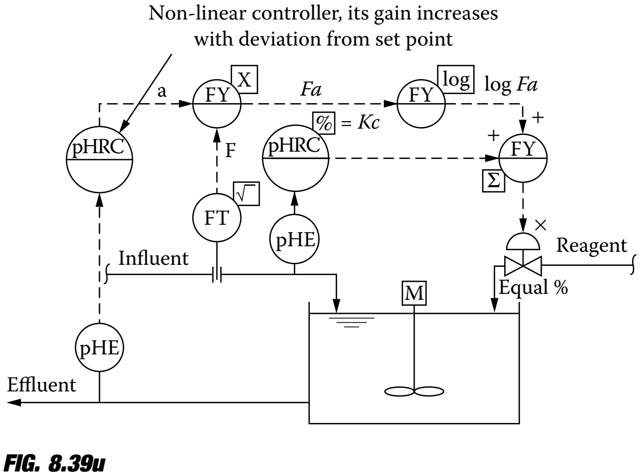


FIG. 8.39u
Low rangeability feedback-trimmed feedforward neutralization control system.

requirement of the reagent is low enough to be met with a single control valve. Here, the reagent valve opening (x) is calculated as follows:

$$X = Kc(\text{setpoint-measurement}) + \log(F.a) \qquad \mathbf{8.39(7)}$$

where

X is the valve position signal to an equal-percentage reagent valve
Kc is the gain of the proportional-only feedforward controller
F is the influent flow
a is the output signal of the nonlinear feedback controller

In this control system, the opening of the equal-percentage reagent valve is obtained by the summing of the feedforward pH controller's output signal with the logarithm of the product of the influent flow (F) and the output of the nonlinear feedback controller (a).

PRECIPITATION AND FILTERING

Precipitation is the creation of insoluble materials by chemical reactions that can then be removed through subsequent liquid-solids separation. Typical of these operations is the removal of sulfates, removal of trivalent chromium, and softening of water with lime. Iron and manganese are removed by a variation of this process, following the treatment discussed earlier in connection with the chemical oxidation process.

Lime Softening

The process of water softening and its controls are discussed in detail in Section 8.40. Figure 8.39v shows a precipitation control system that can be used for the precipitation of calcium carbonate in the process of water softening or for the precipitation of other insoluble crystals.

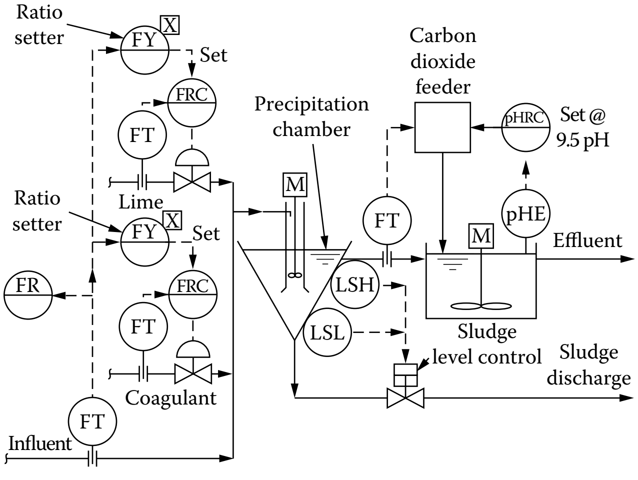


FIG. 8.39v
Calcium carbonate precipitation control system. (For the details of the gas feeder refer to Figure 8.39e.)

In this process, *crystal seeding* involves the acceleration of crystal formation by the presence of previously precipitated crystals. This is accomplished in practice by passing the water being treated through a "sludge blanket" in an upflow precipitation chamber, shown schematically in Figure 8.39v. The resulting crystals are hard, dense, and discrete; therefore, they separate readily.

When colloidal suspended material is also to be removed, which would be the case where surface waters are softened, a coagulant of aluminum, iron salts, or polymers is also added to precipitate the colloids. Dosage is variable, depending on the quantity of suspended material. Application of both coagulant and calcium hydroxide is controlled by flow-ratio modulation.

The resulting sludge, consisting of calcium carbonate, aluminum, or iron hydroxides, and the precipitated colloidal material are discharged to waste continuously. An automatic sludge-level control system is used in Figure 8.39v to keep the sludge level within an optimum range of operation.

Water softened by the excess lime treatment is saturated with calcium carbonate and, therefore, is unstable. Stability is achieved by addition of carbon dioxide to convert a portion of the carbonates into bicarbonate. This process is suited to automatic pH control as shown in Figure 8.39v. The carbon dioxide feeder has two operators; one is controlled in feedforward by the influent flow rate and the other in feedback by effluent pH control (pHRC). The set point is on the order of 9.5 pH.

Fouling of the electrodes is likely to occur as a result of precipitation of crystallized calcium carbonate. Daily maintenance may be expected, unless automated cleaners are used (Figure 8.39w). The farther downstream (from the point of carbon dioxide application) the electrodes can be placed, consistent with acceptable loop time delays, the less will be the maintenance requirement.

Hydroxyde Precipitation

Precipitation is the process used to remove soluble metal ions from solutions as hydroxydes. The process is pH controlled. By raising the pH with lime or sodium hydroxide, the corresponding metallic oxide precipitates out. Figure 8.39x shows the solubility curves of some heavy-metal hydroxides as a function of the pH of the solution.

As can be seen in Figure 8.39x, copper at a pH of 6 will start precipitating from the solution at a concentration of about 20 mg/l, while if the pH is raised to 8, its solubility drops to 0.05 mg/l.

Several metals such as chromium and zinc are amphoteric, being soluble at both alkaline and acidic conditions.

As can be seen in Figure 8.39x, chromium is least soluble in the range of pH 7.5 to 8.0. It is for this reason that in the second stage of chrome treatment in Figure 8.31d, the set point of the pHIC is 8 pH. For a detailed discussion of chrome treatment, refer to Section 8.31 or to the earlier paragraph in this section in connection with Figure 8.39h.

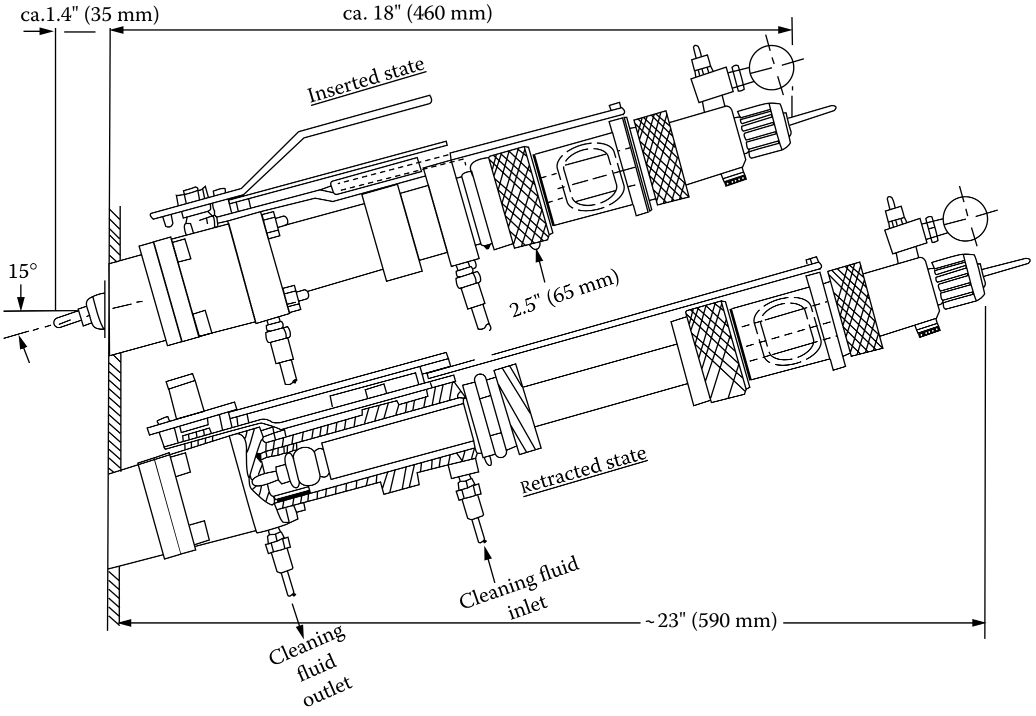


FIG. 8.39w
Self-cleaning, retractable pH probe installation.

Filtering

The solids suspended in the wastewater, including the particles generated by precipitation, can be removed by filtration. A large variety of filter designs are available, including diatomite, microstrainers, gravity, pressure, deep-bed, multilayered, cartridge, moving bed, and membrane. (See Section 7.33 in Chapter 7 of the second edition of the *Environmental Engineers' Handbook* for more details.)

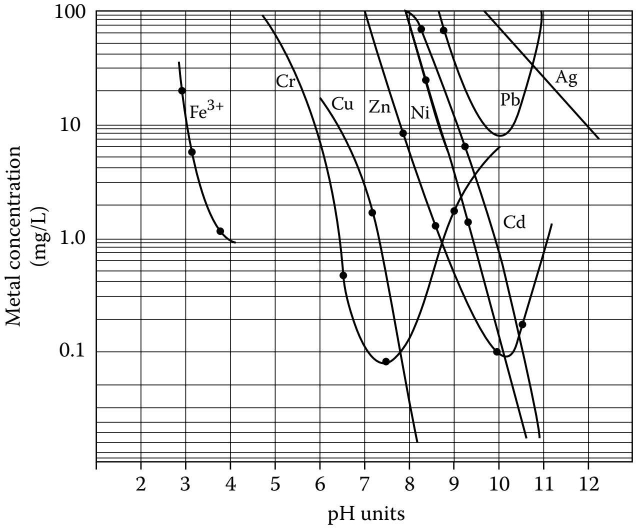


FIG. 8.39x
Solubility curves of a number of metal hydroxides at various concentrations as a function of the pH of the solution. (Courtesy of Hoffland Environmental Inc.)

Figure 8.39y illustrates the design of a multilayered filter, having a flat under-drain deck with long-stem nozzles for washing with backwash and air.

Figure 8.39z describes a deep-bed, pressure type, granular filter that is in forward flow operation. *Forward flow* refers to the state of filtering, while *back flow* refers to the backwashing or wash cycle during which the collected solids are removed and the filter is prepared for another filtering cycle.

Polishing filtration packages are usually designed in duplex configurations. In such designs, under normal "pump down" operation, while one filter is active, the other is being backwashed. The system is operated in its "normal" mode, when the level in the feed buffer tank is within its allowable high and low limits. In this mode, while one filter is filtering the wastewater from the large buffer tank and, after filtering, discharges it to drain, the other filter is being backwashed. The backwash cycle can either be initiated on the basis of the total flow that has passed through the filter or on the basis of the pressure drop that has built up across the filter.

The operation is automatically reconfigured as a function of the level in the large buffering feed tank. If the level in the buffer tank reaches a high level, the normal "pump down" mode of operation is switched to a "dual pump down" mode. In this mode, both filters are operating in their filtering mode, and the discharges of both filters are being sent to drain.

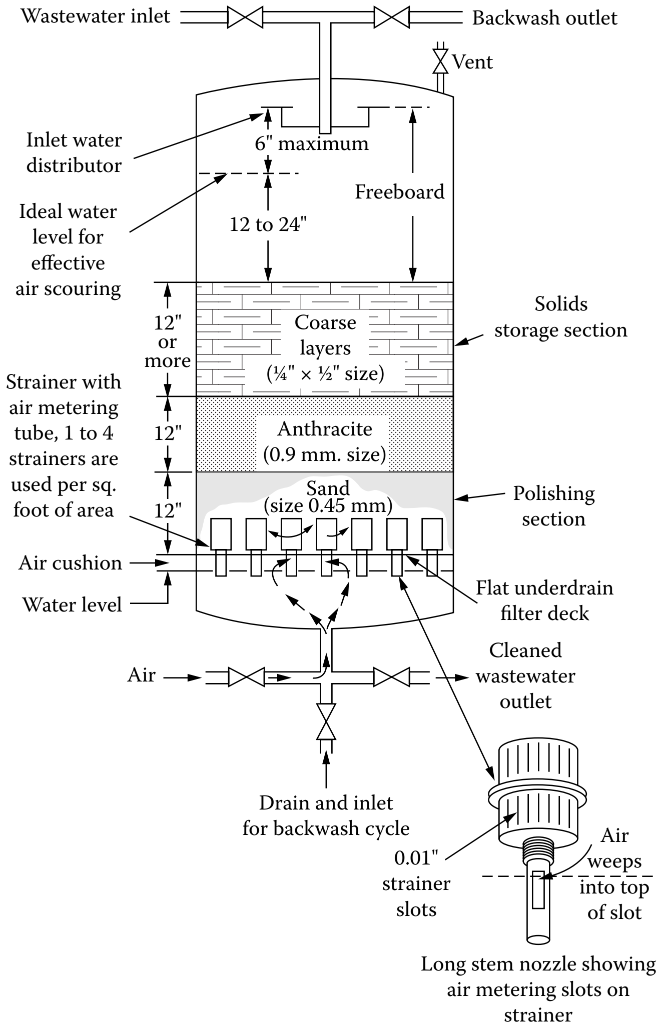


FIG. 8.39y
Construction of a multilayered filter.

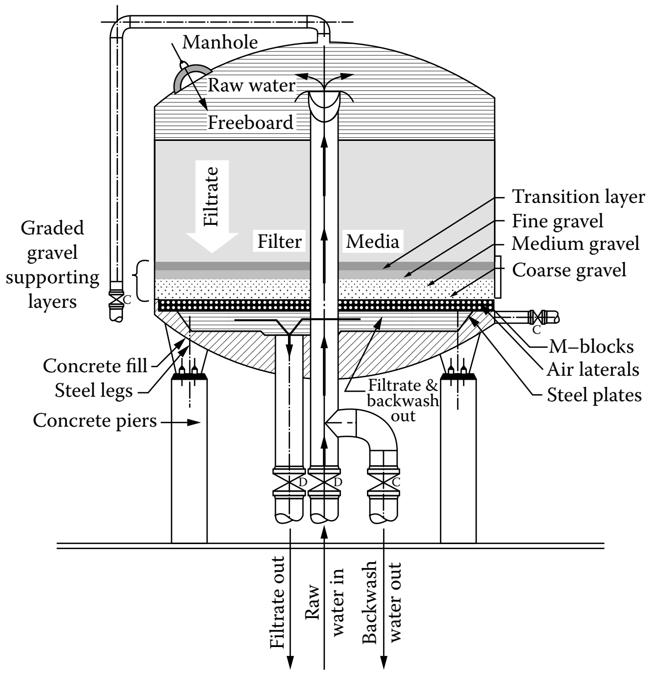


FIG. 8.39z
Construction of a deep-bed granular filter.

The buffer feed tank has three level probes, detecting high, normal, and low levels. If, after a "dual pump down" episode, the level drops to the "normal" probe, the system is returned to the single filter pump down mode and the inactive filter is backwashed. If the level in the buffer feed tank drops further and reaches the low level probe, the filtered wastewater will no longer be drained, but is returned into the holding tank.

Superfilters The required purity of nuclear power plant boiler feed water, in terms of conductivity, is 0.05 microSiemens. The total organic carbon (TOC) content of such feedwaters must be under 50 ppb. Such purity can only be obtained by membrane filtration. These designs are discussed under Tertiary Water Treatment in a later paragraph in this section.

In order to prevent corrosion, oxygen scavengers such as hydrazin are also added to boiler feedwater.

CHLORINATION

Nearly every wastewater treatment process uses chlorination, although drinking water chlorination is being phased out in favor of ozonation. Chlorination may be used either to reduce the possibility of pathogenic bacterial contamination of the receiving water or to prevent biological growth from interfering with other processes. Control of biological slimes that interfere with heat exchange in cooling water systems is often accomplished with chlorine, but also with biocides.

Chlorine is a strong oxidant with many other uses, as were described earlier in connection with chemical oxidation. It is a very effective bactericidal agent readily available at reasonable cost. Its effectiveness has been proved by over 75 years of use.

Several factors contribute to the complexity of chlorination for biological control. These include that the materials in the water that can be chemically oxidized cause an immediate reduction of the available active chlorine into an ineffective chloride form. Therefore, sufficient chlorine must be added to the water to account for this reaction and, in addition, to provide a residual amount, reserving sufficient chlorine for subsequent reactions. The amount of chlorine involved in the initial reduction is called *chlorine demand*. A dosage greater than this amount provides excess chlorine, which is called *residual chlorine.*

When nitrogenous material, particularly ammonia, is present in the water, the residual chlorine will be altered. Two kinds of residual chlorine are recognized: (1) free residual is that remaining after the destruction with chlorine of ammonia or of certain organic nitrogen compounds; (2) combined residual is produced by the reaction of chlorine with natural or added ammonia or with certain organic nitrogen compounds.

Because these two compounds (free residual chlorine and combined residual chlorine) are completely different in their ability to control bacterial organisms, it is important to differentiate which of the two forms of chlorine are involved in a process. Laboratory methods are available for both the measurement and the differentiation among these two.

Continuous analyzers (see Section 8.11 in Chapter 8 in the first volume of this handbook) suitable for control are also available to measure either the level of free residual chlorine or the level of total residual chlorine, meaning the combination of the two when both are present.

The presence of residual chlorine in water does not ensure either disinfection nor biological control. A bacteriological analysis requires several hours, or even days. Through years of experience, it has been determined that measurement of residual chlorine can be a suitable inferential indicator of the effectiveness of biological control. For this reason, such processes are suitably controlled by analysis of residual chlorine.

An example of a typical control system used in the disinfection of wastewater is shown schematically in Figure 8.39aa. The chemical feed system for applying chlorine must be of sufficient capacity both to satisfy the chlorine demand of the waste and to provide sufficient residual after the contact time that is required for the disinfection action has passed. Typical of this for wastewater treatment would be a chlorine demand of 5 mg/l.

Another important factor is that the amount of residual chlorine will decline with time. There is no assurance that the initial residual concentration will persist for the length of time required for the disinfection to be accomplished. Yet, it is important that residual chlorine be present during this entire period. Typical design basis is to size the contact chamber for 30 min retention at maximum flow.

Local regulations usually require that there be a minimum of 1 mg/l of available chlorine at the end of this contact time. It is not uncommon for flows to vary over a range of 6:1. To account for these variations and for the continuing chemical reactions, the control system must automatically control the residual after a short (approximately 5 min) fixed contact time. This is accomplished by building the 5 min retention time into the sampling system.

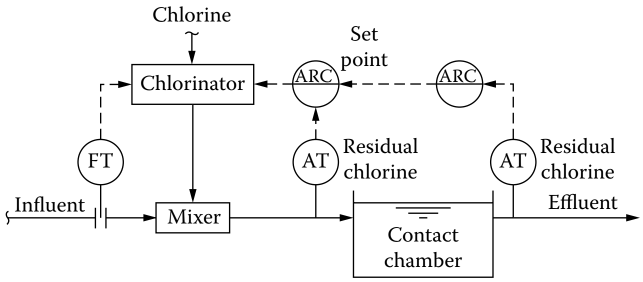


FIG. 8.39aa
Illustration of chlorination-based disinfection controls used when both the flow rate and the quality of the wastewater influent is variable. For the details of the chlorinator, refer to Figure 8.39e.

Feedback control of residual chlorine, in conjunction with flow-proportioning feedforward correction (Figure 8.39aa), establishes a constant residual value at the inlet of the contact chamber. Because of the variable rate at which residual decay occurs, assurance that the residual at the exit of the contact tank is maintained at a minimum value requires a second analyzer. This cascade master adjusts the set point of its slave, the inlet residual chlorine controller, as required.

The amount of residual chlorine decay varies widely as a result of variations in temperature, quality of wastewater, and detection times. It may vary from a minimum of 0.5 to 5 mg/l. To maintain an effluent residual of 1 mg/l, the set point on the controller for the chlorinator may be anywhere from 1.5 to 6 mg/l.

Chlorination of wastewater for disinfection is unique in that it is usually the final process unit prior to discharge. For this reason, detection time is provided as a part of the process. For most other biological control applications, other subsequent unit operations provide sufficient contact time, and the residual decay can be reasonably well predicted.

MUNICIPAL WASTEWATER TREATMENT

As shown in Figure 8.39bb, the conventional household sewage treatment process consists of pretreatment, primary settling, aeration, final clarification, and disinfection steps.

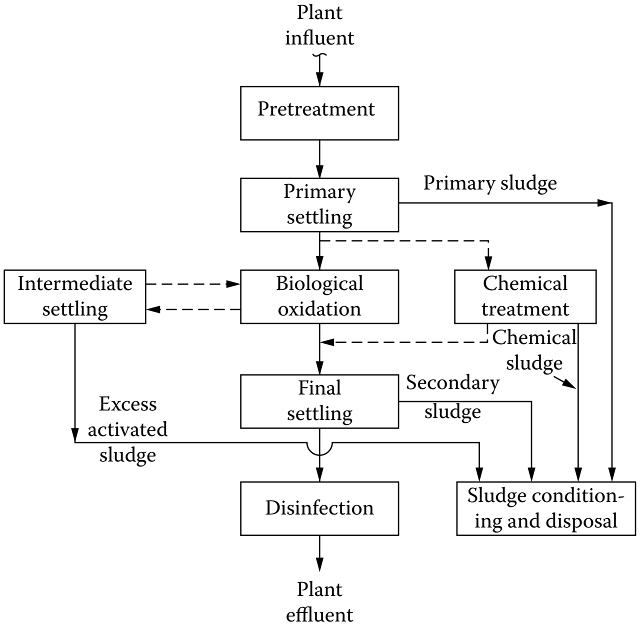


FIG. 8.39bb
The basic stages of municipal wastewater treatment.

Pretreatment equipment includes screens, grinders, skimmers, grit chambers, flow equalization, and primary settling. Traveling grates are also used to remove large debris. Plugging is detected by level differential across the grate. Older systems use bubbler tubes with d/p transmitters for this measurement; newer ones use noncontact ultrasonic level transmitters.

The second step usually is biological oxidation, which can be done either in aeration tanks or in aeration ponds. The aeration process is controlled by dissolved oxygen (DO) analyzers. See Section 8.43 in Chapter 8 of Volume 1 of this handbook for the description of the submersed platinum electrodes operating under diffusing Teflon membranes that are used for this measurement. The biological oxidation step is usually combined with chemical treatment and intermediate settling, which respectively produce chemical and excess activated sludge.

The aerated water is allowed to settle in a settling tank, which is also referred to as a clarifier or a thickener. Figure 8.39cc illustrates a circular basin-type clarifier. Other designs include horizontal flow, solids contact, and inclined surface versions. RF capacitance probes and ultrasonic sludge interface detectors are both often used to detect the sludge blanket level in the clarifier (Sections 3.3 and 3.20 in Chapter 3 Volume 1). The settled sludge is pumped to sludge digester or sludge conditioning units.

Microwave-type density sensors are often used to detect the consistency of the secondary sludge from the final settling tank (see Chapter 6 in the first volume of this handbook). The purpose of the sludge density measurement is to make sure that the secondary sludge that is being pumped to sludge conditioning or sludge digestion is thick enough in its solids content.

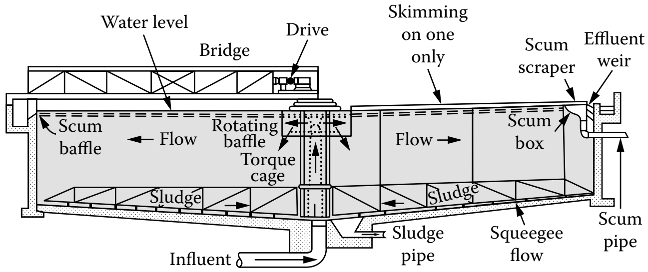


FIG. 8.39cc
Circular basin-type clarifier design.

Either prior to, after, or in parallel with biological oxidation, treatment chemicals such as ferrous chloride or soluble polymers are added to cause flocculation. Figure 8.39dd describes a combined coagulation and settling apparatus used for such dual purposes.

The flocculant and coagulant chemicals are usually added by metering pumps, having either variable-stroke or variable-speed adjustments. The flocculant flow through these pumps is ratioed to the sewage influent flow, so that the chemical dosage is maintained constant. The sewage influent flow is usually measured by a magnetic flowmeter (Figure 8.39ee) if a pipe is used. In open channels, Parshall flume or ultrasonic flow detectors are used (see Chapter 2 in Volume 1 of this handbook).

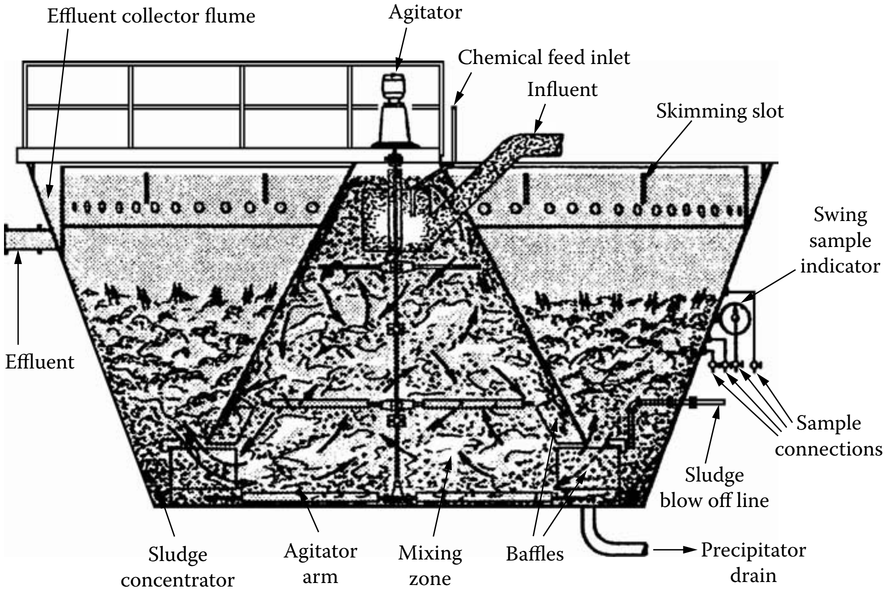


FIG. 8.39dd
Apparatus for the combined processes of coagulation and settling.

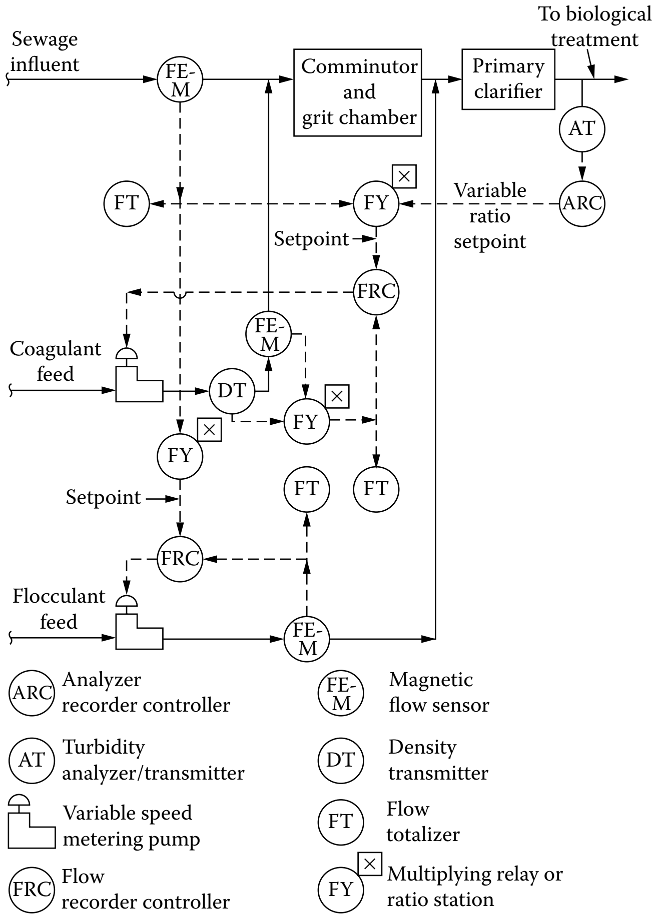


FIG. 8.39ee
Control system for automatic flocculant and coagulant additions.

The coagulant flow is also ratioed to the sewage influent, and the ratio is modulated by the cascade master (ARC in Figure 8.39ee), which keeps the turbidity level of the clarifier effluent constant. If the pH is under 7.5, caustic solution is also added under pH control by a positive displacement pump, because good flocculation requires that the pH in the tank exceed 7.5.

Final settling is usually followed by disinfection and filtration steps, which were discussed in connection with Figures 8.39y and 8.39z. The sand filters are backwashed either on the basis of time or on the basis of totalized flow through them. The clarity of the filtered water is monitored by light-scattering turbidity detectors (see Section 8.60 in Volume 1 of this handbook).

Disinfection While ultrafiltration removes quantities of pathogens, other methods of filtration require disinfection, either by UV radiation, chlorination, or ozonation, mainly to remove coliform bacteria. While coliform tests are lab based, most water treatment plants also use refelometric turbidity meters to bring the purity to 0.06 turbidity units. Chlorination is controlled by residual chlorine analyzers. They measure free chlorine or total chlorine (which includes, e.g., NH_2Cl).

Tertiary Wastewater Treatment

Tertiary treatment is used for the water reclamation where the purification process is followed by disinfection. Membrane filtration separates the influent into two streams. The stream that has passed through a semipermeable membrane will have less solids than the one that did not. The designs used include reverse osmosis (RO), nano-filtration (NF), ultrafiltration (UF), and microfiltration (MF), which is the separation of submicron particulates from dissolved material.

Reverse osmosis uses hollow filter membranes to route ions into one side and pure water into the other side of the membrane. Pumps generate the RO pressure, which usually exceeds 900 psig. In the RO process pH is monitored to prevent the precipitation of alkaline salts and the water purity is detected by conductivity and ion selective electrodes (see Sections 8.17 and 8.28 in Chapter 8 in the first volume of this handbook).

Nano-filtration separates monovalent salts from multivalent and uncharged organic molecules. NF is used for desalting and for concentrating dyes.

Ultrafiltration is used in paper mill, oily, and latex effluent streams. It can also be used as a pretreatment to ion exchange to prevent particulate fouling of the resin bed.

Microfiltration uses ceramic and polymeric membranes for processing liquids in the pharmaceutical and chemical industries, as well as for separating wastewater effluents.

Deionization Ion exchange is mostly used to soften water by the removal of calcium and magnesium. These units can remove not only water hardness, but also other mineral salts and silica. They can provide 2 ppb purity. Ion exchange purified water can be further cleaned by reverse osmosis, combined with ultrafiltration.

Ion exchangers are vessels filled with resins that are saturated with Na+ during regeneration by a salt solution. When the cycle is switched to deionization, the resin captures Ca++ and Mg++ while releasing Na+. When Na+ is exhausted, the resin is backwashed (Figure 8.39ff). During the first phase of regeneration, Ca++ and Mg++ are purged to waste.

The sequencing of the run, regeneration, purge, and backwash phases of the operating cycle are usually PLC controlled, and these phases are started either on the basis of elapsed time or because a conductivity sensor indicates an ion-breakthrough.

When the removal of Cl^- or sulfate is also required, mixed-bed ion exchangers are used that contain both anionic and cationic resins. The anionic resins are regenerated by HCl, which saturates the H+ ions. Conductivity sensors are used to initiate the regeneration of the exhausted unit while switching the water to the stand-by unit.

The sludge from the settling and treatment steps is usually dewatered in rotary filters and is then dried for disposal as soil supplement or is sent to aerobic or anaerobic digesters for fermentation. The H_2S in the generated digester gases is usually monitored by infrared analyzers (see Section 8.26 in Chapter 8 in Volume 1 of this handbook).

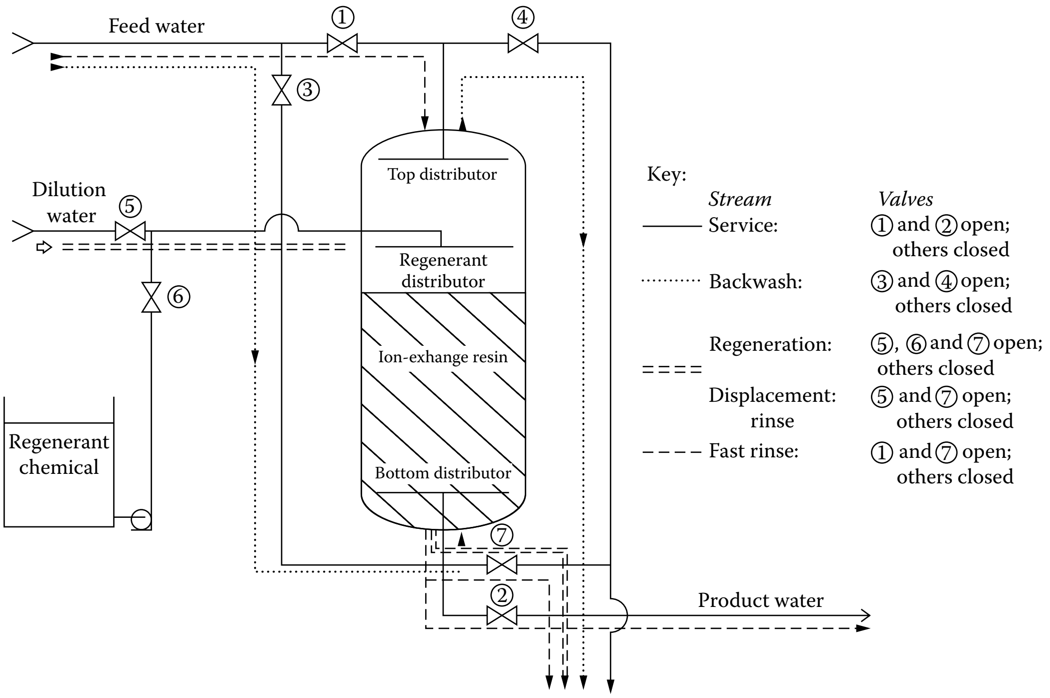


FIG. 8.39ff

Simplified ion-exchange operations cycle. The water used for backwash, dilution water, or displacement rinse can be feed water, softened water, decationized water, or DI water depending on the ion-exchange resin used and the quality of water produced in the service cycle. (Reprinted with permission from Owens, 1985, p. 89.)

Safety

If the treatment plant is in a closed building, the air in the building has to be monitored for toxic H_2S, for combustibles in explosive concentrations, and for oxygen deficiency. The low-oxygen alarm is usually set at 19.5% of oxygen in the air. Combustibles can be detected either by infrared sensors or by catalytic combustion detectors (for details see Section 8.16 in Chapter 8 in Volume 1 of this handbook).

The safety monitoring sensors actuate both local and central alarms. Figure 8.39gg illustrates the remote head-type safety monitoring installation.

In remote and unattended installations, radio transmission in the 920 MHz band is used. The radios can receive both analog and on/off contact signals, and their receivers can transmit them to SCADA, PLC, or PC devices, using Ethernet or Modbus protocols.

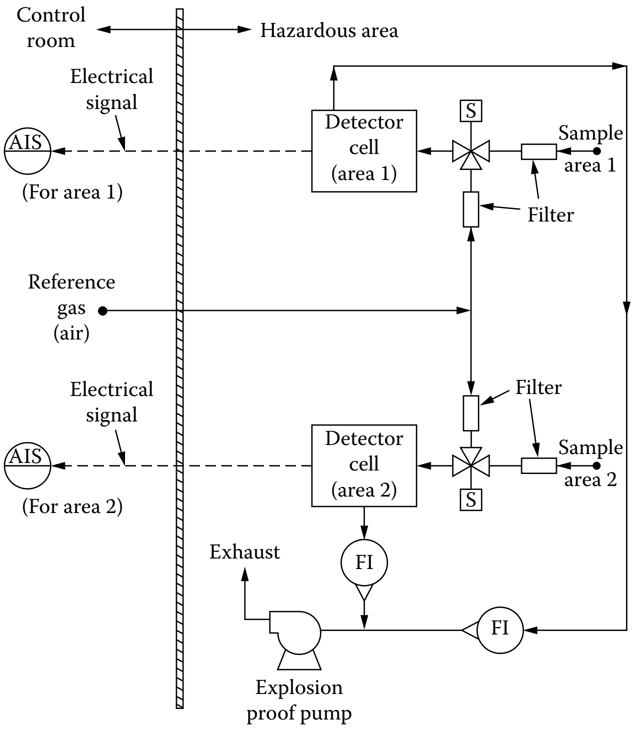


FIG. 8.39gg

Safety monitoring installation using remote detector cells located in the monitored area.

COOLING WATER TREATMENT

For a detailed discussion of cooling tower controls and optimization, refer to Sections 8.16 and 8.17.

Cooling systems can be once-through, cooling tower, or closed-loop cooling systems. Most power plants and refineries use cooling towers with water recirculation, but evaporation losses will cause concentration of scaling compounds such as calcium and magnesium carbonates. Also, biological

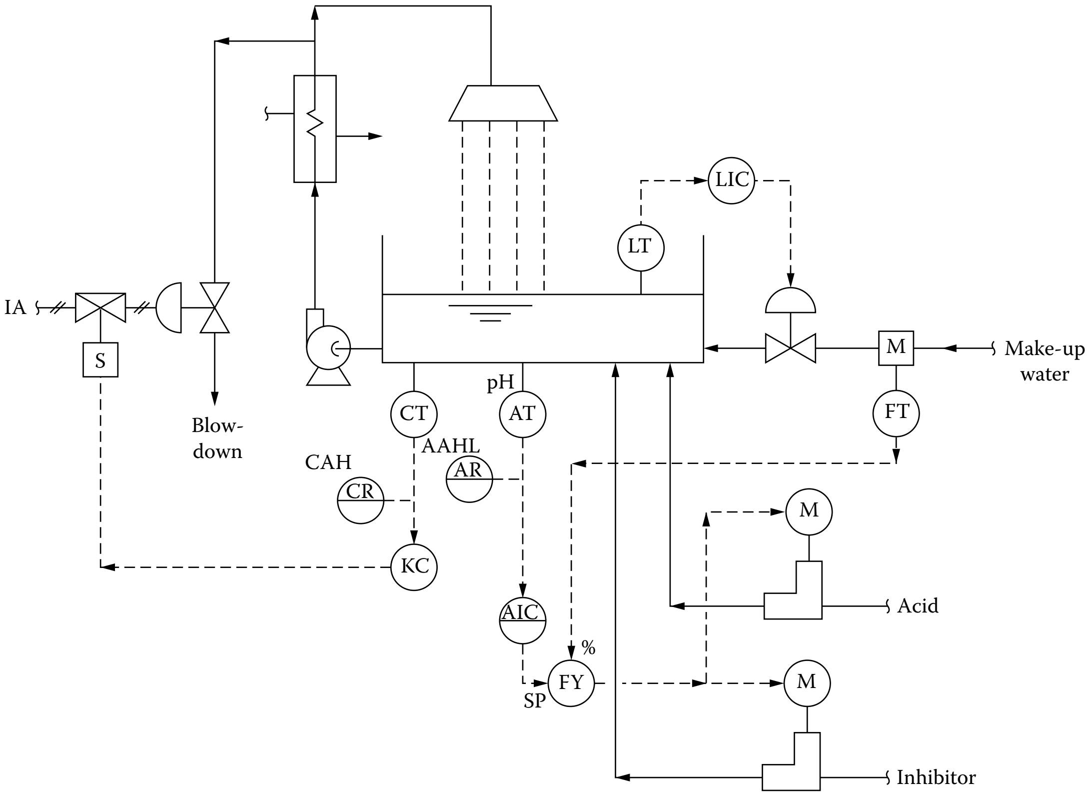


FIG. 8.39hh
Cooling tower water treatment controls. CT: Conductivity transmitter; KC: Intermittent timer.

fouling (algae, slimes) of water pipes and heat exchangers needs to be prevented by using chlorine, chlorine/bromine, or biocides. Sometimes, shock treatment is used, i.e., the periodic introduction of high chlorine concentrations.

Figure 8.39hh describes part of a control system for treating the water in the cooling tower basin.

Some of the considerations in connection with that system are:

- Make-up water needs to be treated with biocides or with chlorine, and these additives should be ratioed to the make-up flow rate. A water meter with totalizer can start an on/off timer, which will start and stop a positive displacement-type, variable-stroke chemical feed pump (interlock not shown in Figure 8.39hh). Periodic biocide shock treatment is normally done manually.
- Make-up water may need pH adjustment (to 7.2–7.6) by injecting an acid (or sometimes caustic). Sulfuric acid is used to destroy scaling bicarbonates of calcium.
- The pH in the basin is monitored and is used to control (AIC in Figure 8.39hh) the acid (caustic) feed rate, which is also proportioned to the make-up water flow rate.
- Conductivity is measured in the basin and used to adjust the blowdown rate. Blowdown may be continuous, using a control valve, or intermittent (as shown in Figure 8.39hh), using an on/off valve.
- Corrosion inhibitors and antiscalants can be mixed in with the acid in the acid feed tank or can be added under the same controls as the acid, as shown in Figure 8.39hh.

The cooling tower's water chemistry can be upset by wind, rain, or drifting spores. Some microorganisms also generate corrosive gases. The treatment process (as was explained in detail in Sections 8.16) has long time constants and dead times. For these and many other reasons, advanced controls (Section 8.17), including statistical process control (Section 2.34 in Chapter 2), can be considered.

Because cooling towers are unattended, upsets detected by the pH and conductivity meters are set to trigger remote alarms, by radio or using a telephone dialer.

CONCLUSIONS

Most wastewater treatment systems are designed for continuous operation, using several process units. Many of the processes consist of chemical treatment. Some of the features of automated effluent and water treatment systems are summarized in Table 8.39ii. The natural laws governing chemical

TABLE 8.39ii

Control Instrumentation Applicable to Various Wastewater Treatment Systems

	Instrumentation			
Treatment Process	*Flow*	*pH*	*ORP*	*Residual Chlorine*
Chemical oxidation of				
Cyanide		✓	✓	
Iron	✓			✓
Manganese	✓			✓
Hydrogen sulfide	✓			✓
Chemical reduction of				
Chromium		✓	✓	
Residual chlorine	✓			✓
Precipitation of				
Chromium	✓	✓		
Iron	✓	✓		
Manganese	✓	✓		
Hardness	✓			
Neutralization of				
Acid and alkali	✓	✓		
Alkalinity (recarbonation)	✓	✓		
Biological control	✓			✓

reactions dictate the design considerations. The most critical of these involve (1) pH, (2) reaction rates, (3) ratios (chemical dosage), (4) concentration, and (5) temperature.

References

1. Shinskey, F. G., *Process Control Systems,* 3rd Edition, New York: McGraw-Hill, 1988.
2. Shinskey, F. G., *pH and pIon Control in Process and Waste Streams,* New York: John Wiley & Sons, 1973.

Bibliography

Baltutis, E. W., “Fieldbus in Water and Wastewater,” ISA/93 Technical Conference, Chicago, IL, September 19–24, 1993.

Bethlehem Regulation Selection Handbook, Bethlehem, PA: The Bethlehem Corporation.

Butler, J. N., *Ionic Equilibrium: A Mathematical Approach,* Reading, MA: Addison-Wesley, 1964.

Corrigan, T. E., and Young, E. F., “General Considerations in Reactor Design—II,” *Chemical Engineering,* October 1955.

Daigger, G. T., Ed., *Biological Wastewater Treatment,* 2nd Edition, New York: Marcel Dekker, 1998.

Devlin, P. M., Hwang, J., Winchell, C. J., Day, S. G., Zura, R. D., and Edlich, R. F., “Automated Hydrotherapy Pool Water Treatment System,” *Hydrotherapy Pool Management,* Vol. 10, No. 1, January/February 1989.

Docherty, P. J., Jr., “Automatic pH Control: Neutralization of Acid Wastes by Addition of Lime Slurry,” Masters thesis, Hanover, NH: Dartmouth College, June 1972.

Eckenfelder, W. W., Jr., *Industrial Water Pollution Control,* 3rd Edition, New York: McGraw-Hill, 1999.

Gray, J. R., “Glass pH Electrode Aging Characteristics,” ISA/93 Technical Conference, Chicago, IL, September 19–24, 1993.

Guisti, A. L., and Hougen, J. O., “Dynamics of pH Electrodes,” *Control Engineering,* April 1961.

Hoyle, D. L., “The Effect of Process Design on pH and pIon Control,” presented at the 18th National Symposium of the Analytical Instrumentation Division of the Instrument Society of America, San Francisco, CA, May 3–5, 1972.

“Ion-Selective Measuring Systems,” *Technical Information Sheet* 43–31a, Foxboro, MA: The Foxboro Company.

Light, T. S., “A Standard Solution for Oxidation-Reduction Potential (ORP) Measurements,” *Analytical Chemistry,* May 1972.

Mahuli, S. K., Rhinehart, R. R., and Riggs, J. B., “Nonlinear Model-Based Control of pH,” AIChE Annual Meeting, 1992, Paper 149g.

McMillan, G. K., *Continuous Control Techniques for Distributed Control Systems,* Research Triangle Park, NC: Instrument Society of America, 1991.

Merriman, D. C., “Junction Potential Variations in pH,” ISA/93 Technical Conference, Chicago, IL, September 19–24, 1993.

Mooney, E. F., “On-Line Photometric Titrations for Process Control,” 1991 ISA Conference in Houston, TX, October 1992.

Moore, F. E., “Taking Errors out of pH Measurement by Grounding and Shielding,” *ISA Journal,* February 1966.

Negus, L. E., and Light, T. S., “Temperature Coefficients and their Compensation in Ion-Selective Measuring Systems,” presented at the 18th National Symposium of Analytical Instrumentation Division of the Instrument Society of America, San Francisco, CA, May 3–5, 1972.

Parsons, W. A., *Chemical Treatment of Sewage and Industrial Wastes,* Washington, D.C.: National Lime Association, 1965.

Piovoso, M. J., and Williams, J. M., “Self-Tuning pH Control,” *InTech,* May 1985.

Shinskey, F. G., and Myron, T. J., “Adaptive Feedback Applied to Feedforward pH Control,” Instrument Society of America, Paper No. 565–70.

Steininger, J. M., “Clean Up with OPR,” *The Business Magazine for the Spa & Pool Professional,* June 1990.

Waarvick, C., “Automated Odor Control Systems in Wastewater Facility,” *InTech,* August 1993.

8.40 Water Supply Plant Controls

P. Y. KESKAR (2005)

Reviewed by B. G. Lipták (2005)

INTRODUCTION

This section describes some of the control strategies and instrumentation systems that can be used in the treatment of raw water in the preparation of drinking water supplies. The described control systems correspond to applications in which PLC-based controls were used. This is not intended to suggest that a PLC-type of control hardware is preferred for these types of applications, but only that the author's experience was based on that type of hardware. The last paragraph of this section provides a brief discussion of some of the PLC limitations of the PID loops described in this section.

The previous section, Section 8.39, discussed wastewater treatment controls. Because some of the unit processes are utilized in both types of plants, there is some overlap between the coverage of these two sections, and it is recommended to review both.

This section will describe the controls of conventional water treatment plants, which perform lime softening, filtration, and chlorine- or ozone-based disinfection. The section will also describe the control of water treatment plants using reverse osmosis or submerged ultrafiltration membrane technologies. Examples of actual treatment plants will also be described by flow diagrams.

CONVENTIONAL WATER TREATMENT PLANTS

Figure 8.40a shows a simplified block diagram of the unit processes involved in a typical groundwater treatment plant, where raw water is pumped from wells into the treatment units that perform lime softening. In this step, the lime for softening (LS) and the polymer (PO) that assists in the coagulation are added by metering pumps.

The coagulated calcium carbonate and colloids, which were used in the water softening step, settle down as chemical sludge (CHS). This sludge is periodically discharged into the washwater recovery and sludge handling unit process by a blow-down pump station.

The softened water (FI in Figure 8.40a) flows into the filter influent channel, where carbon dioxide is added to stabilize it and to bring the alkalinity of the water down to 9.5 pH. The pH-adjusted water is filtered in a series of filtration steps, and the filtered and treated water (FE) is sent to a clear well for transferring into ground storage tanks. A disinfectant chlorine solution (CS) is added in the softening, in the filtration, and again in the clear well steps. A high-service pumping station pumps the water from the ground storage tanks (FW) into the water distribution system.

Filters are periodically backwashed and air scoured to clean the filter media. The backwash water from the filters is drained (BWD) to the washwater recovery process. The washwater recovered from the washwater recovery process is recycled (BWR) back to the treatment units, and the sludge (DWS) is pumped to vacuum filters for dewatering and disposal.

The sludge thickener also receives sludge (DWS) from the washwater recovery basin. Sludge pumps transfer the sludge from the washwater recovery basin to the sludge thickener (see Figure 8.39cc). The thickener decant is received in the decant pump station, where the water is pumped back to the treatment units, and the sludge is pumped to the vacuum filter for dewatering and disposal.

The paragraphs that follow will describe the major unit processes and their associated controls.

Filtration Hardware and Controls

The clarified and softened water from treatment units enters the filter influent channel common to several filters (FI in Figure 8.40a). Here, the water passes through the filter media for the removal of suspended solids from the water such as silt, clay, colloidal particles, and microorganisms.

Filtration removes disinfection-resistant microorganisms and colloidal particles that can cause growth and deposition in piping and can alter the taste of water. Figure 8.40b illustrates a typical filtration process and its related controls. In this configuration, the filter effluent and backwash flows, the filter differential pressure, the influent level, and turbidity are continuously measured, and the effluent flow is continuously throttled.

Signals from all valves and instruments are connected to a local control panel that houses a remote I/O rack. On/off sequencing valves receive open/close commands from the PLC, and their status is reported to the PLC. The controls shown in Figure 8.40b can be typical for various numbers of filters.

PLC Configuration

Figure 8.40c illustrates the configuration of the PLC equipment used in the filtration process that was described in Figure 8.40b. The system consists of a redundant PLC housed

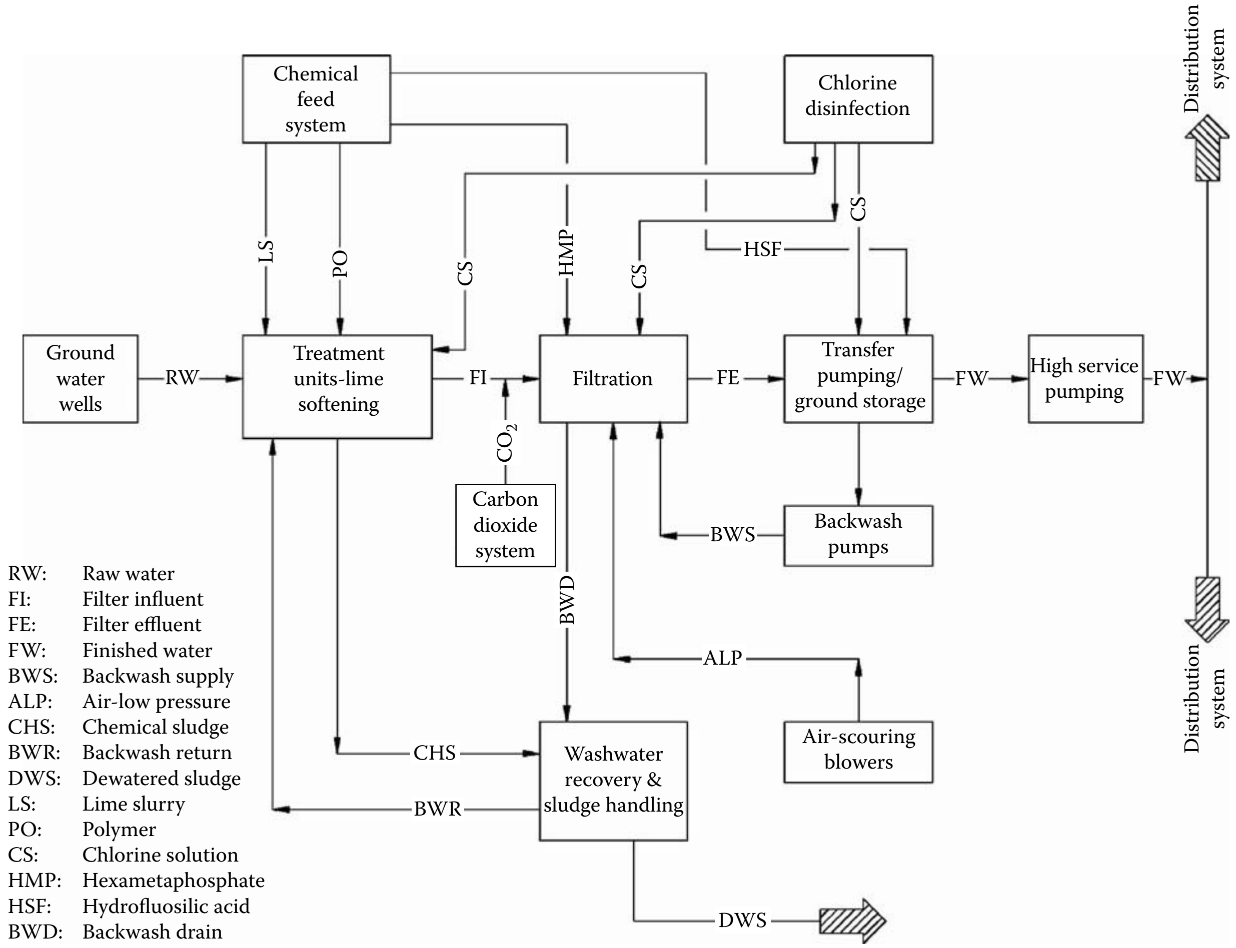


FIG. 8.40a
Block diagram of the unit processes of a drinking water supply plant that uses groundwater as its source of water.

in a main control panel, MCP-1. The redundant PLC communicates with eight remote I/O racks over a self-healing fiber-optic data link. Each remote I/O rack is housed in a local control panel (LCP) and receives I/O from a pair of filters. Thus, the system controls a total of 16 filters via eight I/O racks. One local I/O rack is installed in the main panel to receive common signals related to backwash pumps and air-scouring blowers.

Logic for sequencing of backwash and filter effluent valve control for each filter is resident in the main redundant PLC. The main PLC communicates with the plant computer system via fiber-optic Ethernet links. This allows the plant computer system HMI to display dynamic displays for all filters and to perform SCADA functions like changing set points. The main PLC panel MCP-1 also includes an industrial PC-based operator interface (OI-P) that allows monitoring and control of all filters as well as balance of the plant unit processes via the Ethernet link to the plant computer system. Each local panel has a local operator interface (OI-F) that allows monitoring and local manual control of each pair of filters from the dedicated panel. Other filters can also be monitored from each local operator interface (OI-F).

Filter Control and Sequencing The control and backwash sequencing logic associated with Figure 8.40b is typical of PLC-based constant-rate filtration installations, where the level in the filter is controlled by modulating the effluent flow control. The operating states of the filter include “auto filtration” (Modes 1 and 2) and “backwash.”

When a filter is in its “Auto Filtration” mode, the PLC controls the opening of the effluent valve (FCV-X-3 in Figure 8.40b) as a function of the level in the filter. If the level is very low, a low–low alarm is actuated and the effluent valve closes.

If the level (LIT-X-9) is to be kept between $L_{\min}$ and $L_{\max}$ (say, 7.0 and 8.0 ft) and the filter flow rate is adjustable between $F_{\min}$ and $F_{\max}$ (say, 1.0 and 2.0 million gal per day, or MGD), in *Mode 1 (gap control)* the PLC would set the effluent flow as follows: If level is at or below $L_{\min}$, the flow is set at $F_{\min}$ and stays there until the level reaches $L_{\max}$. At that point, the PLC switches the flow set point to $F_{\max}$, which remains there until the level drops back to $L_{\min}$.

In *Mode 2 (level-proportional),* if the actual level is $x\%$ above $L_{\min}$ (for example, at 7.6 ft, the value of $x = 60\%$), the PLC adjusts the flow set point to $F_{\min}(100 + x)$, or to 1.6 MGD.

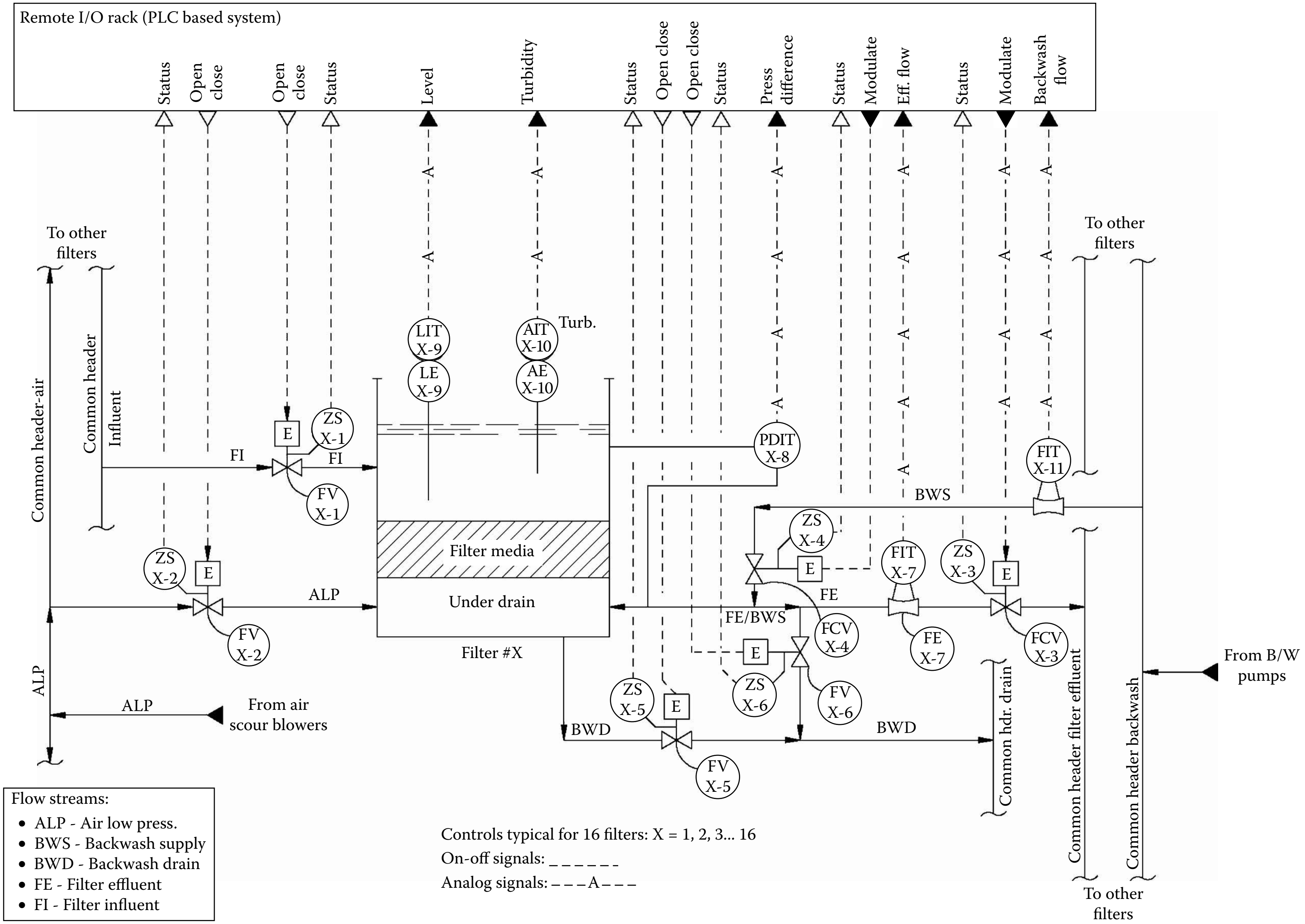


FIG. 8.40b
The sensors and valving of one of 16 filters having PLC-based sequencing controls.

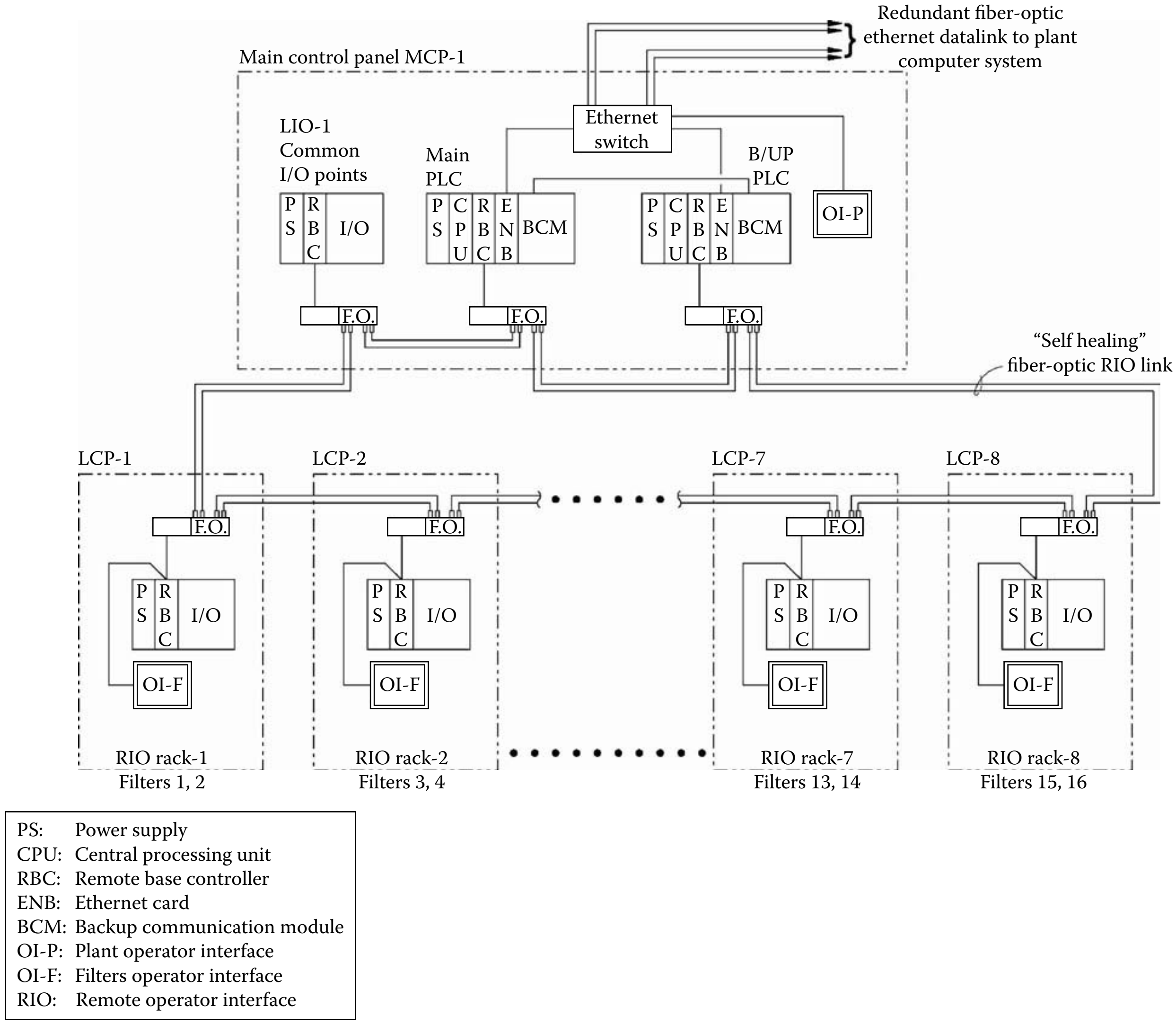


FIG. 8.40c

PLC main and subpanels are connected by "self-healing" fiber-optic links and with redundant FO Ethernet links to the plant computer.

If the effluent valve opening reaches 90% during the auto filtration phase, this initiates an alarm condition, indicating that the pressure available for valve differential is becoming insufficient, probably because the drop across the filter bed is high (PDIT-X-8 in Figure 8.40b). Under such conditions, the operator might initiate a backwash cycle.

Backwash Cycle The sequence of the backwash cycle can be initiated either manually by the operator or automatically by the PLC. The PLC controls the operation of a number of filters (for example, 16) and will allow only one filter to be backwashed at a time.

Automatic backwash is initiated by the PLC, if the preset limits are exceeded on any of the following process variables (Figure 8.40b): level (LIT-X-9), turbidity (AIT-X-10), pressure drop across filter media bed (PDIT-X-8), or when the total run time reaches a preset limit. Table 8.40d lists the steps that occur in a sequence during the automatic backwashing of a filter.

Transfer Pumping and Ground Storage

As shown in Figure 8.40a, the filtered effluent (FE) flows into a clear well, where it is further disinfected by a chlorine solution (CS) until the required residual chlorine concentration is reached in the finished water (FW). For detailed discussions of chlorination controls, refer to Sections 8.31 and 8.39.

Transfer pumps are used to pump the finished water from the clear well to the ground storage tanks, before it is sent into the water distribution system. The transfer pumping station consists of both constant- and variable-speed pumps and is operated under level control. For a detailed discussion

TABLE 8.40d
Definitions of the Steps in the Auto-Backwash Sequence

Step No.	*Description*
0	Out of service
1	Filtration
2	Washwater recovery basin not at high level
3	Close influent valve (FV-X-1)
4	Open drain valve at Falling Level 1 (FV-X-5)
5	Close effluent valve at Falling Level 2 (FCV-X-3)
6	Wait for preset time
7	Start air scour blower
8	Open air supply valve FV-X-2
9	Air scour at preset airflow for preset time
10	Enable washwater control valve (FCV-X-4)
11	Start backwash pump
12	Backwash at low rate with air scour
13	Open air vent valve at Rising Level 3 and close air supply valve (NOTE 2)
14	Stop air scour blower
15	Backwash at mid-rate for preset time (NOTE 1)
16	Backwash at high-rate for preset time (NOTE 1)
17	Backwash at mid-rate for preset time (NOTE 1)
18	Backwash at low rate for preset time (NOTE 1)
19	Close washwater control valve (FCV-X-4)
20	Stop backwash pump
21	Wait for preset time
22	Close drain valve (FV-X-5)
23	Open influent valve (FV-X-1)
24	Open filter-to-waste valve at Rising Level 4
25	Close filter-to-waste valve after preset time (FV-X-6)
26	Select standby/in filtration mode
27	Filter back in filtration

Notes:

1. Low, medium, and high backwash flow rates are set by a flow control algorithm, which detects the actual measured flow (FIT-X-11), changes the flow set point as required and modulates the control valve (FCV-X-4) to meet these set point settings.
2. The blowers are positive displacement (PD) blowers; therefore, the vent valve at the blower discharge header must be opened before closing the air scour valve FV-X-2.
3. All valve positions are confirmed via limit switches before advancing to next step.

of pump station control and optimization, refer to Sections 8.34 and 8.35.

High-Service Pumping

The finished water from the storage tanks is pumped into the distribution system by the high-service pumping station, which consists of both variable- and constant-speed pumps. The flow and residual chlorine concentration in the distribution header is continuously monitored, while the pressure is controlled.

The high-service pumps are controlled and sequenced to maintain the pressure in the distribution header within a ±2 psi band around the desired pressure set point. A PID algorithm is provided in the PLC to modulate the variable-speed pumps to maintain this pressure set point.

In order to properly design a pumping station, one needs to develop reliable system curves for the process, obtain the characteristic curves of the individual pumps, and identify such operating constraints as the surge lines. Based on such information, static and dynamic models can be developed and used to simulate pump station operation using a variety of alternate pump combinations and strategies, before they are implemented.

The key to success is using reliable data to obtain the correct system and pump curves for all possible combinations of load conditions. The goal is to avoid overdesigning, which can cause excessive pump cycling, pressure fluctuation, and waste of pumping energy, yet provide all the capacity that is needed to meet present and future loads. For a detailed discussing of pump station control and optimization, refer to Sections 8.34 and 8.35.

OZONATED WATER TREATMENT

The conventional water treatment system shown in Figure 8.40a utilizes disinfection steps using chlorine. Figure 8.40e describes the configuration of the equipment in a water supply plant where ozone is used for disinfection.

The plant that is illustrated has two ozone generators, each supplying a common header, which takes the ozone to two ozone contactors. Each contactor supply line serves five ozone diffusers, which mix the ozone with the raw water received from sedimentation basins. The ozonated water effluent (OZW) from the contactors is sent for filtration to the biologically activated carbon (BAC) media filters. The excess ozone (OZA) that accumulates above the water surface in each of the contactors is removed and destructed.

Each ozone generator includes a power supply for the generation of high-amplitude, high-frequency voltage, which is applied to the electrodes in the ozone generator, which ionizes the oxygen-rich air (GOX) supplied to the ozone generator. The frequency of the power supply unit output is modulated to vary the rate of ozone production. An ozone production turndown of 10:1 can be achieved by varying the frequency from 1000 to 100 Hz.

In another method of ozone production, the frequency is kept constant and the ozone production rate is adjusted by varying the amplitude of the voltage. The ozone production rate can be modulated either in proportion to the raw water flow rate (feedforward) or to maintain the desired ozone

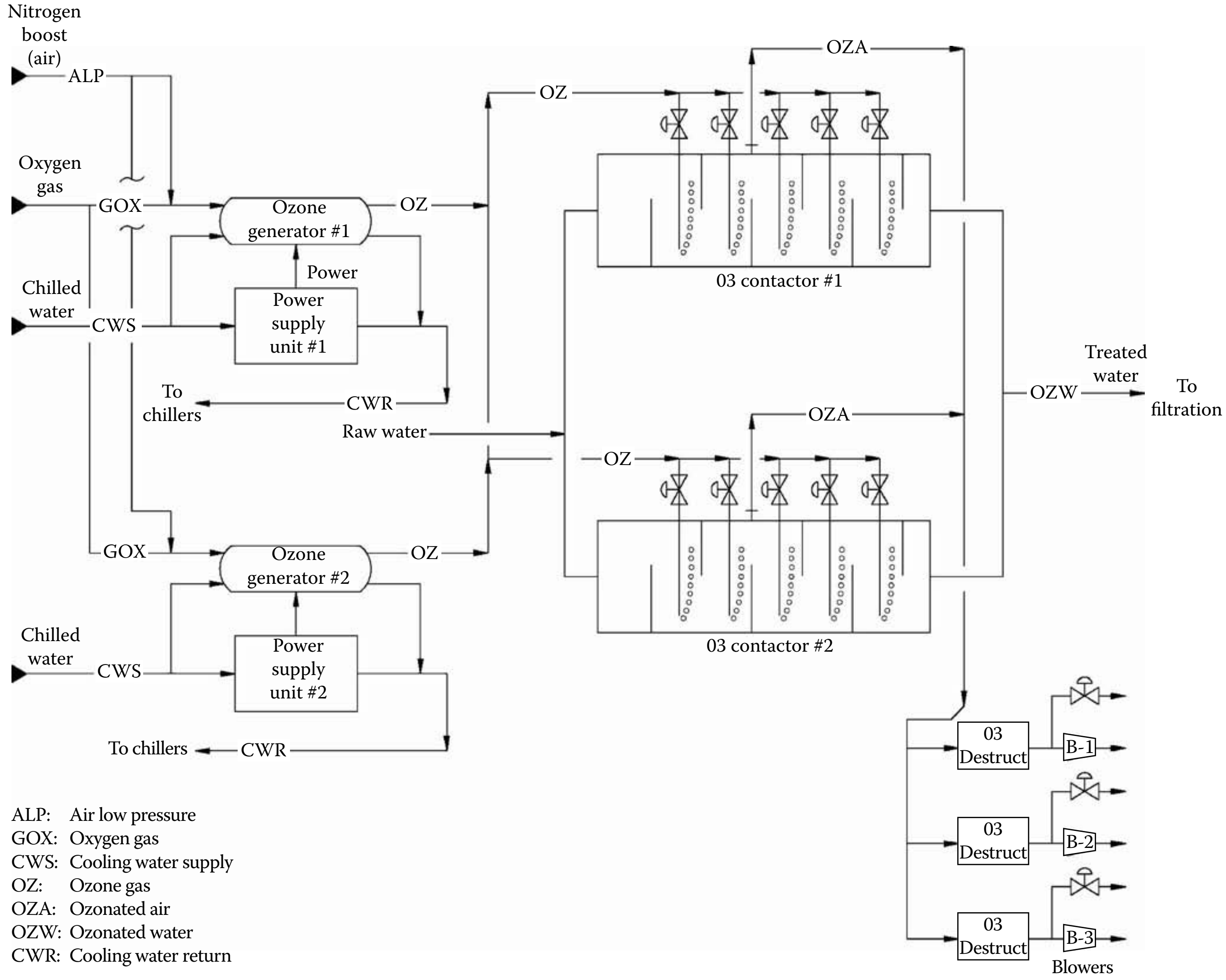


FIG. 8.40e

The configuration of the unit processes in a drinking water supply plant, where ozone is utilized as the disinfectant.

concentration in the water effluent (feedback). In either case, the control signal modulates the power supply inverter circuits.

As shown on Figure 8.40e, both the power supply units and the ozone generators generate significant amount of heat. Chillers and cooling water circulation pumps are provided to remove this heat and to cool this equipment to the desired operating temperature.

Ozone Generator Controls

Figure 8.40f illustrates how a PLC system might approximate, by a series of simple calculations, the operation of a cascade control loop. In this case, the cascade master is the ozone pressure to the contactors (PIC-1-2) and the cascade slave is the oxygen flow (FIC-1-1) into the ozone generator.

As illustrated in Figure 8.40f, the oxygen supply to the ozone generator can be under either "pressure" or "flow" control. The operator can select (FY-1-1-2) the output of either the pressure controller (PIC-1-1) or the flow controller (FIC-1-1) to modulate the oxygen supply valve PCV-1-1. Depending on that selection, either the pressure or the flow in the oxygen supply header will be held on set point by the corresponding PID algorithm in the PLC.

The set point for the oxygen supply pressure controller is set by the operator.

Calculating the Slave Flow Controller Set Point In the following six steps, a sequence of simple mathematical calculations will be described. This procedure illustrates how PLC-based systems can approximate the operation of cascade control systems.

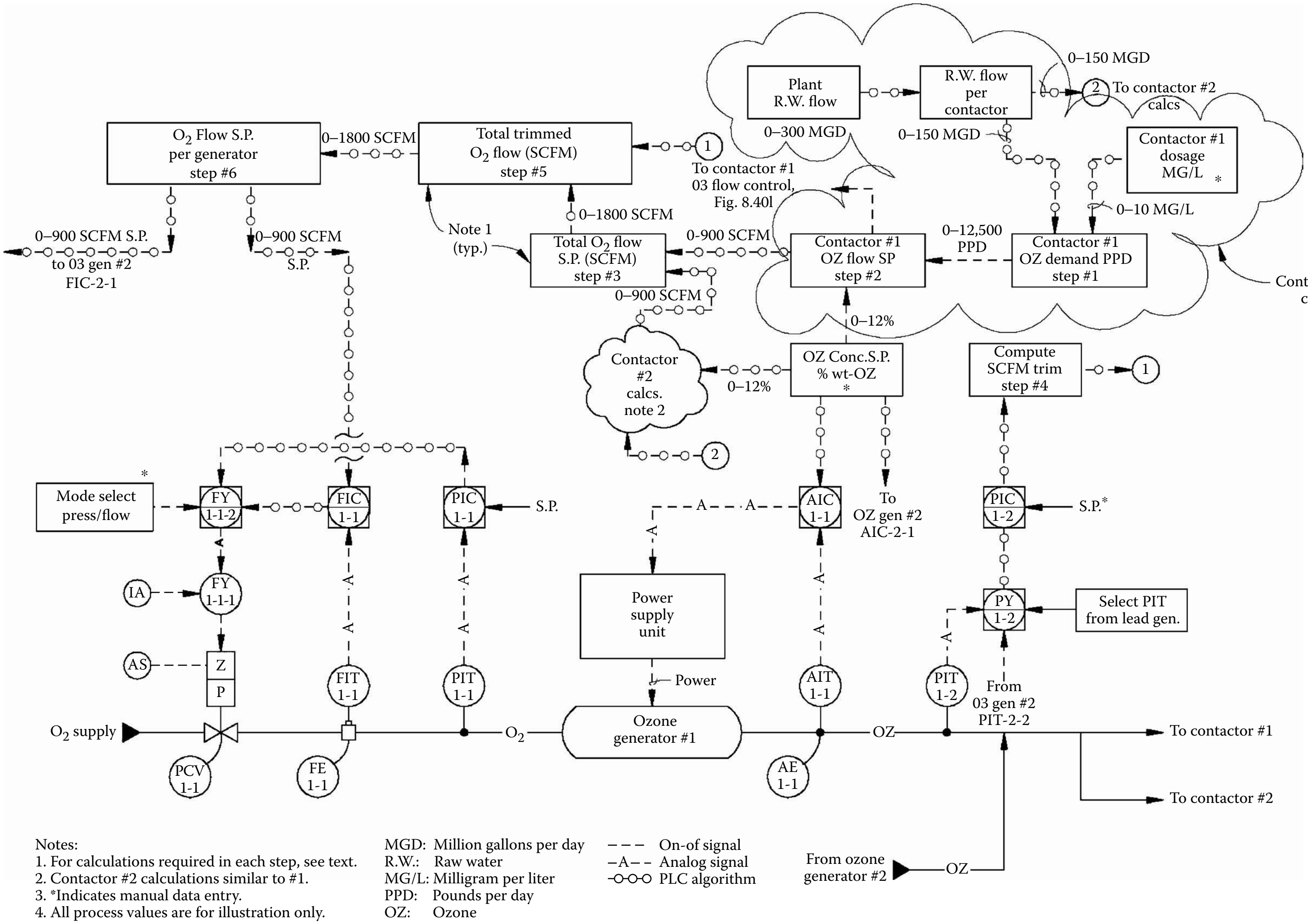


FIG. 8.40f

Ozone concentration is controlled by modulating the frequency of the power supply, while the oxygen feed rate is throttled either on the basis of holding the inlet or the discharge pressure of the generator constant.

The set point of the oxygen flow controller to the ozone generator is calculated on the basis of raw water flow, ozone dosage requirement, and ozone concentration measured in the effluent as follows:

STEP 1: $\text{PPD} = \text{RWF} \times \text{Dosage} \times 8.34$ **8.40(1)**

where

PPD is pounds per day of ozone demand in Contactor 1
RWF is raw water flow per contactor in million gallons per day (mgd)
Dosage is the required ozone dosage in mg/liter

STEP 2: $O_3\,\text{Flow\#1} = [(\text{PPD})/(O_3\%\ \text{wt} \times .083 \times 1440)] \times 100$ **8.40(2)**

where

O_3 Flow#1 is the O_3 flow set point in SCFM for the flow into Contactor 1
O_3% wt is the set point for the ozone concentration in percentage by weight
Note: The ozone flow S.P. for Contactor 2 is calculated the same way as for Contactor 1.

STEP 3: $\text{Total } O_2 \text{ Flow} = O_3\,\text{Flow\#1} + O_3\,\text{Flow\#2}$ **8.40(3)**

where

Total O_2 Flow is the set point for FIC-1-1 in SCFM.

STEP #4: $O_2 \text{ Flow Trim} = [(\text{PIC-1-2 output})/100] \times \text{Trim Range} - (\text{Trim Range}/2)$ **8.40(4)**

where

O_2 Flow Trim is the set point trimming signal in SCFM, received from PIC-1-2.
Trim Range is (−100 SCFM to +100 SCFM) = 200 SCFM.
PIC-1-2 output is the percentage output signal of the ozone generator discharge pressure controller PIC-1-2. This controller algorithm in the PLC detects the ozone pressure of the lead ozone generator (PIT-1-2 or PIT-2-2) and compares that reading to an operator-entered pressure set point.

STEP #5: Total Trimmed O_2 Flow S.P. $= \text{Total } O_2 \text{ Flow (Step \#3)} + O_2 \text{ Flow Trim (Step \#4)}$ **8.40(5)**

where

All flow set points and trim are in SCFM.

STEP #6: O_2 Flow S.P. per generator $= \text{Total Trimmed } O_2 \text{ Flow S.P. (Step \#5)}/2$ **8.40(6)**

Ozone Concentration Control Loop AIC-1-1 in Figure 8.40f controls the percentage concentration of ozone by weight in the ozone generator discharge stream. This controller receives its measurement signal from the ozone analyzer AIT-1-1 and maintains the ozone concentration at the operator-entered set point (percentage) by modulating the output frequency of the inverter part of the power supply unit.

Ozone Flow Distribution Control

Figure 8.40g describes the ozone flow distribution control loops. The set point for the ozone flow to Contactor 1 is calculated by using Equation 8.40(2), described in Steps #1 and #2 earlier. This total ozone flow is distributed among five diffuser tubes in Contactor 1.

Each ozone feeding tube is provided with its own flow transmitter, controller, and control valve. A predetermined percentage of the total ozone flow is sent as the set point of the flow controller of each ozone diffuser tube. The percentages can be changed by the operator; they are provided in a lookup table, and their total must add up to 100%.

Control of Ozone Destruction

As was noted in connection with Figure 8.40e, the excess ozone that accumulates above the water surface in the contactors is exhausted through ozone destructors, which heat it and, by the use of a catalyst, reconvert it back to oxygen.

Each ozone destructor is provided with a continuously running exhaust blower, which maintains a slightly negative pressure within the ozone contactor. A single pressure controller (PIC-4 in Figure 8.40h) maintains a slight vacuum, which is measured at the outlet of the “lead” contactor (PIT-1-4-1 or 2-4-1) and modulates all three vacuum breaker valves (FCVs).

In the destructor, the ozone is preheated to a preset temperature and is converted back to oxygen by a catalyst. An SCR-based solid-state controller controls the heater by changing the SCR firing angle to control the amount of heat generated.

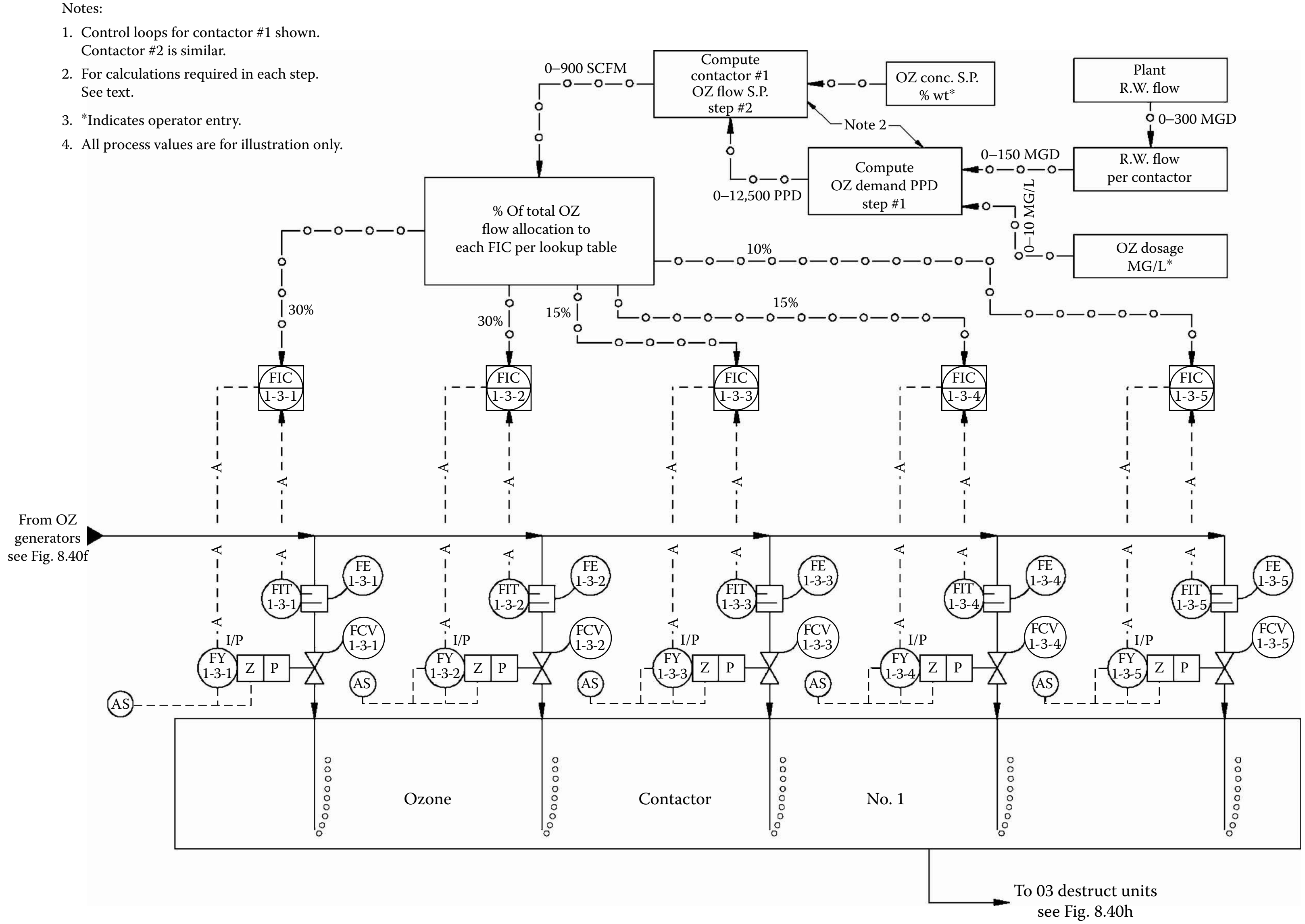
The ozone destructors must be on and running properly before any ozone gas is generated.

Ozone Generator Start-Up Sequencing

The PLC logic also includes an ozone generator automatic sequencer, which initiates and carries out the system start-up and shutdown in a safe manner. Certain prerequisite conditions must be satisfied before an automatic start-up or shutdown sequence can be initiated.

Prior to initiating the automatic start-up sequence of an ozone generator, the following major prerequisites must be satisfied: 1) At least one of the two oxygen supply valves of the two ozone generators (PCV-1-1 or PCV-2-1 in Figure 8.40f) must be in automatic control. 2) At least one of the two ozone generators must be in automatic and ready for operation. 3) At least one of the two contactor trains must be enabled and ready for operation. 4) Emergency stop has not been actuated.

When all of the above prerequisites are satisfied, the “Ready” status is displayed and a run command can be



Notes:

1. Control loops for contactor #1 shown. Contactor #2 is similar.
2. For calculations required in each step. See text.
3. *Indicates operator entry.
4. All process values are for illustration only.

FIG. 8.40g

Ozone flow controls for distributing the gas among five diffuser tubes.

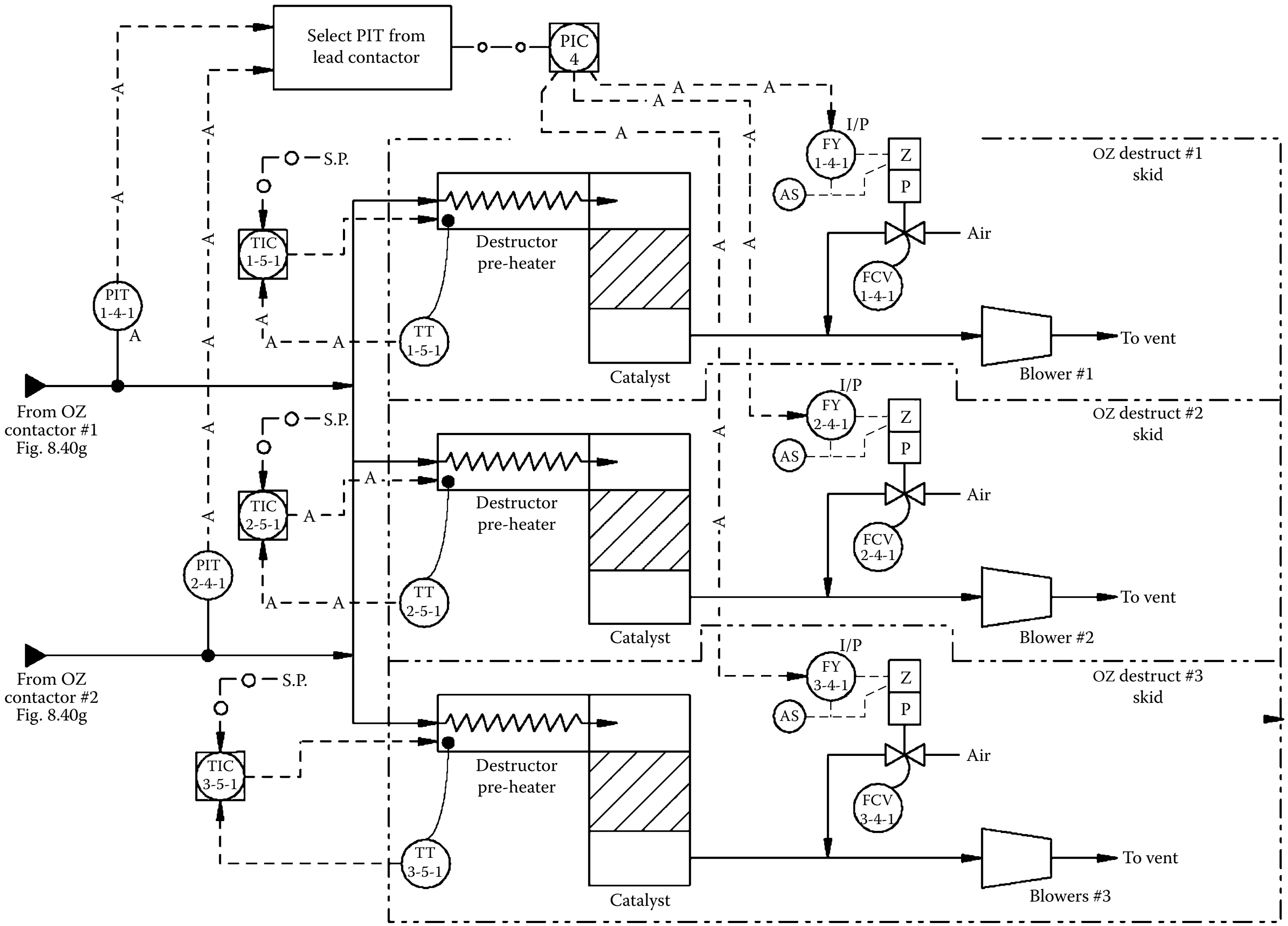


FIG. 8.40h
The ozone is preheated and its vacuum pressure is controlled by throttling FCV valves that admit air into the suction of the blowers.

TABLE 8.40i
Steps in the Ozone Generator Start-Up and Shutdown Sequence

Step 0 — System off, idling. System ready. System run command issued.
Step 1 — Start cooling water pump.
Step 2 — Start chiller.
Step 3 — Start boost air.
Step 4 — Start destruct unit and enable controls.
Step 5 — Open oxygen valves.
Step 6 — Enable ozone flow control loops.
Step 7 — Start ozone generator purge.
Step 8 — Start generator, enable generator control loops.
Step 9 — Enable contactor flow controls, system running.
**Issue system stop command.
Step 10 — Stop generator, disable control loops, purge.
Step 11 — Stop purge.
Step 12 — Disable ozone flow controls.
Step 13 — Close oxygen valves, wait, destructor purge.
Step 14 — Stop destructor.
Step 15 — Stop boost air.
Stop 16 — Stop chiller.
Step 17 — Stop cooling water pump.

issued. Once a run command is issued, the system sequencer goes through a series of steps to start the system. Automatic shutdown is similarly sequenced. The major start-up and shutdown steps in the sequence are listed in Table 8.40i.

As each step is taken in Table 8.40i, the actions of the previous step are verified and confirmed before the current step is initiated.

PLC Hardware

Figure 8.40j shows the hardware of a typical PLC-based control system for ozone generation and disinfection. The control components include a main control panel (MCP) that houses a redundant PLC, a local I/O rack, and an operator interface (industrial PC).

Each generator has a dedicated control panel, OGCP-1 and OGCP-2, housing a PLC, I/O modules, and an operator interface. Each ozone generator is controlled by its PLC with all related sensors connected to the PLC I/O modules. The chillers, contactors, destructors, oxygen flow controls, and overall sequencing is controlled by the redundant PLC part of the main control panel, and the local I/O rack in the main panel receives all signals related to everything else except the ozone generators.

The operator interfaces in each of the panels allow local control and monitoring of the system. The main redundant PLCs, the two generator PLCs, and the three operator interfaces communicate over a high-speed peer-to-peer data link.

The main redundant PLC communicates with the overall water plant computer system via fiber-optic Ethernet links. The Ethernet communication allows SCADA functions from the central control room.

REVERSE OSMOSIS-BASED WATER TREATMENT

A water treatment plant utilizing the reverse osmosis (RO) process includes a raw water pretreatment section, an RO membrane train section, and a post-treatment with transfer pumping subsystem.

Raw Water and Pretreatment

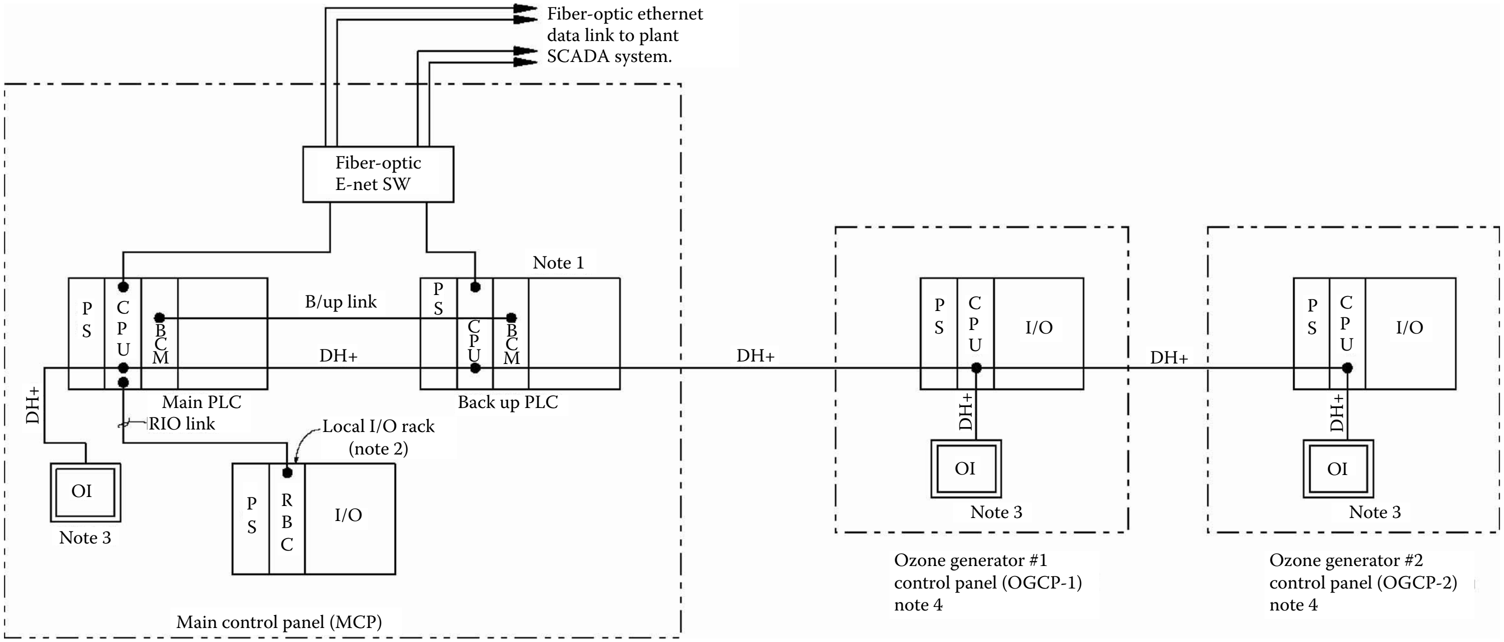
As shown in Figure 8.40k, this subsystem consists of the raw water wells from which the raw water is pumped into the membrane trains, after chemical pretreatment and filtering. Chemical pretreatment consists of the injection of acid for pH control and of scale inhibitor, which prevents scaling inside the membranes.

The starting/stopping and the adjustment of the speed of the raw water well pumps is automatically controlled to maintain the cartridge filter effluent pressure (PT-1-2-2 in Figur 8.40k) and a "well pump sequencer." The following data is manually entered into the pump sequencer: well number, pump number, the order in which pumps should be started, target speed, target flow, maximum flow, and well water conductivity. The real-time data used by the pump sequencer includes the pump status (on/off), actual speeds, flows, and well levels.

Based on the well levels, water conductivity and other process parameters, the operator manually selects the target speeds and flows for each pump and enters this data in the pump sequencer, which then becomes part of the PLC logic. The operator also enters the pump-starting priority (1–8) in the sequencer logic.

In the automatic mode, the well pumps automatically start to maintain the cartridge filter effluent (CFE) pressure in their predetermined sequence and operate at speeds determined by the pump sequencer. As the pressure rises, the pumps are automatically stopped in reverse order, determined by the sequencer. If a pump is switched into the manual mode, it can be started and stopped manually, and its speed can be set by the operator.

Acid and Scale Inhibitor Controls As shown in Figure 8.40k, acid is injected into the raw water, upstream of the mixer and cartridge filters, and after the sulfuric acid has been mixed with the raw water, the resulting pH is measured downstream of the mixer (AIT-1-5). The stroke of the acid metering pump is adjusted as a function of the raw water flow (FT-1-1) using a look-up table. The speed of the acid pump is set by the pH controller, AIC-1-5.



PS:	Power supply
CPU:	Central processing unit-PLC
BCM:	Backup module
RBC:	Remote base communication module
OI:	Operator interface
DH$^+$:	High speed datahiway, peer to peer link

Notes:

1. Redundant PLC with hot backup.
2. Local I/O rack receives signals from ozone destruct units, chillers & other auxiliary equipment.
3. Operator interface is touch screen, industrial environmentally hardened personal computer.
4. Each ozone generator includes a dedicated PLC communicating with main PLC via DH$^+$.

FIG. 8.40j

PLC hardware configuration used for the ozone generation and disinfection processes.

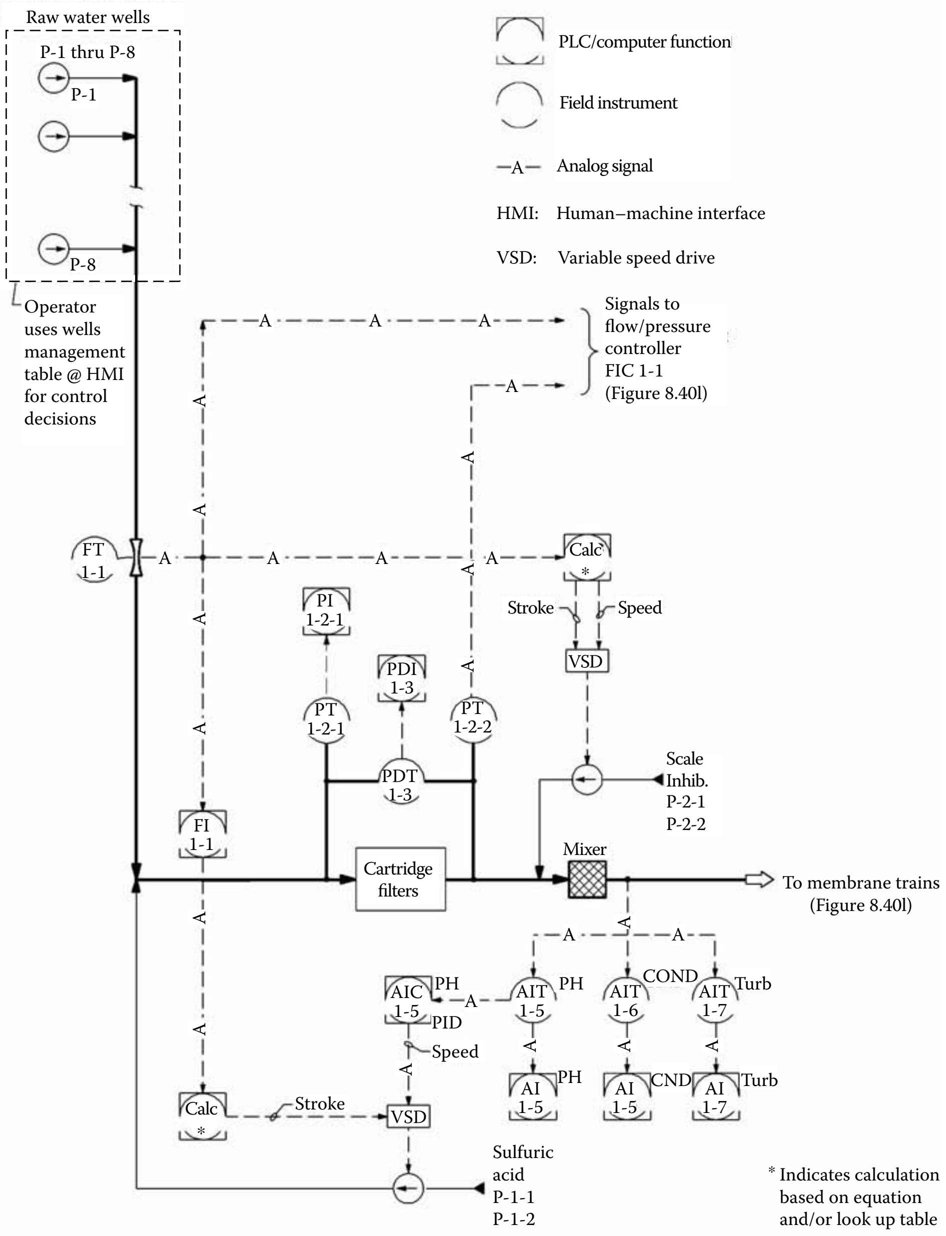


FIG. 8.40k
Raw water pretreatment controls.

The scale inhibitor metering pump is controlled to keep the scale inhibitor concentration (mg/l) in the raw water constant. Here, too, the stroke position of the metering pump is adjusted as a function of the raw water flow (FT-1-1) using a look-up table. The speed of the scaling inhibitor pump is determined by the PLC, using Equation 8.40(7):

$$\%\ \text{speed} = [\ (F \times D \times Ki)/(\#\ \text{of Pumps} \times \%\ \text{Stroke})\ ] \times 100 \tag{8.40(7)}$$

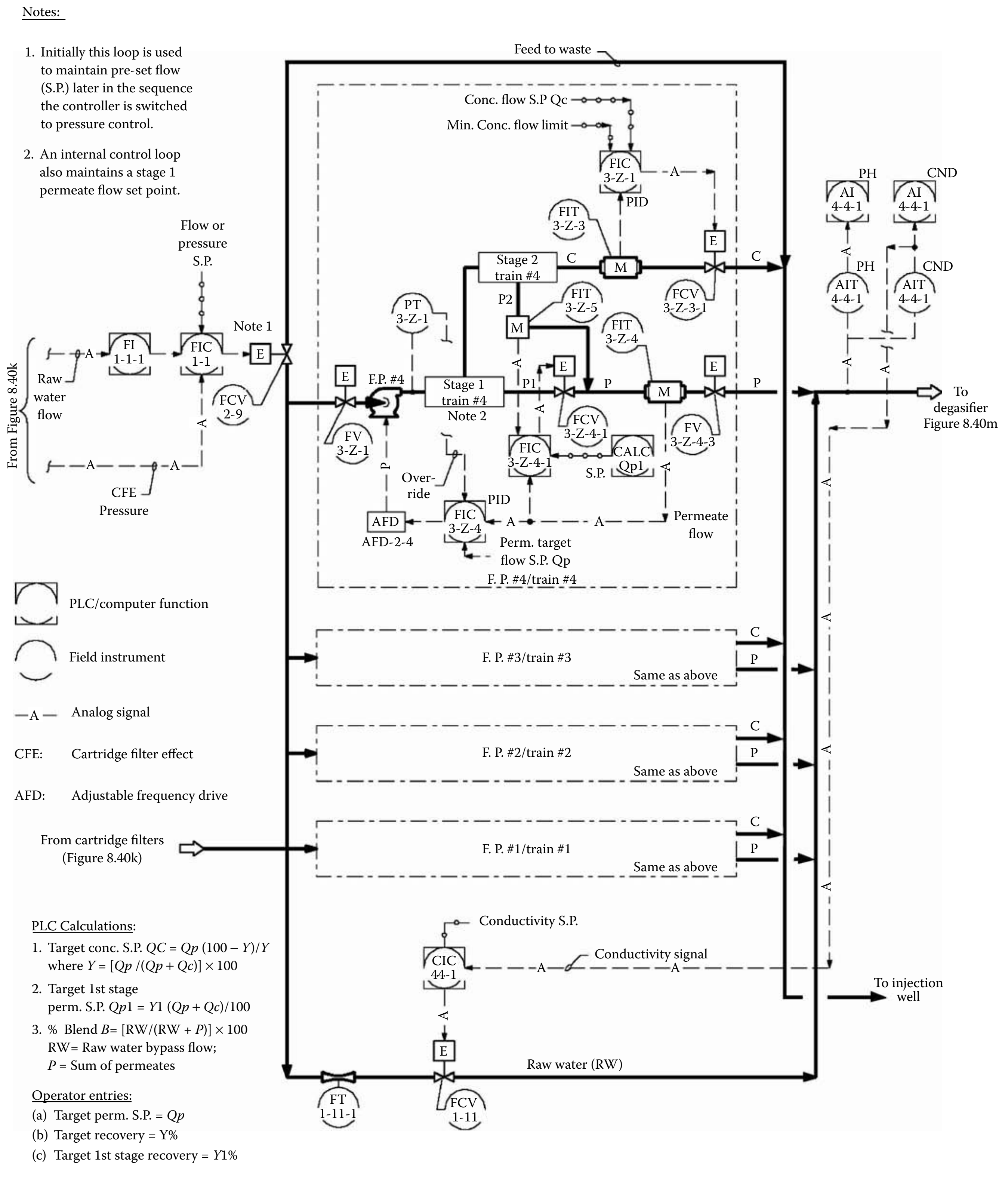
where

F is the raw water flow in mgd
D is the desired inhibitor dose in mg/l
Ki is a constant; in this case, its numerical value is 0.0092

The conductivity and turbidity of the conditioned water feed to the membrane trains is displayed for the operators. The water is allowed to reach the membranes only, if the water quality is within acceptable limits. Any excursions are alarmed, and in extreme conditions, trains can be automatically shut down.

R.O. Membrane Train Controls

Figure 8.40l illustrates a membrane train configuration, where there are four trains working in parallel. Conditioned water from the pretreatment system is received through a common header that supplies the membrane feed pumps.



Notes:

1. Initially this loop is used to maintain pre-set flow (S.P.) later in the sequence the controller is switched to pressure control.
2. An internal control loop also maintains a stage 1 permeate flow set point.

PLC Calculations:

1. Target conc. S.P. $QC = Qp\,(100 - Y)/Y$ where $Y = [Qp/(Qp + Qc)] \times 100$
2. Target 1st stage perm. S.P. $Qp1 = Y1\,(Qp + Qc)/100$
3. % Blend $B = [\mathrm{RW}/(\mathrm{RW} + P)] \times 100$
   RW= Raw water bypass flow;
   P = Sum of permeates

Operator entries:

(a) Target perm. S.P. = Qp
(b) Target recovery = Y%
(c) Target 1st stage recovery = $Y1$%

FIG. 8.40l
The control loops of a number of parallel trains of reverse osmosis membrane unit include the pressure and flow controls of the water supply, the individual concentrate and permeate flows, and the bypass flows to waste and for blending.

Each variable-speed membrane feed pump feeds conditioned water to the corresponding train.

The permeate flow from each train is taken by a common header to a degasifier and, after that, into a clear well. Concentrate from each train is taken by a common header for dumping into injection well. Each train consists of two stages. In Figure 8.40l, the overall permeate and concentrate flows are designated as "*P*" and "*C*," respectively. Stage 1 and Stage 2 permeate flows are designated as *P*1 and *P*2, respectively.

A common feed-to-waste bypass allows the conditioned water to be sent to waste into an injection well by control valve FCV-2-9 during initial start-up or during normal operation, should a high-pressure condition arise. Another bypass allows raw water to be blended into the permeate by control valve FCV-1-11 under conductivity control.

Concentrate and Permeate Flow Controls The target percentage of overall water recovery (Y) can be calculated by Equation 8.40(8). The values of Y, Q_P, and Q_{CMIN} are all entered by the operator into the PLC:

$$Y = [Q_P/(Q_P + Q_C)] \times 100 \qquad \textbf{8.40(8)}$$

where

Y is the target percentage of overall water recovery
Q_P is the target set point for the overall permeate flow
Q_C is the target set point for the overall concentrate flow
Q_{cMIN} is the minimum acceptable overall concentrate flow

The set point for the overall concentrate flow (Q_C) is calculated by Equation 8.40(9) as follows:

$$Q_C = Q_P\,(100 - Y)/Y \qquad \textbf{8.40(9)}$$

where

Y is the recovery in percentage (%)

If $Q_C \Lleftarrow Q_{CMIN}$ then Q_{CMIN} is used as the overall concentrate flow set point.

The overall permeate flow set point Q_P is maintained by FIC-3-Z-4, which is controlling the speed of the variable-speed feed pump. If, during operation, a high-pressure condition is detected (PT-3-Z-1), the output of FIC-3-Z-4 is frozen and an alarm is sounded.

The overall concentrate flow set point Q_C is maintained by FIC-3-Z-1, which throttles FCV-3-Z-3-1. If, during operation, the concentrate flow drops below a minimum preset value, the output of FIC-3-Z-1 is frozen and an alarm is sounded.

First-Stage Permeate Flow Control The set point for the permeate flow rate through the first stage (Q_{P1}) is calculated by Equation 8.40(10):

$$Q_{P1} = Y_1\,(Q_P + Q_C)/100 \qquad \textbf{8.40(10)}$$

where

Y_1 is the target percentage of the first-stage recovery, which is entered by the operator
Q_P and Q_C are overall permeate and concentrate flow set points, respectively

The first-stage permeate flow (FSPF) is calculated as the difference between the overall permeate flow (FIT-3-Z-4) and second-stage permeate flow (FIT-3-Z-5).

$$\text{FSPF} = [(\text{FIT-3-Z-4}) - (\text{FIT-3-Z-5})] \qquad \textbf{8.40(11)}$$

In Figure 8.40l, the flow controller FIC-3-Z-4-1 keeps the permeate flow through the first stage at the set point that was calculated in Equation 8.40(10).

Feed-to-Waste Flow Control During the initial start-up of a train, the raw water has to be bypassed into the injection well until the quality of the conditioned raw water meets all water quality requirements. Therefore, during the initial start-up, FIC-1-1 controls valve FCV-2-9 to keep the bypass flow equal to one train's rated flow. The raw water flow transmitter (FT-1-1 in Figure 8.40k) provides the measurement signal for this control loop.

Later in the start-up sequence, once the conditioned raw water meets the quality requirements, FIC-1-1 is switched to control the pressure detected by PT-1-2-2 in Figure 8.40k. When operating in this configuration, the bypass is used for overpressure relief only, and because the normal pressure is very much below the pressure set point of FIC-1-1, the feed-to-waste valve is closed during normal operation.

Raw Water Blend Control This control loop at the bottom of Figure 8.40l serves to blend the permeate product with conditioned (bypassed) raw water, in order to keep the conductivity of the blend that is being sent to the clear well constant. The conductivity of the permeate/raw water blend (detected by AIT-4-4-1 in Figure 8.40l) is controlled by CIC-4-4-1, which determines the blend percentage (B) by modulating the raw water bypass flow (RW). The PLC determines the blend ratio by using Equation 8.40(12) as follows:

$$B = [(\text{RW})/(\text{RW} + P)] \times 100 \qquad \textbf{8.40(12)}$$

where

B is the percentage blend that is calculated
RW is the raw water bypass flow, measured by FT-1-11-1
P is the sum of the permeate flows of all running trains

If the calculated bypass blend (B) is more than or equal to a preset maximum (say, 15%) then the CIC-4-4-1 controller's output is frozen, the bypass valve remains in its last position, and an alarm is sounded. Even during automatic operation, the operator always has the option of switching the controller in manual and modifying the bypass flow to maintain a blend ratio.

Pumping and Post-Treatment Controls

Figure 8.40m describes the process of transfer pumping and post-chemical treatment applied prior to the water being sent to storage tanks for distribution.

The blended permeate from the membrane trains is sent into the clear well via the degasifier. In the degasifier, hydrogen sulfide (H_2S) gas is removed from the water and passed through an odor control scrubber system, where the odor is eliminated before the cleaned gas stream is released to the atmosphere.

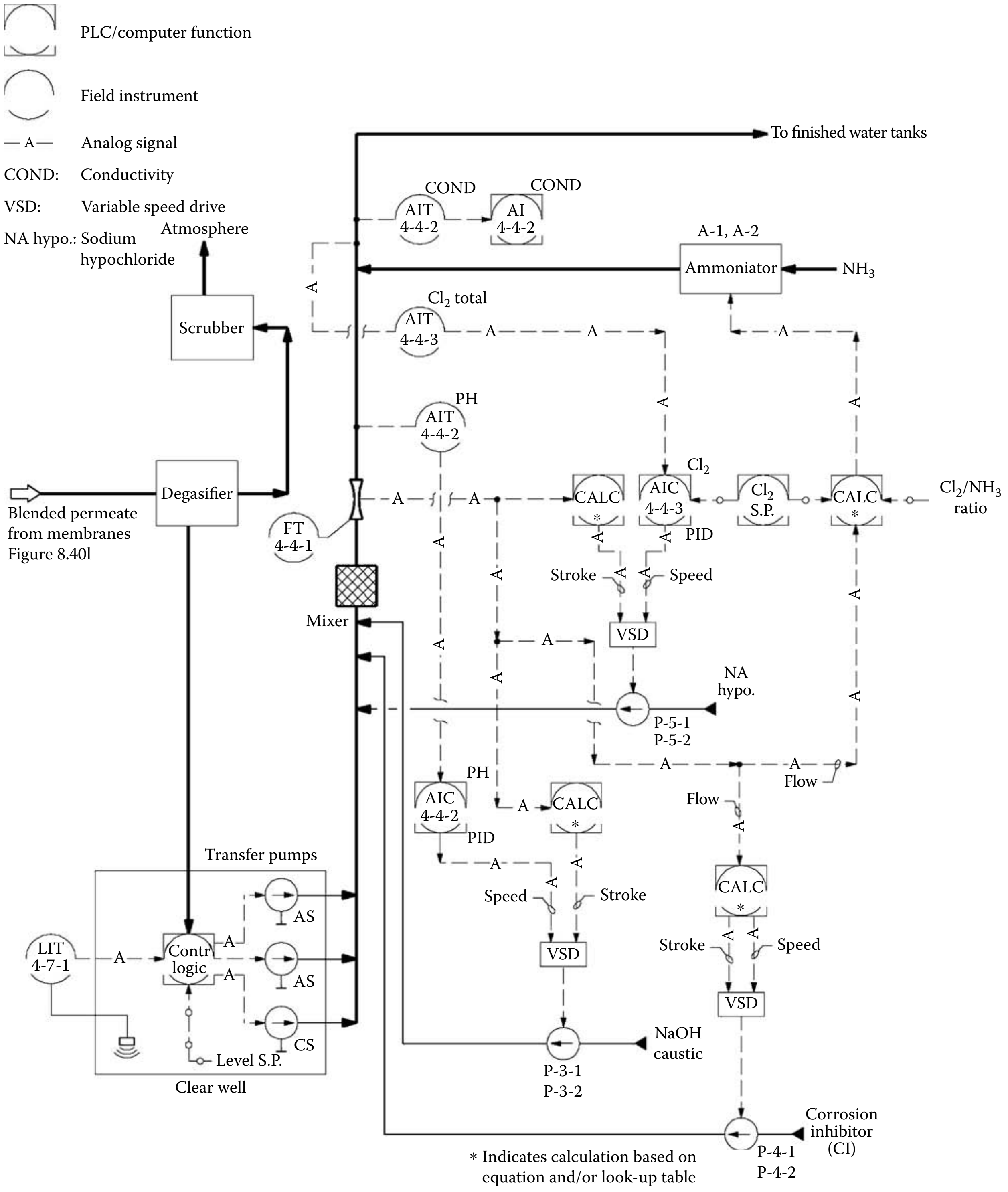


FIG. 8.40m

Transfer pumping and post-treatment controls include the adjusting of the injection rates of pH, residual chlorine, corrosion inhibitor, and ammonia controls.

The water from the clear well is pumped into storage tanks by the transfer pumps. A combination of variable- and constant-speed pumps are operated under level control in the clear well (LIT-4-7-1 in Figure 8.40m). The speed of the variable-speed pump(s) is modulated to maintain the level in the clear well level within a preset band.

The blended permeate from the clear well must be chemically treated before being transferred to the storage tanks. Therefore, the pH of the blended permeate is adjusted to bring it between the limits of 7.5–8.5 pH by the addition of sodium hydroxide (caustic) to the water line.

Similarly, sodium hypochlorite is added to maintain the required chlorine concentration (mg/l) for disinfection. Lastly, a corrosion inhibitor is also added to provide protection against corrosion. All three chemicals are added upstream to the inline mixer to guarantee thorough mixing.

Metering pumps are used to inject the above-mentioned chemicals. The strokes of all metering pumps are controlled according to the values stored in a look-up table to match the additive injection rate to the actual water flow rate (FT-4-4-1 in Figure 8.40m).

Additive Controls Both residual chlorine and pH are measured on-line, downstream of the mixer. The speed of the caustic metering pump is adjusted by AIC-4-4-2, which keeps the finished water pH between 7.5 and 8.5. The speed of the sodium hypochlorite pump is controlled by AIC-4-4-3, which maintains the required total chlorine concentration (mg/l) on set point.

The speed of the corrosion inhibitor pump is controlled in a similar manner as was described for the scale inhibitor metering pump on Figure 8.40k by Equation 8.40(7). Finally, ammonia is added to the finished water to maintain the chlorine-to-ammonia ratio in a range of 3 to 5. Equation 8.40(13) describes how the required feed rate of ammonia is calculated:

$$NH_3 \text{ Feed Rate} = [F \times \text{Hypochlorite Dose} \times 8.34]/\text{Ratio} \qquad \textbf{8.40(13)}$$

where

NH_3 Feed Rate is in lb/day

F is the flow rate of finished water (FT-4-4-1) in mgd

Hypochlorite Dose is in mg/l

Ratio is in the range from 3 to 5 and is entered by the operator

Sequence of Operation Controls

For the reverse osmosis train of equipment, the sequence of operation includes both manually initiated prerequisite steps and automatic steps. Several of the steps are subdivided into mini-sequences. The total reverse osmosis process is subdivided into some ten major sequence steps, listed in Table 8.40n.

TABLE 8.40n
Major Sequence Steps in the RO Treatment Process

Post Chemical Feed
Transfer Pumping/Degasifier/Scrubber
Pretreatment Chemical System
Raw Water Well Management and Automatic Sequence
Train Start-Up Check Lists
First Train Start-up
Subsequent Train Start-up
Automatic/Manual Train Shutdown
Plant Power Failure Scenarios
Train Flush Sequences

Before initiating an automatic train start-up, the operators use two types of checklists, which are displayed on the HMI screens. These are regular and train-specific checklists. The first general checklist display allows operators to verify pump selection, manual valve positions, and selector switch positions for the auxiliary subsystems, e.g., chemical systems, transfer pumping, and scrubber.

The second, train-specific checklist display allows operators to verify major valve positions, feed pump switch positions, controller modes, and manually entered set-point and target values. The automatic train start-up cannot be initiated unless all boxes in the checklists are checked and the correct manual data entries are made.

Start-Up and Shutdown Sequences In the train start-up sequences, there are two types of prerequisite interlocks. The permanent interlocks are active all the time, while the initial condition interlocks are active only during start-up.

Permanent prerequisite interlocks are safety interlocks that must be in place all the time, e.g., feed pump suction and train inlet valves must be open before starting a feed pump, or chemical injection valve must be open before starting a chemical pump.

Initial conditions are the prerequisites required only during the initial start-up, and these conditions can change once the train is in operation. As an example, such initial conditions are the start-up positions of such valves in Figure 8.40l as the raw water bypass valve (FCV-2-9), raw water blend valve (FCV-1-11), and train concentrate valves (FCV-3-Z-3-1).

The logic sequences for "first" train start-up are different from the logic sequences used for "subsequent" train start-up.

Train shutdown sequences can be either automatically initiated by process upsets or manually initiated by operators because of maintenance functions, such as for membrane cleaning or following alarm conditions. Membrane cleaning operation is a completely manual process that is performed by the operators with the aid of a local field panel.

Plant power failure scenarios are included in the sequence of operation. Control strategies are developed for situations when the power failure is only at the raw water wells and not at the plant or vice versa.

Nonoperating trains must be "flushed" every 24 hours and also after each power failure of significant duration. Flush sequences can be initiated automatically, for trains that are not in operation, or manually, based on operator commands.

ULTRAFILTRATION-BASED WATER TREATMENT

For certain types of raw waters, submerged ultrafiltration membranes can be used instead of conventional lime softening water treatment. Ultrafiltration membranes can be substituted for the processing steps of direct filtration, settled water treatment, TOC and color removal, and taste and odor removal, and for pretreatment of the raw water prior to reverse osmosis-based filtration.

Transmembrane Pressure

Equation 8.40(14) can be used to calculate the transmembrane pressure (TMP) difference:

$$\mathrm{TMP} = \mathrm{PP} + K \times [(A + B) - \mathrm{WL}] \qquad \mathbf{8.40(14)}$$

where (referring to Figure 8.40o)

PP is the permeate pressure in psi

A is the elevation of the pressure transmitter above the top of the membranes, in inches

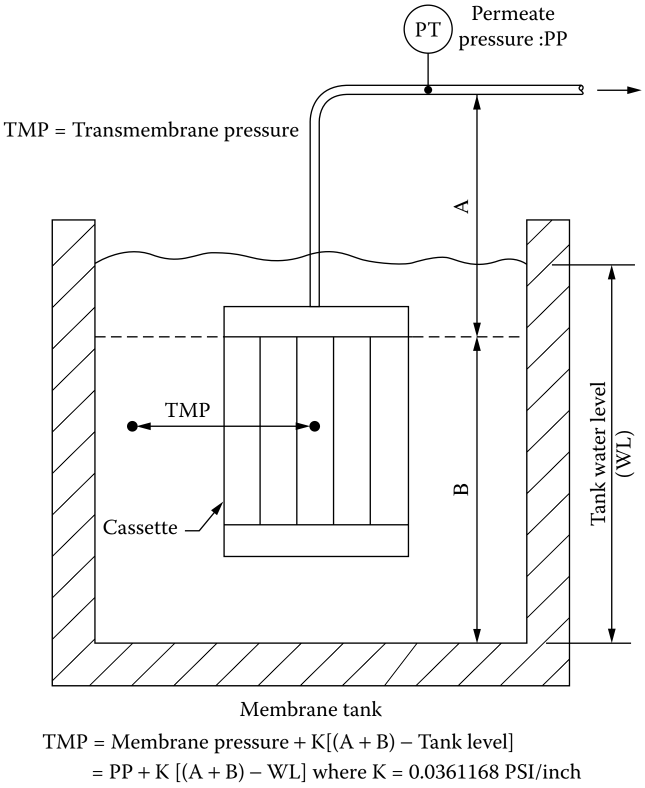


FIG. 8.40o

Illustration for defining the transmembrane pressure (TMP). (Courtesy of ZENON Environmental Inc.)

B is the distance between the top of the membranes and the bottom of the membrane tank, in inches

K is the conversion factor to convert pressure from in. H_2O to psi (0.0361168 psi/inch)

WL is the water level in the tank in inches

The TMP is negative when the unit is in production. During backwashing and cleaning, the TMP value is positive.

Control Loops

Figure 8.40p describes the ultrafiltration process and its controls.

When an ultrafiltration train is to be started up, the state of the train with the highest starting priority is changed from the "standby state" to "backwash operation state." During the backwashing operation, the backpulse pumps under flow control (FIC-6 in Figure 8.40p) start pumping permeate into the membranes for a short, preset time period. During the backwash cycle, FIC-6 maintains the required flow rate by modulating the speed of the backpulse pumps.

Permeate Pump and TMP Controls After being backwashed, the state of the train is changed to the "production state." At the beginning of production, the speed of the permeate pump is held constant by the control logic for a preset period. When that time has passed, the control of the speed of the permeate pump is transferred to be set by the output of a low-signal selector.

This selector compares the output signal of the clear well level controller (LIC-2), divided by the number of operating trains (N) and the output of the transmembrane pressure difference controller (PDC-1), and admits the lower of the two to adjust the speed of the permeate pump. Operating the permeate pump at a lower speed prevents overpressuring and reduces membrane fouling.

The flow controller (LIC-2) and the TMP pressure controller (PDC-1) loops selectively control the permeate pump speed through the low-signal selector. The output of LIC-2 is divided by the number of operating trains to arrive at the flow requirement of the individual train. If the membranes become fouled or if the TMP rises for other reasons, the low-signal selector selects the TMP controller (PDC-1) output to set the pump speed.

Membrane Tank and Clear Well Level Control Water from the distribution channel is fed into the membrane tanks under level control (LIC-1). The water level in the membrane tanks is held by the controller above the top of the cassette troughs.

At the beginning of the backwash cycle, the backwash drain valve (FCV-7) opens and drains out the water from above and in the backwash channel. At the end of the backwash cycle, FCV-7 closes and the train returns to the production state.

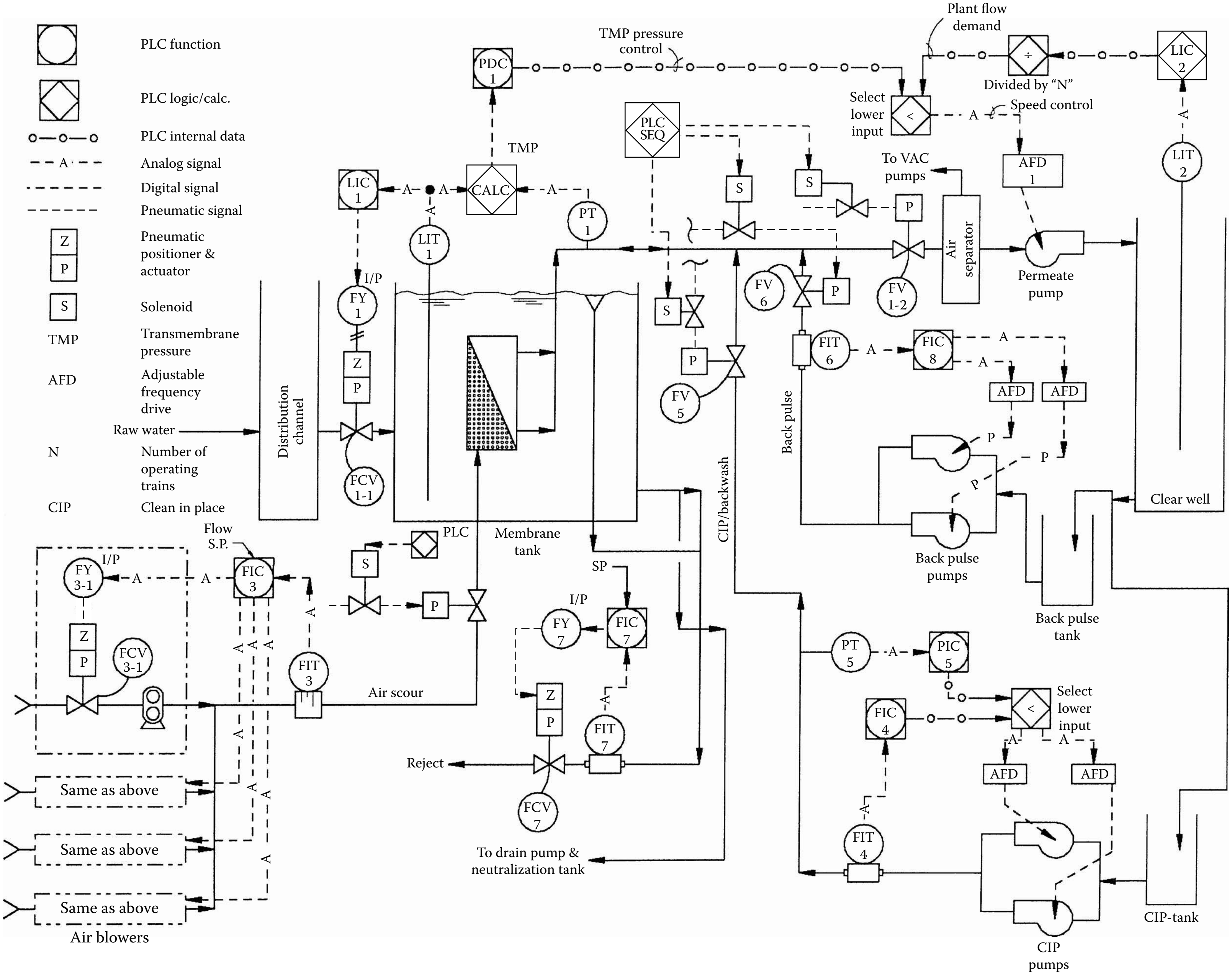


FIG. 8.40p
Ultrafiltration process controls include the selective control of the production rate by adjusting the permeate pump speed, using the lower of the PDC-1 and LIC-2 outputs. The CIP pumps are on overpressure protected selective flow control (FIC-4, PIC-5); the backpulse pumps, air scour, and reject streams are on straight flow control. (Courtesy of ZENON Environmental Inc.)

The total demand for permeate water that is to be produced by the trains of treatment plant is controlled by the clear well level controller (LIC-2). This demand is divided evenly between the operating trains. The instantaneous actual flow from the membranes is set higher than this net flow demand, in order to allow for the time when the trains are not producing water (i.e., during backwash), and to consider the water that is wasted during backwash.

The flow set point can also be entered manually for each train. The operator can set the net production flow rate for each train on the HMI. The control system will maintain the operator-entered production flow rate or the automatically set production flow rate, so long as they do not result in a trans-membrane pressure difference that exceeds the maximum allowed by PDC-1.

Air Scour Flow Control The flow controller, FIC-3, maintains a preset airflow for the scouring of the membranes. Air scouring is required during the backwash and the clean in place (CIP) maintenance cycles. The flow controller (FIC-3) throttles the inlet valves on the operating air blowers. The number of the blowers in operation is a function of the airflow controller set point.

If, because of blower malfunction, the required airflow cannot be maintained, a low-flow alarm is actuated and one or more trains are shut down. This control strategy has been described in Section 8.1.

CIP and Reject Flow Controls The clean in place control loops maintain a preset flow during maintenance cleaning operations. The flow controller (FIC-4) sets the speed of the CIP pump, while the pressure controller (PIC-5) provides standby overpressure protection. As long as the pressure is safely below the PIC-5 set point, the pump speed is set to maintain the required flow. If the pressure exceeds the set point of PIC-5, its output is selected by the low-signal selector, and pressure control takes over the control of the pump speed.

FIC-7 maintains the reject flow rate during the backwash cycle.

For a detailed discussion of all the control strategies and logic sequences that can be programmed in a PLC, the reader is referred to the references at the end of this section.

PLC IMPLEMENTATION OF PID LOOPS

While PLCs are ideally suited for logical and sequential on/off control tasks, when used on PID applications, one should clearly understand their characteristics. It should be understood that when these algorithms are implemented digitally, what used to be integration in the analog world becomes summation, and what used to be time differential becomes difference. The scan period of digital systems is usually fixed at around 0.5 sec or is selectable for each loop from 0.1, 0.2, 0.5, 1.0, 5.0, 10, or 30 sec.

As a digital system looks at the measurements only intermittently, it increases the dead time of the loop by two "scan periods" (the time it takes to return to that measurement), and it makes its algorithm calculations on the basis of the "present" and the "previous" error. This might not be a limitation on slow processes, but can be on fast ones.

There are two basic types of digital algorithms in use. One is called positional. This means that the full output signal to the valve is recalculated every time the measurement is looked at (Equation 8.40(15)).

$$m = K_c\left(e + \frac{1}{T_i}\sum_{o}^{n} e\,\Delta t + \frac{T_d}{\Delta t}\Delta e\right) + b \qquad \textbf{8.40(15)}$$

where

Δt is the scan period

Δe is the change between the previous and the present value of the error

$\sum_o^n e$ is the sum of all previous errors between time zero and time n

The PID algorithm used in this section is called a velocity algorithm. In this case, the value of the previous output signal (m) to the valve is held in memory, and only the required change in that output signal is calculated (Equation 8.40(16)).

$$\Delta m = K_c\left(\Delta e + \frac{1}{T_i}\sum_{o}^{n} e\,\Delta t + \frac{T_d}{\Delta t}\Delta(\Delta e)\right) \qquad \textbf{8.40(16)}$$

where

$\Delta(\Delta e)$ is the change in the error between the previous and the present scan period

The positional algorithm is preferred when the measurement is noisy (flow, for example), because it works with the error and not the change in the error when calculating its proportional correction. Velocity algorithms have the advantages of bumpless transfer and less reset windup, and they are better suited for controlling servomotor-driven devices. Their main limitations include noise sensitivity, likelihood to oscillate, and lack of an internal reference.

In-Line Mixer Controls

In this section, additives are mixed into water streams under the feedback control of analyzers. Because agitated vessels are expensive, simple devices such as in-line mixers are often considered for composition control systems. Properly applied, these devices can be effective, but careful attention to the following design criteria is required: reagent delivery hysteresis, loop gain, and neutralization stage interaction.

In case of pH control, an in-line mixer will control influent streams entering between 4 and 10 pH, and between pH

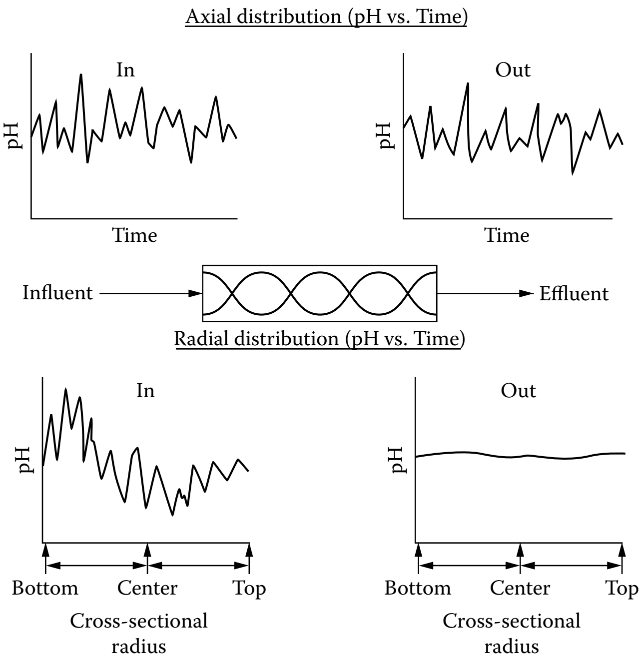


FIG. 8.40q
Axially, the fast disturbances pass through the static mixers, because they provide no back-mixing. This process is dominated by dead time.

6 and 8. The response will always be oscillatory following a disturbance.

An in-line mixer can be a dynamic mixer such as centrifugal pump or a baffled section of pipe called a static mixer. The static mixer provides radial mixing but little back-mixing. It can be considered to be a plug flow process *dominated by dead time.* Disturbances and noise pass through the mixer unattenuated. With such a mixer, the controller response to fast disturbances and noise is no *response at all*, because any corrective action will arrive too late and will create yet another disturbance (Figure 8.40q).

For this reason, a filter is often used on the pH measurement, introducing the largest time constant in the loop to attenuate the type of noise shown in Figure 8.40q. The response to a step change in load of an inline mixer is relatively slow and therefore, when a new train is started up or shut down, upsets are likely to occur.

The advantage of the in-line system is its small dead time loop period and recovery time. Disturbances that have ramp time or time constant greater than that of the in-line mix dead time will reduce the peak error, because the peak error approximates the deviation reached during 150% of the dead time. Averaging tanks downstream from the mixer will attenuate the oscillations that pass through the mixer, and attenuation tanks upstream will slow down concentration upsets. Similarly, valve or set point velocity limits will slow down flow upsets sufficiently to make the use of in-line mixers an attractive economic alternative to well-mixed vessels for even those applications with difficult titration curves.

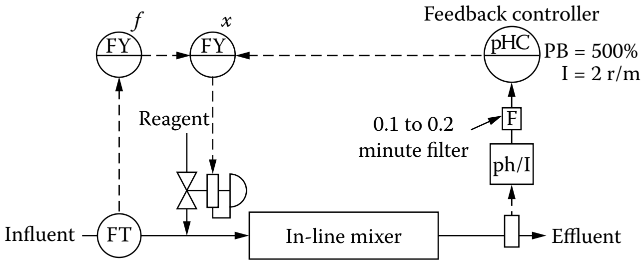


FIG. 8.40r
Flow-based feedforward correction can correct for the sudden changes in influent flow as trains are turned on or off. The reagent should be injected right before the mixer, the sensor should be located at about 10 diameters from the mixer, and its measurement should be filtered.

Filtering and Feedforward In this section, the influents are pumped, and the reagent and influent are both liquids. For pH control, the acceptable control band is greater than 0.5 pH, and an in-line mixer system becomes a low-cost option for continuous pH control. Originally, in-line mixers were relegated to installations with set points on a relatively flat portion of the titration curve. This is no longer the case.

When the set point is on a steep section of the titration curve, a digital filter with a time constant of 0.1–0.2 min can average the severe oscillations sufficiently to provide good control. This technique allowed the in-line mixing method to be used in a wide variety of applications and provided considerable cost savings.

Flow-based feedforward as shown in Figure 8.40r will substantially improve the performance of mixer-based analytical controls. In mixer controls the feedback controller gain is only 0.1–0.2, and the integral (reset) setting varies from two to four repeats per minute while the digital filter is set at 0.2 to 0.1 min. For in-line mixer systems, it is not important to add adaptive gain for flow-based feedforward, because it is a mostly dead time process and the controller gain is already detuned due to measurement noise.

The neutralization reaction is instantaneous once the ions are in intimate contact. The reagents should be injected as close to the mixer inlet as possible, and the measuring electrodes should be located about 10 pipe diameters downstream from the mixer as a compromise between the need to get a more uniform flow profile and the desire to minimize transportation delay.

Override and Selective Control

There are several loops in this section, where the normally operating flow control system is provided with a pressure-based override on a selective basis. In case of such overrides, a pressure controller can take command of a manipulative variable from a flow controller when the pressure would otherwise exceed the equipment limits. Whereas interlocks shut down equipment to avoid exceeding a limit, overrides

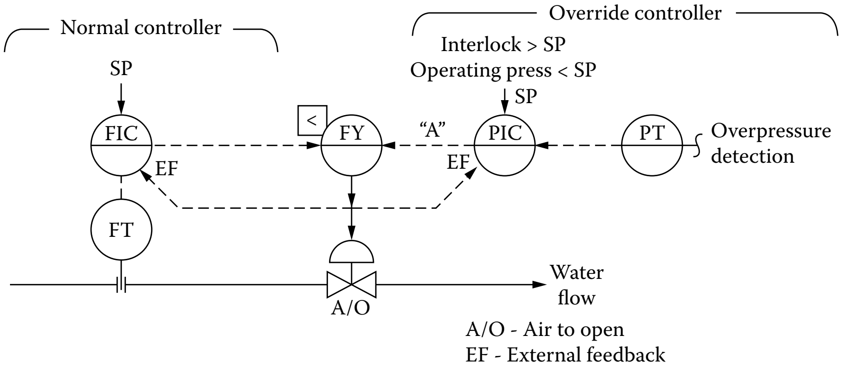


FIG. 8.40s
Flow control with pressure override requires external feedback (EF) to prevent reset windup in the idle controller.

"back off" from the desired operating set points so that the interlock trip point is not reached. Thus, equipment is kept running although perhaps at a suboptimal level.

The signal selectors facilitate the switching from one controller to another. Very often, overrides are preferable to interlocking when safety is not an issue. However, in most applications the overrides should be backed up by interlocks. The definition of overrides can therefore be stated as follows:

"Controllers that remain inactive until a constraint is about to be exceeded, at which point they take over control of the manipulated variable from the normal controller through a selector and, thereby, prevent the exceeding of that constraint."

Figure 8.40s shows a simple override control loop in which the normal controller (FIC) and override controller (PIC) outputs are both fed to a low-signal selector. The override controller set point (in this case, pressure) is set at a maximum value somewhere above the normal operating pressure but below the interlock point. As long as the pressure set point is not exceeded, the output of the override controller "A" is blocked by the low selector and cannot reach the flow control valve.

When the water pressure exceeds the override set point, the controller output decreases and eventually will become lower than the normal controller output and, therefore, will start to throttle the valve to satisfy the override set point. PID control algorithms require "external reset feedback" to avoid integrating the error, i.e., "reset windup," when their outputs are overridden.

CONCLUSIONS

Over the years, water plant controls have become more complex and more automated. The section described some of the PLC-based controls that can be used in conventional water supply treatment plants and in ozone disinfection-, reverse osmosis-, and ultrafiltration-based raw water treatment processes.

The emphasis on PLC-based controls is not intended to suggest that a PLC-type of control hardware is preferred for these types of applications. The treatment in this section reflects the author's experience with a specific application and the type of control hardware that was used in that application.

References

1. Keskar, P. Y., Stark, B., and Elarde, J., "Automation Control and Operation of R.O. Membrane Water Treatment Plants — From Design to Implementation," 2003 Florida Section AWWA Conference, Orlando, FL, November 2003.
2. "Control Narrative for ZW 1000 WTP," Zenon Environmental Inc., Oakville, Ontario, Canada, 2003.

Bibliography

American Water Works Association, American Society of Civil Engineers, *Water Treatment Plant Design,* 3rd edition, New York: McGraw-Hill, 1998.

James M. Montgomery, Consulting Engineers, Inc.; *Water Treatment Principles and Design,* New York: John Wiley & Sons, 1985.

Langlais, B., Reckhow, D. A., and Brink, D. R. (Editor), *Ozone in Water Treatment: Application and Engineering,* Chelsea, MI: Lewis, 1991.

Masschelein, W. J. (Eds.), *Ozonation Manual for Water and Wastewater Treatment,* New York: John Wiley & Sons, 1982.

Mallevialle, J., Odendaal, P. E., and Weisner, M. R., American Water Works Association Research Foundation/Water Research Commission of South Africa, *Water Treatment Membrane Processes,* New York: McGraw-Hill, 1996.

8.41 Well-Supplied Underground Gas Storage Controls

B. G. LIPTÁK (2005) **G. SZECSŐ** (2005)

INTRODUCTION

This section describes the controls of an actual underground natural gas storage system and the associated distribution network that supplies the nation of Hungary with natural gas. This large operating facility takes the gas from several natural geographic layers of sandstone, pressurizes it to 1450 psig (100 bars), and stores it in a spongy, underground sandstone formation (Figure 8.41a). This underground storage layer is about 3000 ft below the ground, it is about 300 ft thick, and it is penetrated by hundreds of natural gas wells, grouped in two distinct regions, referred to as the Northern and the Southern sites.

The storage facility is operated in two cycles. During the winter, it is in its production cycle (PRC), during which the natural gas is removed from storage and is distributed to the users at between 725 and 870 psig (50 and 60 bars) pressure. The pressure difference of 1450 to 870 psig is used up as pressure

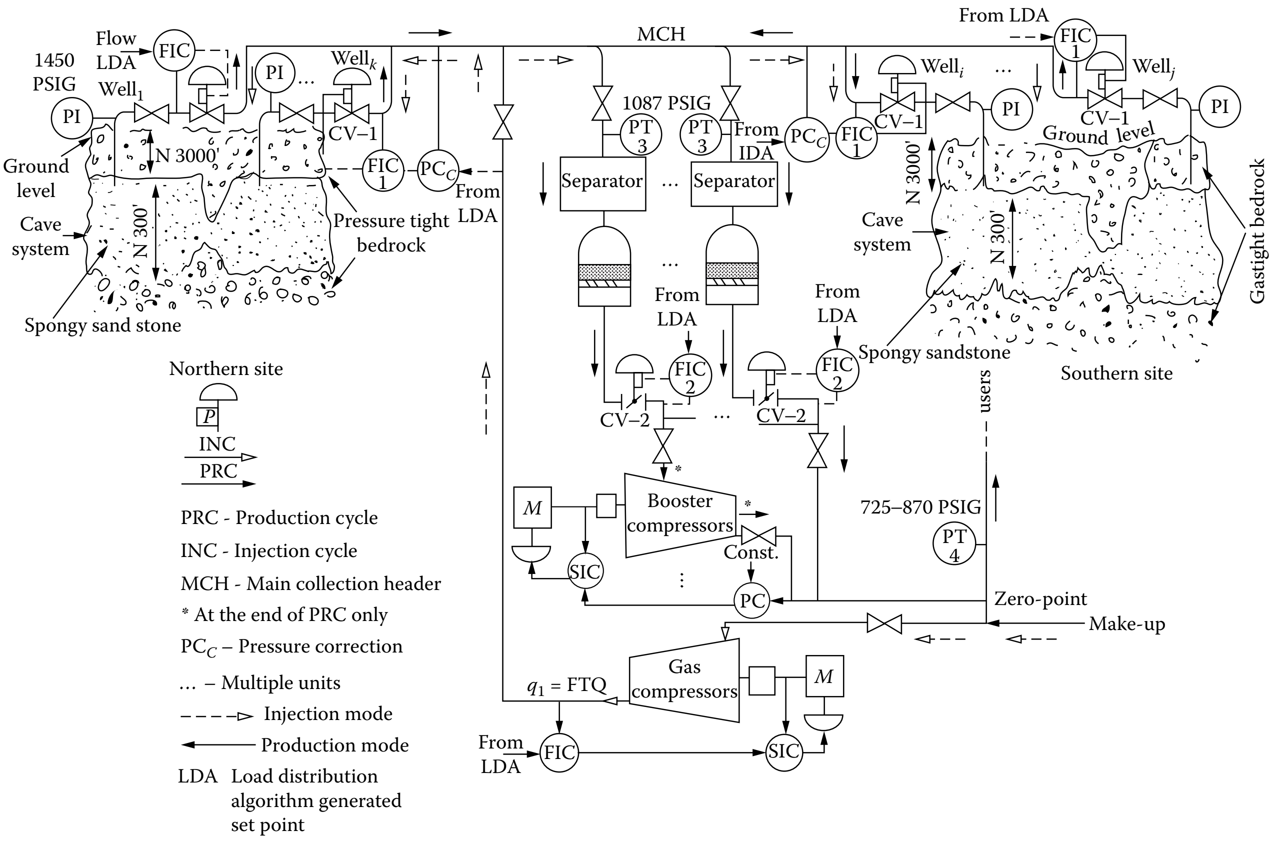


FIG. 8.41a
Schematic representation of an underground natural gas storage facility. It supplies a nationwide natural gas distribution system in the winter, when it is operated in its production mode (PRC). The actual facility consists of 100+ wells, a number of dryers, compressors, and the associated network of piping. Each well connects to a storage area of spongy sandstone. In order not to damage the underground geological formations, the differential pressure between storage areas served by the neighboring one should be minimized.

drop through the well flow control valves (CV-1s in Figure 8.41a), the dryer flow control valves (CV-2s), the dryers, and the connecting system piping. At the end of the production cycle, a booster compressor station can serve to remove the remaining low-pressure gas from storage.

After the production cycle of the winter, the depleted wells are refilled during the injection cycle (INC) of the summer. During this cycle, compressor stations deliver fresh natural gas, obtained from domestic gas containing sandstone layers in the area (and from gas supplied mainly by Russia), into this underground gas storage (UGS) system. Figure 8.41a schematically describes the overall UGS system.

The reader is reminded that the controls for the various unit operations that are utilized in this system are discussed in more detail in the following sections: Section 8.15, Compressor Control and Optimization, Section 8.22, Dryer Controls, and Section 8.28, Header Pressure Control. It should also be kept in mind that what is being described is an operating system that has evolved from a manually controlled process that has been in opreation for nearly a decade.

THE PROCESS

Over millions of years, hermetically sealed multiple hydrocarbon gas layers evolved in the sandstone under Hungary. This region used to be the sea bottom of the ancient Pannon Sea. The natural gas in these layers is extracted and stored in a pressurized artificial storage layer to serve as the gas energy supply of the nation. The shape of the storage layer is elliptical, about 300 ft thick, and is situated about 3000 ft below the surface.

The gas is collected by hundreds of wells, which are grouped at two gas-collecting stations (Northern and Southern sites). The wells are periodically discharged (production mode of operation) and then refilled (injection mode of operation). Piping is provided among the wells and stations, serving the purposes of flow and pressure control. The two stations are also connected by one large-diameter pipe.

At one of these stations, the natural gas is dried and conditioned during the PRC so as to be suited for long-distance transportation. During the reinjection cycle, the natural gas is distributed for reinjection in order to refill the wells during the INC. The whole system is comprised of more than 100 gas wells.

The UGS system therefore consists of the Southern and Northern sites and the above-ground plants, which include the distribution pipelines, separators, dryers, and compressors. The gas supply is divided 60 to 40% between the Southern and the Northern sites, where the majority of above-ground equipment is also located. All gas wells and the above-ground facilities been designed for 1450 psig (100 bars) operation. The gas distribution network is designed for 870 psig (60 bars). If the adequate conditions exist then the above-mentioned two systems will be connected to each other.

The underground gas storage system has two modes of operations, which mostly depend on the seasons. The control system in both modes of operation has to provide protection for the natural geological formations and the sandstone storage, while providing continuous operation either in the injection or the production mode of the underground storage system.

The Operating Modes

The two modes of operation are referred to as the pressure correction control method (PCCM) and the load distribution algorithm (LDA). In both cycles, unique algorithms perform the controls of the process. Digital controls (DCS) are used to implement the algorithms that generate the control signals to the manipulated variables.

Basically, the system has a start-up mode, TWM1 (Technological Working Method #1), during which the controls start the process and supervise the changing operating set points (OP), and TWM2, which stabilizes the operating points (OP) and maintains it in the steady state. During the start-up phase (TWM1), the gas supply to the nationwide distribution network is not stable or tightly controlled. Table 8.41b gives a summary of the two modes of operation.

Injection Cycle

In this mode of operation, single automatic stage flow control (SASFC) is provided with the control loop in automatic (ACM) for the individual gas wells in order to optimize the operation of the whole group of wells. In this mode of operation, the load distribution algorithm (LDA) provides the flow set points for the flow controllers on the individual wells.

TABLE 8.41b
The Two Modes of Operation of the Underground Gas Storage System

Operating Modes (Cycles)	*During the Injection Cycle (INC)*	*During the Production Cycle (PRC)*
Start-up mode: Technological Working Method #1 (TWM1)	Starting process and groups of compressors. Supervising the operating set points (OP).	Starting process for operation in the production cycle (PRC). Supervising the operating set points (OP).
Normal operating mode: Technological Working Method #2 (TWM2)	Stabilizing the operating set points (OP). Supervising gas distribution among wells.	Stabilizing the operating set points (OP). Producing required gas quantity at the required outlet pressure.

During the start-up mode of the INC, the flow into the individual wells is limited by the maximum flows that are allowed for each well. These limits are used in initiating the emergency stoppage of the compressors or the changes of OPs.

One of the goals of the control system is to have the pressure in the main collection header (MCH) approximately approach the average pressure of the underground gas storage (UGS) system in order to minimize the energy consumption of the compressors. This is achieved by the use of weighing factors (WFs), so that the load distribution will be proportional to the capacities of the individual wells.

The compressors are manually controlled, and it is important that the manual changes that modify the injected total gas flow set point do not upset the stable operation of the cascade flow control on the individual wells. When the compressors are restarted after a power failure, they are protected against overpressure.

All other control and safety functions, including control mode switching and emergency responses, are the same as used during the PRC, which is described below.

Production Cycle

In this mode of operation, single automatic stage flow control (SASFC) set points are provided for the single flow loops on the individual wells and also for the wells that are provided with LDA cascaded flow control. The LDA serves to guarantee that the total gas produced matches the total demand. The flow set points of the individual gas wells are set according to their WFs with minimal actuator actions.

During the operation of the system, it is necessary to limit the pressure drop across the control valve (CV-1 in Figure 8.41h) between the wells and the MCH and also on the dryers (CV-2 on Figure 8.41k) to 290 psid.

When the production OP is steady, it is the goal to limit the error between the actual and the desired flow (*Q*) to ±1%. It is desirable to keep the control system operable even if the measurement errors of the ultrasonic flowmeters on the gas wells (F^w in Figures 8.41h, j, and k) increased or if deep groundwater interfered with their operation.

It is also desirable that the upset caused by the manual starting of the booster compressor station should not interfere with the operation of the other control loops.

MODELING AND SIMULATION

The simulation of the operation of the total system requires the modeling of the individual equipment components, which are listed in Table 8.41c.

The overall system model is composed of several submodels. The first submodel is the model of the individual wells (GGW1, GGW2, and so on), which includes the models of the valves, sensors, and their PI controllers. The typical operating conditions of a well are listed in Table 8.41d.

TABLE 8.41c
Submodels Are Required of Each System Component in Order to Model the Total System

System Component	*Function*
Control valve (CV-1)	Control valve with positioner* and its programmable characteristic curve, which controls the gas flow from each individual well
Butterfly valve (CV-2)	Butterfly valve with positioner* to control the flow through the dryers
Flowmeter (F^w and F^d)	A sonic, gas flow-measuring block with programmable timing from 5 to 60 readings
Main collection header (MCH)	The model for the main collection header is a block that is a nonconcentrated and nonlinear, long-distance pipe model
Dryer	The model representing a dryer block with variable capacity
Well group	The model of three groups of gas wells
Dryer group	The model for a group of dryers
Controller (PID)	Proportional and integral control** algorithms for the flow controls of the gas wells and the dryers
Load	A variable load change model for the zero point

* Generally we do not use positioners on flow control valves, because flow is a fast process and the positioner cannot keep up with it, which results in detuning of the controller and in limit-cycling and hunting.

** In cascade configurations, external reset is required to protect against integral windup, when the loop is switched from cascade to manual or direct slave control.

The other three submodels are that of the MCH, the group of dryers (GD), and the simulated load change (LC). While the model of the gas wells (GGW1, GGW2, and so on) can use real, dynamically measured data, the other three

TABLE 8.41d
Typical Operating Conditions Used in Modeling the Wells

Parameters	*Conditions*
Gas temperature	$T = 50°F$
Set point (reference signal) change	$\Delta q = 1310$ yd^3/h
Set point (work point) at steady state	$qst = 3270$ yd^3/h
Average system pressure	943 psig
Sampling time of flow detection	15 sec*
PI controller tuning parameters**	PB = 150%, Integral time = 120s/repeat

* The sampling time on flow loops should be much shorter than 15 sec. Also, because of their noisy nature only positional algorithms should be used.

** These tuning parameters can be improved (PB narrowed, integral time shortened), if the valve positioner is eliminated.

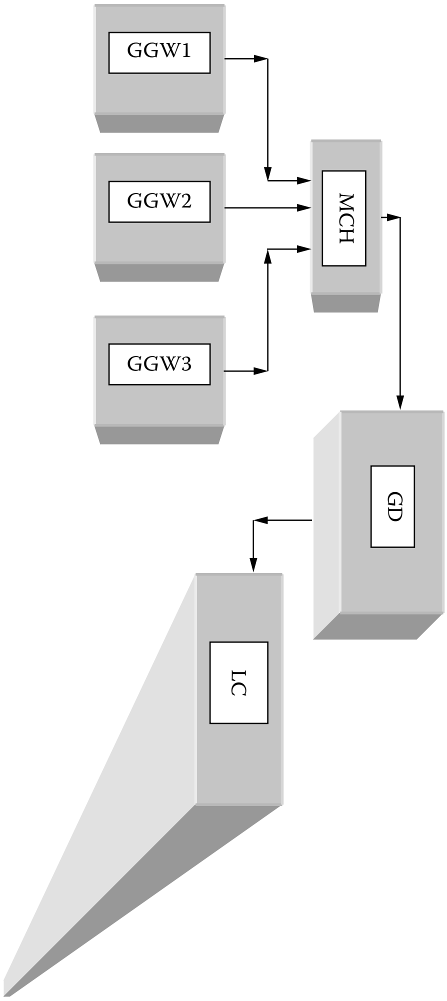


FIG. 8.41e
Functional block diagram of an integrated gas storage model, where GGW1–GGW3 are the submodels of the gas wells, MCH is the submodel of the main collecting header, and GD is the submodel of a group of dryers and represents a simulated load change (LC).

submodels utilize only statically measured inputs. Figure 8.41e illustrates the relationship of these submodels within the total model of the gas storage system model.

In order to decrease dynamic and static deviation, nonlinear control needs to be used. Capacity of the gas well has a long-term (monthly) and a short-term (hourly) characteristic. In either case, the pressure drop across the control valve (CV-1 in Figure 8.41h) affects the dynamics of the flow control loop. For this reason, adaptive control is desired to adapt the tuning of the loop to the changing inlet pressure to the valve. The effect of the variations in outlet pressure can be minimized by pressure compensation.

GENERAL CONTROL CONSIDERATIONS

The quality of control can only be as good as the sensors, the control valves, and the control algorithms used. Therefore, it is desirable to carefully consider their requirements.

Sensor Selection

It is important that when two flowmeters operate in series, such as the flow sensors at the wells (F^w in Figure 8.41h) and at the dryers (F^d in Figure 8.41i), they be calibrated together, so that they will give the same readings at the same flow rates.

It is also important that the sampling time of the flow sensors be short enough so that their contribution to the flow control loop dead time is minimum, because otherwise the flow control loop will cycle. It is also important that only smart, self-diagnosing sensors be used, and only those suppliers be considered that have substantial experience with natural gas service applications. As can be seen from Table 8.41f, the number of ultrasonic flowmeter suppliers is large, but as of this writing the most experienced are Daniel and Instromet.

Control Valve Selection

The proper selection of control valves is equally important. On fast loops such as flow control loops, the use of positioners is rather questionable, because in effect the positioner is the cascade slave of the flow controller, and therefore its time constant should be one tenth of its master, and such high-speed positioners are seldom available.

A controlled process can be considered “fast” if its period of oscillation is less than three times that of the positioned valve. In such situations, the positioned valve is one of the slowest components in the loop, and therefore it slows the loop down (limits the open-loop gain of the loop and lengthens its period of oscillation). In such cases the loop without a positioner can be tuned more tightly (for higher gain and more repeats/minute); such a loop responds better without a positioner. It might also be noted that after a new steady state is reached, the positioned installation can provide more noisy control because of the hunting and limit cycling of the positioner, if it cannot keep up with the process.

Some will argue that all loops can be controlled using positioned valves if the controllers are sufficiently “detuned.” This is true, but “detuning” means that the controller is made less effective (due to a reduction in the amount of proportional and integral correction), which is undesirable.

Controller Algorithms

On fast and noisy processes such as on flow loops, positional digital algorithms are preferred (see Section 2.4 in Chapter 2). It is also recommended that the sampling time of the flow loops should be short, because the sample period is pure dead time. Therefore, for example, if a set point change can occur faster than several sampling periods, the loop will be upset. Consequently, if large set-point changes are anticipated and fast loop response is required, one should increase the sampling frequency.

It is also important that when a cascade configuration is used, the controller be provided with external reset, so that its integral mode will not wind up when its output signal is

TABLE 8.41f
Models and Types of Ultrasonic Flowmeters by Supplier

	Type			*Operating Principle*			*Fluid*		
Company	*SP*	*CL*	*IN*	*TT*	*D*	*H*	*G*	*L*	*S*
American Sigma		x			x			x	
Automated Sonix	x	x						x	
Caldon	x	x		x				x	
Controlotron	x	x		x	x		x	x	x
Danfoss	x			x				x	
Daniel	x			x			x		
Datam Flutec			x	x				x	
D-Flow				x				x	
Durag			x	x			x		
Dynasonics		x		x	x			x	
Eastech Badger	x	x		x				x	
EES		x		x				x	
Elis Plzen	x			x				x	
EMCO		x		x				x	
Endress+Hauser		x		x				x	
Flexim		x		x				x	
Flotek UK		x		x				x	
Fluenta			x	x			x		
FMC Smith Meter	x			x			x		
Fuji Electric		x		x	x			x	
GE Panametrics		x		x		x	x	x	x
Greyline		x			x			x	
Honda		x		x				x	
Instromet	x		x	x			x		
Kaijo	x	x		x			x	x	
Kamstrup	x			x				x	
Krohne	x	x		x			x	x	x
Laaser		x			x			x	
Matelco	x	x	x	x				x	
Mesa Laboratories	x	x		x	x			x	
Micronics		x		x				x	
Monitor Labs				x			x		
Oval Corp.	x			x			x		
Polysonics		x		x	x			x	
Quality Control		x			x			x	
Rittmeyer		x	x	x				x	
Sick			x	x			x		
Siemens	x			x				x	
Solartron Mobrey		x				x		x	
Sparling	x			x				x	
Teksco USA		x			x			x	
Thermo MeasureTech		x			x			x	
Tokimec		x		x	x		x	x	
Tokyo Keiso	x	x		x				x	
Ultraflux		x	x	x				x	
Ultrasound Res. Ctr.		x	x	x				x	
Yokogawa		x		x				x	

SP = spoolpiece
CL = clamp-on
IN = insertion
TT = transit time
D = Doppler
H = hybrid
G = gas
L = liquid
S = steam

blocked from reaching the control valve, because the loop has been switched to manual.

Lastly, one should keep the total loop gain constant, because the proper tuning of the controller cannot be maintained if the loop gain varies. Therefore, if the control valve gain is not constant, which is the case with all valves except the linear ones, valve compensators should be used. This is also the case with butterfly valves.

Pressure Correction

In the following paragraphs, three control loop configurations will be described for controlling the flow from the individual wells: 1) cascaded total flow master setting individual flow loops as its slaves, 2) combination flow/pressure cascade master, and 3) combination cascade master based on flow and pressure correction from the main collection header.

Pressure correction of the flow control loop can improve the quality of flow control both in terms of accuracy and in terms of reducing the size of the upsets caused by sudden load changes. A method of pressure correction will be described in the following paragraphs that can also be useful in keeping the pressure drop across the dryers constant and in keeping the total natural gas flow from the facility constant.

Different control algorithms will be described for use when the load is increasing or decreasing and to achieve bumpless transfer between the manual and automatic modes of controller operation. Figure 8.41g illustrates the response

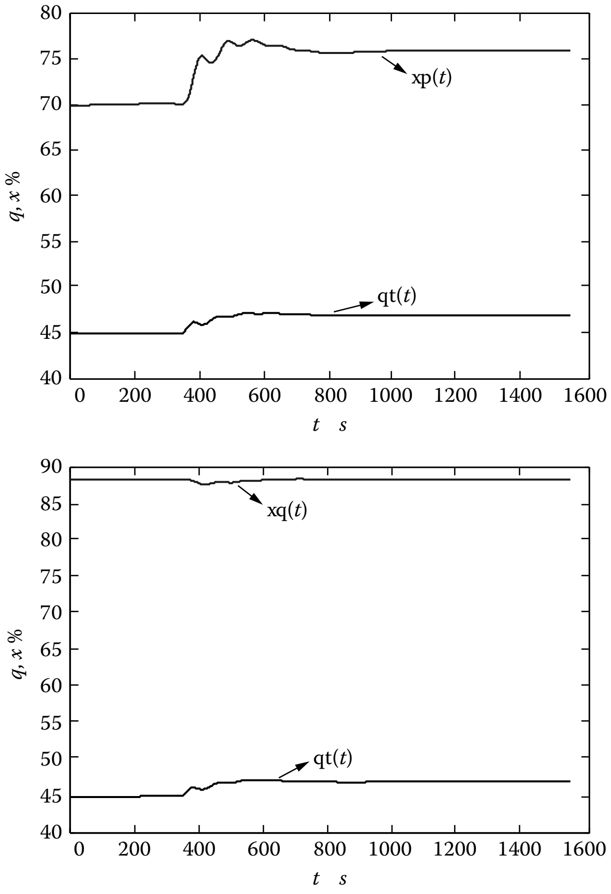


FIG. 8.41g
The response of the pressure-corrected natural gas flow control loop to a sudden change in set point.

of a pressure-corrected flow control system when responding to sudden change in the required flow rate.

CONTROLLING THE INJECTION CYCLE

After the winter season, the exhausted wells must be recharged. This is done by switching the system to the INC, which is shown by the dashed arrows in Figure 8.41a. During the injection cycle, the exhausted storage wells are recharged by natural gas that is lifted by compressors from other layers of the sandstone deposits of the region and from Russian gas imports. The compressors pressurize an MCH, from which the gas is distributed to the individual wells that need recharging. The pressure in the MCH is a function of the flow rate of gas delivered by the compressors and the permeability (the rate at which the spongy sandstone in the wells can absorb the gas).

The wells in the more porous sandstone layers can receive and store the gas faster. These are called “soft wells.” The less porous formations can accept the gas only at a slower rate. These are called “hard wells.” The control task is to refill the underground storage system, while on the one hand using a minimum amount of compression energy and on the other hand protecting the underground geological formations from damage, caused by excessive pressure differences between the wells serving neighboring storage areas.

If one based the operation only on the permeability of the “soft” wells, the MCH pressure could be relatively low and much compression energy would be saved, but the speed of recharging the “hard” wells would suffer. If the MCH pressure was maintained at a level that keeps the rate of recharging of the “hard” wells efficient, compression energy consumption would be high, and much compression energy would be burned up as pressure drops in the gas valves serving the “soft” wells. In addition, the pressure differences between soft and hard wells could cause damage to the geologic formations.

The task of the control system in the reinjection mode (INC) is to find a reasonable compromise.

Prestart-Up Settings

The operator manually initiates either the start-up mode (TWM1) or the operating mode (TWM2) of the control system. Prior to start-up, the operator should enter the following constants:

ETMI	The expected total mass of the gas that is to be stored, in m^3/h
ALSW	The amount of gas that can be admitted into each well, in m^3/h
IWFW	The initial weighing factor for each well
FRSS	The slope at which the total flow rate can change, in (m^3/h)/h
MVSM	The rate at which a single well’s control valve signal can change in %/h
MDPH	The maximum pressure difference between the average well pressures (p^w in Figure 8.41h) and the pressure in the main collection header (P_c in Figure 8.41k) in bars
DFSWi	The direction of flow that can be measurement or production
AFSW	A flag indicating if the particular well is active or on standby
CMW	The control mode for the flow controllers on the individual wells
MVLS	The low-signal limit for the control signal to the valve (CV-1 in Figure 8.41h) in %
MVHS	The high-signal limit for the control signal to the valve (CV-1 in Figure 8.41h) in %

START-UP MODE

The start-up mode (TWM1) is used: 1) during the initialization of the INC when the compressors are being started, 2)

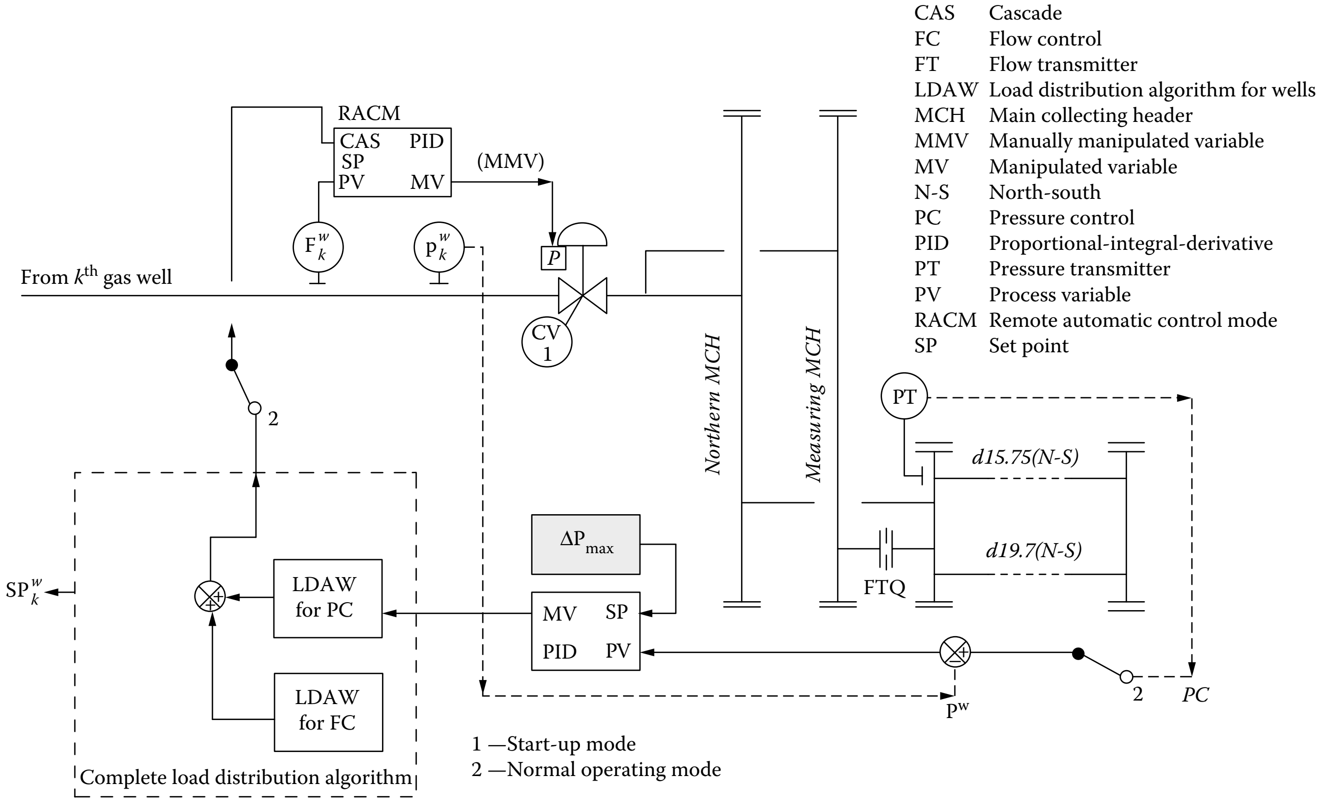


FIG. 8.41h

The flow control loop on each well is cascaded to properly distribute the total flow and to keep the pressure drop across the control valve (CV-1) under a maximum limit.

when the flow distribution among the compressors is being changed, 3) in case of problems that necessitate that the operator initiate manual corrections, 4) prior to the activation of the safety controls to protect against the development of overpressure at the end of the injection operation (INC), and 5) during upsets that result in fast changing (high-transient) operating conditions.

At the beginning of the start-up operation the control valves (CV-1 in Figure 8.41h) are totally opened and the active wells imbibe (absorb) the gas according to their natural permittivity (the rate at which gas can be absorbed). When the flow rate reaches its maximum limit, which is set for each particular well, the flow controller is activated and maintains its set point at that limit. This maximum flow stage of operation continues as long as the compressors can meet the demand by lifting the required quantity of gas.

The pressure that is developed in the MCH is a function of the permittivity of the "soft" gas wells. When the compressors can no longer meet the demand for gas, the pressure in MCH drops and the remaining gas is charged to the soft permittivity gas wells. These remaining gas quantities (lent values) are delivered at flow rates that are controlled at a maximum limit for each well.

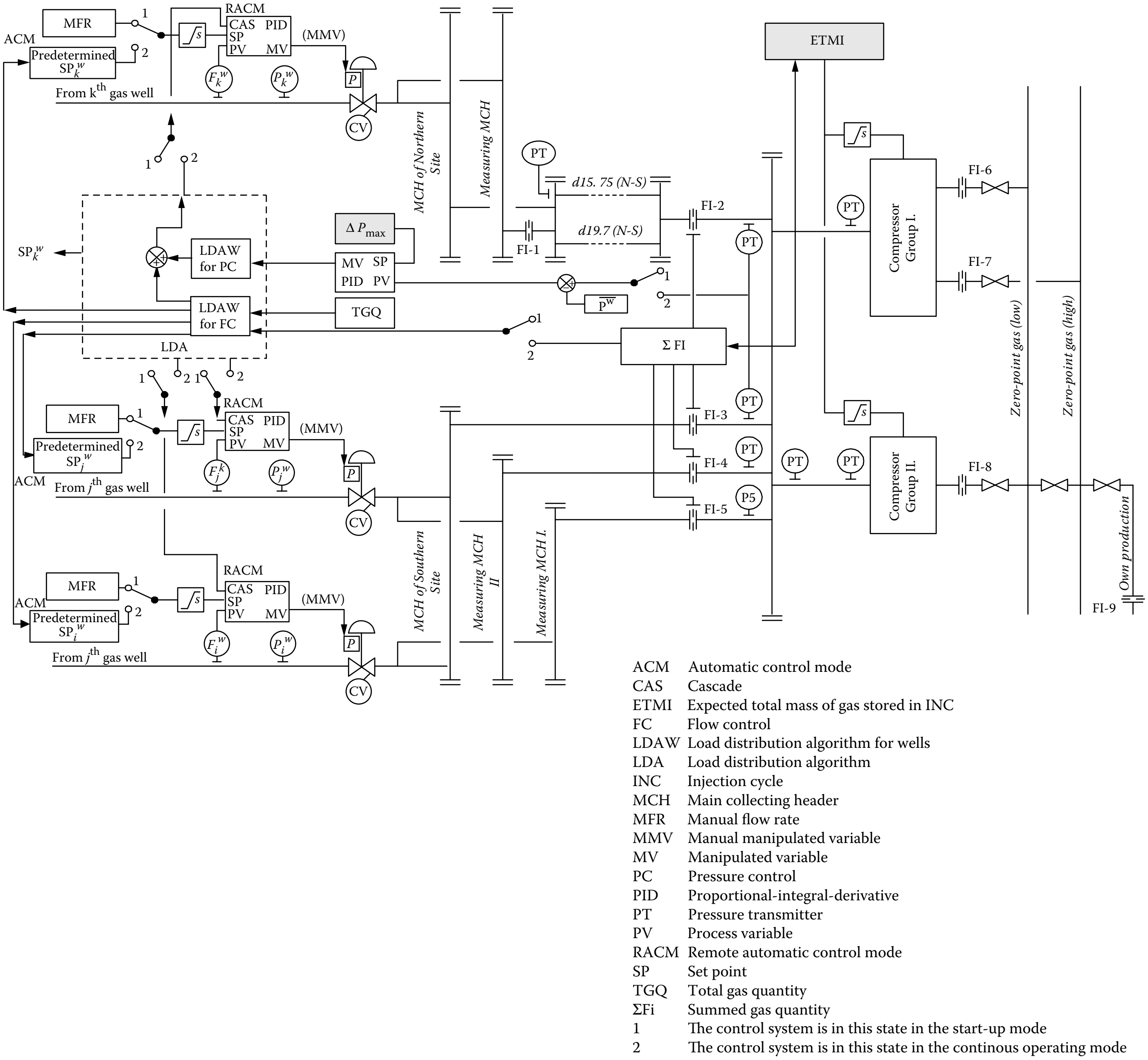
During the reinjection cycle (INC), the pressure in the storage wells is increased, while the actual amount of gas stored is a function of the permittivity of the individual wells. The storage pressures obtained in the individual wells will be different because of the differing permittivity of the wells, but this pressure difference must not be excessive, because the pressure differences between underground storage layers can damage the geological structures.

The requirements for energy-efficient operation of the compressors is different if they are serving hard or soft permittivity wells. In case of hard permittivity wells, the compressors can efficiently charge them at maximum flow rates, but if soft permittivity wells are also served, the pressure loss across their control valves will be large, which represents a waste of energy. In case of the soft permittivity gas wells, the minimum level of compressor operation (minimum lifting energy) is insufficient to develop the pressure required for the wells to efficiently absorb (imbibe) the gas.

The overall control system and equipment used during the INC of this underground gas storage facility is shown in Figure 8.41i.

Operator Actions in the Start-Up Mode

After the parameters that were listed in the previous paragraph have been specified, the system is placed in the start-up mode (TWM1) and is started up in manual, under the supervision of



ACM	Automatic control mode
CAS	Cascade
ETMI	Expected total mass of gas stored in INC
FC	Flow control
LDAW	Load distribution algorithm for wells
LDA	Load distribution algorithm
INC	Injection cycle
MCH	Main collecting header
MFR	Manual flow rate
MMV	Manual manipulated variable
MV	Manipulated variable
PC	Pressure control
PID	Proportional-integral-derivative
PT	Pressure transmitter
PV	Process variable
RACM	Remote automatic control mode
SP	Set point
TGQ	Total gas quantity
ΣFi	Summed gas quantity
1	The control system is in this state in the start-up mode
2	The control system is in this state in the continous operating mode

FIG. 8.41i
The equipment and control system used in the reinjection of natural gas into underground storage.

the operator. The start-up sequence involves the switching of all controllers to the manual mode and setting their output signals (manipulated variable signals to the control valves) to their low limit values (MVLS).

All flow control loops on all wells (AFSWi) are activated. Compressor groups are loaded up at a rate that is less than the rate allowed (TDTR) for reaching stable state of the system. When the flows have stabilized, the operator switches the controllers to automatic and specifies the maximum total flow rate limit for the facility. The operator also adjusts the set point of the flow controller to the admissible load (AL for INC) and switches the controllers to automatic.

The operator checks the rate and stability of the flows into the active wells, while isolating the nonoperative gas wells. The initial flow rate that was set by the operator (AL) is overridden by the LDA, as the set points of the flow controllers are automatically modified. At this point, the local or remote automatic control mode of the flow controllers is switched to cascade control, which provides pressure correction, and the control system is switched from the start-up mode (TMW1) to the production mode (TWM2). When in the production mode, the total gas flow is distributed between the wells according to the weighing factors of the individual wells.

Normal Operating Mode

In the normal operating mode of the injection phase (TWM2), the set point of the individual flow controllers on the individual wells is provided as shown in Figure 8.41h. The charging rate into the wells can be manually determined by the operator, can be automatically maintained as a preset portion of the total flow, or can be adjusted in a cascade manner to also consider the compressor discharge pressure and the openings of the operating control valves. The goal of the control system is to match the rate of natural gas supply generated by the compressors with the rate at which the operating well can accept that flow and to do that without either wasting compressor energy or creating unsafe overpressure conditions.

Artificially determined weighing factors are used to describe the permittivity of each well. The operator can choose to switch the system into the normal mode (TWM2), if the following conditions exist: The injected total gas flow rate is stable, the starting of the compressors or the changing of their loading been completed, and the distribution of the total gas flow among the wells can be handled by making small changes in the set points of the corresponding flow controllers.

In the normal mode of operation, the flow rate set points of the wells are determined by distribution of the total gas flow in proportion to WF of the wells. The flow set points are redistributed if the total injected or supplied quantity of gas changes. The change in the total flow rate has to exceed the limits of a dead zone (dead band) before automatic redistribution is initiated, but the operator can also initiate this redistribution.

Pressure Correction and Flow Balancing The total flow generated by the compressors is detected by orifice plates (FI-1 in Figure 8.41i), while the flow received by the active wells is the sum of the active ultrasonic flow sensors (F^w in Figures 8.41h and i). If the two flows are not the same, the imbalance will change the pressure in the MCH, and to avoid that, the total flow has to be redistributed among the active wells. Equation 8.41(1) describes how the individual flow set point is calculated for a particular well, in order to keep the pressure in the MCH (the reserve pressure in the main collection header) constant.

$$\mathrm{SP}_i^w = F_i^{\min} + \left[\sum_{i=1}^{n} F_i^{\text{injected}} \mathrm{SC} \left(\frac{\sum_{i=1}^{m} F_i^{ns}}{\sum_{i=1}^{m} F_i^{\text{injected}}} \right) - \sum_{i=1}^{k} F_i^{\text{man}} - \sum_{i=1}^{l} F_i^{\text{man, aut, cas}} \right] \frac{\mathrm{WF}_i}{\sum_{i=1}^{m} \mathrm{WF}_i} \qquad \mathbf{8.41(1)}$$

During reinjection (INC), constant pressure in the MCH indicates a balance between the supply and demand for gas, because the gas quantity lifted by the compressors matches the amount of the injected gas mass.

If more gas is lifted by the compressors than the amount of gas pressed into the wells, the pressure of the MCH will increase. When this occurs, the pressure correction cascade algorithm in Figure 8.41h will increase the set points of the flow controllers on the wells that can safely accept more gas.

If the mass of gas lifted by the compressors is less than the amount pressed into the wells, the pressure in the MCH decreases and the control valves serving the soft permittivity wells will not receive enough gas, so they will fully open. When this happens, the operator will switch the system to the PCCM if the flow into the soft permittivity wells is higher than maximum allowed. In this case, the pressure correction controls (PCC) will lower the automatic and cascade set points of the flow control loops of the “soft” wells and redistribute the total flow. If the flow through the fully opened valves falls below their minimum limit for the soft permittivity gas wells, the redistribution process will stop.

Because of flow sensor errors and other considerations, the total gas quantity (TGQ in Figures 8.41a and h) is calculated as follows:

$$q_2 = v q_1 \mathrm{SC} \qquad \mathbf{8.41(2)}$$

where

q_2 is the total gas flow (TGQ) that is measured by ultrasonic flow sensors

v is the average deviation ($q_1 - q_2$) of the previous sampling period

q_1 is the total gas flow (TGQ) that is measured by an orifice-type detector

SC is a system constant

In order to minimize the power consumption of the compressors, the pressure in the main collection header should be minimized. When this minimized pressure is stabilized, the flows into the active wells will differ, as soft permittivity gas wells can accept more and hard ones can take less flow. Consequently, the soft permittivity gas wells will not receive all the flow they could take, and their control valves will fully open (100%).

Once a control valve is fully open, it is out of control and cannot maintain the flow set point requested by the LDA. Under these conditions a new, experience-based flow set point is applied to these soft permittivity wells as their maximum flow rate (MTSP). If the flow into a soft permittivity gas well exceeds this MTSP, the total flow is redistributed among the active wells to maintain it. The goal of this redistribution is the deactivation of the hard permittivity gas wells. The redistribution procedure modifies the set points of the flow controllers according to the following relationship:

$$\mathrm{SP}_i^k = \mathrm{SP}_i^{k-1} - \left(\sum_{i=1}^{n} \mathrm{SP}_i^{\mathrm{AUT+CAS}} - \sum_{i=1}^{n} \mathrm{F}_i^{\mathrm{AUT+CAS}} \right) \frac{1}{2} \frac{\mathrm{WF}_i}{\sum_{i=1}^{n} \mathrm{WF}_i} \qquad \mathbf{8.41(3)}$$

The redistribution procedure based on Equation 8.41(3) is executed every 15 min, until the actual flow into the soft permittivity gas well drops below MTSP.

Manual, Automatic, and Cascade Control Modes

The flow control loop can be in manual (MCM), local automatic (LACM), and remote cascade (RACM). When the controller is in manual, the control valve (CV-1) is manually throttled by the operator. This is the case if the ultrasonic flow sensor (F^w) is out of service. This is referred to as the manual control mode (MCM).

The controller is in local automatic (LACM) at the beginning of both the start-up and the continuous operating modes or when load changes occur. In the start-up phase (TWM1), the set points of the controllers are set to a maximum value that corresponds to the admissible load of the particular well (ALSW). In the continuous operating phase (TWM2), the flow controller set point is determined on the basis of the weighing factor (WF) that is assigned to the particular well by the load distribution algorithm (LDA). This set point is also limited, so that the LDA cannot move the set point higher than the allowable maximum for the particular well. The operator can manually modify this set point when necessary.

The flow controller is in cascade or remote automatic (RACM) when the pressure correction mode is active. In all control modes, the rate of change of the flow measurement signal is limited until the steady state is reached. When in the continuous operating mode (TWM2), the set point of the flow controller is adjusted to protect against overpressuring the system.

Controlling Valve Pressure Drops

In the reinjection mode of operation (INC), the energy consumption of the compressors rises as the pressure differential across the control valves rises. The higher this differential, the more energy is wasted. In such distribution control systems, it is recommended to use valve position-based optimization, which is described in detail in connection with Figure 8.15dd in Section 8.15.

As the system is started up in the reinjection mode (TWM1), the pressure in the MCH is minimum, and therefore the valve pressure drops are also low. During this phase of operation, if the flow controller set point is limited to its maximum flow rate, the differential pressure between that of the MCH and the average pressure at the side of the gas wells will not be excessive.

During the continuous mode of operation (TWM2) the pressure drop across the control valves is limited to a programmed value (dPmax). The development of an overpressure in the MCH can be eliminated by sending more gas to the best permittivity gas wells. This cascade controller configuration for the best permittivity wells is shown in the lower part of Figure 8.41h. The set point of these controllers can be calculated by the following equation:

$$\mathrm{SPi}^{\mathrm{CAS}} = \mathrm{SPi}^{\mathrm{FLOW}} + \mathrm{MV}^{\mathrm{PRES}}(\mathrm{Fi}^{\mathrm{MAX}} - \mathrm{SPi}^{\mathrm{FLOW}}) \times \mathrm{MV}_{\mathrm{max}}^{\mathrm{PRESS}} \qquad \mathbf{8.41(4)}$$

The cascade controller shown in Figure 8.41h serves to lower the pressure in the MCH by increasing the set points of the flow controllers serving the high permeability wells, according to Equation 8.41(4). This nonlinear algorithm stabilizes the MCH pressure at a level that is below the allowable maximum. These algorithms can also respond to fast pressure transients, because they do not contain any ramp functions that would limit the rate of rise of their measurements.

Protecting the Compressors

If the discharge pressure or the rate of change of the discharge pressure of any compressor exceeds its predetermined limit, the control system is automatically switched back into its start-up mode (TWM1). The system is also switched automatically to the start-up mode if the control valves on half of the operating wells exceed their 90% opening.

Alarm signals are actuated if the tale gas flow of the gas well that is under pressure-corrected cascade control reaches its maximum limit. Another condition that initiates a warning signal is when the flow rate to half of the active wells exceeds their predetermined maximum admissible load settings (MVAL). Another element in the safety logic of this control system is that if the opening of a valve exceeds 80%, it cannot be switched to cascade. In that case, the operator has the option of either activating new wells or switching controls of that well to the start-up mode (TWM1).

Additional safety-related logic steps include the stopping of compressors if, after activating the start-up mode (TWM1), the total lifted gas quantity (TGQ) is less than 25%. After the compressors are returned to stable operation at their expected OPs, the system can be returned to the operating mode (TWM2). The control system will automatically switch to the start-up mode (TWM1) if the discharge pressure or its rate of rise exceeds a preadjusted maximum limit value.

CONTROLLING PRODUCTION CYCLE

The natural gas that was compressed into the underground storage facility during the summer months is used up in the winter. Therefore, as the heating season approaches, the gas wells are switched from the injection (INC) to the production

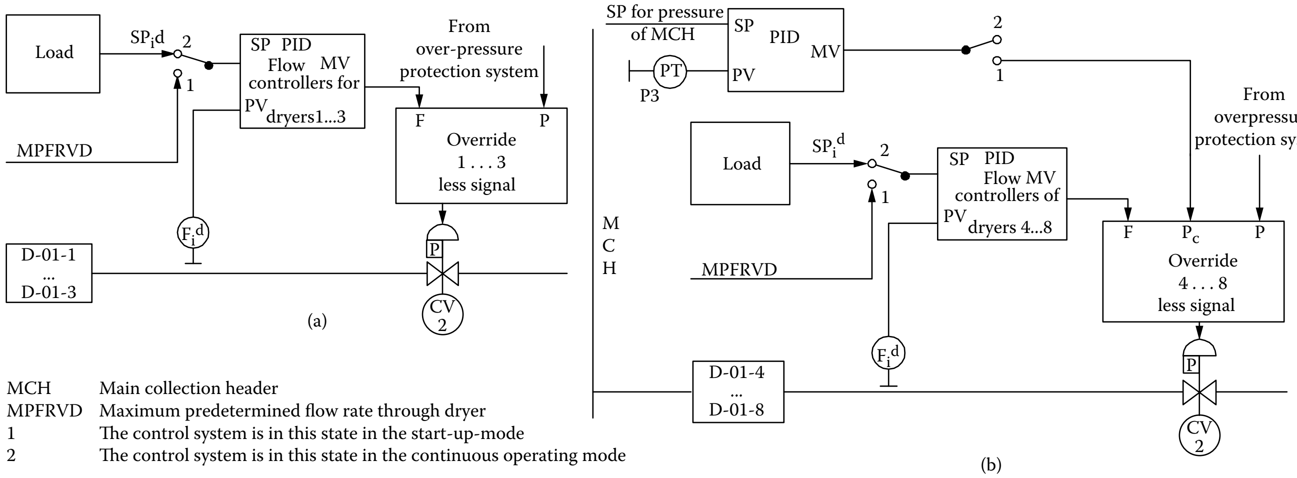


FIG. 8.41j
Dryer flow control algorithms for low- and high-capacity dryers. Note that the high-capacity dryers (b on the right of the figure), when in their start-up mode, are provided with a pressure override, which controls the pressure in the main collection header (MCH).

(PRC) mode. In this mode of operation the gas that is stored at about 1450 psig (100 bars) in the wells is admitted under flow control into the main collection header (MCH in Figures 8.41a, j, and k). This header is held at 1087 psig (75 bars) and supplies the large- and small-capacity dryers (Figure 8.41j).

The gas flow through the individual dryers is under flow control. The set points of the individual flow controllers (FIC-2 in Figure 8.41a) are adjusted so that, on the one hand, their total flow matches the total demand of the users. On the other hand, these set points are corrected to keep the pressure in the MCH (PT-3 in Figure 8.41a) at 1087 psig, while keeping the "zero point" pressure (the pressure at the beginning of the distribution header, detected by PT-4 in Figure 8.41a) below its high limit of 870 psig (60 bars).

Start-Up

After the START command, the flow controllers on the individual gas wells receive their set points on the basis of the LDA and move to that point at a preset rate (FRS). After the pressure of the MCH has stabilized, the required (expected) total gas flow (ETMP) is sent through the dryers in a distribution set by the operator.

In the start-up mode (TWM1) of the production cycle (PRC), the flow control loops on each active well are switched to their local automatic modes (LACM), and their set points are automatically established according to Equation 8.41(5) as follows:

$$SP_i^w = \left(Q_{\text{total}}^{\text{prescribed}} - \sum Q_i^{\text{manual}}\right)\frac{WF_i}{\sum_{i=1}^n WF_i} \qquad \mathbf{8.41(5)}$$

where

SP_i^w is the set point of the flow controller on the individual well
WF_i is the weighing factor for each individual well, which determines its share of the total flow required

The set points of the flow controllers are gradually increased according to the rate determined by a preprogrammed length of transient time (TDTR), which sets the new value of this set point, determining its slope. Similarly, the task of the control valves downstream of the dryers (CV-2 in Figures 8.41a, j, and k) is to maintain the pressure in the main collection header at a pressure of 1087 psig (75 bars). This is necessary in order to provide a steady pressure drop across the dryers. At the same time, the flow rate through the individual dryers is set to its predetermined maximum flow rate, and override controls are actuated if that limit is exceeded.

Pressure Correction During the production cycle, it is necessary to apply pressure correction to the MCH in order to maintain a constant pressure drop across the dryers, but this goal is subordinated to the necessity of meeting the total user's demand for conditioned gas.

The gas flow from the individual wells will be constant if the pressure in the MCH is stable, because this keeps the pressure drop across the well flow control valves (CV-1 in Figures 8.41a and j) constant. It is desirable to keep the MCH pressure relatively high, because that provides the driving force for high production rate. This relatively high pressure can be maintained as long as the pressure in all the active gas wells is greater than the pressure in the MCH.

Continuous Operation

After start-up, the system is switched into continuous operation (TWM2), during which the supply from this UGS must match the gas demand of the countrywide distribution network. The actual total gas flow (TGQ) obtained during start-up (TWM1) is usually not equal to the expected total (EFRD). In order to correctly match the supply to the demand and, thereby, stabilize the pressure in the main collection header, the control system has to be switched into its continuous operating mode (TWM2).

Before changeover, the operator must set the allowable set-point ramp rate for the dryers (SPR) and the admissible or required load or production rate for each dryer (AL). The pressure controllers (PCc in Figure 8.41a) that control the pressure in the MCH are configured as the cascade masters of the well flow controllers that are assigned to pressure correction. Also before changeover to the continuous production mode (TWM2), the individual well flow controllers must be in automatic and should be delivering the gas flows matching their set points. See Figure 8.41j for a schematic description of the two algorithms used for dryer flow controllers.

In the continuous mode of operation (TWM2), the system automatically performs the following functions: It calculates the total gas flow that the UGS should produce (ETMP). The operator then sets the set points of the dryer flow controllers based on this total required flow, using the LDA to make sure that each dryer is assigned the right flow. When the load demand is changing, the algorithm revises its recommended set points, which are then modified by the operator.

In determining the set points for the individual dryer flow controllers, the control system limits that value to the maximum flow rate that is allowable through the particular dryer (MFRD). If the calculated set point is below the MFRD of the dryer, the operator is not allowed to modify that setting. The dryer flow set point is calculated as follows:

$$\mathrm{SP}_i^d = \left(\mathrm{ETMP} - \sum_{i=1}^{n} \mathrm{MPFD}_i \right) \frac{\mathrm{MVAL}_i}{\sum_{i=1}^{n} \mathrm{MVAL}_i} \qquad \mathbf{8.41(6)}$$

where

SP_i^d is the set point for the particular dryer

ETMP is the required (expected) total amount of gas to be produced

MPFD_i is the predetermined maximum flow rate that the particular dryer can handle

MVAL_i is the maximum value of the flow that a particular well can produce

If some of the dryers are already operating at their maximum (MPFD), the required total production (ETMP) must be met by setting the flow controllers on the other dryers, in accordance with Equation 8.41(6).

Switching the Operating Modes When the operation is stable, the operator can switch the system back into the start-up mode (TWM1), but the normal mode of operation is continuous (TWM2). When the system is in the continuous mode, the set point of the flow controllers on the individual wells is cascaded for pressure correction (PCCM), while the flow controllers on the dryers are in their automatic flow control mode.

Switching between operating modes should be bumpless, and the pressure in the MCH should be controlled by pressure correction (PCCM). The set points of the dryer flow controllers remain the same as existed before the mode change. If the operator changes the set-point values, the rate of change is limited by the slope that is predetermined for each dryer (FRSD).

Control Modes and Loops

Figure 8.41k describes the equipment and controls used in the production mode of operation. The main components of this system include the wells that supply the gas; the MCH, which collects and transports the gas to the dryers; and the header, which takes the gas to the users. The demand for the gas is determined by the users, while the supply is a function of the gas flows from the operating wells. If the supply and demand are out of balance, the pressure in the MCH will change.

The control system described in Figure 8.41k serves to match the supply to the demand, while protecting the component subsystems (wells, dryers) from sudden upsets or from overpressure conditions. Because both the accuracy and the rangeability of the flow sensors are limited, pressure correction is used to maintain the balance between gas supply and demand.

Controlling the Wells

If the ultrasonic flowmeter that is serving the flow control loop of a particular well fails, the loop is switched to manual. Under this condition, the flow is estimated on the basis of control valve pressure drop and opening.

The well flow control valve (CV-1 in Figure 8.41a and k) is switched to the local automatic mode (LACM) whenever sudden upsets are experienced or when the operation mode is being switched between the start-up (TWM1) and continuous (TWM2) operating modes of operation. In this case, the flow controller set point is provided by the LDA, but can be modified by the operator. When the set point is being changed, the rate of change is limited. Minimum and maximum flow limits are also provided for each well.

When the well is in continuous operation (TWM2), pressure correction is applied to the flow controller's set point (PCCM). If a controller is in its remote automatic mode (RACM), the operator is not allowed to modify its set point.

EFRVD	Expected flow rate value through dryer
ETMP	Expected total mass produced
LACM	Local automatic control mode
LDAD	Load distribution algorithm for dryers
LDAW	Load distribution algorithm for wells
MPFRVD	Maximun predetermined flow rate value through the dryer
1	The control system is in this state in the start-up mode
2	The control system is in this state in the continuous operarting mode

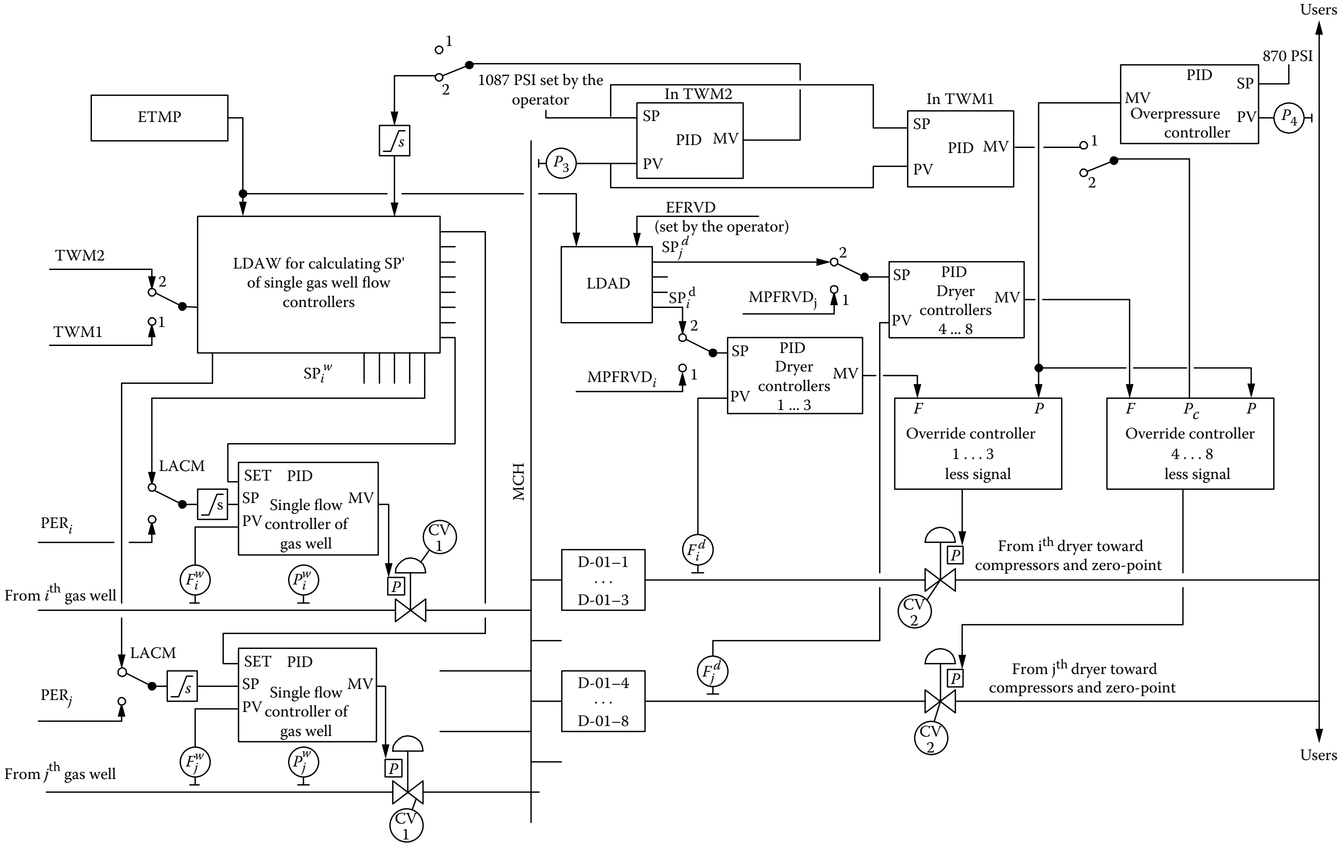


FIG. 8.41k

The overall control system of an underground gas storage facility, operating in its continuous-production mode (PRC).

Dryer Controls As shown on the left side of Figure 8.41j, the three small-capacity dryers are under flow control during start-up (TWM1). In this mode, the operator adjusts the dryer flow controller set point, using the LDA as a reference. When the total flow demanded by the users changes, the set points of the active dryer controllers are recalculated, based on the predetermined flow rate of the total system (PFRU). In addition to flow control, the dryer control loops are also provided with overpressure protection, as shown on the left side of Figure 8.41j.

The right side of Figure 8.41j illustrates the controls used on the five high-capacity dryers. In the start-up mode (TWM1), their controls are in the PCCM, but their set point cannot exceed the maximum flow (MPFD) set by the operator. In the continuous mode of operation (TWM2), the operation of these high-capacity dryers is similar to that of the small-capacity dryers. Therefore, they are operated in the PCCM in order to maintain the pressure in the MCH at the setting specified by the operator.

Pressure Correction In the start-up mode (TWM1), the set points of the control loops on the high-capacity dryers have to satisfy several considerations. They have to be under the maximum flow allowed for the particular dryer. In addition, they have to be pressure corrected. The pressure correction considers the pressure in the MCH (P_3 in the figures) and the pressure at the start of the supply header serving the users, which point is also called the “zero point” (P_4 in the figures).

In the continuous operating mode (TWM2), the flow controllers of the active wells develop a relatively steady pressure in the MCH. The FICs on the wells are in their RACMs, and their set points are adjusted by the operator. If the control valves (CV-1 in the figures) on half of the active the wells are more than 80% open, the operator is warned to lower the pressure in the MCH by reducing the set points of the FICs on the wells.

Warning the Operator Usually, the load redistribution is recommended to be done per Equation 8.41(3). When this is not feasible, the control system advises the operator if the flow set point from a particular well is no longer available for pressure correction. The operator is also advised by the control system if the maximum or minimum allowable flow limit is reached for a particular well. The control system also warns the operator if any well control valve has opened beyond 80% or if the flow through them is outside allowable limits. When the demand for gas cannot otherwise be met, the operator must activate new, previously passive wells.

Normally, the control system keeps the total flow from the wells that are under pressure-corrected control within the minimum and maximum tale flow limits (mCC and MCC). If the flow from the wells that are controlled in the PCCM reaches the preset limit for maximum tale flow (MCC), the control system must warn the operator.

The control system also warns the operator if any well control loop with a pressure-corrected set point (PCCM) opens its control valve to more than 80%.

Controlling the Pressure at "Zero Point" The "zero-point" pressure is the pressure at the beginning of the header that serves the users (P_4 in Figures 8.41a and k). This supply pressure to the users is not controlled directly, as it is the consequence of the balance between the gas supply and demand. Therefore, it is allowed to float between 725 and 870 psig (50 to 60 bars).

At the beginning of the PRC, the outlet pressure of the active wells is about 1450 psig (100 bars), and under such conditions, overpressure conditions can arise. As shown in Figure 8.41k, the overpressure protection control loop is set at 870 psig (60 bars), and if the zero-point pressure reaches that level, the controller will start to throttle down the dryer flow control valves (CV-2). A high-pressure alarm is also provided.

Just as the pressure in the main collection header is a function of the balance between the total flows of the wells and the dryers, the zero-point pressure is an indication of the balance between the flow supplied by the dryers and the flow demanded by the users. In more sophisticated control systems, feedforward control can be used on the basis of the gas demand, so that the system does not wait until a pressure upset is caused by the supply–demand imbalance, but eliminates it as soon as it occurs.

THE TOTAL CONTROL SYSTEM

The control system is so designed that it has to be switched into the start-up mode whenever the OP for the overall facility is changed or when upsets occur. In that mode (TWM1), the individual wells are on straight flow control. The LDA is used to distribute the total gas production (ETMI), according to the WFs of the individual wells.

Because of the limited reliability of the flow sensors under changing flow conditions, the total gas flow obtained from the wells (ETMI) is not accurately known during start-up. Once the flows have stabilized, the system is switched into the continuous operating mode (TWM2). In this mode, when the operating point for the whole system (OP) changes, the pressure of the main collection header (MCH) reflects that change, and the automatic pressure correction (PCCM) adjusts the flow set points at the individual wells.

As shown in Figures 8.41j and k, the control system is reconfigured as the operating mode is switched from start-up (TWM1) to continuous (TWM2). In the figures these modes are indicated by the numbers 1 and 2.

Existing DCS Hardware

The control hardware consists of two fiber-optic bus-connected DCS units. The first DCS contains the controls for the well side of the total UGS, ten field control stations (FCS), four human interface stations (HIS), three operator consoles, one PLC serving data acquisition, another PLC for safety controls, and a number of ultrasonic flowmeters.

The second DCS controls the gas conditioning and drying process. It is provided with two field and two operator consoles.

All the FCSs, including their controllers, digital and analog I/O, CPU, power, and communication boards, are redundant. The HISs consist of an industrial computer with hard disc, 21 in. monitor, trackball, and keyboard.

CONCLUSIONS AND COMMENTS

The goal of the existing control system is to accurately meet the demand for natural gas of the users while minimizing upsets or disturbances. This task is made more difficult by the changing conditions in the wells during both charging (INC) and discharging (PRC) of the wells. An added source of difficulty is the low accuracy and unreliability of the flow sensors. In addition, the capacity and safety limitations of both the wells and the dryers cannot be violated and is desired to minimize the energy cost and maintenance of the operation.

One of the tools used to achieve these goals is to apply pressure correction to the flow controllers at both the wells and the dryers. Another tool is to use the pressure drop ratio ($\Delta p/p_1$) and the control valve openings to estimate the flows passing through them.

The reader might notice that the control system described in this section is one that probably evolved from a manually operated facility; it is still operating with unreliable sensors when water is present in the gas and control valves that have no direct stroke position detectors. Under these conditions, it

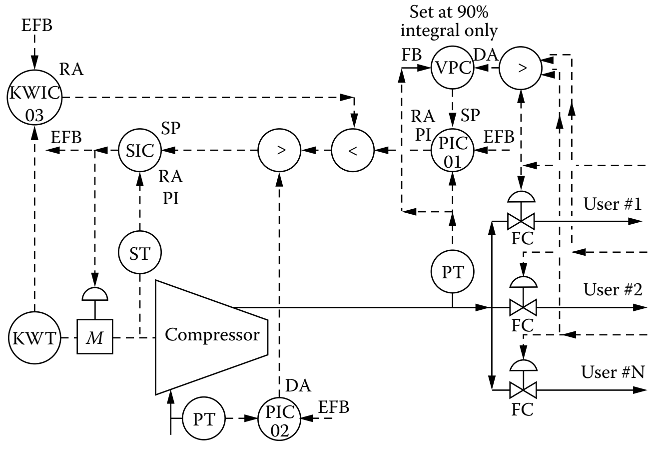


FIG. 8.41l
Optimized load-following controls satisfy the users at minimum energy cost while providing protection against low suction pressure and against overloading the compressor drive.

is not surprising that the addition of DCS controls resulted only in a semimanual mode of operation, with the DCS system serving to correct for the inadequacy of the in-line hardware and fulfilling only supervisory functions.

If the controls of a new gas storage facility were being designed today, in addition to the improvements in the sensors, one would use more model predictive feedforward, and adaptive controls. In the PRC, such feedforward controls would be based on supply-demand balancing. This is because the pressure in the MCH is a function of the balance between the total flow from the wells and the total flow to the dryers. Similarly, the zero-point pressure is an indication of the balance between the total flow from the dryers and the total flow demanded by the users. Consequently, in a modern and automated control system, feedforward control would be based on the gas demand, so that the system does not have to wait until a pressure upset is caused by the imbalance between supply and demand imbalance, but can eliminate it as soon as it occurs.

Similarly, in the reinjection mode (INC), the goal is to refill the wells as quickly as possible without causing damage to them and to do that at a minimum energy cost at the compressors. In this mode, the users are the wells, the suppliers are the compressors, and supply–demand control at minimum energy consumption can be provided by feedforward-based valve position control (Figure 8.41l).

References

1. Szecső, G. and Tornyi, L., Control System of Underground Gas Storage, *Proceedings of 6th International Conference on Electronics and Automation,* April 10–12, 2001, pp. 73–78, p. 165.
2. Szecső, G., Control Strategies of Underground Gas Storage, INES'99 IEEE International Conference on Intelligent Engineering Systems, Nov. 1–3, Poprad, Slovakia, *Proceedings,* pp. 177–181, 1999.
3. Szecső, G., The Function of the Simulation in Control Structure Design of the Underground Gas Storage System, *Proceedings of 4th Workshop on European Scientific and Industrial Collaboration,* Vol. I, pp. 243–250. WESIC 2003 Advanced Technologies in Manufacturing, May 28–30, 2003.
4. Szecső, G., The Function of the Simulation in Control Structure Design of the Underground Gas Storage System, *Proceedings of 5th International Carpathian Control Conference,* Vol. I, pp. 521–526, Zakopane, Poland, May 25–28, 2004.

Bibliography

1. Cox, C. S., Arden, W. J. B., and Fletcher, I., Investigations into the Remote Control of Gas Transmission Pipelines, *Proceed. Int. AMSE Conf. Modeling & Simulation,* Sorrento, Italy, Vol. 3.3, pp. 33–44, 1986.
2. Ocelíková, E., and Madarász, L., Contribution of Creation of Macrosituations of Complex Systems, *Journal of Advanced Computational Intelligence,* 2002, pp. 20–25.
3. Arden. W. J. B., and Cox, C. S., Remote Boundary Pressure Control, *Measurement + Control,* Vol. 20, pp. 56–60, November 1987.
4. Tornyi, L., Realization Plan of Controlling Underground Natural Gas Storage System, Hajdúszoboszló, 1998 (in Hungarian), Hungarian Gas and Oil Company (MOL Rt.).
5. Douglas, J. M., *Process Dynamics and Control,* Volumes 1 and 2, Englewood Cliffs, NJ: Prentice Hall, 1972.

Appendix

A.1 International System of Units

B. G. LIPTÁK

The decimal system of units was conceived in the 16th century when there was a great confusion and jumble of units of weights and measures. It was not until 1790, however, that the French National Assembly requested the French Academy of Sciences to work out a system of units suitable for adoption by the entire world. This system, based on the metre (meter) as a unit of length and the gram as a unit of mass, was adopted as a practical measure to benefit industry and commerce. Physicists soon realized its advantages and it was adopted also in scientific and technical circles. The importance of the regulation of weights and measures was recognized in Article 1, Section 8, when the U.S. Constitution was written in 1787, but the metric system was not legalized in this country until 1866. In 1893, the international meter and kilogram became the accepted standards of length and mass in the United States, both for metric and customary weights and measures. The tables of conversion factors presented in the following pages are intended to serve two purposes:

1. To express the definitions of miscellaneous units of measure as exact numeral multiples of coherent "metric" units. Relationships that are exact in terms of the base unit are followed by an asterisk. Relationships that are not followed by an asterisk are either the results of physical measurements or are only approximate.
2. To provide multiplying factors for converting expressions of measurements given by numbers and miscellaneous units to corresponding new numbers and metric units.

Conversion factors are presented for ready adaptation to computer readout and electronic data transmission. The factors are written as a number equal to or greater than 1 and less than 10 with six or fewer decimal places. This number is followed by the letter E (for exponent), a plus or minus symbol, and two digits that indicate the power of 10 by which the number must be multiplied to obtain the correct value. For example:

$$3.523\,907 \text{ E} - 02 \text{ is } 3.523\,907 \times 10^{-2}$$

or

$$0.035\,239\,07$$

Similarly,

$$3.386\,389 \text{ E} + 03 \text{ is } 3.386\,389 \times 10^{3}$$

or

$$3\,386.389$$

An asterisk (*) after the sixth decimal place indicates that the conversion factor is exact and that all subsequent digits are zero.

When a figure is to be rounded to fewer digits than the total number available, the procedure should be as follows:

1. When the first digit discarded is less than 5, the last digit retained should not be changed. For example, 3.463 25, if rounded to four digits, would be 3.463; if rounded to three digits, 3.46.
2. When the first digit discarded is greater than 5 or if it is a 5 followed by at least one digit other than 0, the last figure retained should be increased by one unit. For example, 8.376 52, if rounded to four digits, would be 8.377; if rounded to three digits, 8.38.
3. When the first digit discarded is exactly 5, followed only by zeros, the last digit retained should be rounded upward if it is an odd number, but no adjustment made if it is an even number. For example, 4.365, when rounded to three digits, becomes 4.36. The number 4.355 would also round to the same value, 4.36, if rounded to three digits.

Where fewer than six decimal places is shown, more precision is not warranted.

TABLE A.1a
International System of Units

Quantity	*Unit*	*SI Symbol*	*Formula*
	Base Units		
length	meter	m	—
mass	kilogram	kg	—
time	second	s	—
electric current	ampere	A	—
thermodynamic temperature	kelvin	K	—
amount of substance	mole	mol	—
luminous intensity	candela	cd	—
	Supplementary Units		
plane angle	radian	rad	—
solid angle	steradian	sr	—
	Derived Units		
acceleration	meter per second squared	—	m/s^2
activity (of a radioactive source)	disintegration per second	—	(disintegration)/s
angular acceleration	radian per second squared	—	rad/s^2
angular velocity	radian per second	—	rad/s
area	square meter	—	m^2
density	kilogram per cubic meter	—	kg/m^3
electric capacitance	farad	F	A·s/V
electrical conductance	siemens	S	A/V
electric field strength	volt per meter	—	V/m
electric inductance	henry	H	V·s/A
electric potential difference	volt	V	W/A
electric resistance	ohm	Ω	V/A
electromotive force	volt	V	W/A
energy	joule	J	N·m
entropy	joule per kelvin	—	J/K
force	newton	N	$kg{\cdot}m/s^2$
frequency	hertz	Hz	(cycle)/s
illuminance	lux	lx	lm/m^2
luminance	candela per square meter	—	cd/m^2
luminous flux	lumen	lm	cd·sr
magnetic field strength	ampere per meter	—	A/m
magnetic flux	weber	Wb	V·s
magnetic flux density	tesla	T	Wb/m^2
magnetomotive force	ampere	A	—
power	watt	W	J/s
pressure	pascal	Pa	N/m^2
quantity of electricity	coulomb	C	A·s
quantity of heat	joule	J	N·m
radiant intensity	watt per steradian	—	W/sr
specific heat	joule per kilogram-kelvin	—	J/kg·K
stress	pascal	Pa	N/m^2
thermal conductivity	watt per meter-kelvin	—	W/m·K
velocity	meter per second	—	m/s
viscosity, dynamic	pascal-second	—	Pa·s
viscosity, kinematic	square meter per second	—	m^2/s
voltage	volt	V	W/A
volume	cubic meter	—	m^3
wavenumber	reciprocal meter	—	(wave)/m
work	joule	J	N·m

TABLE A.1b

Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	*To*	*Multiply by*
	A	
abampere	ampere (A)	1.000 000*E+01
abcoulomb	coulomb (C)	1.000 000*E+01
abfarad	farad (F)	1.000 000*E+09
abhenry	henry (H)	1.000 000*E+09
abmho	siemens (S)	1.000 000*E+09
abohm	ohm (Ω)	1.000 000*E+09
abvolt	volt (V)	1.000 000*E+08
acre foot (U.S. survey)[a]	meter^3 (m^3)	1.233 489 E+03
acre (U.S. survey)[a]	meter^2 (m^2)	4.046 873 E+03
ampere hour	coulomb (C)	3.600 000*E+03
are	meter^2 (m^2)	1.000 000*E+02
angstrom	meter (m)	1.000 000*E−10
astronomical unit	meter (m)	1.495 979 E+11
atmosphere (standard)	pascal (Pa)	1.013 250*E+05
atmosphere (technical = 1 kgf/cm^2)	pascal (Pa)	9.806 650*E+04

[a] Since 1893, the U.S. basis of length measurement has been derived from metric standards. In 1959, a small refinement was made in the definition of the yard to resolve discrepancies both in this country and abroad, which changed its length from 3600/3937 m to 0.9144 m exactly. This resulted in the new value being shorter by two parts in a million.

At the same time it was decided that any data in feet derived from and published as a result of geodetic surveys within the United States would remain with the old standard (1 ft = 1200/3937 m) until further decision. This foot is named the U.S. survey foot.

As a result, all U.S. land measurements in U.S. customary units will relate to the meter by the old standard. All the conversion factors in these tables for units referenced to this footnote are based on the U.S. survey foot, rather than the international foot.

Conversion factors for the land measures given below may be determined from the following relationships:

1 league = 3 miles (exactly)
1 rod = 16$^1/_2$ feet (exactly)
1 section = 1 square mile (exactly)
1 township = 36 square miles (exactly)
1 chain = 66 feet (exactly)

To Convert from	*To*	*Multiply by*
	B	
bar	pascal (Pa)	1.000 000*E+05
barn	meter^2 (m^2)	1.000 000*E−28
barrel (for petroleum, 42 gal)	meter^3(m^3)	1.589 873 E−01
board foot	meter^3(m^3)	2.359 737 E−03
British thermal unit (International Table)[b]	joule (J)	1.055 056 E+03
British thermal unit (mean)	joule (J)	1.055 87 E+03
British thermal unit (thermochemical)	joule (J)	1.054 350 E+03
British thermal unit (39°F)	joule (J)	1.059 67 E+03
British thermal unit (59°F)	joule (J)	1.054 80 E+03
British thermal unit (60°F)	joule (J)	1.054 68 E+03
Btu (International Table) · ft/h · ft^2 · °F (*k*, thermal conductivity)	watt per meter-kelvin (W/m · K)	1.730 735 E+00
Btu (thermochemical) · ft/h · ft^2 · °F (*k*, thermal conductivity)	watt per meter-kelvin (W/m · K)	1.729 577 E+00
Btu (International Table) · in./h · ft^2 · °F (*k*, thermal conductivity)	watt per meter-kelvin (W/m · K)	1.442 279 E−01
Btu (thermochemical) · in./h · ft^2 · °F (*k*, thermal conductivity)	watt per meter-kelvin (W/m · K)	1.441 314 E−01
Btu (International Table) · in./s · ft^2 · °F (*k*, thermal conductivity)	watt per meter-kelvin (W/m · K)	5.192 204 E+02
Btu (thermochemical) · in./s · ft^2 · °F (*k*, thermal conductivity)	watt per meter-kelvin (W/m · K)	5.188 732 E+02
Btu (International Table)/h	watt (W)	2.930 711 E−01
Btu (International Table)/s	watt (W)	1.055 056 E+03
Btu (thermochemical)/h	watt (W)	2.928 751 E+01
Btu (thermochemical)/min	watt (W)	1.757 250 E+01
Btu (thermochemical)/s	watt (W)	1.054 350 E+03
Btu (International Table)/ft^2	joule per meter^2 (J/m^2)	1.135 653 E+04

[b] This value was adopted in 1956. Some of the older International Tables use the value 1.055 04 E+03. The exact conversion factor is 1.055 055 852 62*E+03.

TABLE A.1b Continued

Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	*To*	*Multiply by*
Btu (thermochemical)/ft^2	joule per meter^2 (J/m^2)	1.134 893 E+04
Btu (thermochemical)/ft$^2 \cdot$ h	watt per meter^2 (W/m^2)	3.152 481 E+00
Btu (thermochemical)/ft$^2 \cdot$ min	watt per meter^2 (W/m^2)	1.891 489 E+02
Btu (thermochemical)/ft$^2 \cdot$ s	watt per meter^2 (W/m^2)	1.134 893 E+04
Btu (thermochemical)/in.$^2 \cdot$ s	watt per meter^2 (W/m)^2	1.634 246 E+06
Btu (International Table)/h $\cdot$ ft$^2 \cdot$ °F (*C*, thermal conductance)	watt per meter^2-kelvin (W/m$^2 \cdot$ K)	5.678 263 E+00
Btu (thermochemical)/h $\cdot$ ft$^2 \cdot$ °F (*C*, thermal conductance)	watt per meter^2-kelvin (W/m$^2 \cdot$ K)	5.674 466 E+00
Btu (International Table)/s $\cdot$ ft$^2 \cdot$ °F	watt per meter^2-kelvin (W/m$^2 \cdot$ K)	2.044 175 E+04
Btu (thermochemical)/s $\cdot$ ft$^2 \cdot$ °F	watt per meter^2-kelvin (W/m$^2 \cdot$ K)	2.042 808 E+04
Btu (International Table)/lb	joule per kilogram (J/kg)	2.326 000*E+03
Btu (thermochemical)/lb	joule per kilogram (J/kg)	2.324 444 E+03
Btu (International)/lb $\cdot$ °F (*c*, heat capacity)	joule per kilogram-kelvin (J/kg $\cdot$ K)	4.186 800*E+03
Btu (thermochemical)/lb $\cdot$ °F (*c*, heat capacity)	joule per kilogram-kelvin (J/kg $\cdot$ K)	4.184 000 E+03
bushel (U.S.)	meter^3 (m^3)	3.523 907 E–02
	C	
caliber (inch)	meter (m)	2.540 000*E–02
calorie (International Table)	joule (J)	4.186 800*E+00
calorie (mean)	joule (J)	4.190 02 E+00
calorie (thermochemical)	joule (J)	4.184 000*E+00
calorie (15°C)	joule (J)	4.185 80 E+00
calorie (20°C)	joule (J)	4.181 90 E+00
calorie (kilogram, International Table)	joule (J)	4.186 800*E+03
calorie (kilogram, mean)	joule (J)	4.190 02 E+03
calorie (kilogram, thermochemical)	joule (J)	4.184 000*E+03
cal (thermochemical)/cm^2	joule per meter^2 (J/m^2)	4.184 000*E+04
cal (International Table)/g	joule per kilogram (J/kg)	4.186 800*E+03
cal (thermochemical)/g	joule per kilogram (J/kg)	4.184 000*E+03
cal (International Table)/g $\cdot$ °C	joule per kilogram-kelvin (J/kg $\cdot$ K)	4.186 800*E+03
cal (thermochemical)/g $\cdot$ °C	joule per kilogram-kelvin (J/kg $\cdot$ K)	4.184 000*E+08
cal (thermochemical)/min	watt (W)	6.973 333 E–02
cal (thermochemical)/s	watt (W)	4.184 000*E+00
cal (thermochemical)/cm$^2 \cdot$ min	watt per meter^2 (W/m^2)	6.973 333 E+02
cal (thermochemical)/cm$^2 \cdot$ s	watt per meter^2 (W/m^2)	4.184 000*E+04
cal (thermochemical)/cm $\cdot$ s $\cdot$ °C	watt per meter-kelvin (W/m $\cdot$ K)	4.184 000*E+02
carat (metric)	kilogram (kg)	2.000 000*E–04
centimeter of mercury (0°C)	pascal (Pa)	1.333 22 E+03
centimeter of water (4°C)	pascal (Pa)	9.806 38 E+01
centipoise	pascal second (Pa $\cdot$ s)	1.000 000*E–03
centistokes	meter^2 per second (m^2/s)	1.000 000*E–06
circular mil	meter^2 (m^2)	5.067 075 E–10
clo	kelvin meter^2 per watt (K $\cdot$ m^2/W)	2.003 712 E–01
cup	meter^3 (m^3)	2.365 882 E–04
curie	becquerel (Bq)	3.700 000*E+10
	D	
day (mean solar)	second (s)	8.640 000 E+04
day (sidereal)	second (s)	8.616 409 E+04
degree (angle)	radian (rad)	1.745 329 E–02
degree Celsius	Kelvin (K)	

TABLE A.1b Continued

Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	*To*	*Multiply by*
degree Centigrade	[see footnote[c]]	$t_K = t_{°C}+273.15$
degree Fahrenheit	degree Celsius	$t_{°C} = (t_{°F}-32)/1.8$
degree Fahrenheit	kelvin (K)	$t_K = (t_{°F}+459.67)/1.8$
degree Rankine	kelvin (K)	$t_K = t_{°R}/1.8$
°F·h· ft^2/Btu (International Table) (*R*, thermal resistance)	kelvin meter^2 per watt (K·m^2/W)	1.761 102 E–01
°F·h·ft^2/Btu (thermochemical (*R*, thermal resistance)	kelvin meter^2 per watt (K·m^2/W)	1.762 280 E–01
denier	kilogram per meter (kg/m)	1.111 111 E–07
dyne	newton (N)	1.000 000*E–05
dyne/cm	newton meter (N·m)	1.000 000*E–07
dyne/cm^2	pascal (Pa)	1.000 000*E–01
	E	
electronvolt	joule (J)	1.602 19 E–19
EMU of capacitance	farad (F)	1.000 000*E+09
EMU of current	ampere (A)	1.000 000*E+01
EMU of electric potential	volt (V)	1.000 000*E–08
EMU of inductance	henry (H)	1.000 000*E–09
EMU of resistance	ohm (Ω)	1.000 000*E–09
ESU of capacitance	farad (F)	1.112 650 E–12
ESU of current	ampere (A)	3.335 6 E–10
ESU of electric potential	volt (V)	2.997 9 E+02

[c] The SI unit of thermodynamic temperature is the kelvin (K), and this unit is properly used for expressing thermodynamic temperature and temperature intervals. Wide use is also made of the degree Celsius (°C), which is the SI unit for expressing Celsius temperature and temperature intervals. The Celsius scale (formerly called Centigrade) is related directly to thermodynamic temperature (kelvins) as follows:

1. The temperature interval one degree Celsius equals one kelvin exactly.
2. Celsius temperature (t) is related to thermodynamic temperature (T) by the equation $t = T - T_0$, where $T_0 = 273.15$ K by definition.

To Convert from	*To*	*Multiply by*
ESU of inductance	henry (H)	8.987 554 E+11
ESU of resistance	ohm (Ω)	8.987 554 E+11
erg	joule (J)	1.000 000*E–07
erg/(cm^2·s)	watt per meter^2 (W/m^2)	1.000 000*E–03
erg/s	watt (W)	1.000 000*E–07
	F	
faraday (based on carbon-12)	coulomb (C)	9.648 70 E+04
faraday (chemical)	coulomb (C)	9.649 57 E+04
faraday (physical)	coulomb (C)	9.652 19 E+04
fathom	meter (m)	1.828 8 E+00
fermi (femtometer)	meter (m)	1.000 000*E–15
fluid ounce (U.S.)	meter^3 (m^3)	2.957 353 E–05
foot	meter (m)	3.048 000*E–01
foot (U.S. survey)[a]	meter (m)	3.048 006 E–01
foot of water (39.2 °F)	pascal (Pa)	2.988 98 E+03
ft^2	meter^2 (m^2)	9.290 304*E–02
ft^2/h (thermal diffusivity)	meter^2 per second (m^2/s)	2.580 640*E–05
ft^2/s	meter^2 per second (m^2/s)	9.290 304*E–02
ft^3 (volume; section modulus)	meter^3 (m^3)	2.831 685 E–02
ft^3/min	meter^3 per second (m^3/s)	4.719 474 E–04
ft^3/s	meter^3 per second (m^3/s)	2.831 685 E–02
ft^4 (moment of section)[d]	meter^4 (m^4)	8.630 975 E–03
ft/h	meter per second (m/s)	8.466 667 E–05
ft/min	meter per second (m/s)	5.080 000*E–03
ft/s	meter per second (m/s)	3.048 000*E–01
ft/s^2	meter per second^2 (m/s^2)	3.048 000*E–01

[d] This is sometimes called the moment of inertia of a plane section about a specified axis.

TABLE A.1b Continued

Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	*To*	*Multiply by*
footcandle	lux (lx)	1.076 391 E+01
footlambert	candela per meter^2 (cd/m^2)	3.426 259 E+00
ft·lbf	joule (J)	1.355 818 E+00
ft·lbf/h	watt (W)	3.766 161 E–04
ft·lbf/min	watt (W)	2.259 697 E–02
ft·lbf/s	watt (W)	1.355 818 E+00
ft·poundal	joule (J)	4.214 011 E–02
free fall, standard (g)	meter per second^2 (m/s^2)	9.806 650*E+00
	G	
gal	meter per second^2 (m/s^2)	1.000 000*E–02
gallon (Canadian liquid)	meter^3 (m^3)	4.546 090 E–03
gallon (U.K. liquid)	meter^3 (m^3)	4.546 092 E–03
gallon (U.S. dry)	meter^3 (m^3)	4.404 884 E–03
gallon (U.S. liquid)	meter^3 (m^3)	3.785 412 E–03
gallon (U.S. liquid) per day	meter^3 per second (m^3/s)	4.381 264 E–08
gallon (U.S. liquid) per minute	meter^3 per second (m^3/s)	6.309 020 E–05
gallon (U.S. liquid) per hp·h (SFC, specific fuel consumption)	meter^3 per joule (m^3/J)	1.410 089 E–09
gamma	tesla (T)	1.000 000*E–09
gauss	tesla (T)	1.000 000*E–04
gilbert	ampere (A)	7.957 747 E–01
gill (U.K.)	meter^3 (m^3)	1.420 654 E–04
gill (U.S.)	meter^3 (m^3)	1.182 941 E–04
grad	degree (angular)	9.000 000*E–01
grad	radian (rad)	1.570 796 E–02
grain (1/7000 lb avoirdupois)	kilogram (kg)	6.479 891*E–05
grain (lb avoirdupois/7000)/gal (U.S. liquid)	kilogram per meter^3 (kg/m^3)	1.711 806 E–02
gram	kilogram (kg)	1.000 000*E–03
g/cm^3	kilogram per meter^3 (kg/m^3)	1.000 000*E+03
gram-force/cm^2	pascal (Pa)	9.806 650*E+01
	H	
hectare	meter^2 (m^2)	1.000 000*E+04
horsepower (550 ft·lbf/s)	watt (W)	7.456 999 E+02
horsepower (boiler)	watt (W)	9.809 50 E+03
horsepower (electric)	watt (W)	7.460 000*E+02
horsepower (metric)	watt (W)	7.354 99 E+02
horsepower (water)	watt (W)	7.460 43 E+02
horsepower (U.K.)	watt (W)	7.457 0 E+02
hour (mean solar)	second (s)	3.600 000 E+03
hour (sidereal)	second (s)	3.590 170 E+03
hundredweight (long)	kilogram (kg)	5.080 235 E+01
hundredweight (short)	kilogram (kg)	4.535 924 E+01
	I	
inch	meter (m)	2.540 000*E–02
inch of mercury (32°F)	pascal (Pa)	3.386 38 E+03
inch of mercury (60°F)	pascal (Pa)	3.376 85 E+03
inch of water (39.2°F)	pascal (Pa)	2.490 82 E+03
inch of water (60°F)	pascal (Pa)	2.488 4 E+02
in.^{2}	meter^2 (m^2)	6.451 600*E–04
in.^{3} (volume; section modulus)[e]	meter^3 (m^3)	1.638 706 E–05
in.^{3}/min	meter^3 per second (m^3/s)	2.731 177 E–07
in.^{4} (moment of secion)[4]	meter^4 (m^4)	4.162 314 E–07
in./s	meter per second (m/s)	2.540 000*E–02
in./s^2	meter per second^2 (m/s^2)	2.540 000*E–02
	K	
kayser	1 per meter (1/m)	1.000 000*E+02
kelvin	degree Celsius	$t_{°C} = t_K - 273.15$
kilocalorie (International Table)	joule (J)	4.186 800*E+03

[e] The exact conversion factor is 1.638 706 4*E–05.

TABLE A.1b Continued

Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	To	Multiply by
kilocalorie (mean)	joule (J)	4.190 02 E+03
kilocalorie (thermochemical)	joule (J)	4.184 000*E+03
kilocalorie (thermochemical)/min	watt (W)	6.973 333 E+01
kilocalorie (thermochemical)/s	watt (W)	4.184 000*E+03
kilogram-force (kgf)	newton (N)	9.806 650*E+00
kgf·m	newton meter (N·m)	9.806 650*E+00
kgf·s^2/m (mass)	kilogram (kg)	9.806 650*E+00
kgf/cm^2	pascal (Pa)	9.806 650*E+04
kgf/m^2	pascal (Pa)	9.806 650*E+00
kgf/mm^2	pascal (Pa)	9.806 650*E+06
km/h	meter per second (m/s)	2.777 778 E–01
kilopond	newton (N)	9.806 650*E+00
kW·h	joule (J)	3.600 000*E+06
kip (1000 lbf)	newton (N)	4.448 222 E+03
kip/in^2 (ksi)	pascal (Pa)	6.894 757 E+06
knot (international)	meter per second (m/s)	5.144 444 E–01
	L	
lambert	candela per meter^2 (cd/m^2)	1/π *E+04
lambert	candela per meter^2 (cd/m^2)	3.183 099 E+03
langley	joule per meter^2 (J/m^2)	4.184 000*E+04
league	meter (m)	[see footnote a]
light year	meter (m)	9.460 55 E+15
liter[f]	meter^3 (m^3)	1.000 000*E–03
	M	
maxwell	weber (Wb)	1.000 000*E–08
mho	siemens (S)	1.000 000*E+00
microinch	meter (m)	2.540 000*E–08

To Convert from	To	Multiply by
micron	meter (m)	1.000 000*E–06
mil	meter (m)	2.540 000*E–05
mile (international)	meter (m)	1.609 344*E+03
mile (statute)	meter (m)	1.609 3 E+03
mile (U.S. survey)[a]	meter (m)	1.609 347 E+03
mile (international nautical)	meter (m)	1.852 000*E+03
mile (U.K. nautical)	meter (m)	1.853 184*E+03
mile (U.S. nautical)	meter (m)	1.852 000*E+03
mi^2 (international)	meter^2 (m^2)	2.589 988 E+06
mi^2 (U.S. survey)[a]	meter^2 (m^2)	2.589 998 E+06
mi/h (international)	meter per second (m/s)	4.470 400*E–01
mi/h (international)	kilometer per hour (km/h)	1.609 344*E+01
mi/min (international)	meter per second (m/s)	2.682 240*E+01
mi/s (international)	meter per second (m/s)	1.609 344*E+03
millibar	pascal (Pa)	1.000 000*E+02
millimeter of mercury (0°C)	pascal (Pa)	1.333 22 E+02
minute (angle)	radian (rad)	2.908 882 E–04
minute (mean solar)	second (s)	6.000 000 E+01
minute (sidereal)	second (s)	5.983 617 E+01
month (mean calendar)	second (s)	2.628 000 E+06
	O	
oersted	ampere per meter (A/m)	7.957 747 E+01
ohm centimeter	ohm meter (Ω·m)	1.000 000*E–02
ohm circular-mill per foot	ohm millimeter^2 per meter (Ω·mm^2/m)	1.662 426 E–03
ounce (avoirdupois)	kilogram (kg)	2.834 952 E–02
ounce (troy or apothecary)	kilogram (kg)	3.110 348 E–02
ounce (U.K. fluid)	meter^3 (m^3)	2.841 307 E–05
ounce (U.S. fluid)	meter^3 (m^3)	2.957 353 E–05

[f] In 1964 the General Conference on Weights and Measures adopted the name liter as a special name for decimeter. Prior to this decision the liter differed slightly (previous value, 1.000028 dm^3) and in expression of precision volume measurement this fact must be kept in mind.

TABLE A.1b Continued
Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	*To*	*Multiply by*
ounce-force	newton (N)	2.780 139 E–01
ozf·in.	newton meter (N·m)	7.061 552 E–03
oz (avoirdupois)/gal (U.K. liquid)	kilogram per meter^3 (kg/m^3)	6.236 021 E+00
oz (avoirdupois)/gal (U.S. liquid)	kilogram per meter^3 (kg/m^3)	7.489 152 E+00
oz (avoirdupois)/in^3	kilogram per meter^3 (kg/m^3)	1.729 994 E+03
oz (avoirdupois)/ft^2	kilogram per meter^3 (kg/m^3)	3.051 517 E–01
oz (avoirdupois)/yd^2	kilogram per meter^2 (kg/m^2)	3.390 575 E–02
	P	
parsec	meter (m)	3.085 678 E+16
peck (U.S.)	meter^3 (m^3)	8.809 768 E–03
pennyweight	kilogram (kg)	1.555 174 E–03
perm (0°C)	kilogram per pascal second meter^2 (kg/Pa·s·m^2)	5.721 35 E–11
perm (23°C)	kilogram per pascal second meter^2 (kg/Pa·s·m^2)	5.745 25 E–11
	kilogram per pascal second meter (kg/Pa·s·m)	1.453 22 E–12
perm·in. (0°C)	kilogram per pascal second meter (kg/Pa·s·m)	1.453 22 E–12
perm·in. (23°C)	kilogram per pascal second meter (kg/Pa·s·m)	1.459 29 E–12
phot	lumen per meter^2 (lm/m^2)	1.000 000*E+04
pica (printer's)	meter (m)	4.217 518 E–03
pint (U.S. dry)	meter^3 (m^3)	5.506 105 E–04
pint (U.S. liquid)	meter^3 (m^3)	4.731 765 E–04
point (printer's)	meter (m)	3.514 598*E–04
poise (absolute viscosity)	pascal second (Pa·s)	1.000 000*E–01
pound (lb avoidrupois)[g]	kilogram (kg)	4.535 924 E–01
pound (troy or apothecary)	kilogram (kg)	3.732 417 E–01
lb·ft^2 (moment of inertia)	kilogram meter^2 (kg·m^2)	4.214 011 E–02
lb·in.^{2} (moment of inertia)	kilogram meter^2 (kg·m^2)	2.926 397 E–04
lb/ft·h	pascal second (Pa·s)	4.133 789 E–04
lb/ft·s	pascal second (Pa·s)	1.488 164 E+00
lb/ft^2	kilogram per meter^2 (kg/m^2)	4.882 428 E+00
lb/ft^3	kilogram per meter^3 (kg/m^3)	1.601 846 E+01
lb/gal (U.K. liquid)	kilogram per meter^3 (kg/m^3)	9.977 633 E+01
lb/gal (U.S. liquid)	kilogram per meter^3 (kg/m^3)	1.198 264 E+02
lb/h	kilogram per second (kg/s)	1.259 979 E–04
lb/hp·h (SFC, specific fuel consumption)	kilogram per joule (kg/J)	1.689 659 E–07
lb/in.^{3}	kilogram per meter^3 (kg/m^3)	2.767 990 E+04
lb/min	kilogram per second (kg/s)	7.559 873 E–03
lb/s	kilogram per second (kg/s)	4.535 924 E–01
lb/yd^3	kilogram per meter^3 (kg/m^3)	5.932 764 E–01
poundal	newton (N)	1.382 550 E–01
poundal/ft^2	pascal (Pa)	1.488 164 E+00
poundal·s/ft^2	pascal second (Pa·s)	1.488 164 E+00

[g] The exact conversion factor is 4.535 923 7*E–01

TABLE A.1b Continued

Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	*To*	*Multiply by*
pound-force (inf)[h]	newton (N)	4.448 222 E+00
lbf/ft	newton meter (N·m)	1.355 818 E+00
lbf·s/ft^2	pascal second (Pa·s)	4.788 026 E+01
lbf·s/in^2	pascal second (Pa·s)	6.894 757 E+03
lbf/ft	newton per meter (N/m)	1.459 390 E+01
lbf/ft^2	pascal (Pa)	4.788 026 E+01
lbf/in.	newton per meter (N/m)	1.751 268 E+02
lbf/in.^{2} (psi)	pascal (Pa)	6.894 757 E+03
lbf/lb (thrust/weight [mass] ratio)	newton per kilogram (N/kg)	9.806 650 E+00
	Q	
quart (U.S. dry)	meter^3 (m^3)	1.101 221 E–03
quart (U.S. liquid)	meter^3 (m^3)	9.463 529 E–04
	R	
rad (radiation dose absorbed)	gray (Gy)	1.000 000*E–02
rhe	1 per pascal second (1/Pa·s)	1.000 000*E+01
rod	meter (m)	[see footnote a]
roentgen	coulomb per kilogram (C/kg)	2.58 E–04
	S	
second (angle)	radian (rad)	4.848 137 E–06
second (sidereal)	second (s)	9.972 696 E–01
section	meter^2 (m^2)	[see footnote a]
shake	second (s)	1.000 000*E–08
slug	kilogram (kg)	1.459 390 E+01
slug/ft·s	pascal second (Pa·s)	4.788 026 E+01
slug/ft^3	kilogram per meter^3 (kg/m^3)	5.155 788 E+02
statampere	ampere (A)	3.335 640 E–10
statcoulomb	coulomb (C)	3.335 640 E–10
statfarad	farad (F)	1.112 650 E–12
stathenry	henry (H)	8.987 554 E+11
statmho	siemens (S)	1.112 650 E–12
statohm	ohm (Ω)	8.987 665 E+11
statvolt	volt (V)	2.997 925 E+02
stere	meter^3 (m^3)	1.000 000*E–04
stilb	candela per meter^2 (cd/m^2)	1.000 000*E+04
stokes (kinematic viscosity)	meter^2 per second (m^2/s)	1.000 000*E–04
	T	
tablespoon	meter^3 (m^3)	1.478 676 E–05
teaspoon	meter^3 (m^3)	4.928 922 E–06
tex	kilogram per meter (kg/m)	1.000 000*E–06
therm	joule (J)	1.055 056 E+08
ton (assay)	kilogram (kg)	2.916 667 E–02
ton (long, 2240 lb)	kilogram (kg)	1.016 047 E+03
ton (metric)	kilogram (kg)	1.000 000*E+03
ton (nuclear equivalent of TNT)	joule (J)	4.184 E+09[i]
ton (refrigeration)	watt (W)	3.516 800 E+03
ton (register)	meter^3 (m^3)	2.831 685 E+00
ton (short, 2000 lb)	kilogram (kg)	9.071 847 E+02
ton (long)/yd^3	kilogram per meter^3 (kg/m^3)	1.328 939 E+03
ton (short)/yd^3	kilogram per meter^3 (kg/m^3)	1.186 553 E+03

[h] The exact conversion factor is 4.448 221 615 260 5*E+00.

[i] Defined (not measured) value.

Table A.1b Continued

Alphabetical List of Units (Symbols of SI units given in parentheses)

To Convert from	*To*	*Multiply by*
ton (short)/h	kilogram per second (kg/s)	2.519 958 E–01
ton-force (2000 lbf)	newton (N)	8.896 444 E+03
tonne	kilogram (kg)	1.000 000*E+03
torr (mm Hg, 0°C)	pascal (Pa)	1.333 22 E+02
township	meter^2 (m^2)	[see footnote a]
	U	
unit pole	weber (Wb)	1.256 637 E–07
	W	
W·h	joule (J)	3.600 000*E+03
W·s	joule (J)	1.000 000*E+00
W/cm^2	watt per meter^2 (W/m^2)	1.000 000*E+04
W/in^2	watt per meter^2 (W/m^2)	1.550 003 E+03
	Y	
yard	meter (m)	9.144 000*E–01
yd^2	meter^2 (m^2)	8.361 274 E–01
yd^3	meter^3 (m^3)	7.645 549 E–01
yd^3/min	meter^3 per second (m^3/s)	1.274 258 E–02
year (365 days)	second (s)	3.153 600 E+07
year (sidereal)	second (s)	3.155 815 E+07
year (tropical)	second (s)	3.155 693 E+07

A.2 Engineering Conversion Factors

B. G. LIPTÁK

TABLE A.2a
Conversion Factors

To Convert	*Into*	*Multiply by*
	A	
abcoulomb	statcoulombs	2.998×10^{10}
acre	sq. chain (Gunters)	10
acre	rods	160
acre	sq. links (Gunters)	1×10^{5}
acre	hectare or sq. hectometer	.4047
acres	sq. ft	43,560.0
acres	sq. meters	4,047.
acres	sq. miles	1.562×10^{-3}
acres	sq. yards	4,840.
acre-feet	cu. ft	43,560.0
acre-feet	gallons	3.259×10^{5}
amperes/sq. cm	amps/sq. in.	6.452
amperes/sq. cm	amps/sq. meter	10^{4}
amperes/sq. in.	amps/sq. cm	0.1550
amperes/sq. in.	amps/sq. meter	1,550.0
amperes/sq. meter	amps/sq. cm	10^{-4}
amperes/sq. meter	amps/sq. in.	6.452×10^{-4}
ampere-hours	coulombs	3,600.0
ampere-hours	faradays	0.03731
ampere-turns	gilberts	1.257
ampere-turns/cm	amp-turns/in.	2.540
ampere-turns/cm	amp-turns/meter	100.0
ampere-turns/cm	gilberts/cm	1.257
ampere-turns/in.	amp-turns/cm	0.3937
ampere-turns/in.	amp-turns/meter	39.37
ampere-turns/in.	gilberts/cm	0.4950
ampere-turns/meter	amp/turns/cm	0.01
ampere-turns/meter	amp-turns/in.	0.0254
ampere-turns/meter	gilberts/cm	0.01257
Ångstrom unit	in.	$3,937 \times 10^{-9}$
Ångstrom unit	meter	1×10^{-10}
Ångstrom unit	micron or μm	1×10^{-4}
are	acre (U.S.)	.02471
ares	sq. yards	119.60
ares	acres	0.02471
ares	sq. meters	100.0
Astronomical Unit	kilometers	1.495×10^{8}
atmospheres	ton/sq. in.	.007348
atmospheres	cms of mercury	76.0
atmospheres	ft of water (at 4°C)	33.90
atmospheres	in. of mercury (at 0°C)	29.92
atmospheres	kgs/sq. cm	1.0333
atmospheres	kgs/sq. meter	10,332.
atmospheres	pounds/sq. in.	14.70
atmospheres	tons/sq. ft	1.058
	B	
barrels (U.S., dry)	cu. in.	7056.
barrels (U.S., dry)	quarts (dry)	105.0
barrels (U.S., liquid)	gallons	31.5
barrels (oil)	gallons (oil)	42.0
bars	atmospheres	0.9869
bars	dynes/sq. cm	10^{6}
bars	kgs/sq. meter	1.020×10^{4}
bars	pounds/sq. ft	2,089.
bars	pounds/sq. in.	14.50
baryl	dyne/sq. cm	1.000
bolt (U.S. cloth)	meters	36.576
Btu	liter-atmosphere	10.409
Btu	ergs	1.0550×10^{10}
Btu	foot-lbs	778.3
Btu	gram-calories	252.0
Btu	horsepower-hrs	3.931×10^{-4}
Btu	joules	1,054.8
Btu	kilogram-calories	0.2520
Btu	kilogram-meters	107.5
Btu	kilowatt-hrs	2.928×10^{-4}
Btu/hr	foot-pounds/s	0.2162

TABLE A.2a Continued
Conversion Factors

To Convert	*Into*	*Multiply by*
Btu/hr	gram-cal/s	0.0700
Btu/hr	horsepower-hrs	3.929×10^{-4}
Btu/hr	watts	0.2931
Btu/min	foot-1bs/s	12.96
Btu/min	horsepower	0.02356
Btu/min	kilowatts	0.01757
Btu/min	watts	17.57
Btu/sq. ft/min	watts/sq. in.	0.1221
bucket (U.K. dry)	cu. cm	1.818×10^{4}
bushels	cu. ft	1.2445
bushels	cu. in.	2,150.4
bushels	cu. meters	0.03524
bushels	liters	35.24
bushels	pecks	4.0
bushels	pints (dry)	64.0
bushels	quarts (dry)	32.0
	C	
calories, gram (mean)	Btu (mean)	3.9685×10^{-3}
candle/sq. cm	lamberts	3.142
candle/sq. in.	lamberts	.4870
centares (centiares)	sq. meters	1.0
Centigrade	Fahrenheit	$(\mathrm{C}^\circ \times 9/5) + 32$
centigrams	grams	0.01
centiliter	ounce fluid (U.S.)	.3382
centiliter	cu. in.	.6103
centiliter	drams	2.705
centiliters	liters	0.01
centimeters	feet	3.281×10^{-2}
centimeters	inches	0.3937
centimeters	kilometers	10^{-5}
centimeters	meters	0.01
centimeters	miles	6.214×10^{-6}
centimeters	millimeters	10.0
centimeters	mils	393.7
centimeters	yards	1.094×10^{-2}
centimeter-dynes	cm-grams	1.020×10^{-3}
centimeter-dynes	meter-kgs	1.020×10^{-8}
centimeter-dynes	pound-feet	7.376×10^{-8}
centimeter-grams	cm-dynes	980.7
centimeter-grams	meter-kgs	10^{-5}
centimeter-grams	pound-feet	7.233×10^{-5}
centimeters of mercury	atmospheres	0.01316
centimeters of mercury	feet of water	0.4461
centimeters of mercury	kgs/sq. meter	136.0
centimeters of mercury	pounds/sq. ft	27.85
centimeters of mercury	pounds/sq. in.	0.1934
centimeters/s	feet/min	1.1969
centimeters/s	feet/s	0.03281
centimeters/s	kilometers/hr	0.036
centimeters/s	knots	0.1943
centimeters/s	meters/min	0.6
centimeters/s	miles/hr	0.02237
centimeters/s	miles/min	3.728×10^{-4}
centimeters/s/s	feet/s/s	0.03281
centimeters/s/s	kms/hr/s	0.036
centimeters/s/s	meters/s/s	0.01
centimeters/s/s	miles/hr/s	0.02237
chain	inches	792.00
chain	meters	20.12
chains (surveyors' or Gunter's)	yards	22.00
circular mils	sq. cms	5.067×10^{-6}
circular mils	sq. mils	0.7854
circumference	radians	6.283
circular mils	sq. inches	7.854×10^{-7}
cords	cord feet	8
cord feet	cu. ft	16
coulomb	statcoulombs	2.998×10^{9}
coulombs	faradays	1.036×10^{-5}
coulombs/sq. cm	coulombs/sq. in.	64.52
coulombs/sq. cm	coulombs/sq. meter	10^{4}
coulombs/sq. in.	coulombs/sq. cm	0.1550
coulombs/sq. in.	coulombs/sq. meter	1,550
coulombs/sq. meter	coulombs/sq. cm	10^{-4}
coulombs/sq. meter	coulombs/sq. in.	6.452×10^{-4}
cubic centimeters	cu. ft	3.531×10^{-5}
cubic centimeters	cu. in.	0.06102
cubic centimeters	cu. meters	10^{-6}
cubic centimeters	cu. yards	1.308×10^{-6}
cubic centimeters	gallons (U.S. liq.)	2.642×10^{-4}
cubic centimeters	liters	0.001
cubic centimeters	pints (U.S. liq.)	2.113×10^{-3}
cubic centimeters	quarts (U.S. liq.)	1.057×10^{-3}
cubic feet	bushels (dry)	0.8036
cubic feet	cu. cms	28,320.0
cubic feet	cu. in.	1,728.0

TABLE A.2a Continued
Conversion Factors

To Convert	*Into*	*Multiply by*
cubic feet	cu. meters	0.02832
cubic feet	cu. yards	0.03704
cubic feet	gallons (U.S. liq.)	7.48052
cubic feet	liters	28.32
cubic feet	pints (U.S. liq.)	59.84
cubic feet	quarts (U.S. liq.)	29.92
cubic feet/min	cu. cms/s	472.0
cubic feet/min	gallons/s	0.1247
cubic feet/min	liters/s	0.4720
cubic feet/min	pounds of water/min	62.43
cubic feet/s	million gals/day	0.646317
cubic feet/s	gallons/min	448.831
cubic inches	cu. cms	16.39
cubic inches	cu. feet	5.787×10^{-4}
cubic inches	cu. meters	1.639×10^{-5}
cubic inches	cu. yards	2.143×10^{-5}
cubic inches	gallons	4.329×10^{-3}
cubic inches	liters	0.01639
cubic inches	mil-feet	1.061×10^{5}
cubic inches	pints (U.S. liq.)	0.03463
cubic inches	quarts (U.S. liq.)	0.01732
cubic meters	bushels (dry)	28.38
cubic meters	cu. cms	10^{6}
cubic meters	cu. ft	35.31
cubic meters	cu. in.	61,023.0
cubic meters	cu. yards	1.308
cubic meters	gallons (U.S. liq.)	264.2
cubic meters	liters	1,000.0
cubic meters	pints (U.S. liq.)	2,113.0
cubic meters	quarts (U.S. liq.)	1,057.
cubic yards	cu. cms	7.646×10^{5}
cubic yards	cu. ft	27.0
cubic yards	cu. in.	46,656.0
cubic yards	cu. meters	0.7646
cubic yards	gallons (U.S. liq.)	202.0
cubic yards	liters	764.6
cubic yards	pints (U.S. liq.)	1,615.9
cubic yards	quarts (U.S. liq.)	807.9
cubic yards/min	cu. ft/s	0.45
cubic yards/min	gallons/s	3.367
cubic yards/min	liters/s	12.74

D

To Convert	*Into*	*Multiply by*
dalton	gram	1.650×10^{-24}
days	seconds	86,400.0
decigrams	grams	0.1
deciliters	liters	0.1
decimeters	meters	0.1
degrees (angle)	quadrants	0.01111
degrees (angle)	radians	0.01745
degrees (angle)	seconds	3,600.0
degrees/s	radians/s	0.01745
degrees/s	revolutions/min	0.1667
degrees/s	revolutions/s	2.778×10^{-3}
dekagrams	grams	10.0
dekaliters	liters	10.0
dekameters	meters	10.0
drams (apothecaries or troy)	ounces (avoidupois)	0.1371429
drams (apothecaries or troy)	ounces (troy)	0.125
drams (U.S., fluid or apothecaries)	cu. cm	3.697
drams	grams	1.7718
drams	grains	27.3437
drams	ounces	0.0625
dyne/cm	erg/sq. millimeter	.01
dyne/sq. cm	atmospheres	9.869×10^{-7}
dyne/sq. cm	inch of mercury at 0°C	2.953×10^{-5}
dyne/sq. cm	inch of water at 4°C	4.015×10^{-4}
dynes	grams	1.020×10^{-3}
dynes	joules/cm	10^{-7}
dynes	joules/meter (newtons)	10^{-5}
dynes	kilograms	1.020×10^{-6}
dynes	poundals	7.233×10^{-5}
dynes	pounds	2.248×10^{-6}
dynes/sq. cm	bars	10^{-6}

E

To Convert	*Into*	*Multiply by*
ell	cm	114.30
ell	in.	45
em, pica	in.	.167
em, pica	cm	.4233
erg/s	dyne-cm/s	1.000
ergs	Btu	9.480×10^{-11}
ergs	dyne-centimeters	1.0
ergs	foot-pounds	7.367×10^{-8}
ergs	gram-calories	0.2389×10^{-7}
ergs	gram-cms	1.020×10^{-3}
ergs	horsepower-hrs	3.7250×10^{-14}

TABLE A.2a Continued
Conversion Factors

To Convert	Into	Multiply by
ergs	joules	10^{-7}
ergs	kg-calories	2.389×10^{-11}
ergs	kg-meters	1.020×10^{-8}
ergs	kilowatt-hrs	0.2778×10^{-13}
ergs	watt-hours	0.2778×10^{-10}
ergs/s	Btu/min	$5{,}688 \times 10^{-9}$
ergs/s	ft-lbs/min	4.427×10^{-6}
ergs/s	ft-lbs/s	7.3756×10^{-8}
ergs/s	horsepower	1.341×10^{-10}
ergs/s	kg-calories/min	1.433×10^{-9}
ergs/s	kilowatts	10^{-10}
	F	
farads	microfarads	10^{6}
faraday/s	ampere (absolute)	9.6500×10^{4}
faradays	ampere-hours	26.80
faradays	coulombs	9.649×10^{4}
fathom	meter	1.828804
fathoms	feet	6.0
feet	centimeters	30.48
feet	kilometers	3.048×10^{-4}
feet	meters	0.3048
feet	miles (naut.)	1.645×10^{-4}
feet	miles (stat.)	1.894×10^{-4}
feet	millimeters	304.8
feet	mils	1.2×10^{4}
feet of water	atmospheres	0.02950
feet of water	in. of mercury	0.8826
feet of water	kgs/sq. cm	0.03048
feet of water	kgs/sq. meter	304.8
feet of water	pounds/sq. ft	62.43
feet of water	pounds/sq. in.	0.4335
feet/min	cms/s	0.5080
feet/min	ft/s	0.01667
feet/min	kms/hr	0.01829
feet/min	meters/min	0.3048
feet/min	miles/hr	0.01136
feet/s	cms/s	30.48
feet/s	kms/hr	1.097
feet/s	knots	0.5921
feet/s	meters/min	18.29
feet/s	miles/hr	0.6818
feet/s	miles/min	0.01136
feet/s/s	cms/s/s	30.48
feet/s/s	kms/hr/s	1.097
feet/s/s	meters/s/s	0.3048
feet/s/s	miles/hr/s	0.6818
feet/100 feet	percent grade	1.0
foot-candle	lumen/sq. meter	10.764
foot-pounds	Btu	1.286×10^{-3}
foot-pounds	ergs	1.356×10^{7}
foot-pounds	gram-calories	0.3238
foot-pounds	hp-hrs	5.050×10^{-7}
foot-pounds	joules	1.356
foot-pounds	kg-calories	3.24×10^{-4}
foot-pounds	kg-meters	0.1383
foot-pounds	kilowatt-hrs	3.766×10^{-7}
foot-pounds/min	Btu/min	1.286×10^{-3}
foot-pounds/min	foot-pounds/s	0.01667
foot-pounds/min	horsepower	3.030×10^{-5}
foot-pounds/min	kg-calories/min	3.24×10^{-4}
foot-pounds/min	kilowatt	2.260×10^{-5}
foot-pounds/s	Btu/hr	4.6263
foot/pounds/s	Btu/min	0.07717
foot-pounds/s	horsepower	1.818×10^{-3}
foot-pounds/s	kg-calories/min	0.01945
foot-pounds/s	kilowatts	1.356×10^{-3}
furlongs	miles (U.S.)	0.125
furlongs	rods	40.0
furlongs	feet	660.0
	G	
gallons	cu. cms	3,785.0
gallons	cu. ft	0.1337
gallons	cu. in.	231.0
gallons	cu. meters	3.785×10^{-3}
gallons	cu. yards	4.951×10^{-3}
gallons	liters	3.785
gallons (liq. Br. Imp.)	gallons (U.S. liq.)	1.20095
gallons (U.S.)	gallons (Imp.)	0.83267
gallons of water	pounds of water	8.3453
gallons/min	cu. ft/s	2.228×10^{-3}
gallons/min	liters/s	0.06308
gallons/min	cu. ft/hr	8.0208
gausses	lines/sq. in.	6.452
gausses	webers/sq. cm.	10^{-8}
gausses	webers/sq. in.	6.452×10^{-8}
gausses	webers/sq. meter	10^{-4}
gilberts	ampere-turns	0.7958
gilberts/cm	amp-turns/cm	0.7958

TABLE A.2a Continued
Conversion Factors

To Convert	Into	Multiply by
gilberts/cm	amp-turns/in.	2.021
gilberts/cm	amp-turns/meter	79.58
gills (British)	cu. cm	142.07
gills	liters	0.1183
gills	pints (liq.)	0.25
grade	radian	.01571
grains	drams (avoirdupois)	0.03657143
grains (troy)	grains (avdp)	1.0
grains (troy)	grams	0.06480
grains (troy)	ounces (avdp)	2.0833×10^{-3}
grains (troy)	pennyweight (troy)	0.04167
grains/U.S. gal	parts/million	17.118
grains/U.S. gal	pounds/million gal	142.86
grains/Imp. gal	parts/million	14.286
grams	dynes	980.7
grams	grains	15.43
grams	joules/cm	9.807×10^{-5}
grams	joules/meter (newtons)	9.807×10^{-3}
grams	kilograms	0.001
grams	milligrams	1,000.
grams	ounces (advp)	0.03527
grams	ounces (troy)	0.03215
grams	poundals	0.07093
grams	pounds	2.205×10^{-3}
grams/cm	pounds/inch	5.600×10^{-3}
grams/cu. cm	pounds/ cu. ft	62.43
grams/cu. cm	pounds/cu. in.	0.03613
grams/cu. cm	pounds/mil-foot	3.405×10^{-7}
grams/liter	grains/gal	58.417
grams/liter	pounds/1,000 gal	8.345
grams/liter	pounds/cu. ft	0.062427
grams/liter	parts/million	1,000.0
grams/sq. cm	pounds/sq. ft	2.0481
gram-calories	Btu	3.9683×10^{-3}
gram-calories	ergs	4.1868×10^{7}
gram-calories	foot-pounds	3.0880
gram-calories	horsepower-hrs	1.5596×10^{-6}
gram-calories	kilowatt-hrs	1.1630×10^{-6}
gram-calories	watt-hrs	1.1630×10^{-3}
gram-calories/s	Btu/hr	14.286
gram-centimeters	Btu	9.297×10^{-8}
gram-centimeters	ergs	980.7
gram-centimeters	joules	9.807×10^{-5}
gram-centimeters	kg-cal	2.343×10^{-8}
gram-centimeters	kg-meters	10^{-5}
	H	
hand	cm	10.16
hectares	acres	2.471
hectares	sq. ft	1.076×10^{5}
hectograms	grams	100.0
hectoliters	liters	100.0
hectometers	meters	100.0
hectowatts	watts	100.0
henries	millihenries	1,000.0
hogsheads (U.K.)	cu. ft	10.114
hogsheads (U.S.)	cu. ft	8.42184
hogsheads (U.S.)	gallons (U.S.)	63
horsepower	Btu/min	42.44
horsepower	ft-lbs/min	33,000.
horsepower	ft-lbs/s	550.0
horsepower (metric) (542.5 ft lb/s)	horsepower (550 ft-lb/s)	0.9863
horsepower (550 ft lb/s)	horsepower (metric) (542.5 ft-lb/s)	1.014
horsepower	kg-calories/min	10.68
horsepower	kilowatts	0.7457
horsepower	watts	745.7
horsepower (boiler)	Btu/hr	33.479
horsepower (boiler)	kilowatts	9.803
horsepower-hrs	Btu	2,547.
horsepower-hrs	ergs	2.6845×10^{13}
horsepower-hrs	ft-lbs	1.98×10^{6}
horsepower-hrs	gram-calories	641,190.
horsepower-hrs	joules	2.684×10^{6}
horsepower-hrs	kg-calories	641.1
horsepower-hrs	kg-meters	2.737×10^{5}
horsepower-hrs	kilowatt-hrs	0.7457
hours	days	4.167×10^{-2}
hours	weeks	5.952×10^{-3}
hundredweights (long)	pounds	112
hundredweights (long)	tons (long)	0.05
hundredweights (short)	ounces (avoirdupois)	1,600
hundredweights (short)	pounds	100
hundredweights (short)	tons (metric)	0.0453592
hundredweights (short)	tons (long)	0.0446429

TABLE A.2a Continued
Conversion Factors

To Convert	Into	Multiply by
	I	
inches	centimeters	2.540
inches	meters	2.540×10^{-2}
inches	miles	1.578×10^{-5}
inches	millimeters	25.40
inches	mils	1,000.0
inches	yards	2.778×10^{-2}
inches of mercury	atmospheres	0.03342
inches of mercury	feet of water	1.133
inches of mercury	kgs/sq. cm	0.03453
inches of mercury	kgs/sq. meter	345.3
inches of mercury	pounds/sq. ft	70.73
inches of mercury	pounds/sq. in.	0.4912
inches of water (at 4°C)	atmospheres	2.458×10^{-3}
inches of water (at 4°C)	inches of mercury	0.07355
inches of water (at 4°C)	kgs/sq. cm	2.540×10^{-3}
inches of water (at 4°C)	ounces/sq. in.	0.5781
inches of water (at 4°C)	pounds/sq. ft	5.204
inches of water (at 4°C)	pounds/sq. in.	0.03613
International ampere	ampere (absolute)	.9998
International volt	volts (absolute)	1.0003
	J	
joules	Btu	9.480×10^{-4}
joules	ergs	10^{7}
joules	foot-pounds	0.7376
joules	kg-calories	2.389×10^{-4}
joules	kg-meters	0.1020
joules	watt-hrs	2.778×10^{-4}
joules/cm	grams	1.020×10^{4}
joules/cm	dynes	10^{7}
joules/cm	joules/meter (newtons)	100.0
joules/cm	poundals	723.3
joules/cm	pounds	22.48
	K	
kilograms	dynes	980,665
kilograms	grams	1,000.0
kilograms	joules/cm	0.09807
kilograms	joules/meter (newtons)	9.807
kilograms	poundals	70.93
kilograms	pounds	2.205
kilograms	tons (long)	9.842×10^{-4}
kilograms	tons (short)	1.102×10^{-3}
kilograms/cu. meter	grams/cu. cm	0.001
kilograms/cu. meter	pounds/cu. ft	0.06243
kilograms/cu. meter	pounds/cu. in.	3.613×10^{-5}
kilograms/cu. meter	pounds/mil-foot	3.405×10^{-10}
kilograms/meter	pounds/ft	0.6720
kilograms/sq. cm	dynes	980,665
kilograms/sq. cm	atmospheres	0.9678
kilograms/sq. cm	feet of water	32.81
kilograms/sq. cm	inches of mercury	28.96
kilograms/sq. cm	pounds/sq. ft	2,048
kilograms/sq. cm	pounds/sq. in.	14.22
kilograms/sq. meter	atmospheres	9.678×10^{-5}
kilograms/sq. meter	bars	98.07×10^{-6}
kilograms/sq. meter	feet of water	3.281×10^{-3}
kilograms/sq. meter	inches of mercury	2.896×10^{-3}
kilograms/sq. meter	pounds/sq. ft	0.2048
kilograms/sq. meter	pounds/sq. in.	1.422×10^{-3}
kilograms/sq. mm	kgs/sq. meter	10^{6}
kilogram-calories	Btu	3.968
kilogram-calories	foot-pounds	3,088
kilogram-calories	hp-hrs	1.560×10^{-3}
kilogram-calories	joules	4,186
kilogram-calories	kg-meters	426.9
kilogram-calories	kilojoules	4.186
kilogram-calories	kilowatt-hrs	1.163×10^{-3}
kilogram meters	Btu	9.294×10^{-3}
kilogram meters	ergs	9.804×10^{7}
kilogram meters	foot-pounds	7.233
kilogram meters	joules	9.804
kilogram meters	kg-calories	2.342×10^{-3}
kilogram meters	kilowatt-hrs	2.723×10^{-6}
kilolines	maxwells	1,000.0
kiloliters	liters	1,000.0
kilometers	centimeters	10^{5}
kilometers	feet	3,281
kilometers	inches	3.937×10^{4}
kilometers	meters	1,000.0
kilometers	miles	0.6214
kilometers	millimeters	10^{6}
kilometers	yards	1,094
kilometers/hr	cms/s	27.78
kilometers/hr	ft/min	54.68
kilometers/hr	ft/s	0.9113
kilometers/hr	knots	0.5396
kilometers/hr	meters/min	16.67

TABLE A.2a Continued
Conversion Factors

To Convert	*Into*	*Multiply by*
kilometers/hr	miles/hr	0.6214
kilometers/hr/s	cms/s/s	27.78
kilometers/hr/s	ft/s/s	0.9113
kilometers/hr/s	meters/s/s	0.2778
kilometers/hr/s	miles/hr/s	0.6214
kilowatts	Btu/min	56.92
kilowatts	ft-lbs/min	4.426×10^{4}
kilowatts	ft-lbs/s	737.6
kilowatts	horsepower	1.341
kilowatts	kg-calories/min	14.34
kilowatts	watts	1,000.0
kilowatt-hrs	Btu	3,413
kilowatt-hrs	ergs	3.600×10^{13}
kilowatt-hrs	ft-lbs	2.655×10^{6}
kilowatt-hrs	gram-calories	859,850
kilowatt-hrs	horsepower-hrs	1,341
kilowatt-hrs	joules	3.6×10^{6}
kilowatt-hrs	kg-calories	860.5
kilowatt-hrs	kg-meters	3.671×10^{5}
kilowatt-hrs	pounds of water evaporated from and at 212°F	3.53
kilowatt-hrs	pounds of water raised from 62° to 212°F	22.75
knots	ft/hr	6,080
knots	kilometers/hr	1.8532
knots	nautical miles/hr	1.0
knots	statute miles/hr	1.151
knots	yards/hr	2,027
knots	ft/s	1.689
	L	
league	miles (approx.)	3.0
light year	miles	5.9×10^{12}
light year	kilometers	9.46091×10^{12}
lines/sq. cm	gausses	1.0
lines/sq. in.	gausses	0.1550
lines/sq. in.	webers/sq. cm	1.550×10^{-9}
lines/sq. in.	webers/sq. in.	10^{-8}
lines/sq. in.	webers/sq. meter	1.550×10^{-5}
links (engineer's)	inches	12.0
links (surveyor's)	inches	7.92
liters	bushels (U.S. dry)	0.02833
liters	cu. cm	1,000.0
liters	cu. ft	0.03531
liters	cu. in.	61.02
liters	cu. meters	0.001
liters	cu. yards	1.308×10^{-3}
liters	gallons (U.S. liq.)	0.2642
liters	pints (U.S. liq.)	2.113
liters	quarts (U.S. liq.)	1.057
liters/min	cu. ft/s	5.886×10^{-4}
liters/min	gals/s	4.403×10^{-3}
lumens/sq. ft	foot-candles	1.0
lumen	spherical candle power	.07958
lumen	watt	.001496
lumen/sq. ft	lumen/sq. meter	10.76
lux	foot-candles	0.0929
	M	
maxwells	kilolines	0.001
maxwells	webers	10^{-8}
megalines	maxwells	10^{6}
megohms	microhms	10^{12}
megohms	ohms	10^{6}
meters	centimeters	100.0
meters	feet	3.281
meters	inches	39.37
meters	kilometers	0.001
meters	miles (naut.)	5.396×10^{-4}
meters	miles (stat.)	6.214×10^{-4}
meters	millimeters	1,000.0
meters	yards	1.094
meters	varas	1.179
meters/min	cms/s	1.667
meters/min	ft/min	3.281
meters/min	ft/s	0.05468
meters/min	kms/hr	0.06
meters/min	knots	0.03238
meters/min	miles/hr	0.03728
meters/s	ft/min	196.8
meters/s	ft/s	3.281
meters/s	kilometers/hr	3.6
meters/s	kilometers/min	0.06
meters/s	miles/hr	2.237
meters/s	miles/min	0.03728
meters/s/s	cms/s/s	100.0
meters/s/s	ft/s/s	3.281

TABLE A.2a Continued
Conversion Factors

To Convert	*Into*	*Multiply by*	*To Convert*	*Into*	*Multiply by*
meters/s/s	kms/hr/s	3.6	milliliters	liters	0.001
meters/s/s	miles/hr/s	2.237	millimeters	centimeters	0.1
meter-kilograms	cm-dynes	9.807×10^{7}	millimeters	feet	3.281×10^{-3}
meter-kilograms	cm-grams	10^{5}	millimeters	inches	0.03937
meter-kilograms	pound-feet	7.233	millimeters	kilometers	10^{-6}
microfarad	farads	10^{-6}	millimeters	meters	0.001
micrograms	grams	10^{-6}	millimeters	miles	6.214×10^{-7}
microhms	megohms	10^{-12}	millimeters	mils	39.37
microhms	ohms	10^{-6}	millimeters	yards	1.094×10^{-3}
microliters	liters	10^{-6}	million gals/day	cu. ft/s	1.54723
microns	meters	1×10^{-6}	mils	centimeters	2.540×10^{-3}
miles (naut.)	miles (statute)	1.1516	mils	feet	8.333×10^{-5}
miles (naut.)	yards	2,027	mils	inches	0.001
miles (statute)	centimeters	1.609×10^{5}	mils	kilometers	2.540×10^{-8}
miles (statute)	feet	5,280.	mils	yards	2.778×10^{-5}
miles (statute)	inches	6.336×10^{4}	miner's inches	cu. ft/min	1.5
miles (statute)	kilometers	1.609	minims (U.K.)	cu. cm	0.059192
miles (statute)	meters	1,609	minims (U.S., fluid)	cu. cm	0.061612
miles (statute)	miles (naut.)	0.8684	minutes (angles)	degrees	0.01667
miles (statute)	yards	1,760	miles (naut.)	feet	6,080.27
miles/hr	cms/s	44.70	miles (naut.)	kilometers	1.853
miles/hr	ft/min	88	miles (naut.)	meters	1,853.
miles/hr	ft/s	1.467	minutes (angles)	quadrants	1.852×10^{-4}
miles/hr	kms/hr	1.609	minutes (angles)	radians	2.909×10^{-4}
miles/hr	kms/min	0.02682	minutes (angles)	seconds	60.0
miles/hr	knots	0.8684	myriagrams	kilograms	10.0
miles/hr	meters/min	26.82	myriameters	kilometers	10.0
miles/hr	miles/min	0.1667	myriawatts	kilowatts	10.0
miles/hr/s	cm/s/s	44.70			
miles/hr/s	ft/s/s	1.467		***N***	
miles/hr/s	kms/hr/s	1.609	nepers	decibles	8.686
miles/hr/s	meters/s/s	0.4470	Newton	dynes	1 × 105
miles/min	cms/s	2,682		***O***	
miles/min	ft/s	88	ohm (International)	ohm (absolute)	1.0005
miles/min	kms/min	1.609	ohms	megohms	10^{-6}
miles/min	knots/min	0.8684	ohms	microhms	10^{6}
miles/min	miles/hr	60.0	ounces	drams	16.0
mil-feet	cu. in.	9.425×10^{-6}	ounces	grains	437.5
milliers	kilograms	1,000.	ounces	grams	28.349527
millimicrons	meters	1×10^{-9}	ounces	pounds	0.0625
milligrams	grains	0.01543236	ounces	ounces (troy)	0.9115
milligrams	grams	0.001	ounces	tons (long)	2.790×10^{-5}
milligrams/liter	part/million	1.0	ounces	tons (metric)	2.835×10^{-5}
millihenries	henries	0.001	ounces (fluid)	cu. in.	1.805

TABLE A.2a Continued
Conversion Factors

To Convert	*Into*	*Multiply by*	*To Convert*	*Into*	*Multiply by*
ounces (fluid)	liters	0.02957	pounds	dynes	44.4823×10^{4}
ounces (troy)	grains	480.0	pounds	grains	7,000.
ounces (troy)	grams	31.103481	pounds	grams	453.5924
ounces (troy)	ounces (avdp.)	1.09714	pounds	joules/cm	0.04448
ounces (troy)	pennyweights (troy)	20.0	pounds	joules/meter (newtons)	4.448
ounces (troy)	pounds (troy)	0.08333	pounds	kilograms	0.4536
ounce/sq. in.	dynes/sq. cm	4309	pounds	ounces	16.0
ounces/sq. in.	pounds/sq. in.	0.0625	pounds	ounces (troy)	14.5833
			pounds	poundals	32.17
	P		pounds	pounds (troy)	1.21528
parsec	miles	19×10^{12}	pounds	tons (short)	0.0005
parsec	kilometers	3.084×10^{13}	pounds (troy)	grains	5,760.
parts/million	grains/U.S. gal	0.0584	pounds (troy)	grams	373.24177
parts/million	grains/Imp. gal	0.07016	pounds (troy)	ounces (avdp.)	13.1657
parts/million	pounds/million gal	8.345	pounds (troy)	ounces (troy)	12.0
pecks (U.K.)	cu. in.	554.6	pounds (troy)	pennyweights (troy)	240.0
pecks (U.K.)	liters	9.091901	pounds (troy)	pounds (avdp.)	0.822857
pecks (U.S.)	bushels	0.25	pounds (troy)	tons (long)	3.6735×10^{-4}
pecks (U.S.)	cu. in.	537.605	pounds (troy)	tons (metric)	3.7324×10^{-4}
pecks (U.S.)	liters	8.809582	pounds (troy)	tons (short)	4.1143×10^{-4}
pecks (U.S.)	quarts (dry)	8	pounds of water	cu. ft	0.01602
pennyweights (troy)	grains	24.0	pounds of water	cu. in.	27.68
pennyweights (troy)	ounces (troy)	0.05	pounds of water	gallons	0.1198
pennyweights (troy)	grams	1.55517	pounds of water/min	cu. ft/s	2.670×10^{-4}
pennyweights (troy)	pounds (troy)	4.1667×10^{-3}	pound-feet	cm-dynes	1.356×10^{7}
pints (dry)	cu. in.	33.60	pound-feet	cm-grams	13,825.
pints (liq.)	cu. cms	473.2	pound-feet	meter-kgs	0.1383
pints (liq.)	cu. ft	0.01671	pounds/cu. ft	grams/cu. cm	0.01602
pints (liq.)	cu. in.	28.87	pounds/cu. ft	kgs/cu. meter	16.02
pints (liq.)	cu. meters	4.732×10^{-4}	pounds/cu. ft	pounds/cu. in.	5.787×10^{-4}
pints (liq.)	cu. yards	6.189×10^{-4}	pounds/cu. ft	pounds/mil-foot	5.456×10^{-9}
pints (liq.)	gallons	0.125	pounds/cu. in.	gms/cu. cm	27.68
pints (liq.)	liters	0.4732	pounds/cu. in.	kgs/cu. meter	2.768×10^{4}
pints (liq.)	quarts (liq.)	0.5	pounds/cu. in.	pounds/cu. ft	1,728.
Planck's quantum	erg/s	6.624×10^{-27}	pounds/cu. in.	pounds/mil-foot	9.425×10^{-6}
poise	gram/cm s	1.00	pounds/ft	kgs/meter	1.488
pounds (avoirdupois)	ounces (troy)	14.5833	pounds/in.	gms/cm	178.6
poundals	dynes	13,826	pounds/mil-foot	gms/cu. cm	2.306×10^{6}
poundals	grams	14.10	pounds/sq. ft	atmospheres	4.725×10^{-4}
poundals	joules/cm	1.383×10^{-3}	pounds/sq. ft	feet of water	0.01602
poundals	joules/meter (newtons)	0.1383	pounds/sq. ft	inches of mercury	0.01414
poundals	kilograms	0.01410	pounds/sq. ft	kgs/sq. meter	4.882
poundals	pounds	0.03108	pounds/sq. ft	pounds/sq. in.	6.944×10^{-3}
pounds	drams	256.00	pounds/sq. in	atmospheres	0.06804

TABLE A.2a Continued
Conversion Factors

To Convert	*Into*	*Multiply by*
pounds/sq. in	feet of water	2.307
pounds/sq. in	inches of mercury	2.036
pounds/sq. in	kgs/sq. meter	703.1
pounds/sq. in	pounds/sq. ft	144.0
	Q	
quadrants (angle)	degrees	90.0
quadrants (angle)	minutes	5,400.0
quadrants (angle)	radians	1.571
quadrants (angle)	seconds	3.24×10^{5}
quarts (dry)	cu. in.	67.20
quarts (liq.)	cu. cms	946.4
quarts (liq.)	cu. ft	0.03342
quarts (liq.)	cu. in.	57.75
quarts (liq.)	cu. meters	9.464×10^{-4}
quarts (liq.)	cu. yards	1.238×10^{-3}
quarts (liq.)	gallons	0.25
quarts (liq.)	liters	0.9463
	R	
radians	degrees	57.30
radians	minutes	3,438.
radians	quadrants	0.6366
radians	seconds	2.063×10^{5}
radians/s	degrees/s	57.30
radians/s	revolutions/min	9.549
radians/s	revolutions/s	0.1592
radians/s/s	revs/min/min	573.0
radians/s/s	revs/min/s	9.549
radians/s/s	revs/s/s	0.1592
revolutions	degrees	360.0
revolutions	quadrants	4.0
revolutions	radians	6.283
revolutions/min	degrees/s	6.0
revolutions/min	radians/s	0.1047
revolutions/min	revs/s	0.01667
revolutions/min/min	radians/s/s	1.745×10^{-3}
revolutions/min/min	revs/min/s	0.01667
revolutions/min/min	revs/s/s	2.778×10^{-4}
revolutions/s	degrees/s	360.0
revolutions/s	radians/s	6.283
revolutions/s	revs/min	60.0
revolutions/s/s	radians/s/s	6.283
revolutions/s/s	revs/min/min	3,600.0
revolutions/s/s	revs/min/s	60.0
rod	chain (Gunters)	.25
rod	meters	5.029
rods (surveyors' meas.)	yards	5.5
rods	feet	16.5
	S	
scruples	grains	20
seconds (angle)	degrees	2.778×10^{-4}
seconds (angle)	minutes	0.01667
seconds (angle)	quadrants	3.087×10^{-6}
seconds (angle)	radians	4.848×10^{-6}
slug	kilogram	14.50
slug	pounds	32.17
sphere	steradians	12.57
square centimeters	circular mils	1.973×10^{5}
square centimeters	sq. ft	1.076×10^{-3}
square centimeters	sq. in.	0.1550
square centimeters	sq. meters	0.0001
square centimeters	sq. miles	3.861×10^{-11}
square centimeters	sq. millimeters	100.0
square centimeters	sq. yards	1.196×10^{-4}
square feet	acres	2.296×10^{-5}
square feet	circular mils	1.833×10^{8}
square feet	sq. cms	929.0
square feet	sq. in.	144.0
square feet	sq. meters	0.09290
square feet	sq. miles	3.587×10^{-8}
square feet	sq. millimeters	9.290×10^{4}
square feet	sq. yards	0.1111
square inches	circular mils	1.273×10^{6}
square inches	sq. cms	6.452
square inches	sq. ft	6.944×10^{-4}
square inches	sq. millimeters	645.2
square inches	sq. mils	10^{6}
square inches	sq. yards	7.716×10^{-4}
square kilometers	acres	247.1
square kilometers	sq. cms	10^{10}
square kilometers	sq. ft	10.76×10^{6}
square kilometers	sq. in.	1.550×10^{9}
square kilometers	sq. meters	10^{6}
square kilometers	sq. miles	0.3861
square kilometers	sq. yards	1.196×10^{6}
square meters	acres	2.471×10^{-4}
square meters	sq. cms	10^{4}

TABLE A.2a Continued
Conversion Factors

To Convert	*Into*	*Multiply by*
square meters	sq. cms	10^4
square meters	sq. in.	1,550
square meters	sq. miles	3.861×10^{-7}
square meters	sq. millimeters	10^6
square meters	sq. yards	1.196
square miles	acres	640.0
square miles	sq. ft	27.88×10^6
square miles	sq. kms	2.590
square miles	sq. meters	2.590×10^6
square miles	sq. yards	3.098×10^6
square millimeters	circular mils	1,973
square millimeters	sq. cms	0.01
square millimeters	sq. ft	1.076×10^{-5}
square millimeters	sq. in.	1.550×10^{-3}
square mils	circular mils	1.273
square mils	sq. cms	6.452×10^{-6}
square mils	sq. in.	10^{-6}
square yards	acres	2.066×10^{-4}
square yards	sq. cms	8,361
square yards	sq. ft	9.0
square yards	sq. in.	1,296
square yards	sq. meters	0.8361
square yards	sq. miles	3.228×10^{-7}
square yards	sq. millimeters	8.361×10^5
	T	
temperature (°C) + 273	absolute temperature (K)	1.0
temperature (°C) + 17.78	temperature (°F)	1.8
temperature (°F) + 460	absolute temperature (°R)	1.0
temperature (°F) − 32	temperature (°C)	5/9
tons (long)	kilograms	1,016
tons (long)	pounds	2,240
tons (long)	tons (short)	1.120
tons (metric)	kilograms	1,000
tons (metric)	pounds	2,205
tons (short)	kilograms	907.1848
tons (short)	ounces	32,000
tons (short)	ounces (troy)	29,166.66
tons (short)	pounds	2,000
tons (short)	pounds (troy)	2,430.56
tons (short)	tons (long)	0.89287
tons (short)	tons (metric)	0.9078
tons (short)/sq. ft	kgs/sq. meter	9,765
tons (short)/sq. ft	pounds/sq. in.	2,000
tons of water/24 hrs	pounds of water/hr	83.333
tons of water/24 hrs	gallons/min	0.16643
tons of water/24 hrs	cu. ft/hr	1.3349
	V	
volt/inch	volt/cm	.39370
volt (absolute)	statvolts	.003336
	W	
watts	Btu/hr	3.4129
watts	Btu/min	0.05688
watts	ergs/s	107.
watts	ft-lbs/min	44.27
watts	ft-lbs/s	0.7378
watts	horsepower	1.341×10^{-3}
watts	horsepower (metric)	1.360×10^{-3}
watts	kg-calories/min	0.01433
watts	kilowatts	0.001
watts (abs.)	Btu (mean)/min	0.056884
watts (abs.)	joules/s	1.
watt-hours	Btu	3.413
watt-hours	ergs	3.60×10^{10}
watt-hours	foot-pounds	2,656
watt-hours	gram-calories	859.85
watt-hours	horsepower-hrs	1.341×10^{-3}
watt-hours	kilogram-calories	0.8605
watt-hours	kilogram-meters	367.2
watt-hours	kilowatt-hrs	0.001
watt (International)	watt (absolute)	1.0002
webers	maxwells	10^8
webers	kilolines	10^5
webers/sq. in.	gausses	1.550×10^7
webers/sq. in.	lines/sq. in.	10^8
webers/sq. in.	webers/sq. cm	0.1550
webers/sq. in.	webers/sq. meter	1,550
webers/sq. meter	gausses	10^4
webers/sq. meter	lines/sq. in.	6.452×10^4
webers/sq. meter	webers/sq. cm	10^{-4}
webers/sq. meter	webers/sq. in.	6.452×10^{-4}
	Y	
yards	centimeters	91.44
yards	kilometers	9.144×10^{-4}
yards	meters	0.9144
yards	miles (naut.)	4.934×10^{-4}
yards	miles (stat.)	5.682×10^{-4}
yards	millimeters	914.4

TABLE A.2b
Units of Area

1 cir mil[a]
= .000000785 sq. in.

1 sq. in.
= 1,273,200 cir mils
= .00694 sq. ft
= .000772 sq. yd

1 sq. ft.
= 144 sq. in.
= .01111 sq. yd
= .00002296 acres

1 sq. yd
= 1296 sq. in.
= 9 sq. ft
= .0002066 acres

1 acre
= 43,560 sq. ft
= 4840 sq. yd
= 4096 sq. m

[a] A cir (circular) mil is the area of a circle of 1/1000 in. dia. Thus, a round rod of 1-in. dia. has an area of 1,000,000 cir mils.

TABLE A.2c
Units of Density

1 lb/cu. in.
= 1728 lb/cu. ft
= 0.864 tons[a]/cu. ft
= 23.3 tons/cu. yd
= 231 lb/gal

1 lb/cu. ft
= .000579 lb/cu. in.
= .000500 tons/cu. ft
= .0135 tons/cu. yd
= .1337 lb/gal

1 ton/cu. ft
= 1.157 lb/cu. in.
= 2000 lb/cu. ft
= 27 tons/cu. yd
= 267 lb/gal

1 ton/cu. yd
= .0429 lb/cu. in.
= 74.1 lb/cu. ft
= .0370 tons/cu. ft
= 9.90 lb/gal

1 lb/gal
= .00433 lb/cu. in.
= 7.48 lb/cu. ft
= .00374 tons/cu. ft
= .1010 tons/cu. yd

[a] Tons are short = 2000 lb.

TABLE A.2d
Units of Work, Energy, Heat

1 Btu
= 9340 in. lb
= 778.3 ft-lb
= .0002938 kwhr[a]
= .0003931 hphr

1 in. lb
= .0001070 Btu
= .0833 ft-lb
= .00000003137 kwhr
= .0000000421 hphr

1 ft-lb
= .001284 Btu
= 12 in. lb
= .000000376 kwhr
= .000000505 hphr

1 kwhr
= 3413 Btu
= 31,873,000 in. lb
= 2,656,100 ft-lb
= 1.342 hphr

1 hphr
= 2544 Btu
= 23,760,000 in. lb
= 1,980,000 ft-lb
= 0.7455 kwhr

[a] 1 kilowatthour = 3413 Btu and 1 Btu = 778.3 ft-lb.

TABLE A.2e
Units of Mass Flow

1 lb/s
= 60 lb/min
= 3600 lb/hr
= 86,400 lb/day
= 2,628,000 lb/mo[a]
= 31,536,000 lb/yr

1 lb/min
= .01667 lb/s
= 60 lb/hr
= 1440 lb/day
= 43,800 lb/mo
= 525,600 lb/yr

1 lb/hr
= .0002778 lb/s
= .01667 lb/min
= 24 lb/day
= 730 lb/mo
= 8760 lb/yr

1 lb/day
= .00001157 lb/s
= .000694 lb/min
= .0417 lb/hr
= 30.4 lb/mo
= 365 lb/yr

1 lb/mo
= .000000381 lb/s
= .0000228 lb/min
= .001370 lb/hr
= .0329 lb/day
= 12 lb/yr

1 lb/yr
= .0000000317 lb/s
= .000001903 lb/min
= .0001142 lb/hr
= .002740 lb/day
= .0833 lb/mo

[a] Month used is exactly 1/12 year = 30.4 days.

TABLE A.2f
Units of Volume Flow

1 cu. ft/s
= 60 cu. ft/min
= 3600 cu. ft/hr
= 7.48 gal/s
= 448.8 gal/min
= 26,930 gal/hr

1 cu. ft/min
= .01667 cu. ft/s
= 60 cu. ft/hr
= .1247 gal/s
= 7.48 gal/min
= 448.8 gal/hr

1 cu. ft/hr
= .0002778 cu. ft/s
= .01667 cu. ft/min
= .002078 gal/s
= .1247 gal/min
= 7.48 gal/hr

1 gal/s
= .1337 cu. ft/s
= 8.02 cu. ft/min
= 481 cu. ft/hr
= 60 gal/min
= 3600 gal/hr

1 gal/min
= .002228 cu. ft/s
= .1337 cu. ft/min
= 8.02 cu. ft/hr
= .01667 gal/s
= 60 gal/hr

1 gal/hr
= .0000371 cu. ft/s
= .002228 cu. ft/min
= .1337 cu. ft/hr
= .0002778 gal/s
= .01667 gal/min

TABLE A.2g
Units of Length

1 in.
= .0833 ft
= .0277 yd
= .0000158 miles

1 ft
= 12 in.
= .333 yd
= .000189 miles

1 yd
= 36 in.
= 3 ft
= .000568 miles

1 mile
= 63360 in.
= 5280 ft
= 1760 yd

TABLE A.2h
Linear Conversion

Inches to Millimeters 1 inch = 25.4 millimeters						*Millimeters to Inches 1 millimeter = .03937 inches*							
Inches	*Millimeters*	*Inches*	*Millimeters*	*Inches*	*Millimeters*	*Millimeters*	*Inches*	*Millimeters*	*Inches*	*Millimeters*	*Inches*	*Millimeters*	*Inches*
1/16	1.6	2.0	50.8	4.5	114.3	1.0	.0394	3.5	.1378	6.0	.2362	8.5	.3346
1/8	3.2	2.1	53.3	4.6	116.8	1.1	.0433	3.6	.1417	6.1	.2402	8.6	.3386
3/16	4.8	2.2	55.9	4.7	119.4	1.2	.0472	3.7	.1457	6.2	.2441	8.7	.3425
1/4	6.4	2.3	58.4	4.8	121.9	1.3	.0512	3.8	.1496	6.3	.2480	8.8	.3465
5/16	7.9	2.4	61.0	4.9	124.5	1.4	.0551	3.9	.1535	6.4	.2520	8.9	.3504
3/8	9.5	2.5	63.5	5.0	127.0	1.5	.0591	4.0	.1575	6.5	.2559	9.0	.3543
7/16	11.1	2.6	66.0	5.5	139.7	1.6	.0630	4.1	.1614	6.6	.2598	9.1	.3583
1/2	12.7	2.7	68.6	6.0	152.4	1.7	.0669	4.2	.1654	6.7	.2638	9.2	.3622
9/16	14.3	2.8	71.1	6.5	165.1	1.8	.0709	4.3	.1693	6.8	.2677	9.3	.3661
5/8	15.9	2.9	73.7	7.0	177.8	1.9	.0748	4.4	.1732	6.9	.2717	9.4	.3701
11/16	17.5	3.0	76.2	7.5	190.5	2.0	.0787	4.5	.1772	7.0	.2756	9.5	.3740
3/4	19.1	3.1	78.7	8.0	203.2	2.1	.0827	4.6	.1811	7.1	.2795	9.6	.3780
13/16	20.6	3.2	81.3	8.5	215.9	2.2	.0866	4.7	.1850	7.2	.2835	9.7	.3819
7/8	22.2	3.3	83.8	9.0	228.6	2.3	.0906	4.8	.1890	7.3	.2874	9.8	.3858
15/16	23.8	3.4	86.4	9.5	241.3	2.4	.0945	4.9	.1929	7.4	.2913	9.9	.3898
				10.0	254.0							10.0	.3937
1.0	25.4	3.5	88.9			2.5	.0984	5.0	.1969	7.5	.2953		
1.1	27.9	3.6	91.4			2.6	.1024	5.1	.2008	7.6	.2992		
1.2	30.5	3.7	94.0			2.7	.1063	5.2	.2047	7.7	.3031		
1.3	33.0	3.8	96.5			2.8	.1102	5.3	.2087	7.8	.3071		
1.4	35.6	3.9	99.1			2.9	.1142	5.4	.2126	7.9	.3110		
1.5	38.1	4.0	101.6			3.0	.1181	5.5	.2165	8.0	.3150		
1.6	40.6	4.1	104.1			3.1	.1220	5.6	.2205	8.1	.3189		
1.7	43.2	4.2	106.7			3.2	.1260	5.7	.2244	8.2	.3228		
1.8	45.7	4.3	109.2			3.3	.1299	5.8	.2283	8.3	.3268		
1.9	48.3	4.4	111.8			3.4	.1339	5.9	.2323	8.4	.3307		

By moving the decimal place, conversions for figures larger than 10 may be obtained.

TABLE A.2i
Units of Power

1 kw
= 1.3415 hp
= 738 ft-lb[a]/s
= 44,268 ft-lb/min
= 2,656,100 ft-lb/hr
= .948 Btu/s
= 56.9 Btu/min
= 3413 Btu/hr

1 hp
=.7455 kw
= 550 ft-lb/s
= 33,000 ft-lb/min
= 1,980,000 ft-lb/hr
= .707 Btu/s
= .424 Btu/min
= 2544 Btu/hr

1 ft-lb/s
= .001355 kw
= .001818 hp
= 60 ft-lb/min
= 3600 ft-lb/hr
= .001284 Btu/s
= .0771 Btu/min
= 4.62 Btu/hr

1 ft-lb/min
= .00002259 kw
= .0000303 hp
= .01667 ft-lb/s
= 60 ft-lb hr
= .00002141 Btu/s
= .001284 Btu/min
= .0771 Btu/hr

1 ft-lb/hr
= .000000376 kw
= .000000505 hp
= .000278 ft-lb/s
=. 01667 ft-lb/min
= .000000357 Btu/s
= .00002141 Btu/min
= .001284 Btu/hr

1 Btu/s
= 1.055 kw
= 1.416 hp
= 778 ft-lb/s
= 46,700 ft-lb/min
= 2,802,000 ft-lb/hr
= 60 Btu/min
= 3600 Btu/hr

1 Btu/min
= .01759 kw
= .02359 hp
= 12.98 ft-lb/s
= 778 ft-lb/min
= 46,700 ft-lb/hr
= .01667 Btu/s
= 60 Btu/hr

1 Btu/hr
= .0002931 kw
= .0003932 hp
= .2163 ft-lb/s
= 12.98 ft-lb/min
= 778 ft-lb/hr
= .0002778 Btu/s
= .01667 Btu/min

[a] ft-lb means foot pound, the work done in moving against one pound force a distance of one foot.

Table A.2j
Units of Pressure

1 in. water[a]
= .0833 ft water
= .0735 in. Hg
= .577 oz/sq. in.
= .83.1 oz/sq. ft
= .0361 lb/sq. in.
= 5.20 lb/sq. ft

1 ft water
= 12 in. water
= .882 in. Hg
= 6.93 oz/sq. in.
= 998 oz/sq. ft
= .433 lb/sq. in.
= 62.4 lb/sq. ft

1 in. Hg
= 13.61 in. water
= 1.131 ft water
= 7.84 oz/sq. in.
= 1129 oz/sq. ft
= .491 lb/sq. in.
= 70.5 lb/sq. ft

1 oz/sq. in.
= 1.732 in. water
= .1443 ft water
= .1276 in. Hg
= 144 oz/sq. ft
= .0625 lb/sq. in.
= 9 lb/sq. ft

1 oz/sq. ft
= .01203 in. water
= .001002 ft water
= .000886 in. Hg
= .00694 oz/sq. in.
= .000434 lb/sq. in.
= .0625 lb/sq. ft

1 lb/sq. in.
= 27.71 in. water
= 2.31 ft water
= 2.04 in. Hg
= 16 oz/sq. in.
= 2304 oz/sq. ft
= 144 lb/sq. ft

1 lb/sq. ft
= .1924 in. water
= .01604 ft water
= .01418 in. Hg
= .1111 oz/sq. in.
= 16 oz/sq. ft
= .00694 lb/sq. in.

[a] in. water means inches of water at 60°F.
in. Hg means inches head of mercury at 32°F.

TABLE A.2k
Pressure Conversion

1 pound per square inch = .0703 kilograms per square centimeter

lbs. per sq. in.	*kgs per sq. cm*	*lbs. per sq. in.*	*kgs per sq. cm*	*lbs. per sq. in.*	*kgs per sq. cm*	*lbs. per sq. in.*	*kgs per sq. cm*
1.00	.0703	2.25	.1582	4.50	.3164	8.25	.5800
1.10	.0773	2.30	.1617	4.75	.3339	8.50	.5976
1.20	.0844	2.40	.1687	5.00	.3515	8.75	.6151
1.25	.0879	2.50	.1758	5.25	.3691	9.00	.6327
1.30	.0914	2.60	.1828	5.50	.3867	9.25	.6503
1.40	.0984	2.70	.1898	5.75	.4042	9.50	.6679
1.50	.1055	2.75	.1933	6.00	.4218	9.75	.6854
1.60	.1125	2.80	.1969	6.25	.4394	10.00	.7030
1.70	.1195	2.90	.2039	6.50	.4570		
1.75	.1230	3.00	.2109	6.75	.4746		
1.80	.1265	3.25	.2285	7.00	.4921		
1.90	.1336	3.50	.2461	7.25	.5097		
2.00	.1406	3.75	.2636	7.50	.5273		
2.10	.1476	4.00	.2812	7.75	.5448		
2.20	.1547	4.25	.2988	8.00	.5624		

1 kilogram per square centimeter = 14.223 pounds per square inch

kgs per sq. cm	*lbs. per sq. in.*	*kgs per sq. cm*	*lbs. per sq. in.*	*kgs per sq. cm*	*lbs. per sq. in.*	*kgs per sq. cm*	*lbs. per sq. in.*
1.0	14.22	2.5	35.56	4.0	56.89	7.5	106.67
1.1	15.65	2.6	36.98	4.1	58.31	8.0	113.78
1.2	17.07	2.7	38.40	4.2	59.74	8.5	120.90
1.3	18.49	2.8	39.82	4.3	61.16	9.0	128.01
1.4	19.91	2.9	41.25	4.4	62.58	9.5	135.12
1.5	21.33	3.0	42.67	4.5	64.00	10.0	142.23
1.6	22.76	3.1	44.09	4.6	65.43		
1.7	24.18	3.2	45.51	4.7	66.85		
1.8	25.60	3.3	46.94	4.8	68.27		
1.9	27.02	3.4	48.36	4.9	69.69		
2.0	28.45	3.5	49.78	5.0	71.12		
2.1	29.87	3.6	51.20	5.5	78.23		
2.2	31.29	3.7	52.63	6.0	85.34		
2.3	32.71	3.8	54.05	6.5	92.45		
2.4	34.14	3.9	55.47	7.0	99.56		

TABLE A.2I

Pressure Head Conversion

The center column, marked "Known," may be used either for head in feet or for pressure in pounds per square inch—when used as head, the corresponding pressure is found in the column at the left designated as "Pressure Wanted"; when used as pressure, the corresponding head is found in the column at the right designated as "Head Wanted." For example, a 10-foot head has a pressure of 4.33 pounds per square inch, and a 10-pound pressure has a head of 23.094 feet. By moving the decimal place, quantities larger than 100 may be used.

Pressure Wanted lb. per sq. in.	*Known Pressure or Head*	*Head Wanted ft.* H_2O	*Pressure Wanted lb. per sq. in.*	*Known Pressure or Head*	*Head Wanted ft.* H_2O	*Pressure Wanted lb. per sq. in.*	*Known Pressure or Head*	*Head Wanted ft.* H_2O	*Pressure Wanted lb. per sq. in.*	*Known Pressure or Head*	*Head Wanted ft.* H_2O
0.433	1	2.309	11.259	26	60.044	22.084	51	117.779	32.909	76	175.514
0.866	2	4.619	11.692	27	62.354	22.517	52	120.089	33.342	77	177.824
1.299	3	6.928	12.125	28	64.663	22.950	53	122.398	33.775	78	180.133
1.732	4	9.238	12.558	29	66.973	23.383	54	124.708	34.208	79	182.433
2.165	5	11.547	12.991	30	69.282	23.816	55	127.017	34.642	80	184.752
2.598	6	13.856	13.424	31	71.591	24.249	56	129.326	35.075	81	187.061
3.031	7	16.166	13.857	32	73.901	24.682	57	131.636	35.508	82	189.371
3.464	8	18.475	14.290	33	76.210	25.115	58	133.945	35.941	83	191.680
3.897	9	20.785	14.723	34	78.520	25.548	59	136.255	36.374	84	193.990
4.330	10	23.094	15.156	35	80.829	25.981	60	138.564	36.807	85	196.299
4.763	11	25.403	15.589	36	83.138	26.414	61	140.873	37.240	86	198.608
5.196	12	27.713	16.022	37	85.448	26.847	62	143.183	37.673	87	200.918
5.629	13	30.022	16.455	38	87.757	27.280	63	145.492	38.106	88	203.227
6.062	14	32.332	16.888	39	90.067	27.713	64	147.803	38.539	89	205.537
6.495	15	34.641	17.321	40	92.376	28.146	65	150.111	38.972	90	207.846
6.928	16	36.950	17.754	41	94.685	28.579	66	152.420	39.405	91	210.155
7.361	17	39.260	18.187	42	96.995	29.012	67	154.730	39.838	92	212.465
7.794	18	41.569	18.620	43	99.304	29.445	68	157.039	40.271	93	214.774
8.227	19	43.879	19.053	44	101.614	29.878	69	159.349	40.704	94	217.084
8.660	20	46.188	19.486	45	103.923	30.311	70	161.658	41.137	95	219.393
9.093	21	48.497	19.919	46	106.232	30.744	71	163.967	41.570	96	221.702
9.526	22	50.807	20.352	47	108.542	31.177	72	166.277	42.003	97	224.012
9.959	23	53.116	20.785	48	110.851	31.610	73	168.586	42.436	98	226.321
10.392	24	55.426	21.218	49	113.161	32.043	74	170.896	42.869	99	228.631
10.825	25	57.735	21.651	50	115.470	32.476	75	173.205	43.302	100	230.940

TABLE A.2m
Temperature Conversion

Degrees–Fahrenheit to Centigrade
–°C = 5/9 (°F + 32)
+°C = 5/9 (°F – 32)

°F	*0*	*25*	*50*	*75*
–200	–128.9	–142.8	–156.7	–170.6
–100	–73.3	–87.2	–101.1	–115.0
–0	–17.8	–31.7	–45.6	–59.4
+0	–17.8	–3.9	+10.0	+23.9
100	+37.8	+51.7	65.6	79.4
200	93.3	107.2	121.1	135.0
300	148.9	162.8	176.7	190.6
400	204.4	218.3	232.2	246.1
500	260.0	273.9	287.8	301.7
600	315.6	329.4	343.3	357.2
700	371.1	385.0	398.9	412.8
800	426.7	440.6	454.4	468.3
900	482.2	496.1	510.0	523.9
1000	538.0	551.7	565.6	579.4
1100	593.2	607.2	621.1	635.0

Degrees–Centigrade to Fahrenheit
–°F = (9/5 × °C) – 32
+°F = (9/5 × °C) + 32

°C	*0*	*25*	*50*	*75*
–200	–328	–373	–418	
–100	–148	–193	–238	–283
–0	+32	–13	–58	–103
+0	+32	+77	+122	+167
100	212	257	302	347
200	392	437	482	527
300	572	617	662	707
400	752	797	842	887
500	932	977	1022	1067
600	1112	1157	1202	1247
700	1292	1337	1382	1427

TABLE A.2n
Units of Time

1 s
= .01667 min
= .0002778 hr
= .00001157 days
= .0000003805 mo[a]
= .0000000317 yr

1 min
= 60 s
= .01667 hr
= .000694 days
= .0000228 mo
= .000001903 yr

1 hr
= 3600 s
= 60 min
= .0417 days
= .001370 mo
= .0001142 yr

1 day
= 86,400 s
= 1440 min
= 24 hr
= .0329 mo
= .002740 yr

1 mo
= 2,628,000 s
= 43,800 min
= 730 hr
= 30.4 days
= .0833 yr

1 yr
= 31,536,000 s
= 525,600 min
= 8760 hr
= 365 days
= 12 mo

[a] Month used is exactly 1/12 year.

TABLE A.2o
Units of Velocity

fps
= 60 fpm
= 3600 fph
= .01136 mpm
= .682 mph

1 fpm
= .01667 fps
= 60 fph
= .0001894 mpm
= .01136 mph

1 fph
= .002778 fps
= .01667 fpm
= .00000316 mpm
= .0001894 mph

1 mpm
= 88 fps
= 5280 fpm
= 316,800 fph
= 60 mph

1 mph
= 1.467 fps
= 88 fpm
= 5280 fph
= .01667 mpm

f = feet; h = hours; m = miles or minutes; p = per; s = seconds.

TABLE A.2p
Viscosity Conversion

Kinematic Viscosity Centisokes = K	*Seconds Saybolt Universal*	*Seconds Saybolt Furol*	*Seconds Redwood*	*Seconds Redwood Admiralty*	*Degrees Engler*	*Degrees Bardey*
1.00	31	—	29.1	—	1.00	6200
2.56	35	—	32.1	—	1.16	2420
4.30	40	—	36.2	5.10	1.31	1440
5.90	45	—	40.3	5.52	1.46	1050
7.40	50	—	44.3	5.83	1.58	838
8.83	55	—	48.5	6.35	1.73	702
10.20	60	—	52.3	6.77	1.88	618
11.53	65	—	56.7	7.17	2.03	538
12.83	70	12.95	60.9	7.60	2.17	483
14.10	75	13.33	65.0	8.00	2.31	440
15.35	80	13.70	69.2	8.44	2.45	404
16.58	85	14.10	73.3	8.86	2.59	374
17.80	90	14.44	77.6	9.30	2.73	348
19.00	95	14.85	81.5	9.70	2.88	326
20.20	100	15.24	85.6	10.12	3.02	307
31.80	150	19.3	128	14.48	4.48	195
43.10	200	23.5	170	18.90	5.92	144
54.30	250	28.0	212	23.45	7.35	114
65.40	300	32.5	254	28.0	8.79	95
76.50	350	35.1	296	32.5	10.25	81
87.60	400	41.9	338	37.1	11.70	70.8
98.60	450	46.8	381	41.7	13.15	62.9
110.	500	51.6	423	46.2	14.60	56.4
121.	550	56.6	465	50.8	16.05	51.3
132.	600	61.4	508	55.4	17.50	47.0
143.	650	66.2	550	60.1	19.00	43.4
154.	700	71.1	592	64.6	20.45	40.3
165.	750	76.0	635	69.2	21.90	37.6
176.	800	81.0	677	73.8	23.35	35.2
187.	850	86.0	719	78.4	24.80	33.2
198	900	91.0	762	83.0	26.30	31.3
209	950	95.8	804	87.6	27.70	29.7
220	1000	100.7	846	92.2	29.20	28.2
330	1500	150	1270	138.2	43.80	18.7
440	2000	200	1690	184.2	58.40	14.1

TABLE A.2p Continued

Viscosity Conversion

Kinematic Viscosity Centistokes = K	*Seconds Saybolt Universal*	*Seconds Saybolt Furol*	*Seconds Redwood*	*Seconds Redwood Admiralty*	*Degrees Engler*	*Degrees Bardey*
550	2500	250	2120	230	73.00	11.3
660	3000	300	2540	276	87.60	9.4
770	3500	350	2960	322	100.20	8.05
880	4000	400	3380	368	117.00	7.05
990	4500	450	3810	414	131.50	6.26
1100	5000	500	4230	461	146.00	5.64
1210	5500	550	4650	507	160.50	5.13
1320	6000	600	5080	553	175.00	4.70
1430	6500	650	5500	559	190.00	4.34
1540	7000	700	5920	645	204.50	4.03
1650	7500	750	6350	691	219.00	3.76
1760	8000	800	6770	737	233.50	3.52
1870	8500	850	7190	783	248.00	3.32
1980	9000	900	7620	829	263.00	3.13
2090	9500	950	8040	875	277.00	2.97
2200	10000	1000	8460	921	292.00	2.82

The viscosity is often expressed in terms of viscosimeters other than the Saybolt Universal. The formulas for the various viscosimeters are given opposite.

If viscosity is given at any two temperatures, the viscosity at any other temperature can be obtained by plotting the viscosity against temperature in degrees Fahrenheit on special Log paper. The points for a given oil lie in a straight line.

$$\text{Kinematic viscosity} = \frac{\text{absolute viscosity}}{\text{specific gravity}}$$

$$\text{Redwood } K = 26t - \frac{180}{t}(\text{British})$$

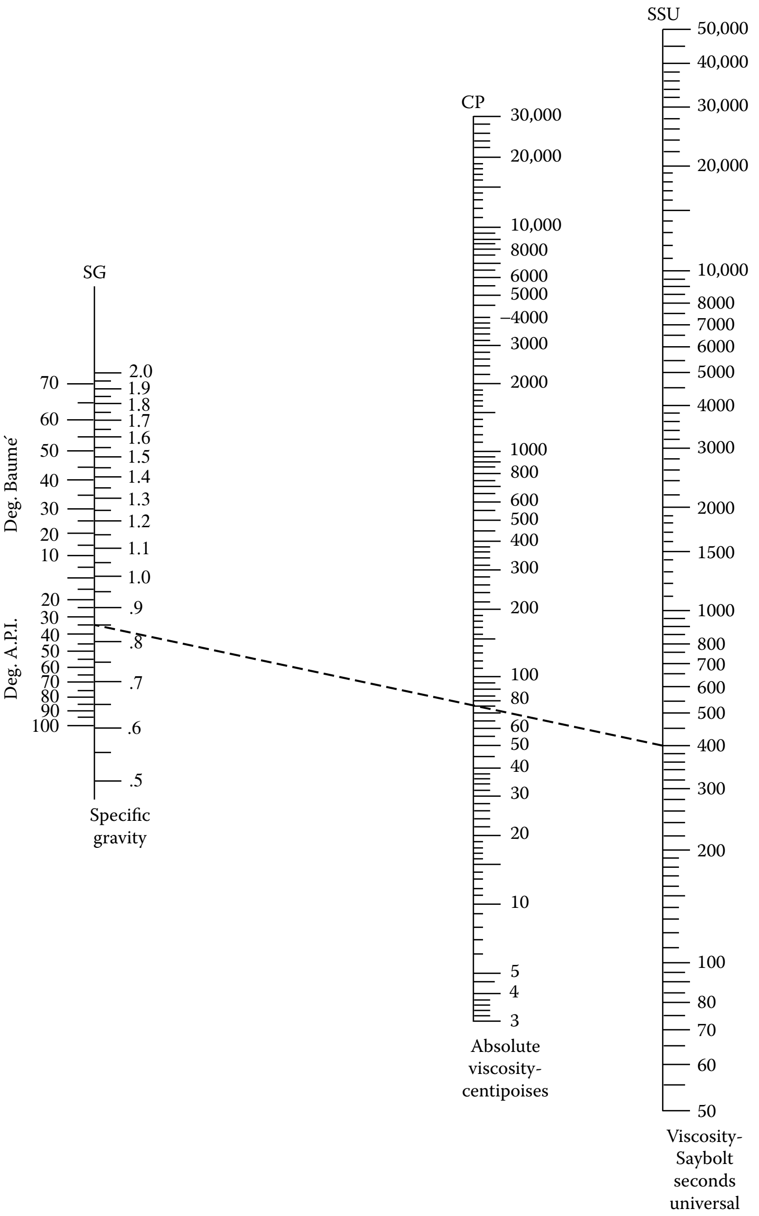
$$\text{Redwood Admiralty } K = 2.7t - \frac{20}{t}(\text{British})$$

$$\text{Saybolt Universal } K = .22t - \frac{195}{t}(\text{American})$$

$$\text{Saybolt Furol } K = 2.2t - \frac{184}{t}(\text{American})$$

$$\text{Engler } K = .147t - \frac{374}{t}(\text{German})$$

TABLE A.2q
Viscosity Conversion Chart



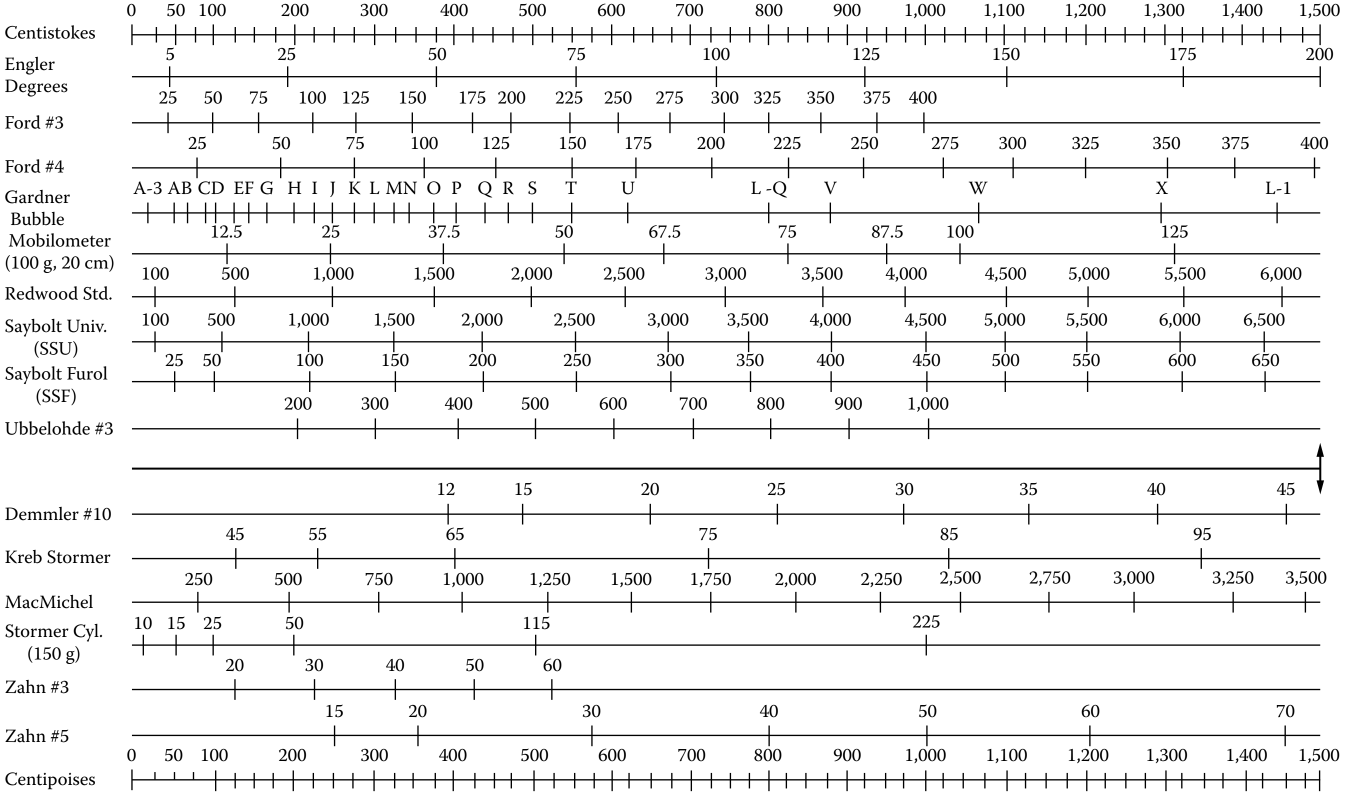
This chart enables the direct conversion of a viscosity in centipoises to a SSU viscosity. As an example, suppose the liquid under consideration has a specific gravity of .85 and a viscosity of 75 centipoises. To determine the viscosity in SSU, lay a straight edge between 75 on the cp scale and .85 on the G scale. The viscosity in SSU can be read directly on the SSU scale. In this instance, the SSU viscosity is 400 (see dotted line).

If the viscosity value is given in centistokes (kinematic viscosity), it can be used directly on the viscosity correction nomograph. The relationship between the absolute viscosity and the kinematic viscosity is expressed by the following formula:

Centipoises = Centistokes × Specific Gravity

TABLE A.2r

Approximate Viscosity Conversion Chart



Note: Scales in lower section read directly in centipoises. Those above read in centistokes (to convert into centipoises, multiply by liquid specific gravity).

TABLE A.2s

Units of Volume

1 cu. in.
= .00433 gal
= .000579 cu. ft
= .0000214 cu. yd

1 gal
= 231 cu. in.
= .1337 cu. ft.
= .00495 cu. yd
= .00000307 acre ft[a]

1 cu. ft
= 1728 cu. in.
= 7.48 gal
= .0370 cu. yd
= .0000230 acre ft

1 cu. yd
= 46,656 cu. in.
= 202.0 gal
= 27 cu. ft.
= .000620 acre ft

1 acre ft
= 325,800 gal
= 43,560 cu. ft
= 1613 cu. yd

[a] Acre ft of water is the volume in 1 ft of depth covering 1 acre.

TABLE A.2t

Units of Weight

1 gr
= .00229 oz[a]
= .0001429 lb
= .0000000714 tons

1 oz
= 438 gr
= .0625 lb
= .00003125 tons

1 lb
= 7000 gr
= 16 oz
= .000500 tons

1 ton
= 14,000,000 gr
= 32,000 oz
= 2000 lb

[a] Avoirdupois oz and lb and short ton of 2000 lb.

TABLE A.2u

Weight Conversion

1 kilogram = 2.2046 pounds 1 pound = 0.4536 kilograms

lbs	*kgs/ lbs*	*kgs*	*lbs*	*kgs/ lbs*	*kgs*	*lbs*	*kgs/ lbs*	*kgs*	*lbs*	*kgs/ lbs*	*kgs*	*lbs*	*kgs/ lbs*	*kgs*
2.205	1.0	.454	6.614	3.0	1.361	11.023	5.0	2.268	15.432	7.0	3.175	19.841	9.0	4.082
2.425	1.1	.499	6.834	3.1	1.406	11.243	5.1	2.313	15.653	7.1	3.221	20.062	9.1	4.128
2.646	1.2	.544	7.055	3.2	1.452	11.464	5.2	2.359	15.873	7.2	3.266	20.282	9.2	4.173
2.866	1.3	.590	7.275	3.3	1.497	11.684	5.3	2.404	16.094	7.3	3.311	20.503	9.3	4.218
3.086	1.4	.635	7.496	3.4	1.542	11.905	5.4	2.449	16.314	7.4	3.357	20.723	9.4	4.264
3.307	1.5	.680	7.716	3.5	1.588	12.125	5.5	2.495	16.535	7.5	3.402	20.944	9.5	4.309
3.527	1.6	.726	7.937	3.6	1.633	12.346	5.6	2.540	16.755	7.6	3.447	21.164	9.6	4.355
3.748	1.7	.771	8.157	3.7	1.678	12.566	5.7	2.586	16.975	7.7	3.493	21.385	9.7	4.400
3.968	1.8	.816	8.377	3.8	1.724	12.787	5.8	2.631	17.196	7.8	3.538	21.605	9.8	4.445
4.189	1.9	.862	8.598	3.9	1.769	13.007	5.9	2.676	17.416	7.9	3.583	21.826	9.9	4.491
												22.046	10.0	4.536
4.409	2.0	.907	8.818	4.0	1.814	13.228	6.0	2.722	17.637	8.0	3.629			
4.630	2.1	.953	9.039	4.1	1.860	13.448	6.1	2.767	17.857	8.1	3.674			
4.850	2.2	.998	9.259	4.2	1.905	13.669	6.2	2.812	18.078	8.2	3.720			
5.071	2.3	1.043	9.480	4.3	1.950	13.889	6.3	2.858	18.298	8.3	3.765			
5.291	2.4	1.089	9.700	4.4	1.996	14.109	6.4	2.903	18.519	8.4	3.810			
5.512	2.5	1.134	9.921	4.5	2.041	14.330	6.5	2.948	18.739	8.5	3.856			
5.732	2.6	1.179	10.141	4.6	2.087	14.550	6.6	2.994	18.960	8.6	3.901			
5.952	2.7	1.225	10.362	4.7	2.132	14.771	6.7	3.039	19.180	8.7	3.946			
6.173	2.8	1.270	10.582	4.8	2.177	14.991	6.8	3.084	19.400	8.8	3.992			
6.393	2.9	1.315	10.803	4.9	2.223	15.212	6.9	3.130	19.621	8.9	4.037			

A.3 Chemical Resistance of Materials

B. G. LIPTÁK

TABLE A.3

Chemical Resistance of Materials

MATERIALS
× — *Very Good Service*
+ — *Moderate Service*
– — *Limited or Variable Service*
O — *Unsatisfactory*
Blank — No Information

CHEMICALS
Solids Assumed in Solution
Room Temperatures Assumed
Unless Otherwise Stated

	METALS														
Chemical	*Carbon Steel; Fe*	*Cast Iron & Ductile Iron; Fe*	*304 Stainless Steel; Fe, 18Cr, 8Ni*	*316 Stainless Steel; Fe 16Cr, 10Ni, 2Mo*	*347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb*	*Ni-Resist Iron; Fe 14Ni, 2Cr, 2Si*	*Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo, 1Si*	*Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si*	*Durorpm; Fe, 14Si; Durichlor, Fe, 14Si, 3Mo**	*Copper; Brass; Bronzes; Everdur*	*Aluminum; Al (and Alloys)*	*Lead; Pb*	*Monel; 67Ni, 30Cu, 1.4Fe*	*Nickel; Ni*	*Inconel; 76Ni, 15Cr, 8Fe*
Acetic Acid, 100% CH_3COOH	O	–	+	×	–	O	×	×	×	O	×	O	–	+	–
Acetic Acid, Dilute	O	O	×	×	–	O	×	×	×	×	–	–	–	O	+
Acetic Anhydride, $(CH_3CO)_2$ O	–	+	×	×	×	–	×	×	×	–	–	O	–	–	+
Acetone, CH_3 $COCH_3$	×		×	×	×	×	×	×	×	×	×	×	×	×	×
Acetyl Chloride, CH_3COCl	O		–	×			×		×	O			+		
Aluminum Chloride, $AlCl_3$	O	O	O	O	O	–	×	O	×	–	O	O	O	–	–
Aluminum Hydroxide, $AL(OH)_3$			×	×			×	×	+	×	×		+	+	
Aluminum Sulfate, Al_2 $(SO_4)_3$	O		–	–	×	–	×	×	×	–	×	×	–	–	–
Alums, Conc., $Al_2(SO_4)_3K_2SO_4$, etc.	O		O	–	–		+		×	–	–	×			
Alums, Dilute	O	O	×	–	–	–	×	×	×	–	×	+	+	+	
Amines, various	×	×	×	×	×		×	×	×	–	–				
Ammonia (Gas), Moist, NH_3	×		×	×	×	×	×		×	O	O	–	–	–	–
Ammonium Carbonate, $(NH_4)_2CO_3$	+	+	+	+	×		×		×	–	–		+	×	+
Ammonium Chloride, NH_4Cl	O	O	O	–	–	×	×	×	×	O	O	–	+	+	–
Ammonium Hydroxide, NH_4OH	×	×	×	×	×	×	×		+	O	–	+	O	O	×
Ammonium Nitrate, NH_4NO_3	–		×	×	×	–	×	×	×	O	O		×		–
Ammonium Persulfate, $(NH_4)_2S_2O_8$	O		×	×	×		×	×	×	–	–		O	O	–
Ammonium Phosphate, $(NH_4)_1H_2PO_4$	O		×	×		–	×		×	O					
Ammonium Phosphate, $(NH_4)_2HPO_4$	–		×	×		×	×		×	–	–				
Ammonium Phosphate, $(NH_4)_3PO_4$	×	–	×	×	×	×	×	×	×	O	–	×	×	×	×
Ammonium Sulfate, $(NH_4)_2SO_4$	×		–	–	+	×	×	×	×	–	–	–	+	+	×
Amyl Acetate, $C_5H_{11}COOCH_3$	–		×	×	×	×	×	×	×	–	×		+	+	+
Amyl Alcohol, $C_5H_{11}OH$	×		×	×			×		×	×	×				

	METALS												
Chemical	*Hastelloy B; Ni, 26Mo, 4Fe*	*Hastelloy C; Ni, 16Mo, 4Fe,14Cr, 4W*	*Hastelloy D; Ni, 8Si, 3Cu*	*Chlorimet 3; 3Fe, 1Si, 60Ni,18Mo, 18Cr*	*Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si*	*Stellite; Co, 28Cr, 4W*	*Zirconium; Z*	*Tantalum; Ta*	*Silver; Ag*	*Platinum; Pt*	*Dowmetal; (Mg alloys)*	*Titanium; Ti*	*Molybdenum; Mo*
Acetic Acid, 100% CH_3COOH	×	×	×	×	×	×	×	×	×	×		×	
Acetic Acid, Dilute	×	×	+	×	×	×	×	×	×	×		×	×
Acetic Anhydride, $(CH_3CO)_2$ O	+	×	+	×	×	×		×	×	×		×	
Acetone, CH_3 $COCH_3$	×	×	×	×	×		×		×			×	
Acetyl Chloride, CH_3COCl				×	×								
Aluminum Chloride, $AlCl_3$	×	+	×	×	×	–	×	×	×	×		×	
Aluminum Hydroxide, $AL(OH)_3$				×	×			×					
Aluminum Sulfate, Al_2 $(SO_4)_3$	×	×	×	×	×	×	×	×	×	×		×	
Alums, Conc., $Al_2(SO_4)_3K_2SO_4$, etc.				×	×			×		×			
Alums, Dilute	+	+		×	×		×	×	×	×			
Amines, various	×	×	×	×	×	×		–	×	×			
Ammonia (Gas), Moist, NH_3	×	×	×	×	×	×		–	×	×	×	–	
Ammonium Carbonate, $(NH_4)_2CO_3$	+	+	+	×	×			×					
Ammonium Chloride, NH_4Cl	+	+	×	×	×	–	×	×	×	×		×	+
Ammonium Hydroxide, NH_4OH	+	+	+	×	×	×	×	–	×	×	×	×	–
Ammonium Nitrate, NH_4NO_3	O	×	O	×	×			×					
Ammonium Persulfate, $(NH_4)_2S_2O_8$	O	×	O	×	×			×					
Ammonium Phosphate, $(NH_4)_1H_2PO_4$				×	×								
Ammonium Phosphate, $(NH_4)_2HPO_4$				×	×								
Ammonium Phosphate, $(NH_4)_3PO_4$	×	×	×	×	×	×		×					
Ammonium Sulfate, $(NH_4)_2SO_4$	×	×	×	×	×	×	×	×	×	×		×	
Amyl Acetate, $C_5H_{11}COOCH_3$	×	×	×	×	×				×				
Amyl Alcohol, $C_5H_{11}OH$				×	×								

	CARBONS & CERAMICS									RUBBERS				
Chemical	*Carbon & Graphite*	*Glass, "Pyrex" brand*	*Silicaware*	*Silicate Cements*	*Chemical Stoneware*	*Transite (asbestos & cement)*	*Chemical Porcelain*	*Concrete—Unbonded*	*Concrete—Motor Bonded*	*Hard Rubber (Natural)*	*Soft Rubber (Natural)*	*Neoprene*	*Butadiene Derivatives*	*Nitrile Rubber (Chemigum)*
Acetic Acid, 100% CH_3COOH	×	×	×	×	×		×			×	–	O	O	O
Acetic Acid, Dilute	×	×	×		×		×			×	–	O	×	O
Acetic Anhydride, $(CH_3CO)_2$ O	×	×	×	×	×		×			×	–	O	–	O
Acetone, CH_3 $COCH_3$	×	×	×		×	×	×			O	O	O	×	–
Acetyl Chloride, CH_3COCl		×					×							
Aluminum Chloride, $AlCl_3$	×	×	×	×	×		×			×	×	×	×	×
Aluminum Hydroxide, $AL(OH)_3$		×					×							
Aluminum Sulfate, Al_2 $(SO_4)_3$	×	×	×		×	×	×			×	×	×	×	–
Alums, Conc., $Al_2(SO_4)_3K_2SO_4$, etc.	×	×					×					×		
Alums, Dilute	×	×	×	×	×	×	×			×	×	×	×	×
Amines, various	×	×			×		×			–				
Ammonia (Gas), Moist, NH_3	×	×	×	O	×	×	×			–	–	×	×	
Ammonium Carbonate, $(NH_4)_2CO_3$		×		×			×			×				O
Ammonium Chloride, NH_4Cl	×	×	×	×	×	×	×	×		×	×	×	×	–
Ammonium Hydroxide, NH_4OH	×	–			×		×			–	×	–		×
Ammonium Nitrate, NH_4NO_3	×	×	×		×		×	×		–	–	–	×	
Ammonium Persulfate, $(NH_4)_2S_2O_8$		×												
Ammonium Phosphate, $(NH_4)_1H_2PO_4$		×					×							
Ammonium Phosphate, $(NH_4)_2HPO_4$		×					×							
Ammonium Phosphate, $(NH_4)_3PO_4$	×	×			×		×			×				
Ammonium Sulfate, $(NH_4)_2SO_4$	×	×	×		×		×			×	×	×	×	
Amyl Acetate, $C_5H_{11}COOCH_3$	×	×	×		×		×	×		O	O	O	O	
Amyl Alcohol, $C_5H_{11}OH$	×	×	×		×	×	×			–	–	×	×	

	THERMOPLASTICS														
Chemical	*Viton*	*Asphaltic, Bitumastic*	*Cellulose Acetate*	*Cellulose Acetatebutyrate*	*Ethyl Cellulose (Ethocel)*	*Cellulose Nitrate*	*Acrylic (Lucite, Plexiglas)*	*Coumarone Resins*	*Polyethylene*	*Polyvinyl Chloride, Rigid or Unplasticized*	*Tygon (PVC & Copolymers)*	*Saran (Vinyl Chloride, Vinylidene Chloride)*	*Kel-F (Polytrifluorochloroethylene)*	*Teflon (Polytetrafluoroethylene)*	*Uscolite CP (Styrene–Acrylonitrile–Butadiene)*
Acetic Acid, 100% CH_3COOH	O	O	O		×	O	O		O	O	O	–	×	×	O
Acetic Acid, Dilute		×	×	×	×	×	×	×	×	–	–	×	×	×	–
Acetic Anhydride, $(CH_3CO)_2$ O	O	O	O		O	O	O			O	O	–		×	
Acetone, CH_3 $COCH_3$	O	O	O	O	O	O	O	O	O	O		O		×	O
Acetyl Chloride, CH_3COCl													×	×	
Aluminum Chloride, $AlCl_3$	×	×				×			×	×	×	×	×	×	×
Aluminum Hydroxide, $AL(OH)_3$	×								×	×			×	×	
Aluminum Sulfate, Al_2 $(SO_4)_3$	×	×				×	×	×	×	×	×	×	×	×	×
Alums, Conc., $Al_2(SO_4)_3K_2SO_4$, etc.									×	×	×	×	×	×	×
Alums, Dilute		×				×	×	×	×	×	×	×	×	×	×
Amines, various	O									O			×	×	
Ammonia (Gas), Moist, NH_3	O		–	×	×		×		×	–		O	×	×	–
Ammonium Carbonate, $(NH_4)_2CO_3$		×							×	×		×	×	×	×
Ammonium Chloride, NH_4Cl	×	×	×	×		×	×	×	×	×	×	×	×	×	×
Ammonium Hydroxide, NH_4OH	+	–					×		×	×	–	O	×	×	×
Ammonium Nitrate, NH_4NO_3	+	×	×			×	×		×	×		×	×	×	×
Ammonium Persulfate, $(NH_4)_2S_2O_8$		×								×	×		×	×	
Ammonium Phosphate, $(NH_4)_1H_2PO_4$										×			×	×	×
Ammonium Phosphate, $(NH_4)_2HPO_4$										×			×	×	×
Ammonium Phosphate, $(NH_4)_3PO_4$									×	×	×		×	×	×
Ammonium Sulfate, $(NH_4)_2SO_4$		×				×	×	×		×	×	×	×	×	×
Amyl Acetate, $C_5H_{11}COOCH_3$	O	O	O	O	O	O	O	O	–	O		O	×	×	
Amyl Alcohol, $C_5H_{11}OH$	×	O	×	O	O	O	×	–	–	×		×	×	×	×

	THERMOSETTING PLASTICS																			
Chemical	*Penton (Chlorinated Polyether)*	*Shellac Compounds*	*Organic Polysulfides*	*Polystyrene (Styron)*	*Vinylidene Chlorides*	*Vinyl Chloride Acetates*	*Cast Phenol Formaldehyde*	*Haveg 41 (Phenolic w. Asbestos)*	*Heresite (Phenol Formaldehyde)*	*Molded Phenolformald (Durez)*	*Phenol Furfural Plastics*	*Urea Formaldehyde*	*Casein Plastics*	*Epoxy Resins*	*Furane Resins (Haveg 61, Duralon)*	*Silicone Resins*	*Permanite (Furan, Glass Fiber)*	*Nylon (Adipic Acid–Hexameth, Diamine)*	*Durcon 6 (Modified Epoxy)*	*Cypress*
Acetic Acid, 100% CH_3COOH	+		×	O	O	O	–	–	×	×	–			–	×	O				
Acetic Acid, Dilute	×	×	×	O	–	–	×	×	×	×	×		×	×	×	×	×	O	×	×
Acetic Anhydride, $(CH_3CO)_2$ O	+			O			–	O	×	×	×				O	–	×			
Acetone, CH_3 $COCH_3$	+	O	×	–	O	O	–	O		–	–	×	×	×		×	×			
Acetyl Chloride, CH_3COCl																×				
Aluminum Chloride, $AlCl_3$	×	×		×	×	×	×	×	×	×	×			×	×	×	×	×		
Aluminum Hydroxide, $AL(OH)_3$	×															×	×			
Aluminum Sulfate, Al_2 $(SO_4)_3$	×	×	×	×	×	×	–	×	×	×			×		×	×	×	×	×	
Alums, Conc., $Al_2(SO_4)_3K_2SO_4$, etc.	×							×	×					×		×				
Alums, Dilute	×	×	×	×	×	×	–	×	×	×			×	×	×	×	×			
Amines, various								×	×							×	×			
Ammonia (Gas), Moist, NH_3	×	×		O	×	×	×			×	–			×	×	–	×			
Ammonium Carbonate, $(NH_4)_2CO_3$	×			×						×	×			×	×	×	×			
Ammonium Chloride, NH_4Cl	×	×		×	×	×	×	×	×	×	×		×	×	×	×	O		+**	
Ammonium Hydroxide, NH_4OH	×			×	O			×	×	O	O			×	×	–	×	×	×	
Ammonium Nitrate, NH_4NO_3	×	–		×	×	×	×			×	×		×	×		×	×			
Ammonium Persulfate, $(NH_4)_2S_2O_8$	×													×	×	×				
Ammonium Phosphate, $(NH_4)_1H_2PO_4$	×															×				
Ammonium Phosphate, $(NH_4)_2HPO_4$	×															×				
Ammonium Phosphate, $(NH_4)_3PO_4$	×							×	×							×				
Ammonium Sulfate, $(NH_4)_2SO_4$	×	×	×		×	×	×	×	×	×	×		×	×		×	O			
Amyl Acetate, $C_5H_{11}COOCH_3$	×	O		O	O	×	–			–	–		×	×		×	×	×		
Amyl Alcohol, $C_5H_{11}OH$	×	O		×	–	–	×			×	×		×	×		×	×			

	WOODS				
Chemical	*Fir*	*Maple*	*Oak*	*Pine*	*Redwood*
Acetic Acid, 100% CH_3COOH	–	×	×	×	–
Acetic Acid, Dilute	×	×	×	×	×
Acetic Anhydride, $(CH_3CO)_2$ O					
Acetone, CH_3 $COCH_3$					
Acetyl Chloride, CH_3COCl					
Aluminum Chloride, $AlCl_3$					
Aluminum Hydroxide, $AL(OH)_3$					
Aluminum Sulfate, Al_2 $(SO_4)_3$	×	×	–	×	×
Alums, Conc., $Al_2(SO_4)_3K_2SO_4$, etc.					
Alums, Dilute	×	×	–	×	×
Amines, various					
Ammonia (Gas), Moist, NH_3					
Ammonium Carbonate, $(NH_4)_2CO_3$					
Ammonium Chloride, NH_4Cl					
Ammonium Hydroxide, NH_4OH					
Ammonium Nitrate, NH_4NO_3					
Ammonium Persulfate, $(NH_4)_2S_2O_8$					
Ammonium Phosphate, $(NH_4)_1H_2PO_4$					
Ammonium Phosphate, $(NH_4)_2HPO_4$					
Ammonium Phosphate, $(NH_4)_3PO_4$					
Ammonium Sulfate, $(NH_4)_2SO_4$					
Amyl Acetate, $C_5H_{11}COOCH_3$					
Amyl Alcohol, $C_5H_{11}OH$					

*Note: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.
**Durcon 5 would be the preferred formula.

TABLE A.3 Continued

Chemical Resistance of Materials

MATERIALS
× — *Very Good Service*
+ — *Moderate Service*
– — *Limited or Variable Service*
O — *Unsatisfactory*
Blank — No Information

CHEMICALS
Solids Assumed in Solution
Room Temperatures Assumed
Unless Otherwise Stated

	METALS																												CARBONS & CERAMICS									RUBBERS					THERMOPLASTICS															THERMOSETTING PLASTICS																				WOODS				
	Carbon Steel; Fe	*Cast Iron & Ductile Iron; Fe*	*304 Stainless Steel; Fe, 18Cr, 8Ni*	*316 Stainless Steel; Fe, 16Cr, 10Ni, 2Mo*	*347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb*	*Ni-Resist Iron; Fe 14Ni, 2Cr, 2Si*	*Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo, 1Si*	*Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si*	*Durorpn; Fe, 14Si; Durichlor, Fe, 14Si, 3Mo**	*Copper; Brass; Bronzes; Everdur*	*Aluminum; Al (and Alloys)*	*Lead; Pb*	*Monel; 67Ni, 30Cu, 1.4Fe*	*Nickel; Ni*	*Inconel; 76Ni, 15Cr, 8Fe*	*Hastelloy B; Ni, 26Mo, 4Fe*	*Hastelloy C; Ni, 16Mo, 4Fe,14Cr, 4W*	*Hastelloy D; Ni, 8Si, 3Cu*	*Chlorimet 3; 3Fe, 1Si, 60Ni,18Mo, 18Cr*	*Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si*	*Stellite; Co, 28Cr, 4W*	*Zirconium; Z*	*Tantalum; Ta*	*Silver; Ag*	*Platinum; Pt*	*Dowmetal; (Mg alloys)*	*Titanium; Ti*	*Molybdenum; Mo*	*Carbon & Graphite*	*Glass, "Pyrex" brand*	*Silicaware*	*Silicate Cements*	*Chemical Stoneware*	*Transite (asbestos & cement)*	*Chemical Porcelain*	*Concrete—Unbonded*	*Concrete—Motor Bonded*	*Hard Rubber (Natural)*	*Soft Rubber (Natural)*	*Neoprene*	*Butadiene Derivatives*	*Nitrile Rubber (Chemigum)*	*Viton*	*Asphaltic, Bitumastic*	*Cellulose Acetate*	*Cellulose Acetatebutyrate*	*Ethyl Cellulose (Ethocel)*	*Cellulose Nitrate*	*Acrylic (Lucite, Plexiglas)*	*Coumarone Resins*	*Polyethylene*	*Polyvinyl Chloride, Rigid or Unplasticized*	*Tygon (PVC & Copolymers)*	*Saran (Vinyl Chloride, Vinyldene Chloride)*	*Kel-F (Polytrifluorochloroethylene)*	*Teflon (Polytetrafluoroethylene)*	*Uscolite CP (Styrene–Acrylonitrile–Butadiene)*	*Penton (Chlorinated Polyether)*	*Shellac Compounds*	*Organic Polysulfides*	*Polystyrene (Styron)*	*Vinyldene Chlorides*	*Vinyl Chloride Acetates*	*Cast Phenol Formaldehyde*	*Haveg 41 (Phenolic w. Asbestos)*	*Heresite (Phenol Formaldehyde)*	*Molded Phenolformald (Durez)*	*Phenol Furfural Plastics*	*Urea Formaldehyde*	*Casein Plastics*	*Epoxy Resins*	*Furane Resins (Haveg 61, Duralon)*	*Silicone Resins*	*Permanite (Furan, Glass Fiber)*	*Nylon (Adipic Acid–Hexameth. Diamine)*	*Durcon 6 (Modified Epoxy)*	*Cypress*	*Fir*	*Maple*	*Oak*	*Pine*	*Redwood*
Amyl Chloride, $C_5H_{11}Cl$	×		–	–	–		×	–	×	×	O		–	+	–	×	×	×	×	×			×						×	×	×		×		×			–	O	–	×		–	O								O		×	×	×		×				×	×	–			–	–		×	×		×										
Antimony Trichloride, $SbCl_3$	O		O	O	O				×	–	O	×	×	×	×	×	×		×	×			×							×	×		×		×						×		×				–					×		×	×	×	×	×													×		O	×								
Arsenic Acid, $HAsO_8$	O		–	–		×	×	×	×	–		–			–				×	×			×	×					×	×					×					×			×								–	×			×	×		×															×	×								
Barium Carbonate, $BaCO_3$			×	×	×		×	×	×	×	O		+	×	×	×	×	×	×		×	×					×			×					×								×									×			×	×	×	×													×		×	×								
Barium Hydroxide, $Ba(OH)_2$	×			×		×	×	×	+	O	O		×	×		+	+	+	×	×		×	×				+			×					×								×	×							×	×			×	×	×	×													×		×	×								
Barium Sulfide, BaS	O		×	×		×	×	×	×	O	O		–						×	×				O						–					×			×					×								×	–	×	×	×	×	×	×															×	×								
Benzaldehyde, C_6H_5CHO	×		×	×			×		×		–	O							×	×									×	×	×		×	×	×	×		O	O	O	×			O							O	O		–	×	×		–	O			×		×			×	×			–		×	×								
Benzene, C_6H_6	+	+	×	×	×	×	×	×	×	×	×	+	+	+	×	+	+	+	×	×	×		×	×	×	×			×	×		×	×		×			O	O	O		O	–	O	–	O	O		O		–	O	O	–	×	×		+		O			O		–	×	×	×			×	×	×	×	$×^{**}$							
Benzoic Acid, C_6H_5COOH	O		×	×	+		×	×	×	+	×	O	+	+	+	×	×	×	×	×			×	×						×					×							×		×								×	×	×	×	×		×													×		O	$+^{**}$								
Borax, $Na_2B_4O_7$	×	–	×	×	×	×	×		×	–	–	×	×	×					×	×				O					×	×	×		×	×	×	×		×	×	×	×	O ×	×	×	×	×	×	×	×	×		×			×	×	×	×	×		×	×	×	×			×	×	×					×	×	×			–	–	–	×	–
Boric Acid, H_3HO_3	O	O	+	+	+	–	×	×	×	+	+	+	+	+	×	×	×	×	×	×	×		×	×	×		×		×	×		×	×		×			×	×	×		× ×	×									–	×	×	×	×	×	×							×	×	×	×				×	×	×	×							
Bromine, Wet, Br_2	O		O	O	O	O	O	O	O	O	O	–	O	O	O	O	×	O	×	O	–		×	O	–		O		O	×	×		×		×	×		–	O		–	× ×	–	O							O	O	–	O	×	×		–			–				O	×	–	–	–		O	O	O	O	$×^{**}$							
Butanol, C_4H_2OH	×		×	×			×		×	×	×		×	×	×	×	×	×	×	×										×					×			–	–	–		× ×							×			×		×	×	×	×	×														–		×	×								
Butyl Acetate, $C_4H_9COOCH_3$	–			×			×	×	×	×	×	×	+			×	×	×	×	×			×			–			×	×	×		×	×	×	×	×	O	O	O	O		O	O	O	O	O	O	O			O		O	×	×		+			O	O	O	–			–	–	O	×	×		×										
Butyric Acid, C_3H_7COOH	O	O	O	×	O		×	×	×	–	–	O	+	+	+	×	×	×	×	×			×				×			×					×			×	–	O			O	O								O		–	×	×		×													–		O	×								
Calcium Bisulfate, $CaHSO_4$	O		O	×		O	×			O	O		O	O					×	×										×					×																				×	×																	×									
Calcium Bisulfate, $CaHSO_3$	×		O	×		O	×	×	O	–	–	×	O						×	×							×			×	×		×		×			×	×	×	×											×		×	×	×	×	×						×			×	×					×									
Calcium Carbonate, $CaCO_3$	×		×	×		×	×	×	×	×	O	×	×	×	×	×	×	×	×	×			×				×			×					×			×	×	–			×									×		×	×	×		×															×	×								
Calcium Chlorate, $CaClO_3$			–	×	O		+	–	+	–			+	+	+		+	+	+	+			×						×	×					×									×								×			×	×	×	×													×		×									
Calcium Chloride, $CaCl_2$	+	+	–	–	–	–	+	×	×	+	–	O	+	+	+	+	×	+	×	×	–	×	×				×		×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	–	×	×	–	–			×	×	×	O	×		×	×	×	×	×	–
Calcium Hydroxide, $Ca(OH)_2$	×		×	×	×	×	×	×	+	+	–		×	×	×	×	×	×	×	×			×				×			×					×			×	×	×			×	×								×		×	×	×	×	×													×		×	×	×							
Calcium Hypochlorite, $Ca(OCl)_2$	–	O	O	–	–	–	–	–	×	O	O	O	O	O	O	O	–	O	–	O	O	×	×		×		×		×	×	×	O	×		×	×		×	×	O	×	–		–							×	×	×	×	×	×	×	×		×	×		×	O	–	–	O	O			×	×	O	×	O			O	O	O	–	O
Calcium Sulfate, $CaSO_4$	–		×	×			×	×	×	×	–		+			+	+	+	×	×			×							×					×			×	×	×			×	×								×		×	×	×	×	×													×		×	×								
Carbon Dioxide (Dry), CO_2	×		×	×	×	×	×		×	×	×	×	×	×	×	×	×		×	×				×				×		×					×			×	×	×				×								×		×		×	×	×													×		O									
Carbon Dioxide (Wet or H_2CO_3)	–	–	×	×	×	–	×	×	×	–	–	–	–	×	×	×	×	×	×	×	×		×	×					×	×	×		×	×	×	×	×	×	×	×	×			×	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	×	×	×	×	×	×			×	×				×	×	×	×	×
Carbon Disulfide, CS_2	×	+	+	+	×	×	×	×	×	–	–	+	+	+	+				×	×			×	–					×	×	×	×	×		×			O	O	O	×		×	O	O	O	O	×	×		O	O		–	×	×		–	–		O	O		×			×	×			O	×	×	×	×							

**Note*: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.

**Durcon 5 would be the preferred formula.

TABLE A.3 Continued

Chemical Resistance of Materials

MATERIALS
× — Very Good Service
+ — Moderate Service
– – Limited or Variable Service
O — Unsatisfactory
Blank — No Information

	METALS																												CARBONS & CERAMICS									RUBBERS					THERMOPLASTICS															THERMOSETTING PLASTICS																				WOODS				
CHEMICALS *Solids Assumed in Sol'n.* *Room Temperatures Assumed* *Unless Otherwise Stated*	Carbon Steel; Fe	Cast Iron & Ductile Iron; Fe	304 Stainless Steel; Fe, 18Cr, 8Ni	316 Stainless Steel; Fe, 16Cr, 10Ni, 2Mo	347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb	Ni-Resist Iron; Fe, 14Ni, 2Cr, 2Si	Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo, 1Si	Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si	Duriron; Fe, 14Si; Durichlor, Fe, 14Si, 3Mo*	Copper; Brass; Bronzes; Everdur	Aluminum; Al (and Alloys)	Lead; Pb	Monel; 67Ni, 30Cu, 1.4 Fe	Nickel; Ni	Inconel; 76Ni, 15Cr, 8Fe	Hastelloy B; Ni, 26Mo, 4Fe	Hastelloy C; Ni, 16Mo, 4Fe, 14Cr, 4W	Hastelloy D; Ni, 8Si, 3Cu	Chlorimet 3; 3Fe, 1Si, 60Ni, 18Mo, 18Cr	Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si	Stellite; Co, 28Cr, 4W	Zirconium; Zr	Tantalum; Ta	Silver; Ag	Platinum; Pt	Dowmetal; (Mg alloys)	Titanium; Ti	Molybdenum; Mo	Carbon & Graphite	Glass, "Pyrex" brand	Silicaware	Silicate Cements	Chemical Stoneware	Transite (asbestos & cement)	Chemical Porcelain	Concrete-Unbonded	Concrete-Mortar Bonded	Hard Rubber (Natural)	Soft Rubber (Natural)	Neoprene	Butadiene Derivatives	Nitrile Rubber (Chemigum)	Viton	Asphaltic, Bitumastic	Cellulose Acetate	Cellulose Acetatebutyrate	Ethyl Cellulose (Ethocel)	Cellulose Nitrate	Acrylic (Lucite, Plexiglas)	Coumarone Resins	Polyethylene	Polyvinyl Chloride, Rigid or Unplasticized	Tygon (P. V. C. & Copolymers)	Saran (Vinyl chloride, vinylidene chloride)	Kel-F (Polytrifluorochloroethylene)	Teflon (Polytetrafluoroethylene)	Uscolite CP (Styrene-acrylonitrile-butadiene)	Penton (Chlorinated Polyether)	Shellac Compounds	Organic Polysulfides	Polystyrene (Styron)	Vinylidene Chlorides	Vinyl Chloride Acetates	Cast Phenol Formaldehyde	Haveg 41 (Phenolic w. asbestos)	Heresite (Phenol formaldehyde)	Molded Phenolformald. (Durez)	Phenol Furfural Plastics	Urea Formaldehyde	Casein Plastics	Epoxy Resins	Furane Resins (Haveg 61, Duralon)	Silicone Resins	Permanite (Furan, Glass Fiber)	Nylon (Adipic Acid—Hexameth. Diamine)	Durcon 6 (Modified Epoxy)	Cypress	Fir	Maple	Oak	Pine	Redwood
Carbon Tetrachloride (Moist) CCl_4	–	–	O	–	–	–	×	×	×	–	–	+	+	+	×	+	×	+	×	×	×	×	×	×			×		×	×	×	×	×	×	×	×		O	O	O	×	–	×	O	×	–		×	×		O	–	O	–	×	×		×	O	–	O	×	×	×	×	×	×	×			×	×		×	×	×						
Chloracetic acid, $ClCH_2CO_2H$	O	O	O	O	O		–	O	×	O	O	O	–	–	+	×	×	–	×	×	–		×	×	×		×			×			×		×			–	–	–	×		×	O			–					×	O	–	×	×		+		×	×				×	×	×	×							O	×						
Chloric Acid, $HClO_3$			O	O																				O						×																									×	×																			O						
Chlorine (Dry), Cl_2	×	×	×	×	×		×	×	×	×	O	×	×	×	×	×	×	×	×	×			×	×	×		O	+	×	×	×		×		×			×	–	O	–		×	O	–	–					×	–	×	–	×	×	×	+			×	×	×	–	–	–	–	–			–		–	–	O	O						
Chlorine (Wet), Cl_2	O	O	O	O	O	O	O	O	–	O	O	–	O	–	O	O	×	O	×	O		O	×	×	×		–	O	–	×	×		×		×			×	×	O	×			–	–	–					×	–	×	–	×	×	×	+			×	×	×	–	–		–	–			O		–	–	O	+						
Chlorobenzene, C_6H_5Cl	×		×	×	×		×	×	×	+			–	–	–				×	×			×							×	×		×		×			O	O	O			–	O					–		O	O		–	×	×		+		O	O	O			–						–	×		×		–**						
Chloroform, $CHCl_3$	×		×	×	×		×	×	×	×	–	×	×	×	×	×	×	×	×	×			×				×		×	×	×		×		×			O	O	O			–	O	O	O	O		O		O	O		–	×	×		+													×				×	×**						
Chromic Acid, CrO_3 solution	–	+	O	–	O	O	+	×	×	O	+	+	O	O	–	O	+	O	+	O	×	×	×	O	×	–	×	+	O	×	×	×	×		×			O	O	O	O	O	–	O	O	O	O		O	O	×	–	×	×	×	×	–	–			O	×		O	O	×	O	O			O	O		O	O	×						
Copper Chloride, $CuCl_2$	O		O	O	O		O	×	O	O	O		O	O	O	×	×	×	–	O		–	×	–			×			×	×		×		×			–	×	×	×		×	×							×	×		×	×	×	×	×	–						×							×	×			O	×						
Copper Cyanide, $Cu(CN)_2$			×	×	×		×	×	×	O	O		O	O	O	×	×	×	×	×			×							+					×									×								×				×	×	×															×				×						
Copper Nitrate, $Cu(NO_3)_2$	O		×	×	–		×	×	×	O	O		O	O	O	O	×	O	×	×			×	O						×					×			×		×			×								×	×		×	×	×	×	×																	×	×						
Copper Sulfate, $CuSO_4$	O	O	×	×	×	–	×	×	×	O	O	–	–	–	–	×	×	×	×	×	×	×	×	–	×				×	×	×	×	×		×			×	×	×	×	O	×	×	×	×		×	O		×	×	×	×	×	×	×	×	×		×			×	×	×	×	×			×	×		×	O	×						
Cresylic acid	O		×	×	×		×	×	×	O	–		–	×	–	–	–	–	×	×										×					×			O	O	O			–									–			+	×		×																	O	×						
Dichlorethane, $C_2H_4Cl_2$	O		×	×	×		×		×				×	×	×	×	×		×	×										×					×																			–	×																			×						
Diethylamine, $(C_2H_5)_2NH$	×			×			×												×	×									×	×					×					O												O			×	×		×																	×						
Diphenyl, $C_6H_5C_6H_5$	×		×				×		×					×					×	×										×					×																				×	×																			×						
Ethers, Various	×		×	×	×		×	×	×	×	×	×	+	+	+	×	×	×	×	×									×	×	×		×	×	×	×	×	O	O	O	O		O	O	O	O	O	O	O		O	O		–	×	×		×			O	O	O	–	–		–	–		×	–	×		×		×						
Ethyl Acetate, $C_2H_5COOCH_3$	×		×	×		×	×	×	×	×	×		×	×	×	×	×	×	×	×			×						×	×	×		×	×	×	×	×	O	O	O	O		O	O	O	O	O	O	O		O	O		–	×	×		+			O	O	O	–	O		–	–		×	×	×		×	×	×						
Ethyl Alcohol, C_2H_5OH	+	+	×	×	×	×	×	×	×	+	–	+	+	+	+	×	×	+	×	×		×	×			×			×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×	–	–		–		O	×	–	×	×	×	×	×	O	–	–	×	×		×		×	×	×		×	–			×	×						
Ethyl Chloride, C_2H_5Cl	×		×	×	×	O	×	–	×	–	–		+	+	×	×	×	×	×	×			×						×	×					×			O		–			×	O								O		–	×	×		×													×				×	×						
Ethylene Chlorohydrin, $Cl(C_2H_4)OH$	×		×				×		×			×							×	×									–	×	×		×		×			–	–	–	–			O	O	O						O		O	×	×		×				–		–			–	–			×					×	×					×
Ethylene Dichloride, $C_2H_4Cl_2$	×		×	×	×		×		×	–	O	×	+	+	–	×			×	×			×				×		×	×	×		×		×			O	O	O	O		–	O	O				O		–	O		–	×	×		+			O	O	O		×							×			×	×	×						
Ethylene Glycol, CH_2OHCH_2OH	–	+	×	×	×	×	×	×	×	+	×	+	+	+	+				×	×									×	×	×	×	×		×			–	–	–	×	×	×	×	×	×		×	×		O	×	–		×	×	×	×			×	–	–	–	×		–	–			×	×			×	×				×		
Ethylene Oxide, CH_2OCH_2	×		×				×		×										×	×										×					×								–									O		O		×		×																	×	×						
Fatty Acids, Various	O	O	×	×	×	–	×	×	×	–	×	O	+	×	×	×	×	×	×	×	×		×	×	×		×	O	×	×	×	×	×	×	×	×	×	O	O	O	×		×	O	×	×		×			O	–	×	×	×	×	×	×			O	×	–	–	×	×	×	–	×		×	×	–	×	×	×	×	×	×	×	×	–
Ferric Chloride, $FeCl_3$	O	O	O	O	O	O	O	O	–	O	O	O	O	O	O	O	+	O	–	O	–	O	×	O			×	O	×	×	×	×	×		×	×		×	×	–	×		×	×			–	×			×	×	×	–	×	×	×	×	×		×	×	×	×	×	×	×	×	×		×	×			O	+**						
Ferric Nitrate, $Fe(NO_3)_3$	O		×	×	×		×	×	×	O	O		O	O	O	O	+	O	×				×	O						×					×			O		×			×	×								×		×	×	×	×	×							×						×	×		×		×						
Ferric Sulfate, $Fe_2(SO_4)_3$	O		–	×	×	O	×	×	×	O	O	×	O	O	O	O	+	O	×		×		×	O			×		–	×	×		×		×			×	×	×	×		×	×			–	×				×	×	×	×	×	×	×	×		×	×	×	×	×	×	×	×	×		×	×		×	×	×						

* NOTE: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.

** Durcon 5 would be the preferred formula.

TABLE A.3 Continued

Chemical Resistance of Materials

MATERIALS
× — Very Good Service
+ — Moderate Service
– — Limited or Variable Service
O — Unsatisfactory
Blank — No Information

	METALS						CARBONS & CERAMICS		RUBBERS	THERMOPLASTICS			THERMOSETTING PLASTICS				WOODS
CHEMICALS Solids Assumed in Sol'n. Room Temperatures Assumed Unless Otherwise Stated	Carbon Steel; Fe Cast Iron & Ductile Iron; Fe 304 Stainless Steel; Fe, 18Cr, 8Ni 316 Stainless Steel; Fe, 16Cr, 10Ni, 2Mo 347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb	Ni-Resist Iron; Fe, 14Ni, 2Cr, 2Si Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo, 1Si Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si Duriron; Fe, 14Si; Durichlor, Fe, 14Si, 3Mo* Copper; Brass; Bronzes; Everdur	Aluminum; Al (and Alloys) Lead; Pb Monel; 67Ni, 30Cu, 1.4 Fe Nickel; Ni Inconel; 76Ni, 15Cr, 8Fe	Hastelloy B; Ni, 26Mo, 4Fe Hastelloy C; Ni, 16Mo, 4Fe, 14Cr, 4W Hastelloy D; Ni, 8Si, 3Cu Chlorimet 3; 3Fe, 1Si, 60Ni, 18Mo, 18Cr Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si	Stellite; Co, 28Cr, 4W Zirconium; Zr Tantalum; Ta Silver; Ag Platinum; Pt	Dowmetal; (Mg alloys) Titanium; Ti Molybdenum; Mo	Carbon & Graphite Glass, "Pyrex" brand Silicaware Silicate Cements	Chemical Stoneware Transite (asbestos & cement) Chemical Porcelain Concrete-Unbonded Concrete-Mortar Bonded	Hard Rubber (Natural) Soft Rubber (Natural) Neoprene Butadiene Derivatives Nitrile Rubber (Chemigum)	Viton Asphaltic, Bitumastic Cellulose Acetate Cellulose Acetatebutyrate Ethyl Cellulose (Ethocel)	Cellulose Nitrate Acrylic (Lucite, Plexiglas) Coumarone Resins Polyethylene Polyvinyl Chloride, Rigid or Unplasticized	Tygon (P. V. C. & Copolymers) Saran (Vinyl chloride, vinylidene chloride) Kel-F (Polytrifluorochloroethylene) Teflon (Polytetrafluoroethylene) Uscolite CP (Styrene-acrylonitrile-butadiene)	Penton (Chlorinated Polyether) Shellac Compounds Organic Polysulfides Polystyrene (Styron) Vinylidene Chlorides	Vinyl Chloride Acetates Cast Phenol Formaldehyde Haveg 41 (Phenolic w. asbestos) Heresite (Phenol formaldehyde) Molded Phenolformald. (Durez)	Phenol Furfural Plastics Urea Formaldehyde Casein Plastics Epoxy Resins Furane Resins (Haveg 61, Duralon)	Silicone Resins Permanite (Furan, Glass Fiber) Nylon (Adipic Acid—Hexameth. Diamine) Durcon 6 (Modified Epoxy) Cypress	Fir Maple Oak Pine Redwood
Ferrous Chloride, $FeCl_2$	OOOOO	×–O+O	OOOOO	–+++–	×O		××××	× ×	××××	××	× ×	–×××	×× ××	××× ×	×× ××	O×	
Ferrous Sulfate, $FeSO_4$	OO+O×	–×××–	+–OOO	+++××	– ×O	×	××××	× ×	×××××	××	× ××	×–×××	×× ××	×××××	×× ××	×O+	
Fluorine, F_2	OO	– OO	O–O××	+× +	O×	OO	OO	O		O	––	–O	O	O	OO	O	
Formaldehyde, CH_2O	–––––	–××××	–O–×–	×××××	× ××	–×	×××	× ××	–––×	O×O×	××××–	–×××–	×× ×	× ××	×××	××	
Formic Acid, HCOOH	OO++–	×××O	–O++–	–×–××	×××	O+	××××	× ×	–O–××	+O –	O ×–	×–××–	×– ××	×××××	×××××	O×	
Fuel Oil	× –××	×× ××	× ××+	×××××	×		×	×		×	O×	××	–×	×		+**	
Gallic Acid, $(OH)_3C_6H_2COOH$	O ×××	× ××	× –++	×××××			×	×		+	×	× ××	×			××	
Gasoline, Refined	× ×××	×× ××	× +++	×××××	×		×××	× ×	–O–×	×O××–	××	–×××–	× ––	––× –	– ×××	×××	
Glycerol, $CH_2OH\ CHOH\ CH_2OH$	+ +++	×× ××	×–+××	×××××	×	×	×××	× ××	××––×	××××	×× ××	×××××	× ×–	–– –	– ×	×××	
Hydrobromic Acid, HBr	O OOO	OO––	OOOOO	×–––+	×	×	×××	× ×	–×–×	+O×	××	–×××–	×××××	××× ×	× ×–	×O×	
Hydrochloric Acid, (Conc.), HCl	OOOOO	OOO+O	OOOOO	+–O+×	×××××	O×	××××	× ×	××–×O	–×OOO	O××××	×××××	××OOO	–××–	– ×–	–×O× O	–OOOO
Hydrochloric Acid, (Dilute)	OOOOO	–OO×O	O––––	+––××	–××××	–	×××	× ×	××–×–	×××–	×××××	×××××	×××××	××××–	– ××	–×O× ×	××××–
Hydrochloric Acid (Dry Gas)	× –×	× ×	O––––	×× ××	××		××	×		O	×	×××			×	×	
Hydrocyanic Acid, (Conc.), HCN	O ×××	××××O	×–×––	×××××	×		×	×		×	×	××–			×	×	
Hydrocyanic Acid, (Dilute & Gas)	OO				×		×	×	– –		×	× ×	–		×	×	
Hydrofluoric Acid, (Conc.), HF	–OOOO	O+–OO	OO+––	+++××	– O××	×O	OO	O O	OOO O		× ×–	OO××	+ ×	–OO–	– –×O	O– ×	
Hydrofluoric Acid, (Dilute)	OOOOO	–+OO–	O+++–	+++××	– O ×	OO	×O O	O O	××– ×	+O	× ×–	––×××	× ×	O–×	× ××	O×O×	
Hydrofluosilicic Acid, H_2SiF_6	O OOO	O+ OO	– –––	–––××	O		×–	O	×			– ×	×			× ×	
Hydrocarbons (Aliphatic)	×××××	×× ××	× ×××	×××××	× ×××	× +	×××	× ×	OO××	––	××	– ××	––×–	––××–	–××	+**	
Hydrocarbons (Aromatic)	×××××	×× ××	× ×××	×××××	× ×××	×	×××	× ×	OO O	×––	–	O +×	OOO–	––××–	– ×	×	
Hydrogen Gas, H_2	×	× ××	××××	××	×	×	×	×		××		××	×		×	××	
Hydrogen Peroxide (Conc.), H_2O_2	OO+O–	O+–+O	×–––	–×–××	× ×O×	×	××	× ×	––O×	––	× ×	– ××–	+ OOO	––××	–	O×	
Hydrogen Peroxide (Dilute)	OO++	×××O	×O+++	××–××	×××××	–	××××	× ×	––O××	––	× ××	×××××	OO–	×××–	– OO	O×	
Hydrogen Sulfide (Dry) H_2S	× –×	–× ×O	O×–×+	–×–×+	×O	+	××	× ××	×× ×	××	×	×××	× – ×	–× ×	– ××	×O×	
Hydrogen Sulfide (Wet)	– –×–	–××+O	OOO×+	–+–×+	O		×	×		×	×	××	×			×	
Iodine, I_4 Wet	OOOOO	O OO	OOOO–	O–O+O	– ×–×	–	O×–	××	––O–	×O–	–O	–××		O×	OO	OO×	
Iodoform, CHI_3	O O××	× ×	×	××			×	×		–		×				××	
Kerosene	× ×××	× ××	× ×××	×××××	×		××	×	OOO	×	× ×	××××	×			××	
Ketones, Various	× ×	× ×	–×	××			–×	×	O O	OO O	– OO	OO+×O	+ O–	×	–	×××	
Lactic Acid, $CH_3CHOHCOOH$	OO––×	–×××–	–O–––	×××××	×××××	××	×××	× ×	×–×	+ –×–	××	××××–	× ×	××	××	×–×	×
Lead Acetate, $Pb(CH_3COO)_2$	O ××+	×××–	O ×+	×××××	×	×	×	×		×	×	×××	×	×	××	× ×	

---

* NOTE: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.

** Durcon 5 would be the preferred formula.

TABLE A.3 Continued
Chemical Resistance of Materials

MATERIALS
× — Very Good Service
+ — Moderate Service
− — Limited or Variable Service
O — Unsatisfactory
Blank — No Information

CHEMICALS Solids Assumed in Solution Room Temperatures Assumed Unless Otherwise Stated	METALS																											
	Carbon Steel; Fe	Cast Iron & Ductile Iron; Fe	304 Stainless Steel; Fe, 18Cr, 8Ni	316 Stainless Steel; Fe 16Cr, 10Ni, 2Mo	347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb	Ni-Resist Iron; Fe 14Ni, 2Cr, 2Si	Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo. 1Si	Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si	Durorpm; Fe, 14Si; Durichlor, Fe, 14Si, 3Mo*	Copper; Brass; Bronzes; Everdur	Aluminum; Al (and Alloys)	Lead; Pb	Monel; 67Ni, 30Cu, 1.4Fe	Nickel; Ni	Inconel; 76Ni, 15Cr, 8Fe	Hastelloy B; Ni, 26Mo, 4Fe	Hastelloy C; Ni, 16Mo, 4Fe,14Cr, 4W	Hastelloy D; Ni, 8Si, 3Cu	Chlorimet 3; 3Fe, 1Si, 60Ni,18Mo, 18Cr	Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si	Stellite; Co, 28Cr, 4W	Zirconium; Z	Tantalum; Ta	Silver; Ag	Platinum; Pt	Dowmetal; (Mg alloys)	Titanium; Ti	Molybdenum; Mo
Magnesium Chloride, $MgCl_2$	−	−	−	+	−	×	×	×	×	+	−	O	+	+	+	×	×	×	×	×	−		×	×			×	
Magnesium Hydroxide, $Mg(OH)_2$	×		×	×	+		×	×	+	−	O		×	×	×	×	×	×	×	×			×	×			×	
Magnesium Sulfate, $MgSO_4$	+	O	+	+	+	×	×	×	×	+	×	+	+	+	+	×	+	×	×	×		×	×	×			×	
Maleic Acid; $CO_2HC_2H_2CO_2H$	O		×	×			×	×	×							×			×	×				×				
Malic Acid, $CO_2HCH_2CHOHCO_2H$	O		×	×	×		+		×	O	−		+	+	+	×	×	×	×				×					
Mercuric Chloride, $HgCl_2$	O		O	O	O		−	−	+	O	O	−	−		+	O	×	O	+	O		×	×	O			×	
Mercury, Hg	×		×	×	×		×		×	O	O	O	+	+	×	×	×	×	×	×				O			−	
Methanol, (Conc.), CH_3OH	+	+	+	+	+	×	×	×	×	+	−	+	+	+	+	×	×	+	×	×		×	×	×				
Methanol, (Dilute)	−						×		×	−	−					×	×	−	×	×			×					
Methyl Chloride, CH_3Cl	O		−	−	+		×	×	×	×		×	×	×	×				×	×			×					
Naphtha, Petroleum	×		×	×	×	×	×		×	×	×		+	+	+	×	×	×	×	×								
Nickel Chloride, $NiCl_2$	O		−	−			+	O	+	−	O		−	×		O	×	O	×	×		×	×				×	
Nickel Sulfate, $NiSO_4$	O	O	−	−			×	×	×	+	−	+	+	−	−	O	+	O	×	×		×	×					
Nitrating Acid (>15% H_2SO_4)	×	×	−	−	−		×	×	×	O	O		O	O		O	×	O	×	×	×		−	×	×		−	
Nitrating Acid (<15% H_2SO_4)	O		−	−	−		×	×	×	O	O		O	O		O	×	O	×	×	×		−	O	×		×	
Nitrating Acid (<15% HNO_3)	O		−	O	−		×		×	O	O		O	O		O	×	O	×	O	×	O	−	O	×		−	
Nitrating Acid (>1% Acid)	O		−	×	−		×		×	O	O		O	O		O	×	O	×		×	O	−	O	×			
Nitric Acid (Conc.), HNO_3	O	O	+	+	×		×	×	×	O	−	−	O	O	−	O	+	O	×	O	×	×	×	O	×		×	+
Nitric Acid, Dilute	O	O	×	×	×	O	×	×	×	O	O	O	O	O	O	O	×	O	×	O	×	×	×	O	×		×	O
Nitrobenzene, $C_6N_5NO_2$	×		×	×			×	×	×			O							×	×			×					
Nitrous Acid, HNO_2	O		×	×	−		×	×	×	×			×	O					×				×					
Oleic Acid, C_8H_{17} CH: $CH(CH_2)_7CO_2H$	−	−	−	×	×	−	×	×	×	−	×	O	×	×	×	×	×	×	×	×	×		×	×	×			
Oxalic Acid, CO_2HCO_2H	O	O	−	−	−	−	×	×	×	+	−	O	+	−	×	+	+	+	×	×	×	×	×	×	×		−	+
Phenol (Conc.), C_6H_5OH	+	+	−	×	−	×	×	×	×	O	−	−	+	+	×	+	+	+	×	×	×	×	×	×	×	×	−	
Phenol (Dilute)	+	+	×	×	×		×		×	−	×	−	+	+	−	+	+	+	×	×		×						
Phosphoric Acid (100%), H_3PO_4	O	O	−	−	×		×		×		O	+	−	O	−	+	×	×	×	×	×	×	×	+	×			×
Phosphoric Acid (>45% Hot)	O	O	O	O	×	O	×	×	×	O	O	+	O	O	O	−	−	−	×	×	×	−	×		×			

CHEMICALS Solids Assumed in Solution Room Temperatures Assumed Unless Otherwise Stated	CARBONS & CERAMICS									RUBBERS					THERMOPLASTICS														
	Carbon & Graphite	Glass, "Pyrex" brand	Silicaware	Silicate Cements	Chemical Stoneware	Transite (asbestos & cement)	Chemical Porcelain	Concrete—Unbonded	Concrete—Motor Bonded	Hard Rubber (Natural)	Soft Rubber (Natural)	Neoprene	Butadiene Derivatives	Nitrile Rubber (Chemigum)	Viton	Asphaltic, Bitumastic	Cellulose Acetate	Cellulose Acetatebutyrate	Ethyl Cellulose (Ethocel)	Cellulose Nitrate	Acrylic (Lucite, Plexiglas)	Coumarone Resins	Polyethylene	Polyvinyl Chloride, Rigid or Unplasticized	Tygon (PVC & Copolymers)	Saran (Vinyl Chloride, Vinyldene Chloride)	Kel-F (Polytrifluorochloroethylene)	Teflon (Polytetrafluoroethylene)	Uscolite CP (Styrene–Acrylonitrile–Butadiene)
Magnesium Chloride, $MgCl_2$		×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×		×		×	×	×	×	×	×	×	×
Magnesium Hydroxide, $Mg(OH)_2$		×					×			×			×		×									×		×	×	×	×
Magnesium Sulfate, $MgSO_4$		×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×		×		×	×	×		×	×	×	×
Maleic Acid; $CO_2HC_2H_2CO_2H$		×					×					×			+	O								−		−	×	×	×
Malic Acid, $CO_2HCH_2CHOHCO_2H$		×	×		×		×			×	×	×	×		−									×			×	×	
Mercuric Chloride, $HgCl_2$		×					×			×	−				×	×	×							−	×	×	×	×	×
Mercury, Hg		×					×								×	×								−		×		×	
Methanol, (Conc.), CH_3OH		×		×			×			−	×	−		×	−	×					O		−	×		×	×	×	×
Methanol, (Dilute)		×					×			×	×													×			×	×	×
Methyl Chloride, CH_3Cl		×					×			O	O				−	O								O		−	×	×	
Naphtha, Petroleum	×	×	×		×		×			−	O	×	×		+	O				×	×			×			×	×	
Nickel Chloride, $NiCl_2$	×	×					×			×	×				×	×							×	×	×	×	×	×	×
Nickel Sulfate, $NiSO_4$	×	×					×			×	×	×			×	×							×	×	×	×	×	×	×
Nitrating Acid (>15% H_2SO_4)	O	×			×		×			−	−					O									O		×	×	
Nitrating Acid (<15% H_2SO_4)	O	×			×		×			−	−					O									−		×	×	
Nitrating Acid (<15% HNO_3)	O	×			×		×			−	−					O									−		×	×	
Nitrating Acid (>1% Acid)	O	×			×		×			−	−					O									−		×	×	
Nitric Acid (Conc.), HNO_3	O	×	×	×	×		×			O	O	O	O	O	O				O				O	O	O	−	×	×	O
Nitric Acid, Dilute	−	×	×		×		×			−	O	O	×	O	−	−			−		×		−	×	−	×	×	×	×
Nitrobenzene, $C_6N_5NO_2$		×					×			O	O	O			O	O							−	O		−	×	×	
Nitrous Acid, HNO_2		×	×		×		×			−	O	O	×			×											×	×	
Oleic Acid, C_8H_{17} CH: $CH(CH_2)_7CO_2H$	×	×	×		×	×	×	×	×	O	O	O	×		+	O	×	×		×			O	×	×	×	×	×	×
Oxalic Acid, CO_2HCO_2H	×	×	×	×	×	×	×	×		×	×	−	×		−	×							×	×	×	×	×	×	×
Phenol (Conc.), C_6H_5OH	×	×	×		×	×	×	×		−	O	−	×		×	O	O	O	−		×			−	−	−	×	×	O
Phenol (Dilute)		×					×			−		−				−								×	−	−	×	×	O
Phosphoric Acid (100%), H_3PO_4	+	−	×		×	×	×	×	×	×	×	×	×			×	×		O				×	×	×		×	×	×
Phosphoric Acid (>45% Hot)	×	−		×			O			−	−	−		×	+								−	×	×		×	×	×

CHEMICALS Solids Assumed in Solution Room Temperatures Assumed Unless Otherwise Stated	THERMOSETTING PLASTICS																				WOODS				
	Penton (Chlorinated Polyether)	Shellac Compounds	Organic Polysulfides	Polystyrene (Styron)	Vinylidene Chlorides	Vinyl Chloride Acetates	Cast Phenol Formaldehyde	Haveg 41 (Phenolic w. Asbestos)	Heresite (Phenol Formaldehyde)	Molded Phenolformald (Durez)	Phenol Furfural Plastics	Urea Formaldehyde	Casein Plastics	Epoxy Resins	Furane Resins (Haveg 61, Duralon)	Silicone Resins	Permanite (Furan, Glass Fiber)	Nylon (Adipic Acid–Hexameth, Diamine)	Durcon 6 (Modified Epoxy)	Cypress	Fir	Maple	Oak	Pine	Redwood
Magnesium Chloride, $MgCl_2$	×	×	×	×	×	×	×	×	×	×	×	×		×	×	×	×								
Magnesium Hydroxide, $Mg(OH)_2$	×															×	×								
Magnesium Sulfate, $MgSO_4$	×	×	×	×	×	×	×			×	×	×		×	×	×	×								
Maleic Acid; $CO_2HC_2H_2CO_2H$	×													×		×	+**								
Malic Acid, $CO_2HCH_2CHOHCO_2H$	×			×												×	×								
Mercuric Chloride, $HgCl_2$	×													×		×									
Mercury, Hg	×													×		×	×								
Methanol, (Conc.), CH_3OH	×			×				−		×	×			×	×	×	×								
Methanol, (Dilute)	×			×						×	×			×		×	×								
Methyl Chloride, CH_3Cl	×													×		×	×								
Naphtha, Petroleum	×			−	−	−	−			−	−		×	×		×	×								
Nickel Chloride, $NiCl_2$	×													×	×	×									
Nickel Sulfate, $NiSO_4$	×			−						×	×			×	×	×	×								
Nitrating Acid (>15% H_2SO_4)								O	×					O		×**									
Nitrating Acid (<15% H_2SO_4)								−	×							×									
Nitrating Acid (<15% HNO_3)								−	×							×									
Nitrating Acid (>1% Acid)								−	×							×**									
Nitric Acid (Conc.), HNO_3	O			O	O	O	O	O	×	O	O			O	O	−	O	O	×						
Nitric Acid, Dilute	+			×	−	−	−	−	×	−	−			O		−	O	×							
Nitrobenzene, $C_6N_5NO_2$	+													−		×									
Nitrous Acid, HNO_2	×		×													×									
Oleic Acid, C_8H_{17} CH: $CH(CH_2)_7CO_2H$	×			×	×	×	−	×	×	×	−	×		×		×	×	×			×	×	×		−
Oxalic Acid, CO_2HCO_2H	×			×			×	×	×	×	×			×	×	×	O	×							
Phenol (Conc.), C_6H_5OH	+			O			−	−	×	−	−				×	×	O	×							
Phenol (Dilute)				O						×	×			−	×	×									
Phosphoric Acid (100%), H_3PO_4					×	×		×	×						−	−	−	−	×						
Phosphoric Acid (>45% Hot)								×	×	O	O			×	×	O	×								

---

*Note: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.
**Durcon 5 would be the preferred formula.

TABLE A.3 Continued

Chemical Resistance of Materials

MATERIALS
× — *Very Good Service*
+ — *Moderate Service*
– — *Limited or Variable Service*
O — *Unsatisfactory*
Blank — *No Information*

CHEMICALS
Solids Assumed in Solution
Room Temperatures Assumed
Unless Otherwise Stated

CHEMICALS	METALS																											
	Carbon Steel; Fe	*Cast Iron & Ductile Iron; Fe*	*304 Stainless Steel; Fe, 18Cr, 8Ni*	*316 Stainless Steel; Fe 16Cr, 10Ni, 2Mo*	*347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb*	*Ni-Resist Iron; Fe 14Ni, 2Cr, 2Si*	*Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo. 1Si*	*Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si*	*Durorpm; Fe, 14Si; Durichlor, Fe, 14Si, 3Mo**	*Copper; Brass; Bronzes; Everdur*	*Aluminum; Al (and Alloys)*	*Lead; Pb*	*Monel; 67Ni, 30Cu, 1.4Fe*	*Nickel; Ni*	*Inconel; 76Ni, 15Cr, 8Fe*	*Hastelloy B; Ni, 26Mo, 4Fe*	*Hastelloy C; Ni, 16Mo, 4Fe,14Cr, 4W*	*Hastelloy D; Ni, 8Si, 3Cu*	*Chlorimet 3; 3Fe, 1Si, 60Ni,18Mo, 18Cr*	*Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si*	*Stellite; Co, 28Cr, 4W*	*Zirconium; Z*	*Tantalum; Ta*	*Silver; Ag*	*Platinum; Pt*	*Dowmetal; (Mg alloys)*	*Titanium; Ti*	*Molybdenum; Mo*
Phosphoric Acid (>45% Cold)	O	O	–	×	×	–	×	–	×	O	O	+	–	–	–	+	×	+	×	×	×	×	×	×	×		–	
Phosphoric Acid (<45% Cold)	O	O	–	×	×	–	×	–	×	O	–	+	–	–	–	+	×	+	×	×	×	×	×	×	×		×	
Phosphoric Anhydride, Dry or Moist	O		×	×			×		×										×	×								
Phosphoric Anhydride Molten, P_2O_5	×		×	×			×		×	O	O		O	O					×	×								
Phthalic Anhydride, $C_6H_4(CO)_2O$	×		×				×	×	×							×	×		×	×			×					
Picric Acid, Solution, $HOC_6H_2(NO_3)_3$			×	×	+		×	×	×	O	–	–	O	O	O	×	×	×	×	×								
Potassium Bromide, KBr	O		–	–				×		×	–	×	+	×		×	×	×	×	×			×	O			×	
Potassium Carbonate, K_2CO_3	–		×	×		–	×	×	×	–	O	×	×	×	+	×	×	×	×	×			–					
Potassium Chlorate, $KClO_3$	×		×	×			×	–	×	–	+		+	+	+	O	×	O	×	O			×					
Potassium Chloride, KCl	×		–	–		×	×	×	×	–	–		+	+		×	×	×	×	×			×				×	
Potassium Cyanide, KCN	×		×	×		×	×		×	O	O		×			×	×	×	×	×			×	O	×			
Potassium Dichromate, $K_2Cr_2O_7$			×	×	–		×	×	×	×	×		–	–	+	O	×	O	×	×		×	×	O			×	
Potassium Ferrocyanide, $K_4Fe(CN)_6$	O		×	×			×	×	×	×	–	×	×	×	×	×	×	×	×	×			×	O				
Potassium Hydroxide, KOH	+	–	+	+	–	–	×	×	+	O	O	O	×	×	×	+	+	+	×	×		×	–	×			–	+
Potassium Nitrate, KNO_3	×		×	×	×		×	×	×	×	×	–	+	+	+	O	×	O	×	×			×	O				
Potassium Permanganate, $KMnO_4$	+	+	+	+	+		×	×	×	×	×	O	–	+	+	O	×	O	×	O			×				×	
Potassium Sulfate, K_2SO_4	×		+	+	+	×	×	×	×	+	×	×	+	+	+	+	+	+	×	×			×				×	
Potassium Sulfide, K_2S	–		×	×		–	×		×										×	×				O				
Pyrogallol, $C_6H_3(OH)_3$			×	×			×		×	×				+		×	×	×	×	×								
Silver Nitrate, $AgNO_3$			×	×			×	×	×	O	O	O			+	+	+	+	×	×		×	×					
Sodium, Molten 210–400°F	×	×	×	×	×		×	×	×	×	×		×	×	×	×	×		×	×		×	×			–		×
Sodium Acetate, $NaCH_3COO$	O		–	×			×	×	×	×	–	×	+			×	×	×	×	×							×	
Sodium Bicarbonate, $NaHCO_3$	–		×	×	×	×	×	×	×	–	×	×	×	×	×	×	×	×	×	×		×	×	×				
Sodium Bisulfate, $NaHSO_4$	O	O	–	×	×		×	×	×	–	O	×	+			×	×	×	×	×	×		–		×		×	
Sodium Bisulfite, $NaHSO_3$	O		×	×			×	×		–	×	×							×	×								

CHEMICALS	CARBONS & CERAMICS									RUBBERS				
	Carbon & Graphite	*Glass, "Pyrex" brand*	*Silicaware*	*Silicate Cements*	*Chemical Stoneware*	*Transite (asbestos & cement)*	*Chemical Porcelain*	*Concrete—Unbonded*	*Concrete—Motor Bonded*	*Hard Rubber (Natural)*	*Soft Rubber (Natural)*	*Neoprene*	*Butadiene Derivatives*	*Nitrile Rubber (Chemigum)*
Phosphoric Acid (>45% Cold)	×	×			–		–			–	–	–		
Phosphoric Acid (<45% Cold)	×	×			–		–			×	×	–		
Phosphoric Anhydride, Dry or Moist	×	×												
Phosphoric Anhydride Molten, P_2O_5		O												
Phthalic Anhydride, $C_6H_4(CO)_2O$		×					×			×	×			
Picric Acid, Solution, $HOC_6H_2(NO_3)_3$		×					×					–		
Potassium Bromide, KBr		×					×			×		×		
Potassium Carbonate, K_2CO_3		–					×			×		×		
Potassium Chlorate, $KClO_3$		×					×							
Potassium Chloride, KCl		×					×			×		×		
Potassium Cyanide, KCN		–					×			×		×		
Potassium Dichromate, $K_2Cr_2O_7$		×					×			–	×	×		
Potassium Ferrocyanide, $K_4Fe(CN)_6$		×					×							
Potassium Hydroxide, KOH		–		O			–			×	×	×		O
Potassium Nitrate, KNO_3		×					×			×		×		
Potassium Permanganate, $KMnO_4$		×					×			–	O	O		
Potassium Sulfate, K_2SO_4		×		×			×			×		×		×
Potassium Sulfide, K_2S		–					×							
Pyrogallol, $C_6H_3(OH)_3$		×					×							
Silver Nitrate, $AgNO_3$		×					×			×	×			
Sodium, Molten 210–400°F		–					O							
Sodium Acetate, $NaCH_3COO$		×					×					×		
Sodium Bicarbonate, $NaHCO_3$	×	×	×		×	×	×	×	×	×	×	×	×	
Sodium Bisulfate, $NaHSO_4$	×	×	×		×		×			×	×	×	×	
Sodium Bisulfite, $NaHSO_3$	×	×	×		×		×			×	×	–	×	

CHEMICALS	THERMOPLASTICS														
	Viton	*Asphaltic, Bitumastic*	*Cellulose Acetate*	*Cellulose Acetatebutyrate*	*Ethyl Cellulose (Ethocel)*	*Cellulose Nitrate*	*Acrylic (Lucite, Plexiglas)*	*Coumarone Resins*	*Polyethylene*	*Polyvinyl Chloride, Rigid or Unplasticized*	*Tygon (PVC & Copolymers)*	*Saran (Vinyl Chloride, Vinyldene Chloride)*	*Kel-F (Polytrifluorochloroethylene)*	*Teflon (Polytetrafluoroethylene)*	*Uscolite CP (Styrene–Acrylonitrile–Butadiene)*
Phosphoric Acid (>45% Cold)		×							×	×	×	×		×	×
Phosphoric Acid (<45% Cold)									×	×	×	×		×	×
Phosphoric Anhydride, Dry or Moist									×					×	
Phosphoric Anhydride Molten, P_2O_5														×	
Phthalic Anhydride, $C_6H_4(CO)_2O$		×										×		×	
Picric Acid, Solution, $HOC_6H_2(NO_3)_3$		O								O	×	O		×	
Potassium Bromide, KBr										×		×		×	×
Potassium Carbonate, K_2CO_3	×	×							×	–		×		×	×
Potassium Chlorate, $KClO_3$	×									×		×		×	×
Potassium Chloride, KCl	×	×							×	×		×		×	×
Potassium Cyanide, KCN	×									×	×	×		×	×
Potassium Dichromate, $K_2Cr_2O_7$		×							×	×	×	×		×	×
Potassium Ferrocyanide, $K_4Fe(CN)_6$		×								×		×		×	
Potassium Hydroxide, KOH	×	×							×	×		–		×	×
Potassium Nitrate, KNO_3		×							×	×		×		×	×
Potassium Permanganate, $KMnO_4$	×	O							×	×		–		×	×
Potassium Sulfate, K_2SO_4	×	×								×		×		×	×
Potassium Sulfide, K_2S										×				×	×
Pyrogallol, $C_6H_3(OH)_3$							×				×			×	
Silver Nitrate, $AgNO_3$	×						×		×	×		×		×	×
Sodium, Molten 210–400°F				O					O	O	O	O		–	
Sodium Acetate, $NaCH_3COO$	O	×								×		×		×	
Sodium Bicarbonate, $NaHCO_3$		×	×	×		×	×			×		×		×	×
Sodium Bisulfate, $NaHSO_4$										×	×	×		×	
Sodium Bisulfite, $NaHSO_3$		×							×	×	×	×		×	×

CHEMICALS	THERMOSETTING PLASTICS																				WOODS				
	Penton (Chlorinated Polyether)	*Shellac Compounds*	*Organic Polysulfides*	*Polystyrene (Styron)*	*Vinylidene Chlorides*	*Vinyl Chloride Acetates*	*Cast Phenol Formaldehyde*	*Haveg 41 (Phenolic w. Asbestos)*	*Heresite (Phenol Formaldehyde)*	*Molded Phenolformald (Durez)*	*Phenol Furfural Plastics*	*Urea Formaldehyde*	*Casein Plastics*	*Epoxy Resins*	*Furane Resins (Haveg 61, Duralon)*	*Silicone Resins*	*Permanite (Furan, Glass Fiber)*	*Nylon (Adipic Acid–Hexameth, Diamine)*	*Durcon 6 (Modified Epoxy)*	*Cypress*	*Fir*	*Maple*	*Oak*	*Pine*	*Redwood*
Phosphoric Acid (>45% Cold)	×			×				×	×					×		–	×								
Phosphoric Acid (<45% Cold)	×							×	×							–	×	O	×						
Phosphoric Anhydride, Dry or Moist																×									
Phosphoric Anhydride Molten, P_2O_5																×									
Phthalic Anhydride, $C_6H_4(CO)_2O$														×		×									
Picric Acid, Solution, $HOC_6H_2(NO_3)_3$	×															×									
Potassium Bromide, KBr	×															×									
Potassium Carbonate, K_2CO_3	×													×		×	×								
Potassium Chlorate, $KClO_3$	×															O	×								
Potassium Chloride, KCl	×							×						×	×	×	×	×							
Potassium Cyanide, KCN	×															×	×								
Potassium Dichromate, $K_2Cr_2O_7$	×													×		×**									
Potassium Ferrocyanide, $K_4Fe(CN)_6$	×													×		×	×								
Potassium Hydroxide, KOH	×			×				O		O	O			×	×	×	×	×							
Potassium Nitrate, KNO_3	×													×		×									
Potassium Permanganate, $KMnO_4$	×			×						×	×			×	×	O	×								
Potassium Sulfate, K_2SO_4	×			×				×		×	×			×	×	×	×	×							
Potassium Sulfide, K_2S																×	×								
Pyrogallol, $C_6H_3(OH)_3$																×**									
Silver Nitrate, $AgNO_3$	×						–			–				×		×	×								
Sodium, Molten 210–400°F	O			O	O			O	O							×									
Sodium Acetate, $NaCH_3COO$	×													×		–	×								
Sodium Bicarbonate, $NaHCO_3$	×	×		×	×	×	×	×		×	×			×	×	×	×	×							
Sodium Bisulfate, $NaHSO_4$	×			×	×	×	–	×	×	×	–		×	×	×	×	×								
Sodium Bisulfite, $NaHSO_3$	×									×			×	×		O					×	×	×	×	

---

*Note: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.

**Durcon 5 would be the preferred formula.

TABLE A.3 Continued

Chemical Resistance of Materials

MATERIALS
× — Very Good Service
+ — Moderate Service
– — Limited or Variable Service
O — Unsatisfactory
Blank — No Information

CHEMICALS Solids Assumed in Solution Room Temperatures Assumed Unless Otherwise Stated	METALS																											
	Carbon Steel; Fe	Cast Iron & Ductile Iron; Fe	304 Stainless Steel; Fe, 18Cr, 8Ni	316 Stainless Steel; Fe 16Cr, 10Ni, 2Mo	347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb	Ni-Resist Iron; Fe 14Ni, 2Cr, 2Si	Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo. 1Si	Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si	Durorpm; Fe, 14Si; Durichlor; Fe, 14Si, 3Mo*	Copper; Brass; Bronzes; Everdur	Aluminum; Al (and Alloys)	Lead; Pb	Monel; 67Ni, 30Cu, 1.4Fe	Nickel; Ni	Inconel; 76Ni, 15Cr, 8Fe	Hastelloy B; Ni, 26Mo, 4Fe	Hastelloy C; Ni, 16Mo, 4Fe,14Cr, 4W	Hastelloy D; Ni, 8Si, 3Cu	Chlorimet 3; 3Fe, 1Si, 60Ni,18Mo, 18Cr	Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si	Stellite; Co, 28Cr, 4W	Zirconium; Z	Tantalum; Ta	Silver; Ag	Platinum; Pt	Dowmetal; (Mg alloys)	Titanium; Ti	Molybdenum; Mo
Sodium Borate $NaBO_2$			×	×			×		×	×	–		+			×	×	×	×	×			×					
Sodium Carbonate, Na_2CO_3	+	+	+	+	×	×	×	×	×	+	O	–	+	+	–	+	×	×	×	×	×		–	×	×	×	×	
Sodium Chlorate, $NaClO_3$	–		×	×			–	–	×	+	+				+	O	×	O	×	O			×				×	
Sodium Chloride, NaCl	+	+	–	–	–	–	×	×	+	+	–	+	+	+	–	+	+	–	×	×	–	×	×	×			×	
Sodium Cyanide, NaCN	×		×	×			×	×	×		O	×	×						×	×				O			×	
Sodium Fluoride, NaF	O		–	–			+	–	O	–	–		+			×	×	×	×	×		×	O				×	
Sodium Hydroxide, (Conc.), NaOH	–	–	–	–	×	×	×	×	–	–	O	O	×	×	×	×	+	×	×	×	×	×	O	×	×		×	
Sodium Hydroxide (Dilute)	×	+	+	+	×		×	×	+	×	O	O	–	–	–	×	+	×	×	×	×	×	–		×	–	×	+
Sodium Hydrosulfite							×		×	–	×		O	O	O	–	×	–	×	×								
Sodium Hypochlorite, NaOCl	O	O	–	–	–	–	–	–	×	O	O	O	O	O	O	O	–	O	+	O	–		–	×	×		×	
Sodium Hyposulfate	–		×	×			×		×	O	O		×						×	×				O				
Sodium Nitrate, $NaNO_3$	×	+	×	×	×	–	×	×	×	+	×	O	+	+	×	O	+	O	×	O	×		–	–	×			
Sodium Peroxide, Na_2O_2	–		×	×	+		×		×	–	–		×	+	+	+	×	+	×	×			O	O				
Sodium Phosphate, (Tri) Na_3PO_4	×		×	×	+		×		×	–	O		+	+		×	×	×	×	×			×	×			×	
Sodium Silicate, Na_2SiO_3	×		×	×	+	×	×		×	–	–		+	×	+	×	×	×	×	×			×					
Sodium Sulfate, Na_2SO_4	×	×	–	–	×	×	×	×	×	×	×	–	+	×	×	×	×	×	×	×	–		–	×			×	
Sodium Sulfide, Na_2S	+	×	–	×	×	×	×	×	×	O	O	–	+	+	+	×	×	×	×	×	×		–	O			×	
Sodium Sulfite, Na_2SO_3	+	×	–	–	–	×	×	×	O	–	–	×	–	–	–	O	×	O	×	×	×		–				×	
Stannic Chloride, $SnCl_4$	O		O	O	O		×	–	×	O	O	×	O	O	O	O	×	O	–	O			×				×	
Stannous Chloride, $SnCl_2$	O		O	–			×	–	×	O	O		O	O	O	×	×	×	×	×		×	×				×	
Stearic Acid $CH_3(CH_2)_{16}COOH$	–	–	×	×	×	–	×	×	×	–	–	–	×	×	×	×	×	×	×	×	×		×	×			×	
Sulfur, Molten, S	–	×	–	–	–	–	×	×	×	O	×	O	–	–	×	O	O	O	×	–	O		×	O			×	+
Sulfur Chloride, (Wet), S_2Cl_2	–		O	–		×	–	–	×	O	O		–	–					×					O				
Sulfur Dioxide (Dry), SO_2	×	×	×	×	–	×	×	×	×	×	×	×	O	O	O	O	×	O	×	–	×		×	O		×		
Sulfur Dioxide (Wet)	O		–	×			×		–	O		×	O	O	O	O	×	O	×				×				+	
Sulfur Trioxide, SO_3	×		×	×		×	×	×	–	–	×		O	O	O				×				O	O				
Sulfuric Acid (Fuming to 98%)	–	–	O	O		–	×	×	O	O	–	O	O	O	O	×	O	×	×	×		O	O	O			O	
Sulfuric Acid (Hot Conc.) H_2SO_4	O		O	O			+		×	O	O	–	O	O	O	O	O	×	+	+	O	O	–				O	

CHEMICALS	CARBONS & CERAMICS									RUBBERS				
	Carbon & Graphite	Glass, "Pyrex" brand	Silicaware	Silicate Cements	Chemical Stoneware	Transite (asbestos & cement)	Chemical Porcelain	Concrete—Unbonded	Concrete—Motor Bonded	Hard Rubber (Natural)	Soft Rubber (Natural)	Neoprene	Butadiene Derivatives	Nitrile Rubber (Chemigum)
Sodium Borate $NaBO_2$		×	×			×	×	×		×	×	×	×	
Sodium Carbonate, Na_2CO_3	×	–	–	O	×	×	×	×		×	×	×	×	×
Sodium Chlorate, $NaClO_3$		×	×		×	×	×	×	×	×	×	×	×	
Sodium Chloride, NaCl	×	×		×	×		×			×	×	×		×
Sodium Cyanide, NaCN		–					×			×	×	×		
Sodium Fluoride, NaF		–					×			×	×			
Sodium Hydroxide, (Conc.), NaOH	×	–	O	O	O		O			×	×	×	×	×
Sodium Hydroxide (Dilute)	×	–	O	O	–		–			×	×	×	×	×
Sodium Hydrosulfite		×	×		×		×			×	–		×	
Sodium Hypochlorite, NaOCl	O	×			×		×			×	–	O	×	
Sodium Hyposulfate		×					×							
Sodium Nitrate, $NaNO_3$	×	×	×	×	×		×			×	–	×	×	×
Sodium Peroxide, Na_2O_2		–					×							
Sodium Phosphate, (Tri) Na_3PO_4		–	–		×	×	×			×	×	–	×	
Sodium Silicate, Na_2SiO_3		–					×	×		×	×	×	×	
Sodium Sulfate, Na_2SO_4	×	×			×		×			×	×	×		
Sodium Sulfide, Na_2S	×	–	–		×	×	×			×	–	×	×	
Sodium Sulfite, Na_2SO_3	×	×	×		×		×			×	×	O	×	
Stannic Chloride, $SnCl_4$		×					×							
Stannous Chloride, $SnCl_2$		×					×			×	×	–		
Stearic Acid $CH_3(CH_2)_{16}COOH$	×	×			×		×			–		O		
Sulfur, Molten, S	–	×				–		–						
Sulfur Chloride, (Wet), S_2Cl_2		×					×							
Sulfur Dioxide (Dry), SO_2	×	×	×		×		×			–	–	O	–	
Sulfur Dioxide (Wet)	×	×	×				×			–		O		
Sulfur Trioxide, SO_3		×			×		×			×	×	O	×	
Sulfuric Acid (Fuming to 98%)	O	×	×	×								O		O
Sulfuric Acid (Hot Conc.) H_2SO_4		×								O	O	O		

CHEMICALS	THERMOPLASTICS														
	Viton	Asphaltic, Bitumastic	Cellulose Acetate	Cellulose Acetatebutyrate	Ethyl Cellulose (Ethocel)	Cellulose Nitrate	Acrylic (Lucite, Plexiglas)	Coumarone Resins	Polyethylene	Polyvinyl Chloride, Rigid or Unplasticized	Tygon (PVC & Copolymers)	Saran (Vinyl Chloride, Vinyldene Chloride)	Kel-F (Polytrifluorochloroethylene)	Teflon (Polytetrafluoroethylene)	Uscolite CP (Styrene–Acrylonitrile–Butadiene)
Sodium Borate $NaBO_2$	×	×	×	×		×	×			×	×			×	×
Sodium Carbonate, Na_2CO_3	×	×	–	×		–	O		×	×	×	×	×	×	×
Sodium Chlorate, $NaClO_3$	×	×	×	×		×	×		×	×		×		×	×
Sodium Chloride, NaCl	×	×	×	×			×		×	×	×	×	×	×	×
Sodium Cyanide, NaCN	×	×								×	×	–		×	×
Sodium Fluoride, NaF	×									×		×		×	×
Sodium Hydroxide, (Conc.), NaOH	–	×	O	–	×	O	×		×	×	–	O	×	×	×
Sodium Hydroxide (Dilute)		×	–	×	×	–	×		×	×	×	–	×	×	×
Sodium Hydrosulfite														×	
Sodium Hypochlorite, NaOCl	–	–	–			×				×	–	×		×	×
Sodium Hyposulfate														×	
Sodium Nitrate, $NaNO_3$	×	×	×	×		×				×	×	×		×	×
Sodium Peroxide, Na_2O_2										×				×	×
Sodium Phosphate, (Tri) Na_3PO_4		×			×		×			×	×			×	×
Sodium Silicate, Na_2SiO_3	×	×								×				×	×
Sodium Sulfate, Na_2SO_4	×	×							×	×	×	×		×	×
Sodium Sulfide, Na_2S	×	×							×	×	×	×		×	×
Sodium Sulfite, Na_2SO_3	×	×								×	×	×		×	×
Stannic Chloride, $SnCl_4$	×	O								×	×		×	×	×
Stannous Chloride, $SnCl_2$		×							×	–		×		×	×
Stearic Acid $CH_3(CH_2)_{16}COOH$	×	O							×	×	×	×		×	×
Sulfur, Molten, S										O	O			×	
Sulfur Chloride, (Wet), S_2Cl_2														×	
Sulfur Dioxide (Dry), SO_2		×			–					×	–	–		×	
Sulfur Dioxide (Wet)	×									–		–		×	
Sulfur Trioxide, SO_3	×	O			O					×				×	
Sulfuric Acid (Fuming to 98%)		×	O	O	O	O	O			O		–		×	O
Sulfuric Acid (Hot Conc.) H_2SO_4	O													×	O

CHEMICALS	THERMOSETTING PLASTICS																				WOODS				
	Penton (Chlorinated Polyether)	Shellac Compounds	Organic Polysulfides	Polystyrene (Styron)	Vinylidene Chlorides	Vinyl Chloride Acetates	Cast Phenol Formaldehyde	Haveg 41 (Phenolic w. Asbestos)	Heresite (Phenol Formaldehyde)	Molded Phenolformald (Durez)	Phenol Furfural Plastics	Urea Formaldehyde	Casein Plastics	Epoxy Resins	Furane Resins (Haveg 61, Duralon)	Silicone Resins	Permanite (Furan, Glass Fiber)	Nylon (Adipic Acid–Hexameth, Diamine)	Durcon 6 (Modified Epoxy)	Cypress	Fir	Maple	Oak	Pine	Redwood
Sodium Borate $NaBO_2$		×		×	×	×	–			–	–					×	×**								
Sodium Carbonate, Na_2CO_3	×			×			×	×	×	×	×		O	×	×	×	×	×**		+	+	+	×	+	+
Sodium Chlorate, $NaClO_3$	×	×	×	×	×	×	×			×	×	×	×			O	×	×			×	×	×	×	
Sodium Chloride, NaCl	×			×				×	×	×	×			×	×	×	×	×	×	×	×	×	×	+	+
Sodium Cyanide, NaCN	×													×		–	×								
Sodium Fluoride, NaF	×															×	×								
Sodium Hydroxide, (Conc.), NaOH	×			×			O	O	O	O	O	O		×	×	–	×	×	×	–	–	O	O	O	
Sodium Hydroxide (Dilute)	×		×	×	×	×	–	O	–	O	O	×		×	×	–	×	×	–		–	O	O	–	–
Sodium Hydrosulfite							×			×	×					×									
Sodium Hypochlorite, NaOCl	×			×	×	×	–	O	×	–	–			O	O	O	O								
Sodium Hyposulfate																									
Sodium Nitrate, $NaNO_3$	×			×			×	×	×	×	×			×	×	O	×								
Sodium Peroxide, Na_2O_2																×									
Sodium Phosphate, (Tri) Na_3PO_4				–			–	–		×	–			×	×	×	×	×							
Sodium Silicate, Na_2SiO_3				×			×			×	×					×	×								
Sodium Sulfate, Na_2SO_4	×							×	×					×	×	×	×	×							
Sodium Sulfide, Na_2S				×				×	×				–	×		×	+**	–			–	O	O	–	–
Sodium Sulfite, Na_2SO_3	×						–	×	×	–	–			×	×	×	×	×							
Stannic Chloride, $SnCl_4$	×							×								×	O	×							
Stannous Chloride, $SnCl_2$	×													×		×									
Stearic Acid $CH_3(CH_2)_{16}COOH$	×							×	×					×		×	×								
Sulfur, Molten, S	O							×	O							×									
Sulfur Chloride, (Wet), S_2Cl_2																×									
Sulfur Dioxide (Dry), SO_2	×				×	×		×						×		–	×								
Sulfur Dioxide (Wet)	×															O	×								
Sulfur Trioxide, SO_3					×	×								×		×	×								
Sulfuric Acid (Fuming to 98%)	O			O			O			O	O			O	O	O	×	O			O	O	O	O	O
Sulfuric Acid (Hot Conc.) H_2SO_4	–							O						O	O	×	O				O	O	O	O	O

---

*Note: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.

**Durcon 5 would be the preferred formula.

TABLE A.3 Continued

Chemical Resistance of Materials

MATERIALS
× — *Very Good Service*
+ — *Moderate Service*
– — *Limited or Variable Service*
O — *Unsatisfactory*
Blank — No Information

CHEMICALS
Solids Assumed in Solution
Room Temperatures Assumed
Unless Otherwise Stated

	METALS														
	Carbon Steel; Fe	*Cast Iron & Ductile Iron; Fe*	*304 Stainless Steel; Fe, 18Cr, 8Ni*	*316 Stainless Steel; Fe 16Cr, 10Ni, 2Mo*	*347 Stainless Steel; Fe, 17Cr, 9Ni; (Cx10)Cb*	*Ni-Resist Iron; Fe 14Ni, 2Cr, 2Si*	*Durimet 20; Carpenter 20; Fe; 4Cu, 20Cr, 29Ni, 2Mo, 1Si*	*Worthite; 3Mo, 2Cu, Fe, 20Cr, 24Ni, 3Si*	*Durorpm; Fe, 14Si; Durichlor, Fe, 14Si, 3Mo**	*Copper; Brass; Bronzes; Everdur*	*Aluminum; Al (and Alloys)*	*Lead; Pb*	*Monel; 67Ni, 30Cu, 1.4Fe*	*Nickel; Ni*	*Inconel; 76Ni, 15Cr, 8Fe*
Sulfuric Acid (Cold Conc.)	+	+	×	×	–		×	×	×	O	–	×	+		
Sulfuric Acid (75%–95%)	+	–	O	O	–	–	×	×	×	O	O	×	O	O	–
Sulfuric Acid (10%–75%)	O	O	O	O	–	–	×	×	×	O	O	×	–	O	–
Sulfuric Acid (<10%)	O	O	O	–	–	–	–	×	×	O	–	×	–	–	–
Sulfurous Acid, H_2SO_3	O	O	–	×	–	O	×	×	O	O	–	–	O	O	O
Sulfuryl Chloride, SO_2Cl_2							+		×						
Tannic Acid	–		×	×			×	×	×	–	×		+	+	+
Tartaric Acid, (CHOH COOH)$_2$	–		×	×			×	×	×	–	O	×	+	+	
Toluene, $CH_3C_6H_5$	×		×				×		×	×	×	×	×		
Trichlorethylene, Dry, Cl_2CCHCl	–	–	–	×	–	–	×	×	×	–	–		×	×	×
Water, Fresh, H_2O	×		×	×			×		×	×	×		×	×	
Water, Distilled Lab.	O		×	×		O	×		×	×	O		O	–	
Zinc Chloride, $ZnCl_2$	O	O	O	–	O	–	×	×	×	O	O	–	+	+	
Zinc Sulfate, $ZnSO_4$	O	O	+	+	+	×	×	×	×	+	–	+	+	+	+

	METALS												
	Hastelloy B; Ni, 26Mo, 4Fe	*Hastelloy C; Ni, 16Mo, 4Fe,14Cr, 4W*	*Hastelloy D; Ni, 8Si, 3Cu*	*Chlorimet 3; 3Fe, 1Si, 60Ni,18Mo, 18Cr*	*Chlorimet 2; 63Ni, 32Mo, 3Fe, 1Si*	*Stellite; Co, 28Cr, 4W*	*Zirconium; Z*	*Tantalum; Ta*	*Silver; Ag*	*Platinum; Pt*	*Dowmetal; (Mg alloys)*	*Titanium; Ti*	*Molybdenum; Mo*
Sulfuric Acid (Cold Conc.)	×	×	×	×	×	–	O	–		×		O	–
Sulfuric Acid (75%–95%)	×	×	–	×	×	–	–	–	O	×		O	
Sulfuric Acid (10%–75%)	×	×	×	×	×	–	×	–	×	×		–	×
Sulfuric Acid (<10%)	×	×	×	×	×	–	×	×	×	×		×	
Sulfurous Acid, H_2SO_3	O	+	+	×	×	×	×	×	×			×	
Sulfuryl Chloride, SO_2Cl_2				+	+			×					
Tannic Acid	×	×	×	×	×		×	×	×			×	
Tartaric Acid, (CHOH COOH)$_2$	×	×	×	×	×		×	×	×			×	+
Toluene, $CH_3C_6H_5$				×	×		×						
Trichlorethylene, Dry, Cl_2CCHCl	×	×	×	×	×	×		×			×	×	
Water, Fresh, H_2O				×	×			×	×				×
Water, Distilled Lab.				×	×				×				
Zinc Chloride, $ZnCl_2$	+	+	+	×	×		×	×	×			×	
Zinc Sulfate, $ZnSO_4$	+	+	+	×	×			×	×			×	

	CARBONS & CERAMICS									RUBBERS				
	Carbon & Graphite	*Glass, "Pyrex" brand*	*Silicaware*	*Silicate Cements*	*Chemical Stoneware*	*Transite (asbestos & cement)*	*Chemical Porcelain*	*Concrete—Unbonded*	*Concrete—Motor Bonded*	*Hard Rubber (Natural)*	*Soft Rubber (Natural)*	*Neoprene*	*Butadiene Derivatives*	*Nitrile Rubber (Chemigum)*
Sulfuric Acid (Cold Conc.)	O	×					×			O	O	O		
Sulfuric Acid (75%–95%)	×	×			×		×					O		O
Sulfuric Acid (10%–75%)	×	×	×		×		×			×	×	–	×	–
Sulfuric Acid (<10%)	×	×	×		×		×			×	×	×	×	×
Sulfurous Acid, H_2SO_3	×	×	×	×	×		×			–	–	×		
Sulfuryl Chloride, SO_2Cl_2		×	×		×		×							
Tannic Acid	×	×			×		×	×		×	–	×		
Tartaric Acid, (CHOH COOH)$_2$	×	×	×		×		×			×	×	×		
Toluene, $CH_3C_6H_5$	×	×	×		×		×			O	O	O	O	
Trichlorethylene, Dry, Cl_2CCHCl	×	×			×		×			O	O	O		
Water, Fresh, H_2O	×	×	×		×	×	×	×	×	×	×	×	×	
Water, Distilled Lab.		+					×			×		×		
Zinc Chloride, $ZnCl_2$	×	×		×			×	×		×	×	×	×	×
Zinc Sulfate, $ZnSO_4$	×	×					×			×	×	×		

	THERMOPLASTICS														
	Viton	*Asphaltic, Bitumastic*	*Cellulose Acetate*	*Cellulose Acetatebutyrate*	*Ethyl Cellulose (Ethocel)*	*Cellulose Nitrate*	*Acrylic (Lucite, Plexiglas)*	*Coumarone Resins*	*Polyethylene*	*Polyvinyl Chloride, Rigid or Unplasticized*	*Tygon (PVC & Copolymers)*	*Saran (Vinyl Chloride, Vinyldene Chloride)*	*Kel-F (Polytrifluorochloroethylene)*	*Teflon (Polytetrafluoroethylene)*	*Uscolite CP (Styrene–Acrylonitrile–Butadiene)*
Sulfuric Acid (Cold Conc.)	+						–			×	O	–		×	O
Sulfuric Acid (75%–95%)		×							×	×	O	–		×	O
Sulfuric Acid (10%–75%)	×	×	×	×	O	O	×		×	×	–	×	×	×	–
Sulfuric Acid (<10%)	×	×	×	×	–	–	×		×	×	×	×	–	×	×
Sulfurous Acid, H_2SO_3		×							O	×	–	–		×	×
Sulfuryl Chloride, SO_2Cl_2									O					×	
Tannic Acid		×			–				×	×		×		×	×
Tartaric Acid, (CHOH COOH)$_2$	×	×	×	×						×		×		×	×
Toluene, $CH_3C_6H_5$	–		–	–	O		O		–	O	O	–	–	×	
Trichlorethylene, Dry, Cl_2CCHCl	–	O					O		O	O	O	–	O	×	
Water, Fresh, H_2O	×	×	×	×	×	×	×	×		×		×		×	×
Water, Distilled Lab.							×			×				×	×
Zinc Chloride, $ZnCl_2$	×	×			×				×	×		×		×	×
Zinc Sulfate, $ZnSO_4$	×	×							×	×	×	×		×	×

	THERMOSETTING PLASTICS																			
	Penton (Chlorinated Polyether)	*Shellac Compounds*	*Organic Polysulfides*	*Polystyrene (Styron)*	*Vinylidene Chlorides*	*Vinyl Chloride Acetates*	*Cast Phenol Formaldehyde*	*Haveg 41 (Phenolic w. Asbestos)*	*Heresite (Phenol Formaldehyde)*	*Molded Phenolformald (Durez)*	*Phenol Furfural Plastics*	*Urea Formaldehyde*	*Casein Plastics*	*Epoxy Resins*	*Furane Resins (Haveg 61, Duralon)*	*Silicone Resins*	*Permanite (Furan, Glass Fiber)*	*Nylon (Adipic Acid–Hexameth, Diamine)*	*Durcon 6 (Modified Epoxy)*	*Cypress*
Sulfuric Acid (Cold Conc.)	–							–						O		O	O	×	O	
Sulfuric Acid (75%–95%)	+			–				–		–	–			–	–	–	×			
Sulfuric Acid (10%–75%)	×	×	×	×	×	×	–	×		–	–		O	×	×	–	–	×	–	
Sulfuric Acid (<10%)	×	×	×	×	×	×	–	×		×	×	×	–		×	–	×	–	×	–
Sulfurous Acid, H_2SO_3	×			×			×	×		×	×		×	×	×	–	×	O	+**	
Sulfuryl Chloride, SO_2Cl_2										×						×				
Tannic Acid	×			×			×	×		×	×	×	×	×	×	×	×			
Tartaric Acid, (CHOH COOH)$_2$	×						–	×		–	–	×		×	×	×	×	×		
Toluene, $CH_3C_6H_5$	+	O	O	O	–	–	–	–		–	–		×		×	×	×	×	–	
Trichlorethylene, Dry, Cl_2CCHCl	+							×						×		×	×			
Water, Fresh, H_2O	×	×	×	×	×	×	×			×	×	×	O			×	×			
Water, Distilled Lab.	×															×				
Zinc Chloride, $ZnCl_2$	×			×			–	×		×	–			×	×	×	–	×		
Zinc Sulfate, $ZnSO_4$	×			×				×		×	×			×	×	×	×	×		

	WOODS				
	Fir	*Maple*	*Oak*	*Pine*	*Redwood*
Sulfuric Acid (Cold Conc.)	O	O	O	O	O
Sulfuric Acid (75%–95%)					
Sulfuric Acid (10%–75%)	–	–	–	–	–
Sulfuric Acid (<10%)	–	×	–	–	–
Sulfurous Acid, H_2SO_3	×	×			
Sulfuryl Chloride, SO_2Cl_2					
Tannic Acid	×	×			
Tartaric Acid, (CHOH COOH)$_2$					
Toluene, $CH_3C_6H_5$	–	×	×	×	
Trichlorethylene, Dry, Cl_2CCHCl					
Water, Fresh, H_2O	×	×	×	×	×
Water, Distilled Lab.					
Zinc Chloride, $ZnCl_2$					
Zinc Sulfate, $ZnSO_4$					

**Note*: Duriron is as shown. Durichlor is also satisfactory on chlorides and HCl.

**Durcon 5 would be the preferred formula.

A.4 Composition of Metallic and Other Materials

B. G. LIPTÁK

TABLE A.4
Composition of Metallic and Other Materials

No.	Material	Manufacturer	Composition or Description
		Metals	
17	Aluminum		
19	Aloyco-20	Alloy Steel Products Co.	Fe; 19–21 Cr; 28–30 Ni; 4.0–4.5 Cu; 2.5–3.0 Mo; 1.5 max. Si; 0.65–0.85 Mn; 0.07 max. C
19a	720 Alloy	General Plate	20 Mn; 20 Ni; Cu
54–60	Brass		Various commercial grades ranging 60–65 Cu; 35–40 Zn; 0.5–3.0 Pb
63	Brass, red		85 Cu; 15 Zn
66	Bronze, comm.		90 Cu; 10 Zn
73	Bronze, phosphor, 5% A		94.8–95.5 Cu; 4.3–5 Sn; P
74	Bronze, phosphor, 8% C		Cu; 7–9 Sn; 0.03–0.25 P
75	Bronze, phosphor 10% D		89.5–90 Cu; 10–10.5 Sn; P
76	Bronze, phosphor, spec. free cutting		88 Cu; 4 Zn; 4 Sn; 4 Pb
81	CA-FA20	Cooper Alloy	Fe; 19–21 Cr; 28–30 Ni; 3.5 Mo; 4–4.5 Cu; 0.07 max. C
82	CA-MM	Cooper Alloy	67 Ni; 30 Cu; 1.4 Fe; 0.1 Si; 0.15 C
86	Cast iron		Ordinary unalloyed cast iron
88	Chlorimet 2	Duriron Co.	63 Ni; 32 Mo; 3 max. Fe; 0.15 max. C; 1 Si; 1 Mn
89	Chlorimet 3	Duriron Co.	60 Ni; 18 Mo; 18 Cr; 2 Fe; 0.07 max. C; 1 Si; 1 Mn
111	Copper		99.9+ Cu
112	Copper, Be		97.5 Cu; 2.15 Be; 0.35 Ni
119	Corrosiron	Pacific Fdry.	Fe; 14.5 Si
140	Durichlor	Duriron Co.	Fe; 0.85 C; 14.5 Si; 3 Mo; 0.35 Mn
141	Durimet 20	Duriron Co.	Fe; 20 Cr; 29 Ni; 0.07 max. C; 2 Mo; 4 Cu; 1 Si
142	Durimet T	Duriron Co.	Fe; 19 Cr; 22 Ni; 0.07 max. C; 2 Mo; 1 Cu; 1 Si
143	Duriron	Duriron Co.	Fe; 0.80 C; 14.5 Si; 0.35 Mn
148	Everdur 1000	Amer. Brass	94.9 Cu; 4 Si; 1.1 Mn
149	Everdur 1010	Amer. Brass	95.8 Cu; 3.1 Si; 1.1 Mn
150	Everdur 1015	Amer. Brass	98.25 Cu; 1.5 Si; 0.25 Mn
156	Gold		99.99 Au
156a	Green gold		75% Au; 25% Ag
159	Hastelloy A	Haynes Stellite	Ni; 17–21 Mo; 17–21 Fe
160	Hastelloy B	Haynes Stellite	Ni; 24–32 Mo; 3–7 Fe; 0.02–0.12 C
161	Hastelloy C	Haynes Stellite	Ni; 14–19 Mo; 4–8 Fe; 0.04–0.15 C; 12–16 Cr; 3–5.5 W
162	Hastelloy D	Haynes Stellite	Ni; 8–11 Si; 2–5 Cu; 1 max. Al
163	Stellite 1	Haynes Stellite	Co; 28–34 Cr; 11–15 W

TABLE A.4 Continued
Composition of Metallic and Other Materials

No.	Material	Manufacturer	Composition or Description
165	Stellite 6	Haynes Stellite	Co; 25–31 Cr; 3–6 W
184	Inconel	Int'l Nickel	79.5 Ni; 13 Cr; 6.5 Fe; 0.08 C; 0.2 Cu; 0.25 Mn
191	Lead		99.9 + Pb
192	Lead, antimonial		94 Pb; 6 Sb
193	Lead, antimonial		Pb; 4–12 Sb
196	Lead, chemical		99.93 Pb; 0.06 Cu
200	Lead, Te		99.88 Pb; 0.045 Te; 0.06 Cu
216	Monel	Int'l Nickel	67 Ni; 30 Cu; 1.4 Fe; 0.1 Si; 0.15 C
219	Muntz Metal		60 Cu; 40 Zn
224	Nickel	Int'l Nickel	99.4 Ni; 0.2 Mn; 0.1 Cu; 0.15 Fe; 0.05 Si
224a	Z-Nickel	Int'l Nickel	95 + Ni
226	Nickel–Silver 18% A		65 Cu; 18 Ni; 17 Zn
227	Nickel–Silver 18% B		55 Cu; 18 Ni; 27 Zn
227a	Ni-Span	Int'l Nickel	Ni, Ti, Cr, C, Mn, Si, Al
229	Ni-Hard	Int'l Nickel	Fe; 3.4 C; 1.5 Cr; 4.5 Ni; 0.6 Si
231	Ni-Resist	Int'l Nickel	Fe; 2.8 C; 14 or 20 Ni; 6 Cu (optional); 2 Cr; 2 Si
240	Platinum		99.99 Pt
268	Silver		99.9+ Ag
275	S.S. 301		Fe; 16–18 Cr; 6–8 Ni; 0.08–0.15 C
276	S.S. 302		Fe; 17–19 Cr; 8–10 Ni; 0.08–0.15 C
278	S.S. 303		Fe; 17–19 Cr; 8–10 Ni; 0.15 max. C; 0.07 min. P, S, Se; 0.6
279	S.S. 304		Fe; 18–20 Cr; 8–11 Ni; 0.08 max. C; 2 max. Mn
282	S.S. 310		Fe; 24–26 Cr; 19–22 Ni; 0.25 max. C
283	S.S. 316		Fe; 16–18 Cr; 10–14 Ni; 0.1 max. C; 1.75–2.75 Mo
284	S.S. 317		Fe; 17.5–20 Cr; 10–14 Ni; 0.1 max. C; 3–4 Mo
285	S.S. 321		Fe; 17–19 Cr; 8–11 Ni; Ti, 5xC min.
286	S.S. 347		Fe; 17–19 Cr; 9–12 Ni; Cb, 10xC min.
287	S.S. 403		Fe; 11.5–13 Cr; 0.15 max. C
290	S.S. 410		Fe; 11.5–13.5 Cr; 0.15 max. C
292	S.S. 416		Fe; 12–14 Cr; 0.15 max. C; 0.07 min. P, S, Se; 0.6 max. Zr, Mo
295	S.S. 430		Fe; 14–18 Cr; 0.12 max. C
303	S.S. 446		Fe; 23–27 Cr; 0.35 max. C; 0.25 max. N
360a	Steel		Plain carbon steel
368	Tantalum	Fansteel	99.9+ Ta
390	Worthite	Worthington Pump	Fe; 20 Cr; 24 Ni; 0.07 max. C; 3.25 Si; 3 Mo; 1.75 Cu; 0.5 Mn
		Carbon and Graphite	
401	Karbate (carbon)	National Carbon	Impervious carbon
402	Karbate (graphite)	National Carbon	Impervious graphite
		Ceramics	
611	Lapp Porcelain	Lapp Insulator Co.	Chemical porcelain
614	Pfaudler Glass Lining	Pfaudler Co.	Glass-lined steel equipment
615	Plate Glass		Polished plate glass, flat or bent
616	Pyrex	Corning Glass Wks.	Glass

TABLE A.4 Continued

Composition of Metallic and Other Materials

No.	Material	Manufacturer	Composition or Description
		Plastics	
700	Ace Saran	American Hard Rubber	Vinylidene chloride
710	Geon	B. F. Goodrich	Polyvinyl chloride
711	Haveg 41	Haveg Corp.	Phenolic-asbestos
712	Haveg 43	Haveg Corp.	Phenolic-graphite
713	Haveg 60	Haveg Corp.	Furan-asbestos
714	Haveg 63	Haveg Corp.	Furan-graphite
715	Heresite M 66	Heresite & Chem. Co.	Transparent molding powder
716	Heresite MF 66	Heresite & Chem. Co.	Black molding powder
717a	Kel-F	M. W. Kellogg	Polymerized trifluoroethylene
718	Koroseal	B. F. Goodrich	Plasticized polyvinyl chloride
731	Nylon FM-101	E. I. du Pont	Injection, compression and extrusion moldings (tubing, sheeting, wire covering, gasketing)
731a	Plastisol		Polyvinyl chloride
735	Polythene	E.I. du Pont	Polyethylene
740	Saran	Dow Chemical	Vinyl chloride-vinylidene chloride copolymer
740a	Sirvene	Chicago Rawhide	Synthetic rubber
742	Teflon	E.I. du Pont	Polymerized tetrafluoroethylene
746	Tygon	U.S. Stoneware	Synthetic compounds
		Rubber	
800	Ace Hard Rubber	American Hard Rubber	Vulcanized rubber
805	Butyl (GR-I)	Stanco Distributors	Solid copolymer of isobutylene and isoprene
820	Hycar (GR-A)	B. F. Goodrich	Nitrile type synthetic rubber
829	Neoprene	E. I. du Pont	Polymer of chloroprene
836	Natural (soft)		
837	Natural (hard)		
838	GR-S (soft)		
839	GR-S (hard)		
853	Thiokol (GR-P)	Thiokol Corp.	

A.5 Steam and Water Tables

B. G. LIPTÁK

TABLE A.5a
Dry Saturated Steam: Temperature Table

		Specific Volume, ft^3/lbm^a			Enthalpy, Btu/lbm^a			Entropy, Btu/lbm R^a		
Temp., °F/°C	Abs. Press., PSIA P^a	Sat. Liquid v_f	Evap. v_{fg}	Sat. Vapor v_g	Sat. Liquid h_f	Evap. h_{fg}	Sat. Vapor h_g	Sat. Liquid s_f	Evap. s_{fg}	Sat. Vapor s_g
32/0	0.08854	0.01602	3306	3306	0.00	1075.8	1075.8	0.0000	2.1877	2.1877
35/1.7	0.09995	0.01602	2947	2947	3.02	1074.1	1077.1	0.0061	2.1709	2.1770
40/4.4	0.12170	0.01602	2444	2444	8.05	1071.3	1079.3	0.0162	2.1435	2.1597
45/7.2	0.14752	0.01602	2036.4	2036.4	13.06	1068.4	1081.5	0.0262	2.1167	2.1429
50/10	0.17811	0.01603	1703.2	1703.2	18.07	1065.6	1083.7	0.0361	2.0903	2.1264
60/15.6	0.2563	0.01604	1206.6	1206.7	28.06	1059.9	1088.0	0.0555	2.0393	2.0948
70/21.1	0.3631	0.01606	867.8	867.9	38.04	1054.3	1092.3	0.0745	1.9902	2.0647
80/26.7	0.5069	0.01608	633.1	633.1	48.02	1048.6	1096.6	0.0932	1.9428	2.0360
90/32.2	0.6982	0.01610	468.0	468.0	57.99	1042.9	1100.9	0.1115	1.8972	2.0087
100/37.8	0.9492	0.01613	350.3	350.4	67.97	1037.2	1105.2	0.1295	1.8531	1.9826
110/43	1.2748	0.01617	265.3	265.4	77.94	1031.6	1109.5	0.1471	1.8106	1.9577
120/49	1.6924	0.01620	203.25	203.27	87.92	1025.8	1113.7	0.1645	1.7694	1.9339
130/54	2.2225	0.01625	157.32	157.34	97.90	1020.0	1117.9	0.1816	1.7296	1.9112
140/60	2.8886	0.01629	122.99	123.01	107.89	1014.1	1122.0	0.1984	1.6910	1.8894
150/66	3.718	0.01634	97.06	97.07	117.89	1008.2	1126.1	0.2149	1.6537	1.8685
160/71	4.741	0.01639	77.27	77.29	127.89	1002.3	1130.2	0.2311	1.6174	1.8485
170/77	5.992	0.01645	62.04	62.06	137.90	996.3	1134.2	0.2472	1.5822	1.8293
180/82	7.510	0.01651	50.21	50.23	147.92	990.2	1138.1	0.2630	1.5480	1.8109
190/88	9.339	0.01657	40.94	40.96	157.95	984.1	1142.0	0.2785	1.5147	1.7932
200/93	11.526	0.01663	33.62	33.64	167.99	977.9	1145.9	0.2938	1.4824	1.7762
210/90	14.123	0.01670	27.80	27.82	178.05	971.6	1149.7	0.3090	1.4508	1.7598
212/100	14.696	0.01672	26.78	26.80	180.07	970.3	1150.4	0.3120	1.4446	1.7566
220/104	17.186	0.01677	23.13	23.15	188.13	965.2	1153.4	0.3239	1.4201	1.7440
230/110	20.780	0.01684	19.365	19.382	198.23	958.8	1157.0	0.3387	1.3901	1.7288
240/116	24.969	0.01692	16.306	16.323	208.34	952.2	1160.5	0.3531	1.3609	1.7140
250/121	29.825	0.01700	13.804	13.821	218.48	945.5	1164.0	0.3675	1.3323	1.6998
260/127	35.429	0.01709	11.746	11.763	228.64	938.7	1167.3	0.3817	1.3043	1.6860
270/132	41.858	0.01717	10.044	10.061	238.84	931.8	1170.6	0.3958	1.2769	1.6727
280/138	49.203	0.01726	8.628	8.645	249.06	924.7	1173.8	0.4096	1.2501	1.6597
290/143	57.556	0.01735	7.444	7.461	259.31	917.5	1176.8	0.4234	1.2238	1.6472
300/149	67.013	0.01745	6.449	6.466	269.59	910.1	1179.7	0.4369	1.1980	1.6350
310/154	77.68	0.01755	5.609	5.626	279.92	902.6	1182.5	0.4504	1.1727	1.6231
320/160	89.66	0.01765	4.896	4.914	290.28	894.9	1185.2	0.4637	1.1478	1.6115
330/166	103.06	0.01776	4.289	4.307	300.68	887.0	1187.7	0.4769	1.1233	1.6002
340/171	118.01	0.01787	3.770	3.788	311.13	879.0	1190.1	0.4900	1.0992	1.5891

TABLE A.5a Continued

Dry Saturated Steam: Temperature Table

		Specific Volume, ft^3/lbm^a			*Enthalpy, Btu/lbm^a*			*Entropy, Btu/lbm R^a*		
Temp., °F/°C	*Abs. Press., PSIA P^a*	*Sat. Liquid* υ_f	*Evap.* υ_{fg}	*Sat. Vapor* υ_g	*Sat. Liquid* h_f	*Evap.* h_{fg}	*Sat. Vapor* h_g	*Sat. Liquid* s_f	*Evap.* s_{fg}	*Sat. Vapor* s_g
350/177	134.63	0.01799	3.324	3.342	321.63	870.7	1192.3	0.5029	1.0754	1.5783
360/182	153.04	0.01811	2.939	2.957	332.18	862.2	1194.4	0.5158	1.0519	1.5677
370/188	173.37	0.01823	2.606	2.625	342.79	853.5	1196.3	0.5286	1.0287	1.5573
380/193	195.77	0.01836	2.317	2.335	353.45	844.6	1198.1	0.5413	1.0059	1.5471
390/199	220.37	0.01850	2.0651	2.0836	364.17	835.4	1199.6	0.5539	0.9832	1.5371
400/204	247.31	0.01864	1.8447	1.8633	374.97	826.0	1201.0	0.5664	0.9608	1.5272
410/210	276.75	0.01878	1.6512	1.6700	385.83	816.3	1202.1	0.5788	0.9386	1.5174
420/216	308.83	0.01894	1.4811	1.5000	396.77	806.3	1203.1	0.5912	0.9166	1.5078
430/221	343.72	0.01910	1.3308	1.3499	407.79	796.0	1203.8	0.6035	0.8947	1.4982
440/227	381.59	0.01926	1.1979	1.2171	418.90	785.4	1204.3	0.6158	0.8730	1.4887
450/232	422.6	0.0194	1.0799	1.0993	430.1	774.5	1204.6	0.6280	0.8513	1.4793
460/238	466.9	0.0196	0.9748	0.9944	441.4	763.2	1204.6	0.6402	0.8298	1.4700
470/243	514.7	0.0198	0.8811	0.9009	452.8	751.5	1204.3	0.6523	0.8083	1.4606
480/249	566.1	0.0200	0.7972	0.8172	464.4	739.4	1203.7	0.6645	0.7868	1.4513
490/254	621.4	0.0202	0.7221	0.7423	476.0	726.8	1202.8	0.6766	0.7653	1.4419
500/260	680.8	0.0204	0.6545	0.6749	487.8	713.9	1201.7	0.6887	0.7438	1.4325
520/271	812.4	0.0209	0.5385	0.5594	511.9	686.4	1198.2	0.7130	0.7006	1.4136
540/282	962.5	0.0215	0.4434	0.4649	536.6	656.6	1193.2	0.7374	0.6568	1.3942
560/293	1133.1	0.0221	0.3647	0.3868	562.2	624.2	1186.4	0.7621	0.6121	1.3742
580/304	1325.8	0.0228	0.2989	0.3217	588.9	588.4	1177.3	0.7872	0.5659	1.3532
600/316	1542.9	0.0236	0.2432	0.2668	617.0	548.5	1165.5	0.8131	0.5176	1.3307
620/327	1786.6	0.0247	0.1955	0.2201	646.7	503.6	1150.3	0.8398	0.4664	1.3062
640/338	2059.7	0.0260	0.1538	0.1798	678.6	452.0	1130.5	0.8679	0.4110	1.2789
660/349	2365.4	0.0278	0.1165	0.1442	714.2	390.2	1104.4	0.8987	0.3485	1.2472
680/360	2708.1	0.0305	0.0810	0.1115	757.3	309.9	1067.2	0.9351	0.2719	1.2071
700/371	3093.7	0.0369	0.0392	0.0761	823.3	172.1	995.4	0.9905	0.1484	1.1389
705.4/374.1	3206.2	0.0503	0	0.0503	902.7	0	902.7	1.0580	0	1.0580

Source: Abridged from *Thermodynamic Properties of Steam*, by Joseph H. Keenan and Frederick G. Keyes. © 1936, by Joseph H. Keenan and Frederick G. Keyes. Published by John Wiley & Sons, Inc., New York.

[a] PSIA = 0.069 bar (abs); ft^3/lbm = 62.4 l/kg; Btu/lbm = 0.556 Kcal/kg

TABLE A.5b

Properties of Superheated Steam

Abs. Press., PSIA[a] *(Sat. Temp. °F)*	*Temperature, °F/°C*												
	200/93	*220/104*	*300/149*	*350/177*	*400/204*	*450/232*	*500/260*	*550/288*	*600/316*	*700/371*	*800/427*	*900/482*	*1000/538*
υ	392.6	404.5	452.3	482.2	512.0	541.8	571.6	601.4	631.2	690.8	750.4	809.9	869.5
1 *h*	1150.4	1159.5	1195.8	1218.7	1241.7	1264.9	1288.3	1312.0	1335.7	1383.8	1432.8	1482.7	1533.5
(101.74) *s*	2.0512	2.0647	2.1153	2.1444	2.1720	2.1983	2.2233	2.2468	2.2702	2.3137	2.3542	2.3923	2.4283
υ	78.16	80.59	90.25	96.26	102.26	108.24	114.22	120.19	126.16	138.10	150.03	161.95	173.87
5 *h*	1148.8	1158.1	1195.0	1218.1	1241.2	1264.5	1288.0	1311.7	1335.4	1383.6	1432.7	1482.6	1533.4
(162.24) *s*	1.8718	1.8857	1.9370	1.9664	1.9942	2.0205	2.0456	2.0692	2.0927	2.1361	2.1767	2.2148	2.2509
υ	38.85	40.09	45.00	48.03	51.04	54.05	57.05	60.04	63.03	69.01	74.98	80.95	86.92
10 *h*	1146.6	1156.2	1193.9	1217.2	1240.6	1264.0	1287.5	1311.3	1335.1	1383.4	1432.5	1482.4	1533.1
(193.21) *s*	1.7927	1.8071	1.8595	1.8892	1.9172	1.9436	1.9689	1.9924	2.0160	2.0596	2.1002	2.1383	2.1744
υ		27.15	30.53	32.62	34.68	36.73	38.78	40.82	42.86	46.94	51.00	55.07	59.13
14.696 *h*		1154.4	1192.8	1216.4	1239.9	1263.5	1287.1	1310.9	1335.8	1383.2	1432.3	1482.3	1533.1
(212.00) *s*		1.7624	1.8160	1.8460	1.8743	1.9008	1.9261	1.9498	1.9734	2.0170	2.0576	2.0958	2.1319
υ			22.36	23.91	25.43	26.95	28.46	29.97	31.47	34.47	37.46	40.45	43.44
20 *h*			1191.6	1215.6	1239.2	1262.9	1286.6	1310.5	1334.4	1382.9	1432.1	1482.1	1533.0
(227.96) *s*			1.7808	1.8112	1.8396	1.8664	1.8918	1.9160	1.9392	1.9829	2.0235	2.0618	2.0978
υ			11.040	11.843	12.628	13.401	14.168	14.93	15.688	17.198	18.702	20.20	21.70
40 *h*			1186.8	1211.9	1236.5	1260.7	1284.8	1308.9	1333.1	1381.9	1431.3	1481.4	1532.4
(267.25) *s*			1.6994	1.7314	1.7608	1.7881	1.8140	1.8384	1.8619	1.9058	1.9467	1.9850	2.0214
υ			7.259	7.818	8.357	8.884	9.403	9.916	10.427	11.441	12.449	13.452	14.454
60 *h*			1181.6	1208.2	1233.6	1258.5	1283.0	1307.4	1331.8	1380.9	1430.5	1480.8	1531.9
(292.71) *s*			1.6492	1.6830	1.7135	1.7416	1.7678	1.7926	1.8162	1.8605	1.9015	1.9400	1.9762
υ				5.803	6.220	6.624	7.020	7.410	7.797	8.562	9.322	10.077	10.830
80 *h*				1204.3	1230.7	1256.1	1281.1	1305.8	1330.5	1379.9	1429.7	1480.1	1531.3
(312.03) *s*				1.6475	1.6791	1.7078	1.7346	1.7598	1.7836	1.8281	1.8694	1.9079	1.9442
υ				4.592	4.937	5.268	5.589	5.905	6.218	6.835	7.446	8.052	8.656
100 *h*				1200.1	1227.6	1253.7	1279.1	1304.2	1329.1	1378.9	1428.9	1479.5	1530.8
(327.81) *s*				1.6188	1.6518	1.6813	1.7085	1.7339	1.7581	1.8029	1.8443	1.8829	1.9193
υ				3.783	4.081	4.363	4.636	4.902	5.165	5.683	6.195	6.702	7.207
120 *h*				1195.7	1224.4	1251.3	1277.2	1302.5	1327.7	1377.8	1428.1	1478.8	1530.2
(341.25) *s*				1.5944	1.6287	1.6591	1.6869	1.7127	1.7370	1.7822	1.8237	1.8625	1.8990
υ					3.468	3.715	3.954	4.186	4.413	4.861	5.301	5.738	6.172
140 *h*					1221.1	1248.7	1275.2	1300.9	1326.4	1376.8	1427.3	1478.2	1529.7
(353.02) *s*					1.6087	1.6399	1.6683	1.6945	1.7190	1.7645	1.8063	1.8451	1.8817

υ – specific volume (ft^3/lbm); *h* – enthalpy (BTU/lbm); *s* – entropy (BTU/lbm °F)

TABLE A.5b Continued

Properties of Superheated Steam

Abs. Press., PSIA[a] *(Sat. Temp. °F)*	*Temperature, °F/°C*												
	200/93	*220/104*	*300/149*	*350/177*	*400/204*	*450/232*	*500/260*	*550/288*	*600/316*	*700/371*	*800/427*	*900/482*	*1000/538*
υ					3.008	3.230	3.443	3.648	3.849	4.244	4.631	5.015	5.396
160 *h*					1217.6	1246.1	1273.1	1299.3	1325.0	1375.7	1426.4	1477.5	1529.1
(363.53) *s*					1.5908	1.6230	1.6519	1.6785	1.7033	1.7491	1.7911	1.8301	1.8667
υ					2.649	2.852	3.044	3.229	3.411	3.764	4.110	4.452	4.792
180 *h*					1214.0	1243 5	1271.0	1297.6	1323.5	1374.7	1425.6	1476.8	1528.6
(373.06) *s*					1.5745	1.6077	1.6373	1.6642	1.6894	1.7355	1.7776	1.8167	1.8534
υ					2.361	2.549	2.726	2.895	3.060	3.380	3.693	4.002	4.309
200 *h*					1210.3	1240.7	1268.9	1295.8	1322.1	1373.6	1424.8	1476.2	1528.0
(381.79) *s*					1.5594	1.5937	1.6240	1.6513	1.6767	1.7232	1.7655	1.8048	1.8415
υ					2.125	2.301	2.465	2.621	2.772	3.066	3.352	3.634	3.913
220 *h*					1206.5	1237.9	1266.7	1294.1	1320.7	1372.6	1424.0	1475.5	1527.5
(389.86) *s*					1.5453	1.5808	1.6117	1.6395	1.6652	1.7120	1.7545	1.7939	1.8308
υ					1.9276	2.094	2.247	2.393	2.533	2.804	3.068	3.327	3.584
240 *h*					1202.5	1234.9	1264.5	1292.4	1319.2	1371.5	1423.2	1474.8	1526.9
(397.37) *s*					1.5319	1.5686	1.6003	1.6286	1.6546	1.7017	1.7444	1.7839	1.8209
υ						1.9183	2.063	2.199	2.330	2.582	2.827	3.067	3.305
260 *h*						1232.0	1262.3	1290.5	1317.7	1370.4	1422.3	1474.2	1526.3
(404.42) *s*						1.5573	1.5897	1.6184	1.6447	1.6922	1.7352	1.7748	1.8118
υ						1.7674	1.9047	2.033	2.156	2.392	2.621	2.845	3.066
280 *h*						1228.9	1260.0	1288.7	1316.2	1369.4	1421.5	1473.5	1525.8
(411.05) *s*						1.5464	1.5796	1.6087	1.6354	1.6834	1.7265	1.7662	1.8033
υ						1.6364	1.7675	1.8891	2.005	2.227	2.442	2.652	2.859
300 *h*						1225.8	1257.6	1286.8	1314.7	1368.3	1420.6	1427.8	1525.2
(417.33) *s*						1.5360	1.5701	1.5998	1.6268	1.6751	1.7184	1.7582	1.7954
υ						1.3734	1.4923	1.6010	1.7036	1.8980	2.084	2.266	2.445
350 *h*						1217.7	1251.5	1282.1	1310.9	1365.5	1418.5	1471.1	1523.8
(431.72) *s*						1.5119	1.5481	1.5792	1.6070	1.6563	1.7002	1.7403	1.7777
υ						1.1744	1.2851	1.3843	1.4770	1.6508	1.8161	1.9767	2.134
400 *h*						1208.8	1245.1	1277.2	1306.9	1362.7	1416.4	1.469.4	1522.4
(444.59) *s*						1.4892	1.5281	1.5607	1.5894	1.6398	1.6842	1.7247	1.7623

υ – specific volume (ft^3/lbm); *h* – enthalpy (BTU/lbm); *s* – entropy (BTU/lbm °F)

TABLE A.5b Continued
Properties of Superheated Steam

Abs. Press., PSIA[a] *(Sat. Temp. °F)*	*Temperature, °F/°C*													
	500/260	*550/288*	*600/316*	*620/327*	*640/338*	*660/349*	*680/360*	*700/371*	*800/427*	*900/482*	*1000/538*	*1200/649*	*1400/760*	*1600/871*
υ	1.1231	1.2155	1.3005	1.3332	1.3652	1.3967	1.4278	1.4584	1.6074	1.7516	1.8928	2.170	2.443	2.714
450 *h*	1238.4	1272.0	1302.8	1314.6	1326.2	1337.5	1348.8	1359.9	1414.3	1467.7	1521.0	1628.6	1738.7	1851.9
(456.28) *s*	1.5095	1.5437	1.5735	1.5845	1.5951	1.6054	1.6153	1.6250	1.6699	1.7108	1.7486	1.8177	1.8803	1.9381
υ	0.9927	1.0800	1.1591	1.1893	1.2188	1.2478	1.2763	1.3044	1.4405	1.5715	1.6996	1.9504	2.197	2.442
500 *h*	1231.3	1266.8	1298.6	1310.7	1322.6	1334.2	1345.7	1357.0	1412.1	1466.0	1519.6	1627.6	1737.9	1851.3
(467.01) *s*	1.4919	1.5280	1.5588	1.5701	1.5810	1.5915	1.6016	1.6115	1.6571	1.6982	1.7363	1.8056	1.8683	1.9262
υ	0.8852	0.9686	1.0431	1.0714	1.0989	1.1259	1.1523	1.1783	1.3038	1.4241	1.5414	1.7706	1.9957	2.219
550 *h*	1223.7	1261.2	1294.3	1306.8	1318.9	1330.8	1342.5	1354.0	1409.9	1464.3	1518.2	1626.6	1737.1	1850.6
(476.94) *s*	1.4751	1.5131	1.5451	1.5568	1.5680	1.5787	1.5890	1.5991	1.6452	1.6868	1.7250	1.7946	1.8575	1.9155
υ	0.7947	0.8753	0.9463	0.9729	0.9988	1.0241	1.0489	1.0732	1.1899	1.3013	1.4096	1.6208	1.8279	2.033
600 *h*	1215.7	1255.5	1289.9	1302.7	1315.2	1327.4	1339.3	1351.1	1407.7	1462.5	1516.7	1625.5	1736.3	1850.0
(486.21) *s*	1.4586	1.4990	1.5323	1.5443	1.5558	1.5667	1.5773	1.5875	1.6343	1.6762	1.7147	1.7846	1.8476	1.9056
υ		0.7277	0.7934	0.8177	0.8411	0.8639	0.8860	0.9077	1.0108	1.1082	1.2024	1.3853	1.5641	1.7405
700 *h*		1243.2	1280.6	1294.3	1307.5	1320.3	1332.8	1345.0	1403.2	1459.0	1513.9	1623.5	1734.8	1848.8
(503.10) *s*		1.4722	1.5084	1.5212	1.5333	1.5449	1.5559	1.5665	1.6147	1.6573	1.6963	1.7666	1.8299	1.8881
υ		0.6154	0.6779	0.7006	0.7223	0.7433	0.7635	0.7833	0.8763	0.9633	1.0470	1.2088	1.3662	1.5214
800 *h*		1229.8	1270.7	1285.4	1299.4	1312.9	1325.9	1338.6	1398.6	1455.4	1511.0	1621.4	1733.2	1847.5
(518.23) *s*		1.4467	1.4863	1.5000	1.5129	1.5250	1.5366	1.5476	1.5972	1.6407	1.6801	1.7510	1.8146	1.8729
υ		0.5264	0.5873	0.6089	0.6294	0.6491	0.6680	0.6863	0.7716	0.8506	0.9262	1.0714	1.2124	1.3509
900 *h*		1215.0	1260.1	1275.9	1290.9	1305.1	1318.8	1332.1	1393.9	1451.8	1508.1	1619.3	1731.6	1846.3
(531.98) *s*		1.4216	1.4653	1.4800	1.4938	1.5066	1.5187	1.5303	1.5814	1.6257	1.6656	1.7371	1.8009	1.8595
υ		0.4533	0.5140	0.5350	0.5546	0.5733	0.5912	0.6084	0.6878	0.7604	0.8294	0.9615	1.0893	1.2146
1000 *h*		1198.3	1248.8	1265.9	1281.9	1297.0	1311.4	1325.3	1389.2	1448.2	1505.1	1617.3	1730.0	1845.0
(544.61) *s*		1.3961	1.4450	1.4610	1.4757	1.4893	1.5021	1.5141	1.5670	1.6121	1.6525	1.7245	1.7886	1.8474
υ			0.4532	0.4738	0.4929	0.5110	0.5281	0.5445	0.6191	0.6866	0.7503	0.8716	0.9885	1.1031
1100 *h*			1236.7	1255.3	1272.4	1288.5	1303.7	1318.3	1384.3	1444.5	1502.2	1615.2	1728.4	1843.8
(556.31) *s*			1.4251	1.4425	1.4583	1.4728	1.4862	1.4989	1.5535	1.5995	1.6405	1.7130	1.7775	1.8363
υ			0.4016	0.4222	0.4410	0.4586	0.4752	0.4909	0.5617	0.6250	0.6843	0.7967	0.9046	1.0101
1200 *h*			1223.5	1243.9	1262.4	1279.6	1295.7	1311.0	1379.3	1440.7	1499.2	1613.1	1726.9	1842.5
(567.22) *s*			1.4052	1.4243	1.4413	1.4568	1.4710	1.4843	1.5409	1.5879	1.6293	1.7025	1.7672	1.8263
υ			0.3174	0.3390	0.3580	0.3753	0.3912	0.4062	0.4714	0.5281	0.5805	0.6789	0.7727	0.8640
1400 *h*			1193.0	1218.4	1240.4	1260.3	1278.5	1295.5	1369.1	1433.1	1493.2	1608.9	1723.7	1840.0
(587.10) *s*			1.3639	1.3877	1.4079	1.4258	1.4419	1.4567	1.5177	1.5666	1.6093	1.6836	1.7489	1.8083

υ – specific volume (ft^3/lbm); *h* – enthalpy (BTU/lbm); *s* – entropy (BTU/lbm °F)

TABLE A.5b Continued

Properties of Superheated Steam

Abs. Press., PSIA[a] (Sat. Temp. °F)	Temperature, °F/°C 500/260	550/288	600/316	620/327	640/338	660/349	680/360	700/371	800/427	900/482	1000/538	1200/649	1400/760	1600/871
υ				0.2733	0.2936	0.3112	0.3271	0.3417	0.4034	0.4553	0.5027	0.5906	0.6738	0.7545
1600 *h*				1187.8	1215.2	1238.7	1259.6	1278.7	1358.4	1425.3	1487.0	1604.6	1720.5	1837.5
(604.90) *s*				1.3489	1.3741	1.3952	1.4137	1.4303	1.4964	1.5476	1.5914	1.6669	1.7328	1.7926
υ					0.2407	0.2597	0.2760	0.2907	0.3502	0.3986	0.4421	0.5218	0.5968	0.6693
1800 *h*					1185.1	1214.0	1238.5	1260.3	1347.2	1417.4	1480.8	1600.4	1717.3	1835.0
(621.03) *s*					1.3377	1.3638	1.3855	1.4044	1.4765	1.5301	1.5752	1.6520	1.7185	1.7786
υ					0.1936	0.2161	0.2337	0.2489	0.3074	0.3532	0.3935	0.4668	0.5352	0.6011
2000 *h*					1145.6	1184.9	1214.8	1240.0	1335.5	1409.2	1474.5	1596.1	1714.1	1832.5
(635.82) *s*					1.2945	1.3300	1.3564	1.3783	1.4576	1.5139	1.5603	1.6384	1.7055	1.7660
υ							0.1484	0.1686	0.2294	0.2710	0.3061	0.3678	0.4244	0.4784
2500 *h*							1132.3	1176.8	1303.6	1387.8	1458.4	1585.3	1706.1	1826.2
(668.13) *s*							1.2687	1.3073	1.4127	1.4772	1.5273	1.6088	1.6775	1.7389
υ								0.0984	0.1760	0.2159	0.2476	0.3018	0.3505	0.3966
3000 *h*								1060.7	1267.2	1365.0	1441.8	1574.3	1698.0	1819.9
(695.36) *s*								1.1966	1.3690	1.4439	1.4984	1.5837	1.6540	1.7163
υ									0.1583	0.1981	0.2288	0.2806	0.3267	0.3703
3206.2 *h*									1250.5	1355.2	1434.7	1569.8	1694.6	1817.2
(705.40) *s*									1.3508	1.4309	1.4874	1.5742	1.6452	1.7080
υ								0.0306	0.1364	0.1762	0.2058	0.2546	0.2977	0.3381
3500 *h*								780.5	1224.9	1340.7	1424.5	1563.3	1689.8	1813.6
S								0.9515	1.3241	1.4127	1.4723	1.5615	1.6336	1.6968
υ								0.0287	0.1052	0.1462	0.1743	0.2192	0.2581	0.2943
4000 *h*								763.8	1174.8	1314.4	1406.8	1552.1	1681.7	1807.2
S								0.9347	1.2757	1.3827	1.4482	1.5417	1.6154	1.6795
υ								0.0276	0.0798	0.1226	0.1500	0.1917	0.2273	0.2602
4500 *h*								753.5	1113.9	1286.5	1388.4	1540.8	1673.5	1800.9
S								0.9235	1.2204	1.3529	1.4253	1.5235	1.5990	1.6640
υ								0.0268	0.0593	0.1036	0.1303	0.1696	0.2027	0.2329
5000 *h*								746.4	1047.1	1256.5	1369.5	1529.5	1665.3	1794.5
S								0.9152	1.1622	1.3231	1.4034	1.5066	1.5839	1.6499
υ								0.0262	0.0463	0.0880	0.1143	0.1516	0.1825	0.2106
5500 *h*								741.3	985.0	1224.1	1349.3	1518.2	1657.0	1788.1
S								0.9090	1.1093	1.2930	1.3821	1.4908	1.5699	1.6369

Source: Abridged from *Thermodynamic Properties of Steam*, by Joseph H. Keenan and Frederick G. Keyes. © 1936, by Joseph Keenan and Frederick G. Keyes. Published by John Wiley & Sons, Inc., New York.

[a] For SI units see Section A.1.

υ – specific volume (ft^3/lbm); *h* – enthalpy (BTU/lbm); *s* – entropy (BTU/lbm °F)

TABLE A.5c

Properties of Water at Various Temperatures from 40 to 540°F (4.4 to 282.2°C)

Temp., °F	Temp., °C	Specific Volume[a] ft^3/lb	Specific Gravity	Weight[a] (lb/ft^3)	Vapor Pressure[a] PSIA
40	4.4	.01602	1.0013	62.42	0.1217
50	10.0	.01603	1.0006	62.38	0.1781
60	15.6	.01604	1.0000	62.34	0.2563
70	21.1	.01606	0.9987	62.27	0.3631
80	26.7	.01608	0.9975	62.19	0.5069
90	32.2	.01610	0.9963	62.11	0.6982
100	37.8	.01613	0.9944	62.00	0.9492
120	48.9	.01620	0.9901	61.73	1.692
140	60.0	.01629	0.9846	61.39	2.889
160	71.1	.01639	0.9786	61.01	4.741
180	82.2	.01651	0.9715	60.57	7.510
200	93.3	.01663	0.9645	60.13	11.526
212	100.0	.01672	0.9593	59.81	14.696
220	104.4	.01677	0.9565	59.63	17.186
240	115.6	.01692	0.9480	59.10	24.97
260	126.7	.01709	0.9386	58.51	35.43
280	137.8	.01726	0.9293	58.00	49.20
300	148.9	.01745	0.9192	57.31	67.01
320	160.0	.01765	0.9088	56.66	89.66
340	171.1	.01787	0.8976	55.96	118.01
360	182.2	.01811	0.8857	55.22	153.04
380	193.3	.01836	0.8736	54.47	195.77
400	204.4	.01864	0.8605	53.65	247.31
420	215.6	.01894	0.8469	52.80	308.83
440	226.7	.01926	0.8328	51.92	381.59
460	237.8	.0196	0.8183	51.02	466.9
480	248.9	.0200	0.8020	50.00	566.1
500	260.0	.0204	0.7863	49.02	680.8
520	271.1	.0209	0.7674	47.85	812.4
540	282.2	.0215	0.7460	46.51	962.5

[a] ft^3/lb = 62.4 l/Kg; lb/ft^3 = 0.016 Kg/l; PSIA = 0.069 bar (abs.).

Computed from Keenan & Keyes Steam Table.

A.6 Friction Loss in Pipes

B. G. LIPTÁK

Friction loss moduli for laminar flow are shown by the 45° lines in the upper left-hand portion of each chart. Moduli for turbulent flow are shown by the steeper curves in the lower right-hand portion. Both of these regions represent stable states of flow. A diagonal line separates the regions of laminar and turbulent flow and represents the critical zone, a region in which it is difficult to predict the state of flow and, hence, the friction loss. The critical zone usually represents a region of unstable flow. The critical zone line gives approximate moduli on the high side for this region of unstable flow.

The bottom scale of each chart represents flow in gallons per minute, GPM. An auxiliary top scale shows the average velocity in the pipe in feet per second. Read vertically from the GPM scale to find the corresponding velocity in feet per second. The vertical scales, labeled "Friction Loss Modulus for 100 Feet of Pipe," represent values of the ratio:

$$M = \frac{\Delta p}{SG}$$

where

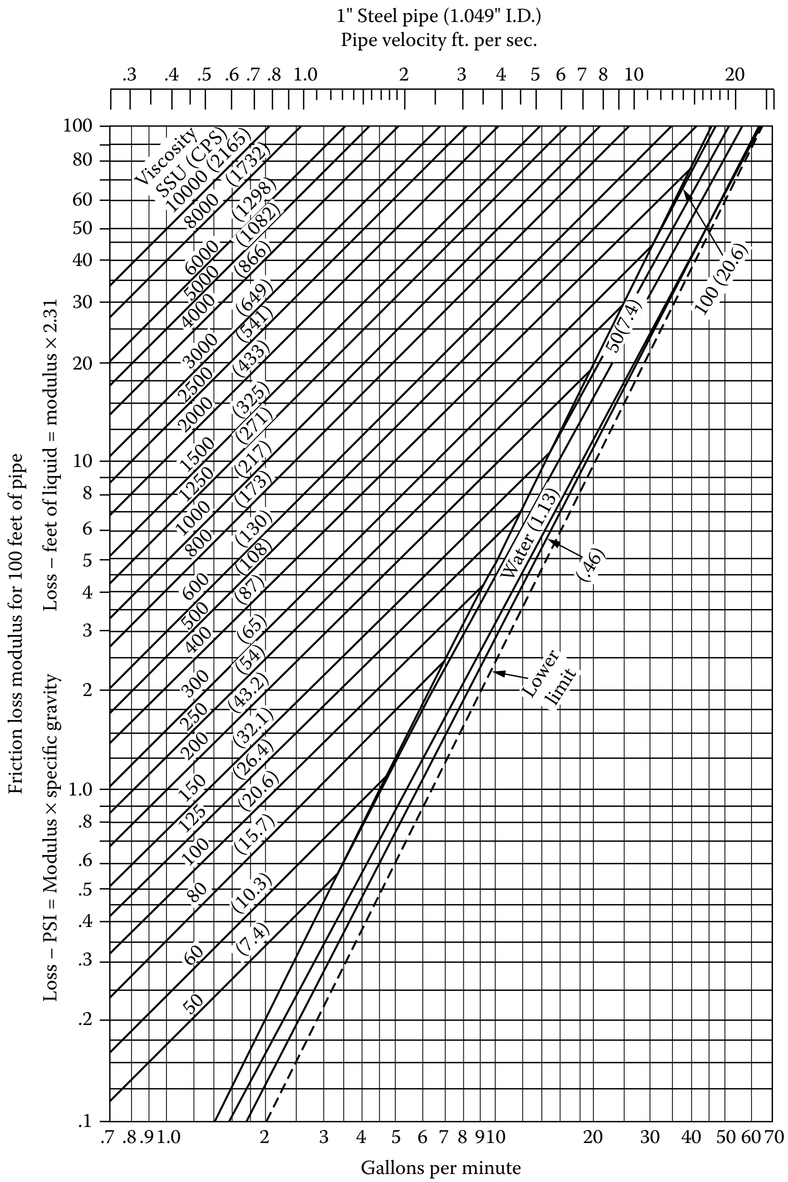
M = friction loss modulus for 100 feet (30 m) of pipe
Δp = pressure loss in pounds per square inch per 100 feet of pipe
SG = specific gravity of fluid at 60°F (15.6°C)

The loss due to pipe friction may be obtained as follows:

$$\Delta p = M \times SG$$

To use the charts, proceed as follows:

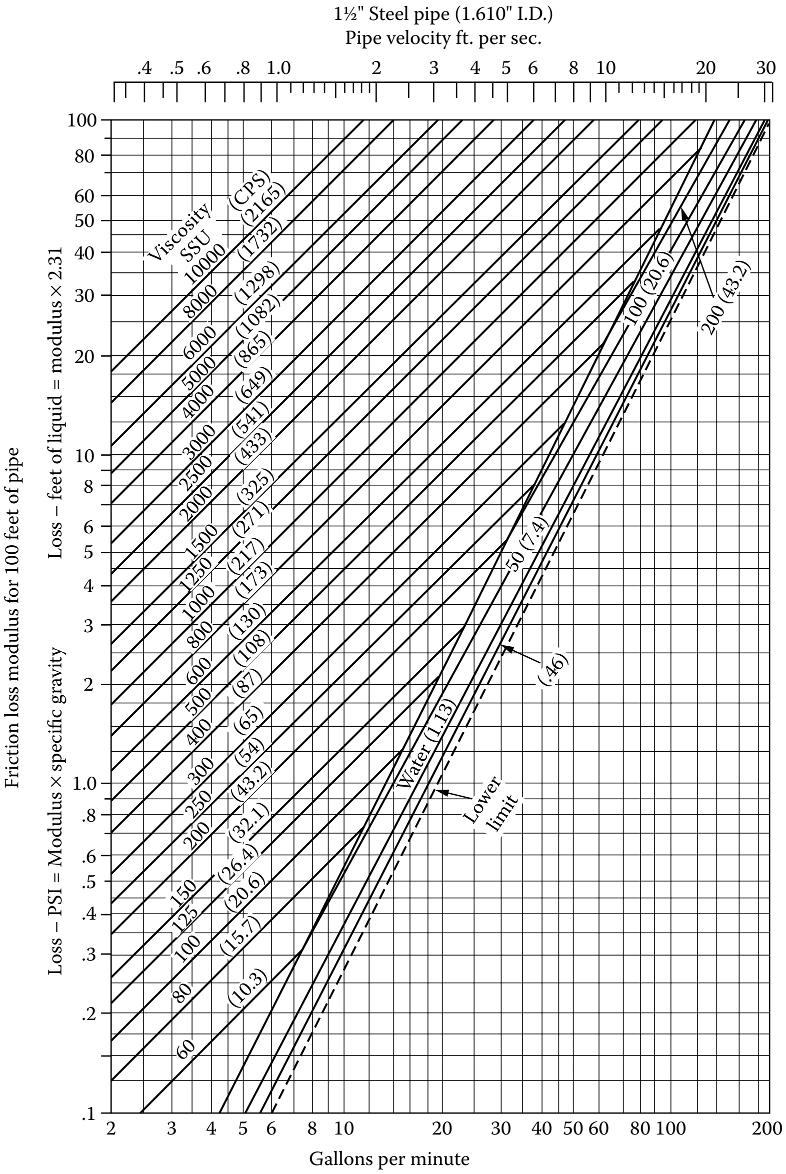
1. Select the chart for the size of pipe in question.
2. Follow the vertical line representing the flow in GPM to its intersection with the desired viscosity curve, and read the modulus at the left.
3. If the vertical line representing the flow in GPM does not intersect the viscosity line in either turbulent or laminar flow, use the intersection with the critical zone line.
4. Compute the friction loss in pressure drop from the equation above.



Loss – lbs. per sq. in. = modulus × specific gravity
Loss – feet of liquid = modulus × 2.31

*For SI units see Section A.1.

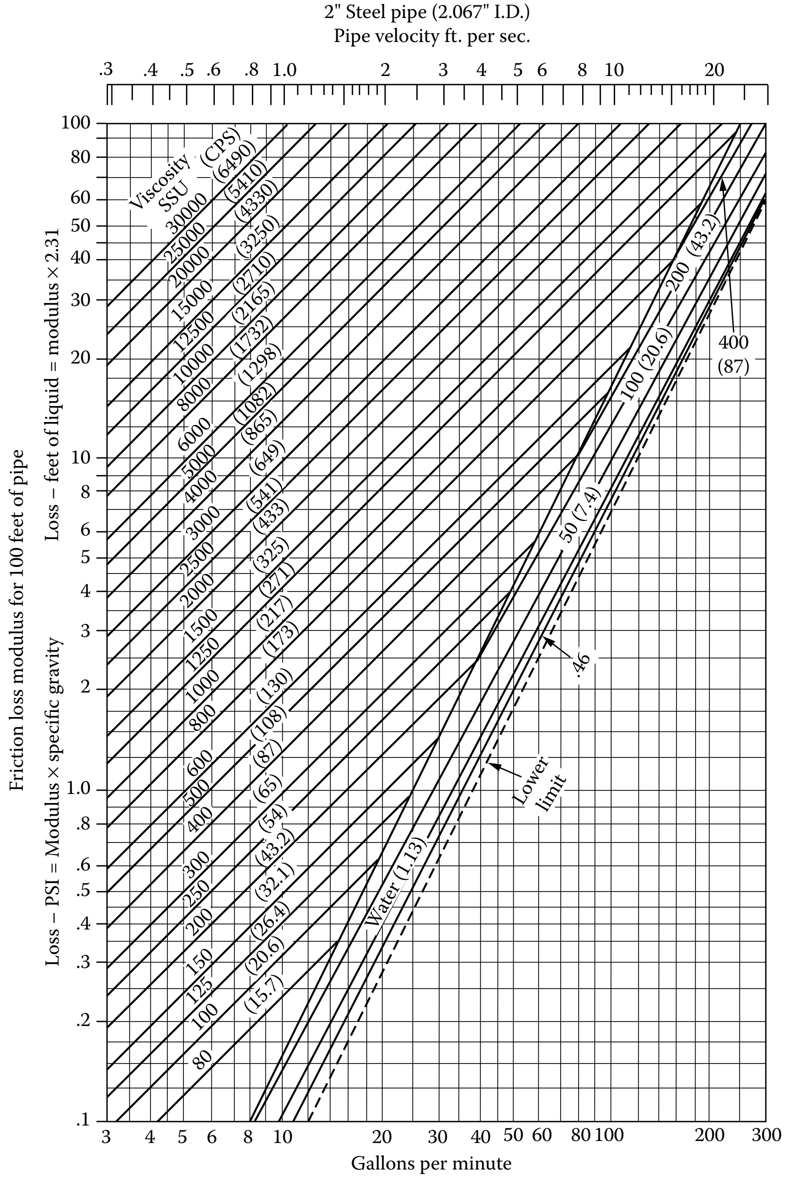
FIG. A.6a
*Friction loss modulus for 100 ft of 1″ steel pipe.**



Loss – lbs. per sq. in. = modulus × specific gravity
Loss – feet of liquid = modulus × 2.31

*For SI units see Section A.1.

FIG. A.6b
*Friction loss modulus for 100 ft of 1½″ steel pipe.**

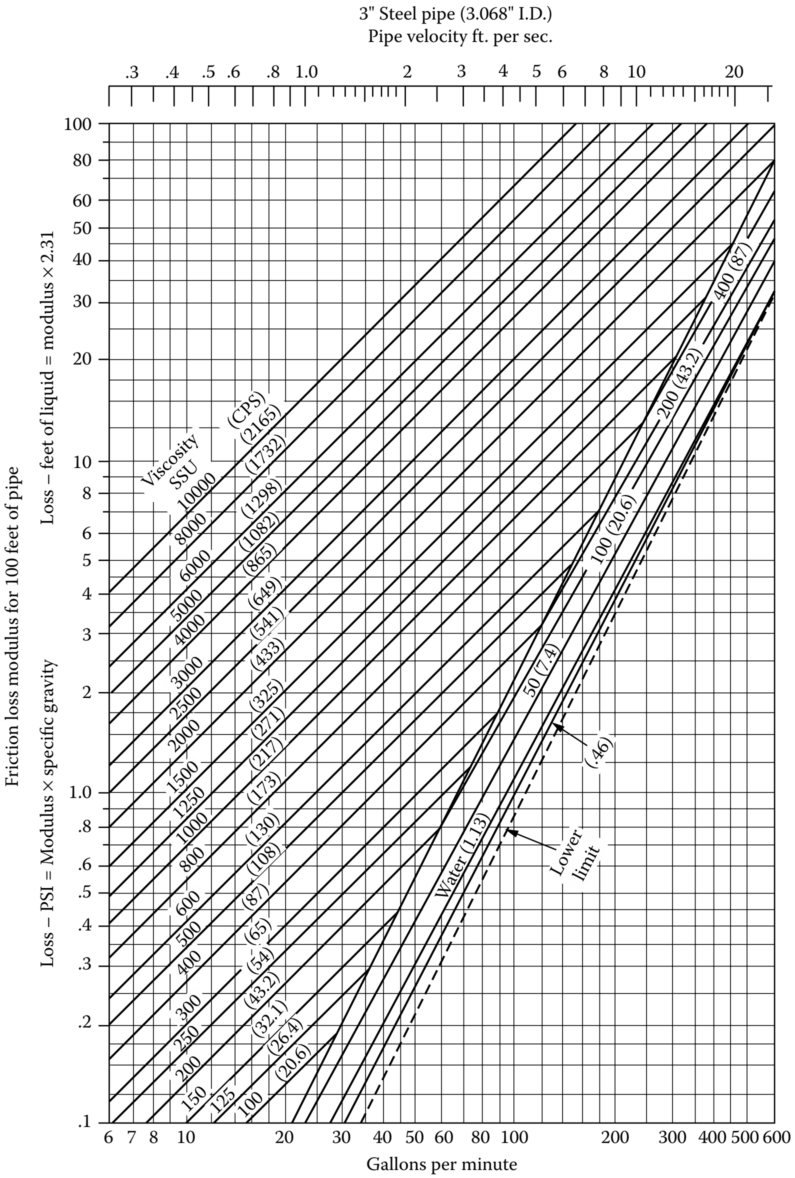


Loss – lbs. per sq. in. = modulus × specific gravity
Loss – feet of liquid = modulus × 2.31

*For SI units see Section A.1.

FIG. A.6c

*Friction loss modulus for 100 ft of 2″ steel pipe.**

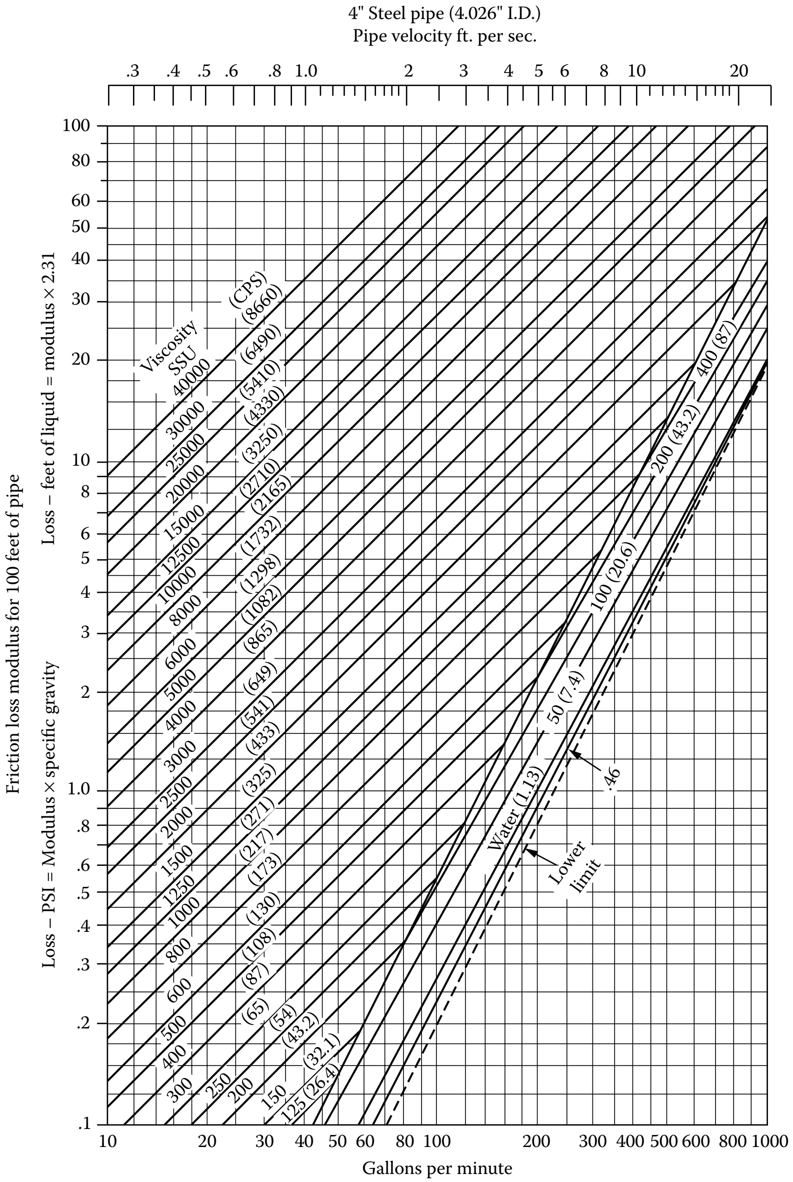


Loss – lbs. per sq. in. = modulus × specific gravity
Loss – feet of liquid = modulus × 2.31

*For SI units see Section A.1.

FIG. A.6d

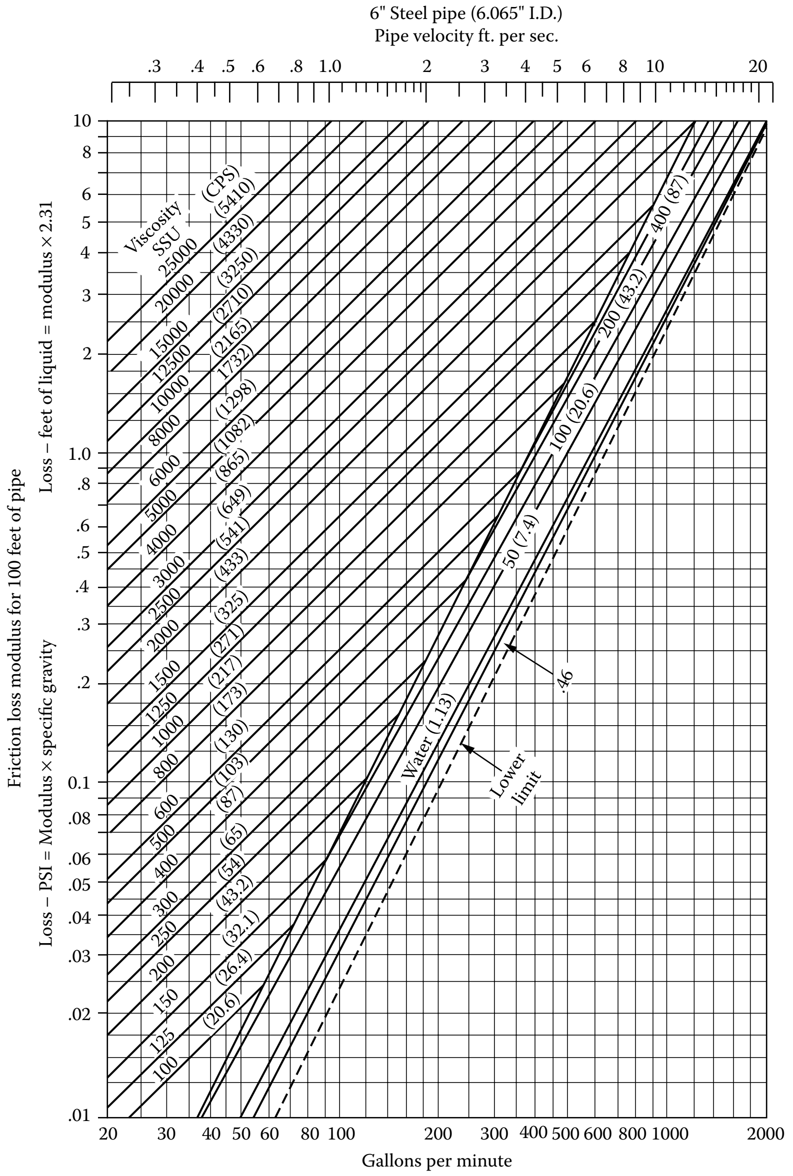
*Friction loss modulus for 100 ft of 3″ steel pipe.**



Loss – lbs. per sq. in. = modulus × specific gravity
Loss – feet of liquid = modulus × 2.31

*For SI units see Section A.1.

FIG. A.6e
*Friction loss modulus for 100 ft of 4″ steel pipe.**



Loss – lbs. per sq. in. = modulus × specific gravity
Loss – feet of liquid = modulus × 2.31

*For SI units see Section A.1.

FIG. A.6f
*Friction loss modulus for 100 ft of 6″ steel pipe.**

A.7 Tank Volumes

B. G. LIPTÁK

TABLE A.7a
Capacity of Round Tanks[a] (per foot of depth)

Diam.	*Gals.*	*Area sq. ft*	*Diam.*	*Gals.*	*Area sq. ft*	*Diam.*	*Gals.*	*Area sq. ft*	*Diam.*	*Gals.*	*Area sq. ft*
1′	5.87	.785	4′	94.00	12.566	11′	710.90	95.03	22′	2843.60	380.13
1′1″	6.89	.922	4′1″	97.96	13.095	11′3″	743.58	99.40	22′3″	2908.60	388.82
1′2″	8.00	1.069	4′2″	102.00	13.635	11′6″	776.99	103.87	22′6″	2974.30	397.61
1′3″	9.18	1.227	4′3″	106.12	14.186	11′9″	811.14	108.43	22′9″	3040.80	406.49
1′4″	10.44	1.396	4′4″	110.32	14.748	12′	846.03	113.10	23′	3108.00	415.48
1′5″	11.79	1.576	4′5″	114.61	15.321	12′3″	881.65	117.86	23′3″	3175.90	424.56
1′6″	13.22	1.767	4′6″	118.97	15.90	12′6″	918.00	122.72	23′6″	3244.60	433.74
1′7″	14.73	1.969	4′7″	123.42	16.50	12′9″	955.09	127.68	23′9″	3314.00	443.01
1′8″	16.32	2.182	4′8″	127.95	17.10	13′	992.91	132.73	24′	3384.10	452.39
1′9″	17.99	2.405	4′9″	132.56	17.72	13′3″	1031.50	137.89	24′3″	3455.00	461.86
1′10″	19.75	2.640	4′10″	137.25	18.35	13′6″	1070.80	142.14	24′6″	3526.60	471.44
1′11″	21.58	2.885	4′11″	142.02	18.99	13′9″	1110.80	148.49	24′9″	3598.90	481.11
2′	23.50	3.142	5′	146.76	19.62	14′	1151.50	153.94	25′	3672.00	490.87
2′1″	25.50	3.409	5′3″	161.86	21.64	14′3″	1193.00	159.48	25′3″	3745.80	500.74
2′2″	27.58	3.687	5′6″	177.66	23.75	14′6″	1235.30	165.13	25′6″	3820.30	510.71
2′3″	29.74	3.976	5′9″	194.27	25.97	14′9″	1278.20	170.87	25′9″	3895.60	520.77
2′4″	31.99	4.276	6′	211.51	28.27	15′	1321.90	176.71	26′	3971.60	530.93
2′5″	34.31	4.587	6′3″	229.50	30.68	15′3″	1366.40	182.65	26′3″	4048.40	541.19
2′6″	36.72	4.909	6′6″	248.23	35.18	15′6″	1411.50	188.69	26′6″	4125.90	551.55
2′7″	39.21	5.241	6′9″	267.69	35.78	15′9″	1457.40	194.83	26′9″	4204.10	562.00
2′8″	41.78	5.585	7′	287.88	38.48	16′	1504.10	201.06	27′	4283.00	572.66
2′9″	44.43	5.940	7′3″	308.81	41.28	16′3″	1551.40	207.39	27′3″	4362.70	583.21
2′10″	47.16	6.305	7′6″	330.48	44.18	16′6″	1599.50	213.82	27′6″	4443.10	593.96
2′11″	49.98	6.681	7′9″	352.88	47.17	16′9″	1648.40	220.35	27′9″	4524.30	604.81
3′	52.88	7.069	8′	376.01	50.27	17′	1697.21	226.87	28′	4606.20	615.75
3′1″	55.86	7.467	8′3″	399.80	53.46	17′6″	1798.51	240.41	28′3″	4688.80	626.80
3′2″	58.92	7.876	8′6″	424.48	56.75	18′	1902.72	254.34	28′6″	4772.10	637.94
3′3″	62.06	8.296	8′9″	449.82	60.13	18′6″	2009.92	268.67	28′9″	4856.20	649.18
3′4″	65.28	8.727	9′	475.89	63.62	19′	2120.90	283.53	29′	4941.00	660.52
3′5″	68.58	9.168	9′3″	502.70	67.20	19′6″	2234.00	298.65	29′3″	5026.60	671.96
3′6″	71.97	9.621	9′6″	530.24	70.88	20′	2350.10	314.16	29′6″	5112.90	683.49
3′7″	75.44	10.085	9′9″	558.51	74.66	20′6″	2469.10	330.06	29′9″	5199.90	695.13
3′8″	78.99	10.559	10′	587.52	78.54	21′	2591.00	346.36	30′	5287.70	706.86

TABLE A.7a Continued

Capacity of Round Tanks[a] (per foot of depth)

Diam.	Gals.	Area sq. ft	Diam.	Gals.	Area sq. ft	Diam.	Gals.	Area sq. ft	Diam.	Gals.	Area sq. ft
3′9″	82.62	11.045	10′3″	617.26	82.52	21′3″	2653.00	354.66	30′3″	5376.20	718.69
3′10″	86.33	11.541	10′6″	640.74	86.59	21′6″	2715.80	363.05	30′6″	5465.40	730.62
3′11″	90.13	12.048	10′9″	678.95	90.76	21′9″	2779.30	371.54	30′9″	5555.40	742.64

To find the capacity of tanks greater than shown above, find a tank of one-half the size desired, and multiply its capacity by four, or find one one-third the size desired, and multiply its capacity by 9.

[a] foot = 0.3048 m; gals = 3.785 l; sq. ft = 0.0929 m^2; inch = 25.4 mm

TABLE A.7b

Capacity of Partially Filled Horizontal Tanks[a]

Diam. (in ft)	Gallons per ft of Length When Tank Is Filled to Noted Fraction of Diameter								
	1/10	1/5	3/10	2/5	1/2	3/5	7/10	4/5	9/10
1	.3	.8	1.4	2.1	2.9	3.6	4.3	4.9	5.5
2	1.2	3.3	5.9	8.8	11.7	14.7	17.5	20.6	22.2
3	2.7	7.5	13.6	19.8	26.4	33.0	39.4	45.2	50.1
4	4.9	13.4	23.8	35.0	47.0	59.0	70.2	80.5	89.0
5	7.6	20.0	37.0	55.0	73.0	92.0	110.0	126.0	139.0
6	11.0	30.0	53.0	78.0	106.0	133.0	158.0	182.0	201.0
7	15.0	41.0	73.0	107.0	144.0	181.0	215.0	247.0	272.0
8	19.0	52.0	96.0	140.0	188.0	235.0	281.0	322.0	356.0
9	25.0	67.0	112.0	178.0	238.0	298.0	352.0	408.0	450.0
10	30.0	83.0	149.0	219.0	294.0	368.0	440.0	504.0	556.0
11	37.0	101.0	179.0	265.0	356.0	445.0	531.0	610.0	672.0
12	44.0	120.0	214.0	315.0	423.0	530.0	632.0	741.0	800.0
13	51.0	141.0	250.0	370.0	496.0	621.0	740.0	850.0	940.0
14	60.0	164.0	291.0	430.0	576.0	722.0	862.0	989.0	1084.0
15	68.0	188.0	334.0	494.0	661.0	829.0	988.0	1134.0	1253.0

[a] foot = 0.3048 m; gallon = 3.785 l.

TABLE A.7c

Capacities of Various Cylinders in U.S. Gallons[a]

Diam. (in inches)	Length of Cylinder																	
	1″	1′	5′	6′	7′	8′	9′	10′	11′	12′	13′	14′	15′	16′	17′	18′	20′	22′	24′
1		0.04	0.20	0.24	0.28	0.32	0.36	0.40	0.44	0.48	0.52	0.56	0.60	0.64	0.68	0.72	0.80	0.88	0.96
2	0.01	0.16	0.80	0.96	1.12	1.28	1.44	1.60	1.76	1.92	2.08	2.24	2.40	2.56	2.72	2.88	3.20	3.52	3.84
3	0.03	0.37	1.84	2.20	2.56	2.92	3.30	3.68	4.04	4.40	4.76	5.12	5.48	5.84	6.22	6.60	7.36	8.08	8.80
4	0.05	0.65	3.26	3.92	4.58	5.24	5.88	6.52	7.18	7.84	8.50	9.16	9.82	10.5	11.1	11.8	13.0	14.4	15.7
5	0.08	1.02	5.10	6.12	7.14	8.16	9.18	10.2	11.2	12.2	13.3	14.3	15.3	16.3	17.3	18.4	20.4	22.4	24.4
6	0.12	1.47	7.34	8.80	10.3	11.8	13.2	14.7	16.1	17.6	19.1	20.6	22.0	23.6	25.0	26.4	29.4	32.2	35.2
7	0.17	2.00	10.0	12.0	14.0	16.0	18.0	20.0	22.0	24.0	26.0	28.0	30.0	32.0	34.0	36.0	40.0	44.0	48.0
8	0.22	2.61	13.0	15.6	18.2	20.8	23.4	26.0	28.6	31.2	33.8	36.4	39.0	41.6	44.2	46.8	52.0	57.2	62.4
9	0.28	3.31	16.5	19.8	23.1	26.4	29.8	33.0	36.4	39.6	43.0	46.2	49.6	52.8	56.2	60.0	66.0	72.4	79.2
10	0.34	4.08	20.4	24.4	28.4	32.6	36.8	40.8	44.8	48.8	52.8	56.8	61.0	65.2	69.4	73.6	81.6	89.6	97.6
11	0.41	4.94	24.6	29.6	34.6	39.4	44.4	49.2	54.2	59.2	64.2	69.2	74.0	78.8	83.8	88.8	98.4	104.	118.
12	0.49	5.88	29.4	35.2	41.0	46.8	52.8	58.8	64.6	70.4	76.2	82.0	87.8	93.6	99.6	106.	118.	129.	141.
13	0.57	6.90	34.6	41.6	48.6	55.2	62.2	69.2	76.2	83.2	90.2	97.2	104.	110.	117.	124.	138.	152.	166.
14	0.67	8.00	40.0	48.0	56.0	64.0	72.0	80.0	88.0	96.0	104.	112.	120.	128.	136.	144.	160.	176.	192.
15	0.77	9.18	46.0	55.2	64.4	73.6	82.8	92.0	101.	110.	120.	129.	138.	147.	156.	166.	184.	202.	220.
16	0.87	10.4	52.0	62.4	72.8	83.2	93.6	104.	114.	125.	135.	146.	156.	166.	177.	187.	208.	229.	250.
17	0.98	11.8	59.0	70.8	81.6	94.4	106.	118.	130.	142.	153.	163.	177.	189.	201.	212.	236.	260.	283.
18	1.10	13.2	66.0	79.2	92.4	106.	119.	132.	145.	158.	172.	185.	198.	211.	224.	240.	264.	290.	317.
19	1.23	14.7	73.6	88.4	103.	118.	132.	147.	162.	177.	192.	206.	221.	235.	250.	265.	294.	324.	354.
20	1.36	16.8	81.6	98.0	114.	130.	147.	163.	180.	196.	212.	229.	245.	261.	277.	294.	326.	359.	392.
21	1.50	18.0	90.0	108.	126.	144.	162.	180.	198.	216.	238.	252.	270.	288.	306.	324.	360.	396.	432.
22	1.65	19.8	99.0	119.	139.	158.	178.	198.	218.	238.	257.	277.	297.	317.	337.	356.	396.	436.	476.
23	1.80	21.6	108.	130.	151.	173.	194.	216.	238.	259.	281.	302.	324.	346.	367.	389.	432.	476.	518.
24	1.96	23.5	118.	141.	165.	188.	212.	235.	259.	282.	306.	330.	353.	376.	400.	424.	470.	518.	564.
25	2.12	25.5	128.	153.	179.	204.	230.	255.	281.	306.	332.	358.	383.	408.	434.	460.	510.	562.	612.
26	2.30	27.6	138.	166.	193.	221.	248.	276.	304.	331.	359.	386.	414.	442.	470.	496.	552.	608.	662.
27	2.48	29.7	148.	178.	208.	238.	267.	297.	326.	356.	386.	416.	426.	476.	504.	534.	594.	652.	712.
28	2.67	32.0	160.	192.	224.	256.	288.	320.	352.	384.	416.	448.	480.	512.	544.	576.	640.	704.	768.
29	2.86	34.3	171.	206.	240.	274.	309.	343.	377.	412.	446.	480.	514.	548.	584.	618.	686.	754.	824.
30	3.06	36.7	183.	220.	257.	294.	330.	367.	404	440.	476.	514.	550.	588.	624.	660.	734.	808.	880.
32	3.48	41.8	209.	251.	293.	334.	376.	418.	460.	502.	544.	586.	668.	668.	710.	752.	836.	920.	1004.
34	3.93	47.2	236.	283.	330.	378.	424.	472.	520.	566.	614.	660.	708.	756.	802.	848.	944.	1040.	1132.
36	4.41	52.9	264.	317.	370.	422.	476.	528.	582.	634.	688.	740.	793.	844.	898.	952.	1056.	1164.	1268.

[a] inch = 25.4 mm; foot = 0.3048 m; gallon = 3.785 1.

A.8 Partial List of Suppliers[*]

3M TOUCH SYSTEMS, 800 Carleton Ct., Annarcis Island, New Westminster, BC, V3L-6L3 Canada, 604/521-3962, Fax: 604/521-4629, www.dynapro.com

4B COMPONENTS LTD, 729 Sabrina Dr., East Peoria, IL 61611, 309/698-5611, Fax: 309/698-5615, www.go4b.com

A

AALBORG INSTRUMENTS & CONTROLS, 20 Corporate Dr., Orangeburg, NY 10962, 800/866-3837, Fax: 845/398-3165, www.aalborg.com

AALIANT, 150 Venture Blvd., Spartanburg, SC 29306, 864/574-8060, Fax: 864/574-8063, www.bindicator.com

ABB AUTOMATION—ANALYTICAL DIV., 843 N. Jefferson St., Lewisburg, WV 24901, 304/647-1761, Fax: 304/645-4236, www.abb.com/usa

ABB AUTOMATIONS, INC., DRIVES & POWER PRODUCTS, 16250 W. Glendale Dr., New Berlin, WI 53151, 800/752-0696, Fax: 262/785-0397, www.abb-drives.com

ABB, INC., 29801 Euclid Ave., Wickliffe, OH 44092, 800/626-4999, Fax: 716/273-7014, www.abb.com

ABB INSTRUMENTATION, 125 E. County Line Rd., Warminster, PA 18974, 215/674-6000, Fax: 215/674-7183, www.abb.com/us/instrumentation

ABB WATER METERS, INC., 1100 SW 38th Ave., Ocala, FL 34474, 800/874-0890, Fax: 352/368-1950, www.abbwatermeters.com

ABSOLUTE PROCESS INSTRUMENTS, INC., 1029 Butterfield Rd., Vernon Hills, IL 60061, 800/942-0315, Fax: 800/949-7502, www.api-usa.com

AC DATA SYSTEMS, 806 Clearwater Loop, Suite C, Post Falls, ID 83854, 800/890-2569, Fax: 208/777-4466, www.surgeblox.com

ACCES I/O PRODUCTS, INC., 10623 Roselle St., San Diego, CA 92121, 858/550-9559, Fax: 858/550-7322, www.acces-usa.com

ACCUTECH, 577 Main St., Hudson, MA 01749, 800/879-6576, Fax: 978/568-9085, www.savewithaccutech.com

ACME ELECTRIC CORP., 4815 W. 5th St., Lumberton, NC 28358, 910/738-1121, Fax: 910/739-0024, www.acmepowerdist.com

ACOPIAN, PO Box 638, Easton, PA 18044, 610/258-5441, Fax: 610/258-2842, www.acopian.com

ACP, 6865 Shiloh Rd. E., Alpharetta, GA 30005, 770/205-2475, Fax: 770/888-5362, www.acpthinclient.com

ACROMAG, INC., 30765 South Wixom Rd., Wixom, MI 48393, 248/624-1541, Fax: 248/624-9234, www.acromag.com

ACTION INSTRUMENTS, 8601 Aero Dr., San Diego, CA 92123, 585/279-5726, Fax: 858/279-6290, Millard Schewe, millards@actionio.com, www.actionio.com

AD PRODUCTS CO., 4799 W. 150th St., Cleveland, OH 44135, 800/325-4935, Fax: 216/267-5392, www.adproductsco.com

ADALET, 4801 W. 150th St., Cleveland, OH 44135, 216/267-9000, Fax: 216/267-1681, www.adalet.com

ADAPTIVE MICRO SYSTEMS, 7840 N. 86th St., Milwaukee, WI 53224, 414/357-2020, Fax: 414/357-2029, www.adaptivedisplays.com

ADAPTIVE RESOURCES, 2 Park Dr., Lawrence, PA 15055, 724/746-4969, Fax: 724/746-9260, www.adaptiveresources.com

ADLINK TECHNOLOGY INC., 15279 Alton Pkwy, Ste. 400, Irvine, CA 92618, 949/727-2077, Fax: 949/727-2099, www.adlinktech.com

ADTECH, 3750 Monroe Ave., Pittsford, NY 14534, 716/383-8280, Fax: 716/383-8386, www.adtech-inst.com

ADVANCED CONTROL TECHNOLOGY, INC., 7050 E. Hwy. 101, Shakopee, MN 55379, 952/882-0000, Fax: 952/890-3644

ADVANCED MOTION CONTROLS, 3805 Calle Tecate, Camarillo, CA 93012, 805/389-1935, Fax: 805/389-1165, www.a-m-c.com

ADVANCED SEPARATIONS AND PROCESS SYSTEMS, INC., 6111 Pepsi Way, Windsor, WI 53716, 608/846-1130, Fax: 608/846-1144, www.asapsys.com

ADVANCED SYSTEMS & DESIGNS, INC., 1100 Owendale #J, Troy, MI 48083, 248/689-4800, Fax: 248/689-8811, www.asdspc1.com

ADVANTECH AUTOMATION CORP., 1320 Kemper Meadow Dr. #500, Cincinnati, OH 45240, 513/742-8895, Fax: 513/742-8892, www.advantech.com

ADVANTECH TECHNOLOGIES, INC., 15375 Barranca Pkwy., Irvine, CA 92618, 949/789-7178, Fax: 949/789-7179, www.advantech.com/epc

---

[*] From *CONTROL Magazine Buyer's Guide*, May 2005. With permission.

AEROTECH, INC., 101 Zeta Dr., Pittsburgh, PA 15238, 412/963-7470, Fax: 412/963-7459, www.aerotech.com

AFCON CONTROL & AUTOMATION, INC., 1014 E. Algonquin Rd. #102, Schaumburg, IL 60173, 847/397-6900, Fax: 847/397-6987, www.afcon-inc.com

AGILE SYSTEMS, 575 Kumpf Dr., Waterloo, Ontario, N2V 1K3 Canada, 519/886-2000, Fax: 519/886-2075, www.agile-systems.com

AIR DIMENSIONS, INC., 1015 W. Newport Center Dr., Deerfield Beach, FL 33442, 954/428-7333, Fax: 954/360-0987, www.airdimensions.com

AIR INSTRUMENTS & MEASUREMENTS, INC., 13300 Brooks Dr. Suite A, Baldwin Park, CA 91706, 626/813-1460, Fax: 626/338-2565, www.aimanalysis.com

AIRPAX CORP., 807 Woods Rd., Cambridge, MD 21613, 410/228-1500, Fax: 410/228-3456, www.airpax.net

ALLEN-BRADLEY, 1201 S. 2nd St., Milwaukee, WI 53204, 414/382-2000, Fax: 414/382-4444, www.ab.com

ALLEN-BRADLEY CO. LLC, Two Executive Dr., Chelmsford, MA 01824, 978/446-3476, Fax: 978/446-3322, www.ab.com/sensors

ALLIED ELECTRONICS, 7410 Pebble Dr., Ft. Worth, TX 76118, 800/ 433-5700, www.alliedelec.com

ALLIED MOULDED PRODUCTS, INC., 222 N. Union St., Bryan, OH 43506, 419/636-4217, Fax: 419/636-2450, Valerie Hageman, www.enclosures.alliedmoulded.com

ALNOR INSTRUMENT CO., 7555 N. Linder Ave., Skokie, IL 60077, 847/677-3249, Fax: 847/677-3514, www.alnor.com

ALTECH CORP., 35 Royal Rd., Flemington, NJ 08822, 908/806-9400, Fax: 908/806-9490, www.altechcorp.com

ALTEK INDUSTRIES CORP., 35 Vantage Point Dr., Rochester, NY 14624, 716/349-3520, Fax: 716/349-3510, www.altekcalibrators.com

ALTERSYS, INC., 555 d'Auvergne St., Longueuil, Quebec, J4H 4A3 Canada, 405/674-7774, Fax: 405/674-7344, www.altersys.com

AMERICAN ELTEC INC., 2810 W. Charleston Ave., Las Vegas, NV 89102, 702/878-4085, Fax: 702/878-4735, www.americaneltec.com

AMERICAN INNOVATIONS, 12112 Technology, Ste., 100, Austin, TX 78727, 512/249-3400, Fax: 512/249-3444, www.aimonitoring.com

AMERICAN LED-GIBLE, INC., 1776 Lone Eagle St., Columbus, OH 43228, 614/851-1100, Fax: 614/851-1121, www.ledgible.com

AMERIPACK, 70 S. Main St., Cranbury, NJ 08512, 609/395-6969, Fax: 609/395-8753, www.ameripack.com

AMETEK DIXSON, 207 27 Road, Grand Junction, CO 81503, 970/242-8863, Fax: 970/245-6267, www.dixson.com

AMETEK DREXELBROOK, 205 Keith Valley Rd., Horsham, PA 19044, 215/674-1234, Fax: 215/674-2731, www.drexelbrook.com

AMETEK NATIONAL CONTROLS, 1725 Western Dr., W. Chicago, IL 60185, 630/231-5900, Fax: 630/231-1377, www.nationalcontrols.com

AMETEK PATRIOT SENSORS, 1080 N. Crooks Rd., Clawson, MI 48017, 248/435-0700, Fax: 248/435-8120, www.patriotsensors.com

AMETEK POWER INSTRUMENTS, 255 N. Union St., Rochester, NY 14605, 716/238-4054, Fax: 716/454-7805, www.rochester.com

AMETEK PROCESS INSTRUMENTS, 455 Corporate Blvd., Newark, DE 19702, 302/456-4400, Fax: 302/456-4444, www.ametekpi.com

AMETEK TEST + CALIBRATION INSTRUMENTS, 8600 Somerset Dr., Largo, FL 33773, 727/536-7831, Fax: 727/539-6882, www.ametek.com/tci

AMETEK THERMOX, 150 Freeport Rd., Pittsburgh, PA 15238, 412/828-9040, Fax: 412/826-0399, www.thermox.com

AMETEK U.S. GAUGE, 900 Clymer Ave., Sellersville, PA 18960, 215/257-6531, Fax: 215/257-3058, www.ametekusg.com

AMETEK U.S. GAUGE/PMT, 820 Penna Blvd., Feasterville, PA 19053, 215/355-6900, Fax: 215/355-2937, www.ametekusg.com

AMOT CONTROLS, 401 N. First St., Richmond, CA 94801, 510/307-8315, Fax: 510/234-9950, www.amotusa.com

AMPRO COMPUTERS, INC., 4757 Hellyer Ave., San Jose, CA 95138, 408/360-0200, Fax: 408/360-0222, www.ampro.com

ANALAB LLC, PO Box 34, Sterling, PA 18463, 570/689-3919, Fax: 570/689-9360, www.analab1.com

ANALOG DEVICES, INC., 3 Technology Way, Norwood, MA 02062, 781/987-1428, Fax: 781/937-1021, www.analog.com

ANALOGIC CORP., 8 Centennial Dr., Peabody, MA 01915, 978/977-3000, Fax: 978/977-6809, www.analogic.com

ANALYTICAL TECHNOLOGY, 6 Iron Bridge Dr., Collegeville, PA 19426, 610/917-0991, Fax: 610/917-0992, www.analyticaltechnology.com

ANDERSON INSTRUMENT CO., 3950 Greenbriar, Fultonville, NY 12072, 518/922-5315, Fax: 518/922-8997, www.andinst.com

ANIMATICS CORP., 3050 Tasman Dr., Santa Clara, CA 95054, 408/748-8721, Fax: 408/748-8725, www.animatics.com

ANN ARBOR TECHNOLOGIES, PO Box 1247, Ann Arbor, MI 48106, 734/995-1360, www.a2t.com

ANORAD CORP., 110 Oser Ave., Hauppauge, NY 11788, 631/231-1990, Fax: 631/435-1612, www.anorad.com

ANORAD, ROCKWELL AUTOMATION, 100 Precision Dr., Shirley, NY 11967, 631/344-6600, Fax: 631/344-6601, www.anorad.com

ANSIMAG, INC., 1090 Fox Ct., Bloomingdale, IL 60108, 303/940-3111, Fax: 303/940-3141, www.sundyne.com

ANT COMPUTER, INC., 20760 E. Carrey Rd., Walnut, CA 91789, 909/598-3315, Fax: 909/598-3215, www.antcomputer.com

ANTEK INDUSTRIAL INSTRUMENTS, 700 Gateway Parkway, Marble Falls, TX 78654, 830/693-5671, Fax: 830/798-8208, www.antekhou.com

ANTON PAAR USA, 10201 Maple Leaf Court, Ashland, VA 23005, 804/550-1051, Fax: 804/550-1057, www.anton-paar.com
API MOTION, INC., 45 Hazelwood Dr., Amherst, NY 14228, 716/691-9100, Fax: 716/691-9181, www.apimotion.com
APPLICOM INTERNATIONAL, INC., 4340 Redwood Hwy., Suite D-309, San Rafael, CA 94903, 415/472-1595, Fax: 415/472-1596, www.applicom-int.com
APPLIED CHEMOMETRICS, INC., PO Box 100, Sharon, MA 02067, 781/784-7700, www.chemometrics.com
APPLIKON ANALYZERS, INC., 1701 North Park Drive, #20, Kingwood, TX 77339, 281/354-2211, Fax: 281/354-0050, www.applikin.com
APT INSTRUMENTS, PO Box 345, Litchfield, IL 62056, 217/324-5444, Fax: 217/324-3858, www.aptinstruments.com
APV, 9525 W. Bryn Mawr Ave., Rosemont, IL 60018, 847/678-4300, Fax: 847/678-4313, www.apv.com
APW ZERO CASES, 500 W. 200 N., North Salt Lake City, UT 84054, 801/298-5900, Fax: 801/292-9450, www.zerocases.com
AQUA MEASURE INSTRUMENTS CO., 1712 Earhard Ct., La Verne, CA 91750, 909/392-5833, Fax: 909/392-5838, www.moisturerregisterproducts.com
ARI INDUSTRIES, 381 ARI Ct., Addison, IL 60101, 630/953-9100, Fax: 630/953-0590
ARIES ELECTRONICS INC., PO Box 130, Frenchtown, NJ 08825, 908/996-6841, Fax: 908/996-3891, www.arieselec.com
ARISTA CORP., 41300 Boyce Rd., Fremont, CA 94538, 510/226-1890, www.aristaipc.com
ARIZONA INSTRUMENT CO., 1912 W. 4th St., Tempe, AZ 85281, 602/470-1414, Fax: 480/804-0656, www.azic.com
AROMAT CORP., 629 Central Ave., New Providence, NJ 07974, 800/228-2350, Fax: 908/464-4128, www.aromat.com/acsd
ASAHI/AMERICA, INC., PO Box 653, Malden, MA 02148-6834, 781/321-5409, Fax: 781/321-4421, www.asahi-america.com
ASCO, 50 Hanover Rd., Florham Park, NJ 07932, 973/966-2372, Fax: 973/966-2448, www.ascovalve.com
A.S.I., PO Box 1230, Carlisle, PA 17013, 717/249-5542, www.asi-ez.com
ASI INSTRUMENTS, 8570 Katy Freeway #117, Houston, TX 77024, 713/461-4535, Fax: 713/461-7348, www.asiinstruments.com
ASL, INC., 100 Brickstone Sq., Andover, MA 01810, 978/658-0000, Fax: 978/658-5444, www.aslinc.com
ASPEN TECHNOLOGY, 10 Canal Park, Cambridge, MA 02141, 617/949-1000, Fax: 617/949-1030, www.aspentech.com
ASTEC, 6339 Paseo Del Lago, Carlsbad, CA 92009, 760/930-4745, Fax: 460/930-4700, www.astec.com
ASTEC POWER, 5810 Van Allen Way, Carlsbad, CA 92008, 888/41ASTEC, Fax: 760/930-4700, www.astecpower.com
ASTRO-MED., INC., 600 E. Greenwich Ave., W. Warwick, RI 02893, 401/828-4000, Fax: 401/822-2430, www.astro-med.com
ASTRODYNE CORP., 300 Myles Standish Blvd., Taunton, MA 02780, 508/823-8080, Fax: 508/823-8181, www.astrodyne.com
ATA SENSORS, 4500 Anaheim Ave., B-6, Albuquerque, NM 87113, 505/823-1320, Fax: 505/823-1560, www.atasensors.com
ATHENA CONTROLS, 5145 Campus Dr., Plymouth Meeting, PA 19462, 610/828-2490, Fax: 610/828-7084, www.athenacontrols.com
ATS RHEOSYSTEMS, 52 Georgetown Rd., Bordentown, NJ 08505, 609/298-2522, Fax: 609/298-2795, www.atsrheosystems.com
AUMA ACTUATORS, INC., 4 Zesta Dr., Pittsburgh, PA 15205, 412/787-1340, Fax: 412/787-1223, www.auma-usa.com
AUTODESK INC., 601 N. Baldwin Ave., Marion, IN 46952, 765/651-3200, Fax: 765/651-3214, www.autodesk.com
AUTOMATA, INC., 104 New Mohawk Rd., Suite A, Nevada City, CA 95959, 530/478-5882, Fax: 530/478-5881, www.automata-inc.com
AUTOMATED CONTROL CONCEPTS, INC., 3535 Route 66, Neptune, NJ 07753, 732/922-6611, Fax: 732/922-9611, www.automated-control.com
AUTOMATED SOLUTIONS, INC., 1415 Fulton Rd., Suite 205-A12, Santa Rosa, CA 95403, 707/578-5882, Fax: 707/579-5756, www.automatedsolutions.com
AUTOMATIC TIMING & CONTROLS, 1827 Freedom Rd., Lancaster, PA 17557, 717/295-0500, Fax: 717/481-7240, www.automatictiming.com
AUTOMATIONDIRECT, 3505 Hutchinson Rd., Cumming, GA 30040, 678/455-1845, Fax: 770/844-4212, www.automationdirect.com
AUTOMATION PRODUCTS, INC., 3030 Maxmy St., Houston, TX 77008, 800/231-2062, Fax: 713/869-7332, www.dynatrolusa.com
AUTOMATION SYSTEMS INTERCONNECT, INC., PO Box 1230, Carlisle, PA 17013, 877/650-5160, Fax: 717/249-5542, www.asi-ez.com
AUTOMATIONDIRECT.COM, 3505 Hutchinson Rd., Cumming, GA 30040, 800/633-0405, Fax: 770/889-7876, www.automationdirect.com
AUTOMATIONTECHIES.COM, PO Box 44759, Eden Prairie, MN 55344, 877/300-6792, Fax: 877/593-6792, www.automationtechies.com
AVERY WEIGH-TRONIX, 1000 Amstrong Dr., Fairmont, MN 56031, 507/238-4461, Fax: 507/238-8258, www.wtxweb.com
AVG PRESS AUTOMATION SYSTEMS, 343 St. Paul Blvd., Carol Stream, IL 60188, 800/TEC-ENGR, Fax: 630/668-4676, www.avg.net
AVO INTL., PO Box 9007, Valley Forge, PA 19485, 610/676-8500, Fax: 610/676-8610, www.avointl.com
AVX CORP., 801 17th Ave. S., Myrtle Beach, SC 29578, 843/946-0601, Fax: 843/626-5814, www.avxcorp.com
AXIOM TECHNOLOGY, INC., 18138 Rowland St., City of Industry, CA 91748, 626/581-3232, Fax: 626/581-3552, www.axiomtek.com

AYDIN DISPLAYS, INC., 700 Dresher Rd., Horsham, PA 19044, 215/784-5335, Fax: 215/830-9545, www.aydindisplays.com

AZONIX CORP., 900 Middlesex Turnpike, Billerica, MA 01821, 978/670-6300, Fax: 978/670-8855, www.azonix.com

B

B&R INDUSTRIAL AUTOMATION, 1325 Northmeadow, Pkwy., S-130, Roswell, GA 30076, 770/772-0400, Fax: 770/772-0243, www.br-automation.com

BABBITT INTERNATIONAL, INC., PO Box 70094, Houston, TX 77270, 800/835-8012, Fax: 713/467-8736, www.babbittlevel.com

BACHARACH, INC., 625 Alpha Dr., Pittsburgh, PA 15238, 412/963-2161, Fax: 412/963-2091, www.bacharach-inc.com

BADGER METER, INC., PO Box 245036, Milwaukee, WI 53224, 800/876-3837, Fax: 414/371-5932, www.badgermeter. com

BALDOR ELECTRIC CO., 5711 R.S. Boreham, Jr. St., Fort Smith, AR 72903, 501/646-4711, Fax: 501/648-5792, www.baldor.com

BALLUFF, 8125 Holton Dr., Florence, KY 41042, 606/727-2200, Fax: 606/727-4823, www.balluff.com

BANNER ENGINEERING CORP., 9714 10th Ave. N., Minneapolis, MN 55441, 763/544-3164, Fax: 763/544-3213, www.baneng.com

BASS-TRIGON, 7810 Shaffer Pkwy #150, Littleton, CO 80127, 303/948-0119, Fax: 303/948-0479, www.bass-trigon.com

BARBER COLMAN INDUSTRIAL INSTRUMENTS, 741-F Miller Dr., Leesburg, VA 20175, 703/443-0000, Fax: 703/669-1300, www.barber-colman.com

BARKSDALE, 3211 Fruitland, Los Angeles, CA 90058, 323/583-6218, Fax: 323/586-3060, www.barksdale.com

BARNANT CO., 28W092 Commercial Ave., Barrington, IL 60010, 847/381-7050, Fax: 847/381-7053, www.barnant.com

BARNETT ENGINEERING LTD, 215 7710 5th St. S.E., Calgary, Alberta, T2H 2L9 Canada, 403/255-9544, Fax: 403/259-2343, www.barnett-engg.com

BARTEC U.S. CORP., 9902 E. East 43rd St., Tulsa, OK 74146, 918/627-1889, Fax: 918/627-1890, www.bartecus.com

BARTON INSTRUMENT SYSTEMS, 900 S. Turnbull Canyon Rd., City of Industry, CA 91745, 626/961-2547, Fax: 626/961-4452, www.barton-instruments.com

BAUMANN INC., 130 International Dr., Portsmouth. NH 03801, 603/766-8507, Fax: 603/766-8595

BAUMER ELECTRIC LTD, 122 Spring St., Southington, CT 06489, 860/621-2121, Fax: 860/628-6280, www.baumerelectric.com

BEAMEX, INC., 2270 Northwest Pkwy., Marietta, GA 30067, 800/888-9892, Fax: 770/951-1927, www.beamex.com

BEBCO INDUSTRIES, 600 Gulf Frwy., Texas City, TX 77591, 800/OK-BEBCO, Fax: 409/938-4189, www.okbebco.com

BECKHOFF AUTOMATION LLC, 12204 Nicollet Ave. S., Minneapolis, MN 55337, 952/890-0000, Fax: 952/890-2888, www.beckhoff.com

BEI—INDUSTRIAL ENCODER DIV., 7230 Hollister Ave., Goleta, CA 93117, 800/350-2727, Fax: 805/968-3154, www.beiied.com

BELDEN ELECTRONICS, 2200 U.S. Hwy. 27 S., Richmond, IN 47374, 765/983-5200, Fax: 765/983-5294, www. belden.com

BELL TECHNOLOGY, INC., 6120 Hanging Moss Rd., Orlando, FL 32807, 407/678-6900, Fax: 407/678-0578, www.belltechinc.com

BENTLY NEVADA, 1631 Bently Pkwy S., Minden, NV 89423, 775/782-3611, Fax: 775/782-9337, www.bently.com

BERGOTECH, INC., 32 Clarissa Dr., PH 22, Richmond Hill, Ontario, L4C 9R7 Canada, 416/456-7533, Fax: 908/883-6202

BERKELEY PROCESS CONTROL, 4124 Lakeside Dr., Richmond, CA 94806, 510/243-3375, Fax: 510/222-8737, www.berkeleyprocess.com

BETTIS/EMERSON VALVE AUTOMATION, 18703 GH Circle, Waller, TX 77484, 281/463-5100, Fax: 281/463-5103, www.emersonvalveautomation.com

BI TECHNOLOGIES, 4200 Bonita Pl., Fullerton, CA 92835, 714/447-2666, Fax: 714/447-2400, www.bitechnologies.com

BI-LOK TUBE FITTINGS, 18254 Technology Dr., Meadville, PA 16335, 814/337-0380, Fax: 814/337-8468, www.bilok.com

BINDICATOR, 150 Venture Blvd., Spartanburg, SC 29306, 864/574-8060, Fax: 864/574-8063, www.bindicator.com

BINMASTER, 4200 N. 48th St., Lincoln, NE 68504, 402/434-9100, Fax: 402/434-9133, www.binmaster.com

BIZWAREDIRECT, 2450 Atlanta Highway, #1202, Cumming, GA 30040, 770/886-5878, Fax: 770/886-1390, www.bizwaredirect.com

BLACOH FLUID CONTROL INC., 601 Columbia Ave., Ste. D, Riverside, CA 92507, 909=342-3100, Fax: 909/342-3101, www.blacoh.com

BLACK & VEATCH, PO Box 8405, Kansas City, MO 64114, 913/458-2000, Fax: 913/458-2934, www.bv.com

BLANCETT, 100 E. Felix St., Suite 190, Fort Worth, TX 76115, 817/920-9998, Fax: 817/921-5282, www.blancett.com

BLUE BOX VIDEO, 3101 Bee Caves #290, Austin, TX 78746, 512/330-9990, Fax: 512/330-9996, www.bluevideo.com

BLUE MOUNTAIN QUALITY RESOURCES, INC., 208 W. Hamilton Ave., State College, PA 16801, 814/234-2417, Fax: 814/234-7077, www.coolblue.com

BLUE-WHITE INDUSTRIES, 14931 Chestnut St., Westminster, CA 92683, 714/893-8529, Fax: 714/894-9492, www.bluwhite.com

BOSCH AUTOMATION TECHNOLOGY, 7505 Durand Ave., Racine, WI 53406, 262/554-7100, Fax: 262/554-8103, www.boschat.com

BOSCH REXROTH PNEUMATICS, 1953 Mercer Rd., Lexington, KY 40511, 859/254-8031, Fax: 857/234-7077, www.boschrexrothus.com

BOURNS INC., 1200 Columbia Ave., Riverside, CA 92507, 909/781-5004, Fax: 909/781-5122, www.bourns.com

BRAY CONTROLS, 13333 Westland East Blvd., Houston, TX 77041, 281/894-5454, Fax: 281/894-0022, www.bray.com

BRINKMANN INSTRUMENTS, INC., One Cantiague Rd., Westbury, NY 11590-0207, 516/334-7500, Fax: 516/334-7506, www. brinkmann.com

BRISTOL BABCOCK, 1100 Buckingham St., Watertown, CT 06795, 860/945-2295, Fax: 860/945-2278, www.bristolbabcock.com

BRISTOL EQUIPMENT CO., PO Box 696, Yorkville, IL 60560, 630/553-7161, Fax: 630/553-5981, www.bristolequipment.com

BROOKFIELD ENGINEERING LABORATORIES, 11 Commerce Blvd., Middleboro, MA 02346, 508/946-6200, Fax: 508/946-6262, www.brookfieldengineering.com

BROOKS AUTOMATION, 15 Elizabeth Dr., Chelmsford, MA 01824, 978/262-2400, Fax: 978/262-2500, www.brooks.com

BROOKS INSTRUMENT, 407 W. Vine St., Hatfield, PA 19440, 215/362-3700, Fax: 215/362-3745, www.brooksinstrument.com

BRUEL & KJAER, 2815 Colonnades Ct., Norcross, GA 30071, 800/332-2040, Fax: 770/447-8440, www.bkhome.com

BRUKER DALTRONICS, INC., 15 Fortune Drive, Billerica, MA 01821, 978-667-9580, Fax: 978-667-5993, www.daltronics.bruker.com

BRUKER OPTICS, INC., 19 Fortune Drive, Billerica, MA 01821, 978-667-9580, Fax: 978-663-9177, www.bruker.com/optics

BURKERT, 2602 McGaw Ave., Irvine, CA 92614, 949/223-3139, Fax: 949/223-3198, www.burkert-usa.com

BURNS ENGINEERING, INC., 10201 Bren Rd. E., Minnetonka, MN 55343, 952/935-4400, Fax: 952/935-8782, www.burnsengineering.com

BW TECHNOLOGIES, 242, 3030-3rd Ave. NE, Calgary, Alberta, T2A 6T7 Canada, 403/248-9226, Fax: 403/273-3708, www.gasmonitors.com

C

C3 CONTROLS, PO Box 496, Beaver, PA 15009, 724/775-7926, Fax: 724/775-5283, www.c3controls.com

CAIG LABORATORIES, INC., 12200 Thatcher Ct., Poway, CA 92064, 858/486-8388, Fax: 858/486-8398, www.caig.com

CAL CONTROLS, INC., 1580 S. Milwaukee Ave., Libertyville, IL 60048, 847/680-7080, Fax: 847/816-6852, www.cal-controls.com

CALEX MFG. CO., INC., 2401 Stanwell Dr., Concord, CA 94520, 925/687-4411, Fax: 925/687-3333, www.calex.com

CALIBRON SYSTEMS, INC., 7861 E. Gray Rd., Scottsdale, AZ 85260, 480/991-3550, Fax: 780/998-5589, www.calibron.com

CAMSOFT CORP., 32295-8 Mission Trail #299, Lake Elsinore, CA 92530, 909/674-8100, Fax: 909/674-3110, www.camsoftcorp.com

CANARY LABS, Brownstone Bldg., Martinsburg, PA 16662, 814/793-3770, Fax: 814/793-3145, www.canarylabs.com

CAPE SOFTWARE, INC., 650 N. Sam Houston Pkwy. #313, Houston, TX 77060, 713/661-7100, Fax: 281/448-2607, www.capesoftware.com

CAPITAL CONTROLS CO., 3000 Advance Lane, Colmar, PA 18915, 215/997-4000, Fax: 215/997-4062, www.capitalcontrols.com

CARDINAL SCALE MFG. CO., 203 E. Daugherty, Webb City, MO 64870, 417/673-4631, Fax: 417/673-5001, www.ardinalscale.com

CARLO GAVAZZI AUTOMATION COMPONENTS, 750 Hastings Ln., Buffalo Grove, IL 60089, 847/465-6100, Fax: 847/465-7373, www.gavazzionline.com

CARLON, 25701 Science Park Dr., Cleveland, OH 44122, 800/3-CARLON, Fax: 216/831-5579, www.carlon.com

CAROLIN CORP., PO Box 2649, Woburn, MA 01888, 781/935-0146, Fax: 781/937-3499

CASHCO, 607 W. 15th St., Ellsworth, KS 67439, 785/472-4461, Fax: 785/472-3539, www.cashco.com

CCI, 22591 Avenida Empresa, Rancho Santa Margarita, CA 92688, 949/858-1877, Fax: 949/858-1878, www.ccivalve.com

CEA INSTRUMENTS, INC., 16 Chestnut St., Emerson, NJ 07630, 201/967-5660, Fax: 201/967-8450, www.ceainstr.com

CHINO WORKS AMERICA INC., 18005 Savarona Way, Carson, CA 90746, 323/321-3943, Fax: 310/532-7195, www.chinoamrica.com

CHRISTENSEN DISPLAY, 8126-B 204th Ave. SE, Preston, WA 98050, 425/222-3800, www.christensendisplay.com

CHROMALOX—PRECISION HEAT & CONTROL, 701 Alpha Dr., 3rd Fl., Pittsburgh, PA 15238, 412/967-3800, Fax: 412/967-5148, www.chromalox.com

CI TECHNOLOGIES, 4828 Parkway Plaza Blvd. #140, Charlotte, NC 28217, 704/329-3838, Fax: 704/329-3839, www.citect.com

CICOIL CORP., 24960 Avenue Tibbitts, Valencia, CA 91350, 661/295-1295, Fax: 661/295-0813, www.cicoil.com

CIDRA CORP., 50 Bames Park N., Wallingford, CT 06492, 203/626-3421, Fax: 203/294-4211, www.cidra.com

CIMLOGIC, INC., 402 Amherst St., Suite 203, Nashua, NH 03063, 603/881-9918, Fax: 603/595-0381, www.cimlogic.com

CIRCLE SEAL CONTROLS, 2301 Wardlow Cir., Corona, CA 92880, 909/270-6280, Fax: 909/270-6201, www.circle-seal.com

CIRCUIT COMPONENTS, INC., 2400 S. Roosevelt, Tempe, AZ 85282, 480/967-0624, Fax: 480/967-9385, www.surgecontrol.com

CIRRONET, INC., 5375 Oakbrook Pkwy., Norcross, GA 30093, 678/684-2000, Fax: 678/684-2001, www.cirronet.com

CISCO SYSTEMS INC., 170 W. Tasman Dr., San Jose, CA 95134, 800/553-6387, Fax: 408/527-8048, www.cisco. com

CITADEL COMPUTER CORP., 29 Armory Rd., Milford, NH 03055, 603/672-5500, Fax: 603/672-5590, www.citadelcomputer.com

CITECT INC., 30000 Mill Creek Ave.,Ste. 300, Alpharetta, GA 30022, 770/521-7511, Fax: 770/521-7512, www.citect.com

CITEL, INC., 1111 Parkcentre Blvd, Suite 340, Miami, FL 33169, 305/621-0022, Fax: 305/621-0766, www.citelprotection.com

CJM CONSULTING, 107 Oak Newel, Peachtree City, GA 30269, 678/637-9062, Fax: 770/487-1712

CLAROSTAT SENSORS & CONTROLS, INC., 12055 Rojas Dr., Suite K, El Paso, TX 79936, 915/858-2632, Fax: 915/872-3333, www.clarostat.com

CLEVELAND MOTION CONTROLS, 7550 Hub Pkwy., Cleveland, OH 44125, 216/642-2178, Fax: 216/642-2100, www.cmccontrols.com

CLIFFORD OF VERMONT, INC., Rte. 107, PO Box 51, Bethel, VT 05032, 802/234-9921, Fax: 802/234-5006, www.cliffordvt.com

CLIPPARD INSTRUMENTS LAB., INC., 7390 Colerain Ave., Cincinnati, OH 45235, 513/521-4261, www.clippard.com

COGNEX CORP., One Vision Dr., Natick, MA 01760, 800/677-2646, Fax: 508/650-3344, www.cognex.com

COLDER PRODUCTS CO., 1001 Westgate Dr., St. Paul, MN 55114, 800/444-2474, Fax: 651/645-5404, www.colder.com

COLE-PARMER INSTRUMENT CO., 625 E. Bunker Ct., Vernon Hills, IL 60061, 800/323-4340, Fax: 847/247-2929, www.coleparmer.com

COLLINS INSTRUMENT CO., Inc., PO Box 938, Analeton, TX 77516, 979/849-8266, Fax: 979/848-0783, www.collinsinst.com

COMARK CORP., 93 West St., Medfield, MA 02052, 508/359-8161, Fax: 508/359-2267, www.comarkcorp.com

COMPUTER DYNAMICS, 7640 Pelham Rd., Greenville, SC 29615, 864/627-8800, Fax: 864/675-0106, www.cdynamics

COMPUTER INSTRUMENTS CORP., 1000 Shames Dr., Westbury, NY 11590, 516/876-8400, Fax: 516/876-9153, www.computerinstruments.com

CONAX BUFFALO TECHNOLOGIES, 2300 Walden Ave., Buffalo, NY 14225, 716/684-4500, Fax: 716/684-7433, www.conaxbuffalo.com

CONBRACO INDUSTRIES, INC., 701 Matthews-Mint Hill Rd., Matthews, NC 28105, 704/841-6000, Fax: 704/ 841-6020, www.conbraco.com

CONCOA, 1501 Harpers Rd., Virginia Beach, VA 23454, 757/422-8330, Fax: 757/422-3125, www.concoa.com

CONDEC, 3 Simm Lane, Newtown, CT 06470, 888/295-8475, Fax: 203/364-1556, www.4condec.com

CONDUANT CORP., 1501 S. Sunset St., Ste. C, Longmont, CO 80501, 888/497-7327, Fax: 303/485-5104, www.conduant.com

CONSILIUM US, INC., 59 Porter Rd., Littleton, MA 01460, 978/486-9800, Fax: 978/486-0170, www.consiliumus.com

CONSTANT POWER MFG., INC., 600 Century Plaza, Bldg. 140, Houston, TX 77073-6033, 281/821-3211, Fax: 281/821-6093, www.constantpowermfg.com

CONTEC MICROELECTRONICS, 744 S. Hillview Dr., Milpitas, CA 95035, 800/888-8884, Fax: 408/719-6750, www.contecusa.com

CONTEMPORARY CONTROLS, 2431 Curtiss St., Downers Grove, IL 60515, 630/963-7070, Fax: 630/963-0109, www.ccontrols.com

CONTREX, INC., Box 9000, Maple Grove, MN 55311-9000, 612/424-7800, Fax: 612/424-8734, www.contrexinc.com

CONTROL CHIEF, 200 Williams St., Bradford, PA 16701, 814/362-6811, Fax: 814/368-4133, www.controlchief.com

CONTROL CONCEPTS, 328 Water St., Binghamton, NY 13901, 607/724-2484, Fax: 607/722-8713, www.control-concepts.com

CONTROL FOR LESS, 101 Copperwood Way, Suite L, Oceanside, CA 92054, 760/433-7633, Fax: 760/433-6859, www.controlforless.com

CONTROL & MEASUREMENT INTL., INC., 421 Homewood Blvd., Delray Beach, FL 33445, 561/330-8144, Fax: 561/330-8134, www.cmi-temp.com

CONTROL INSTRUMENTS CORP., 25 Law Dr., Fairfield, NJ 07004, 973/575-9114, Fax: 973/575-0013, www.controlinstruments.com

CONTROL MICROSYSTEMS, 28 Steacie Dr., Kanata, Ontario, K2K 2A9 Canada, 613/591-1943, Fax: 613/591-1022, www.controlmicrosystems.com

CONTROL SYSTEMS INTERNATIONAL, 8040 Nieman Rd., Lenexa, KS 66214, 913/599-5010, Fax: 913/599-5013, www.csiks.com

CONTROL TECHNOLOGY CORP., 25 South St., Hopkinton, MA 01748, 508/435-9595, Fax: 508/435-2373, www.ctc-control. com

CONTROLAIR, INC., 8 Columbia Dr., Amherst, NH 03031, 800/216-3636, Fax: 603/889-1844, www.controlair.com

CONTROL.COM, INC., 134 Flanders Rd., Westborough, MA 01581, 508/898-9111, Fax: 508/621-3614, www.control.com

CONTROLOTRON, 155 Plant Ave., Hauppauge, NY 11788, 631/231-3600, Fax: 631/231-3334, www.controlotron.com

CONTROLSOFT, INC., 14077 Cedar Rd., Suite 200, Cleveland, OH 44118, 216/397-3900, Fax: 216/381-5001, www.controlsoftinc.com

CONVEYOR COMPONENTS CO., 130 Seltzer Rd., Croswell, MI 48422, 800/233-3233, Fax: 810/679-4510, www.conveyorcomponents.com

COOPER BUSSMANN, 114 Old State Rd., Ellisville, MO 63021, 636/527-1642, Fax: 636/527-1340, www.bussmann.com

CORECO IMAGING, 7075 Place Robert-, Joncas, Ste. 142, Saint-Laurent, QB, H4M, 2Z2 Canada, 51/333-1301, Fax: 514-333-1388, www.imaging.com

COSA INSTRUMENT CORP., 55 Oak St., Norwood, NJ 07648, 201/767-6600, Fax: 201/767-8604

COSENSE, INC., 155 Rice Field Ln., Hauppauge, NY 11788, 631/231-0735, Fax: 631/231-0838, www.cosence.com

CROMPTON INSTRUMENT, INC., 1640 Airport Rd., Suite 109, Kennesaw, GA 30144, 770/425-8903, Fax: 770/423-7194, www.crompton-instruments.com

CROUSE-HINDS, PO Box 4999, Syracuse, NY 13221, 315/477-5110, Fax: 315/477-5118, www.crouse-hinds.com

CROUSE-HINDS MOLDED PRODUCTS, 4758 Washington St., LaGrange, NC 28551, 252/566-3014, Fax: 252/556-9337

CRYSTAL ENGINEERING, 1450 Madonna Rd., San Luis Obispo, CA 93405, 800/444-1850, Fax: 805/595-5466, www.crystalengineering.net

CTC PARKER AUTOMATION, 50 W. TechneCenter Dr., Milford, OH 45150, 513/831-2340, Fax: 513/831-5042, www.ctcusa.com

CTI ELECTRONICS CORP., 110 Old South Ave., Stratford, CT 06615, 203/386-9779, Fax: 203/378-4986, www.ctielectronics.com

CURTISS-WRIGHT FLOW CONTROL CORP., 1966 E. Broadhollow Rd., E. Farmingdale, NY 11735, 631/293-3800, Fax: 631/293-4949, www.cwfc.com

CUTLER-HAMMER, 4201 N. 27th St., Milwaukee, WI 53216, 414/449-6000, Fax: 414/449-7319, www.cutlerhammer.eatoncom

CYBOSOFT, 2868 Prospect Park Dr., Suite 300, Rancho Cordova, CA 95670, 916/631-6313, Fax: 916/631-6312, www.cybosoft.com

D

DAISY DATA, INC., 2850 Lewisberry Rd., York Haven, PA 17370, 717/932-9999, Fax: 717/932-8000, www.daisydata.com

DANAHER CONTROLS, 1675 Delany Rd., Gurnee, IL 60031, 800/873-8731, Fax: 847/662-6633, www.dancon.com

DANIEL MEASUREMENT AND CONTROL, PO Box 19097, Houston, TX 77224, 713/827-5184, Fax: 713/827-4360, www.danielind.com

DATA INDUSTRIAL CORP., 11 Industrial Dr., Mattapoisett, MA 02739, 508/758-6390, Fax: 508/758-4057, www.dataindustrial.com

DATAFORTH CORP., 3331 E. Hemisphere Loop, Tucson, AZ 85706, 520/741-1404, Fax: 520/741-0762, www.dataforth.com

DATALUX CORP., 155 Aviation Dr., Winchester, VA 22602, 540/662-1500, Fax: 540/662-7385, www.datalux.com

DATANET QUALITY SYSTEMS, 24567 Northwestern Hwy., 4th Fl., Southfield, MI 48075, 248/357-2200, Fax: 248/357-4933, www.winspc.com

DATARADIO, 299 Johnson Ave., Waseca, MN 56093, 507/833-6740, Fax: 507/833-6748, www.dataradio.com

DATA TRANSLATION, 100 Loche Dr., Marlboro, MA 01752, 800/525-8528, Fax: 508/481-8620, www.datatranslation.com

DATATRONICS, INC., 28151 Hwy 74, Romoland, CA 92585, 909/928-7700, Fax: 909/928-7701, www.datatronics.com

DATAVIEWS CORP., 47 Pleasant St., Northhampton, MA 01060, 413/586-4144, Fax: 413/586-3805, www.dvcorp.com

DATEL, INC., 11 Cabot Blvd., Mansfield, MA 02048, 508/339-3000, Fax: 508/339-6356, www.datel.com

DAYTRONIC CORP., 2211 Arbor Blvd., Dayton, OH 45439, 937/293-2566, Fax: 937/293-2586, www.daytronic.com

DELTA F. CORP., 4 Constitution Way, Woburn, MA 01801, 781/935-4600, Fax: 781/938-0531, www.delta-f.com

DELTRONIC, INC., 290 Wissahickon Ave., North Wales, PA 19454, 215/699-2310, www.deltronic.com

DENSITRON CORP., 10430-2 Pioneer Blvd., Santa Fe Springs, CA 90670, 562/941-5000, Fax: 562/941-5757, www.densitron.com

DESERT MICROSYSTEMS, INC., 3381 Chicago Ave., Riverside, CA 92507, 800/633-0448, Fax: 909/982-8506, www.desertmicrosys.com

DETCON, INC., 3200 A1 Research Forest, The Woodlands, TX 77381, 281/367-4100, Fax: 281/292-2860, www.detcon.com

DETECTOR ELECTRONICS CORP., 6901 W. 110th St., Minneapolis, MN 55438, 800/468-3244, Fax: 952/829-8745, www.detronics.com

DEVAR, INC., 706 Bostwick Ave., Bridgeport, CT 06605-2396, 203/368-6751, Fax: 203/368-3747, www.devarinc.com

DEZURIK/COPES—VULCAN, A Unit of SPX Corp., 250 Riverside Ave. N., Sartell, MN 56377, 320/259-2000, Fax: 320/259-2227, www.dezurikcopesvulcan.com

DGH CORP., PO Box 5638, Manchester, NH 03108, 603/622-0452, Fax: 603/622-0487, www.dghcorp.com

DH INSTRUMENTS, INC., 1905 W. 3rd St., Tempe, AZ 85281, 480/967-1555, Fax: 480/968-3574, www. dhinstruments.com

DIAMOND POWER INTL., INC., 2600 E. Main St., Lancaster, OH 43130, 740/687-4277, Fax: 740/687-4304, www.diamondpower.com

DIETRICH STANDARD, INC., 5601 N. 71st, Boulder, CO 80301, 720/622-2626, Fax: 303/530-7064, www.annubar.com

DIGI INTERNATIONAL, 11001 Bren Rd. E., Minnetonka, MN 55343, 952/912-3361, Fax: 952/912-4953, www.digi.com

DIGITAL SYSTEMS ENGINEERING, 4325 S. 34th St., Phoenix, AZ 85040, 602/426-8588, Fax: 602/426-8688, www.digitalsys.com

DIGITAL VIEW INC., 18440 Technology Dr., Morgan Hills, CA 95037, 408/782-7773, www.digitalview.com

DIGITAL WIRELESS CORP., 1 Mega Way, Norcross, GA 30093, 770/564-5540, Fax: 770/564-5541, www.digital-wireless.com

DIMENSION TECHNOLOGIES, INC., 315 Mt. Read Blvd., Rochester, NY 14611, 716/436-3530, Fax: 716/436-3280, www.dti3d.com

DIONEX CORP., 1228 Titan Way, Sunnyvale, CA 94086, 408/737-0700, Fax: 408/739-4398, www.dionex.com

DIRECT MEASUREMENT CORP., 4040 Coriolis Way, Longmont, CO 80026, 303/702-7400, Fax: 303/702-7488, www.directmeasurement.com

DIVELBISS CORP., 9778 Mt. Golend Rd., Fredericktown, OH 43019, 740/694-9015, Fax: 740/694-9035, www.divelbiss.com

DJSCIENTIFIC, 5200 Dickey-John Rd., Auburn, IL 62563, 217/438-3371, Fax: 217/438-2609, www.djscientific.com

DORIC INSTRUMENTS, 4750 Viewridge Ave., San Diego, CA 92123, 888/423-6742, Fax: 858/569-8474, www.doric-vas.com

DPL SYSTEMS ENGINEERING, 1216 Sand Cove Rd., Saint John, New Brunswick, E2M 5V8 Canada, 506/635-1055, Fax: 506/ 635-1057, www.dpl.ca

DRAEGER SAFETY, INC., 101 Technology Dr., Pittsburgh, PA 15275, 412/787-8383, Fax: 412/787-2207, www.draeger.net

DRANETZ-BMI, 1000 New Durham Rd., Edison, NJ 08818, 800/372-6832, Fax: 732/248-1834, www.dranetz-BMI.com

DRESSER, INC., Masoneilan Operations, Dresser Flow Control, 85 Bodwell St., Avon, MA 02322, 508/586-4600, Fax: 508/427-8971, www.masoneilan.com

DRESSER INSTRUMENT DIV., 250 E. Main St., Stratford, CT 06614, 203/378-8281, www.dresserinstruments.com

DRIVE CONTROL SYSTEMS, 6111 Blue Circle Dr., Minnetonka, MN 55343, 952/930-0196, Fax: 952/930-0180, www.drivecontrolsystems

DRUCK, INC., 4 Dunham Dr., New Fairfield, CT 06812, 203/746-0400, Fax: 203/746-2494, www.pressure.com

DUTECH, PO Box 964, Jackson, MI 49204, 800/248-1632, Fax: 517/750-4740, www.dutec.net

DWYER INSTRUMENTS, 102 Indiana Hwy., Suite 212, Michigan City, IN 46360, 219/879-8000, Fax: 219/872-9057, www.dwyer-inst.com

DYNAMIC DISPLAYS, INC., 1625 Westgate Rd., Eau Claire, WI 54703, 715/835-9440, Fax: 715/835-2436, www.dynamicdisplay.com

DYNAPRO, 800 Carleton Ct., Annacis Island, New Westminster, BC, U3M 6L3, Canada, 604/521-3962, Fax: 604/521-4629, www.dynapro.com

DYNASONICS, 8635 Washington Ave., Racine, WI 53406, 262/639-6770, Fax: 262/639-2267, www.dynasonics.com

E

E INSTRUMENTS GROUP, 172 Middletown Blvd., #B201, Langhome, PA 19047, 215/750-1212, Fax: 215/750-1399, www.einstrumentsgroup.com

E-MON CORP., One Oxford Valley, Suite 418, Langhorne, PA 19047, 800/334-3666, Fax: 215/752-3094, www. emon.com

E-T-A CIRCUIT BREAKERS, 1551 Bishop Ct., Mt. Prospect, IL 60056, 847/827-7600, Fax: 847/827-7655, www.etacbc.com

EASON TECHNOLOGY, 241 Center St., Heardsburg, CA 95448, 707/433-2854, Fax: 707/433-3706, www.eason. com

EATON CORP., 15 Durant Ave., Bethel, CT 06801, 800/736-1557, Fax: 203/796-6196, www.eaton.com

ECHELON CORP., 4015 Miranda Ave., Palo Alto, CA 95050, 650/855-7456, Fax: 650/843-1242, www.echelon.com

ECHIP, INC., 724 Yorklyn Rd., Hockessin, DE 19702, 302/239-5429, Fax: 302/239-6227, www.echip.com

ECKARDT—INVENSYS FLOW CONTROL, 1790 Satelite Blvd., Suite 100-04, Duluth, GA 30097, 678/474-1500, Fax: 678/474-1515

ECOM INSTRUMENTS, INC., 2000 Dairy Ashford, Suite 3061, Houston, TX 77077, 281/496-5930, Fax: 281/496-2321, www.ecom-ex.com

ECT INTERNATIONAL, INC., 4100 N. Calhoun Rd., Brookfield, WI 53005, 262/781-1511, Fax: 262/781-8411, www.ecti.com

EDGETECH, 455 Fortune Blvd., Milford, MA 01757, 800/276-3729, Fax: 508/634-3010, www.edgetech.com

EDO ELECTRO-CERAMIC PRODUCTS, 2645 South 300 West, Salt Lake City, UT 84115, 801/461-9259, Fax: 801/484-3301, www.edocorp.com

EDWARD VOGT VALVE CO-INVENSYS FLOW CONTROL, 1900 S. Saunders St., Raleigh, NC 27603, 800/225-6989, Fax: 919/831-3254

EL-O-MATIC, 135 English St., Hackensack, NJ 07601, 201/489-5550, Fax: 201/489-9171, www.elomaticusa.com

ELCON INSTRUMENTS, 2700 Garden Rd., Swuanee, GA 30024, 770/271-5519, Fax: 770/271-5049, www.elconinst.com

ELDRIDGE PRODUCTS, INC., 2700 Garden Rd., Monterey, CA 93940, 800/321-3569, Fax: 831/648-7780, www.epiflow.com

ELECTRIC POWER & HEAT CO., 65 White Oak Dr., Smithfield, NC 27577, 919/934-9448, Fax: 919/934-0345

ELECTRO CAM CORP., 13647 Metric Rd., Roscoe, IL 61115, 815/389-2620, Fax: 815/389-3304, www.electrocam.com

ELECTRO SENSORS, INC., 6111 Blue Circle Dr., Minnetonka, MN 55343, 952/930-0100, Fax: 952/930-0130, www.electrosensors.com

ELECTRO STANDARDS LABORATORIES, 36 Western Industrial Dr., Cranston, RI 02921, 401/943-1164, Fax: 401/946-5790, www.electrostandards.com

ELECTRO-MECH COMPONENTS, INC., 1826 Floradale Ave., South El Monte, CA 91733, 626/442-7180, Fax: 626/350-8070, www.electromechcomp.com

ELECTRONIC SENSORS, INC., 1611 W. Harry, Wichita, KS 67213, 800/886-2511, Fax: 316/267-2819, www.leveldevil.com

ELECTROSWITCH, 180 King Ave., Weymouth, MA 02188, 781/607-3303, Fax: 781/335-4253, www.electroswitch.com

ELECTROSWITCH ELECTRONIC PRODUCTS, 2010 Yonkers Rd., Raleigh, NC 27604, 888/768-2797, Fax: 800/909-9171, www.electro-nc.com

ELGAR, 9250 Brown Deer Rd., San Diego, CA 92121, 858/450-0085, Fax: 858/458-0267, www.elgar.com

ELIPSE SOFTWARE, 40190 Jarvis Gray Ln., Avon, NC 27915, 252/995-6885, Fax: 252/995-5686, www.elipse-software.com

ELO TOUCHSYSTEMS, INC., 6500 Kaiser Dr., Fremont, CA 94555, 800/ELO-TOUCH, Fax: 510/790-0627, www.elotouch.com

ELPRO TECHNOLOGIES LTD., 9/12 Billabong St., Stafford City, QLD,Australia 4053, 61 7 3352 4533, Fax: 61 7 33 524577, www.elprotech.com

EMATION, 89 Forbes Blvd., Mansfield, MA 02048, 508/337-9200, Fax: 508/337-9201, www.emation.com

EMCO FLOW SYSTEMS, 600 Diagonal Hwy., Longmont, CO 80501, 303/651-0550, Fax: 303/678-7152, www.emcoflow.com

EMERSON PROCESS MANAGEMENT ROSEMOUNT ANALYTICAL, 6565 P. Davis Industrial Pkwy, Solon, OH 44139, 440/914-1261, Fax: 440/914-1271, www.raihome.com

EMERSON PROCESS MANAGEMENT, PROCESS SYSTEMS, 12301 Research Blvd., Austin, TX 78759, 512/418-7505, www.easydeltav.com

EMERSON PROCESS MANAGEMENT, ROSEMOUNT DIV., 8200 Market Blvd., Chanhassen, MN 55317, 800/999-9307, Fax: 952/949-7001, www.rosemount.com

EMERSON PROCESS MANAGEMENT, FISHER CONTROLS, 205 S. Center St., Marshalltown, IA 50158, 641/754-3288, Fax: 641/754-2179, www.fisher.com

EMERSON PROCESS MANAGEMENT, MICRO MOTION, 7070 Winchester Circle, Boulder, CO 80301, 800/760/8119, www.micromotion.com

EMERSON PROCESS MANAGEMENT, ROSEMOUNT ANALYTICAL INC., 2400 Barrance Pkwy., Irvine, CA 92606, 800/854-8257, Fax: 949/863-9159, www.raihome.com.com

EMERSON PROCESS MANAGEMENT, SAAB ROSEMOUNT TANK GAUGING, 10700 Hammerly Blvd.,Ste. 115, Houston, TX 77043, 713/722-9199, Fax: 713/722-9115, www.saabradar.com

EMERSON PROCESS MANAGEMENT, VALVE AUTOMATION DIV., PO Box 508, Waller, TX 77484, 281/727-5300, Fax: 281/727-5303, www.emersonprocess.com

ENDEVCO, 30700 Rancho Viejo Rd., San Juan Capistrano, CA 92675, 949/493-8181, Fax: 949/661-7231, www.endevco.com

ENDICOTT RESEARCH GROUP INC., 2601 Wayne St., Endicott, NY 13760, 607/754-9187, Fax: 607/754-9255, www.ergpower.com

ENDRESS + HAUSER, INC., 2350 Endress Place, Greenwood, IN 46143, 317/535-1391, Fax: 317/535-2171, www.us.endress.com

ENDRESS + HAUSER SYSTEMS & GAUGING, INC., 5834 Peachtree East, Norcross, GA 30096, 770/447-9202, Fax: 770/622-8939, www.systems.endress.com

ENEA OSE SYSTEMS, 5949 Sherry Ln., Suite 625, Dallas, TX 75225, 214/346-9339, Fax: 214/346-9344, www. enea.com

ENERPRO, INC., 5780 Thornwood Dr., Goleta, CA 93117, 805/683-2114, Fax: 805/964-0798, www.enerpro.thomasregister.com

ENRAF, INC., 4333 W. Sam Houston Pkwy N., Houston, TX 77043, 832/467-3422, Fax: 832/467-3441, www.enrafinc.com

ENTERTRON INDUSTRIES, INC., 3857 Orangeport Rd., Gasport, NY 14067, 716/772-7216, Fax: 716/772-2604, www.entertron.com

ENTIVITY, 935 Technology Dr., Suite 200, Ann Arbor, MI 48108, 734/205-5000, Fax: 734/205-5100, www.entivity.com

ENTRAN DEVICES, INC., 10 Washington Ave., Fairfield, NJ 07004, 973/227-1002, Fax: 973/227-6865, www.entran.com

ENTRELEC, INC., 1950 Hurd Dr., Irving, TX 75038, 800/431-2308, Fax: 800/862-5066, www.entrelec.com

ENTRON COMPUTER CORP., 9001 Airport Blvd., Suite 409, Houston, TX 77061, 713/941-7007, Fax: 713/941-3852, www.entron.com

EPLAN, 16650 Bluemound Rd., Suite 600, Brookfield, WI 53005, 262/789-0428, Fax: 262/789-0428, www.eplan.org

ERGOTRON, 1181 Trapp Rd., St. Paul, MN 55121, 800/888-8458, www.ergotron.com

ESSENTIALS CONTROL, INC., 128 Elmore Dr., Acton, Ontario, L7J 1T2 Canada, 519/853-3830, Fax: 519/853-5073, www.essentials-control.com

E-T-A CIRCUIT PROTECTION & CONTROL, 1551 Bishop Ct., Mt. Prospect, IL 60056, 847/827-7600, Fax: 847/827-7655, www.e-t-a.com

ETI SYSTEMS, INC., 2251 Las Palmas Dr., Carlsbad, CA 92009, 760/929-0749, Fax: 760/929-0748, www.etisystems.com

EUROTHERM CHESSELL, 8601 Aero Dr., San Diego, CA 92123, 800/801-5099, Fax: 858/514-0426, www.chessell.com

EUROTHERM CONTROLS, INC., 741-F Miller Dr., Leesburg, VA 20175, 703/443-0000, Fax: 703/669-1300, www.eurotherm.com

EUTECH INSTRUMENTS, 925 E. Bunker Ct., Vemon Hills, IL 60061, 888/538-4710, Fax: 847/327/2971, www.eutechinstamerica.com

EVANS CONSOLES, INC., 1616 27th Ave. N.E., Calgary, Alberta, T2E 8W4 Canada, 403/717-3009, Fax: 403/717-3320, www.evansonline.com

EVEREST INTERSCIENCE, 1891 N. Oracle Rd., Tucson, AZ 85705, 520/797-0927, Fax: 520/797-0938, www.everestinterscience.thomasregister.com

EXERGEN CORP., 51 Water St., Watertown, MA 02472, 617/923-9900, Fax: 617/923-9911, www.exergen.com

EXOR ELECTRONIC R&D, INC., 3420 Fairlane Farms Rd., Wellington, FL 33414, 561/753-2250, Fax: 561/753-2291, www.exor-rd.com

EXPERTUNE, INC., 4734 Sonseeahray Dr., Hubertus, WI 53033, 262/628-0088, Fax: 262/628-0087, www.expertune.com

EXPO-TELEKTRON SAFETY SYSTEMS, INC., 67 E. Washington St., Chagrin Falls, OH 44022, 440/247-5314, Fax: 440/247-5409, www.expotelektron.com

EXTECH INSTRUMENTS, 285 Bear Hill Rd., Waltham, MA 02451, 781/890-7440, Fax: 781/890-7864, www.extech.com

F

FACTORYSOFT, INC., 89 Forbes Blvd., Mansfield, MA 02048, 508/337-9200, Fax: 508/337-9201, www.emation.com

FAIRBANKS SCALES, 821 Locust, Kansas City, MO 64106, 816/471-0231, Fax: 816/471-5951, www.fairbanks.com

FAIRCHILD INDUSTRIAL PRODUCTS CO., 3920 West Point Blvd., Winston-Salem, NC 27103, 336/659-3400, Fax: 336/659-9323, www.fairchilproducts.com

FASTECH, A SUBSIDARY OF BROOKS AUTOMATION, 15 Elizabeth Dr., Chelmsford, MA 01824, 978/262-2400, Fax: 978/262- 2500, www.fastech.com

FASTRAK SOFTWORKS, INC., 6659 W. Mill Road, Milwaukee, WI 53218, 414/358-8088, Fax: 414/358-8066, www.fast-soft.com

FLUID COMPONENTS INTL., 1755 La Costa Meadows Dr., San Marcos, CA 92069, 760/744-6950, Fax: 760/736-6250, www.fluidcomponents.com

FEDERAL PRODUCTS CO., 1144 Eddy St.; PO Box 9400, Providence, RI 02940, 401/784-3100, Fax: 401/784-3246, www.fedgage.com

FESTO CORP., 395 Moreland Rd., Hauppauge, NY 11788, 631/435-0800, Fax: 631/435-8026, www.festo-usa.com

FIBOX ENCLOSURES, 6675 Santa Barbara, Elkridge, MD 21075, 888/FIBOXUS, Fax: 410/496-3298, www.fiboxusa.com

FIELDBUS FOUNDATION, 9390 Research Blvd. Suite 11-250, Austin, TX 78759, 512/794-8890, Fax: 512/794-8893, www.fieldbus.org

FIELDSERVER TECHNOLOGIES, 1991 Tarob Ct., Milpitas, CA 95035, 408/262-2299, Fax: 408/262-9042, www.fieldserver.com

FIKE CORP., 704 South 10th St., Blue Springs, MO 64015, 816/229-3405, Fax: 816/228-9277, www.fike.com

FISHER CONTROLS INTERNATIONAL, INC., 205 S. Center St., Marshalltown, IA 50158, 641/754-3011, Fax: 641/754-2830

FISHER-ROSEMOUNT, 12001 Technology Dr., Eden Prairie, MN 55344, 952/828-3180, Fax: 952/828-3033, www.assetweb.com

FISHER-ROSEMOUNT SYSTEMS, 8301 Cameron Rd., Austin, TX 78754, 512/835-2190, Fax: 512/418-7505, www.frco.com

FLEX-CORE, 6625 McVey Blvd., Columbus, OH 43235, 614/889-6152, Fax: 614/876-8538, www.flex-core.com

FLIR SYSTEMS INC., 16 Esquire Rd., North Billerica, MA 01862, 978/901-8000, Fax: 978/901-8887, www.flirthermography.com

FLO-TORK, INC., 1701 N. Main St., Orrville, OH 44667, 330/682-0010, Fax: 330/683-6857, T.M. Leaver, executive@flo-tork.com, www.flo-tork.com

FLOW AUTOMATION, 9303 W. Sam Houston Pkwy. S., Houston, TX 77099-5298, 713/272-0404, Fax: 713/272-2272, www.flowautomation.com

FLOW RESEARCH, 27 Water St., Wakefield, MA 01880, 781/245-3200, Fax: 781/224-7552, www.flowresearch.com

FLOW TECHNOLOGY, INC., 4250 E. Broadway, Phoenix, AZ 85040, 602/437-1315, Fax: 602/437-4459, www.ftimeters.com

FLOW-TECH INC., 50 Scott Adarn, #212, Hunt Valley, MD 21030, 410/666-3200, Fax: 410/666-3631, www.flowtechnoline.com

FLOW-TEK, INC., a subsidiary of Bray Intl., Inc., 7404 Fairfield, Columbia, SC 29203, 803/754-8201, Fax: 803/754-2501

FLOWLINE, 10500 Humbolt St., Los Alamitos, CA 90720, 562/598-3015, Fax: 562/431-8507, www.flowline.com

FLOWSERVE, FLOW CONTROL DIV., PO Box 2200, Springville, UT 84663, 801/489-8611, Fax: 801/489-3719, www.flowserve.com

FLUID COMPONENTS INTL., 1755 La Costa, Meadows Dr., San Marcos, CA 92078, 760/744-6950, Fax: 760/736-6250, www.fluidcomponents.com

FLUID METERING, INC., 5 Aerial Way, Suite 500, Syosset, NY 11791, 800/223-3388, Fax: 516/624-8261, www.fmipump.com

FLUKE CORP., PO Box 9090, Everett, WA 98206, 800/44-Fluke, Fax: 425/446-5116, Sales & Applications, fluke-info@fluke.com, www.fluke.com

FMC BLENDING & TRANSFER, 20 N. Wacker Dr., Suite 1300, Chicago, IL 60606, 805/495-7111, Fax: 805/379-3365, www.fmcblending-transfer.com

FORMOSA USA, INC., 21540 Prairie St., Unit A, Chatsworth, CA 91311, 818/407-4965, Fax: 818/407-4966, www.goformosa.com

FOSS NIRSYSTEMS, 12101 Tech Rd., Silver Spring, MD 20904, 301/680-9600, Fax: 301/236-0157, www.foss-nirsystems.com

FOX INDUSTRIES, 505 Mayock Rd., Suite A4, Gilroy, CA 95020, 408/847-2090, Fax: 408/847-1806

FOX THERMAL INSTRUMENTS, 399 Reservation Road, Marina, CA 93933, 831/384-4300, Fax: 813/384-4312, www.foxthermalinstruments.com

FOXBORO CO., The, 33 Commercial St., Foxboro, MA 02035, 508/549-6240, Fax: 508/549-4834, www.foxboro.com

FRANEK TECHNOLOGIES, INC., 13821 Newport Ave., Suite 100, Tustin, CA 92780, 714/734-6957, Fax: 714/544-6957, www. franek-tech.com

FRANK W. MURPHY MFR., PO Box 470248, Tulsa, OK 74147, 918/627-3550, Fax: 918/664-6146, www.fwmurphy.com

FUJI ELECTRIC CORP. OF AMERICA, Park 80 West, Plaza II, Saddle Brook, NJ 07663, 201/712-0555, Fax: 201/368-8258, www.fujielectric.com

FUJIKIN OF AMERICA, INC., 4 Alsan Way, Little Ferry, NJ 07643, 201/641-1119, Fax: 201/641-1137, www.fujikin. com

FURNESS CONTROLS, INC., 3801-A Beam Rd., Charlotte, NC 28217, 800/898-5325, Fax: 704/357-1103, www.furnesscontrols.com

G

G&L MOTION CONTROL, 672 S. Military Rd., Fond du Lac. WI 54935, 920/921-7100, Fax: 920/906-7669, www.glcontrols.com

GALIL MOTION CONTROL, 203 Ravendale Dr., Mountain View, CA 94043, 650/967-1700, Fax: 650/967-1751, www.galilmc.com

GARRETTCOM INC., 213 Hammond Ave., Fremont, CA 94539, 510/438-9071, Fax: 510/438-9072, www.garrettcom.com

GAUGING SYSTEMS INC., 910 Industrial Blvd., Ste. A, Sugarland, TX 77478, 281/980-3999, Fax: 281/980-6929, www.gaugingsystemsinc.com

GAUMER PROCESS, 13616 Hempstead Rd., Houston, TX 77040, 800/460-5200, Fax: 800/460-1444, www.gaumer.com

GE CONTINENTAL CONTROLS, 10,000 Richmond, Houston, TX 77042, 713/978-4401, Fax: 713/978-4444, www.geindustrial.com

GE DRUCK, 4 Dunham Dr., New Fairfield, CT 06812, 203/746-0400, Fax: 203/746-2494, www.pressure.com

GE ENERGY, 1631 Bently, Parkway South, Minden, NV 89423, 775/215-2226, Fax: 775/215-2864, www.geenergy.com

GE FANUC AUTOMATION, Rte. 29N & Rte. 606, Charlottesville, VA 22911, 804/978-5508, Fax: 804/978-5620, www.gefanuc.com

GEMS SENSORS, INC., 1 Cowles Rd., Plainville, CT 06062, 800/378-1600, Fax: 860/793-4500, www.gemssensors.com

GENERAL EASTERN, 20 Commerce Way, Woburn, MA 01801, 781/938-7070, Fax: 781/938-1071, www.geinet.com

GENERAL MICRO SYSTEMS, PO Box 2689, Rancho Cucamonga, CA 91729, 800/307-4863, Fax: 909/987-4863, www.gms4ume.com

GENERAL MONITORS, 26776 Simpatica Circle, Lake Forest, CA 92630, 949/581-4464, Fax: 949/581-1151, www.generalmonitors.com

GENSYM CORP., 125 Cambridge Park Dr., Cambridge, MA 02140, 617/547-2500, Fax: 617/547-1962, www.gensym.com

GEORGE FISCHER, INC., 2882 Dow Ave., Tustin, CA 92780, 800/854-4090, Fax: 714/731-6923, www.us.piping.georgefischer.com

GEOSPHERE EMERGENCY, RESPONSE SYSTEM, 100 S. Main St., Doylestown, PA 18901, 215/340-2204, Fax: 215/340-2205, www.plantsafe.com

GESTRA STEAM PRODUCTS—INVENSYS FLOW CONTROL, 5171 Maritime Rd., Jeffersonville, IN 47130, 800/225-6989, Fax: 812/218-7777, www.edwardvogt.com

GIDDINGS & LEWIS CONTROLS, MEASUREMENT & SENSING, 660 S. Military Rd., Fond du Lac, WI 54935, 920/921-7100, Fax: 920/906-7669, www.giddings.com

GLI, A HACH CO. BRAND, 5600 Lindbergh Dr., Loveland, CO 80539, 800/227-4224, Fax: 970/669-2932, www.hach.com

GLI INTERNATIONAL, 9020 W. Dean Dr., Milwaukee, WI 53224, 414/355-3601, Fax: 414/355-8346, www.gliint.com

GLOBAL LIGHTING TECHNOLOGIES INC., 55 Andrews Circle, Brecksville, OH 44141, 440/922-4584, Fax: 440/992-4585, www.glthome.com

GLOBAL WEIGHING, 5110 Old Ellis Pointe, Suite 200, Roswell, GA 30076, 678/393-9960, Fax: 678/393-9961, www.global-weighing.com

GMC INSTRUMENTS, INC., 250 Telser Rd., Unit F, Lake Zurich, IL 60047, 800/462-4040, Fax: 847/540-7242, www.gmcinc.com

GO REGULATOR, 2301 Wardlow Circle, Corona, CA 92880, 909/270-6280, Fax: 909/270-6201, www.circle-seal.com

GORDON PRODUCTS, 67 Del Mar Dr., Brookfield, CT 06804, 800/315-9233, Fax: 203/775-1162, www.gordonproducts.com

GP:50, 2770 Long Rd., Grand Island, NY 14072, 716/773-9300, Fax: 716/773-5019, www.gp50.com

GRACE ENGINEERED PRODUCTS, 5000 Tremont, #203, Davenport, IA 52807, 319/386-9596, Fax: 319/386-9639, www.grace-eng.com

GRAYBAR, 34 N. Meramec, Clayton, MO 63105, 314/573-9200, Fax: 314-573-9456, www.graybar.com

GRAYHILL, INC., 561 Hillgrove Ave., LaGrange, IL 60525, 708/354-1040, Fax: 708/354-2820, www.grayhill.com

GREAT PLAINS INDUSTRIES, INC., 5252 E. 36th St. N., Wichita, KS 67220, 316/686-7361, Fax: 316/686-6746, www.greatplainsindustries.com

GRECO SYSTEMS, 372 Coogan Way, El Cajon, CA 92020, 619/442-0205, Fax: 619/447-8982, www.grecosystems.com

GSE SCALE SYSTEMS, 23900 Haggerty Rd., Farmington Hills, MI 48335, 248/471-5880, Fax: 248/471-5844, www.gse-inc.com

GTS INC., PO Box 799, Shalimar, FL 32579, 850/651-3388, Fax: 850-651-4777, www.onthelevel.com

H

HACH CO., PO Box 389, Loveland, CO 80539, 970/669-3050, Fax: 970/669-2932, www.hach.com

HALLIBURTON ENERGY SERVICES, 410 17th Street, Suite 900, Denver, CO 80202, 800-654-3760, Fax: 303-573-7856

HALLIBURTON ENERGY SERVICES, 2600 S. 2nd St., Duncan, OK 73536, 303/899-4715, Fax: 303/573-7856, www.halliburton.com

HAMILTON SUNDSTRAND, 2771 N.Garey Ave., Pomona, CA 92675, 909/593-3581, Fax: 909/392-3207, www.hs-ait.com

HAM-LET VALVES & FILTERS, 2153 O'Toole Ave., Suite B, San Jose, CA 95131, 408/473-8880, Fax: 408/473-8882, www.ham-let.com

HAMMOND ENCLOSURES, 394 Edinburgh Rd. N., Guelph, Ontario, N1H 1E5 Canada, 519/822-2960, Fax: 519/822-2301, www.hammfg.com

HARDY INSTRUMENTS INC., 3860 Calle Fortunada, San Diego, CA 92123, 800/821-5831, Fax: 858/278-6700, www.hardyinstruments.com

HART COMMUNICATIONS FOUNDATION, 9390 Research Blvd., Suite 1-350, Austin, TX 78759, 512/794-0369, Fax: 512/ 794-3904, www.hartcomm.org

HATHAWAY PROCESS INSTRUMENTATION BETA CALIBRATORS DIV., 2309 Springlake Rd., Suite 600, Farmers Branch, TX 75234, 972/241-2200, Fax: 972/241-6752, www.hathawayprocess.com

HAWKE INTERNATIONAL, 600 Kenrick #C-10, Houston, TX 77060, 972/335-0176, Fax: 972/712-2511, www.ehawke.com/fieldbus

HAYS CLEVELAND, 1111 Brookpark Rd., Cleveland, OH 44109, 216/398-4414, Fax: 216/398-8553, www.hayscleveland.com

HBM, 2815 Colonnades Ct., Norcross, GA 30071, 888/816-9006, Fax: 770/447-8440, www.hbmhome.com

HEIM DATA SYSTEMS, Po Box N, Belmar, NJ 07719, 248/347-4423, Fax: 248/347-9463, www.heimdata.com

HELM INSTRUMENT CO., 361 W. Dussel Dr., Maumee, OH 43537, 419/893-4356, Fax: 419/893-1371, www.helminstrument.com

HEM DATA CORP., 17336 W. 12 Mile Rd. #200, Southfield, MI 48076, 248/559-5607, Fax: 248/559-8008, www.hemdata.com

HERTZLER SYSTEMS, INC., 2312 Eisenhower Dr. N., Goshen, IN 46526, 219/533-0571, Fax: 219/533-3885, www.hertzler.com

HEWLETT-PACKARD/AGILENT, 100, 3030-3rd Ave. N.E., Calgary, Alberta, T2A 6T7 Canada, 403/299-4756, Fax: 403/272-2299, www.agilent.com/find/ais

HI-SPEED CHECKWEIGHER, 5 Barr Rd., Ithaca, NY 14850, 607/257-6000, Fax: 607/257-6396, www.hispeedcheckweigher.com

HIMA PAUL HILDEBRANDT GMBH & CO., Albert-Bassermann-Str. 28, Gruehl, 68782 Germany, 49/6202-709-145, Fax: 49/6202-709-123, www.hima.com

HIRSCHMANN ELECTRONICS, INC., 30 Hook Mountain Rd., Pine Brook, NJ 07058, 973/830-2000, Fax: 973/830-1470, www.hirschmann-usa.com

HMW ENTERPRISES, INC., 207 N. Franklin St., Waynesboro, PA 17268, 717/765-4690, Fax: 717/765-4660, www.hmwent.com

HNU SYSTEMS, INC., 160 Charlemont St., Newton Highlands, MA 02461, 617/964-6690, Fax: 617/558-0056, www.hnu.com

HOFFER FLOW CONTROLS, INC., 107 Kitty Hawk Ln., Elizabeth City, NC 27909, 252/331-1997, Fax: 252/331-2886, www.hofferflow.com

HOFFMAN, 2100 Hoffman Way, Anoka, MN 55303, 800/355-3560, Fax: 612/942-6940, www.hoffmanonline.com

HOKE, INC., 2301 Wardlow Circle, Corona, CA 92880, 909/270-6280, Fax: 909/270-6201, www.circle-seal.com

HONEYWELL INDUSTRIAL CONTROL, 16404 N. Black Canyon Hwy., Phoenix, AZ 85053, 800/288-7491, Fax: 319/294-2968, www.iac.honeywell.com

HONEYWELL SENSING AND CONTROL, 11 W. Spring St., Freeport, IL 61032, 800/537-6945, Fax: 815/235-6545, www.honeywell.com/sensing

HORIBA INSTRUMENTS, INC., 17671 Armstrong Ave., Irvine, CA 92614, 949/250-4811, Fax: 949/250-0924, www.horiba.com

HORNER APG, LLC, 7821 Carinthian Dr., Indianapolis, IN 46201, 317/916-4274, Fax: 317/916-4280, www.heapg.com

HTM ELECTRONICS INDUSTRIES, 8651 Buffalo Ave., Niagara Falls, NY 14304, 800/644-1756, Fax: 888/283-2127, www.htm-sensors.com

HYDE PARK ELECTRONICS, INC., 1875 Founders Dr., Dayton, OH 45420-4017, 937/252-2121, Fax: 937/258-5830, www. hpsensors.com

HYPROTECH LTD, 707/Eighth Ave. SW, Suite 800, Calgary, Alberta, T2P 1H5 Canada, 403/870-6193, Fax: 403/520-6060, www.hyprotech.com

I

IAI AMERICA, INC., 2360 W. 205th St., Torrance, CA 90501, 310/320-3978, Fax: 310/320-4553, www.iaiamerica.com

IBOCO CORP., 10 Alvin Ct., Suite 100, East Brunswick, NJ 08816, 732/238-0200, Fax: 732/238-0304, www.iboco.com

IC FLUID POWER, 63 Drive Hwy., Rossford, OH 43460, 419/661-8811, Fax: 419/661-8844, www.icfluid.com

ICE QUBE INC., 9111 S. Main St., Greensburg, PA 15601, 888/867-8234, Fax: 724/837-3386, www.iceqube.com

ICONICS, 100 Foxborough Blvd., Foxborough, MA 02035, 508/543-8600, Fax: 508/543-1503, www.iconics.com

ICS ADVENT, 6260 Sequence Dr., San Diego, CA 92121, 800/523-2320, Fax: 858/677-0898, www.icsadvent.com

ICS TRIPLEX, 16400-A Park Row, Houston, TX 77084, 281/492-0604, Fax: 281/578-3268, www.icstriplex.com

IDEC CORP., 1175 Elko Dr., Sunnyvale, CA 94089, 800/262-4332, Fax: 800/635-6246, www.idec.com

IDS ENGINEERING, INC., 600 Century Plaza Dr., Houston, TX 77073-6033, 281/821-7100, Fax: 281/821-3230, www.idsengr.com

IFM EFECTOR, INC., 805 Springdale Dr., Exton, PA 19341, 610/524-2004, Fax: 610/524-2745, www.ifmefector.com

IKEY, 2621 Ridgepoint Dr. #235, Austin, TX 78754, 800/866-6506, Fax: 512/837-0207, www.ikey.com

IMAGEVISION, INC., PO Box F, La Grange, TX 78945, 979/247-4068, Fax: 979/247-4062, www.imagevisioninc.com

IMAGING & SENSING TECHNOLOGY, 14737 NE 87th St., Redmond, WA 98052, 425/881-0778, Fax: 425/869-0667, www.istimaging.com

INDUSOFT, PO Box 164073, Austin, TX 78716, 877/INDUSOFT, Fax: 512/527-0792, www.indusoft.com

INDUSTRIAL DATA SYSTEMS, INC., 15031 Woodham, Suite 360, Houston, TX 77073-6026, 281/821-3200, Fax: 281/821-5488, www.inddata.com

INDUSTRIAL INDEXING SYSTEMS, 626 Fishers Rd., Victor, NY 14564, 716/924-9181, Fax: 716/924-2169, www.iis-servo.com

INDUSTRIAL NETWORKING SOLUTIONS, PO Box 540, Addison, TX 75001, 972/248-7466, Fax: 972/248-9533, www.industrialnetworking.com

INDUSTRIAL SCIENTIFIC CORP., 1001 Oakdale Rd., Oakdale, PA 15071, 412/788-4353, Fax: 412/788-8353, www.indsci.com

INFRARED SOLUTIONS, 3550 Annapolis Ln. N #70, Plymouth, MN 55447, 800/760-4525, Fax: 763/551-0038, www.infraredsolutions.com

INNOVA ELECTRONICS, 9207 Emmott Rd., Suite 105, Houston, TX 77423, 713/690-9909, Fax: 713/466-0210, www.innovaelec.com

INNOVATIVE SENSORS, INC., 4745 E. Bryson St., Anaheim, CA 92807, 800/835-5474, Fax: 714/779-9315, www.innovativesensors.com

INNOVENTION INDUSTRIES, 4018 Kingston Ct., Burlington, ON L7L 4Y5, Canada, 905/631-8293, Fax: 905/631-8761, www.innovin.com

INSPIRED SOLUTIONS LLC, 2451 Cumberland Pkwy, Suite 3455, Atlanta, GA 30339, 770/803-0555, Fax: 770/803-0580, www.inspiredsolutions.net

INTECOLOR, 2150 Boggs Rd., Duluth, GA 30096, 770/622-6242, Fax: 770/622-6370, www.intecolor.com

INTEGRATED CONTROL CORP., 748 Park Ave., Huntington, NY 11743, 631/673-5100, Fax: 631/673-6756, www.goicc.com

INTEK, INC., 751 Intek Way, Westerville, OH 43082, 614/895-0301, Fax: 614/895-0319, www.intekflow.com

INTELLIGENT INSTRUMENTATION, 3000 S. Valencia Rd., Suite 100, Tucson, AZ 85706, 800/685-9911, Fax: 520/573-0522, www.instrument.com

INTELLIGENT MOTION SYSTEMS, 370 N. Main St., Marlborough, CT 06447, 860/295-6102, Fax: 860/295-6107, www.imshome.com

INTELLUTION, 325 Foxborough Blvd., Foxborough, MA 02035, 508/698-3322, Fax: 508/698-6973, www.intellution.com

INTERBUS CLUB, THE, PO Box 25141, Philadelphia, PA 19147, 888/281-2871, www.interbusclub.com

INTERFACE, INC., 7401 E. Butherus Dr., Scottsdale, AZ 85260, 480/948-5555, Fax: 480/948-1924, www.interfaceforce. com

INTERGRAPH PROCESS & BUILDING SOLUTIONS, 300 Intergraph Way, Madison, AL 35758, 800/260-0246, Fax: 256/730-3028, www.intergraph.com/pbs

INTERLINKBT, LLC, 3000 Campus, Dr., Plymouth, MN 55441, 763/694-2332, Fax: 763/694-2399, www.interlinkbt.com

INTERNATIONAL RECTIFIER CORP., 233 Kansas St., El Segundo, CA 90245, 310/252-7019, Fax: 310/252-7171, www.irf.com

INTERSTATE ELECTRONICS CORP., 602 East Vermont St., Anaheim, CA 92803, 714/758-0500, Fax: 714/758-4148, www.iechome.com

INTEWORX.NET, 2450 Atlanta Hwy., #1202, Cumming, GA 30040, 770/886-6166, Fax: 770/886-6187, www.inteworx.net

INTRINSYC, 700 W. Pender St., 10th Fl., Vancouver, BC V6C 1E8, 604/801-6461, Fax: 604/801-6417, www.intrinsyc.com

INVENSYS FLOW CONTROL, 1900 S. Saunders St., Raleigh, NC 27603, 800/225-6989, Fax: 919/831-3254, www.invensysflowcontrol.com

INVENSYS FOXBORO AUTOMATION & INSTRUMENTATION, 33 Commercial St., Foxboro, MA 02703, 508/549-3354, Fax: 508-549-4834, www.foxboro.com

INVENSYS SENSOR SYSTEMS/CLAROSTAT, 12055 Rojas Dr., Suite K, El Paso, TX 79936, 915/858-2632, Fax: 915/872-3333, www.clarostat.com

INVENSYS TRICONEX, 15345 Barranca Pkwy., Irvine, CA 92618, 281/709-1200, Fax: 949/753-9136, www.triconex.com

IOTECH, INC., 25971 Cannon Rd., Cleveland, OH 44146, 440/439-4091, Fax: 440/439-4093, www.iotech.com

IRCON, INC., 7300 N. Natchez Ave., Niles, IL 60714, 847/967-5151, Fax: 847/647-0948, www.ircon.com

ISA—THE INSTRUMENTATION, SYSTEMS & AUTOMATION SOCIETY, 67 Alexander Dr., Research Triangle Park, NC 27709, 919/990-9215, Fax: 919/549-8411, www.isa.org

IST-QUADTEK, 14737 NE 87th St., Redmond, WA 98052, 425/881-0778, Fax: 425/869-0667, www.istimaging.com

ITAC SYSTEMS, INC., 3113 Benton St., Garland, TX 75042, 972/494-3073, Fax: 972/494-4159, www.itacsystems.com

ITS ENCLOSURES, 271 Westech Dr., Mt. Pleasant, PA 15666, 800/423-9911, Fax: 724/696-3333, www.itsenclsures.com

ITT CONOFLOW, 5154 Highway 78, St. George, SC 29477, 843/563-9281, Fax: 843/563-2131, www.ittconoflow.com

ITT INDUSTRIES, CANNON SWITCH PRODUCTS, 15 Riverdale Ave., Newton, MA 02458, 617/969-3700, Fax: 617/969-9600, www.ittcannon.com

ITW SWITCHES, 7301 W. Ainslie St., Harwood Heights, IL 60706, 708/667-3370, Fax: 708/667-3440, www.itwswitches.com

J

J-TEC ASSOCIATES, INC., 5255 Blairs Forest Ln. N.E., Suite L, Cedar Rapids, IA 52402, 319/393-5200, Fax: 319/393-5211, www.j-tecassociates.com

J. H. INSTRUMENTS, 9800 N. Southern Pine Blvd., Charlotte, NC 28273, 704/527-6920, Fax: 704/527-8406, www.jhinstrument.com

J. R. MERRITT CONTROLS, 55 Sperry Ave., Stratford, CT 06615, 203/296-2272, Fax: 203/381-0400, www.jrmerritt.com

JAVLYN, INC., 3136 Winton Rd. South, Rochester, NY 14623, 716/697-5212, Fax: 716/424-3110, www.javlyn.com

JC SYSTEMS, INC., 11535 Sorrento Valley Blvd. #400, San Diego, CA 92121, 858/793-7117, Fax: 858/793-1931, www.jcsystemsinc.com

JENSEN TOOLS, INC., 7815 S. 46th St., Phoenix, AZ 85044, 800/426-1194, Fax: 800/366-9662, www.jensentools. com

JH TECHNOLOGY, INC., 4233 Clark Rd., Unit 9, Sarasota, FL 34233, 800/808-0300, Fax: 941/925-8774, www.jhtechnology.com

JOGLER, INC., 9715 Derrington Rd., Houston, TX 77064, 281/469-6969, Fax: 281/469-0422, www.jogler.com

JOWA CONSILIUM US INC., 59 Porter Rd., Littleton, MA 01460, 978/786-9800, Fax: 978/486-0170, www.jowa-consilium.com

JUMO PROCESS CONTROL, INC., 885 Fox Chase, Coatesville, PA 19320, 610/380-8002, Fax: 610/380-8009, www.jumousa.com

K

K-PATENTS, INC., 1804 Centre Point Cir., Suite 106, Naperville, IL 60563, 630/955-1545, Fax: 630/955-1585, www.kpatents.com

K-TEK, 18321 Swamp Rd., Prairieville, LA 70769, 225/673-6100, Fax: 225/673-2525, www.ktekcorp.com

KAHN INSTRUMENTS, INC., 885 Wells Rd., Wethersfield, CT 06109, 860/529-8643, Fax: 860/529-1895, www.kahn.com

KALGLO ELECTRONICS CO., 5911 Colony Dr., Bethlehem, PA 18017-9348, 610/837-0700, Fax: 610/837-7978, www.kalglo.com

KAMAN INSTRUMENTATION OPERATIONS, 3450 N. Nevada Ave., Colorado Springs, CO 80907, 719/635-6979, Fax: 719/634-8093, www.kamaninstrumentation.com

KAVLICO CORP., 14501 Los Angeles Ave., Moorpark, CA 93021, 805/523-2000, Fax: 805/523-7125, www.kavlico.com

KAY-RAY/SENSALL, 1400 Business Center Dr., Suite 100, Mt. Prospect, IL 60056, 800/323-7594, Fax: 847/803-5466, www.kayray-sensall.com

KEITHLEY INSTRUMENTS, INC., 28775 Aurora Rd., Cleveland, OH 44139, 440/248-0400, Fax: 440/248-6168, www.keithley.com

KEPWARE, 81 Bridge St., Yarmouth, ME 04096, 207/846-5881, Fax: 207/846-5947, www.opcsource.com

KESSLER ELLIS PROD., 10 Industrial Way E., Eatontown, NJ 07724, 732/935-1320, Fax: 732/935-0184, www.kep.com

KEY INSTRUMENTS, 250 Andrews Rd., Trevose, PA 19053, 215/357-0893, Fax: 215/357-9239, www.keyinstruments.com

KEYENCE CORP. OF AMERICA, 50 Tice Blvd., Woodcliff Lake, NJ 07675, 201/930-0100, Fax: 201/930-1883, www.keyence.com

KIN-TEK, 504 Laurel, LaMarque, TX 77568, 409/938-3627, Fax: 409/938-3710

KINETROL USA, INC., 4557 Westgrove Dr., Addison, TX 75001, 972/447-9443, Fax: 972/447-9720, www.kinetrol.com

KING ENGINEERING CORP., 3201 S. State St., Ann Arbor, MI 48108, 734/662-5691, Fax: 734/662-6652, www.king-gage.com

KING INSTRUMENT CO., 12700 Pala Dr., Garden Grove, CA 92841, 714/891-0008, Fax: 714/891-0023

KISTLER-MORSE CORP., 19021 120th Ave. NE, Bothell, WA 98011, 425/486-6600, Fax: 425/402-1500, www.kistler-morse.com

KNF NEUBERBER, INC., 2 Black Forest Rd., Trenton, NJ 08691, 609/890-8600, Fax: 609/890-8323, www.knf.com

KNOWLEDGESCAPE SYSTEMS, 669 W. 200 S., Salt Lake City, UT 84106, 801/526-2351, Fax: 801/526-2701, www.kscape.com

KOBOLD INSTRUMENTS, INC., 1801 Parkway View Dr., Pittsburgh, PA 15205, 800/998-1020, Fax: 412/788-4890, www.koboldusa.com

KOKUSAI CHART CORP. OF AMERICA, 2 Dart Road, Newnan, GA 30265, 770/251-6496, Fax: 70/304-1917, www.kkusai-usa.com

KOLLMORGEN, 201 Rock Road, Radford, VA 24141, 540/633-4124, Fax: 540/731-0847, www.kollmorgen.com

KONTRON, 6260 SEQUENCE Dr., San Diego, CA 92121, 858/977-0877, Fax: 858/677-0615, www.kontron.com

KPSI TRANSDUCERS, PRESSURE SYSTEMS, 34 Research Dr., Hampton, VA 23666, 757/865-1243, Fax: 757/865-8744, www.pressuresystems.com

KRAISSL CO., INC., 299 Williams Ave., Hackensack, NJ 07601-5225, 201/342-0008, Fax: 201/342-0025, www.strainers.com

KROHNE, INC., 7 Dearborn Rd., Peabody, MA 01960, 978/535-6060, Fax: 978/535-1720, www.krohne.com

KSR KUEBLER, 7516 Precision Dr., Raleigh, NC 27613, 919/596-3800, Fax: 919/598-8128, www.ksr-usa.com

K-TEK, 18321 Swamp Rd., Prairieville, LA 70769, 225/673-6100, Fax: 225-673-2525, www.ktekcorp.com

KUKA CONTROLS, 17821 E. 17th St. #160, Tustin, CA 92780, 714/505-1485, Fax: 714/505-1149, www.kuka-controls.com

KURT MANUFACTURING CO., Electronics Div., 2130 107th Ln. NE, Minneapolis, MN 55449, 763/572-4597, Fax: 763/784-6055, www.kurt.com

KURZ INSTRUMENTS, INC., 2411 Garden Rd., Monterey, CA 93940, 800/424-7356, Fax: 831/646-0427, www.kurz-instruments.com

KW SOFTWARE, 3536 Edwards Rd., Cincinnati, OH 45208, 513/321-9385, Fax: 513/321-6992

L

L-3 COMMUNICATIONS, INTERSTATE ELECTRONICS CORP., 602 E. Vermont St., Anaheim, CA 92803, 714/758-4158, Fax: 714/758-4148, www.iechome.com

L & J TECHNOLOGIES, 5911 Butterfield Rd., Hillside, IL 60162, 708/236-6000, Fax: 708/236-6006, www.ljtechnologies.com

LABTECH, 2 Dundee Park, Suite B09, Andover, MA 01810, 978/470-0099, Fax: 978/470-3338, www.labtech.com

LAKE SHORE CRYOTRONICS, INC., 575 McCorkie Blvd., Westerville, OH 43082, 614/891-2243, Fax: 614/818-1600, www.lakeshore.com

LAMBDA, 3055 Del Sol Blvd., San Diego, CA 92154, 619/628-2871, Fax: 619/429-1011, www.lambdapower.com/control

LAND INFRARED DIV. OF LAND INSTRUMENTS INTL., INC., 10 Friends Ln., Newtown, PA 18940, 800/523-8989, Fax: 215/781-0723, www.landinst.com

LANTRONIX, 15353 BARRANCA PKWY., IRVINE, CA 92618, 949/453-3995, www.lantronix.com

LAPP USA, 29 Hanover Rd., Florham Park, NJ 07932, 800/774-3539, Fax: 973/660-9678, www.lappusa.com

LEVITON MFG. CO., 59-25 Little Neck Pkwy., Little Neck, NY 11362, 718/281-6031, Fax: 718/631-6508, www.leviton.com

LIEBERT CORP., 1050 Dearborn Dr., Columbus, OH 43229, 614/841-5798, Fax: 614/841-6022, www.liebert.com

LIGHTHAMMER SOFTWARE, 690 Stockton Dr., Exton, PA 19341, 610/903-8000, Fax: 610/903-8006, www.lighthammer.com

LIMITORQUE, PO Box 11318, Lynchburg, VA 24506, 800/366-4401, Fax: 804/522-9858, www.limitorque.com

LINSEIS, INC., PO Box 666, Princeton Junction, NJ 08550, 609/799-6282, Fax: 609/799-7739, www.linseis.com

LION PRECISION, 563 Shoreview, St. Paul, MN 55126, 651/484-6544, Fax: 651/484-6824, www.liorprecision.com

LIPTAK ASSOC. P.E., 84 Old N. Stamford Rd., Stamford, CT 06905, 203/357-7614, Fax: 203/325-3922, http://hometown.aol.com/liptakbela

LIQUID CONTROLS, INC., A UNIT OF IDEX, 105 Albrecht Dr., Lake Bluff, IL 60044, 847/295-1050, Fax: 847/295-1057, www.lcmeter.com

LOAD CONTROLS, INC., 10 Picker Rd., Storbrodge, MA 01566, 508/347-2606, Fax: 508/347-2064, www. loadcontrols.com

LOCKWOOD GREENE, PO Box 491, Spartanburg, SC 29304, 864/578-2000, Fax: 864/599-4117, www.lg.com

LOCON SENSOR SYSTEMS, INC., PO Box 789, Holland, OH 43528, 419/865-7651, Fax: 419/865-7756, www.locon.net

LOGIC BEACH, INC., 8363-6F Center Dr., La Mesa, CA 91942, 619/698-3300, Fax: 619/469-8604, www.logicbeach.com

LUCENT SPECIALTY FIBERTECHNOLOGIES, 55 Darling Dr., Avon, CT 06001, 860/678-0371, Fax: 860/674-8818, www.lucent.com/ofs/specialtyfiber

LUMBERG, INC., 14121 Justice Rd., Midlothian, VA 23113, 804/379-2010, Fax: 804/379-3232, www.lumbergusa.com

LUMENITE CONTROL TECH., 2331 N. 17th Ave., Franklin Park, IL 60131, 847/455-1450, Fax: 847/455-0127

LUMITREX INC., 8443 Dow Circle, Strongsville, OH 44136, 440/243-8401, Fax: 440/243-8402, www.lumitex.com

LUTZE, INC., 13330 S. Ridge Dr., Charlotte, NC 28273, 704/504-0222, Fax: 704/504-0223, www.lutze.com

LYNX REAL-TIME SYSTEMS, INC., 2239 Samaritan Dr., San Jose, CA 95124, 408/879-3900, Fax: 408/879-3920, www.lynx.com

M

M. DAVIS & SONS INC., 200 Hadco Rd., Wilmington, DE 19702, 800/91-DAVIS, Fax: 302/633-0845, www.mdavisinc.com

M-SYSTEM, 15028 Beltway Dr., Addison, TX 75001, 800/544-3181, Fax: 972/385-2277, www.m-system.com

MACTEK CORP., 7380 Stoneham Rd., Gates Mills, OH 44040, 440/423-0955, Fax: 440/423-0967, www.mactek-corp.com

MAGNETEK, 16555 W. Ryerson Rd., New Berlin, WI 53151, 414/782-0200, Fax: 414/782-3418, www.magnetekdrives.com

MAGNETROL INTL., INC., 5300 Belmont Rd., Downers Grove, IL 60516, 630/969-4000, Fax: 630/969-9489, www.magnetrol.com

MALUERN/INSITEC, 2110 Omega Rd., Suite D, San Ramar, CA 94583, 925/837-1330, Fax: 925/837-3864, www.insitec.com

MANNESMANN REXROTH CORP., REXROTH MECMAN DIV., 1953 Mercer Rd., Lexington, KY 40511, 606/254-8031, Fax: 606/281-3491, www.rexrothmecman.com

MAPLE SYSTEMS, INC., 808 134th St. SW, Suite 120, Everett, WA 98204, 425/745-3229, Fax: 425/745-3429, www.maple-systems.com

MARKLAND SPECIALTY ENG. LTD, 48 Shaft Rd., Toronto, Ontario M9W 4M2 Canada, 416/244-4980, Fax: 416/244-2287, www.sludgecontrols.com

MARSH BELLOFRAM, State Route 2; Box 305, Newell, WV 26050, 304/387-1200, Fax: 304/387-4417, www.marshbellofram.com

MARSH-MCBIRNEY, INC., 4539 Metropolitan Ct., Frederick, MD 21704, 301/874-5599, Fax: 301/874-2172, www.marsh-mcbirney.com

MARTEL ELECTRONICS, PO Box 897, Windham, NH 03087, 603/893-0886, Fax: 603/898-6820, www.martelcorp.com

MASONEILAN/DRESSER, 85 Bodwell St., Avon, MA 02322, 508/586-4600, Fax: 508/427-8971, www.masoneilan.com

MASSA PRODUCTS CORP., 280 Lincoln St., Hingham, MA 02043, 781/749-4800, Fax: 781/740-2045, www.massa.com

MATHESON INSTRUMENTS, 166 Keystone Dr., Montgomeryville, PA 18936, 215/648-4026, Fax: 215/641-0656, www.mathesoninstruments.com

MATRIC, RDI, Box 421A, Seneca, PA 16346, 800/462-8742, Fax: 814/678-1301, www.matric.com

MATRIKON CONSULTING, INC., Suite 1800, 10405 Jasper Ave., Edmonton, Alberta, T5J 3A4 Canada, 780/448-1010, Fax: 780/448-9191, www.matrikon.com

MATROX IMAGING, 1055 St.-Regis Blvd., Durva, QC H9P 2T4, CANADA, 800/804-6243, Fax: 514/822-6273, www.matrox.com/imaging

MAX CONTROL SYSTEMS, INC., 1180 Church Rd., Lansdale, PA 19446, 215/393-3900, Fax: 215/393-3921, www.maxcontrols.com

MAX MACHINERY, INC., 1420 Healdsburg Ave., Healdsburg, CA 95448, 707/433-7281, Fax: 707/433-0571, www.maxmachinery.com

MAXITROL CO., 23555 Telegraph Rd., Southfield, MI 48037-2230, 248/356-1400, Fax: 248/356-0829, www.maxitrol.com

MCCROMETER, 3255 W. Stetson Ave., Hemet, CA 92545, 909/652-6811, Fax: 909/652-3078, www.mccrometer.com

MCG SURGE PROTECTION, 12 Burt Drive, Deer Park, NY 11729, 800/851-1508, Fax: 516/586-5120, www.mcgsurge.com

MCLEAN MIDWEST, 11611 Business Park Blvd. N., Champlin, MN 55316, 612/323-8200, Fax: 612/576-3200, www.2010corp.com

MCMILLAN CO., PO Box 1340, Gerogetown, TX 78627, 800/861-0231, Fax: 512/863-0671, www.mcmillancompany.com

MDSI, 220 E. Huron, Suite 600, Ann Arbor, MI 48104, 734/769-9900, Fax: 734/769-9112, www.mdsi2.com

MDT SOFTWARE, 2520 NorthWind Pkwy., Suite 100, Alpharetta, GA 30004, 678/297-1000, Fax: 678/297-1003, www.mdtsoft.com

MEASUREMENT COMPUTING CORP., 16 Commerce Blvd., Middleboro, MA 02346, 508/946-5100, www.measurementcomputing.com

MEE INDUSTRIES, INC., 204 W. Pomona, Monrovia, CA 91016, 626/359-4550, Fax: 626/359-4660, www.meefog.com

MEECO, INC., 250 Titus Ave., Warrington, PA 18976, 215/343-6600, Fax: 215/343-4194, www.meeco.com

MELCOR, 1040 Spruce St., Trenton, NJ 08648, 609/393-4178, Fax: 609/393-9461, www.melcor.com

MEN MICRO, INC., 1940 Camden Way, Suite 100, Carrollton, TX 75007, 972/939-2675, Fax: 972/466-5986, www.menmicro.com

MENSOR CORP., 201 Barnes Dr., San Macros, TX 78666, 512/396-4200, Fax: 512/396-1820, www.mensor.com

MERIAM INSTRUMENT, 10920 Madison Ave., Cleveland, OH 44102, 216/281-1100, Fax: 216/281-0228, www.meriam.com

METERSANDINSTRUMENTS.COM, 10 Vantage Point Dr., Unit 5, Rochester, NY 14624, 800/773-0370, Fax: 800/773-0371, www.metersandinstruments.com

METRIX INSTRUMENT CO., 1711 Townhurst Dr., Houston, TX 77043, 713/461-2131, Fax: 713/461-8223, www.metrix1.com

METSO AUTOMATION, 44 Bearfoot Rd., Northborough, MA 01532, 508/852-0200, Fax 508/393-0978, www.metsoautomation.com

METTLER-TOLEDO, 1900 Polaris Pkwy., Columbus, OH 43240, 800/523-5123, Fax: 614/438-4544, www.mt.com

METTLER-TOLEDO HI-SPEED INC., 5 Barr Rd., Ithaca, NY 14850, 607/257-6000, Fax: 607/266-5478, www.hispeedcheckweigher.com

METTLER-TOLEDO PROCESS/INGOLD, 299 Washington St., Woburn, MA 01801, 800/352-8763, Fax: 781/939-6392, www.mt.com/pro

MGE UPS SYSTEMS, 1660 Scenic Ave., Costa Mesa, CA 92626-1410, 714/557-1636 Fax: 714/434-7652, www.mgeups.com

MGR INDUSTRIES, INC., 3013 E. Mulberry St., Ft. Collins, CO 80524, 970/221-2201, Fax: 970/484-4078, www.mgrind.com

MICON SYSTEMS, 4955 Gulf Frwy., Houston, TX 77023, 713/921-1899, Fax: 713/921-1882, www.miconsystems.com

MICRO DECISIONS CORP., 3206-A Cascade Dr., Valparaiso, IN 46383, 219/477-2002, Fax: 219/477-3910, www.micro-decisions.com

MICRO MOTION, INC., 7070 Winchester Circle, Boulder, CO 80301, 800/760-8119, Fax: 303/530-8459, www.micromotion.com

MICROGROUP, INC., 7 Industrial Park Rd., Medway, MA 02053, 508/533-4925, Fax: 508/533-5691, www.microgroup.com

MICROTOUCH SYSTEMS, 300 Griffin Brook Pk., Methuen, MA 01844, 978/659-9000, Fax: 978/659-9051, www.microtouch.com

MICROWAVE DATA SYSTEMS INC., 175 Science pkwy., Rochester, NY 14620, 585/242-9600, Fax: 585/242-9620, www.microwavedata.com

MID-WEST INSTRUMENT, 6500 Dobry Dr., Sterling Heights, MI 48314, 810/254-6500, Fax: 810/254-6509, www.midwestinstrument.com

MIDAC CORP., 17911 Fitch Ave., Irvine, CA 92614, 949/660-8558, Fax: 949/660-9334, www.midac.com

MIE, INC., 7 Oak Park, Redford, MA 01730, 781/275-1919, Fax: 781/275-2121, www.mieinc.com

MIKRON INSTRUMENT CO., Inc., 16 Thornton Rd., Oakland, NJ 07436, 201/405-0900, Fax: 201/405-0090, www.mikroninst.com

MIL-RAM TECHNOLOGY INC., 4135 Business Center Dr., Fremont, CA 94538, 510/656-2001, Fax: 510/656-2004, www.mil-ram.com

MILLTRONICS, INC., 709 Stadium Dr., Arlington, TX 76011, 817/277-3543, Fax: 817/277-3894, www.milltronics.com

MITAC INDUSTRIAL CORP., 42001 Christy St., Fremont, CA 94538, 510/656-5288, Fax: 510/656-2669, www.mitacinds.com

MITSUBISHI ELECTRIC AUTOMATION, INC., 500 Corporate Woods Pkwy., Vernon Hills, IL 60061, 847/478-2419, Fax: 847/478-2396, www.meau.com

MITSUBISHI ELECTRIC & ELECTRONICS USA, 1050 E. Arques Ave., Sunnyvale, CA 94085, 408/730-5900, Fax: 408/245-2690, www.angleview.com

MKS INSTRUMENTS, INC., 6 Shattuck Rd., Andover, MA 01810, 978/975-2350, Fax: 978/975-0093, www.mksinst.com

MODCOMP, INC., 1650 W. McNab Rd., Ft. Lauderdale, FL 33309, 954/977-1380, Fax: 954/977-1900, www.modcomp.com

MODULAR INDUSTRIAL COMPUTERS, INC., 6025 Lee Hwy., Suite 340, Chattanooga, TN 37421, 423/499-0700, Fax: 423/892-0000, www.mic.com

MOELLER ELECTRIC CORP., 25 Forge Pkwy., Franklin, MA 02038, 508/520-7080, Fax: 508/520-7084, www.moellerusa.net

MOISTURE REGISTER PROD., 1712 Earhart Ct., La Verne, CA 91750, 909/392-5833, Fax: 909/392-5838, www.moistureregisterproducts.com

MONARCH INSTRUMENT, 15 Columbia Dr., Amherst, NH 03031, 603/883-3390, Fax: 603/886-3300, www.monarchinstrument.com

MONITOR TECHNOLOGIES, LLC, 44W320 Keslinger Rd., Elburn, IL 60119, 630/365-9403, Fax: 630/365-5646, www.monitortech.com

MONITROL MFG. CO., INC., PO Box 6296, Tyler, TX 75711, 903/561-0742, Fax: 903/561-3559, www. monitrolmfg.com

MOOG, INC., Seneca & Jaminson Rd., E. Aurora, NY 14052, 716/687-4785, Fax: 716/687-4467, www.moog.com/imc/product

MOORE INDUSTRIES-INTERNATIONAL, INC., 16650 Schoenborn St., Sepulveda, CA 91343, 818/894-7111, Fax: 818/891-2816, www.miinet.com

MOORE PROCESS AUTOMATION SOLUTIONS, 1201 Sumneytown Pike, Spring House, PA 19477, 215/646-7400, Fax: 215/ 283-2802

MOREHOUSE INSTRUMENT CO., 1742 South Ave., York, PA 17403, 717/843-0081, Fax: 717/846-4193

MOSAIC INDUSTRIES, 5437 Central Avenue, Suite 1, Newark, CA 94560, 510/790-1255, Fax: 510/790-0925, www.mosaic-industries.com

MOTORTRONICS, 13214-38th St. N., Clearwater, FL 33762, 727/573-1819, Fax: 727/573-1803, www.motortronics.com

MOYNO RKL CONTROLS, 1895 W. Jefferson Ct., Springfield, OH 45506, 937/327-3540, Fax: 937/327-3619, www.moyno.com

MSA INSTRUMENT DIV., PO Box 427, Pittsburgh, PA 15230, 800/MSA-4678, Fax: 724/776-3280, www.msanet.com

MSDI, 220 E. Huron St., Ann Arbor, MI 48104, 734/327-8246, Fax: 734/769-9112, www.mdsi2.com

MSE – TETROGENICS, 65 E. Broadway, Butte, MT 59701, 406/533-6800, Fax: 406/533-6818, www.tetragenics.com

MSI SENSORS, 1000 Lucas Way, Hampton. VA 23666, 757/766-1500, Fax: 757/766-4297, www.msisensors.com

M-SYSTEM CO., LTD., 660 Fargo Ave., Elk Grove Village, IL 60007, 800/544-3181, Fax: 847/364-1140, www. m-system.com

MTL, INC., 9 Merrill Industrial Dr., Hampton, NH 03842, 603/926-0090, Fax: 603/926-1899, www.mtl-inst.com

MTS SYSTEMS CORP., 3001 Sheldon Dr., Cary, NC 27513, 919/677-0100, Fax: 919/677-0200, www.levelplus.com

MUSTANG ENGINEERING, INC., 16001 Park Ten Pl., Houston, TX 77084, 713/215-8000, Fax: 713/215-8590, www.mustangeng.com

MYPLANT.COM, a Honeywell business, 7047 E. Greenway Parkway, Suite 400, Scottsdale, AZ 85254, 877/848-8831, Fax: 480/850-6301, www.myplant.com

N

NAF—INVENSYS FLOW CONTROL, 1790 Satelite Blvd., Ste 100-04, Duluth, GA 30097, 678/474-1500, Fax: 678/474-1515, www.nafcontrols.com

NATIONAL INSTRUMENTS, 11500 N. Mopac Expwy., Austin, TX 78759, 512/683-6863, Fax: 512/683-5759, www.ni.com

NATIONWIDE PERSONNEL RECRUITING & CONSULTATION, 20834 SW Martinazzi Ave., Tualatin, OR 97062, 503/692-4925

NDC INFRARED ENGINEERING, 5314 N. Irwindale Ave., Irwindale, CA 91706, 626/960-3300, Fax: 626/939-3870, www.ndc.com

NELES AUTOMATION, 7000 Hollister, Fl. 3, Houston, TX 77040, 713/346-0600, Fax: 713/346-0602, www.nelesautomation.com

NEMATRON, 5840 Interface Dr., Ann Arbor, MI 48103, 734/214-2000, Fax: 734/994-8074, www.nematron.com

NETSILICON, INC., 411 Waverly Oaks Rd., Bldg. 26, Waltham, MA 02452, 781/893-1234, Fax: 781/893-1338, www.netsilicon.com

NEURALWARE, 230 E. Main St., Suite 200, Carnegie, PA 15106, 412/278-6280, Fax: 412/278-6289, www.neuralware.com

NEUTRONICS, INC., 456 Creamery Way, Exton, PA 19341, 610/524-8800, Fax: 610/524-8807, www.neutronicsinc.com

NEWPORT ELECTRONICS, INC., 2229 S. Yale St., Santa Ana, CA 92704, 714/540-4914, Fax: 714/546-3022, www.newportinc.com

NEXUS ENGINEERING, 900 Rockmead, Suite 250, Kingwood, TX 77339, 281/359-5190, Fax: 281/359-5278, www.nexusengineering.com

NIBCO, INC., 1516 Middlebury St., Elkhart, IN 46516, 219/295-3000, Fax: 219/295-3307, www.nibco.com

NKK SWITCHES, 7850 E. Gelding Dr., Scottsdale, AZ 85260, 480/991-0942, Fax: 480/998-1435, www.nkkswitches.com

NOAX TECHNOLOGIES, 2937 Bee Ridge Rd., Sarasota, FL 34239, 941/922-1150, Fax: 941/330-8422, www.noax.com

NORDSTROM/AUDCO—INVENSYS FLOW CONTROL, 1511 Jefferson St., Sulphur Springs, TX 75482, 903/885-4691, Fax: 903/439-3411, www.nordstromaudco.com

NORRISEAL, 11122 W. Little York, Houston, TX 77041, 713/466-3552, Fax: 713/896-7386, www.norriseal.com

NORTECH FIBRONIC, INC., 240-500 Avenue St. Jean-Baptiste, Quebec, QC, G2E 5R9 Canada, 418/872-4686, Fax: 418/872-2894, www.nortech.ca

NORTHWIRE, 110 Prospect Way, Osceola, WI 54020, 800/468-1516, Fax: 715/294-3727, www.northwire.com

NORTH EAST ELECTRONICS CONTROLS, 1545 Holland Rd., Maumee, OH 43537, 419/893-4158, Fax: 419/893-4171, www.nee-controls.com

NORTHSTAR TECHNOLOGIES, INC., 575 McCorkle Blvd., Westerville, OH 43082, 614/818-1150, Fax: 614/891-6909, www.northstarencoders.com

NORTHWEST ANALYTICAL, INC., 519 SW Park Ave., Portland, OR 97205, 503/224-7727, Fax: 503/224-5236, www.nwasoft.com

NOSHOK, INC., 1010 W. Bagley Rd., Berea, OH 44133, 440/243-0888, Fax: 440/243-3472, www.noshok.com

N-TRON, 578 Azalea Rd., Suite 105, Mobile, AL 36609, 334/666-9878, Fax: 334/666-9833, www.n-tron.com

NUMATICS, INC., 1450 N. Milford Rd., Highland, MI 48357, 248/889-6227, Fax: 248/887-4768, www.numatics.com

O

OCEANA SENSOR TECHNOLOGIES, 1632 Corporate Landing Pkwy., Virginia Beach, VA 23454, 757/426-3678, Fax: 757/426-3633, www.oceanasensor.com

OGDEN MANUFACTURING CO., 64 W. Seegers Rd., Arlington Heights, IL 60005, 847/593-8050, Fax: 847/593-8062, www.ogdenmfg.com

OHIO SEMITRONICS, INC., 4242 Reynolds Dr., Hilliard, OH 43026, 614/777-1005, Fax: 614/777-4511, www.ohiosemitronics.com

OHMART/VEGA CORP., 4241 Allendorf Dr., Cincinnati, OH 45209, 513/272-0131, Fax: 513/272-0133, www.ohmartvega.com

OI DIRECT, 343 St. Paul Blvd., Carol Stream, IL 60188, 888/OI-DIRCT, Fax: 888/OI-FAX-US, www.oidirect.com

OLFLEX WIRE & CABLE, INC., 30 Plymouth St., Fairfield, NJ 07004, 973/575-1101, Fax: 973/575-1267, www.olflex.com

OMEGA ENGINEERING, INC., One Omega Dr., Stamford, CT 06907, 203/359-1660, Fax: 203/359-7700, www.omega.com

OMEGA VANZETTI, INC., 6 Merchant St., Sharon, MA 02067, 781/784-4733, Fax: 781/784-2447, www.vanzetti.com

OMNI FLOW COMPUTERS, 10701 Corporate Dr, #300, Stafford, TX 77477, 281/240-6161, Fax: 281/240-6162, www.omniflow.com

OMRON ELECTRONICS INC., One E. Commerce Dr., Schaumburg, IL 60173, 800/55-OMRON, Fax: 847/843-7787, www.omron.com/oei

ONCUITY, INC., 1410 Blalcok Rd., Houston, TX 77055, 888/271-6726, Fax: 713/682-8066, www.oncuity.com

ONLINE DEVELOPMENT, 7209 Chapman Hwy., Knoxville, TN 37920, 865/251-5222, Fax: 865/579-4740, www.oldi.com

ONTRAK CONTROL SYSTEMS, INC., 422 Arnley St., Sudbury, Ontario, P3C 1E7 Canada, 705/671-2652, Fax: 705/671-6127, www.ontrak.net

OPTEK-DANULAT, INC., 279 S. 17th Ave., Suite 10, West Bend, WI 53095, 800/371-4288, Fax: 262/335-4299, www.optek.com

OPTIMATION, INC., PO Box 4107, Huntsville, AL 35815, 256/883-3050, Fax: 256/883-3070, www.optimate.com

OPTO 22, 43044 Business Park Dr., Temecula, CA 92590, 909/695-3000, Fax: 909/695-3095, www.opto22.com

OPTREX AMERICA INC., 46723 Five Mile Rd., Polymouth, MI 48170, 734/416-8500, Fax: 734/416-8520, www.optrex.com

OP/STATION, A DIV. OF AUTOMATED CONTROL CONCEPTS, INC., 3535 Route 66, Neptune, NJ 07753, 732/922-6611, Fax: 732/922-9611, www.opstation.com

ORBITAL SCIENCES CORP., 2771 N. Garey Ave., Pomona, CA 91767, 909/593-3581, Fax: 909/392-3207, www.orbital-ait.com

ORENA CONTRONET, INC., 8320 NW Hawkins Blvd., Portland, OR 97229, 503/297-1854, Fax: 503/297-1914, www.orena.com

ORSI-GROUP, 502 Earth City Expwy., Suite 139, St. Louis, MO 63045, 314/298-9100, Fax: 314/298-1729, www.orsigroup.com

ORTRONICS INC., 125 Eugene O’Neil Dr., New London, CT 06320, 800/934-5432, Fax: 888/282-0043, www.ortronics.com

OSI SOFTWARE, 777 Davis St., Suite 250, San Leandro, CA 94577, 510/297-5836, Fax: 510/357-8136, www.osisoft.com

OSMONICS, INC., 5951 Clearwater Dr., Minnetonka, MN 55343-8995, 719/264-3937, Fax: 719/536-3301, www.osmonics.com

OTEK CORP., 4016 E. Tennessee St., Tucson, AZ 85714, 520/748-7900, Fax: 520/790-2808, www.otekcorp.com

OTTO ENGINEERING, 2 E. Main St., Carpenterville, IL 60110, 847/428-7171, Fax: 847/428-1956, www.ottoeng.com

P

PACIFIC SCIENTIFIC, 4301 Kishwaukee St., Rockford, IL 61105-0106, 815/226-3100, Fax: 815/226-3080, www.pacsci.com

PAI PARTNERS, 135 Fort Lee Rd., Leonia, NJ 07605, 201/585-2050, Fax: 201/585-1968

PANAMETRICS, INC., 221 Crescent St., Waltham, MA 02453, 781/899-2746, Fax: 781/894-8582, www.panametrics.com

PARKER COMPUMOTOR, 5500 Business Park Dr., Rohnert Park, CA 94928, 800/358-9068, Fax: 707/584-8015, www.compumotor.com

PARKER FILTRATION, 100 Ames Pond Dr., Tewksbury, MA 01876, 800/343-4048, Fax: 978/858-0625, www.parker.com/balston

PARKER FLUID CONTROL DIV., 95 Edgewood Ave., New Britain, CT 06051, 860/827-2300, Fax: 860/827-2884, www.parker.com/fcd

PARKER HANNIFIN, 6035 Parkland Blvd., Cleveland, OH 44124, 256/435-2130, Fax: 256/435-7710, www.parker.com

PARKER HANNIFIN CORP., INSTRUMENTATION CONNECTORS DIV., 9400 S. Memorial Pkwy., Huntsville, AL 35803-2197, 256/881-2040, Fax: 256/881-5730, www.parker.com/icd/

PARKER HANNIFIN CORP., INSTRUMENTATION VALVE DIV., 2651 Alabama Hwy. 21 North, Jacksonville, AL 36265-9681, 256/435-2130, Fax: 256/435-7718, www.parker.com/ivd

PARKER HANNIFIN CORP., SKINNER VALVE DIV., 250 Canal Blvd., New Britain, CT 06051, 860/827-2300, Fax: 860/827-2384, www.parker.com/skinner

PARKER HANNIFIN CORP., VERIFLO DIVISION, 250 Canal Blvd., Richmond, CA 94947, 510/412-1166, Fax: 510/232-7396, www.veriflo.com

PATLITE CORP., 3860 Del Amo Blvd., Suite 404, Torrance, CA 90503, 310/214-5286, Fax: 310/214-5288, www.patlite.com

PAVE TECHNOLOGY CO., 2751 Thunderhawk Ct., Dayton, OH 45414-3445, 937/890-1100, Fax: 937/890-5165, www.pavetechnologies.com

PAVILION TECHNOLOGIES, INC., 11100 Metric Blvd., Austin, TX 78758, 800/886-8432, Fax: 512/438-1401, www.pavtech.com

PAYNE ENGINEERING CO., PO Box 70, Scott Depot, WV 25302, 304/757-7353, Fax: 304/757-7305, www.payneg.com

PC SOFT INTERNATIONAL, 89 Forbes Blvd., Mansfield, MA 02048, 508/337-9200, Fax: 508/337-9201, www.emation.com

PCB PIEZOTRONICS INC., 3425 Walden Ave., Depew, NY 14043, 716/684-0002, Fax: 716/684-0987, www.pcb.com

PCD, INC., 2 Technology Dr., Peabody, MA 01960, 978/532-8800, Fax: 978/532-6800, www.pcdinc.com

PCME LTD, Clearview Building, Edison Rd., St. Ives, Cambs PE27 3GH UK, 01480 468200, Fax: 01480 463400, www.pcme.co.uk

PELICAN PRODUCTS, INC., 23215 Early Ave., Torrance, CA 90505, 310/326-4700, Fax: 310/326-3311, www.pelican.com

PENBERTHY, 320 Locust St., Prophetstown, IL 61277, 815/537-2311, Fax: 815/537-5764, www.pcc-penberthy.com

PENTADYNE POWER CORP., 20750 Lassen St., Chatsworth, CA 91311, 818/350-0370, Fax: 818/350-0384, www.pentadyne.com

PENTAIR ELECTRONIC PACKAGING, 170 Commerce Dr., Warwick, RI 02886, 800/451-8755, Fax: 401/738-2904, www.pentair-ep.com

PENTEK, One Park Way, Upper Saddle River, NJ 07458, 201/818-5900, Fax: 201/818-5904, www.pentek.com

PEP MODULAR COMPUTERS, 750 Holiday Dr., Pittsburgh, PA 15220, 412/921-3322, Fax: 412/921-3356, www.pepusa.com

PEPPERL+FUCHS, INC., 1600 Enterprise Pkwy., Twinsburg, OH 44087, 330/425-3555, Fax: 330/425-4607, www.am.pepperl-fuchs.com

PERMA PURE, INC., PO Box 2105, Toms River, NJ 08754, 732/244-0010, Fax: 732/244-8140, www.permapure.com

PGI INTERNATIONAL, 16101 Vallen Dr., Houston, TX 77041, 713/466-0056, Fax: 713/744-9899, www.pgiint.com

PHOENIX CONTACT, INC., PO Box 4100, Harrisburg, PA 17111, 800/322-3225, Fax: 717/944-1625, www.phoenixcon.com

PHOENIX DIGITAL, 7650 E. Evans Rd., Bldg. A, Scottsdale, AZ 85260, 480/483-7393, Fax: 480/483-7391, www.phoenixdigitalcorp.com

PHONETICS, INC., 701 Tryens Rd., Aston, PA 19014, 610/558-2700, Fax: 610/558-0222, www.sensaphone.com

PILZ AUTOMATION SAFETY L.P., 24850 Drake Rd., Farmington Hills, MI 48335, 248/473-1133, Fax: 248/473-3997, www.pilzusa.com

PLANAR SYSTEMS, 1400 NW Compton, Beaverton, OR 97006, 503/690-1100, Fax: 503/690-1493, www.planar. com

PLAST-O-MATIC VALVES, INC., 1384 Pompton Ave., Cedar Grove, NJ 07009, 973/256-3000, Fax: 973/256-4745, www.plastomatic.com

PMV-USA, INC., 1440 Lakefront Circle #160, The Woodlands, TX 77380, 281/292-7500, Fax: 281/292-7760, www.pmvusa.com

POMONA ELECTRONICS, 1500 E. Ninth St., Pomona, CA 91766-3835, 909/469-2900, Fax: 909/469-3317, www.pomonaelectronics.com

POWERCUBE, 9340 Owensmouth Ave., Chatsworth, CA 91311, 818/734-6500, Fax: 818/734-6540, www.powercube.com

POWERS PROCESS CONTROLS, 3400 Oakton St., Skokie, IL 60076, 847/568-6256, Fax: 847/673-9044, www.powerscontrols.com

PQ SYSTEMS, INC., 10468 Miamisburg-Springbordo Rd., Miamisburg, OH 45342, 800/777-3020, Fax: 937/885-2252, www.pqsystems.com

PRECIDIA TECHNOLOGIES, 10A Hearst Way, Kanata, Ontario K2L 2P4, 613/592-7557, Fax: 613/482-5770, www.precidia.com

PRECISION DIGITAL CORP., 19 Strathmore Rd., Natick, MA 01760, 508/655-7300, Fax: 508/655-8990, www.predig.com

PRECISION SOLUTIONS, INC., 3101 Bee Caves Rd. #290, Austin, TX 78746, 512/330-9990, Fax: 512/330-9996, www.psivideo.com

PRESSURE SYSTEMS, INC., 34 Research Dr., Hampton, VA 23666, 757/865-1243, Fax: 757/865-8744, www.pressuresystems.com

PRESYS INSTRUMENTS, INC., 3000 SW 77 Pl., Miami, FL 33155, 305/262-8488, Fax: 305/262-7225, www.presys.com.br.

PRINCO INSTRUMENTS, INC., 1020 Industrial Blvd., Southampton, PA 18966, 215/355-1500, Fax: 215/355-7766, www.princoinstruments.com

PRO-TECH, 3600A Swiftwater Park Dr., Suwanee, GA 30024, 770/271-0048, Fax: 770/271-2796, www.protech1.com

PROCONSUL, INC., PO Box 1823, Georgetown, TX 78627, 512/863-8000, Fax: 512/869-4999, www.proconsul.net

PROFIBUS TRADE ORGANIZATION (PTO), 16101 N. 82nd St., Suite 3B, Scottsdale, AZ 85260, 480/483-2456, Fax: 480/483-7202, www.profibus.com

PROSOFT TECHNOLOGY, INC., 9801 Camino Media, Suite 105, Bakersfield, CA 93311, 661/664-7208, Fax: 661/664-7233, www.prosoft-technology.com

PROSYS, INC., 11814 Coursey Blvd., Suite 408, Baton Rouge, LA 70816, 225/291-9591, Fax: 225/291-9594, www.prosysinc.com

PRO-TECH, 3600A Swiftwater Park Dr, Suwanee, GA 30024, 770/271-0048, Fax: 770/271-2796, www.protech1.com

PSI-TRONIX TECHNOLOGIES, INC., 3950 South K St., Tulare, CA 93274, 559/686-0558, Fax: 559/686-0609, www.psi-tronix.com

PTR CONNEX, 7330 Executive Way, Frederick, MD 21704, 301/696-9433, Fax: 301/696-9493, www.ptr-usa.com

PULSE, A TECHNITROL CO., 12220 World Trade Dr., San Diego, CA 92128, 619/385-8031, Fax: 619/674-8262, www.pulseeng.com

PULIZZI ENGINEERING INC., 3200 S. Susan St., Santa Ana, CA 92704, 714/540-4229, Fax: 714/641-9062, www.pulizzi.com

PULS, 2560 Foxfield Rd. #320, St. Charles, IL 60174, 630/587-9780, Fax: 630/587-9735, www.puls.power.com/us

PULSAR INC., PO Box 799, Shalimar, FL 32579, 850/609-1777, Fax: 850/651-4777, www.pulsar-us.com

PYROMATION, INC., 209 Industrial Pkwy., Fort Wayne, IN 46825, 219/484-2580, Fax: 218/482-6805, www.pyromation.com

PYROMETER INSTRUMENT CO., 209 Industrial Pkwy., Northvale, NJ 07647, 201/768-2000, Fax: 201-768-2570, www.pyrometer.com

Q

QSI CORP., 2212 South W. Temple, #50, Salt Lake City, UT 84115, 801/466-8770, Fax: 801/466-8792, www.qsicorp.com

QUATECH INC., 662 Wolf Ledges Pkwy., Akron, OH 44311, 800/553-1170, Fax: 330/434-1409, www.quatech.com

QUEST INTERNATIONAL, INC., 65 Parker, Irvine, CA 92618, 800/231-6777, Fax: 949/581-4011, www.questinc.com

R

RAE SYSTEMS, 1339 Moffett Park Dr., Sunnyvale, CA 94089, 408/752-0723, Fax: 408/752-0724, www.raesystems.com

R. STAHL, INC., 45 Northwestern Dr., Salem, NH 03079, 603/870-9500, Fax: 603/870-9290, www.rstahl.com

RACO MANUFACTURING ENGINEERING, 1400 62nd St., Emeryville, CA 94608, 510/658-6713, Fax: 510/658-3153, www.racoman.com

RADSTONE TECHNOLOGY CORP., 50 Craig Rd., Montvale, NJ 07645, 201/391-2700, Fax: 201/391-2899, www.radstone.com

RAECO, INC., 9324 Gulfstream Rd., Frankfort, IL 60423, 815/464-6200, Fax: 815/464-8720, www.raeco.com

RAYTEK CORP., 1201 Shaffer Rd., Bldg. 2, Santa Cruz, CA 95061, 831/458-1110, Fax: 831/458-1239, www.raytek.com

RC ELECTRONICS, INC., 6464 Hollister Ave., Santa Barbara, CA 93117, 805/685-7770, Fax: 805/685-5853, www.rcelectronics.com

RC SYSTEMS CO., 2513 Hwy. 646, Santa Fe, TX 77510, 409/925-7808, Fax: 409/925-1078, www.members.aol.com/rcsci/res.htm

RCM INDUSTRIES, INC., 110 Mason Cir., Suite D, Concord, CA 94520-1238, 925/687-8363, Fax: 925/671-9636, www.flo-gage.com

RDF CORP., 23 Elm Ave., Hudson, NH 03051, 603/882-5195, Fax: 603/882-6925, www.rdfcorp.com

REAL-TIME INNOVATIONS, 155A Moffett Park Dr., Sunnyvale, CA 94089, 408/734-4200, Fax: 408/734-5009, www.rti.com

REBIS, 1600 Riviera Ave., Suite 300, Walnut Creek, CA 94596, 925/933-2525, Fax: 925/933-1920, www.rebis.com

RED LION CONTROLS, 20 Willow Springs Circle, York, PA 17402, 717/767-6511, Fax: 717/764-0839, www.redlion-controls.com

RED VALVE CO., 700 N. Bell Ave., Carnegie, PA 15106, 412/279-0044, Fax: 412/279-7878, www.redvalve.com

RELIABLE POWER METERS, 400 Blossom Hill Rd., Los Gatos, CA 95032, 408/358-5100, Fax: 408/358-4420, www.reliablemeters.com

REMOTE CONTROL, INC., 386 Dry Bridge Rd., North Kingstown, RI 02852, 401/294-1400, Fax: 401/294-3388, www.rciactuators.com

REXROTH MECMAN PNEUMATICS, 1953 Mercer Rd., Lexington, KY 40511, 859/254-8031, Fax: 859/281-3491, www.us.rexroth.com

RICE LAKE WEIGHING SYSTEMS, 230 W. Coleman St., Rice Lake, WI 54868, 715/234-9171, Fax: 715/234-6967, www.rlws.com

RITEPRO, INC., A SUBSIDIARY OF BARY INTL., INC., One Rittal Place, Montreal-North, Quebec, H1G 3K7 Canada, 514/324-8900, Fax: 514/324-9525

RITTAL CORP., One Rittal Place, Springfield, OH 45504, 937/390-0500, Fax: 937/390-8392, www.rittal-corp.com

ROBERTSHAW IPD, AN INVENSYS CO., 1602 Mustang Dr., Maryville, TN 37801, 865/981-3100, Fax: 865/981-3168

ROBICON, 500 Hunt Valley Dr., New Kensington, PA 15068, 724/339-9500, Fax: 724/339-9505, www.robicon.com

ROCHESTER INSTRUMENT SYSTEMS, 255 N. Union St., Rochester, NY 14605, 716/238-4078, Fax: 716/454-7805, www.rochester.com

ROCKWELL AUTOMATION, 1201 S. 2nd St., Milwaukee, WI 53204, 414/382-2000, Fax: 414/382-4444, www.automation.rockwell.com

ROCKWELL AUTOMATION, 1 Allen-Bradley Dr., Mayfield Heights, OH 44124, 440/646-5000, Fax: 414/382-4444, www.ab.com

ROCKWELL AUTOMATION PRESENCE SENSING BUSINESS, Two Executive Dr., Chelmsford, MA 01824, 978/446-3476, Fax: 978/446-3322, www.ab.com/sensors

ROCKWELL SOFTWARE, 2424 S. 102nd St., West Allis, WI 53227, 414/321-8000, Fax: 414/321-9647, www.software.rockwell.com

RONAN ENGINEERING CO., PO Box 1275, Woodland Hills, CA 91367, 818/883-5211, Fax: 818/992-6435, www.ronan.com

ROSE & BOPLA, 7330 Executive Way, Frederick, MD 21704, 301/696-9800, Fax: 301/696-9494, www.rose-bopia.com

ROSEMOUNT ANALYTICAL PROCESS ANALYTIC DIV., 1201 N. Main St., Orrville, OH 44667, 330/682-9010, Fax: 330/684-4434, www.processanalytic.com

ROSEMOUNT ANALYTICAL, INC., Uniloc Div., 2400 Barrance Pkwy., Irvine, CA 92606, 949/863-1181, Fax: 949/474-7250, www.rauniloc.com

ROSEMOUNT, INC., 8500 Market Blvd., Chanhassen, MN 55317, 952/949-5165, Fax: 952/949-5114, www.rosemount.com

ROTEK INSTRUMENT CORP., 390 Main St., Waltham, MA 02452, 781/899-4611, Fax: 781/894-7273, www.rotek.com

ROTORK CONTROLS, INC., 19 Jetview Dr., Rochester, NY 14624, 716/328-1550, Fax: 716/328-5848, www.rotork.com

ROTRONIC INSTRUMENT CORP., 160 E. Main St., Huntington, NY 11743, 631/427-3898, Fax: 631/427-3902, www.rotronic-usa.com

ROYCE INSTRUMENT CORP., 13555 Gentilly Rd., New Orleans, LA 70129, 800/347-3505, Fax: 504/254-6855

RTP CORP., 2705 Gateway Dr., Pompano Beach, FL 33069, 954/974-7210, Fax: 954/975-9815, www.rtpcorp.com

RVSI ACUITY CI MATRIX, 5 Shawmur Rd., Canton, MA 02021, 781/821-0830, Fax: 781/828-8942, www.rvsi.com

S

S. HIMMELSTEIN & CO., 2990 Pembroke Ave., Hoffman Estates, IL 60195, 847/843-3300, Fax: 847/843-8938, www.himmelstein.com

SAAB TANK CONTROL, INC., 10700 Hammerly Blvd., Suite 115, Houston, TX 77043, 713/722-9199, Fax: 713/722-9115, www.saabradar.com

SAFEPLEX SYSTEMS INC., 5213 B Tacoma Dr., Houston, TX 77041, 713/983-8410, Fax: 713/983-8428, www.safeplexsystems.com

SAGE METERING INC., 8 Harris Ct., Bldg D-1, Monterey, CA 93940, 866/677-7243, Fax: 831/655-4965, www.sagemetering. com

SAMSON CONTROLS, INC., 4111 Cedar Blvd., Baytown, TX 77520, 281/383-3677, Fax: 281/383-3690, www.samson-usa.com

SANDELIUS INSTRUMENTS, INC., PO Box 30098, Houston, TX 77249, 713/861-1100, Fax: 713/861-9136, www.sandelius.com

SARTORIUS CORP., 131 Heartland Blvd., Edgewood, NY 11717, 631/254-4249, Fax: 631/254-4253, www.sartorius.com

SBS TECHNOLOGIES, INC., 8371C Central Ave., Newark, CA 94560, 510/742-2500, Fax: 510/742-2501, www.sbs.com

SCANIVALVE CORP., 1722 N. Madson St., Liberty Lake, WA 99019, 800/935-5151, Fax: 509/891-9481, www.scanivalve.com

SCHLUMBERGER MEASUREMENT DIV., 1000 Lucas Way, Greenwood, SC 29646-8800, 800/833-3357, Fax: 864/223-0341, www.slb.com/rms/measurement

SCHNEIDER ELECTRIC, 1415 S. Roselle Rd., Palatine, IL 60067, 800/392-8781, Fax: 800/824-7151, www.schneiderautomation.com
SCHROFF NORTH AMERICAN, 170 Commerce Dr., Warwick, RI 02886, 800/451-8755, Fax: 401/738-7988, www.schroffus.com
SCIENTIFIC TECHNOLOGIES, 6550 Dumbarton Circle, Fremont, CA 94555, 800/991-4946, Fax: 510/744-1442, www.sti.com
SCIENTIFIC TECHNOLOGIES INC., 1025 W. 1700 N. Logan, UT 84321, 888/525-7300, Fax: 435/753-7490, www.stiapg.com
SDRG CONTROLS, 8234 Braniff, Houston, TX 77061, 713/242-0822, Fax: 713/644-8294, www.sdrg.com
SEALEVEL SYSTEMS, INC., 155 Technology Place, Liberty, SC 29657, 864/843-4343, Fax: 864/843-3067, www.sealevel.com
SEIKO INSTRUMENTS U.S.A., INC., 2990 West Lomita Blvd., Torrance, CA 90505, 909/975-5637, Fax: 909/975-5699, www.seiko-usa-ecd.com
SELCO PRODUCTS CO., 709 N. Poplar St., Orange, CA 92868, 714/712-6200, Fax: 714/712-6222, www.selcoproducts.com
SENSIDYNE, 16333 Bay Vista Dr., Clearwater, FL 33760, 800/451-9444, Fax: 727/530-3602, www.sensidyne.com
SENSO-METRICS, INC., 4584 Runway St., Simi Valley, CA 93063, 805/527-3640, Fax: 805/584-2960, www.senso-metrics.com
SENSOR ELECTRONICS CORP., 5500 Lincoln Dr., Minneapolis, MN 55436, 952/938-9486, Fax: 952/938-9617, www.sensorelectronic.com
SENSOR PRODUCTS, INC., 188 Rte. 10 W., Suite 307, E. Hanover, NJ 07936, 973/884-1755, Fax: 973/884-1699, www.sensorprod.com
SENSOTEC, INC., 2080 Arlingate Ln., Columbus, OH 43204, 614/850-5000, Fax: 614/850-1111, www.sensotec.com
SEQUENCIA CORP., 15458-B N. 28th Ave., Phoenix, AZ 85053, 602/896-3700, Fax: 602/896-3896, www.sequencia.com
SERVOMEX, 90 Kerry Place, Norwood, MA 02062, 781/769-7710, Fax: 781/769-2834, www.servomex.com
SETRA SYSTEMS, INC., 159 Swanson Rd., Boxborough, MA 01719, 978/266-3629, Fax: 978/264-0292, www.setra.com
SEVERN TRENT SERVICES, Ste. 300, 580 Virginia Dr., Ft. Washington, PA 19034, 215/646-9201, Fax: 215/283-6138, www.severntrentservics.com
SEW EURODRIVE, INC., 1295 Old Spartanburg Hwy., Lyman, SC 29365, 864/439-7537, Fax: 864/661-1276, www.seweurodrive.com
SICK, INC., 6900 W. 110th St., Bloomington, MN 55438, 952/941-6780, Fax: 952/941-9287, www.sickoptic.com
SIEMENS APPLIED AUTOMATION, 500 W. Highway 60, Bartlesville, OK 74003, 918/622-7000, Fax: 918/662-7052, www.aai-us.com
SIEMENS ENERGY & AUTOMATION, INC., 3333 Old Milton Pkwy., Alpharetta, GA 30005, 800/964-4114, Fax: 678/475-5840, www.sea.siemens.com
SIERRA INSTRUMENTS, INC., 5 Harris Court, Bldg. L, Monterey, CA 93940, 800/866-0200, Fax: 831/373-4402, www.sierrainstruments.com
SIERRA MONITOR CORP., 1991 Tarob Court, Milpitas, CA 95035, 408/262-6611, Fax: 408/262-9042, www.sierramonitor.com
SIMPSON ELECTRIC CO., 853 Dundee Ave., Elgin, IL 60120, 847/697-2260, Fax: 847/697-2272, www.simpsonelectric.com
SIMULATION SCIENCES, INC., 601 Valencia Ave., Brea, CA 92823, 714/579-0412, Fax: 714/579-7927, www.simsci.com
SIXNET, PO Box 767, Clifton Park, NY 12065, 518/877-5173, Fax: 518/877-8346, www.sixnetio.com
SL CORP., 240 Tamal Vista Blvd., Corte Madera, CA 94925, 415/927-8400, Fax: 415/927-8401, www.sl.com
SMAR INTERNATIONAL CORP., 7240 Brittmoore #118, Houston, TX 77041, 713/849-2021, Fax: 713/849-2022, www.smar.com
SMITH METER, INC., an FMC Energy Systems Business, 1602 Wagner Ave., Erie, PA 16514, 814/898-5264, Fax: 814/899-8927, www.smithsystems-inc.com
SMOOT CO., 1250 Seminary, Kansas City, KS 66103, 913/362-1710, Fax: 913/362-7863, www.smootco.com
SNELL INFRARED, PO Box 6, Montpelier, VA 05601, 800/636-9820, Fax: 802/223-0460, www.snellinfrared.com
SOFTPLC CORP., 25603 Red Brangus Dr., Spicewood, TX 78669, 512/264-8390, Fax: 512/264-8399, www.softplc.com
SOFTWARE HORIZONS, INC., 100 Treble Cove Rd., N. Billerica, MA 01862, 978/670-8700, Fax: 978/670-8787, www.instanthmi.com
SOFTWARE TOOLBOX, INC., 148A E. Charles St., Matthews, NC 28105, 704/849-2773, Fax: 704/849-6388, www.softwaretoolbox.com
SOLA/HEVI-DUTY, 7770 N. Frontage Rd., Skokie, IL 60077, 800/377-4384, Fax: 800/367-4384, www.sola-hevi-duty.com
SOLAREX, 630 Solarex Ct., Frederick, MD 21703, 301/698-4200, Fax: 301/698-4201, www.solarex.com
SOLARTRON, INC., 19408 Park Row, Suite 320, Houston, TX 77084, 281/398-7890, Fax: 281/398-7891, www.solartron.com
SOLBERG MFG., INC., 1151 W. Ardmore Ave., Itasca, IL 60143, 630/773-1363, Fax: 630/773-0727, www.solbergmfg.com
SONY PRECISION TECHNOLOGY AMERICA, 20381 Hermana Circle, Lake Forest, CA 92630, 949/770-8400, Fax: 888/910-7669, www.sonypt.com
SORENSEN, 9250 Brown Deer Rd., San Diego, CA 92121-2294, 858/450-0085, Fax: 858/458-0267, www.sorensen.com

SPARLING INSTRUMENTS, INC., 4097 N. Temple City Blvd., El Monte, CA 91731, 626/444-0571, Fax: 626/452-0723, www.sparlinginstruments.com

SPC PRESS/STATISTICAL PROCESS CONTROLS, INC., 5908 Toole Dr., Suite C, Knoxville, TN 37919, 423/584-5005, Fax: 423/588-9440, www.spcpress.com

SPECTRUM CONTROLS, 1605 MW Sammamish, Rd. #200, Issaquam, WA 98027, 425/746-9481, Fax: 425/641-9473, www.spectrumcontrols.com

SPENCE ENGINEERING CO., INC., 150 Coldenham Rd., Walden, NY 12586, 800/398-2493, Fax: 914/778-1072, www.spenceengineering.com

SPIRAX SARCO, INC., 1150 Northpoint Blvd., Blythewood, SC 29016, 803/714-2071, Fax: 803/714-2224, www.spiraxsarco-usa.com

SPONSLER CO., INC., 2363 Sandifer Blvd., Westminster, SC 29693, 864/647-2065, Fax: 864/647-1255, www.sponsler.com

SST, A PART OF WOODHEAD CONNECTIVITY, 50 Northland Rd., Waterloo, Ontario, NV2 IN3 Canada, 519/725-5136, Fax: 519/725-1515, www.sstech.on.ca

STACOSWITCH, 1139 Baker St., Costa Mesa, CA 92626, 714/549-3041, Fax: 714/549-0930, www.stacoswitch.com

STAHLIN, 500 Maple St., Belding, MI 48809, 616/794-0700, Fax: 616/794-7564, www.stahlin.com

STANLEY ELECTRIC, 2660 Barranca Pkwy., Irvine, CA 92606, 949/222-0777, Fax: 949/222-0555, www.stanleyelec.com

STAT-EASE, INC., 2021 E. Hennepin Ave, #191, Minneapolis, MN 55413, 612/378-4449, Fax: 612/378-2152, www.statease.com

STATISTICAL PROCESS CONTROLS, INC., 5908 Toole Dr., Suite C, Knoxville, TN 37919, 865/584-5005, Fax: 865/588-9440, www.spcpress.com

STEALTH COMPUTER CORP., 4-530 Roundtree Dairy Rd., Woodbridge, ON,CANADA L4L 8H2, 905/264-9000, Fax: 905/264-7440, www.stealthcomputer.com

STOCHOS, INC., 14 N. College St., Schenectady, NY 12305, 518/372-5426, Fax: 518/372-4789, www.stochos.com

STONEL, One StoneL Dr., Fergus Falls, MN 56537, 218/739-5774, Fax: 218/739-5776, www.stonel.com

SUN ELECTRONIC SYSTEMS, INC., 1900 Shepard Dr., Titusville, FL 32780, 321/383-9400, Fax: 321/383-9412, www.sunelectronics.com

SUNX SENSORS, 1207 Maple St., West Des Moines, IA 50265, 800/280-6933, Fax: 515/225-0063, www.sunx-ramco.com

SUPERIOR ELECTRIC, 383 Middle St., Bristol, CT 06010, 860/585-4500, Fax: 860/584-1483, www.superiorelectric.com

SVF FLOW CONTROLS, INC., 13560 Larwin Cir., Santa Fe Springs, CA 90670, 562/802-2255, Fax: 562/802-3114, www.svfflowcontrols.com

SWAGELOK, 31400 Aurora Rd., Solon, OH 44139, 440/349-5934, Fax: 440/349-5843, www.swagelok.com

SYMCOM, INC., 2880 N. Plaza Dr., Rapid City, SD 57702, 605/348-5580, Fax: 605/348-5685, www.symcominc.com

SYNERGETIC, 2506 Wisconsin Ave., Downers Grove, IL 60515, 630/434-1770, Fax: 630/434-1987, www.synergetic.com

T

TA ENGINEERING CO., INC., 1150 Moraga Way, Moraga, CA 94556, 925/376-8500, Fax: 925/376-4977, www.aimax.com

TAIYO YUDEN (USA), INC., Power Systems Group, 1770 La Costa Meadows Dr., San Marcos, CA 95131, 760/510-3200, Fax: 760/471-4021, www.t-yuden.com

TEAC AMERICA, INC., 7733 Telegraph Rd., Montebello, CA 90640, 323/726-0303, Fax: 323/727-7674, www.teac.com

TECHNE, INC., 743 Alexander Rd., Princeton, NJ 08540, 609/452-9275, Fax: 609/987-8177, www.techneusa.com

TECHNICAL MARINE SERVICE, INC., 6040 N. Cutter Cir., Suite 302, Portland, OR 97217, 503/285-8947, Fax: 503/285-1379, www.tms-usa.com

TECHNOLAND, INC., 1050 Stewart Dr., Sunnyvale, CA 94086, 408/992-0888, Fax: 408/992-0808, www.technoland.com

TECNOMATIX TECHNOLOGIES, 21500 Haggerty Rd., Suite 300, Noriville, MI 48167, 248/471-6140, Fax: 248/471-6147, www.tecnomatix.com

TEL-TRU MANUFACTURING CO., 408 St. Paul St., Rochester, NY 14605, 716/232-1440, Fax: 716/232-3857, www.teltru.com

TELEDYNE ANALYTICAL INSTRUMENTS, 16830 Chestnut St., City of Industry, CA 91748, 626/934-1507, Fax: 626/961-2538, www.teledyne-ai.com

TELETROL SYSTEMS, INC., 286 Commercial St., Manchester, NH 03101, 603/645-6061, Fax: 603/645-6174, www.teletrol.com

TEMPROX, 2915 Parkway St., Lakeland, FL 33811, 863/619-5999, Fax: 863/619-5274, www.temprox.com

TESCOM CORP., 12616 Industrial Blvd., Elk River, MN 55330, 763/441-6330, Fax: 763/241-3224, www.tescom.com

TESTO, INC., 35 Ironia Rd., Flanders, NJ 07836, 800/227-0724, Fax: 973/252-1724, www.testo.com

TEXAS INDUSTRIAL PERIPHERALS, 2621 Ridgepoint Dr. Suite 235, Austin, TX 78754, 800/866-6506, Fax: 512/837-0207, www.ikey.com

TEXAS MICRO, INC., 5959 Corporate Dr., Houston, TX 77036, 713/541-8200, Fax: 713/541-8226, www. texasmicro.com

THE SIEMON COMPANY, 27 Siemon Company Dr., Watertown, CT 06795, 860/945-4200, Fax: 860/945-8503, www.siemon.com

THERMACAL, INC., 30275 Bainbridge Rd., Solon, OH 44139, 440/498-1005, Fax: 440/498-1062, www.thermacal.com

THERMAL INSTRUMENTS CO., 217 Sterner Mill Rd., Trevose, PA 19053, 215/355-8400, Fax: 215/355-1789, www.thermalinstrument.com

THERMO ANDERSEN, 500 Technology Court, Smyrna, GA 30082, 770/319-9999, Fax: 770/319-0336

THERMO BLH, 75 Shawmut Rd., Canton, MA 02021, 781/821-2000, Fax: 781/828-1451, www.thermoblh.com

THERMO BRANDT INSTRUMENTS, 3333 Air Dark Rd., Fuquay, NC 27526, 919/552-9011, Fax: 919/552-9717, www.brandtinstruments.com

THERMO ELECTRIC CO., INC., 109 N. Fifth Ave., Saddle Brook, NJ 07663, 201/843-5800, Fax: 201/843-4568, www.thermo-electric-direct.com

THERMO ELECTRON CORP., PROCESS INSTRUMENTS DIV., 9303 W. Sam Houston Pkwy. S., Houston, TX 77099, 713/272-0404, Fax: 713/272-4573, www.thermo.com

THERMO GASTECH, 8407 Central Ave., Newark, CA 94560, 510/745-8700, Fax: 510/794-6201, www.thermogastech.com

THERMO MEASURETECH, 2555 North IH-35, Round Rock, TX 78664, 800/736-0801, Fax: 512/388-9200, www.thermomt.com

THERMO NICOLET CORP., 5225 Verona Rd., Madison, WI 53711, 608/276-6100, Fax: 608/273-5046, www.thermonicolet.com

THERMO ONIX, 1201 N. Velasco, Angleton, TX 77515, 979/849-2344, Fax: 979/849-2166, www.thermoonix.com

THERMO POLYSONICS, 10335 Landsbury Dr. #300, Houston, TX 77099, 281/879-3700, Fax: 281/498-7721, www.thermopolysonics.com

THERMO RAMSEY, 501-90th Ave. NW, Minneapolis, MN 55433, 763/783-2500, Fax: 763/780-2315, www.thermoramsey.com

THERMO WESTRONICS, 22001 North Park Dr., Suite 100, Kingwood, TX 77339, 281/348-1800, Fax: 281/348-1288, www.thermowestronics.com

THIELSCH ENGINEERING, 195 Frances Ave., Cranston, RI 02910, 401/467-6454, Fax: 401/467-6454, www.thielsch.com

TIPS, INC., 2402 Williams Dr., Georgetown, TX 78628, 512/863-3653, Fax: 512/863-5392, www.tipswweb.com

TISCOR, 12250 Parkway Centre Dr., Poway, CA 92064, 800/227-6379, Fax: 858/513-8497, www.tiscor.com

TOL-O-MATIC, 3800 CR 116, Hamel, MN 55340, 612/478-4322, Fax: 612/478-8080, www.tolomatic.com

TOPWORX, 3300 Fern Valley Rd., Louisville, KY 40213, 502/969-8000, Fax: 502/964-5911, www.topworx.com

TOSHIBA INTERNATIONAL CORP., 13131 West Little York Rd., Houston, TX 77041, 713/466-0277, Fax: 713/896-5225, www.tic.toshiba.com

TOTAL CONTROL PRODUCTS, INC., 2001 N. Janice Ave., Melrose Park, IL 60160, 708/345-5500, Fax: 708/345-5670, www.total-control.com

TOUCH CONTROLS, INC., 520 Industrial Way, Fallbrook, CA 92028, 800/848-4385, Fax: 760/723-7910, www.touchcontrols.com

TRACEWELL SYSTEMS, INC., 567 Enterprise Dr., Columbus, OH 43081, 800/848-4525, Fax: 614/846-4450, www.tracewellsystems.com

TRANSCAT, 10 Vantage Pt. Dr., Rochester, NY 14624, 800/828-1470, Fax: 800/395-0543, www.transcat.com

TRANSDUCER TECHNIQUES, 43178 Business Park Dr., Temecula, CA 92590, 909/676-3965, Fax: 909/676-1200, www.ttloadcells.com

TRANSICOIL, 43178 Business Park Dr., Norristown, PA 19403, 800/323-7115, Fax: 616/539-3400, www.transicoil.com

TRANSMATION, 35 Vantage Point Dr., Rochester, NY 14624, 716/349-3520, Fax: 716/349-3510, www.transmation.com

TRANSTECTOR SYSTEMS, 10701 Airport Dr., Hayden Lake, ID 83835, 208/762-6055, Fax: 208/762-6080, www.transtector.com

TRANSYSOFT, INC., 11 Merrill Dr., Suite C5, Hampton, NH 03842, 603/929-6330, Fax: 603/929-6331, www.transyssoft.com

TRENDVIEW RECORDERS, PO Box 141489, Austin, TX 78714-1489, 512/927-7800, Fax: 512/834-4333, www.trendview.com

TRENTON TECHNOLOGY, INC., 2350 Centennial Dr., Gainesville, GA 30504, 800/875-6031, Fax: 770/287-3150, www.trentonprocessors.com

TRICONEX CORP., 15345 Barranca Pkwy., Irvine, CA 92618, 949/885-0714, Fax: 949/753-9101, www.triconex.com

TRIHEDRAL ENGINEERING LTD, 1160 Bedford Hwy., Suite 400, Bedford, Nova Scotia, B4A ICI Canada, 902/835-1575, Fax: 902/835-0369, www.trihedral.com

TRIPLETT CORP., One Triplett Dr., Bluffton, OH 45817, 800/TRIPLETT, Fax: 419/358-7956, www.triplett.com

TRW SENSORS & COMPONENTS SCHAEVITZ SENSORS, 1000 Lucas Way, Hampton, VA 23666, 757/766-1500, Fax: 757/766-4297, www.schaevitz.com

TSI, INC., 500 Cardigan Rd., Shoreview, MN 55126, 651/490-2711, Fax: 651/490-2874, www.tsi.com

TTI, 8 Leroy Rd., Williston, VT 05495, 800/235-8367, Fax: 802/863-1193, www.ttiglobal.com

TURCK, INC., 3000 Campus Dr., Minneapolis, MN 55441, 800/544-7769, Fax: 612/553-0708, www.turck.com

TUTHILL TRANSFER SYSTEMS, 8825 Aviation Dr., Fort Wayne, IN 46809, 919/460-6000, Fax: 919/460-7595, www.tuthill.com

TWO TECHNOLOGIES, INC., 419 Sargon Way, Horsham, PA 19044, 215/441-5305, Fax: 215/441-0423, www.2t.com

TYCO ELECTRONICS, PO Box 3608, MS38-41, Harrisburg, PA 17105, 800/522-6752, Fax: 717/986-7575, www.tycoelectronics.com

TYCO THERMAL CONTROLS, 300 Constitution Dr., Menlo Park, CA 94025, 800/545-6258, Fax: 800/527-5703, www.tycothermal.com

TYCO VALVES & CONTROLS LP, 9700 W. Gulf Rd., Houston, TX 77040, 713/466-1176, www.tycovalves.com

U

UE SYSTEMS, INC., 14 Hayes St., Elmsford, NY 10523, 800/223-1325, Fax: 914/347-2181, www.uesystems.com

ULTRAFLO CORP., a Subsidiary of Bray Intl., Inc., PO Box 423, St. Genevieve, MO 63670, 573/883-8881, Fax: 573/883-8882

ULTRAMAX CORP., 110 Boggs Ln., Suite 255, Cincinnati, OH 45246, 513/771-8629, Fax: 513/771-7185, umaxcorp.com

UNICOM, 908 Canada Ct., City of Industry. CA, 91748, 800/346-6668, Fax: 626/964-7880, www.unicomlink.com

UNITED ELECTRIC CONTROLS, 180 Dexter Ave., Watertown, MA 02471, 617/926-1000, Fax: 617/926-4354, www.ueonline.com

UNIVERSAL DYNAMICS TECHNOLOGIES, INC., 100-13700 International Place, Richmond, British Columbia, V6V 2X8 Canada, 604/214-3456, Fax: 604/214-3457, www.brainwave.com

UNIVERSAL FLOW MONITORS INC., 1755 E. 9 Mile Rd., Hazel Park, MI 48030, 248/542-9635, Fax: 248/398-4274, www.flowmeters.com

USDATA, 2435 N. Central Expwy., Richardson, TX 75080, 972/497-0233, Fax: 972/669-9556, www.usdata.com

USFILTER/WALLACE & TIERNAN PRODUCTS, 1901 W. Garden Rd., Vineland, NJ 08360, 856/507-9000, Fax: 856/507-4125, www.wallaceandtiernan.usfilter.com

V

VAISALA, INC., 100 Commerce Way, Woburn, MA 01801, 781/933-4500, Fax: 781/933-8029, www.vaisala-usa.com

VALCOR SCIENTIFIC, 2 Lawrence Rd., Springfield, NJ 07081, 973/467-8400, Fax: 973/467-9592, www.valcor.com

VALIDYNE ENGINEERING, 8626 Wilbur Ave., Northridge, CA 91324, 818/886-2057, Fax: 818/886-6512, www.validyne.com

VALVE ACCESSORIES & CONTROLS, 12585 Old Hwy. 280, Unit 101, Chelsea, AL 35043, 205/678-0507, Fax: 205/678-0510, www.vacaccessories.com

VAREC INC., 5834 Peachtree, Comers East, Norcross, GA 30096, 281/498-9202, Fax: 281/498-0183, www.varec.com

VAS ENGINEERING, 4750 Viewridge Ave., San Diego, CA 92123, 619/569-1601, Fax: 619/569-8474, www.doric-vas.com

VERANO, 310 E. Caribbean Dr., Sunnyvale, CA 94089, 408/541-7658, Fax: 408/541-7601, www.verano.com

VERIFLO DIVISION, PARKER HANNIFIN CORP., 250 Canal Blvd., Richmond, CA 94947, 510/412-1166, Fax: 510/232-7396, www.veriflo.com

VERIS, INC., 6315 Monarch Park Place, Niwot, CO 80503, 303/652-8550, Fax: 303/652-8552, www.veris-inc.com

VERSALOGIC CORP., 3888 Stewart Rd., Eugene, OR 97402, 541/485-8575, Fax: 541/485-5712, www.versalogic.com

VERTACROSS, INC., PO Box 14526, Research Triangle Park, NC 27709, 866/248-4088, Fax: 919/248-4099, www.vertacross.com

VIA DEVELOPMENT CORP., PO Box 3268, Marion, IN 46953, 765/677-3232, Fax: 765/674-3964, www.viadevelopment.com

VIATRAN CORP., 300 Industrial Dr., Grand Island, NY 14072, 716/773-1700, Fax: 716/773-2488, www.viatran.com

VICOR CORP., 25 Frontage Rd., Andover, MA 01810, 978/470/2900, Fax: 978/475-6715, www.vicr.com

VIKING PUMP, INC., 406 State St., Cedar Falls, IA 50613, 319/266-1741, Fax: 319/273-8157, www.vikingpump.com

VISHAY BLH, 75 Shawmut Rd., Canton, MA 02021, 781/298-2282, Fax: 781/828-1451, www.blh.com

VISIONEX, 430 Tenth St. NW, #N205, Atlanta, GA 30318, 404/873-9775 x.14, Fax: 404/873-0535, www.mindspring.com/~visionex

VISTA CONTROLS SYSTEMS, INC., 176 Central Park Square, Los Alamos, NM 87544, 505/662-2484, Fax: 505/662-3956, www.vista-control.com

VISUAL SOLUTIONS, INC., 487 Groton Rd., Westford, MA 01886, 978/392-0100, Fax: 978/692-3102, www.vissim.com

VMIC, 12090 S. Memorial Pkwy., Huntsville, AL 35803, 256/880-0444, Fax: 256/882-0859, www.vmic.com

VMR SOFTWARE, PO Box 1463, Edmonds, WA 98020, 425/774-2483, Fax: 206/727-8650, www.vmrsoftware.com

VORNE INDUSTRIES, INC., 1445 Industrial Dr., Itasca, IL 60143, 888/DISPLAYS, Fax: 630/875-3609, www.vorne.com

VYNCKIER ENCLOSURE SYSTEMS LTD, 249 McCarty Dr., Houston, TX 77029, 713/374-7850, Fax: 713/672-8632, www.enclosuresonline.com

W

WAGO CORP., N120 W19129 Freistadt Rd., Germantown, WI 53022, 800/DINRAIL, Fax: 262/255-3232, www.wago.com

WALCHEM CORP., 5 Boynton Rd., Holliston, MA 01746, 508/429-1110, Fax: 508/429-7433, www.walchem.com

WARNER ELECTRIC MOTORS & CONTROLS, 383 Middle St., Bristol, CT 06010, 860/585-4500, Fax: 860/582-3784, www.warnernet.com/sev_main.html

WATERS EQUIPMENT CO., PO Box 576, Lansdale, PA 19446, 215/699-8700, Fax: 215/699-8795, www.watersequipment.com

WATLOW, 12001 Lackland Rd., St. Louis, MO 63146, 800/4-WATLOW, Fax: 314/878-6814, www.watlow.com

WEED INSTRUMENT, 707 Jeffrey Way, Round Rock, TX 78664, 512/434-2844, Fax: 512/434-2851, www.weedinstrument.com

WEG MOTORS AND DRIVES, 1327 Northbrook Pkwy., Suwanee, GA 30024, 770/338-5656, Fax: 770/338-1632, www.webelectric.com

WEIDMULLER, INC., 821 Southlake Blvd., Richmond, VA 23236, 804/379-6027, Fax: 804/379-2593, www.weidmuller.com

WEIGH-TRONIX, 1000 Armstrong Dr., Fairmont, MN 56031, 507/238-8253, Fax: 507/238-8258, www.weigh-tronix.com

WESCO DISTRIBUTION, INC., 1600 N. Sixth St., Milwaukee, WI 53212, 414/264-6400, www.wescodist.com

WESTERN RESERVE CONTROLS, 1485 Exeter Dr., Akron, OH 44306, 330/733-6662, Fax: 330/733-6663, www.wrcakron.com

WESTINGHOUSE PROCESS CONTROL, 200 Beta Dr., Pittsburgh, PA 15238, 412/963-2727, Fax: 412/963-3644, www.westinghousepc.com

WESTLOCK CONTROLS CORP., 280 Midland Ave., Saddle Brook, NJ 07663, 201/794-7650, Fax: 201/794-0913

WHESSOE VAREC, 10800 Valley View St., Cypress, CA 90630, 714/761-1300, Fax: 714/952-2701, www.whessoevarec.com

W.I. GORE-ELECTRONICS, 402 Vieve's Way, Elkton, MD 21922, 800/445-4673, www.gore.com/electronics

WIELAND ELECTRIC, 49 International Rd., Burgaw, NC 28425, 910/259-5050, Fax: 910/259-3691, www.wielandinc.com

WIKA INSTRUMENT CORP., 1000 Wiegand Blvd., Lawrenceville, GA 30043, 770/338-5229, Fax: 770/277-2668, www.wika.com

WILCOXON RESEARCH, 21 Firstfield Rd., Gaithersburg, MD 20878, 800/Wilcoxon, Fax: 301/330-8873, www.wilcoxon.com

WILKERSON INSTRUMENT CO., INC., 2915 Parkway St., Lakeland, FL 33811, 800/234-1343, Fax: 863/644-5318, www.wici.com

WILLIAMSON CORP., 70 Domino Dr., Concord, MA 01742, 978/369-9607, Fax: 978/369-5485, www.williamsonir.com

WINTERS INSTRUMENTS, 600 Ensminger Rd., Buffalo, NY 14150, 716/874-8700, Fax: 716/874-8700, www.winters.ca

WOLFRAM RESEARCH, INC., 100 Trade Center Dr., Champaign, IL 61820, 217/398-0700, www.wolfram.com

WONDERWARE, 100 Technology Dr., Irvine, CA 92618, 949/453-6568, Fax: 949/450-5098, www.wonderware.com

WOODHEAD CONNECTIVITY, 3411 Woodhead Dr., Northbrook, IL 60062, 847/272-7990, Fax: 847/272-8133, www.connector.com

WORCESTER CONTROLS CORP., 33 Lock Drive, Marlborough, MA 01752, 508/481-4800, Fax: 508/481-4454, www.worcestercc.com

X

XENTEK POWER SYSTEMS, 1770 La Costa Meadows Dr., San Marcos, CA 92069, 760/471-4001, Fax: 760/471-4021, www.xentek.com

XYCOM AUTOMATION, INC., 750 N. Maple Rd., Saline, MI 48176, 734/429-4971, Fax: 734/429-3087, www.xycom.com

XYMOX TECHNOLOGIES, INC., 9099 W. Dean Rd., Milwaukee, WI 53224, 414/362-9000, Fax: 414/362-9090, www.xymoxtech.com

XYNTEK, INC., 301 Oxford Valley Rd., Bldg. 1402A, Yardley, PA 19067, 215/493-7091, Fax: 215/493-7094, www.xyntekinc.com

Y

Y2 SYSTEMS, 3101 Pollok Dr., Conroe, TX 77303, 936/788-5526, Fax: 936/788-5698, www.y2systems.com

YASKAWA ELECTRIC AMERICA, INC., 2121 Norman Dr. S., Waukegan, IL 60085, 800/YASKAWA, Fax: 847/887-7310, www.yaskawa.com

YCV, A YAMATAKE CO., 11225 North 28th Dr., Suite A106, Phoenix, AZ 85029, 602/548-1800, Fax: 602/548-1127, www.ycv.com

YOKOGAWA CORP. OF AMERICA, 2 Dart Road, Newnan, GA 30265, 770/258-2552, www.yca.com

Z

Z-WORLD, 2900 Spafford St., Davis, CA 95616, 530/757-3737, www.zworld.com

ZELLWEGER ANALYTICS, INC., 405 Barclay Blvd., Lincolnshire, IL 60069, 847/955-8200, Fax: 847/955-8208, www.zelana.com

ZENITH PRODUCTS DIV., PARKER HANNIFIN CORP., 5910 Elww Buchanan Dr., Sanford, NC 27330, 919/775-4600, Fax: 919/774-5952, www.zenithpumps.com

ZETEC INC., 1370 NW Mall St., Issaquah, WA 98027, 425/392-5316, Fax: 425/392-2086, www.zetec.com

ZONEWORX, 40925 County Center Dr., Suite 200, Temecula, CA 92591, 909/296-1226, Fax: 909/506-9309, www.zoneworx.com

ZONTEC INC., 1389 Kemper Meadow Dr., Cincinnati, OH 45240, 513/648-0088, Fax: 513/648-9007, www.zontec-spc.com

A.9 Directory of "Lost" Companies

W. H. BOYES WITH S. EDVI (2005)

Between 1990 and 2005, the instrumentation, systems and automation industries were the subject of significant, not to say massive, change. Not only was this change of a technical nature, with significant advances in many product areas, but also this period has been one of consolidation among supplier companies who furnish instruments and control systems. It appears that this consolidation will continue for the foreseeable future.

Identified in this appendix are over 100 companies whose name, ownership, or product line changed significantly over the period from 1990 to 2005. This list is not complete, and the authors believe that there are perhaps as many as 50 more companies, most outside North America, whose names have also changed in the same fashion.

This turmoil among suppliers has had a significant effect on the ability of Instrument Engineers to find familiar products, and locate companies for service of existing installed equipment. For example, the Fischer & Porter Company, a well-known name in 1990, ceased to exist as a marque in 2002. The company itself was purchased by Elsag BV, and merged with Bailey Controls to form Bailey-Fischer-Porter. Later, Elsag sold the combined companies to ABB, who elected to drop both trade names in favor of ABB Process Instrumentation. So, an Instrument Engineer whose company installed Fischer & Porter instrumentation or control systems in the early 1990s may now find it impossible to easily find information, manuals, and source spare parts and service for Fischer & Porter systems. Fischer & Porter, as you will see in the table below, is far from the only example of this problem. The authors therefore provide this information as a public service to the profession.

Prior Name	Intermediate, If Any	2005
Advanced Pollution Instrumentation Inc.		Teledyne Instruments - API
AEG Automation Systems		Cegelec
AEG Schneider Automation	Groupe Schneider/Schneider Electric	Schneider Automation
AGEMA Infrared Systems		FLIR Systems Inc.
Altek Calibrators		Transmation
American Sigma		American Sigma div. of Danaher
Ametek McCrometer	Ketema-McCrometer	McCrometer div. of Danaher
Amscor	Fluid Data; ThermoONIX	ThermoElectron Measurement and Control
Anatel	Anatel Div. of Danaher	Pacific Scientific Instruments div. of Danaher
Anchor Darling		Flowserve Corporation
Applied Automation Inc	Elsag-Bailey	ABB Process Automation (much of AAI was sold to Siemens)
Aquamatic Inc.	Osmonics, Inc.	GE Infrastructure Sensing
ARCOM Control Systems		Spectris PLC
Argus	Invensys Flow Control	Flowserve Corporation
Arthur Technology		Respirometry Plus
ASAP Software	Xycom Automation/Pro-face America div. of Digital-Japan	Schneider Electric
Ashcroft-Heise		Ashcroft-Heise div. of Dresser Industries Instruments
Astro	Hach Co.	Hach div. of Danaher
Atomac		Flowserve Corporation
Automax		Flowserve Corporation
Azonix		Crane Co.
BTG	BTG Spectris	Spectris PLC BTG div.
B/W Controls	Ametek Patriot Sensors	Ametek Automation & Process Technology
Baan Software	Invensys	SSA Global Technologies
Badger Meter Ultrasonic div.		Badger Eastech Inc.
Bailey Controls	Elsag-Bailey	ABB Process Automation
Barber-Coleman	Eurotherm	Invensys
Battig		Flowserve Corporation
Bebco		Pepperl+Fuchs
Beta Products	Hathaway-Beta and Beta Calibrators div. Hathaway Process Instrumentation Corp.	Martell Electronics
Bettis Actuators & Controls		Emerson Process Measurement
Bindicator	Aaliant	Venture Meas. Div. of Danaher
Biosynergy Inc		Biosynergy Inc div. of America Clinical Laboratory
Brooks Instruments	Rosemount-Brooks	Emerson Process management (no longer a div. only a trademark)
Bruel & Kjaer	B&K div. Of Fairey Group	B&K div of Spectris PLC
Edmond Buhler GMBH	Buhler Montec	Buhler Montec div. of Danaher
CSI	Emerson CSI	Emerson Process Management
Cannonbear Technologies	TN Technologies; ThermoMeasureTech	ThermoElectron Measurement and Control
Celtek	Bindicator Celtek, Aaliant	Venture Meas. div. of Danaher
Chessell	Eurotherm	Invensys
Chino Works Ltd.		Chino Corporation

Prior Name	Intermediate, If Any	2005
Combustion Engineering		ABB Process Automation
Danfoss Flow Control	Siemens Danfoss	Siemens
Daniel Industries	Rosemount	Emerson Process Measurement, Daniel div.
Detector Electronics Corp		Det-Tronics
DeZurik Valve Corp.	DeZurik Div. Of SPX	SPX Process Equipment
Dietrich Standard (Annubar)	Rosemount	Emerson Process Measurement, Dietrich div.
Direct Measurement Corp.	DMC	FMC Energy Systems Group
Dolch Computer Co.	Dolch Kontron	Kontron Inc.
Drexelbrook Eng. Co.		Ametek Drexelbrook
Druck Inc.	GE Druck	GE Infrastructure Sensing
DuPont Process Instruments		Ametek Process Instruments Div.
Durco		Flowserve Corporation
Eastech Controls	Neptune Eastech	Badger Eastech Inc.
Eckhardt	Siebe	Invensys
Edward Valve	Invensys Flow Control	Flowserve Corporation
Elsag-Bailey		ABB Process Automation
Elcon Inc.		Pepperl+Fuchs
Eurcontrol	BTG	Spectris div. of Fairey
Ever Ready Thermometer Company		Apogent Technologies Co.
Fairey Group PLC	Purchased Spectris	Spectris PLC
Fenwal		Kidde-Fenwal
Fischer and Porter Co.	Elsag-Bailey	ABB Process Automation
Fischer and Porter Co. (chlorination div. only)		Capital Controls Co. div. of Severn Trent
Fisher Controls Inc.	Fisher-Rosemount Inc.	Emerson Process Measurement
Flow Research	Polysonics	ThermoElectron Measurement and Control
Fluenta AB		Roxar Flow Measurement
Fluid Data Inc.	Thermo ONIX	Thermo Electron Measurement and Control
Forma Scientific Co.		Forma Scientific Co. div. of Thermo Electron
Foxboro	Siebe	Invensys
FTS Systems Inc.		FTS Kinetics
Gas Measuring Technology		RMG Group - Germany
Gemco Controls	Ametek Patriot Sensors	Ametek Automation & Process Technology
GEMS Sensors	IMO-Delaval Gems	GEMS-Danaher
General Eastern Inc.	GE General Eastern	GE Infrastructure Sensing
GE Kaye Instruments	GE Kaye Instruments div. of GE Industrial Systems	GE Infrastructure Sensing
Gestra	Invensys Flow Control	Flowserve Corporation
Gilbarco		Gilbarco div. of Danaher
Great Lakes Instruments	GLI-Viridor/Danaher-GLI	Hach div. Danaher
Griffith Industrial		Winters Instrumentation
Groupe Schneider	Schneider Electric	Schneider Automation
HBM	HBM div. Of Fairey Group	Spectris PLC
Hach Chemical Co.	Hach Inc.	Hach div. of Danaher
Hart Scientific	Hart Scientific div. of Fluke	Hart Scientific div. of Danaher Electronic Test and Measurement
Hartmann and Braun	Elsag-Bailey	ABB Process Automation

Prior Name	Intermediate, If Any	2005
Hastings Instrument Co.	Teledyne Hastings	Teledyne Instruments Group- Hastings
Heinrichs Messtechnik		Bopp and Reuther
Heise Instruments		Ashcroft-Heise div. of Dresser Industries Instruments
Helicoid Instruments		Helicoid Instruments div. of Bristol Babcock
Hewlett Packard	HP Instruments	Agilent Technologies
Hiac-Royco	Pacific Scientific Inc.	Pacific Scientific Instrumentation div. of Danaher
Honeywell Control Valve div.	DeZurik Corp.	SPX Process Equipment
Hydrolab Inc.	Hydrolab div. of Danaher	Hach div. of Danaher
Hyprotech	Aspen Technology HYSIS and OTS div.	Honeywell Unisym
Hyprotech AXSYS	Aspen Technology	Bentley Systems Inc.
InLine Measurements		Honeywell Inc.
Innovative Sensors Inc.	ISI	Endress+Hauser
Intellution	Emerson Process Management	GE Fanuc Automation
Invalco Inc (Smith Meters)	FMC Invalco	FMC Energy Systems Group
IRCON Inc.	IRCON div of Fairey Group	Spectris PLC
ISA Instruments	Solartron ISA	Solartron ISA div. of Roxboro
ISCO Environmental		Recently sold to Teledyne Instrument Group
John Fluke Company	Fluke Inc.	Fluke div. of Danaher Electronic Test and Measurement
Fluke Inc. (networks and communications only)	Fluke Networks Inc.	Fluke Networks div. of Danaher Electronic Test and Measurement
Joseph Kaye Co.		GE Kaye Instruments
Kammer		Flowserve Corporation
Kay-Ray	Rosemount-Kay-Ray; ThermoMeasureTech	ThermoElectron Measurement and Control
KDG Mobrey Ltd.	KDG Mobrey Ltd. div. of Solartron/div. of Roxboro	Mobrey Measurements div. of Emerson Process Mgt.
Kent Meter Co.	Kent Taylor	ABB Process Automation
Kistler-Morse	(ultrasonics to American Sigma)/Kistler-Morse div. of Danaher	Venture Meas. Div. of Danaher
Kollmorgen	Kollmorgen div. of Danaher	Dahaher Motion- General Purpose Systems
Lakewood Instruments	Fisher-Rosemount Inc.; Osmonics Inc.; GE Osmonics	GE Infrastructure Sensing
Lear Siegler Measurement Systems	Bowthorpe plc	Spirent plc
Leeds and Northrop Inc.	Honeywell AIC	Honeywell Inc.
Leeds and Northrup Inc. (DCS only)		ICS
Leeman Labs Inc.		Teledyne Instruments Leeman Labs
Lighthammer, Inc.		SAP
Limitorque	Invensys Flow Control	Flowserve Corporation
Liquid Metronics Inc.	LMI	Milton Roy Flow Controls, LMI div.
Loveland Controls Ltd.	Love Controls	Dwyer Instruments
Lumidor Safety Products	Invivo Corporation Zellweger Analytical	Honeywell Inc.
M & J Valves Inc.		SPX Process Equipment
MagneTEK	Ametek Patriot Sensors	Ametek Automation & Process Technology
Maihak		Sick Maihak
Malvern Instruments		Spectris PLC

Prior Name	Intermediate, If Any	2005
Markal Co.		LA-CO /Markal
Marsh Instruments		Desco Inc.
Marshalltown Instruments		Desco Inc.
Masoneilan	Masoneilan div. of Dresser	Masoneilan div. of Dresser Flow Control
Mass Evolution		Brechbüehler AG
McCANNA	Invensys Flow Control	Flowserve Corporation
MDA Scientific	Zellweger Analytics	Honeywell Inc.
Measurex		Honeywell Inc.
Mercoid		Mercoid div. of Dwyer Instruments Ltd.
Merlin Gerin	Groupe Schneider	Schneider Automation
Micro Motion Inc.		Micro Motion Inc. div. of Emerson Process Measurement
MicroSwitch Inc.	Honeywell MicroSwitch	Pepperl+Fuchs
Milltronics	Siemens Milltronics	Siemens
Modicon (Gould Modicon)	Groupe Schneider	Schneider Automation
Monitek Technologies		Monitek Technologies div. of Martisa Inc.
Monitor Labs		Teledyne Monitor Labs
Montedoro-Whitney Corp.	Badger Meter Ultrasonic div.	Badger Eastech Inc.
Moore Industries Inc.		Moore Industries-International, Inc.
Moore Products Co.	Siemens Moore	Siemens
Moore Products — flow and pressure transmitter line only	MycroSensor Technologies Inc.	Prime Measurement Inc.
MTI		Agilent Technologies
MultiFluid International		Roxar Flow Measurement
NAF	Invensys Flow Control	Flowserve Corporation
NDC Infrared Engineering		Spectris PLC
Nametre Inc.		Nametre div. of Metrisa Inc.
Naval Oy	Invensys Flow Control	Flowserve Corporation
Neotronics		Zellweger Analytics
Neptune Meter Co.	Neptune Schlumberger	Actaris
Neptune-Hersey	Hersey Meter Div., Aaliant	Venture Meas. div. of Danaher
Neslab Instruments Inc.	Thermo Electron	ThermoNeslab
Newport Electronics		Newport div. Omega Eng.
Noble Alloy Valve, Inc. (NAV)		Flowserve Corporation
Nordstrom Audco	Invensys Flow Control	Flowserve Corporation
Norriseal		Norriseal a Dover Resources Company
NovaSensor	Lucas Novasensor	GE Infrastructure Sensing
Osmonics	GE Osmonics	GE Infrastructure Sensing
Pacific Scientific ATC	Pacific Scientific div. of Danaher	Danaher Motion, General Purpose Systems
Panalarm		Panalarm div. of Ametek Inc.
Panametrics Inc	GE Panametrics	GE Infrastructure Sensing
Patriot Sensors Inc.	Ametek Patriot Sensors	Ametek Automation & Process Technology
Penny and Giles Ltd.	TRENDview	Honeywell
PID	Sequencia	Rockwell Automation
PMV	Invensys Flow Control	Flowserve Corporation
Polymetron		Hach div. of Danaher
Powers Process Controls		Moore Industries-International Inc.
Profimatics	Honeywell Inc.	KBC Advanced Technologies plc

Prior Name	Intermediate, If Any	2005
Promac Inc.	Hathaway Beta	Martell Electronics
Red Lion Controls Inc.	Redlion div. Fairey Group	Redlion div. Spectris PLC
Robertshaw Controls		Robertshaw Controls div. of Invensys
Rochester Instrument Systems	RIS	Ametek Inc.
Rosemount Inc.	Fisher-Rosemount Inc.	Emerson Process Measurement
Ruska Inc.	Druck Ruska; GE Ruska	GE Infrastructure Sensing
Saab Level Control	Rosemount Saab	Emerson Process Management
Schmidt Armaturen	Invensys Flow Control	Flowserve Corporation
Seaflow Inc.		Seametrics Inc.
Serck Audco	Invensys Flow Control	Flowserve Corporation
Sereg Vannes		Flowserve Corporation
Servomex		Servomex div. Spectris PLC
Setpoint		Aspen Technology
Setra		Setra div. of Danaher
Sieger	Zellweger Analytics	Honeywell
Siemens Electronic Personal Dosimetry div.	Siemens EPD	Environmental Instruments div.
Sievers Instruments		Ionics Instrument Business Group
Signet Scientific Company	George Fischer Signet	GF Signet Inc.
Smith Meters		Smith Meters a FMC Energy Systems Business
Solartron Analytical		Solartron div. of Roxboro
Square D	Groupe Schneider	Schneider Automation
Steeplechase Software	Schneider	Entivity
Taylor Instruments	Combustion Engineering/Sybron Corp.	ABB Process Automation
TBI	TBI-Bailey; Bailey-Fischer & Porter	ABB Process Automation
Telemechanique	Groupe Schneider	Schneider Automation
Tempil		Tempil div. of Illinois Tool Worka Inc. (ITW)
Tenny Engineering Inc. (bankrupt 2000)	Tenny Environmental	Tenny Environmental div. of Lunaire Ltd.
Texas Nuclear Corp	TN Technologies; ThermoMeasureTech	ThermoElectron Measurement Products
Thermometrics		GE Infrastructure Sensing
Thornton Instruments		Mettler-Toledo Thornton
Tracor	Tracor Westronics	ThermoFinnegan Austin
Trend Instruments		Wika
Triconex	Siebe	Invensys
Tytronics Inc.		Tytronics div. of Metrisa Inc.
UES	Neptune Schlumberger	Liquidated the division — out of business
UFM Vortex meter only		Asahi-America Inc.
Valtek	Durco International	Flowserve Corporation
Varec div. of Emerson	Whessoe Varec	Endress+Hauser
Veeder-Root		Veeder-Root div. of Danaher
Vogt	Invensys Flow Control	Flowserve Corporation
Water Specialties	McCrometer-Water Specialties	McCrometer div. of Danaher
Weksler Instruments Corp.		Weksler Instruments Corp div. of Dresser Industries Instruments
Western Research Div. of Bow Valley Resources Ltd.	BOVAR	Ametek Process Instruments

Prior Name	Intermediate, If Any	2005
Westinghouse Process Control		Westinghouse Process Control div. of Emerson Process Management
Westronics Inc.	Tracor Westronics; ThermoWestronics	ThermoElectron Measurement Systems
Wonderware Inc.	Siebe Wonderware	Wonderware div. of Invensys
Worcester Valve	Invensys Flow Control	Flowserve Corporation
Xomox	Emerson	Crane Corp.
Zellweger Analytics		Honeywell Inc.

A.10 ISA Standards

B. G. LIPTÁK (2005)

The latest standards and related technical publications available from the Instrument Society of America (ISA):

1-55617-531-0 — Standards Library for Measurement and Control (ISA)
12.00.01 — Electrical Apparatus for Use in Hazardous Locations (ISA)
12.01.01 — Electrical Instruments in Hazardous (Classified) Locations (ISA)
12.02.01 — Electrical Apparatus for use in Hazardous Locations (ISA)
12.10 — Area Classification in Hazardous (Classified) Dust Locations (ISA)
12.12 — Nonincendive Electrical Equipment for Hazardous Locations (ISA)
12.13-1 — Performance Requirements, Combustible Gas Detectors (ISA)
12.16.01 — Electrical Apparatus for Use in Hazardous Locations (ISA)
12.22.01 — Electrical Apparatus for Use in Hazardous Locations (ISA)
12.23.01 — Electrical Apparatus for Use in Hazardous Locations (ISA)
12.25.01 — Electrical Apparatus for Use Hazardous Locations (ISA)
12.26.01 — Electrical Apparatus for Use in Hazardous Locations (ISA)
18.1 — Annunciator Sequences and Specifications (ISA)
20 — Process Measurement, Instruments, Primary Elements, Valves (ISA)
26 — Dynamic Response Testing of Process Control Instrumentation (ISA)
37.1 — Electrical Transducer Nomenclature and Terminology (ISA)
37.10 — Piezoelectric Pressure and Sound-Pressure Transducers: Specs (ISA)
37.12 — Potentiometric Displacement Transducers: Specs, Tests (ISA)
37.3 — Strain Gage Pressure Transducers: Specs, Tests (ISA)
37.6 — Potentiometric Pressure Transducers: Specs, Tests (ISA)
37.8 — Specifications and Tests for Strain Gage Force Transducers (ISA)
5.1 — Instrumentation Symbols and Identification (ISA)
5.2 — Binary Logic Diagrams For Process Operations (ISA)
5.3 — Graphic Symbols for Distributed Control/Shared Display Systems (ISA)
5.4 — Instrument Loop Diagrams (ISA)
5.5 — Graphic Symbols for Process Displays (ISA)
50.02.2 — Fieldbus Standard for Use in Industrial Control Systems (ISA)
50.02.3 — Fieldbus Standard for Use in Industrial Control Systems (ISA)
50.02.4 — Fieldbus Standard for Use in Industrial Control Systems (ISA)
50.02.5 — Fieldbus Standard for Use in Industrial Control Systems (ISA)
50.02.6 — Fieldbus Standard for Use in Industrial Control Systems (ISA)
51.1 — Process Instrumentation Terminology (ISA)
67.01 — Transducer and Transmitter Installation for Nuclear Safety (ISA)
67.02.01 — Nuclear Safety Instrument Sensing Line Piping, Tubing (ISA)
67.03 — Light Water Reactor Coolant Pressure Boundary Leak Detection (ISA)
67.04.01 — Setpoints for Nuclear Safety-Related Instrumentation (ISA)
67.06 — Response Time Testing of Nuclear Safety Instrument Channels (ISA)
67.14.01 — Certification of Instrumentation and Control Technicians (ISA)
71.01 — Environmental Conditions Process Measurement, Control Systems (ISA)
71.02 — Environmental Conditions Process Measurement, Control Systems (ISA)
71.03 — Environmental Conditions Process Measurement, Control Systems (ISA)
71.04 — Environmental Conditions Process Measurement, Control Systems (ISA)
72.01 — PROWAY-LAN Industrial Data Highway (ISA)
72.02 — Manufacturing Message Specification: Companion Standard (ISA)
75.01.01 — Flow Equations for Sizing Control Valves (ISA)
75.02 — Control Valve Capacity Test Procedures (ISA)
75.03 — Integral Flanged Globe-Style Control Valve Bodies (ISA)
75.04 — Dimensions for Flangeless Control Valves (ISA)

75.05 — Control Valve Terminology (ISA)
75.11 — Inherent Flow Characteristics, Rangeability of Control Valves (ISA)
75.12 — Socket Weld-, Screwed-End Globe-Style Control Valves (ISA)
75.13 — Method of Evaluating the Performance of Positioners (ISA)
75.14 — Dimensions for Buttweld-End Globe-Style Control Valves (ISA)
75.15 — Dimensions for Buttweld-End Globe-Style Control Valves (ISA)
75.16 — Dimensions for Flanged Globe-Style Control Valve Bodies (ISA)
75.17 — Control Valve Aerodynamic Noise Prediction (ISA)
75.19 — Hydrostatic Testing of Control Valves (ISA)
75.22 — Dimensions for Flanged Globe-Style Angle Control Valve Bodies (ISA)
77.13.01 — Fossil Fuel Power Plant Steam Turbine Bypass System (ISA)
77.20 — Fossil Fuel Power Plant Simulators – Functional Requirements (ISA)
77.41 — Fossil Fuel Power Plant Boiler Combustion Controls (ISA)
77.42.01 — Fossil Fuel Power Plant Feedwater Control System (ISA)
77.43 — Fossil Fuel Power Plant Unit/Plant Demand Development (ISA)
77.44 — Fossil Fuel Plant Steam Temperature Control System (ISA)
77.70 — Fossil Fuel Power Plant Instrument Piping Installation (ISA)
82.02.01 — Safety Standard for Electrical and Electronic Equipment (ISA)
82.02.02 — Safety Requirements for Electrical Equipment (ISA)
82.02.04 — Safety Requirements for Electrical Equipment (ISA)
82.03 — Safety Standard for Electrical and Electronic Equipment (ISA)
84.01 — Safety Instrumented Systems for the Process Industries (ISA)
88.01 — Batch Control: Models and Terminology (ISA)
91.01 — Identification of Emergency Shutdown Systems and Controls (ISA)
92.0.01 — Toxic Gas-Detection Instruments: Hydrogen Sulfide (ISA)
92.02.01-1 — Carbon Monoxide Detection Instruments (ISA)
92.03.01 — Ammonia Detection Instruments (25–500 ppm) (ISA)
92.04.01-1 — Detecting Oxygen-Deficient/Oxygen-Enriched Atmospheres (ISA) 93.00.01 – External Leakage of Manual and Automated On-Off Valves (ISA)
95.00.01 — Enterprise-Control System Integration (ISA)
95.00.02 — Enterprise-Control System Integration (ISA)
REWIC03 — Management of the Development of Critical Computer Systems (ISA) MC96.1 – Temperature Measurement Thermocouples (ISA)
RA8425 — Standards Library for Automation and Control 2003, CD (ISA)
REWIC01 — Programmable Logic Controllers in Safety-Related Systems (ISA)
REWIC02 — Achieiving Safety in Distributed Systems (ISA)
REWIC03 — Development of Critical Computer Systems (ISA)
RP12.13-2 — Combustible Gas Detection Instruments (ISA)
RP12.2.02 — Intrinsic Safety Control Drawings (ISA)
RP12.4 — Pressurized Enclosures (ISA)
RP12.6 — Wiring Practices for Hazardous Locations Instrumentation (ISA)
RP16.1,2,3 — Indicating Variable Area Meters (Rotameters) (ISA)
RP16.4 — Extension-Type Variable Area Meters (Rotameters) (ISA)
RP16.5 — Glass Tube Variable Area Meters (Rotameters) (ISA)
RP16.6 — Calibration of Variable Area Meters (Rotameters) (ISA)
RP2.1 — Manometer Tables (ISA)
RP31.1 — Turbine Flowmeters: Specification, Installation, Calibration (ISA)
RP42.1 — Nomenclature for Instrument Tube Fittings (ISA)
50.1 — Analog Signals for Electronic Industrial Process Instruments (ISA)
RP52.1 — Recommended Environments for Standards Laboratories (ISA)
RP60.1 — Control Center Facilities (ISA)
RP60.11 — Crating, Shipping, and Handling for Control Centers (ISA)
RP60.2 — Control Center Design Guide and Terminology (ISA)
RP60.3 — Human Engineering for Control Centers (ISA)
RP60.4 — Documentation for Control Centers (ISA)
RP60.6 — Nameplates, Labels, and Tags for Control Centers (ISA)
RP60.8 — Electrical Guide for Control Centers (ISA)
RP60.9 — ISA-RP60.9-1981 – Piping Guide for Control Centers (ISA)
RP67.04.02 — Setpoints for Nuclear Safety-Related Instrumentation (ISA)
RP74.01 — Continuous-Belt Weighbridge Scales (ISA)
RP75.21 — Process Data Presentation for Control Valves (ISA)
RP75.23 — Considerations for Evaluating Control Valve Cavitation (ISA)
RP76.0.01 — Analyzer System Inspection and Acceptance (ISA)
RP92.0.02 — Toxic Gas-Detection Instrument: Hydrogen Sulfide (ISA)

RP92.02.02 — Carbon Monoxide Detection Instruments (ISA)

RP92.03.02 — Ammonia Detection: Installation, Operation, Maintenance (ISA)

RP92.06.02 — Chlorine Detection: Installation, Operation, Maintenance (ISA)

TR12.06.01 — Electrical Equipment in Hazardous Locations (ISA)

TR12.13.01 — Flammability of Combustible Gases, Vapors (ISA)

TR12.13.02 — Fire, Explosion Accidents in the Chemical, Mining, Fuel (ISA)

TR12.2 — Intrinsically Safe System Assessment: the Entity Concept (ISA)

TR12.24.01 — Classification of Locations for Electrical Installations (ISA)

TR50.02-9 — Fieldbus Standard for Use in Industrial Control Systems (ISA)

TR67.04.08 — Setpoints for Sequenced Actions (ISA)

TR75.04.01 — Control Valve Position Stability (ISA)

TR77.60.04 — Fossil Fuel Power Plant Human-Machine Interface (ISA)

TR77.81.05 — Interfaces for CEMS Relative Accuracy Test Audit Data (ISA)

TR88.0.03 — Possible Recipe Procedure Presentation Formats (ISA)

TR92.06.03 — Feasibility of Chlorine Detection Instrument Testing (ISA)